The Development of Forensic DNA Techniques to Assist with Criminal Investigations involving the Australasian Carpet Python (*Morelia spilota*)

by

Sherryn A. Ciavaglia

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy
College of Science and Engineering
May 2019

TABLE OF CONTENTS

Thesis abstract

The analysis of non-human DNA to provide evidence in criminal cases involving wildlife species is a relatively young area of forensic science. Forensic genetic tests have been developed for only a small handful of the plethora of species persecuted and exploited, despite their statutory protection. Poor resourcing is ever an issue for wildlife crime, which most often falls below crimes against humans in the financial priorities of governments. Forensic DNA profiling techniques are resource intensive to develop, so the wildlife forensic scientist is continually working to develop and validate new techniques and expand their capacity for forensic testing between casework.

Furthermore, wildlife species bring additional challenges not met by forensic scientists specialising in human DNA. Species identification is peculiar to wildlife forensic science and the elevated genetic structure and fragmented populations and low sample sizes encountered when validating individualisation tests for wildlife species provide difficulties of interpretation and application unique to this field. This PhD thesis has progressed the field of wildlife forensic science by providing novel phylogenetic and population genetic analyses of Australasian python species commonly encountered in wildlife crime, and in particular one species commonly encountered by Australian enforcement officials, the carpet python, *Morelia spilota*. Forensic tests for both species identification and individualisation of carpet pythons are developed and validated for direct application to casework involving this species.

This thesis explores the arrangement and sequence level variability of mitochondrial genomes in a wide range of snake species to identify a locus with characteristics desirable for species identification of individuals belonging to Pythonidae. Species identification by DNA requires priming sites located in DNA regions highly conserved across snakes taxa to facilitate wide amplification utility. The amplified region ideally represents a 'Goldilocks' sequence, containing a fine balance of sequence variability between species and conservation within a species. Such a locus is identified to separate all eleven species within the genus *Morelia,* excluding *M. bredli*. However, mitochondrial and nuclear DNA analyses within this thesis suggest that the species differentiation of *M. spilota* and *M. bredli* is not reflected at a genetic level. In the context of reporting species identification, any match to either of these species should be reported as *M. spilota* or *M. bredli*. A more conservative approach would be to report to genus rather than species, however the former approach is far more useful and this decision falls to the comfort of the reporting officer. Furthermore, this project has identified a small target region for species identification of *Morelia*, suitable for application to degraded forensic remains. A developmental validation of the resulting assay is provided to facilitate direct application to criminal casework involving python species. Mock case samples demonstrate the utility of this test on unknown samples suspected to be carpet pythons.

The primary focus of this thesis is the development and validation of a novel STR test to individualise carpet pythons. This DNA test aids investigations involving matching of evidentiary items, parentage testing and ultimately also repatriation of a poached python to the wild through identification of the most likely population of origin, which also facilitates concentration of enforcement efforts on identified poaching 'hotspots'. New STR markers are identified with characteristics desirable for forensic application. A set of 24 tetra- and penta-nucleotide markers are optimised into three 11-plex multiplex assays with inbuilt mechanisms to maintain high quality assurance. Allelic ladders accompany the three assays for robust single base-pair genotyping and inter-laboratory utility.

Sequence level characteristics of the STR loci are explored and certain features are identified that can provide information subsequent to genotyping by size migration, where case scenarios involve *M. bredli* or *M. spilota imbricata.* The applicability of these STR markers to other Australasian python species of forensic significance is also explored.

A DNA reference database of carpet pythons representative of their native range is presented to facilitate the calculation of statistical conclusions. However, the complete database violates assumptions made when calculating the forensic statistics ultimately reported in casework. Population genetic analyses are conducted to identify more appropriate sub-populations with which to calculate forensic statistics. The dataset offers a number of limitations that confound straight-forward reporting of the forensic statistics, caused by small sample sizes, sparse sampling across large geographic areas and high genetic structuring within sub-populations. Approaches to application involving varying degrees of conservativism are presented, but the best approach to apply will need to be assessed on a case by case basis, depending on the scenario and population involved.

The overall reference DNA dataset is divided into six sub-populations that are more appropriate than the total database for calculating forensic statistics. Samples unsuitable for inclusion in the reference database (for example, missing data at >2 loci) are used to test the effectiveness of geographic assignment to the six sub-populations and the resulting population specific statistical calculations. The resulting statistics offer strong discrimination power in all sub-populations, even when samples in poor condition yield DNA profiles of around half a dozen loci. Geographic assignment to the six populations is also effective even when less than half the total loci have been obtained. Only one interface between two populations showed limitations for assignment to sub-population. A limited developmental validation of the STR assays lays the foundation for application in criminal casework involving the Australasian carpet python.

Author declaration

The research presented in this doctoral thesis is my own original work. I certify that, to the best of my knowledge and belief, it does not include any content previously published or written by any other person, unless acknowledged or specified in the publication author declarations. The work within this thesis has been developed by myself and facilitated by discussions with colleagues in wildlife forensic science and human forensic genetics as elaborated upon in the acknowledgements section. I am the principal author and contributor to all publications included in this work, except the book chapter that was authored by Prof. Adrian Linacre and contributed to by myself. No part of this thesis has been submitted for any previous degree.

Author declaration regarding published works

With co-authors in various locations around the world, it is impractical to have each personally sign a declaration. The following is my declaration of the contribution they made to the submission of our manuscripts:

Adrian Linacre contributed to the experimental design, elaboration of ideas and made comments towards the final manuscript of all publications included in this thesis, except for Linacre & Ciavaglia, 2018 for which he was first author of the book chapter.

Julianne Henry contributed to the experimental design of the mitochondrial assays and made comments towards the final manuscript of the publications on which she is included as an author.

Shanan Tobe gave technical advice about the SYBR Green real time PCR assay, genetic analysis software, in particular Mr. Bayes and made comments towards the final manuscript of the publications on which he is included as an author.

Stephen Donnellan contributed to discussions about the overall aims of the PhD project, offered advice about the experimental design of the mitochondrial and STR assays, provided samples for the project and made comments towards the final manuscript of the publications on which he is included as an author.

K. Paul Kirkbride offered advice about the confiscation of assets and made comments towards the final manuscript of the publication on which he is include as an author.

Hannah Dridan contributed ideas and case studies and made comments towards the final manuscript of the publications on which she is included as an author.

Sherryn A. Ciavaglia

2018

Author contributions per manuscript

Manuscript: Sherryn Ciavaglia, Julianne Henry, Adrian Linacre, Profiling pythons to combat common illegal wildlife activities, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e31–e32

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	60%	100%	70%	100%	75%
J. Henry	10%	-	10%	-	10%
A. Linacre	30%	-	20%	-	15%

Manuscript: Sherryn Ciavaglia, Stephen Donnellan, Julianne Henry, Adrian Linacre, Species identification of protected carpet pythons suitable for degraded forensic samples, Forensic Sci., Med. Pathol. (2014) 10:295-305

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	40%	100%	50%	100%	45%
S. Donnellan	15%	-	25%	-	25%
J. Henry	5%	-	10%	-	15%
A. Linacre	40%	-	15%	-	15%

Manuscript: Sherryn Ciavaglia, Shanan Tobe, Stephen Donnellan, Julianne Henry, Adrian Linacre, Molecular identification of python species: Development and validation of a novel assay for forensic investigations, Forensic Sci. Int. Genet. (2015) 16:64-70

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	55%	100%	55%	100%	40%
S. Tobe	5%	-	10%	-	15%
S. Donnellan	5%	-	5%	-	15%
J. Henry	5%	-	10%	-	5%
A. Linacre	30%	-	20%	-	25%

Manuscript: Sherryn Ciavaglia, Hannah Dridan, K. Paul Kirkbride, Adrian Linacre, Current Issues with the Investigation of Wildlife Crime in Australia: Problems and Opportunities for Improvement, J. Int. Wildl. Law Policy (2015) 18: 244-263.

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	40%	100%	68%	100%	55%
H. Dridan	30%	-	10%	-	10%
KP. Kirkbride	-	-	2%	-	10%
A. Linacre	30%	-	20%	-	25%

Manuscript: Sherryn Ciavaglia, Stephen Donnellan, Shanan Tobe, Julianne Henry and Adrian Linacre, A novel forensic DNA profiling technique for protected species, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e258–e260

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	65%	100%	75%	100%	84%
S. Donnellan	5%	-	5%	-	-
S. Tobe	5%	-	5%	-	2%
J. Henry	5%	-	-	-	2%
A. Linacre	20%	-	15%	-	10%

Manuscript: Sherryn Ciavaglia, Hannah Dridan and Adrian Linacre, Getting more for less: can forensic tools for Australian wildlife enforcement support international compliance efforts? Australian Journal of Forensic Sciences, (2017) DOI: 10.1080/00450618.2017.1384060

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	65%	100%	80%	100%	60%
H. Dridan	10%	-	-	-	5%
A. Linacre	25%	-	20%	-	35%

Manuscript: Sherryn Ciavaglia, Adrian Linacre, OzPythonPlex: An optimised and forensically validated STR multiplex assay set for the Australasian carpet python (Morelia spilota), Forensic Sci. Int. Genet. (2018) 34:231-248

	Conception and experimental design	Performed laboratory work	Results analysis & interpretation	Manuscript preparation	Manuscript editing
S. Ciavaglia	60%	100%	65%	100%	65%
A. Linacre	40%	-	35%	-	35%

Acknowledgements

I dedicate this PhD thesis to my dad, Richard Ciavaglia, as well as the previous director of Forensic Science SA, Dr. Ross Vining who was very supportive of my desire to start a PhD, and without whom the project would not have happened. I wish that both of these great men were here to see its completion.

Stephen Donnellan helped to get the project underway and invaluably provided the samples to make the project possible. He was a great source of advice and wealth of knowledge about snake systematics and genetics. I am grateful for the time that he gave for discussions and the loan of various out of print textbooks.

My sincerest thank you for the support that Lucy Webster, David Kenyon and the Senior Management Team at Science and Advice for Scottish Agriculture (SASA) gave, allowing me to complete some of my data analysis and thesis drafting in work time. A thesis of this size could not possibly have happened in the required timeframe without this assistance.

Duncan Taylor was another major contributor of support and voice of reason in the face of difficult population genetics. Despite publishing at least 45 peer-reviewed papers, most of which he first-authored, co-authoring one textbook and revolutionising the field of forensic genetic statistics by the development and continual update of STRmix over the course of my PhD project, Duncan still found time in his life to respond to my queries and offer assistance regarding snake population genetics and drawing statistical conclusions.

With similar availability to Duncan, Rob Ogden graciously allowed me some of his scant time between international trips to discuss population genetic analyses. He offered particularly helpful discussion around the various approaches to dealing with the difficult scenarios in my dataset and how different measures of conservativism can enable some form of forensic statistical results to be presented to the courts.

Shane Tobe, Lucy Webster and Gillian Murray-Dickson offered technical advice about wildlife forensic techniques and population genetic analyses. John Butler offered advice about STR multiplex development. I thank him for his genuine interest in my project. Mike Gardner provided assistance in the initial MPS component of my project. Terry Bertozzi and Leanne Wheaton offered assistance in obtaining mitochondrial DNA sequence from the MPS data and getting the MRT screening assays working, respectively. Nicole White and Mike Bunce at Murdoch University offered redesigned primer sequences for some of the MS primer sets and advice about these markers and M13 tagging. Paul Kirkbride provided encouragement in the short time that we worked together at Flinders University and hereafter.

I received a great deal of technical support at Forensic Science SA from Julianne Henry, Chris Hefford, Alice Stephenson, Denise Ward and Bridget Wright. Julianne Henry also helped with the planning of the PhD thesis overall, improving aspects of experimental design, quality assurance and offering input towards the discussion of the mitochondrial DNA publications. Beatrice Goh from Applied Biosystems/Life Technologies (now ThermoFisher Scientific) also provided much needed technical support regarding the creation, import and correct formatting of bin sets and marker sets in Genemapper-ID. I had practical assistance from Ine van Velzen, an intern working with us in the Flinders Forensic DNA Laboratory for a few months, who assisted me with laboratory work involving DNA extraction, PCR amplification and sample preparation for Sanger sequencing that contributed to the species identification assay data.

Hannah Dridan was a continual source of support, enthusiasm and energy, possessing enviable knowledge and experience about Australian wildlife enforcement and prosecutions, of which she is

always only too happy to share. She provided an excellent source of motivation for getting the techniques developed. Elaine Kellett deserves special mention for offering encouragement and proof-reading various documents. Hannah and Elaine's friendships have been particularly special assets to come out of this PhD project.

Other enforcement officers who are/have been a great asset to wildlife enforcement in Australia and also contributed ideas and experiences to my own work are: Jason Bentley, Luke Bond, Mike Devery, Brett Easton, Sally Heaton, David Mell, Samantha Moore, Mike Sverns, Brad Wade, and, in particular, Steven James.

Funding sources that enabled the work to be done are: the Attorney General's Department, South Australian Government and Science and Advice for Scottish Agriculture, Scottish Government (for work time). Small grants were obtained from the Nature Foundation of South Australia, the Field Naturalist Society of South Australia Lirabenda Endowment Fund and the Department of Environment, Water and Natural Resources, South Australian Government. Flinders University and the Australian and New Zealand Forensic Science Society (both Executive and SA branch) funded a number of travel scholarships to allow me to present aspects of my work at international and national conferences, respectively.

My mother-in-law Clare Harvey has been here before (here being the world of a part-time PhD while working full-time) and was a welcome ear for lamentation and source of ongoing support and encouragement. My mum has likewise shared the oral burden of my PhD for many years and deserves recognition for her ongoing encouragement and faith in my abilities. My husband Dale has been an unwavering source of moral and programming support, writing Excel macros and pieces of code here and there to create snake genome maps and geographic maps of snake locations. If a marriage can survive a PhD candidature, then I imagine it can survive anything.

Lastly, my Honours project taught me the difficultly of completing a thesis with a very unsupportive supervisor and, to my fortune, my PhD has taught me the exact opposite; that is, how much a student can achieve when they are encouraged and motivated by a supervisor that offers unfaltering support and reassurance. In seven long years, Adrian has never hassled me to have a draft completed, never pressured me to produce results by a particular deadline and trusted in me to get what was necessary done in the required timeframe. For that I am eternally grateful and I hope I have gone above and beyond expectations in terms of what this project has achieved. More than any other person I've ever work with, he has constantly made himself available for advice and consistently turned around my drafts and queries in incredibly quick timeframes. His generosity has allowed me to present my work at multiple international conferences. He has been the primary figure responsible for changing the direction of my career and indeed my life – being able to work in a field for which I am passionate in the beautiful city of Edinburgh. For this, and for being the most ideal PhD supervisor that any student could ever wish for, I will always owe him a great debt of gratitude.

List of Abbreviations

µg	microgram
µL	microlitre
µM	micromolar
° C	degrees Celcius
θ	also coancestry coefficient, correction for degree of population substructure, relates directly to F_{ST}
ABTC	Australian Biological Tissue Collection
AGRF	Australian Genome Research Facility
ANZFSS	Australian and New Zealand Forensic Science Society
AQIS	Australian Quarantine Inspections Service
ASPE	Allele Specific Primer Extension
ATP	Adenosine Triphosphate
bp	Base pair
CITES	Convention on International Trade in Endangered Species of flora and fauna
COI	cytochrome oxidase c 1
COII	cytochrome oxidase c 2
COIII	cytochrome oxidase c 3
cyt b	cytochrome b
DNA	Deoxyribonucleic Acid
dNTP	dinucleotide triphosphate
EPBCA	Environmental Protection and Biodiversity Conservation Act
fg	femtogram
FSSA	Forensic Science SA
F_{ST}	measure of population substructure, equates to Θ in the forensic context when mating within subpopulations is random
H_e	Expected heterozygosity
HMW	High molecular weight

H_o	Observed heterozygosity
HWE	Hardy-Weinberg Equilibrium
IBD	Isolation By Distance
ISFG	International Society for Forensic Genetics
K	number of populations (terminology specific to STRUCTURE software package)
LD	Linkage Disequilibrium
LMW	Low molecular weight
Ln	natural logarithm
MCMC	Markov Chain Monte Carlo
mg	milligram
$MgCl_2$	Magnesium Chloride
mL	millilitre
mM	millimolar
MPS	Massively Parallel Sequencing
MRT	Multiplex Ready Technology
MS	*Morelia spilota*
MsF	*Morelia spilota* forensic
Msi GI	*Morelia spilota imbricata* Garden Island
Msi ML	*Morelia spilota imbricata* Mainland
Msi SFI	*Morelia spilota imbricata* Saint Francis Island
mtDNA	mitochondrial DNA
ncDNA	nuclear DNA
ND	NADH dehydrogenase subunit gene
NEC	North Eastern Coastal
ng	nanogram
nM	nanomolar
PCoA	Principal Co-ordinate Analysis

PCR	Polymerase Chain Reaction
pg	picogram
PIC	Polymorphic Information Content
pM	pico Molar
QA	Quality Assurance
RFU	Relative Fluorescent Unit
SEC	South Eastern Coastal
SIB	Southern Internal and *Bredli*
SMM	Stepwise Mutation Model
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
SWFS	Society for Wildlife Forensic Science
SWGDAM	Scientific Working Group on DNA Analysis Methods
SWGWILD	SWFS Scientific Working Group for Wildlife Forensic Sciences
TEAM	Traditional East Asian Medicine
UNODC	United Nations Office on Drugs and Crime
VNTR	Variable Number of Tandem Repeats

List of Figures

List of Tables

List of Appendices

Achievements

Peer reviewed publications

1. **Sherryn Ciavaglia**, Julianne Henry, Adrian Linacre, Profiling pythons to combat common illegal wildlife activities, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e31–e32.
2. **Sherryn Ciavaglia**, Stephen Donnellan, Julianne Henry, Adrian Linacre, Species identification of protected carpet pythons suitable for degraded forensic samples, Forensic Sci., Med. Pathol. (2014) 10:295-305
3. **Sherryn Ciavaglia**, Shanan Tobe, Stephen Donnellan, Julianne Henry, Adrian Linacre, Molecular identification of python species: Development and validation of a novel assay for forensic investigations, Forensic Sci. Int. Genet. (2015) 16:64-70
4. **Sherryn Ciavaglia**, Hannah Dridan, K. Paul Kirkbride, Adrian Linacre, Current Issues with the Investigation of Wildlife Crime in Australia: Problems and Opportunities for Improvement, J. Int. Wildl. Law Policy (2015) 18: 244-263.
5. **Sherryn Ciavaglia**, Stephen Donnellan, Shanan Tobe, Julianne Henry and Adrian Linacre, A novel forensic DNA profiling technique for protected species, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e258–e260
6. **Sherryn Ciavaglia**, Hannah Dridan and Adrian Linacre, Getting more for less: can forensic tools for Australian wildlife enforcement support international compliance efforts? Australian Journal of Forensic Sciences, (2017) DOI: 10.1080/00450618.2017.1384060
7. **Sherryn Ciavaglia**, Adrian Linacre, OzPythonPlex: An optimised and forensically validated STR multiplex assay set for the Australasian carpet python (*Morelia spilota*), Forensic Sci. Int. Genet. (2018) 34:231-248

Book Chapter

Adrian Linacre, **Sherryn Ciavaglia**, Chapter 18: Wildlife Forensic Science. In, Amorim A. & B. Budowle (2016) Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation, Imperial College Press.

Successful grant applications

- Nature Foundation South Australia, October 2012, 1 year duration, $2660
- Department of Environment, Water and Natural Resources, June 2013, 6 months duration, $800
- Field Naturalist Society of South Australia Lirabenda Endowment Fund, March 2014, 1 year duration, $2760

Scholarships

- ANZFSS National Travel Award, awarded by ANZFSS Executive Group to attend ANZFSS 2012 conference in Hobart. Full registration fee ($1150) covered. Accommodation funding ($600) awarded by Forensic Science SA, part-time employer of SC.
- ANZFSS National Travel Award, awarded by ANZFSS Executive Group to attend ANZFSS 2014 conference in Adelaide. Full registration fee ($1250) covered.

- Flinders University Semester 2 Travel Scholarship, awarded by Flinders University for attendance at Society of Wildlife Forensic Science meeting, Missoula, Montana, June 22-26, 2015. $1500

- Flinders University Semester 1 Travel Scholarship, awarded by Flinders University for attendance at International Society of Forensic Genetics meeting in Krakow, Poland, August 31- September 5, 2015. $1500

- Flinders University Research Student Conference Travel Grant, awarded by Flinders University to attend the International Society of Forensic Genetics meeting in Krakow, Poland, August 31- September 5, 2015. $2200 awarded to contribute to flights.

Presentations

Oral presentations

- Ciavaglia, Donnellan & Linacre (2012) Reptile wildlife crime in Australia: local problems, global solutions? Society for Wildlife Forensic Science 1st triennial symposium, Jackson Hole, Wyoming, USA, 22-25 May

- Ciavaglia, Henry & Linacre (2012) Applications of DNA technology to wildlife enforcement involving python species, 21st International ANZFSS Symposium, Hobart, Australia, 23-27 September

    - Joint Winner of Best Oral Presentation in Wildlife Forensics, Entomology and Botany Session

- Ciavaglia, Dridan & Linacre (2013) Utilising scientific tools to enhance regulation and support successful legal outcomes; applications using python DNA, Australasian Environmental Law Enforcement and Regulators Network (AELERT) conference, Melbourne, Australia, 13-14 November

- Ciavaglia, Henry & Linacre (2014) A novel forensic STR profiling assay to aid investigations of criminal activities involving carpet pythons, 22nd International ANZFSS Symposium, Adelaide, Australia, 31 August – 4 September

- Ciavaglia, Tobe, Donnellan, Henry & Linacre (2015) Development and validation of a forensic species identification assay for Australasian python species, 3rd Meeting of the Society for Wildlife Forensic Science, Missoula, Montana, 22-26 June

- Ciavaglia, Tobe & Linacre (2015) A DNA profiling assay for carpet pythons, 3rd Meeting of the Society for Wildlife Forensic Science, Missoula, Montana, 22-26 June

- Ciavaglia, Donnellan, Tobe, Henry & Linacre (2016) A novel forensic STR profiling assay to investigate allegations of false paternity and the associated poaching of native carpet pythons, 23nd International ANZFSS Symposium, Auckland, New Zealand, 19-23 September

- Ciavaglia, Tobe & Linacre (2017) The perplexity of the carpet python, 4th Meeting of the Society for Wildlife Forensic Science, Edinburgh, Scotland, 5-9 June

- 3 annual lectures on Wildlife Crime for the Graduate Certificate in Environmental Compliance coursee at TAFE SA, 2012, 2013, 2014.

Poster presentations

- Ciavaglia, Henry & Linacre (2013) Profiling pythons to combat common illegal wildlife activities, 25th World Congress of the International Society of Forensic Genetics, Melbourne, Australia, 2-7 September

- Ciavaglia, Dridan & Linacre (2013) Getting more for less: Can forensic tools for national problems aid international compliance efforts? AELERT conference, Melbourne, Australia, 13-14 November

- Ciavaglia, Donnellan, Tobe, Henry & Linacre (2015) A novel forensic DNA profiling technique for protected species, 26th World Congress of the International Society of Forensic Genetics, Krakow, Poland, 31 August – 5 September

Casework

The following is a Statement of Witness from a case for which I completed the laboratory work:

STR marker assessment of Boa constrictor samples for Northern Territory Parks and Wildlife Commission

Content has been removed due to confidentiality reasons.

Content has been removed due to confidentiality reasons.

Content has been removed due to confidentiality reasons.

Content has been removed due to confidentiality reasons.

CHAPTER 1

Introduction

Australian Snakes in Illegal Activities and Relevant Wildlife Forensic DNA Techniques

Appendix included in Chapter 1:

Appendix 1 Publication - Sherryn Ciavaglia, Hannah Dridan, K. Paul Kirkbride, Adrian Linacre, Current Issues with the Investigation of Wildlife Crime in Australia: Problems and Opportunities for Improvement, J. Int. Wildl. Law Policy (2015) 18: 244-263.

1.1. Protecting Australia's unique species

1.1.1 The importance of the Australian biota

Australia has had a long history of isolation from the other major continents, stretching back to the breaking up of the supercontinent Gondwana, around 180 to 130 Mya [1]. This has caused the diversification of unique biota and over 80% of flora and fauna are found only in Australia [2]. The vast majority of our herpetofaunal complement is unique in the world, as 93% of all Australian amphibian and 89% of all Australian reptile species are endemic [2]. The South-West corner of Australia constitutes one of the world's 25 biodiversity hotspots [3].This range of unusual creatures causes great interest in Australian species from fauna and flora collectors worldwide [4] and has led to a considerable black market international trade in Australian species.

At a national and state level in Australia, there are many opportunities for wildlife crime to occur. These include the regulation of the local trade in native fauna; the keeping, breeding and selling is regulated at different levels for different species and can lead to illegal take from the wild. Section 1.6 provides a review of the issues that Australian wildlife enforcement officers face and presents suggestions to mitigate some of these issues. Essentially, Australian wildlife is a unique and valuable asset to the country that warrants protection.

1.1.2 Forensic science and the Australian legal system

Forensic science is the application of scientific techniques to aid legal investigations. Historically, it has been a reactive science, borrowing techniques that have been developed in other sciences and applying these to answer judicial questions, rather than a proactive science that invests time and effort in developing new techniques (although novel forensic development has become much more commonplace). Australian judicial proceedings follow the adversarial system, which involves a prosecution and defence council presenting alternative cases in relation to particular facts in issue [5,6]. These facts in issue pertain to one or more allegations of breach in legislation. So, forensic science must be conducted in response to the relevant legislation. Legislation varies between jurisdictions and also over time, so a forensic question that is relevant to one case will not necessarily be relevant to other investigations.

1.1.3 Wildlife forensic science

A specialised area of forensic science has emerged in recent years specifically for the investigations of crimes involving wildlife. The term wildlife embodies both flora and fauna, although fauna and especially the class Mammalia are most frequently publicised, reported and subsequently investigated, due to the public popularity of high profile 'flagship' species. This polarity in subject matter does not mirror the affected victims of wildlife crime.

Wildlife forensic investigations have gained momentum in recent years due to changing attitudes towards the sustainability of environmental resources. The involvement of wildlife in breaches of legislation can occur in a number of forms: where the animal is the suspect such as dog bite; where the animal is an accomplice to a crime, such as deliberate poisoning via lizard derivatives or the use of hounds in badger baiting; or the animal might be associated with a crime e.g. [7,8]. The most common form of wildlife offence (though not mostly highly publicised) are those where the animal is the victim.

1.1.4 Relevant Australian Legislation

Affairs in Australia are governed by both federal and state legislation, the federal legislation overruling state legislation if any conflict is encountered between these. International trade in endangered species and the control of international borders and ports are dealt with in the federal jurisdiction.

The Washington Convention on International Trade in Endangered Species of Fauna and Flora (CITES) is an international agreement that provides recommendations for legislation in order to regulate the use of wildlife in trade and promote the sustainability of wildlife resources (www.cites.org). As a signatory to CITES, Australia is obliged to incorporate the guidelines into federal legislation and provide the appropriate enforcement of this legislation.

The Environmental Protection & Biodiversity Conservation Act (EPBCA) 1999 is Australia's primary piece of federal legislation dealing with environmental and wildlife crime. Laws pertaining to CITES can be found in Section 13 – Species and Communities. This legislation deals with import and export of internationally endangered (CITES listed) species, export of native species and import of live plants and animals and all other environmental matters of national significance.

Divisions 1 & 2 of the EPBCA 1999 outline the processes of nomination and listing, the permit system and protection of critical habitat for threated species and ecological communities, and also migratory species, respectively. Division 3 outlines offences against whales and cetaceans and laws regarding permits. The statutes for protected marine habitats including the Australian Whale Sanctuary are also included. Division 4 concludes with laws regarding other marine species and permit systems. The above laws cover the taking of species from the wild, harm to species in the wild and any trade involving these species. Australia also imposes stricter domestic measures on some species than that which would be considered the minimum by CITES. This means that certain species listed in CITES Appendix II are treated by authorities as if they were Appendix I listed species [9].

Another key national legislation document concerning wildlife is the Customs Act 1901. This is monitored and enforced primarily by the Australian Customs and Border Protection Service and deals with wildlife crossing the international border of Australia.

The Australian Quarantine and Inspection Service (AQIS) administers a number of legislative Acts. The Quarantine Act 1908 governs the import of wildlife into Australia in order to control biosecurity and prevent the entry of infectious diseases in live imports. The Export Control Act 1982 contains numerous regulations regarding the export of Australian wildlife and wildlife derivatives.

The Therapeutic Goods Act 1989 applies to both the export of therapeutic goods and use of therapeutic goods within the country. An Australian Register of Therapeutic Goods exists on which goods must be listed before they can be imported. The import of traditional medicines would be regulated by this piece of legislation, among others (e.g. EPBCA 1999 for CITES listed species).

For more information on national legislation regarding wildlife trade, refer to: http://www.environment.gov.au/biodiversity/wildlife-trade/legislation/index.html (accessed 03/06/2018).

Each state and territory also has a list of species that are and are not allowed to be traded across borders, as well as other state-specific legislation. South Australian legislation relevant to wildlife includes the Animal Welfare Act, SA 1985; Animal Welfare Regulations, SA 2000; Dog and Cat Management Act, SA 1995; Livestock Act, SA 1997; Marine Parks Act, SA 2007; Mining Act, SA 1971; National Parks and Wildlife Act, SA 1972; Native Vegetation Act, SA 1991; Pastoral Land Management and Conservation Act, SA 1989 and Wilderness Protection Act, SA 1992.

1.1.5 The nature of wildlife forensic investigations – how do these differ to human investigations?

Wildlife forensic science has become a forensic discipline in its own right much more recently than human forensic science. However, it has benefited from the advances made in human forensic science to develop much more rapidly, despite having a much smaller community of practitioners. The biology of the respective victims is the same (see Section 1.2: Forensic DNA techniques), so techniques applied to human forensic investigation can also be applied to wildlife forensic investigation. Depending on the question being answered though, wildlife forensic work requires a degree of specialisation beyond the human field and any human forensic practitioners interested to investigate wildlife crimes must be aware of differences between the two (discussed in the following section).

One major distinction between human and wildlife forensic investigations is that most wildlife crime involves a silent victim. While industries, such as the legal wildlife trade and legitimate breeding programs can suffer loss as a result of illegal activity, the animal or plant almost certainly suffers direct hardship from illegal activity. Especially in cases of trade, the welfare of the animal goes without consideration. For example, two personal suitcases were intercepted at Mopah Airport in Merauke, Papua New Guinea containing 1495 pig-nose turtles that were most likely destined for the pet trade or the dinner table [10].

The animal or plant cannot report the offence and so wildlife crime relies on report or detection by a secondary party. The majority of wildlife crime being conducted is probably never detected. For this reason, it is impossible to quantify the amount of wildlife crime that actually occurs, and the subsequent cost to governments and the public. A recent estimate has placed this range between US$7 billion and US$23 billion per annum [11] (in [12]). As most crime goes undetected, estimates based directly on the level of criminal activity identified are likely to be vast underestimates.

The victimless nature of the crime is largely responsible for lacking resources allocated to wildlife crime. If the crime is not investigated, then there will be no retaliation on the part of the animal or plant. A higher value is placed on the human life than the life of an animal or plant by most people

and so resources will always be allocated to fighting crimes against humans before they are allocated to crimes against wildlife. Countries often affected the most by wildlife crime due to their tropical climate and resulting high compliment of species endemism by and large are third world countries with struggling economies, some rife with problems of governmental corruption and civil war. Sixteen of the world's 25 biodiversity hotspots (categorised by species endemism) are located in developing countries where threats to biodiversity are the greatest and resources for conservation and environmental enforcement most limited [3]. The human forensic work in these countries is severely underfunded and many crimes against humans go without investigation. What chance then is there of justice for the wildlife victimised in these countries?

While beyond the scope of this discussion, a mention should be made of an ethical problem confounding wildlife enforcement in developing countries. Often, wildlife exploitation is the sole form of income available to citizens living below the poverty line and a balance must be struck between conservation interests and the allowance of some exploitation for those whose lives rely on farming or trade of wildlife for income [13,14].

Yet, much wildlife crime in third world countries is conducted by wealthy individuals or organisations and motivated by greed. The organised nature of illegal trade is another similarity between crimes seen in human and wildlife forensic investigations [12]. While recognised in general for at least twenty years [15], the ties between wildlife crime and organised crime syndicates have become increasingly apparent to the law enforcement community in more recent years [16].

The circumstances surrounding wildlife smuggling often suggest that criminal networks must be in operation [17]. Take for example the arrest of a Czech national caught while trying to smuggle almost 250 animals out of Buenos Aires in December 2011 [18]. The animals were stowed in the man's luggage, which officials noticed was moving during security screening and included snakes, snails, spiders, lizards and insects, some of which are CITES listed. All animals were native to Argentina or surrounding countries, yet the man's travel records show he had entered Argentina just a few days prior. This has drawn authorities to suspect that the wildlife had been collected by other individuals acting in an organised animal smuggling syndicate and the Czech man was acting as courier between groups.

Illegal wildlife trade is suggested to follow similar patterns to the trade of narcotics and arms. Often these trades are conducted simultaneously by organisations. Wildlife trade is enticing to criminals for a number of reasons. It can result in high returns, yet the risk of getting caught is far lower than that of the narcotics trade. The risk of prosecution of those caught is minimal [16]. The resulting penalties of prosecutions that have taken place are far lower than the expected value of the trafficked stock on the black market [2], leading questions to be raised over the effectiveness of the relatively lenient sentences actually handed to deter further criminal activity [19].

With the south western corner of Australia home to one of the world's 25 biodiversity hotspots[3], where 80% of mammal, reptile and flowering plant species from this region are endemic [20], there is much reason for interest in Australian biota by the international pet trade. Australia is not exempt from organised criminal activity and illegal wildlife trade operations have been suggested to link with much larger international organisations [2].

1.1.6 Applying scientific techniques in a forensic context

The issues surrounding the straddling of wildlife forensic science between the human forensic field and the conservation biology laboratories where many of the techniques applied to wildlife forensic cases are developed have been discussed elsewhere [16,21]. One extremely important attribute that separates wildlife forensic science from science conducted in a standard university laboratory is the application of strict Quality Assurance procedures and the conclusion of assay development with a validation study. Forensic science is a process ultimately conducted as part of the criminal justice system and must be undertaken in accordance with judicial proceedings. The process of validation involves conducting a number of tests on a new technique to establish the capabilities and limitations of the test under the given set of conditions. This process is used to demonstrate that the technique is robust, can reliably achieve its intended purpose and is ultimately fit for application to judicial proceedings.

The validation procedure for forensic genetic investigations has previously been described for human work by the Scientific Working Group on DNA Analysis Methods (SWGDAM) [22] and the most recent edition of the Forensic Science Regulator (UK) Codes of Practice and Conduct [23] provides extensive detail about what a validation conducted by any forensic practitioner should involve. As evidence will be presented under the same level of scrutiny, non-human forensic work should follow the same validation guidelines as human forensic work.

Routine validation tests include:

Robustness: Ability of the technique to perform in an identical manner under varying circumstance.

Repeatability: Consistency of the technique to produce a result for the target substrate every time. This includes the precision of the test (ability to produce an identical result every time, such as the same basepair sizing) and accuracy of the test (ability to produce an accurate result every time, such the true basepair sizing).

Reproducibility: Replication of the technique between independent laboratories to yield the same result.

Specificity: The ability of the technique to produce a result for the target substrate of interest (e.g. species) and any other substrate (e.g. closely related species) has been documented. Note: amplification result from a closely related species does not necessarily negate the usefulness of the test, but must be recorded and disclosed as part of the validation study.

Sensitivity: Threshold amount of starting material below which the technique will not give a result has been determined, in addition to the range of product amount that will yield a reliable result.

Stability: Effects of environmental fluctuations on the results of the techniques have been documented. This step might also include the application of test materials on different substrates and when subjected to various environmental insults in order to detect whether post-depositional changes to the substrate have any effect on the technique.

Selectivity: In a biological sense, the same result is obtained from different tissues of origin, such as blood, hair, bone, scale, liver, etc.

Some form of peer review, usually through publication in a peer reviewed journal is not essential to validation, but would certainly increase the admissibility of the technique in court. The Frye standard, superseded by the Daubert standard adopted in many states in the US legal system, dictate that a scientific technique is admissible in court only if it has already been generally accepted by the relevant scientific community. Australian common law references a similar prescription, provided by King in *The Queen v. Bonython* (1984) 38 SASR 45, that a judge's assessment of expert testimony must consider whether it forms part of "*a body of knowledge….sufficiently organised or recognised to be accepted as a reliable body of knowledge*". Peer review publication serves to demonstrate this acceptance of the technique in question.

Other recommendations for non-human DNA typing have been published. Budowle and co-authors [24] presented the first guidelines on animal forensic DNA analysis, based largely on the human system, in order to educate about quality systems necessary for all forensic work. The guidelines serve to provide a baseline structure for quality assurance and quality control from which laboratories can build more detailed workplace specific standard operating procedures, while also promoting standardisation of these fundamental aspects across laboratories.

This theme has since been updated by discussions of the International Society of Forensic Genetics and subsequent publication of thirteen recommendations for best practice when typing non-human DNA for forensic purposes [25]. This discussion applies stronger emphasis on the importance of standardisation between laboratories as interest in non-human forensic work has increased markedly in the intermediate six years and different groups are introducing increasing variations into methodology. The specified recommendations are much more precise than those of the previous publication in order to ensure that wildlife forensic work keeps on track with court requirements. Likewise, the Society for Wildlife Forensic Science (SWFS) Scientific Working Group for Wildlife Forensic Sciences (SWGWILD) has produced Standards and Guidelines for best practice (accessed from https://www.wildlifeforensicscience.org/wp-content/uploads/2016/07/swgwild-standards_and_guidelines_2-0_12192012.pdf, accessed 03/06/2018).

Ultimately, wildlife forensic genetics work should be conducted to the same standards as human forensic work and the ISFG recommendations strive to ensure that best practice is in line with concurrent human casework. Many obstacles still prevent this congruence, such as under-resourcing and the infancy of the wildlife forensic industry. With time and increased interest in this expanding field, wildlife forensic investigations will develop in a direction to bring it in line with current human forensic best practices. A guilty verdict from either investigation will result in a sentencing and potentially incarceration for the defendant, so vigilance in stringent quality practices must be maintained.

1.1.7 The Wildlife Forensic Scientist's Toolkit

As mentioned earlier, the questions asked in wildlife criminal investigations are often quite different from those that arise in investigations of crimes against humans and accordingly the techniques applied to answer these questions differ. One important distinction when transferring human forensic DNA techniques to wildlife is that human investigations deal with one focal species, whereas wildlife investigations deal with many. Short Tandem Repeat (STR) profiling (these DNA markers are commonly referred to as microsatellites outside of the human forensic field), the workhorse of human forensic DNA investigations, is much more resource intensive for the associated returns in wildlife forensic investigations. Each species being dealt with requires population screening and the construction of an allelic frequency database in order to calculate forensic statistics reported in support of the competing propositions. This drain on resources has contributed to the much slower growth of wildlife forensic investigations, which are conducted by a fraction of the current organisations practicing human forensic work. Wildlife forensic DNA investigations also commonly apply a universal test for species identification, not relevant to human forensic investigations, which is relatively quick to develop and cheap to conduct.

The biological sample types encountered can be specific to wildlife forensic science, such as feather, bile, processed (e.g. pickled, cooked, dried, tanned and taxidermied) tissues, organs and skins, and powders claimed to contain derivatives of protected species. This thesis will focus on biological remains and in particular the application of DNA techniques to provide investigative answers in crimes against wildlife.

Every crime has at least one allegation and forensic investigations work to provide evidence relating to the allegation. Questions to be answered by wildlife forensic science are peculiar to wildlife crime and often involve the following:

- *What species is this?*
  This question often arises when an evidentiary item is not morphologically identifiable, but is suspected to derive from a certain protected species. The answer in this case requires a molecular test that can identify a species to the exclusion of all other species.

- *Does this crime scene sample match that from an individual?*
  A forensic investigation might ask to match meat from a protected species in a freezer to blood on a suspect's toolkit or perhaps discarded remains at the scene of an illegal poaching event. This technique requires a test for individualisation; that is, a test that can identify a particular individual and differentiate it from all other individuals in a population.

- *What is the relationship of this individual to that individual?*
  A parentage test is the wildlife equivalent of a human paternity test and might be required to investigate a fraudulent captive breeding program, where native young were collected from the wild and claimed to be the offspring of captive parents. This uses the same individualisation technique as the matching test.

- *Where did it come from?*

In the above scenario, where young have been identified as unrelated to the alleged captive parents, potential repatriation back into the wild can only be attempted if the most likely native source population is identified. The prediction of geographic origin is also useful for identifying poaching 'hotspots' of heavily persecuted and traded species, allowing limited resources to be focused on patrolling these areas. To answer this question using DNA analysis requires an extension of the individualisation technique and extensive surveying of native populations across their entire geographic range, a resource intensive undertaking that has yet been successfully achieved for very few species.

A forensic case should avoid reliance on a single line of evidence and techniques provided by different disciplines ideally provide independent support for a particular scenario. Many wildlife forensic techniques can provide evidentiary answers to the above questions, for example microscopy [26,27] and mass spectrometry [28] for species identification and stable isotope analysis for the determination of geographic provenance [29,30]. However, for the purpose of this doctoral thesis the discussion herein will focus specifically on the use of molecular genetic techniques to answer these questions. The next section will examine the underlying characteristics of the different DNA markers available and their application to the above questions.

1.2. Forensic DNA techniques

Put simply, the animal genetic blueprint comprises two types of DNA molecule, the single-copy chromosomal nuclear genome found in the cell nucleus (although nuclear fragments can be found repeated throughout the genome) and the multi-copy plasmid mitochondrial genome found in the mitochondrion organelle. Both mitochondrial DNA and nuclear DNA are of use to wildlife forensic investigations, as they can answer different questions based on their unique characteristics.

1.2.1 Mitochondrial DNA

Mitochondrial DNA is a small (~16 000 bp) circular molecule of double-stranded DNA that exists in multiple copies in most cells, rendering it an excellent candidate for ancient DNA and trace evidence investigations where only a small amount of potentially degraded material remains. Anywhere from 1 to 15 copies of mitochondrial DNA are contained in a mitochondrion. Given that each cell contains around 100 mitochondria, estimates of 500 copies of the mitochondrial genome per cell have been made, although this figure is possibly an underestimate due to potential mitochondrial destruction during the visualisation process [31]. It is a robust organelle, protected by an organellar wall, so mitochondrial DNA endures in the environment due to this characteristic as well as its high copy number [32].

Mitochondrial DNA is maternally inherited, so it can only inform about ancestry along the maternal lineage, not paternal relationships. Due to this pattern of inheritance, the recombination process that affords nuclear DNA the large phenotypic variation seen between individuals even between single generations is not present in mitochondrial DNA [33], so mitochondrial genotypes will remain very similar over many generations. Despite these mechanisms producing a more conserved mitochondrial genome, mitochondrial DNA has an evolutionary rate around five to ten times higher than nuclear DNA [34]. Mitochondrial DNA does not have the proof reading mechanism built into

nuclear DNA replication, so all changes created by the spontaneous mutation rate become incorporated into the new sequence at replication time, providing a degree of variation that is useful to phylogenetic and forensic studies.

Owing to maternal inheritance, relatively slow mutation rates (when compared to much more rapidly mutating STR markers) and lack of genetic recombination, the mitochondrial genome is highly conserved across a range of taxonomic levels and is thus a good candidate marker for examining deep phylogenetic relationships; it allows examination of relatedness at the coarse-scale resolution of species, genera, families and orders. A genetic signal will endure over many generations, which could amount to thousands of years of descent. These characteristics make it a popular marker for species identification.

1.2.1.1 Species identification using mitochondrial DNA

Regardless of the target region, the technique used for species identification follows the same process (figure 1.1). An evidentiary item of questioned species origin is submitted. All DNA is liberated and purified from the remainder of the cellular contents through a process called DNA extraction. Short, synthetic DNA oligonucleotides called primers are designed to target highly conserved DNA regions and produce an amplification product from a large range of taxa (universal primers), or a specific target region known only to amplify in a subset of taxa (species or group specific primers).

Note that if the species is completely unknown then universal primers are applied, but in a forensic context a species of interest is usually provided for which targeted primers might be available. PCR amplification is used to replicate the target region up to a suitable quantity to enable sequencing. The PCR product is purified, a Sanger sequencing reaction performed and the product further purified and the underlying sequence determined using a genetic analyser.

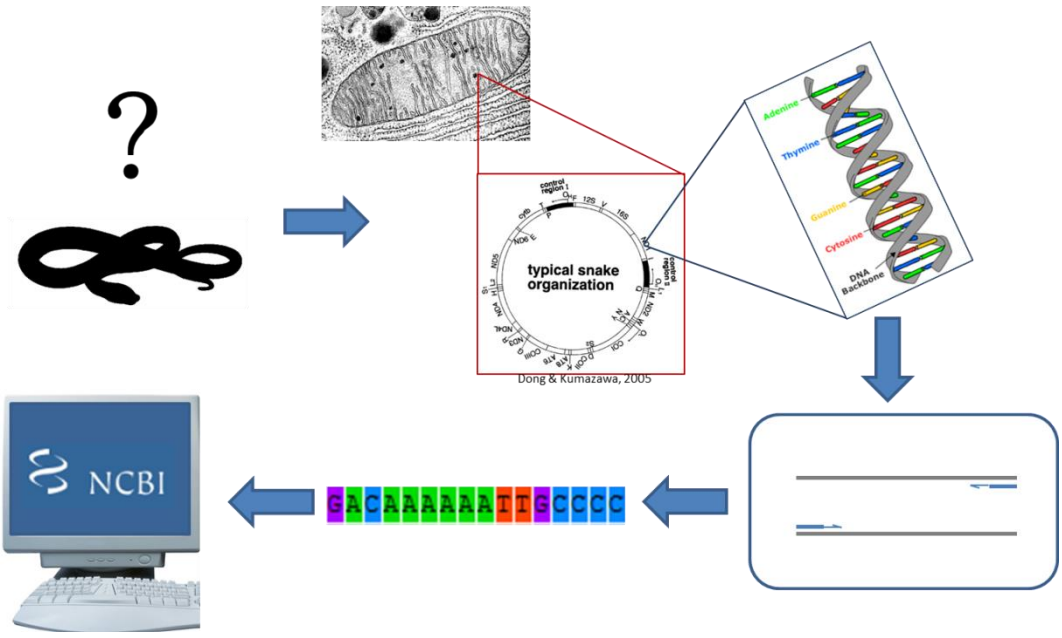


Figure 1.1: Process of species identification by sequencing of a mitochondrial locus, simplified for illustrative purposes. Mitochondrial DNA is isolated from the unknown sample and a specific target locus amplified and sequenced. The resulting sequence of the unknown is searched against a database of knowns to identify the most closely related species (indicated by highest sequence similarity).

The resulting sequence of unknown origin is compared against a collection of known homologous reference sequences from many different species. If the species from which the evidentiary item originated is represented in the reference sequences, then the unknown sequence should show the highest similarity with this species. If it is not, then the sequence should show the highest percent sequence homology with the species most closely related to actual source species. Phylogenetic reconstruction (arranging the sequences into a phylogenetic tree) is a desirable way to visualise the results because it reveals the phylogenetic relationships between the different reference sequences and the unknown. It also shows whether the locus is suitable for accurately reconstructing the phylogenies of the associated species (do the sequences group taxonomically in the expected hierarchical manner?). The ability of the locus to accurately reflect the phylogenetic relationship of the target taxon must be established as part of the validation of the technique prior to casework being conducted. This investigation should involve representation of all closely related species in a genus where possible, if the technique is going to be used to report a species identification result in criminal casework.

1.2.1.2 Choice of locus for species identification

The mitochondrial genome comprises almost entirely coding DNA. This functional constraint maintains highly conserved sequence. Most variation within protein coding regions occurs every third base (due to the codon redundancy in transfer RNAs). It contains genes coding 13 proteins, 2 ribosomal RNA subunits and 22 transfer RNA molecules, as well as the D-loop that does not appear to code for any functional genes. It is not subject to the same constraints as coding sequence, but does contain elements for initiation and control of replication and transcription [35]. Genome length and gene order can vary between taxa [36]. Most snake species have a duplicated control region [37–39].

Mitochondrial genes exhibit different degrees of conservation in mammals in accordance with their function. Gene conservation is generally ordered: D-loop < protein coding genes < ribosomal subunits < tRNA genes [40] (in [34]), but figure 1.2 provides a much more detailed summary from mammalian species. This affects the choice of gene examined for a particular purpose. By the advent of forensic DNA investigations, the mitochondrial genome had been well characterised across different taxa (e.g. [41–46]). As the D-loop (also called the control region) is most variable, it is not generally suitable for species identification because it exhibits too much intra-species variation (however it has historically been applied in some taxa such as cetaceans e.g. [47–49]). Highly conserved sequence such as either of the ribosomal subunits would be more useful for comparisons of highly divergent taxa, but less suitable for congeneric comparisons (although see [50]). Early studies used conserved regions in the cyt *b*, 12S rRNA and control region loci for priming sites [46,47] and examined cyt b as a potential target region for phylogenetically differentiating many mammalian taxa [51,52] (Irwin, Kocher & Wilson, 1991; Hsieh et al. 2001).

The issue of which locus to use has seen much recent debate in the species identification field. The amount of variability at most mitochondrial genes results in moderate to high inter-species variability (depending on the degree of divergence), but low intra-species variability. This is desirable, as a species identification test ideally targets a piece of DNA which is highly conserved within a species, but differs in base composition from all other species. Primers need to be placed in

highly conserved regions, either across all taxa universally or across the particular taxon of interest. In practice, finding the mitochondrial region appropriate to the different phylogenetic histories of different taxa is not a trivial task.

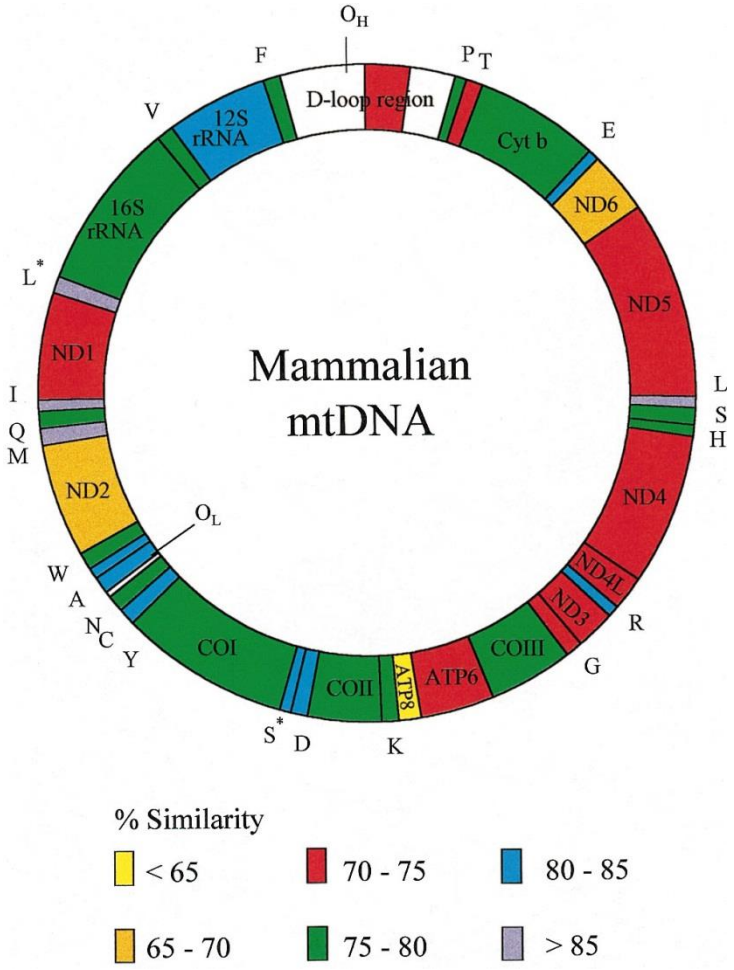


Figure 1.2: Degree of conservation in mitochondrial genes compared in 33 mammalian species (from [35]).

1.2.1.3 The advent of DNA 'barcoding' & debate over the best locus

In 2003, the cytochrome c oxidase I (COI) gene was advanced as a potential alternative to the cytochrome b gene for species identification [53]. The basis of this paper was to propose a single gene region which could be used as a 'barcode'[1] to identify any species to the exclusion of all other species - despite the fact that Palumbi & Cipriano [48] had already expressed the need for multiple locus investigations to address cases of species hybridisation. The COI region was proposed to fit this purpose because, in this particular demonstration and other research [54], it was particularly good at distinguishing between species of lepidopterans and had also been used successfully in planktonic copepods [55] and onychophora [56]. One should note that the aforementioned taxa are invertebrates and COI has been successfully applied to species identification of many invertebrate taxa e.g. [57], yet the technique was subsequently proposed as a universal species indicator for all

[1] I note here a terminological aside that the 'barcoding' label is broadly reserved for use of the COI gene alone.

taxa! This gene sequence is proposed as a superior alternative to cyt *b* as it exhibits higher sequence conservation and should therefore be more informative when looking at deeply divergent species [53]. However, a locus maintaining a high degree of conservation can confound the species identification tests required in forensic investigations if species have diverged at very recent timescales.

This rising interest in species barcoding by COI also sparked raging debate about various issues, including: the suitability of COI in comparison to other mitochondrial genes, such as cytochrome b [58], ND1 [59], 12S rRNA [60,61] and 16S rRNA [62,63]; it's use as a substitute to taxonomy (see Trends in Ecology and Evolution, issue 18, February 2003, [64]); and the appropriateness of a DNA system to be able to categorise and distinguish all life on earth [65,66]. From a forensic perspective, cyt *b* has been argued as the more useful locus for many of the mammalian taxa often encountered in casework [58], a conclusion reiterated in avian species [67]. Wilson-Wilde and colleagues [68] compared two mitochondrial genes (COI and ND2) and three nuclear genes (APOB, IRBP and GAPD) across Australian marsupials; they claim that ND2 was found to be the superior gene for species identification.

It is clear that the choice of gene will depend on the taxon of interest (and the reference sequences available, a point to be discussed shortly). Validation must be used to identify a region with a gene tree that accurately reflects the species tree. Both cytochrome b [69] and COI [70] have been validated for use in forensic casework. However, when applying these tests to a new species, the practitioner must undertake an appropriate assessment of the species' taxonomy and investigate whether it is correctly reflected within the genetic signal of the locus in order to confidently report the species identification results.

As hinted at earlier [48], application of the single mitochondrial locus without the combined use of nuclear DNA can create difficulties in identification of parapatric hybridising species where introgression of mitochondrial genomes is occurring [71]. Wiemers & Fiedler [72] emphasise the need to use any 'barcoding' region in conjunction with other non-mitochondrial DNA data in order to avoid misidentification that can occur through use of a single DNA locus. The use of nuclear DNA is problematic when forensic evidence involves trace amounts of DNA, but the above issue is one of which wildlife forensic biologists must be aware.

1.2.1.4 Online databanks and reference sequence repositories

The development of online repositories or databases of species sequences over time has allowed the refinement of the species identification technique. The three predominant genetic code databanks – the European Nucleotide Archive, Genbank run by the National Centre for Biotechology in the US and the DNA Databank of Japan – have joined to form one collaborative resource, the International Nucleotide Sequence Database. This database can be searched from any of the three collaborators websites. Using Genbank (www.ncbi.nlm.nih.gov) for example, the sequence is searched by a Basic Local Alignment Search Tool (BLAST) algorithm against the online database, which in 2010 already contained over one hundred million uploaded nucleotide sequences from more than 300 000 taxa listed at genus level or below [73]. Highly similar or 100 % homologous sequences are returned and enable the putative species of origin to be identified. If the sequences are not identical then a percent homology between sequences is also given. The use of mitochondrial DNA sequencing

results has been validated for human genetics [74], however the validation of specific searchable databases was not included in this report and is indeed rarely encountered.

One must exercise caution when using this technique though, as the source of the reference material has not been verified and can often contain mistakes [32,70,75]. Human contamination should be obvious if the sequence also retrieves results matching *Homo sapiens*. An assumption is made that the reference sample is representative of the species, rather than constituting a rare genotypic variant or sub-species [58]. Further, this technique relies on only a relatively short stretch (i.e. 400-550bp) of DNA and an identical match in the sequence of interest does not exclude possible variation between the two individuals at DNA sequence outside this section.

The Barcode of Life Database (BOLD, www. http://www.boldsystems.org/) resulting from the COI barcoding work described earlier, is often applied to forensic casework. It is considered a more reliable source of accurate sequence data than Genbank, due to increased measures of authenticating samples prior to curation onto the databank and inbuilt tools to flag and investigate data anomalies [76]. Despite or even in direct contradiction of these measures, many sequences represented on BOLD have been 'mined from Genbank' and BOLD no longer overcomes the issues of sequence authenticity encountered on Genbank. Another disadvantage is that BOLD only contains COI locus data (for animals), so precludes utility in species for which this locus is not an appropriate phylogenetic discriminator.

A novel fully-regulated online repository of whole mitochondrial genome sequences has recently been proposed, to be used for the identification of wildlife species commonly encountered in forensic investigations [77]. This ForCyt tool will allow wildlife forensic practitioners to access vouchered and validated reference sequences from any mitochondrial locus for comparison with their unknown casework samples. The ForCyt initiative will provide a highly improved, quality controlled and universally accessible resource for forensic casework involving species identification analyses.

1.2.1.5 Applications of mitochondrial DNA tests for species identification

The question of species identification is by far the most commonly addressed question in wildlife investigations, due to the relative simplicity with which DNA methods can be developed and validated. There are two possibilities in forensic casework: the allegation will stipulate a specified species; or the allegation will provide that legislation has been breached, but the particular species in question is uncertain, for example the smuggling of bird eggs from unknown taxa. Universal tests can be applied when there is no a priori knowledge of the species, as the primers are designed with low annealing stringency in highly conserved mitochondrial sequence to amplify a wide breadth of taxa. Nucleotide sequencing is required to determine the identity of the unknown sample. Universal tests are also useful when a certain species is suspected, but tests for this species have not yet been developed.

Species-specific tests might be more suitable when a particular species requires identification, such as a confiscated meat or pelt sample that is suspected to be from a protected species. If the primers have been designed to target only a specific subset of species, then their presence or absence is indicated by a positive or negative PCR product without the need for sequencing. Mixed DNA samples are frequently encountered in forensic investigations, particularly when Traditional

Medicines are involved and in these cases species specific assays are required to obtain an interpretable result, as DNA sequence from multiple sources cannot be deconvoluted without the use of new and expensive technologies that are not yet routinely accessed by wildlife forensic laboratories.

Universal tests

DNA sequencing and phylogenetic reconstruction

Baker and Palumbi [47] and Malik and colleagues [78] used DNA sequencing and phylogenetic reconstruction to identify protected marine species sold in marketplaces and indicate the incidence of illegal poaching for trade purposes. Wu and colleagues [79] were able to differentiate between the four extant species of sika deer using the mitochondrial control region and consequently identify the presence of the wild sub-species in traditional medicines that legally should only contain the two captive sub-species. This indicated the continual illegal harvesting of wild sika deer for the traditional medicine industry.

Traditional East Asian Medicine (TEAM) investigations constitute a recurring theme in wildlife investigations, due to their frequent incorporation of protected species. Phylogenetic reconstruction involving the cyt b locus has been applied to species identification of skinned and bleached turtle shells [80], seahorse [81] and processed and powdered rhino horn [82]. Due to the degraded nature of the horn, a nested PCR was designed to amplify only a small internal region of cytochrome b – a novel method to deal with difficult TEAM samples. It should be noted that the visualisation phylogenetic trees only gives a relative determination of how similar two sequences are to one another, unlike direct sequence comparison which can give a percent homology quantification of similarity (corresponding only to the small region begin compared), so both data should be considered.

DNA sequencing and sequence homology comparison

This DNA databank search technique is used in species 'barcoding' [53], yet the term barcoding also embraces the use of a standardised gene region (chosen as COI) and the BOLD sequence repository (see [83] for a critique of barcoding using the usual distance-based analyses). Due to the conceptual differences of the two methods, barcoding examples will be grouped separately succeeding other examples using traditional DNA sequence homology comparisons.

Parson and colleagues [84] simulated the types of poor quality or unusual samples seen in casework and demonstrated the successful species identification from taxa across the five major vertebrate groups; this work provided the earliest example of direct applicability of species identification by sequence homology comparison to forensic casework involving wildlife. A cyt b fragment was shown by sequence homology comparison to be diagnostic in uniquely identifying each of 221 different species considered forensically relevant [85]. Turtle derivatives for sale in US and Asian market places were also identified as having been substituted with other turtle species and even alligator using cyt *b* and control region sequence homology [86].

Pyrosequencing of the two mitochondrial ribosomal RNA subunits was attempted for 28 forensically relevant European mammals, but species identification of all species was unsuccessful due to the highly conserved nature of these genes [87]. This result reiterates the importance of conducting

multi-gene conservation rate studies when developing species identification tests to determine which genes are most suitable to identify the species in question.

Nested PCR analysis was used to identify the species origin of elephant ivory [88] and treated turtle shells [89]. This research group took the small amplicon idea further in avian species by designing five internal cyt *b* primers to amplify subsequently smaller fragments in degraded samples, increasing their chances of amplifying a product approaching maximal obtainable sequence size [90]. Even the smallest fragment of 220 nucleotide bases contained enough sequence variability to differentiate between all 40 bird species tested.

Marsupial specific primers have been designed to amplify very small fragments of cyt *b* in order to obtain macropod DNA from degraded samples such as museum specimens and forensic samples. Three sets of cytochrome b primers were combined into a triplex reaction to amplify a total of 554 bp from difficult samples with limited starting material, but these small fragments still have to be sequenced separately in order to identify particular macropod species [91].

While providing important trade intelligence, DNA sequence comparison has also led to successful criminal convictions. Identification of remnant tissue on a chopping board as originating from a protected peacock species led to the confession and apprehension of the accused [92]. One Korean forensic investigation involved allegations of illegal hunting that prompted the genetic examination of 11 evidentiary items [93]. The investigators applied cytochrome b sequence comparison to identify two of the six game meats as illegally hunted roe deer. The other exhibits were five skinned pheasants, in which hunting of females is illegal. Pheasants are easily sexed by plumage dimorphism, but the skinned exhibits had to be sexed by molecular genetic techniques. Fragment size polymorphism in the CDH gene of the pheasants identified two of the five exhibits to be female (also see [94] discussed below). The suspect was found guilty based on the DNA evidence and sentenced accordingly.

DNA Barcoding

Barcoding the COI locus has been applied to the African bushmeat trade [95] and to legitimise labelling of blue fin tuna in restaurants and markets around San Francisco [83]. Endangered species were found to be present and, while no direct prosecutions resulted, information about the continuing illegal import into California of protected tuna species destined for the sushi menu was obtained. Three forensic cases of poaching in South Africa were investigated using COI barcoding. Two exhibits were identified as beef, but the third carcass returned a 100% match to the reference samples of the reedbuck, indicating illegal poaching had taken place [96].

Sequencing is an unsuitable technique when DNA mixtures are involved and in these cases species-specific tests have proven valuable.

Species-specific tests

The following tests are species-specific. Unlike sequence homology comparison and barcoding using database searches, these techniques require prior knowledge of the species likely to be present.

Species specificity by Restriction Enzymes

The analysis of Restriction Fragment Length Polymorphism (PCR-RFLP) patterns offers a cheaper alternative to Sanger sequencing and has therefore been applied in forensic case investigations. Restriction enzymes are used to create digestion patterns specific to particular species. It can however be difficult to compare band sizes between gels, precluding the standardisation of techniques across laboratories, and its utility has been superseded by the availability of much more robust and standardised techniques.

Nonetheless, the potential effectiveness of PCR-RFLP has been demonstrated for a number of forensically relevant species tests including the differentiation of red and sika deer species [97], Indian crocodile species [98] and distinction between human and domesticate species [99]. This technique has also been combined with two sex-specific markers for use in identifying species and sex of game meat [100]. Combining information obtained from independent loci can add value to poaching investigations.

Single Nucleotide Polymorphisms in mitochondrial DNA

Single Nucelotide Polymorphisms (SNPs) are single base changes occurring throughout both genomes that can be diagnostic of particular species. While they offer great utility for species identification and population genetics, their application in this project has been limited and their discussion within this thesis will therefore be kept brief.

Identifying single nucleotide sequence polymorphisms that are species and sub-species specific allows a test to be developed to interrogate just these sites rather than the whole DNA sequence (especially useful when the remainder of the genome asides these bases is conserved). This simple test can determine in one step whether a particular species is present in a sample, providing greater efficiency than universal primers. Of equal importance, this test can determine the presence of a species in a mixed sample containing DNA from multiple species.

While originally developed for conservation genetics [101], SNP testing by Allele Specific Primer Extension (ASPE) offers a simple but excellent forensic test for presence or absence of a species. Development of the primers must involve an extensive validation process of screening individuals within the species and other species that might reasonably be expected to be present, in order to ensure the robustness of the test to give a product for every species member and no false positive results for other species.

This species-specific primer technique has been applied to monitor the illegal trade and false labelling of sturgeon [102,103] as well as a criminal investigation into tiger meat sold illegally by a circus to which the tiger had belonged six months earlier [104]. Single species-specific mitochondrial 12S rRNA primers have been applied in Real Time PCR to detect trace amounts of a species in mixed and degraded samples and quantify the amount of target species present [105] (see [106] for methodology involved). Nuclear gene species-specific primers have also been applied to differentiate European carnivores [107], yet this approach is less common due to high gene conservation caused by selective pressure.

Species-specific primers have been combined with other informative tests to provide more information for forensic investigations. Y chromosome markers have been combined with a control

region species-specific primer pair in elephants to obtain both species and sexing data [108]. This study is a rare application of the Y chromosome to wildlife work, even though Y markers are becoming routine in human casework.

A dual species/gender test has also been developed for birds [94], which is quite unique because birds exhibit Z/W female heterogamety and these chromosomes have been far less characterised than the X/Y male heterogamety system (but also see [93] discussed above). In this case, two avian specific primer sets amplify Z and W regions, enabling sex identification. These products happen to exhibit species-specific size polymorphism, allowing the simultaneous determination of avian species. Similar species identification based on fragment size polymorphism has been found in a nuclear gene, TP53, but the underlying variation of the intronic region has not been well characterised and validation studies are necessary [109].

Whole mitochondrial genome sequencing for SNP identification

Species-specific SNP sites are very useful for species identification, but can be difficult to locate. This is particularly true when looking at sub-species between which there is low genetic variation, such as the four dwindling sub-species of extant tiger. Kitpipit and colleagues [110,111] designed an overlapping set of 26 primers to amplify the entire mitochondrial genome of the tiger and found that they contained very few polymorphic mitochondrial sites. Targeting just a small number of species informative SNP sites resulted in a more precise species identification test than comparing the whole of the mitochondrial genome between individuals, due to ambiguities caused by non-species informative mutations [112]. Tiger sub-species could be identified by a haplotype containing five particular SNPs not seen together in any other species tested. Sub-species polymorphisms were identified as indicative of two of the four tiger sub-species, but no SNPs specific only to either of the other two sub-species were identified.

Similar primer design has been undertaken on the Asiatic Brown Bear, culminating in an 18 overlapping primer set test to amplify the entire mitochondrial genome [113]. This test would likewise prove useful in identifying SNPs from which to design a species diagnostic SNP test, given that the Asiatic brown bear exhibits 87-92% mitochondrial genome sequence similarity with the three other extant *Ursus* species.

With this application in mind, the prospects of the ForCyt project become even more exciting.

SNP multiplex tests

More information can be gained from a multiplex of many SNPs in one reaction. This is extremely beneficial in forensic investigations, which are often working with very valuable and limited trace samples.

Optimal species-specific primers for each of six shark species were combined into an octaplex with the two universal shark primers [114]. Chapman and colleagues [115] have similarly designed a highly robust species test for white shark. A pentaplex amplifying two fragments from the cytochrome b gene and three fragments from the nuclear ribosomal internal transcribed spacer 2 (ITS2) gene introduces bi-organellar redundancy of the test to provide heightened stability, robustness and quality control. One of each primer pair is species-specific to the white shark and the other primer sets are universal, acting as internal positive controls. This test has since been used to

monitor shipments for white shark [116], and detect illegal white shark trade in Hong Kong markets [117]. Similar tests have also been applied to the Chilean shark fin trade [118], and designed for the basking shark to detect derivatives in Hong Kong and Japanese market places as well as international trade worldwide [119]. Validation of the test for forensic application has been reported previously [120,121].

Tobe & Linacre [122,123] have further refined the species-specific SNP multiplex concept by combining universal and species-specific primers in an amplification pair to incorporate more species into a single test. A universal forward primer amplifies a region in conjunction with a species-specific reverse primer. If the species and therefore the reverse annealing site is not present, then no product will result. Reverse primers are designed so that species products size uniquely. These products can then be separated in a single capillary electrophoresis lane. By including multiple universal forward primers amplifying different regions, more species can be combined into a single test. This approach, using three different gene regions, has enabled 18 different European mammals of domestic and forensic importance to be identified with a single test. Further, this test has been fully validated and demonstrates sensitivity to detect product down to 17 femtograms of DNA. Robustness is increased by inbuilt dual primer set redundancy for each species, safeguarding against polymorphism at one priming site.

A similar test to the above Allele Specific Primer Extension (ASPE) SNP multiplex kit for European mammals has since been tailored specifically for species commonly found in TEAM and their common domesticate substitutes [123,124]. A universal primer pair is also built into the test to detect whether negative results are due to inhibition caused by some medicinal product substrates. This technique has also been developed for the species identification of 19 terrestrial New Zealand mammals [125].

The above technique applies to SNP detection. Single Base Extension (SBE) has also been applied to multiplex testing for species identification. A SBE multiplex technique, known as SNaPshot, has been applied to test for the four extant wild subspecies of tiger [112,126]. SNaPshot is more appropriate for species testing than ASPE when there are very few species-informative SNP sites present and samples are increasingly degraded. For instance, in the whole tiger mitochondrial genome five SNP sites were identified as species informative and a further six could differentiate between sub-species. The SNaPshot test therefore interrogates these specific sites to detect the exact nucleotide polymorphism present and this can distinguish population of origin in tigers. The small size of the amplification product (primer + 1 base) makes this test ideal for highly degraded remains – in this particular test the maximum product size was a mere 71 base pairs [112]. A SNaPshot based assay has also been developed to identify the species origin of elephant ivory [127]. Allele Specific Primer Extension is cheaper to use than SNaPshot kits that are purchased through biotechnology companies, but the initial developmental stage is time consuming.

1.2.2 Nuclear DNA

In brief, double-stranded DNA is found in the nucleus of every eukaryotic cell (with some exceptions, although snakes possess nucleated red blood cells). In humans, this genetic material is packaged into diploid cells (containing two sets of 22 autosomal chromosomes and two sex chromosomes), or in the sex gametes, a haploid set of 22 autosomal chromosomes and one sex-linked chromosome. Humans and most other mammals exhibit male heterogamety, where all individuals exhibit an X

chromosome and the inheritance of a Y chromosome as the second chromosome determines the male sex. Some avian and squamatan species exhibit female heterogamety, where a male has two Z chromosomes and the inheritance of a W chromosome determines the individual to be female [128,129]. The Y and W chromosomes are referred to as lineage markers. Lineage markers have their own favourable characteristics and Y chromosome markers are used extensively in forensic human identification [130–132]. Equivalent applications for the W lineage marker have not yet been developed for routine use in wildlife investigations beyond [94], but this is one area with potential for development in the future.

The human nuclear genome contains 3.2 billion nucleotide bases, though only approximately 5 % actually code for functional proteins [133]. These gene regions consist of DNA that is related directly to the generation of a biological product. Natural selection acts upon coding DNA, so limited variation is seen within gene regions between individuals, otherwise the functionality of the products would be affected. A eukaryotic gene complex includes areas that initiate transcription, promoters, coding sections termed exons, and non-coding sections that separate the exons, termed introns. Both exons and introns are transcribed, but the introns are spliced out of the primary transcript prior to translation [134]. These intronic regions are responsible for regulatory functions.

While less than 5 % of genomic DNA is comprised from exons, a further 20 % of the genome is comprised from introns and sections of DNA up-stream of the exons relating to initiation of transcription. This leaves approximately 75 % of the genome that is comprised from extragenic DNA (see Figure 1.3). Extragenic DNA is invisible to selective forces and the freedom to mutate produces highly polymorphic markers that are highly informative about relationships between individuals.

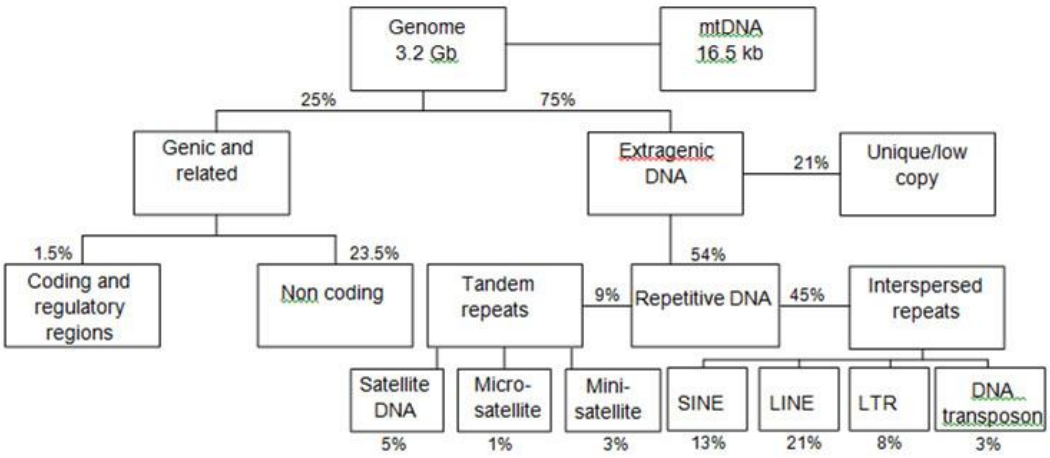


Figure 1.3: DNA type and proportion within the human genome (from [135]).

Repetitive DNA actually comprises a large proportion of the extragenic DNA. These repeat regions (first termed satellite repeat regions as a result of the density gradient centrifugation study that led to their discovery [136]) vary in repeat motif size, repeat number and frequency within the genome. Variable Number of Tandem Repeats (or VNTR [137]) is another overarching name for all DNA containing tandemly repeated motifs of varying size [138]. Moderately repetitive regions tend to be of large motif sizes and many are found interspersed between coding regions in genes. Some moderately repetitive regions in the genome are known as transposable elements and are found

duplicated throughout the genome (see [139] for further reading). These are in fact known to move throughout the genome, but further discussion of this phenomenon is not within the scope of this introduction.

The first application of DNA to a forensic investigation involved the use of telomeric repetitive motifs termed minisatellites by Jeffreys [140]. Minisatellites contain repeat motifs that range between 10 and 100 bases in length [133]. Recognising that these size variable regions (size varying due to motif repeat number at a given site) were heritable and could therefore resolve individual identity and relationships between individuals, Professor Sir Alec Jeffreys isolated these markers in humans using Restriction Fragment Length Polymorphism (RFLP) primers to target the variable regions [141]. Their first application in forensic casework was to resolve an immigration dispute [142]. This technique he termed 'DNA fingerprinting' was shortly thereafter used to link evidence from two murdered girls to the otherwise unsuspected offender and this became the first successful apprehension of a criminal via DNA evidence [133]. While DNA provides critical evidence to apprehend, it plays just as important a role in exonerating the accused, although these outcomes often don't receive as high a profile in the media (see innocenceproject.org for more information).

A smaller alternative marker has since superseded the use of minisatellites in forensic casework. Microsatellites exhibit similarly highly repetitive DNA, but the repeat motifs vary between two and six base pairs in length (hence the name). These markers offer increased ease of genotyping as the two microsatellite fragments found at a particular site in any individual are similar in size due to the much smaller repeat motifs, even though repeat number can ranges between 5 and 40 units [143].

1.2.2.1 Short Tandem Repeats

Microsatellites are currently the routine marker used in human forensic biology worldwide. Refer to Goldstein & Schlotterer [144] for an in depth discussion of underlying microsatellite biology, function and applications. For reviews of microsatellites, see [143] and [145].

Having been identified shortly after minisatellites and thus named microsatellites [146], these regions are often also called Short Tandem Repeats (STRs) and Simple Sequence Repeats (SSRs) [143]. Due to the term's wider use in forensic applications, STR will be used to refer to these markers for the remainder of this thesis. STRs appear within the nuclear genomes of most taxa, although at varying frequencies ([147] provides a good comparison of STR proportions in different taxa). Homologous STR regions are seen in closely related species (and become less comparable as taxa increase in genetic distance), which makes these excellent target areas for numerous applications including fairly recent phylogenetic studies of population differentiation and contemporary evaluations of relatedness.

STRs are lengths of repeated motifs. The repeat number varies between individuals and this variation has arisen through slipped strand mispairing at the time of chromosomal replication [148]. Accepted allele nomenclature in forensic science denotes the number of repeats in the underlying sequence. Dinucleotide repeats (consisting of two base units, e.g $(CA)_n$) are most frequent within the genome [149], followed by tri-nucleotide repeats, which will not cause a frameshift mutation in a neighbouring coding region [143]. Yet, tetra- and penta-nucleotide repeats are far more useful for forensic studies, due to the lower frequency of associated analytical artefacts, namely stutter peaks (described shortly).

STRs are surmised to experience mutation most frequently as a result of polymerase slippage during replication. The mutational pattern, put simply, generally follows a Stepwise Mutation Model (SMM, [150]), where a whole single repeat unit is gained or lost, resulting in a new allele one repeat unit longer or shorter than the ancestral allele. This mutational pattern endows STRs high allelic diversity (some loci showing >0.90 heterozygosity [151]), but also facilitates homoplasy (hidden past allelic diversity), as single step mutations are just as likely to step back to a previous state as give rise to a new state.

Although SMM is the dominant mutational force, STRs also undergo other mutational events such as point mutations, insertion/deletions (indels) and anomalies involved with recombination, such as unequal crossing over. STRs are selectively neutral. The lacking selective restraint, recombination during meiosis and mutational model afford STRs a much higher mutation rate than most other markers (between 10^{-2} and 10^{-6} mutations per generation), which can be either beneficial or detrimental to analyses, depending on the question being asked. This high mutation rate further intensifies the problem of homoplasy, as more frequent mutations lead to quicker saturation of mutational possibilities.

1.2.2.2 Stutter

Both alleles exhibited by autosomal STRs are easily detectable through electrophoresis. Due to size differences which might only involve one or two base pairs, STR alleles are most commonly viewed by capillary electrophoresis, which gives much better resolution of similar size alleles than traditional gel electrophoresis. Certain artefacts such as stutter can arise from the analytical process. The same slipped strand mispairing that causes STRs to be highly polymorphic occurs during PCR when the polymerase enzyme has difficulty replicating the repetitive DNA [133,152]. As the complementary DNA strands separate during denaturation, a hairpin loop of one repeat unit can form upon reannealing causing a daughter DNA strand differing by one repeat. If this loop forms in the template strand, then the resulting molecule will contain one less repeat unit. This phenomenon is a common artefact of STR analysis, evident as a small amount of product one repeat unit shorter than the allele of interest on the viewing platform, and is known as 'back-stutter' or simply 'stutter'. Very occasionally, the synthesised strand will loop out during replication causing a fraction of product one repeat unit larger and this is known as 'forward stutter' or ‘over-stutter', although it is much less commonly seen than back-stutter (see figure 1.4). Stutter is much more pronounced in di- and tri-nucleotide repeat units, and could easily be confused with genuine alleles in a profile with more than one contributor. Therefore, larger repeat units such as tetra- and penta-nucleotide repeats are favoured in forensic studies.

1.2.2.3 STR repeat class and implications

Table 1.1 shows of the nomenclature relating to STR motif arrangements as agreed upon by the European DNA Profiling group [153], although terminology and indeed definition of a STR locus still endures debate [154]. Simple (a.k.a. pure or perfect) repeats comprise a string of a single motif type, surrounded by non-repeating flanking DNA. When a simple repeat motif exhibits a micro-variant (an incomplete repeat unit created by an Indel), the result is called a simple non-consensus repeat. Compound STRs contain two different motif types. The size of each

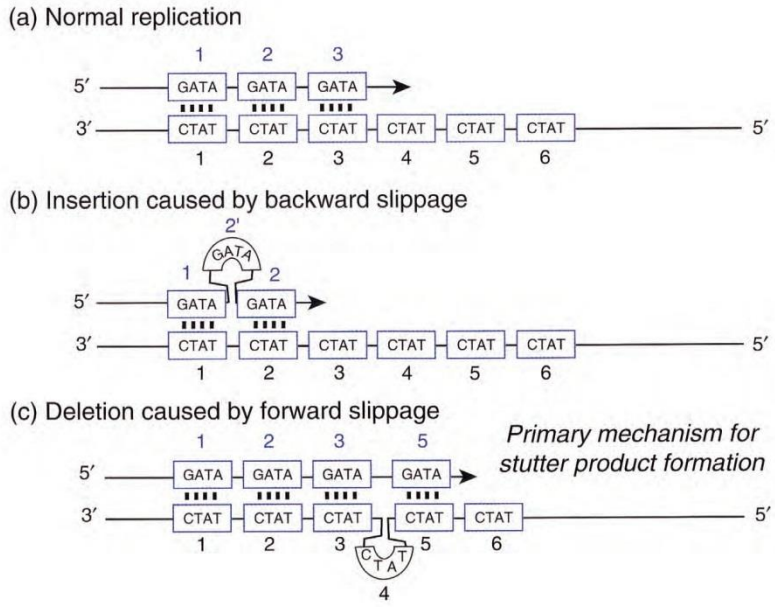


Figure 1.4: Mechanism of slipped strand mispairing proposed to cause stutter bands during STR amplification (from [133]).

motif is not necessarily equal. Complex STRs represent more than two repeat motifs flanked by non-repetitive DNA. These might contain incomplete motifs and invariable motifs (see Table 1.1 for examples of STR nomenclature).

Table 1.1: Examples of STR nomenclature as described in [153] .

Nomenclature	Example motif
Simple	$(ATTT)_{8-14}$
Simple non-consensus	$(TCAT)_4CAT(TCAT)_5$
Compound	$(TCTG)_3(TCTA)_5(TCTG)_4$
Complex	$(TCTA)_5(TCTG)_6TCA(TCTA)_3TCCATA(TCTG)_3$

Simple STR motifs and STRs with longer tracts are more likely to exhibit higher allelic diversity [155]. It's possible that longer tracts of simple repeats are more effective at creating polymerase slippage. Work on human Y-chromosome STRs by Kayser and colleagues [156] suggests that complex arrangements might exhibit increased mutability over simple repeats of the same motif number due to effects caused by neighbouring motifs.

1.2.2.4 Wildlife forensic applications of nuclear DNA

Hypervariable STR loci provide a rich source of information for answering various forensic questions. By screening populations for alleles at multiple loci, allelic frequency databases can be constructed. Using the genotype information, examinations for Hardy Weinberg Equilibrium, Linkage Disequilibrium and measures of population differentiation (F_{ST}) can inform about the suitability of the data for forensic applications. The allele frequency database can be used to calculate likelihoods that two profiles taken at random in the population would match, allowing the probability of a

reference sample and crime scene sample originating from the same source to be determined. Relatedness of individuals can also be calculated e.g. [157], providing evidence about parentage in investigations of breeding programs or individual identity where the victim's remains might be compromised (see [158]). A statistical assessment of allelic frequency differences between populations can culminate in a probabilistic assignment of individual to a source population [159].

Individualisation of crime scene samples

Unlike species identification tests, individual identification tests require separate development for each species. This question is often preceded by a species identification test, as the correct species needs to be determined in order to assess whether individual identification tests appropriate to the species in question have been developed. If profiling markers exist for a species, but an allelic frequency population database does not, profiling could still be of use to exclude the individual as a source of the crime scene sample. However, the alternative outcome of a match is of no use if a database is not available to calculate supporting match probabilities.

Individual identification of wildlife is conducted using the same DNA profiling technique applied in human forensic investigations. Multiple independent STR loci (as with human work) are used to construct an allelic frequency database from reference samples of the species of interest. This database shows relative incidence of alleles within the population. A statistic can be calculated by multiplying the allele frequencies (following the product rule if loci are independent) to evaluate the probability that an evidentiary profile originated from the particular individual of interest, rather than another unknown individual within the population exhibiting an identical profile by chance. The addition of independent loci (and especially loci with greater numbers of alleles) decreases the likelihood of seeing the same combination of alleles in two individuals in the population, subsequently increasing the power of discrimination of the dataset.

In this way, animals can be matched to evidentiary items, such as blood and hair, in the same manner as humans. Linkage between loci should also be examined as this will dictate which statistical analysis should be used (for instance, the product rule is used for the statistical analyses of unrelated loci, but related loci such as the lineage markers require the counting method of statistical analysis as they are not inherited independently of one another).

It is crucial that allelic frequency databases are large (or capture the majority of extant individuals where large populations no longer exist) to accurately capture the actual allelic diversity and frequency within the population; otherwise the resulting statistics will be inaccurate. Population structure also becomes a greater issue when calculating probability of identity for wildlife than with humans [160]. Wildlife populations tend to exhibit much higher sub-structuring, due to behavioural factors or restricted environmental mobility. This results in increased sharing of profiles within a population (due to common descent) and causes a bias in favour of the prosecution hypothesis if not addressed. Further, if a population database actually contains two or more subpopulations (Wahlund Effect [161]), then the use of the greater dataset to calculate a match probability for an individual belonging to a subpopulation will underestimate the population substructure and incidence of profiles matching by chance, biasing the results in favour of the prosecution hypothesis. Increased population structure must be factored into statistical calculations appropriately. This is facilitated by the use of theta (θ) to correct for population substructure. Theta represents population F_{ST} values (Wright's [162] measure of population subdivision), given random mating

within subpopulations [133]. This correction factor is determined from the underlying breeding behaviour and size of the population in question and must be included in the validation of wildlife allelic frequency databases. Endangered species with few surviving individuals will inescapably be inbreeding and thus require a higher θ adjustment. As a stronger θ value is applied, the probability that profiles match by chance increases, favouring the defence hypothesis.

Selective mating and inbreeding are issues relevant to wildlife populations that are usually discounted in human statistical calculations due to human behaviour favouring outbreeding. Animals may exhibit hierarchical mating systems or live in superfamilies, causing elevated inbreeding levels. The breeding behaviour of animals might otherwise be unknown, but inbreeding is likely to be more prominent than in human populations and should be factored into statistical calculations. A factor *f* exists to account for inbreeding in the match probability [163], however the best way to apply this statistic is debated [164]. Polygamy is an added issue in wildlife species, not so commonly encountered in human investigations.

An exclusion of an individual as a match to a crime scene profile demonstrates that the crime scene sample did not originate from the individual in question. A non-exclusion of an individual that matches every allele at every locus is not as definitive, as the loci chosen are only a subset of the entire genome and it is possible that the individual could be excluded through use of alternative loci. An evaluation of the match probability of two individuals sharing a profile at the given loci is always necessary in the case of a non-exclusion to provide statistical support to the conclusions of the test based on the known population allelic content. In the forensic context, presentation of the inverse of the match probability, the likelihood ratio, is preferable to present the results in terms of the competing prosecution and defence hypotheses.

The requirement of an allelic frequency database for each separate species is reiterated, as this elevates the cost of wildlife profiling investigations in comparison to human. However, many of these databases have already been developed by molecular ecologists and conservation geneticists (e.g. [165]) and can be made suitable for forensic investigations through additional validation studies – a much more cost effective approach than entire marker development.

Due to their cost, allelic frequency databases have frequently been developed for species with high commercial returns, such as game animals (e.g. deer: [166–168]; wild boar: [169] and prominent fisheries species (e.g. Salmon: [170]). Animals often in conflict with humans have also been targeted due to the high frequency of problems arising. This includes animals threatening to human safety (e.g. bears: [171]; leopards: [172]) as well as those considered a pest for their impact on human industry (e.g badgers: [173,174] foxes: [175]).

Dawnay and colleagues demonstrate the procedure of constructing an allelic frequency database for forensic use, using the example of the Eurasian Badger [173]. The process of database validation as outlined by the Scientific Working Group on DNA Analysis Methods (SWGDAM) is presented and difficulties following these guidelines when pedigree records and chromosomal maps are not available addressed.

A commercial STR kit produced for domestic dogs was used to construct an allelic frequency database of the UK dog population including 15 STR loci and one sex differentiation marker [176]. Three loci included in the kit were discarded from the database as inheritance behaviour deviated

from Hardy-Weinberg Equilibrium, suggesting loci to be linked. The authors calculated the random match probability between unrelated individuals across breeds to be 2.8×10^{-17}. Both the badger and dog DNA profiling sets are currently applied to forensic casework in the UK.

Wild boars are a popular game hunting target in the UK and are frequently poached illegally. An STR allele frequency database using 11 tetrameric markers was constructed to calculate the probability of exclusion for individual identity testing [169]. Wild boar and domesticated pigs constitute the two sub-species of *Sus scrofa*. In this situation, a domesticated pig reference database was also typed; results demonstrated STR typing to be capable of discriminating between subspecies, which will aid investigations involving illegally poached boar meat that is claimed to have originated from domestic pig.

Northern European brown bears were sampled remotely (hair trapping) to construct a 12 locus database, which was successfully applied to forensic casework [171]. A profile match was obtained between the bear carcass and a blood stain from the suspect's firearm in an illegal poaching investigation, providing important evidence for the prosecution. Another case involving a bear located at a taxidermy shop saw the suspect exonerated when genetic evidence identified the bear to have been a legally hunted individual.

STR loci were used for individual identification of leopards in Western India, which are often in conflict with humans due to their adaptability to environments bordering settlements. Remotely sampled (scat collection) specimens were profiled and compared against a database constructed of 37 wild-caught individuals to identify a further seven individuals [172]. Population structure is evident in the leopard population and the probability of identity assuming all individuals were siblings was 5×10^{-4} or 1 in 2000 individuals.

Tigers have also been the subject of individual identity testing due to their critically endangered status and prevalence in smuggling and poaching cases. Seven loci developed in cats cross-amplified in tigers to enable screening of 37 reference individuals [177]. Although three loci were found to depart from Hardy-Weinberg Equilibrium, match probabilities of the test including all seven loci were calculated (details of calculations not given). This test was applied to casework samples of eight bone pieces and results suggested all fragments to have originated from one individual.

<u>Relationships between individuals and parentage testing</u>

Mitochondrial DNA is termed a lineage marker due to single parent inheritance and is therefore only useful for looking at familial relationships down the maternal line. The conserved nature of the almost entirely protein coding genome and limited genomic size make mitochondrial DNA unsuitable for examining relationships between individuals. This investigative tool has been useful in human forensic investigation left only with degraded remains of important figures e.g. [178,179], yet it can provide limited use in wildlife forensic investigations and should be used in conjunction with nuclear markers for more definitive outcomes, see [158].

The applicability of DNA profiling to answer questions regarding individual relationships and parentage lies in the diploid nature of nuclear DNA inheritance in sexually reproducing individuals. This inheritance mode introduces genetic variation among generations and between siblings from the same parental pair – variation further enhanced by the integrated mechanism of recombination.

Not long after the advent of human DNA fingerprinting, the first published parentage investigation was applied to wildlife. Minisatellite profiling was used to assess parentage claims in sparrows and to reach the conclusion that an offspring showed DNA fingerprinting patterns not seen in either of the alleged parents and was therefore from an 'extrapair copulation' [180].

This question of parentage is relevant to wildlife investigations of captive breeding programs where the provenance of an individual as captive or wild-caught is questioned. Breeders require a permit to take individuals from the wild and, on occasion, an individual might be taken from the wild to supplement breeding stock and be claimed to be an offspring of a breeding pair in the captive population. A paternity test can be conducted by comparing alleles between the progeny and the alleged parents. Alleles belonging to the offspring not seen in a parent result in a straight forward exclusion of the offspring as having been sired by that individual. When all alleles match each parent, then an allele frequency database of the population must be used to calculate the probability that the offspring was sired by each parent as opposed to any other member of the population [157]. The rate of mutation should also be taken into account and, if the markers have a high mutation rate as STRs do, then one or two mismatches might be acceptable.

STR profiling databases have been developed for six raptor species in an initiative funded by the UK government to directly counter illegal laundering in avian captive breeding programs [164,181]. Allele frequency databases were constructed and validated for forensic application. While four of the genetic systems demonstrated high discriminatory power and robustness, tests for two species were found to be inadequate for legal proceedings due to low sample sizes used to screen markers [181]. These techniques have since been applied to forensic investigations, one such case involving a peregrine falcon breeder suspected of providing false parentage documents. One chick was excluded as offspring to the parental pair, but the breeder subsequently admitted to the possibility of record keeping errors and was able to produce a chick that matched the parents at all loci [164].

Renshaw and colleagues [182] applied 25 STR loci to wild and hatchery populations of the marine fish red drum, including in the analysis computer simulated progeny for any mating combination within and between these two groups. Exclusion and likelihood tests could identify wild versus captive bred individuals within error margins acceptable for forensic application.

Parentage tests have also been conducted to identify the origin of poached animal remains. A tigress at an Indian zoological park was poached for its hide and claws in 2000, the crime receiving high media interest and remaining unsolved for five years [158]. In 2005, a framed claw and tiger hide suspected to belong to the poached individual were located, but the carcass was not available for direct individual testing. Seven STR loci and mitochondrial D-loop testing were applied to the evidence samples and a parentage test conducted using the extant parents and sibling of the tigress. All alleles were shared between the alleged family members and likelihood ratios of these relationships calculated. The study also included 34 unrelated individuals and one allele was specific to the alleged tigress family to the exclusion of all other individuals. The D-loop sequence matched that of the alleged mother. These findings resulted in charges being laid against the accused.

Another wildlife investigation involved the poisoning of a protected vulture through intentional baiting with a laced lambs carcass [183]. STR parentage analysis was used to identify the parental flock of the poisoned lamb to the precision of the particular mating pair, the flock origin thereby

revealing the identity of the perpetrator. Individual identification and paternity testing has also been applied to koala poaching [184].

Population assignment and geographic provenance

This scenario is addressed last because the techniques used build on those already discussed. It is a particularly relevant question in wildlife enforcement because species ranges rarely coincide with legal jurisdictions and populations or stock are only protected within the political boundaries of the protective legislation. In a forensic genetic context, the question 'from which geographic origin did this come' equates to 'from which reproductive population did this individual originate?' [164].

Population differentiation using DNA requires populations to exhibit diverging genetic characteristics. Natural populations show varying levels of admixture and relatedness. As population differentiation decreases, additional genetic markers are necessary to distinguish between them, until the point is reached where two populations are so genetically similar that differentiation cannot be achieved and they are essentially one population. This was the case of the red drum discussed earlier [182], where 25 STR loci were unsuccessful in assigning source population, due to the low genetic divergence between the two groups.

Statistical analyses are then applied to the allele frequencies to calculate the probability of an individual originating from one population and also surrounding populations. This allows an assignment to be made based on the most probable event. Allelic frequency data must be available for all of the populations under question in order to place an individual in one population to the exclusion of all other populations (although assignment techniques that apply smoothing methods to estimate missing allele frequencies are being designed to account for sampling gaps that include the source population, see [185]). Numerous statistical tests applying different models have been developed for population assignment and there is no consensus about the best technique to apply [186]. Regardless of the technique applied, the forensic scientist must ensure that they understand the underlying assumptions of the chosen model and are able to explain these if challenged in court.

Illegal cases of animal translocations have been aided by geographic assignment tests. Four red deer individuals were suspected to have been translocated into Luxembourg for game hunting [187]. Using multilocus STR genotyping, each of the individuals had alleles in at least one locus that excluded them from any of the three reference populations that were sampled in Luxembourg. While this does not assign their population of origin, it does support the allegation of illegal translocation. Wild pig populations in Western Australia were also genotyped during population genetic investigations. Some individuals exhibited multilocus genotypes that did not fit with their current population. Their genotypes suggested that they had originated from populations distantly removed, indicating that illegal translocation for recreational hunting had occurred [188].

Game hunting fraud has also been detected in fishing competitions using STR markers for Salmon [170]. The winning individual's genotype was excluded from the population that the fishing competition targeted. The individual could not be assigned to a source population, and the fisherman confessed to having purchased it from the local shop. In this case, the population exclusion was enough to press charges.

STR markers have also been identified in Ramin [189], red cedar [190] and Sapelli timbers [191] that could aid investigations of illegal logging and source population, should the appropriate reference populations exist to conduct population assignment tests. Dormontt *et al.* [192] notes that STRs often exhibit poor amplification from timber derived DNA template.

Historically, assignment tests have encountered limitations caused by sampling design. If the population of interest has sparse coverage over a large geographic range, then it becomes difficult to sample the reference population over the entire geographic range to ensure that the resulting frequency database is accurate. The power of the assignment test decreases with inadequate sampling. If the individual of interest's source population has not been sampled, the assignment will be even more inaccurate as it will be assigned with most probability to a population from which it did not come.

A smoothing method has been developed to fill the sampling gaps by estimating allele frequencies in areas from which no individuals have been sampled. This technique was first demonstrated with African Elephants, a classic example of a species difficult to sample for an accurate reference database due to their high geographic range and light distribution in any particular locality [185]. Allele frequencies from 16 STR loci were determined over the entire geographic origin, including both savannah and forest populations.

This information has been used to pinpoint the origin of illegally traded ivory and focus anti-poaching and law enforcement efforts [193,194]. The largest seizure of ivory since the trade ban in 1989 was suspected to originate from various elephant populations in Africa. The above assignment test indicated that ivory actually originated entirely from savannah elephants and could be pinpointed to a particular East-West band of elephants located around Zambia. Interestingly, the 16 STR markers also indicated that rare hybridisation events had occurred between savannah and forest elephants[185].

1.3. Which marker is best?

Each technique can benefit wildlife forensic investigations. There is overlap in the questions that each technique can answer, so it's not a simple situation of one technique for one application. When choosing a technique to answer a given question, one should weigh up the advantages of each technique, alongside consideration of current access to equipment and readily available resources for the question, and in wildlife casework species, at hand.

Mitochondrial DNA markers are ideal for degraded or trace remains often found at crime scenes due to their high copy number and more robust packaging than nuclear DNA. The rate of sequence conservation is high enough that comparison of more deeply divergent species can be made and so this is a good candidate tool for species identification tests. The mitochondrial genome is a linked single locus, so non-mitochondrial markers should also be assessed to provide increased statistical significance to relationships between individuals. Maternal inheritance makes this molecule unsuitable for paternity evaluations [195] and inquiries about hybridisation. Paternal leakage can cause heteroplasmy within individuals and pseudogenes can appear as problematic nuclear paralogues to the true mitochondrial target locus [196]. In wildlife forensic investigations,

mitochondrial DNA is a useful marker for species identification tests and can also provide an alternate source of information about phylogenetic relationships of taxa that have diverged further back in time, should these need to be clarified prior to development of forensic assays.

In rare cases, usually where population numbers have declined severely and there is no longer genetic migration between fragmented populations that have now become isolated, genetic variation between populations will have become discrete enough that mitochondrial markers can be applied to identify the geographic origin of an individual. This was the case above with the subspecies identification of Sika deer [79] and tigers [112]. While this trait intensifies as species near extinction, most species under study do not exhibit such extreme isolation.

STR loci have also been applied for the purpose of species identification[197,198]. One locus has been applied to Indian felid species [197]. The marker amplifies across species, exhibits very low polymorphism within species, and length polymorphisms group discretely for each species. Encountering an STR marker with these characteristics happens rarely. McCusker & colleagues [198] qualify that wolf-fish exhibit very low genetic differentiation and the application of STRs to species identification might only be suitable in large ranging species, with few geographic barriers and low genetic diversity, such as marine species.

STR regions are found in nuclear DNA, so there is a higher abundance of markers to characterise. STR markers have been demonstrated to detect hybridisation events e.g. [185]. In most cases they are free of selection and their very high mutation rate makes them excellent candidates for individual identification and paternity studies. Yet, homoplasy and null alleles [199] can confound their use and the uncertainty of the underlying mutation model makes them less suitable than Single Nucleotide Polymorphisms (SNPs) for statistical models that rely on underlying assumptions [195]. STR sites and frequency are variable between genomes and can be difficult to characterise in some taxa [147].

Single Nucleotide Polymorphisms (SNPs) have not been discussed in detail herein (although they might be the underlying cause of variable RFLP patterns) because they are not being investigated within this project. However, they must be acknowledged for offering advantages over species identification by sequencing and STR analysis. SNPs are found in both nuclear DNA and mitochondrial DNA. Like STRs, SNP sites can be multiplexed and used to construct allelic frequency databases from which to calculate an individual match probability or geographic assignment statistic. SNPs occur more frequently throughout the genome (every 300-1000bp) than STR loci (every 5000-50000bp) [195]. Ascertainment bias used to be considered a disadvantage of SNP characterisation [195], but this has been resolved by Mass Parallel Sequencing (MPS) technology. Dormontt and colleagues [192] observed that STR markers perform poorly in timber species and SNPs provide a favourable alternative. SNPs offer a versatile substitute to the other markers discussed and their prominence in forensic investigations is steadily increasing.

STR loci exhibit much higher allelic diversity than bi-allelic or tri-allelic SNPs and so many more SNP sites are required to give the same statistical confidence in paternity exclusions and relatedness estimates [195,200]. An STR assay involving 4-6 loci has around the same discriminatory power as a SNP test involving 10-20 loci for individual identification, dependant on the measure of Heterozygosity at each locus [195]. Chakraborty and colleagues [201] validated 13 STR loci for human individual identification. They found that many more SNPs were needed to provide the same

power of resolution, however SNP loci alone would not resolve more complex cases unless they were supplemented by the 13 validated STRs.

To date, many more allelic frequency databases for non-model species have been published using STR markers rather than SNPs, so wildlife forensic investigators should review and utilise any existing data before considering the development of a new screening test for the species of interest. Interest is rapidly increasing in SNP typing of non-model organisms (see the special issue of Molecular Ecology Resources titled “SNP Development in Non-Model Organisms” [202]).

In fact, using a combination of techniques can produce improved outcomes. By itself, mitochondrial DNA sequencing cannot be applied to mixed DNA samples. Applying the SNP technique of ASPE can isolate DNA from the species of interest and this DNA fragment can then be sequenced to give increased accuracy in the species identification [159]. This technique is ideal, provided there is only one individual of the particular species present in the mixed sample.

Ramakrishnan & Mountain [203] present the utility of SNPSTRs, STR tracts that are closely linked with SNP sites in the flanking DNA, to increase the precision and accuracy of estimating divergence times. The associated SNP can help clarify the uncertainties caused by homoplasy in the hypervariable STR. Grimaldi & Crouau-Roy [204] also discuss the benefit of STR flanking region mutations assisting separation of populations with otherwise inseparable evolutionary histories due to homoplasious events. While rare, this sort of combined marker when identified could provide increased benefits over traditional STR typing of wildlife samples, especially in the context of geographic assignment tests.

The utility of MPS to forensic marker investigation and casework has not been discussed in detail in this chapter, as it is of minimal focus within this project. Application of MPS is however gaining momentum by wildlife forensic scientists. This technology is most often applied for novel marker discovery (as demonstrated in Chapter 3). However, it offers great potential for contributor sequence deconvolution from mixed human DNA samples (discussed in an MPS workshop presented by D. McNevin & R. Daniel at the ANZFSS Symposium, 2014 in Adelaide, Australia) as well as determination of taxonomic composition of complex mixed samples such as traditional medicines [205].

1.3.1 Forensic investigations involving multiple questions

Many published forensic case studies demonstrate that investigations frequently ask more than one question, requiring the application of multiple genetic approaches:

A wild boar carcass was located in an Italian wildlife park shortly after it had been illegally killed [206]. On return to collect the carcass, the poacher was apprehended. Necropsy determined the boar to have been fatally wounded by knife cut. A knife was located at the suspect’s house that displayed blood-like staining on microscopic analysis. No other evidence was found to link the suspect with the carcass. The blood from the knife was examined with PCR-RFLP, which indicated it to be of single source from *Sus scrofa* species, but restriction analysis cannot differentiate between domestic pig and wild boar. Sequencing of cytochrome b identified a sequence match to wild boar. Morphological analysis refuted the possibility of a wild-domesticate hybrid. Twelve STR loci were applied and demonstrated a complete match of alleles between the bloodstain and the carcass.

Using these markers, a reference database of 30 wild boar from the park estimates a Probability of Identity between siblings of $6.7x10^{-6}$ or one in 150 000 individuals, which is substantial given the small population size of a few hundred individuals. The suspect was subsequently convicted of illegal poaching and animal cruelty charges.

Enforcement officers apprehended a tourist attempting to smuggle four iguanas from Galapagos [207]. Analysis of the Cyt b gene allowed species identification of the four iguanas to *Conolophus subcristatus*. An existing mitochondrial DNA database for the Galapagos populations of iguanas assisted the research group to identify the population of origin of each individual and perform subsequent repatriation.

A protected Sardinian mouflon was found poached and the farmer claimed it to be one of his sheep [208]. Mouflon are the wild counterpart of sheep, both sub-species of *Ovis aries*, and in rare cases interbreeding occurs. High relatedness of these animals rendered identification by mitochondrial DNA inappropriate, so STR based population assignment was undertaken to assess whether the individual was a domestic sheep or wild mouflon. Sixteen sheep STR loci from an online genomics site cross-amplified in mouflon. Reference databases were constructed for both sheep and mouflon, although admixture was identified enough structure was present to identify the carcass as that of pure mouflon. The database was also used to demonstrate that evidentiary bloodstains from the farmer's sheepfold matched the poached carcass. In this investigation, different statistical analyses of genetic data were able to answer the questions of both individual identity and population assignment; the farmer was indicted for the crime and the case is reported to be ongoing.

Poaching of the endemic and threatened Cypriot mouflon (*Ovis orientalis ophion*) was also investigated by local authorities using DNA evidence [209]. A car containing several bloodstained exhibits was intercepted by police; three protected mouflon carcasses were located on a nearby roadside. All bloodstains and the three carcasses were sequenced for cytochrome b and exhibited identical sequences identifying them as mouflon species. A reference database was utilised that contained genotypes of ten STR loci in 47 wild mouflon. This was used to match nine of the exhibits with two of the three genotypes with a Likelihood Ratio of >3000 (which offers strong support given the low number of extant Cypriot mouflon). The driver was charged with illegal poaching of the mouflon.

Expansions of wolf populations in Italy are causing frequent conflict with humans. For fifteen years, DNA data has been collected by remote sampling and genotyping of deceased wild animals as an initiative to develop wolf and domestic dog population genetic databases and also inform the management of protected wolves. The database contained 417 individuals at time of publication. A necklace containing ten canine teeth was confiscated from a villager in a locality where a wolf was found poached and missing its entire muzzle [210]. The reference databases were used to determine species origin of the teeth as domestic dog or wild wolf, individually genotype each tooth to assess the minimum number of individuals, assess whether any teeth matched the wolf carcass and assign a likely population of origin to each tooth. Mitochondrial control region wolf-specific primers were also used to amplify a region that contains an Italian wolf diagnostic haplotype. Nine teeth gave single source control region sequences that matched the Italian wolf haplotype (the remaining tooth did not produce amplifiable DNA).

STR data also identified the presence of six different individuals. Analysis of the sexing markers ZFX/ZFY indicated three female and three male wolves. The three male samples were examined with three wolf Y-STR markers that confirming their identity as Italian wolves. Profiles from one tooth matched the de-muzzled wolf carcass, while two other profiles were direct matches to two wolves remotely sampled within the area of the suspect's village. A Bayesian analysis of genotypes to assign population of origin supported the previous findings that all individuals were protected Italian wolves. This case demonstrates well the different types of evidence that genetic techniques can provide wildlife forensic investigations.

To conclude, the value of forensic genetic markers to wildlife investigations has also been demonstrated in an Australian endemic species of Black Cockatoo [211]. Reference databases developed for the Red Tailed-Black Cockatoo and White Tailed-Black Cockatoo were used to investigate three allegations of illegal activities involving these species. A Red Tailed-Black Cockatoo nestling (not yet able to fly) cockatoo was confiscated from a residence after the homeowner was witnessed inspecting a Red Tailed-Black Cockatoo nest in a roadside tree. Using seventeen STR loci from a database of 30 individuals, eggshell from the nest was a direct match to the DNA profile obtained from the nestling. The reference database was used to calculate a likelihood ratio of 1 in 389 billion that the samples match by shared origin rather than chance. The suspect pled guilty to illegal poaching and Red Tailed-Black Cockatoo nestling was confiscated.

This cockatoo STR database was also used to investigate the paternity of 15 Red Tailed- and White Tailed-Black Cockatoos found in a breeder's backyard; foul play was suspected, as these birds are difficult to rear in captivity. Paternity testing found two of the White Tailed-Black Cockatoos not to be related, and at least one bird was found to match a genotype from a wild nest site on the White-Tailed Black Cockatoo database. The suspect pled guilty to illegal poaching and all birds were confiscated.

The last example presented by White and colleagues [211] involves the illegal shooting of Black Cockatoos in an orchard. Using STR analysis, at least 13 separate individuals were identified as having been shot. None of the individuals matched the 347 profiles on the population database of White Tailed-Black Cockatoos. Unfortunately, no evidence existed to link the crime to a perpetrator so no charges could be laid.

The wildlife forensic investigations discussed in this section demonstrate the range of questions for which DNA analysis can provide evidence in criminal investigations and also that successful convictions can result if the evidence is presented in an appropriate manner. The different DNA markers can be applied complimentarily and all comprise the wildlife forensic biologist's tool kit for providing evidentiary information in a judicial context. In some cases, even the suggestion of DNA evidence might be enough to provoke an admission of foul play [212]. The publication of successful convictions of offenders in wildlife crimes by DNA analysis could also act to deter the activities of potential wildlife poachers.

Australian snakes are popular in international legal and illegal trade, yet limited work has been conducted to develop forensic tests and deter the poaching of protected native species. The development of forensic genetic tests for pythons and other protected snake species will form the final section of this chapter.

1.4. Snakes in wildlife crime

1.4.1 Reptiles in wildlife crime

Media coverage concerned with wildlife crime tends to focus on a handful of high profile species, such as the tiger, rhino, and great panda [32]. High profile species have dominated a large part of forensic investigations as they are able to attract funding due to either public popularity e.g. [52,82,104,197,213] or their role in an important economic or commercial industry (e.g. abalone [214], game hunting [215]).

Yet, herpetofauna (reptiles and amphibians) are a large constituent of the international wildlife trade and any product traded legally begs the opportunity for a black market supplying illegally acquired cheaper alternatives. Around the turn of this century, the international trade in reptile skins was already alleged to consume 10 million reptiles each year [15,216]. The legal trade has increased over time [14] and was reportedly worth $6 million per annum worldwide in 2002 [13]. The global net legal trade in reptiles was reported to be 20 116 616 individuals between 1995 and 1999 [217]. The most prevalent legally traded species are not protected by legislation [14,218]. At least 44% of the top 25 imported or exported in the United States live reptile trade were not protected by CITES [217].

Legal trade records can only leave one guessing at the extent of illegal trade, but prosecution records can provide much better insight into the magnitude of the problem. Eighty-nine percent of Australian reptiles are considered endemic [20]. It is no surprise then that reptiles were the most targeted taxa in illegal wildlife smuggling detected by the Australian Customs Service between 1997 and 2004. They were involved in 43% of prosecutions cases during this time as recorded on the Australian Customs Prosecutions Database [2] (note that the category of reptile appears to include amphibians in the study).

Reptiles are a popular target to illegal traders as they can come in very unique forms and peculiar morphs, fetching high prices on the international pet market. Being typically small they are relatively easy to conceal either in the post, luggage or on one's body [4] and their quiet nature renders them much more likely to avoid detection than birds, for instance. They tend to be fairly robust and more likely to survive transport in compromised conditions than other animals – although that's not to say that mortality rates are not high during smuggling attempts [4]. There is high demand for reptile products, such as leathers and skins [219]. Clearly, there is a need to tightly regulate trade in reptiles and prosecute those responsible for illegal activities to break the cycle driving many endangered species to extinction.

1.4.2 Snakes in illegal international trade

Snake derivatives are also commonly found in traditional medicines [220–224]. Consumption of snake products is believed to convey certain qualities associated with snakes and also provide therapeutic benefits for numerous ailments [221,225]. Snake derivatives are frequently traded at markets in China [221,222], Hong Kong [225], Brazil [226,227], Vietnam (where snake wine is hugely popular [228]) and South Africa [229] to name a tiny subset of localities. This popularity can cause heavy exploitation of particular species (Gao, 1996, as cited in [221]. Many traditional medicines

make their way into Australian markets and alternative medicine shops opening the possibility that these products may contain snake derivatives that possibly include protected species.

Snake skins and leathers are also a popular commodity and reason that snakes are heavily targeted in legal trade [216,230], which often leads to illegal markets. While snakes are also traded as food in many developing countries [231], this is does not generally pose a problem for Australian wildlife enforcement except if travellers enter Australia carrying snake products for personal consumption; this creates a quarantine risk (as well as a CITES breach if a python).

Over-collection for the pet trade has also contributed to severe declines in wild populations of many snake species [232]; the Australian snake species *Hoplocephalus bungaroides* and the focal point of this study, *Morelia spilota* are included [233,234].

1.4.3 Protected snake species

1.4.3.1 CITES listed snakes

Snakes protected by CITES can be seen in Appendix 1. Note that all species of the family Pythonidae are listed in Appendix II. Species of the genus *Python* are common in the international skin trade [216]. The Royal Python (*Python regius*) was the second most commonly traded reptile between 1996-2012 and accounted for 14% of total reptile trade according to CITES records [219]. Heavy harvesting has been linked with reduced physical and reproductive health in wild populations [4]. A TRAFFIC trade report on the live reptile trade in Europe [235] highlighted *Morelia bredli (Bredl's python)* to be of great trade demand to EU markets. Furthermore, both *Morelia bredli* and *Morelia oenpelliensis (Oenpelli python)* are claimed to be abundant in illegal international trade in general. Aside from the pythons, *Cerberus rynchops* (a water snake) and *Hoplocephalus bungaroides* (Broad Headed Snake) are the only other CITES listed snakes that are found naturally in Australia, the latter endemic to the country. International trade in these species is regulated through mandatory trade permits.

1.4.3.2 Snakes of interest to South Australian enforcement authorities

Snakes protected in South Australia are any species indigenous to Australia, any species that migrates regularly to Australia and any species listed in the National Parks and Wildlife Act, 1972 (Hannah Dridan, Department of Environment and Natural Resources, *pers. comm.* 2012).

According to the National Parks and Wildlife Act, 1972, no snake species are listed as endangered, the Desert Death Adder (*Acanthopis pyrrhus*) is listed as vulnerable and the Spotted Brown Snake (*Pseudonaja guttata*) and Western Black-naped Snake (*Simoselaps bimaculatus*) are listed as rare species and therefore protected by this legislation.

It is an offence to take any of the species mentioned in table 1.2 from the wild. One individual can be kept without requirement of a permit, but attainment must be through a licensed fauna dealer or appropriate permit holder and from captive stock. Permits are required to sell or otherwise dispose of an individual.

Table 1.2: 'Basic' species that require a Fauna Permit to be taken from the wild in South Australia

Common name	**Species**	**Common name**	**Species**
Black-Headed Python	*Aspidites melanocephalus*	Master's Snake	*Drysdalia masterii*
Brown Tree Snake	*Boiga irregularis*	Olive Python	*Liasis olivaceus*
WA Carpet Python	*Morelia spilota imbricata*	Pygmy Python	*Antaresia perthensis*
Carpet Python	*Morelia spilota*	Slaty-grey Snake	*Stegonotus cucullatus*
Centralian Python	*Morelia bredli*	Stimson's Python	*Antaresia stimsoni*
Children's Python	*Antaresia childreni*	Spotted Python	*Antaresia maculosus*
Common Tree Snake	*Dendrelaphis punctualata*	Water Python	*Liasis fuscus*
Diamond Python	*Morelia spilota spilota*	Woma Python	*Aspidites ramsayi*

The fact that legislation dictates rules pertaining to species and subspecies definitions introduces problems into enforcement efforts. The definitions of these categories have to be unambiguous in order that molecular genetic evidence can provide support for or refute the allegations made in forensic investigations. Unfortunately, in the case of pythons, they are not.

1.4.4 Molecular genetic investigations of snakes

Much molecular genetic based research has been conducted on snakes in the evolutionary biology disciplines of molecular ecology and conservation genetics. Evolutionary relationships among snake taxa have been examined typically using mitochondrial DNA sequences with more recent studies incorporating nuclear loci as single locus sequence data (e.g. c-mos gene [236]), multi locus STR data for population level studies (e.g. [209,237] or whole genome sequencing to study evolution and adaptation [238–240].

Some of the published data might be applicable to forensic investigations once the appropriate validation has taken place, yet very few studies of snakes have undertaken investigations with the intent of applying the results directly to wildlife forensic investigation. A collaborative approach is required between Australian wildlife enforcement officers and research scientists in order to develop tests for use in wildlife forensic investigations that will directly answer the questions asked during investigations and fulfil and comply with the requirements of the legal system.

1.4.5 Taxonomic difficulties of Pythonidae

Historically, the situation has not been helped by the taxonomic debate of what exactly constitutes a species and a subspecies [241–245]. Traditionally, species boundaries have been based on the biological species definition, which utilises behavioural and morphological traits to distinguish between species. Genetic investigations into species boundaries have revealed the limitations of species determination based on behavioural and morphological characteristics [246]. Behavioural distinctions might reflect selection on a fewf traits due to local environmental conditions rather than the widespread genetic differentiation that populations exhibit when on their own unique evolutionary trajectory.

Morphological characters can exhibit phenotypic plasticity and won't necessarily reflect the underlying genetic divergence. Different disciplines use different variable characters to classify species, which don't necessarily concur. Further, there are always exceptions of species identifications based any one of these characters. De Queiroz [247] argues that, while all of these traits should contribute to the recognition of different species, none is integral to defining a separate species. The sub-species category is problematic as it treats the species construct as another taxonomic rank rather than a unique entity in its own right, as it is now commonly accepted to be [247]. Examination of genetic variability provides but another alternate means to assessing the presence of separately evolving lineages, from which conclusions about systematics can be drawn.

Different classifications of python taxonomy based on phenotypic characteristics have been debated (e.g. [248,249]) and new molecular techniques have provided an important alternate line of evidence about evolutionary relationships [236,250–254]. Schleip & O'Shea [255] published a review of current python taxonomy and one can see from the amount of synonymous names in this document that defining python systematics has not been a clear or easy task.

Pythonidae is still accepted to contain the eight genera proposed by Kluge [249,255]; see [254] for further insight into phylogenetic relationships within Pythonidae). Seven of these genera (comprising 22 species) are found in Australia and New Guinea, with *Python* being the only exclusion. *Antaresia* and *Aspidites* are confined to Australia, while *Apodora* and *Bothrochilus* are only found on the New Guinean mainland and island territories [254]. The remaining genera *Leiopython*, *Liasis* and *Morelia* can pose quarantine difficulties for Australian enforcement agencies if traders falsely claim that an imported specimen originated in the country of import. Molecular markers might be able to aid these investigations, but only if enough differentiation exists between cross-territory populations to distinguish them from one another.

Python curtus, P. reticulatus, *P. breitensteini* and *P. brongersmai* are prevalently harvested for leather, meat and medicinal uses [231]. Due to the heavy exploitation of this group, phylogenetic studies were undertaken to determine whether its previous taxonomic grouping as one species containing three subspecies was accurate [231]. The data indicated the populations to be genetically divergent for each subspecies to be elevated to the species level.

1.4.5.1 Systematics of Morelia

Within pythonidae, the genus *Morelia* was proposed to contain six species by Kluge [249]: *Morelia amethistina*, *M. boeleni*, *M. carinata*, *M. oenpelliensis*, *M. spilota* and *M. viridis*. This view is still largely accepted, although central Australian python *M. bredli* was grouped by Kluge as a subspecies of *M. spilota*. *Morelia bredli* is now considered a separate species [256,257] with *Morelia* comprising seven species.

Limited molecular genetic investigation has been conducted specifically on the systematics of *Morelia*; most systematic classifications are based on morphological assessments e.g. [258]. Molecular genetic work using direct sequencing of two mitochondrial and one nuclear locus examined the genetic variability in Boelen's python (*M. boeleni*), native to New Guinea, and compared it with congeneric and co-habiting *M. amethistina* and *M. viridis* [259]. Boelen's python exhibited reduced genetic variation in contrast with the other species, suggesting a recent population size reduction and raising concerns about its use in captive breeding programs.

Morelia viridis is popular in trade, but for the pet market (sought after for novelty rather than utility value). At least 80% of individuals exported from Indonesia as captive bred are suspected to have actually been wild caught [260]. This species provides the added complication for Australian enforcement that populations occur naturally on both sides of Torres Strait, inhabiting New Guinea and also the Australian northern tropical rainforests. Although these two populations represent one species, translocation between them poses a biosecurity risk for Australia. The northern Australian population exhibit a very close genetic relationship to the southern New Guinean snakes and form a single lineage, expressed as a clade nested within the New Guinean populations [261].

1.4.5.2 Morelia spilota

Within *Morelia*, the *Morelia spilota* species complex poses its own problems to taxonomists and therefore legislators that rely on correct taxonomic boundaries. Commonly referred to as Carpet Pythons and Diamond Pythons, the underlying systematics of this species complex has been debated over the last thirty years [257,262,263] and still remains unresolved. Its natural range crosses seven different state jurisdictions, each with their own differing legislation. This vast range, which is largely continuous down the eastern third of Australia, spans many varying habitats from rainforest, through temperate woodlands and grasslands to arid and vegetatively sparse rocky outcrops.

Covering the largest geographic range of all the *Morelia* species, *Morelia spilota* has correspondingly been split into the largest number of subspecies, seven as proposed by Barker & Barker [257]. This may be as expected as the large diversity of habitats could cause a wide variety of selective pressures and phenotypic plasticity not mirrored in the underlying genotypic record. *Morelia spilota* exhibits a variety of morphological characteristics across its range. Colouration and patterning, behavioural traits [264], sexual dimorphism [265] and mating systems [266] have been used in attempts to categorise these snakes [234]. *Morelia spilota* taxonomy has been confounded by the use of morphological characters which have overlapping ranges and for which areas of intergradation exist between the putative subspecies. Shine [234] has also argued that some populations cannot be simply classified using these features.

Of the seven subspecies recognised within the Carpet Python complex, some are geographically isolated from each of the other subspecies (i.e. south-western WA, New Guinea), but the remainder are found within a continuous distribution across northern and eastern Australia. Most of the subspecies have a contentious taxonomic status and whether these subspecies represent separate populations or evolutionary lineages has yet to be established.

A doctoral dissertation [246] was the first molecular genetic study undertaken to elucidate the taxonomy of the *Morelia spilota* complex in order to aid wildlife conservation, management planning and law enforcement efforts. Taylor used allozymes, mitochondrial DNA (*control region*) nucleotide sequence data and nuclear STR allele frequency data from 334 individuals across the range of *M. spilota* to reassess the systematic basis of the species classification, including the seven subspecies. His study confirmed *M. bredli* as a separate species, but also demonstrated genetic differentiation of *M. spilota imbricata* that warrants its reclassification into a separate species. Classifications of the putative subspecies residing in northern and eastern coastal Australia remained unresolved; a conservative conclusion was made that these populations form one species *M. spilota* and subspecies ranks should be discarded, but evidence exists for potential population genetic structuring within this large group and further work is required. Insufficient sampling was

highlighted as a confounding factor of analyses in and between certain populations and increased sampling of these areas recommended, which would greatly benefit future studies of *Morelia spilota* systematics. The discontinuation of the subspecies ranking has important implications for law enforcement, as current legislation includes guidelines specific to subspecies (see National Parks and Wildlife Act SA, 1972; National Parks & Wildlife (Wildlife) Regulations, SA, 2001, Regulation 19).

1.4.6 Problems that pythons pose to enforcement agencies

Pythons are a popular trade commodity for many reasons. Due to symbolic associations (often mythologically derived) and the vast range of distinctive colour and pattern morphs that occur naturally or are purposely bred, they maintain high novelty value as pets or trophies. Australasian pythons display many morphological characteristics not seen elsewhere, adding to their popularity in overseas markets. Pythons are also popular to keep as pets and python breeding for the local commercial trade is a substantial industry in Australia. A large number of pythons exist in captivity, estimated to be around 25 000 carpet pythons in South Australia alone (H. Dridan, *pers. comm.*). Currently in South Australia, a permit is required to breed and trade protected pythons.

As discussed earlier, pythons inhabit a large geographic range that covers most of Australia. This range includes fairly continuous populations down the eastern coast with morphological intergradation. Problems arise when captive breeders illegally take wild caught individuals to supplement their own captive collection and doctor breeding records to reflect the false parentage (an issue common for reptiles, see [4]). An alternative means of pedigree analysis is required. Molecular genetic testing can assist with parentage analyses, whilst also holding the potential to provide information about the most likely population of origin if the appropriate reference databases have been developed and validated to demonstrate suitability.

Isotope analysis and chemical profiling has also shown promising results in differentiating wild and captive pythons from South-East Asia [267], but like DNA profiling the development of a reference dataset representing the native population is resource intensive. Information about an individual's population of origin is important as it can help to identify poaching hotspots and channel enforcement efforts, while also assisting possible repatriation efforts.

Pythons further present a complicated scenario in legal settings when their native range crosses political boundaries; the carpet python's natural range traverses Australia's northern border with New Guinea. Not only might legislation involving international trade be in breach, but the trade scenario creates a biosecurity threat to Australia.

1.4.7 Australia's close proximity with New Guinea – problems of local enforcement and biosecurity

A publication entitled "Snakes across the Straits" [268] describes exactly this scenario and examines the phylogeography of three Australasian elapid snake genera that naturally occur on both sides of Torres Strait. With the fluctuating sea levels of the late Quaternary, a land bridge has existed at times between Australia and New Guinea, allowing multiple migrations between the countries. Interestingly, no genetic variation was seen in 1360 bp of the mitochondrial *ND4* and *cytochrome b* genes between Taipan individuals examined, which originated from Merauke, Indonesia, Central Province in PNG and Cairns, Queensland. Only one individual from each locality was sampled, yet the

results suggest very limited genetic variation between these populations. This suggests that nuclear DNA might also be of limited use to support allegations of transfer between these populations.

A similar mitochondrial DNA based phylogeographic study was undertaken by Williams and colleagues [269] to examine relationships between Brown snakes sampled from Papua New Guinea, Indonesian Papua and Australia. The research aimed to determine number of colonisation events and whether these were a result of natural migration or human intervention. Again, population divergence rates indicate continental crossing by land bridge during the Pleistocene. Interestingly, the eastern New Guinea populations are most closely related to the northern Queensland populations, while the Arnhem Land populations are genetically most similar to those from Merauke, Indonesian Papua, indicating that two independent dispersal events occurred, separated by Lake Carpentaria. These mitochondrial DNA sequence data suggest that these populations could contain enough genetic variation for a DNA-based test to be able to assign individuals to their population, or country, of origin.

The distribution of *Morelia viridis* also spans the Torres Strait and might likewise cause problems for enforcement agencies. As discussed earlier, mitochondrial haplotypes were not shared between populations inhabiting Australia and New Guinea [261]. The haplotype diversity between these two groups suggests there is enough underlying genetic differentiation to distinguish between New Guinean and Australian Green Pythons, allowing molecular genetic support of law enforcement and border security.

1.4.8 Previous DNA Analyses

1.4.8.1 Species Identification

The prevalence of snake products and derivatives in highly processed goods such as leathers, dried meats, snake wines and a multitude of traditional medicine products directs the need for molecular genetic tests that can help to identify whether a species of protected snake is present within a sample. Many of these goods are no longer morphologically identifiable and morphology can also be misleading in some cases if different species exhibit similar or clinal variation in characters.

Limited work has already been conducted regarding snake species identification. Yau and colleagues [220] multiplexed a technique called sequence-characterized amplified region (SCAR) to identify three species of protected snake commonly encountered in traditional medicines. In this technique, an extension of the Randomly Amplified Polymorphic DNA (RAPD) method, products are subsequently sequenced and species-specific primers designed. Species-specific primers were developed for three snake species (*Zaocys dhumnades, Agkistrodon acutus* and *Bungarus multicinctus multicinctus*) and validated against another six species not used in Chinese pharmacopeia.

Work subsequently published by the same research group [225] involved sequencing of the *cytochrome b* gene in six snakes species including 3 elapids, 2 colubrids and a python (*Python regius*). Four of these species are protected by CITES and all are commonly consumed in Hong Kong [225]. Universal primers (from [46]) were used to sequence a *cytochrome b* gene fragment in these six taxa and these sequences, along with sequence data from GenBank for another 84 snake species, were used to construct a *cytochrome b* database containing 90 different snake species.

Wong and colleagues [225] conclude with a case report describing the application of the database to identify six fresh and dried meat samples confiscated by Chinese authorities. Other species commonly traded as meat were included in the comparison and the phylogenetic reconstruction using all sequences demonstrated the unknown samples to group most closely with *Python reticulatus* and *Python molurus* – reiterating the popularity of pythons as a meat source in Asian countries. The paper claimed that the snake *cytochrome b* database was available for researchers to access at time of publication, but this database has since been decommissioned due to insufficient resources for maintenance (P.C. Shaw, *pers. comm.*).

The DNA research group of the Central Forensic Science Laboratory in West Bengal, India also conducted investigations into the development of different species identification techniques for three endangered species of snake: Indian Rock Python (*Python molurus*), Indian Cobra (*Naja naja*) and Checkered Keelback (*Xenochrophis piscator*). Species identification techniques included a PCR-RFLP test targeting the *cytochrome b* region [270] and the design of a multiplex SNP test that uses ASPE (as described by [122]) involving one universal forward and three species-specific reverse primers to demonstrate presence or absence of a particular species [271]. Barcoding of two *COI* regions (175bp and 245bp) was also demonstrated to identify 11 different snake species with 98% to 100% sequence homology for each specific species [272].

1.4.8.2 STR based analyses

Gibbs & Weatherhead [273] review four investigations that used STR markers to examine parentage in colubrid and viperid snakes. All were ecological enquiries, exploring such themes as demographic connectedness of populations and individual reproductive success. To date, STR markers have not been used to study parentage in pythons for forensic purposes.

Technological advances are further aiding STR research in snakes. MPS analysis has been used to identify Copperhead snake STR markers [155] and compare repeat element content between the Copperhead snake and Burmese python. Allozymes have been used to look at parental relationships in rattlesnakes, but for the purpose of identifying artificial insemination rather than verifying breeding records [274].

1.4.8.3 STR markers in closely related species

Historically, research has relied heavily on the use of STR loci characterised previously due to the prohibitive resourcing involved in finding new markers. Many studies have performed cross-species amplification with pre-discovered loci in the hope of finding a variable marker for their own study organism. This has enabled research of many closely related species, but the disadvantages include a higher frequency of null alleles in the species for which the primers were not designed and reduced polymorphism in related species (caused by sampling bias), even though amplification is successful. The emergence of new whole genome sequencing technologies has made STR discovery much quicker, cheaper and more accessible (refer to [155] for further reading). For research groups with greater resources, MPS analysis of each species provides a more efficient means of locating effective STR markers. Until whole genome sequencing prices reduce, cross-species screening during marker development remains a cheaper alternative to provide potential markers in closely related species.

STR markers have been developed and tested for cross-species amplification and polymorphism in rattlesnakes [275–279], adders [280], Australian tiger snakes [281] and Australian Eastern Small Eyed snakes [282]. Jordan, Goodman & Donnellan [283] took this approach when characterising 27 new STR loci in the focal species *Morelia spilota*. When screening STR loci for polymorphism, members of twelve other python species were included along-side the focal species, representing all recognised python genera [255]. With the exception of one species (*Bothrochilus boa*), pythons showed high STR locus conservation (81-95%) across all individuals. Nine STR loci (47%) were conserved between *Morelia spilota* and *Bothrochilus boa*.

Development of STR tests for wildlife forensic applications is restricted by the heavy resource requirement in locating new STR loci, screening for polymorphism, selecting putative markers for population screening and subsequently constructing population allelic frequency databases for each different wildlife species. If the cost of MPS prohibits its use to identify STRs in multiple species, then cross-species testing during the first two stages should be encouraged to increase efficiency in development of these tests for closely related species. The MPS approach does however overcome the disadvantages of cross-species marker screening, such as ascertainment bias.

1.4.8.4 Geographic assignment testing

Taylor [246] further assessed the effectiveness of a geographic assignment technique (developed by [185]) in order to determine the most likely geographic origin of a particular unknown individual. In accordance with the phylogenetic assessment previously discussed, the assignment testing was reliable for geographically isolated mainland and island populations. As this technique requires population divergence at the genetic markers of interest, the limited genetic structuring seen in the northern and eastern populations of *M. spilota* combined with insufficient density of geographic sampling cause poor geographic assignment resolution. Taylor [246] developed a new two-phase assignment methodology which helped to improve the precision of assignment in these populations, but not to a level of reliability necessary for court purposes. His research used eight independent STR loci. The use of more STR loci exhibiting greater numbers of alleles and a denser sampling of snakes from the eastern and northern part of the range will improve the effectiveness of this assignment test to establish a probable geographic origin of an unknown Carpet Python.

1.5. Aim and Objectives of my research

This chapter has demonstrated the need for further development of molecular genetic techniques to aid successful prosecutions of crimes involving snake species. My project aims specifically to develop and validate a range of forensic DNA tests for the carpet python that can be applied directly to forensic casework involving this species.

My PhD project objectives are to:

- develop and validate a forensic species identification assay specific to the family Pythonidae that can differentiate all species of the genus *Morelia* and is suitable for use on degraded forensic samples;

- develop, optimise and validate multiple forensic DNA profiling multiplex assays to survey a marker set of around two dozen highly polymorphic STR loci;

- apply the optimised multiplex assays to screen native carpet pythons across their native geographic range and compile a reference DNA database of carpet python genotypes that can be used to calculate forensic statistics for application to criminal casework involving matching of evidentiary items and questions of parentage;

- assess the ability of the developed dataset to assign a carpet python suspected to be poached from the wild to its most likely geographic origin.

1.6. References

[1] S. McLoughlin, The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism, Aust. J. Bot. 49 (2001) 271–300.

[2] E. Alacs, A. Georges, Wildlife across our borders: a review of the illegal trade in Australia, Aust. J. Forensic Sci. 40 (2008) 147–160.

[3] N. Myers, R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities, Nature. 403 (2000) 853–858.

[4] M. Auliya, S. Altherr, D. Ariano-Sanchez, E.H. Baard, C. Brown, R.M. Brown, J.C. Cantu, G. Gentile, P. Gildenhuys, E. Henningheim, J. Hintzmann, K. Kanari, M. Krvavac, M. Lettink, J. Lippert, L. Luiselli, G. Nilson, T.Q. Nguyen, V. Nijman, J.F. Parham, S.A. Pasachnik, M. Pedrono, A. Rauhaus, D.R. Córdova, M.E. Sanchez, U. Schepp, M. van Schingen, N. Schneeweiss, G.H. Segniagbeto, R. Somaweera, E.Y. Sy, O. Türkozan, S. Vinke, T. Vinke, R. Vyas, S. Williamson, T. Ziegler, Trade in live reptiles, its impact on wild populations, and the role of the European market, Biol. Conserv. 204 (2016) 103–119.

[5] J.D. Heydon, Cross on Evidence, 7th ed., LexisNexis Butterworths, Sydney, 2004.

[6] R.A. Hughes, G.W.G. Leane, A. Clarke, Australian Legal Institutions: principles, structure and organisation, 2nd ed., Lawbook Co., Pyrmont, NSW, 2003.

[7] M.A. Menotti-Raymond, V.A. David, S.J. O'Brien, Pet cat hair implicates murder suspect, Nature. 386 (1997) 774.

[8] J.H. Wetton, J.E. Higgs, A.C. Spriggs, C.A. Roney, C.S.F. Tsang, A.P. Foster, Mitochondrial profiling of dog hairs, Forensic Sci. Int. 133 (2003) 235–241.

[9] Department of the Environment and Energy, Australia's stricter domestic measures, Aust. Gov. (2018). http://www.environment.gov.au/biodiversity/wildlife-trade/cites/stricter-measures (accessed June 3, 2018).

[10] TRAFFIC, Almost 1500 turtles "crammed like sardines" into suitcases, (2012). http://www.traffic.org/home/2012/3/2/almost-1500-turtles-crammed-like-sardines-into-suitcases.html (accessed June 3, 2018).

[11] C. Nellemann, R. Henriksen, P. Raxter, N. Ash, E. Mrema, The environmental crime crisis: threats to sustainable development from illegal exploitation and trade in wildlife and forest resources, United Nations Environment Programme (UNEP), 2014.

[12] N. Ahlers, J. Creecy, G. Frankham, R.N. Johnson, A. Kotze, A. Linacre, R. McEwing, M. Mwale, J.J. Rovie-Ryan, F. Sitam, L.M.I. Webster, "ForCyt" DNA database of wildlife species, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) e466–e468. doi:10.1016/j.fsigss.2017.09.195.

[13] D. Roe, T. Mulliken, S. Milledge, J. Mremi, S. Mosha, M. Grieg-Gran, Making a Killing or Making a Living? Wildlife Trade, Trade Controls & Rural Livelihoods, Biodiversity and Livelihoods Issues No. 6, IIED, London, UK, 2002.

[14] B. Tapley, R.A. Griffiths, I. Bride, Dynamics of the trade in reptiles and amphibians within the United Kingdom over a ten-year period, Herpetol. J. 21 (2011) 27–34.

[15] J. Holden, By Hook or by Crook: A reference manual on illegal wildlife trade and prosecutions in the United Kingdom, 1998.

[16] R. Ogden, N. Dawnay, R. McEwing, Wildlife DNA forensics - Bridging the gap between conservation genetics and law enforcement, Endanger. Species Res. 9 (2009) 179–195.

[17] J. Sellar, Illegal Trade and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), in: A. Linacre (Ed.), Forensic Sci. Wildl. Investig., CRC Press, Taylor & Francis Group, 2009.

[18] A. Chambers, Illegal wildlife trade is far more terrifying than just snakes on a plane, Guard. (2012). https://www.theguardian.com/commentisfree/2011/jul/28/illegal-wildlife-trade-snakes-plane-bangkok (accessed June 3, 2018).

[19] R.N. Johnson, The use of DNA identification in prosecuting wildlife-traffickers in Australia : do the penalties fit the crimes ?, Forensic Sci. Med. Pathol. (2010) 211–216.

[20] J. Williams, C. Read, T. Norton, S. Dovers, M. Burgman, W. Proctor, H. Anderson, Biodiversity, CSIRO Publishing on behalf of the Department of the Environment and Heritage, Canberra, 2001.

[21] R. Ogden, Forensic science, genetics and wildlife biology: Getting the right mix for a wildlife DNA forensics lab, Forensic Sci. Med. Pathol. 6 (2010) 172–179.

[22] Scientific Working Group on DNA Analysis Methods (SWGDAM), Validation Guidelines for DNA Analysis Methods, (2012). https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf (accessed October 6, 2017).

[23] Forensic Science Regulator, Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System, London, UK, 2017. doi:978-1-84987-624-7.

[24] B. Budowle, P. Garofano, A. Hellman, M. Ketchum, S. Kanthaswamy, W. Parson, W. Van Haeringen, S. Fain, T. Broad, Recommendations for animal DNA forensic and identity testing, Int. J. Legal Med. (2005) 295–302.

[25] A. Linacre, L. Gusmao, W. Hecht, A.P. Hellmann, W.R. Mayr, W. Parson, M. Prinz, P.M. Schneider, N. Morling, ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations, Forensic Sci. Int. Genet. 5 (2011) 501–505.

[26] B.C. Yates, E.O. Espinoza, B.W. Baker, Forensic species identification of elephant (Elephantidae) and giraffe (Giraffidae) tail hair using light microscopy, Forensic Sci. Med. Pathol. (2010) 165–171.

[27] V. Sahajpal, S.P. Goyal, Microscopic Examinations in Wildlife Investigations, in: A. Linacre (Ed.), Forensic Sci. Wildl. Investig., CRC Press, Taylor & Francis Group, 2009.

[28] E.O. Espinoza, N.C. Lindley, K.M. Gordon, J.A. Ekhoff, M.A. Kirms, Electrospray Ionization Mass Spectrometric Analysis of Blood for Differentiation of Species, Anal. Biochem. 268 (1999) 252–261.

[29] G.J. Bowen, L.I. Wassenaar, K.A. Hobson, Global application of stable hydrogen and oxygen isotopes to wildlife forensics, Oecologia. 143 (2005) 337–348.

[30] S.S. Tobe, Determining the Geographic Origin of Animal Samples, in: A. Linacre (Ed.), Forensic Sci. Wildl. Investig., CRC Press, Taylor & Francis Group, 2009.

[31] M. Satoh, T. Kuroiwa, Organization of multiple nucleoids and DNA molecules in mitochondria

of a human cell, Exp. Cell Res. 196 (1991) 137–140.

[32] A. Linacre, Forensic Science in Wildlife Investigations, CRC Press, Taylor & Francis Group, 2009.

[33] D.-X. Zhang, G.M. Hewitt, Highly conserved nuclear copies of the mitochondrial con- trol region in the desert locust Schistocerca gregaria: some implications for population studies, Mol. Ecol. 5 (1996) 295–300.

[34] M.W. Gray, Origin and Evolution of Mitochondrial DNA, Annu. Rev. Cell Biol. 5 (1989) 25–50.

[35] C. Saccone, C. De Giorgi, C. Gissi, G. Pesole, A. Reyes, Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system, Gene. 238 (1999) 195–209.

[36] Y. Kumazawa, M. Nishida, Variations in Mitochondrial tRNA Gene Organization of Reptiles as Phylogenetic Markers, Mol. Biol. Evol. 12 (1995) 759–772.

[37] Y. Kumazawa, M. Nishida, T. Ozawa, Gene Rearrangements in Snake Mitochondrial Genomes: Highly Concerted Evolution of Control-Region-Like Sequences Duplicated and Inserted into a tRNA Gene Cluster, Mol. Biol. Evol. 13 (1996) 1242–1254.

[38] Y. Kumazawa, Mitochondrial DNA Sequences of Five Squamates: Phylogenetic Afflilation of Snakes, DNA Res. 11 (2004) 137–144.

[39] J. Yan, H. Li, K. Zhou, Evolution of the mitochondrial genome in snakes: Gene rearrangements and phylogenetic relationships, BMC Genomics. 9 (2008) 1–7.

[40] W.M. Brown, The mitochondrial genome of animals, in: R.J. MacIntyre (Ed.), Monogr. Evol. Biol. Mol. Evol. Genet., Plenum, New York/London, 1985: pp. 95–130.

[41] S. Anderson, A.T. Bankier, B.J. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, I.G. Young, Sequence and Organisation of the human mitochondrial genome, Nature. 290 (1981) 457–465.

[42] S. Anderson, M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger, I.G. Young, Complete Sequence of Bovine Mitochondrial DNA - Conserved features of the mammalian mitochondrial genome, J. Mol. Biol. 156 (1982) 683–717.

[43] M.J. Bibb, R.A. Vanetten, C.T. Wright, M.W. Walberg, D.A. Clayton, Sequence and Gene Organization of Mouse Mitochondrial DNA, Cell. 26 (1981) 167–180.

[44] B.A. Roe, D.P. Ma, R.K. Wilson, J.F.H. Wong, The Complete Nucleotide Sequence of the Xenopus Laevis Mitochondrial Genome, J. Biol. Chem. 260 (1985) 9759–9774.

[45] D.O. Clary, E.D.R. Wolstenholme, The Mitochondrial DNA Molecule of Drosophila Yakuba Nucleotide Sequence, J. Mol. Evol. 22 (1985) 252–271.

[46] T.D. Kocher, W.K. Thomas, A. Meyer, S. V. Edwards, S. Paabo, F.X. Villablanca, A.C. Wilson, Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers., Proc. Natl. Acad. Sci. 86 (1989) 6196–6200.

[47] C.S. Baker, S.R. Palumbi, Which Whales Are Hunted? A Molecular Genetic Approach to Monitoring Whaling, Science (80-.). 265 (1994) 1538–1539.

[48] S.R. Palumbi, F. Cipriano, Species Identification Using Genetic Tools: The Value of Nuclear and Mitochondrial Gene Sequences in Whale Conservation, J. Hered. 89 (1998) 459–464.

[49] A.E. Dizon, C.S. Baker, F. Cipriano, G.M. Lento, P.J. Palsbøll, R.R. Reeves, Molecular genetic identification of whales, dolphins, and porpoises: Proceedings of a workshop on the forensic use of molecular techniques to identify wildlife products in the marketplace, La Jolla, CA, 2000.

[50] L. Hadas, D. Hermon, G.K. Bar-Gal, Before they are gone - Improving gazelle protection using wildlife forensic genetics, Forensic Sci. Int. Genet. 24 (2016) 51–54.

[51] D.M. Irwin, T.D. Kocher, A.C. Wilson, Evolution of the cytochrome b of mammals., J. Mol. Evol. 32 (1991) 128–144.

[52] H.-M. Hsieh, H.-L. Chiang, L.-C. Tsai, S.-Y. Lai, N.-E. Huang, A. Linacre, J.C.-I. Lee, Cytochrome b gene for species identification of the conservation animals, Forensic Sci. Int. 122 (2001) 7–18.

[53] P.D.N. Hebert, A. Cywinska, S.L. Ball, J.R. deWaard, Biological identifications through DNA barcodes, Proc. R. Soc. London B Ser. 270 (2003) 313–321.

[54] B. Brown, R.M. Emberson, A.M. Paterson, Mitochondrial COI and II provide useful markers for Weiseana (Lepidoptera, Hepialidae) species identification, Bull. Entomol. Res. 89 (1999) 287–294.

[55] A. Bucklin, M. Guarnieri, R.S. Hill, A.M. Bentley, S. Kaartvedt, Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR, Hydrobiology. 401 (1999) 239–254.

[56] S.A. Trewick, Mitochondrial DNA sequences support allozyme evidence for cryptic radiation of New Zealand Peripatoides (Onychophora), Mol. Ecol. 9 (2000) 269–282.

[57] O. Folmer, M. BLACK, W. HOEH, R. Lutz, R. Vrijenhoek, DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates, Mol. Mar. Biol. Biotechnol. 3 (1994) 294–299.

[58] S.S. Tobe, A.C. Kitchener, A.M.T. Linacre, Reconstructing mammalian phylogenies: A detailed comparison of the cytochrome b and cytochrome oxidase subunit i mitochondrial genes, PLoS One. 5 (2010). doi:10.1371/journal.pone.0014156.

[59] F. Mayer, C. Dietz, A. Kiefer, Molecular species identification boosts bat diversity, Front. Zool. 4 (2007) 1–5.

[60] B. Balitzki-Korte, K. Anslinger, C. Bartsch, B. Rolf, Species identification by means of pyrosequencing the mitochondrial 12S rRNA gene, Int. J. Legal Med. 119 (2005) 291–294.

[61] T. Melton, C. Holland, Routine forensic use of the mitochondrial 12S ribosomal RNA gene for species identification, J. Forensic Sci. 52 (2007) 1305–1307.

[62] M. Vences, M. Thomas, R.M. Bonett, D.R. Vieites, Deciphering amphibian diversity through DNA barcoding: chances and challenges, Philos. Trans. R. Soc. B Biol. Sci. 360 (2005) 1859–1868.

[63] T. Mitani, A. Akane, T. Tokiyasu, S. Yoshimura, Y. Okii, M. Yoshida, Identification of animal species using the partial sequences in the mitochondrial 16S rRNA gene, Leg. Med. 11 (2009) S449–S450.

[64] C. Ebach, Malte, C. Holdrege, DNA barcoding is no substitute for taxonomy, Nature. 434 (2005) 697.

[65] C. Moritz, C. Cicero, DNA Barcoding: Promise and Pitfalls, PLoS Biol. 2 (2004) 1529–1531. doi:10.1371/journal.pbio.0020354.

[66] J. Waugh, DNA barcoding in animal species: progress, potential and pitfalls, Phylogenetics Syst. 29 (2007) 188–197.

[67] S. Boonseub, S.S. Tobe, A.M.T. Linacre, The use of mitochondrial DNA genes to identify closely related avian species, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 275–277. doi:10.1016/j.fsigss.2009.08.050.

[68] L. Wilson-wilde, J. Norman, J. Robertson, S. Sarre, A. Georges, Australian marsupial species identification, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e543–e544. doi:10.1016/j.fsigss.2011.10.014.

[69] W. Branicki, T. Kupiec, R. Pawlowski, Validation of cytochrome b sequence analysis as a method of species identification, J. Forensic Sci. 48 (2003) 83–87.

[70] N. Dawnay, R. Ogden, R. Mcewing, G.R. Carvalho, R.S. Thorpe, Validation of the barcoding gene COI for use in forensic genetic species identification, Forensic Sci. Int. 173 (2007) 1–6.

[71] M. Aliabadian, M. Kaboli, V. Nijman, M. Vences, Molecular Identification of Birds: Performance of Distance-Based DNA Barcoding in Three Genes to Delimit Parapatric Species, PLoS One. 4 (2009) e4119. doi:10.1371/journal.pone.0004119.

[72] M. Wiemers, K. Fiedler, Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae), Front. 4 (2007) 1–16.

[73] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.. Sayers, Genbank, Nucleic Acids Res. 38 (2010) D46–D51.

[74] M.. Wilson, J.A. Dizinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (1995) 68–74.

[75] M.K. Moore, M. Burnham-Curtis, Genbank: the “Gen” does not stand for “Genuine”., in: Soc. Wildl. Forensic Sci. 1st Trienn. Meet., Jackson Hole, Wyoming, 2012.

[76] S. Ratnasingham, P.D.N. Hebert, BOLD: The Barcode of Life data system (www.barcodinglife.org), Mol. Ecol. Notes. 7 (2007) 355–364.

[77] N. Ahlers, J. Creecy, G. Frankham, R.N. Johnson, A. Kotze, A. Linacre, R. McEwing, M. Mwale, J.J. Rovie-Ryan, F. Sitam, L.M.I. Webster, “ForCyt” DNA database of wildlife species, Forensic Sci. Int. Genet. Suppl. Ser. 6 (2017) E466–E468. doi:10.1016/j.fsigss.2017.09.195.

[78] S. Malik, P.J. Wilson, R.J. Smith, D.M. Lavigne, B.N. White, Pinniped Penises in Trade: A Molecular-Genetic Investigation, Conserv. Biol. 11 (1997) 1365–1374.

[79] H. Wu, Q.-H. Wan, S.-G. Fang, S.-Y. Zhang, Application of mitochondrial DNA sequence analysis in the forensic identification of Chinese sika deer subspecies, Forensic Sci. Int. 148 (2005) 101–105.

[80] C.-F. Lo, Y.-R. Lin, H.-C. Chang, J.-H. Lin, Identification of turtle shell and its preparations by PCR-DNA sequencing method, J. Food Drug Anal. 14 (2006) 153–158.

[81] J. Sanders, J. Cribbs, H. Fienberg, G. Hulburd, L. Katz, S. Palumbi, The tip of the tail: molecular identification of seahorses for sale in apothecary shops and curio stores in California, Conserv. Genet. 9 (2008) 65–71.

[82] H. Hsieh, L. Huang, L. Tsai, Y. Kuo, H. Meng, A. Linacre, J.C. Lee, Species identification of rhinoceros horns using the cytochrome b gene, Forensic Sci. Int. 136 (2003) 1–11.

[83] J.H. Lowenstein, G. Amato, S.-O. Kolokotronis, The Real maccoyii : Identifying Tuna Sushi with DNA Barcodes – Contrasting Characteristic Attributes and Genetic Distances, PLoS One. 4 (2009) e7866. doi:10.1371/journal.pone.0007866.

[84] W. Parson, K. Pegoraro, H. Niederstatter, M. Foger, M. Steinlechner, Species identification by means of the cytochrome b gene, Int. J. Legal Med. 114 (2000) 23–28.

[85] S.K. Verma, L. Singh, Novel universal primers establish identity of an enormous number of animal species for forensic application, Mol. Ecol. Notes. 3 (2003) 28–31.

[86] J. Roman, B.W. Bowen, The mock turtle syndrome: genetic identification of turtle meat purchased in the south-eastern United States of America, Anim. Conserv. 3 (2000) 61–65.

[87] A.O. Karlsson, G. Holmlund, Identification of mammal species using species-specific DNA pyrosequencing, Forensic Sci. Int. 173 (2007) 16–20.

[88] J.C. Lee, H.-M. Hsieh, L.-H. Huang, Y.-C. Kuo, J.-H. Wu, S.-C. Chin, A.-H. Lee, A. Linacre, L. Tsai, Ivory identification by DNA profiling of cytochrome b gene, Int. J. Legal Med. 123 (2009) 117–121.

[89] J.C.-I. Lee, L.-C. Tsai, S.-P. Liao, A. Linacre, H.-M. Hsieh, Species identification using the cytochrome b gene of commercial turtle shells, Forensic Sci. Int. 3 (2009) 67–73.

[90] J.C. Lee, L. Tsai, M. Huang, J. Jhuang, C. Yao, S. Chin, L. Wang, A. Linacre, H. Hsieh, A novel strategy for avian species identification by cytochrome b gene, Electrophoresis. 29 (2008) 2413–2418.

[91] P.B.S. Spencer, D. Schmidt, S. Hummel, Identification of historical specimens and wildlife seizures originating from highly degraded sources of kangaroos and other macropods, Forensic Sci. Med. Pathol. 6 (2010) 225–232.

[92] S.K. Gupta, S.K. Verma, L. Singh, Molecular insight into a wildlife crime: the case of a peafowl slaughter, Forensic Sci. Int. 154 (2005) 214–217.

[93] J. An, M. Lee, M.-S. Min, M.-H. Lee, H. Lee, A molecular genetic approach for species identification of mammals and sex determination of birds in a forensic case of poaching from South Korea, Forensic Sci. Int. 167 (2007) 59–61.

[94] J.C. Lee, L. Tsai, P. Hwa, C. Chan, A. Huang, S. Chin, L. Wang, J. Lin, A. Linacre, H. Hsieh, A novel strategy for avian species and gender identification using the CHD gene, Mol. Cell. Probes. 24 (2010) 27–31.

[95] M.J. Eaton, G.L. Meyers, M.S. Kolokotronis, Sergios-Orestis Leslie, A.P. Martin, G. Amato, Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates, Conserv. Genet. 11 (2010) 1389–1404.

[96] D.L. Dalton, A. Kotze, DNA barcoding as a tool for species identification in three forensic wildlife cases in South Africa, Forensic Sci. Int. 207 (2011) e51–e54.

[97] T. Matsunaga, K. Chikuni, R. Tanabe, S. Muroya, H. Nakai, K. Shibata, J. Yamada, Y. Shinmura, Determination of Mitochondrial Cytochrome B Gene Sequence for Red Deer (Cervus Elaphus) and the Differentiation of Closely Related Deer Meats, Meat Sci. 49 (1998) 379–385.

[98] P.R. Meganathan, B. Dubey, I. Haque, Molecular identification of Indian crocodile species: PCR-RFLP method for forensic authentication, J. Forensic Sci. 54 (2009) 1042–1045.

[99] Y.S. El-sayed, O.I. Mohamed, K.M. Ashry, S.M.A. El-Rahman, Using species-specific repeat and PCR–RFLP in typing of DNA derived from blood of human and animal species, Forensic Sci. Med. Pathol. 6 (2010) 158–164.

[100] A. Malisa, P. Gwakisa, S. Balthazary, S. Wasser, B. Mutayoba, Species and gender differentiation between and among domestic and wild animals using mitochondrial and sex-linked DNA markers, African J. Biotechnol. 4 (2005) 1269–1274.

[101] G. Amato, J. Gatesy, PCR assays of variable nucleotide sites for identification of conservation units, in: Mol. Ecol. Evol. Approaches Appl., Birkhäuser, Basel, 1994: pp. 215–226.

[102] R. DeSalle, V.J. Birstein, PCR identification of black caviar, Nature. 381 (1996) 197–198.

[103] V.J. Birstein, P. Doukakis, B. Sorkin, R. DeSalle, Population Aggregation Analysis of Three Caviar-Producing Species of Sturgeons and Implications for the Species Identification of Black Caviar, Conserv. Biol. 12 (1998) 766–775.

[104] Q.-H. Wan, S.-G. Fang, Application of species-speci ® c polymerase chain reaction in the forensic identi ® cation of tiger species, Forensic Sci. Int. 131 (2003) 75–78.

[105] M. Rojas, I. Gonzalez, M.A. Pavon, N. Pegels, P.E. Hernandez, T. Garcia, R. Martin, Development of a real-time PCR assay to control the illegal trade of meat from protected capercaillie species (Tetrao urogallus), Forensic Sci. Int. 210 (2011) 133–138.

[106] O. Berry, M.D. Tocher, S.D. Sarre, Can assignment tests measure dispersal ?, Mol. Ecol. 13 (2004) 551–561.

[107] R. Oliveria, D. Castro, R. Godinho, G. Luikart, P.C. Alves, Species identification using a small nuclear gene fragment : application to sympatric wild carnivores from South-western Europe, Conserv. Genet. 11 (2010) 1023–1032.

[108] S.K. Gupta, K. Thangaraj, L. Singh, A Simple and Inexpensive Molecular Method for Sexing and Identification of the Forensic Samples of Elephant Origin, J. Forensic Sci. 51 (2006) 805–807.

[109] C. Bellis, K.J. Ashton, L. Freney, B. Blair, L.R. Griffiths, A molecular genetic approach for forensic animal species identification, Forensic Sci. Int. 134 (2003) 99–108.

[110] T. Kitpipit, S.S. Tobe, A.C. Kitchener, P. Gill, A. Linacre, The development and validation of a single SNaPshot multiplex for tiger species and subspecies identification-Implications for forensic purposes, Forensic Sci. Int. 6 (2012) 250–257.

[111] T. Kitpipit, A. Linacre, S.S. Tobe, Tiger species identification based on molecular approach, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 310–312. doi:10.1016/j.fsigss.2009.08.041.

[112] T. Kitpipit, S.S. Tobe, A.C. Kitchener, P. Gill, A. Linacre, The development and validation of a single SNaPshot multiplex for tiger species and subspecies identification — Implications for forensic purposes, Forensic Sci. Int. Genet. 6 (2011) 250–257.

[113] W. Hou, Y. Chen, X. Wu, J. Hu, Z. Peng, J. Yang, Z. Tang, C.-Q. Zhou, Y. Li, S. Yang, Y. Du, L. Kong, Z. Ren, H. Zhang, S. Shuai, A complete mitochondrial genome sequence of Asian black bear Sichuan subspecies (Ursus thibetanus mupinensis), Int. J. Biol. Sci. 3 (2007) 85–90.

[114] M. Shivji, S. Clarke, M. Pank, L. Natanson, N. Kohler, M. Stanhope, Genetic Identification of

Pelagic Shark Body Parts for Conservation and Trade Monitoring, Conserv. Biol. 16 (2002) 1036–1047.

[115] D.D. Chapman, D.L. Abercrombie, C.J. Douady, E.K. Pikitch, M.J. Stanhope, M.S. Shivji, A streamlined, bi-organelle, multiplex PCR approach to species identification: Application to global conservation and trade monitoring of the great white shark, Carcharodon carcharias, Conserv. Genet. 4 (2003) 415–425.

[116] M.S. Shivji, D.D. Chapman, E.K. Pikitch, P.W. Raymond, Genetic profiling reveals illegal international trade in fins of the great white shark , Carcharodon carcharias, Conserv. Genet. 6 (2005) 1035–1039

[117] S.C. Clarke, J.E. Magnussen, D.L. Abercrombie, M.K. Mcallister, M.S. Shivji, Identification of Shark Species Composition and Proportion in the Hong Kong Shark Fin Market Based on Molecular Genetics and Trade Records, Conserv. Biol. 20 (2006) 201–211.

[118] H. Sebastian, P.A. Haye, M.S. Shivji, Characterization of the pelagic shark-fin trade in north-central Chile by genetic identification and trader surveys, J. Fish Biol. 73 (2008) 2293–2304.

[119] J.E. Magnussen, E.K. Pikitch, S.C. Clarke, C. Nicholson, A.R. Hoelzel, M.S. Shivji, Genetic tracking of basking shark products in international trade, Anim. Conserv. 10 (2007) 199–207.

[120] M. Pank, M. Stanhope, L. Natanson, N. Kohler, M. Shivji, Rapid and Simultaneous Identification of Body Parts from the Morphologically Similar Sharks Carcharhinus obscurus and Carcharhinus plumbeus (Carcharhinidae) Using Multiplex PCR, Mar. Biotechnol. 3 (2001) 231–240.

[121] D.L. Abercrombie, S.C. Clarke, M.S. Shivji, Global-scale genetic identification of hammerhead sharks: Application to assessment of the international fin trade and law enforcement, Conserv. Genet. 6 (2005) 775–788.

[122] S.S. Tobe, A.M.T. Linacre, A multiplex assay to identify 18 European mammal species from mixtures using the mitochondrial cytochromeb gene, Electrophoresis. 29 (2008) 340–347.

[123] S.S. Tobe, A. Linacre, A method to identify a large number of mammalian species in the UK from trace samples and mixtures without the use of sequencing, Forensic Sci. Int. Genet. Suppl. Ser. (2008) 625–627. doi:10.1016/j.fsigss.2007.10.147.

[124] S.S. Tobe, A. Linacre, A new assay for identifying endangered species in Traditional East Asian Medicine, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e232–e233. doi:10.1016/j.fsigss.2011.08.115.

[125] A. Ramón-Laca, A.M.T. Linacre, D.M. Gleeson, S.S. Tobe, Identification multiplex assay of 19 terrestrial mammal species present in New Zealand, Electrophoresis. 34 (2013) 3370–3376.

[126] T. Kitpipit, S.S. Tobe, A.C. Kitchener, P. Gill, A. Linacre, Where does this tiger come from ?— A robust molecular technique for simultaneous identification of endangered species and subspecies, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e532–e533. doi:10.1016/j.fsigss.2011.10.011.

[127] T. Kitpipit, K. Thongjued, K. Penchart, K. Ouithavon, W. Chotigeat, Mini-SNaPshot multiplex assays authenticate elephant ivory and simultaneously identify the species origin, Forensic Sci. Int. Genet. 27 (2017) 106–115.

[128] N. Takagi, M. Itoh, M. Sasaki, Chromosome studies in four species of Ratitae (Aves),

Chromosoma. 36 (1972) 281–291.

[129] K.W. Jones, L. Singh, Snakes and the evolution of sex chromosomes, TRENDS Genet. (1985) 55–61.

[130] M.A. Jobling, A. Pandaya, C. Tyler-Smith, The Y chromosome in forensic analysis and paternity testing, Int. J. Legal Med. 110 (1997) 118–124.

[131] P. Gill, C. Brenner, B. Brinkman, B. Budowle, A. Carracedo, M.A. Jobling, P. De Knijff, M. Kayser, M. Krawczak, W.R. Mayr, N. Morling, B. Olaisen, V. Pascali, M. Prinz, L. Roewer, P.M. Schneider, A. Sajantila, C. Tyler-Smith, DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, Forensic Sci. Int. 124 (2001) 5–10.

[132] L. Roewer, Y chromosome STR typing in crime casework, Forensic Sci. Med. Pathol. 5 (2009) 77–84.

[133] J.M. Butler, Forensic DNA Typing: biology, technology and genetics of STR markers, Elsevier Academic Press, 2005.

[134] D.P. Snustad, M.J. Simmons, Principles of Genetics, Wiley, 2000.

[135] W. Goodwin, A. Linacre, S. Hadi, An Introduction to Forensic Genetics, John Wiley & Sons, 2011.

[136] W. Britten, D.E. Kohne, Repeated Sequences in DNA, Science (80-.). 161 (1968).

[137] Y. Nakamura, M. Leppert, P. OConnell, P. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, R. White, Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping, Science (80-.). 235 (1987) 1616–1622.

[138] G.K. Chambers, E.S. MacAvoy, Microsatellites: consensus and controversy, Comp. Biochem. Physiol. Part B. 126 (2000) 455–476.

[139] J. Jurka, V. V Kapitonov, O. Kohany, M. V Jurka, Repetitive Sequences in Complex Genomes: Structure and Evolution, Annu. Rev. Genomics Hum. Genet. 8 (2007) 241–259.

[140] A.J. Jeffreys, V. Wilson, S.L. Thein, Hypervariable “minisatellite” regions in human DNA, Nature. 314 (1985) 67–73.

[141] A.J. Jeffreys, V. Wilson, S.L. Thein, Individual-specific “fingerprints” of human DNA, Nature. 316 (1985) 76–79.

[142] P. Gill, A.J. Jeffreys, J. Werrett, David, Forensic Applications of DNA “fingerprints,” Nature. 318 (1985) 577–579.

[143] K.A. Selkoe, R.J. Toonen, Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers, Ecol. Lett. 9 (2006) 615–629.

[144] D.B. Goldstein, C. Schlotterer, Microsatellites: Evolution and Applications, Oxford University Press, 1999.

[145] E. Guichoux, L. Lagache, S. Wagner, P. Chaumeil, P. Leger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin, R.J. Petit, Current trends in microsatellite genotyping, Mol. Ecol. Resour. 11 (2011) 591–611.

[146] M. Litt, J.A. Luty, A Hypervariable Microsatellite Revealed by In Vitro Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene, Am. J. Hum. Genet. 44 (1989) 397–401.

[147] M.G. Gardner, A.J. Fitch, T. Bertozzi, A.J. Lowe, Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development, Mol. Ecol. Resour. 11 (2011) 1093–1101.

[148] G. Levinson, G.A. Gutman, Slipped-strand mispairing: a major mechanism for DNA sequence evolution, Mol. Biol. Evol. 4 (1987) 203–221.

[149] Y.-C. Li, A.B. Korol, T. Fahima, A. Beiles, E. Nevo, Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, Mol. Ecol. 11 (2002) 2453–2465.

[150] T. Ohta, M. Kimura, A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population, Genet. Res. 22 (1973) 201–204.

[151] P.S. Walsh, N.J. Fildes, R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA, Nucleic Acids Res. 24 (1996) 2807–2812.

[152] X.Y. Hauge, M. Litt, A study of the origin of “shadow bands” seen when typing dinucleotide repeat polymorphisms by the PCR, Hum. Mol. Genet. 2 (1993) 411–415.

[153] P. Gill, B. Brinkmann, E. D’Aloja, J. Andersen, W. Bar, A. Carracedo, B. Dupuy, B. Eriksen, M. Jangblad, V. Johnsson, A.D. Kloosterman, P. Lincoln, N. Morling, S. Rand, M. Sabatier, R. Scheithauer, P. Schneider, M.C. Vide, Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature, Forensic Sci. Int. 87 (1997) 185–192.

[154] E. Buschiazzo, N.J. Gemmell, The rise, fall and renaissance of microsatellites in eukaryotic genomes, BioEssays. 28 (2006) 1040–1050.

[155] T.A. Castoe, A.W. Poole, W. Gu, A.P.J. De Koning, J.M. Daza, E.N. Smith, D.D. Pollock, Rapid identification of thousands of copperhead snake (Agkistrodon contortrix) microsatellite loci from modest amounts of 454 shotgun genome sequence, Mol. Ecol. Resour. 10 (2010) 341–347.

[156] M. Kayser, R. Kittler, A. Erler, M. Hedman, A.C. Lee, A. Mohyuddin, S.Q. Mehdi, Z. Rosser, M. Stoneking, M.A. Jobling, A. Sajantila, C. Tyler-smith, A Comprehensive Survey of Human Y-Chromosomal Microsatellites, Am. J. Hum. Genet. 74 (2004) 1183–1197.

[157] D.C. Queller, J.E. Strassmann, C.R. Hughes, Microsatellites and Kinship, TRENDS Ecol. Evol. 8 (1993) 285–288.

[158] S.K. Gupta, J. Bhagavatula, K. Thangaraj, L. Singh, Establishing the identity of the massacred tigress in a case of wildlife crime, Forensic Sci. Int. Genet. 5 (2011) 74–75.

[159] R. Ogden, Fisheries forensics : the use of DNA tools for improving compliance , traceability and enforcement in the fishing industry, (2008) 462–472.

[160] L.P. Waits, G. Luikart, P. Taberlet, Estimation the probability of identity among genotypes in natural populations: cautions and guidelines, Mol. Ecol. 10 (2001) 249–256.

[161] S. Wahlund, The combination of populations and the appearance of correlation examined from the standpoint of the study of heredity, Hereditas. 11 (1928) 65–106.

[162] S. Wright, The genetical structure of populations, Ann. Eugen. 15 (1951) 323–354.

[163] K.L. Ayres, A.D.J. Overall, API-CALC 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives, Mol. Ecol. Notes. 4 (2004) 315–318.

[164] R. Ogden, DNA profiling markers in wildlife forensic science, in: A. Linacre (Ed.), Forensic Sci. Wildl. Investig., CRC Press, Taylor & Francis Group, 2009.

[165] P. Timms, J. Kato, M. Maugeri, N. White, DNA Fingerprint Analysis of a Free-Range Koala Population, 31 (1993) 363–374.

[166] M. Poetsch, S. Seefeldt, M. Maschke, E. Lignitz, Analysis of microsatellite polymorphism in red deer, roe deer, and fallow deer–possible employment in forensic applications, Forensic Sci. Int. 116 (2001) 1–8.

[167] R.M. Jobin, D. Patterson, Y. Zhang, DNA typing in populations of mule deer for forensic use in the Province of Alberta, Forensic Sci. Int. Genet. 2 (2008) 190–197.

[168] E. Socratous, E. a. M. Graham, G.N. Rutty, Forensic DNA profiling of Cervus elaphus species in the United Kingdom, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 281–282. doi:10.1016/j.fsigss.2009.08.127.

[169] S. Caratti, L. Rossi, B. Sona, S. Origlia, S. Viara, G. Martano, C. Torre, C. Robino, Analysis of 11 tetrameric STRs in wild boars for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 339–342.

[170] C.R. Primmer, M.T. Koskinen, J. Piironen, The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud, Proc. R. Soc. London. Ser. B Biol. Sci. 267 (2000) 1699–1704.

[171] H.G. Eiken, R.J. Andreassen, A. Kopatz, S.G. Bjervamoen, I. Wartiainen, C. Tobiassen, P.M. Knappskog, P.E. Aspholm, M.E. Smith, J. Aspi, Population data for 12 STR loci in Northern European brown bear (Ursus arctos) and application of DNA profiles for forensic casework, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 273–274. doi:10.1016/j.fsigss.2009.07.007.

[172] S. Mondol, N. R, V. Athreya, K. Sunagar, V. Selvaraj, U. Ramakrishnan, A panel of microsatellites to individually identify leopards and its application to leopard monitoring in human dominated landscapes, BMC Genet. 10 (2009) 79.

[173] N. Dawnay, R. Ogden, R.S. Thorpe, L.C. Pope, D.A. Dawson, R. McEwing, A forensic STR profiling system for the Eurasian badger: A framework for developing profiling systems for wildlife species, Forensic Sci. Int. Genet. 2 (2008) 47–53.

[174] R. Ogden, N. Dawnay, R. McEwing, Development of STR profiling systems for individual identification in wildlife: A case study of the Eurasian badger, Meles meles, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 612–613. doi:10.1016/j.fsigss.2007.10.224.

[175] M. Wesselink, I. Kuiper, Individual identification of fox (Vulpes vulpes) in forensic wildlife investigations, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e214–e215. doi:10.1016/j.fsigss.2011.08.107.

[176] R. Ogden, R.J. Mellanby, D. Clements, A.G. Gow, R. Powell, R. McEwing, Genetic data from 15 STR loci for forensic individual identification and parentage analyses in UK domestic dogs (Canis lupus familiaris), Forensic Sci. Int. Genet. 6 (2012) e63–e65.

[177] Y.C. Xu, B. Li, W.S. Li, S.Y. Bai, Y. Jin, X.P. Li, M.B. Gu, S.Y. Jing, W. Zhang, Individualization of

tiger by using microsatellites, Forensic Sci. Int. 151 (2005) 45–51.

[178] P. Gill, L. Ivanov, Pavel, C.P. Kimpton, R. Piercy, N. Benson, G. Tully, I.W. Evett, E. Hagelberg, K. Sullivan, Identification of the remains of the Romanov family by DNA analysis, Nat. Genet. 6 (1994) 130–135.

[179] L. Ivanov, Pavel, M.J. Wadhams, R.K. Roby, M. Holland, Mitchell, V.W. Weedn, T.J. Parsons, Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgi Romanov establishes the authenticity of the remains of Tsar Nicholas II, Nat. Genet. 12 (1996) 417–420.

[180] T. Burke, M.W. Bruford, DNA fingerprinting in birds, Nature. 327 (1987) 149–152.

[181] N. Dawnay, R. Ogden, J.H. Wetton, R.S. Thorpe, R. McEwing, Genetic data from 28 STR loci for forensic individual identification and parentage analyses in 6 bird of prey species, Forensic Sci. Int. Genet. 3 (2009).

[182] M.A. Renshaw, E. Saillant, R.E. Broughton, J.R. Gold, Application of hypervariable genetic markers to forensic identification of “wild” from hatchery-raised red drum, Sciaenops ocellatus, Forensic Sci. Int. 156 (2006) 9–15.

[183] F. Rendo, M. Iriondo, C. Manzano, A. Estonba, Microsatellite based ovine parentage testing to identify the source responsible for the killing of an endangered species, Forensic Sci. Int. Genet. 5 (2011) 333–335.

[184] R.A. Cocciolone, P. Timss, DNA Profiling of Queensland Koalas reveals Sufficient Variability for Individual Identification and Parentage Determination, Wildl. Res. 19 (1992) 279–287.

[185] S.K. Wasser, A.M. Shedlock, K. Comstock, E. a Ostrander, B. Mutayoba, M. Stephens, Assigning African elephant DNA to geographic region of origin: applications to the ivory trade., Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 14847–14852.

[186] S. Manel, O.E. Gaggiotti, R.S. Waples, Assignment methods: matching biological questions with appropriate techniques, TRENDS Ecol. Evol. 20 (2005) 136–142.

[187] A.C. Frantz, J.T. Pourtois, M. Heuertz, L. Schley, M.C. Flamand, A. Krier, S. Bertouille, F. Chaumont, T. Burke, Genetic structure and assignment tests demonstrate illegal translocation of red deer (Cervus elaphus) into a continuous population, Mol. Ecol. 15 (2006) 3191–3203.

[188] P.B.S. Spencer, J.O. Hampton, Illegal translocation and genetic structure of feral pigs in Western Australia, J. Wildl. Manage. 69 (2005) 377–384.

[189] M.J.M. Smulders, W.P.C. Van ’T Westende, B. Diway, G.D. Esselink, P.J. Van Der Meer, W.J.M. Koopman, Development of microsatellite markers in Gonystylus bancanus (Ramin) useful for tracing and tracking of wood of this protected species, Mol. Ecol. Resour. 8 (2008) 168–171.

[190] E. White, J. Hunter, C. Dubetz, R. Brost, A. Bratton, S. Edes, R. Sahota, Microsatellite markers for individual tree genotyping: application in forest crime prosecutions, J. Chem. Technol. Biotechnol. 75 (2000) 923–926.

[191] C. Jolivet, B. Degen, Use of DNA fingerprints to control the origin of sapelli timber (Entandrophragma cylindricum) at the forest concession level in Cameroon, Forensic Sci. Int. Genet. 6 (2012) 487–493.

[192] E.E. Dormontt, M. Boner, B. Braun, G. Breulmann, B. Degen, E. Espinoza, S. Gardner, P. Guillery, J.C. Hermanson, G. Koch, S.L. Lee, M. Kanashiro, A. Rimbawanto, D. Thomas, A.C.

Wiedenhoeft, Y. Yin, J. Zahnen, A.J. Lowe, Forensic timber identification: It's time to integrate disciplines to combat illegal logging, Biol. Conserv. 191 (2015) 790–798.

[193] S.K. Wasser, C. Mailand, R. Booth, B. Mutayoba, E. Kisamo, B. Clark, M. Stephens, Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 4228–4233.

[194] S.K. Wasser, W. Joseph Clark, O. Drori, E. Stephen Kisamo, C. Mailand, B. Mutayoba, M. Stephens, Combating the illegal trade in African elephant ivory with DNA forensics, Conserv. Biol. 22 (2008) 1065–1071.

[195] P.A. Morin, G. Luikart, R.K. Wayne, workshop group SNP, SNPs in ecology, evolution and conservation, TRENDS Ecol. Evol. 19 (2004) 208–216.

[196] E.A. Alacs, A. Georges, N.N. Fitzsimmons, J. Robertson, DNA detective: a review of molecular approaches to wildlife forensics, Forensic Sci. Med. Pathol. 6 (2010) 180–194.

[197] A. Singh, A. Gaur, K. Shailaja, B.S. Bala, L. Singh, A novel microsatellite (STR) marker for forensic identification of big cats in India, Forensic Sci. Int. 141 (2004) 143–147.

[198] M.R. McCusker, I.G. Paterson, P. Bentzen, Microsatellite markers discriminate three species of North Atlantic wolffishes (Anarhichas spp.), J. Fish Biol. 72 (2008) 375–385.

[199] J.M. Pemberton, J. Slate, D.R. Bancroft, J.A. Barrett, Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies, Mol. Ecol. 4 (1995) 249–252.

[200] R.J. Haasl, B.A. Payseur, Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites, Heredity (Edinb). 106 (2011) 158–171.

[201] R. Chakraborty, D.N. Stivers, B. Su, Y. Zhong, B. Budowle, The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems, Electrophoresis. 20 (1999) 1682–1696.

[202] Special Issue: SNP Development in Non-Model Organisms, Mol. Ecol. Resour. March (2011).

[203] U. Ramakrishnan, J.L. Mountain, Precision and Accuracy of Divergence Time Estimates from STR and SNPSTR Variation, Mol. Biol. Evol. 21 (2004) 1960–1971.

[204] M. Grimaldi, B. Crouau-roy, Microsatellite Allelic Homoplasy Due to Variable Flanking Sequences, J. Mol. Evol. 44 (1997) 336–340.

[205] M.L. Coghlan, J. Haile, J. Houston, D.C. Murray, N.E. White, P. Moolhuijzen, M.I. Bellgard, M. Bunce, Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns, PLoS Genet. 8 (2012). doi:10.1371/journal.pgen.1002657.

[206] R. Lorenzini, DNA forensics and the poaching of wildlife in Italy: A case study, Forensic Sci. Int. 153 (2005) 218–221.

[207] G. Gentile, M. Ciambotta, W. Tapia, Illegal wildlife trade in Galápagos: Molecular tools help the taxonomic identification of confiscated iguanas and guide their rapid repatriation, Conserv. Genet. Resour. 5 (2013) 867–872.

[208] R. Lorenzini, P. Cabras, R. Fanelli, G.L. Carboni, Wildlife molecular forensics: Identification of the Sardinian mouflon using STR profiling and the Bayesian assignment test, Forensic Sci. Int. Genet. 5 (2011) 345–349.

[209] F. Barbanera, M. Guerrini, C. Beccani, G. Forcina, P. Anayiotos, P. Panayides, Conservation of endemic and threatened wildlife: Molecular forensic DNA against poaching of the Cypriot mouflon (Ovis orientalis ophion, Bovidae), Forensic Sci. Int. Genet. 6 (2012) 671–675.

[210] R. Caniglia, E. Fabbri, C. Greco, M. Galaverni, E. Randi, Forensic DNA against wildlife poaching: Identification of a serial wolf killing in Italy, Forensic Sci. Int. Genet. 4 (2010) 334–338.

[211] N.E. White, R. Dawson, M.L. Coghlan, S.R. Tridico, P.R. Mawson, J. Haile, M. Bunce, Application of STR markers in wildlife forensic casework involving Australian black-cockatoos (Calyptorhynchus spp.), Forensic Sci. Int. Genet. 6 (2012) 664–670.

[212] D. Mell, J. Wetherall, To catch a thief, Landscope. 7 (1992) 28.

[213] F. Burragato, S. Materazzi, R. Curini, G. Ricci, New forensic tool for the identification of elephant or mammoth ivory, Forensic Sci. Int. 96 (1998) 189–196.

[214] N.A. Sweijd, R.C.K. Bowie, B.S. Evans, A.L. Lopata, Molecular genetics and the management and conservation of marine organisms, Hydrobiologia. 420 (2000) 153–164.

[215] E.A. Guglich, P.J. Wilson, B.N. White, Forensic Application of Repetitive DNA Markers to the Species Identification of Animal Tissues, J. Forensic Sci. 39 (1994) 353–361.

[216] M. Jenkins, S. Broad, International Trade in Reptile Skins: A Review and Analysis of the Main Consumer Markets, 1983-1991, 1994.

[217] M.A. Schlaepfer, C. Hoover, C.K.D. Jr, Challenges in Evaluating the Impact of the Trade in Amphibians and Reptiles on Wild Populations, 55 (2005).

[218] C. Hoover, Amazon tree boas to Zululand dwarf chameleons, the US role in the international live reptile trade, 1999.

[219] J.E. Robinson, R.A. Griffiths, F.A. V St. John, D.L. Roberts, Dynamics of the global trade in live reptiles: Shifting trends in production and consequences for sustainability, Biol. Conserv. 184 (2015) 42–50.

[220] F.C.F. Yau, K.L. Wong, P.C. Shaw, P.P.H. But, J. Wang, Authentication of snakes used in Chinese medicine by sequence characterized amplified region (SCAR), Biodivers. Conserv. 11 (2002) 1653–1662.

[221] Z.H. Zhou, Z.G. Jiang, International trade status and crisis for snake species in China, Conserv. Biol. 18 (2004) 1386–1394.

[222] Z. Zhou, Z. Jiang, Identifying snake species threatened by economic exploitation and international trade in China, Biodivers. Conserv. 14 (2005) 3525–3536.

[223] R.R. da N. Alves, W.L. da Silva Vieira, G.G. Santana, Reptiles used in traditional folk medicine: conservation implications, Biodivers. Conserv. 17 (2008) 2037–2049.

[224] R.R.N. Alves, I.L. Rosa, Trade of Animals Used in Brazilian Traditional Medicine: Trends and Implications for Conservation, Hum. Ecol. 38 (2010) 691–704.

[225] K.L. Wong, J. Wang, P.P.H. But, P.C. Shaw, Application of cytochrome b DNA sequences for the authentication of endangered snake species, Forensic Sci. Int. 139 (2004) 49–55.

[226] R.R. da N. Alves, G.A.P. Filho, Commercialization and use of snakes in North and Northeastern Brazil: implications for conservation and management, Biodivers. Conserv. 16 (2007) 969–

985.

[227] R.R. da N. Alves, G.A.P. Filho, Y.C.C. Delima, Snakes used in ethnomedicine in northeast brazil, Environ. Dev. Sustain. 9 (2007) 455–464.

[228] R. Somaweera, N. Somaweera, Serpents in jars: the snake wine industry in Vietnam, J. Threat. Taxa. 2 (2010) 1251–1260.

[229] M.J. Whiting, V.L. Williams, T.J. Hibbitts, Animals traded for traditional medicine at the Faraday market in South Africa: species diversity and conservation implications, J. Zool. 284 (2011) 84–96.

[230] N.S. Sodhi, L.P. Koh, B.W. Brook, P.K.L. Ng, Southeast Asian biodiversity: an impending disaster, TRENDS Ecol. Evol. 19 (2004) 654–660.

[231] J.S. Keogh, D.G. Barker, R. Shine, Heavily exploited but poorly known: systematics and biogeography of commercially harvested pythons (Python curtus group) in Southeast Asia, Biol. J. Linn. Soc. 73 (2001) 113–129.

[232] J. Whitfield Gibbons, D.E. Scott, T.J. Ryan, K.A. Buhlmann, T.D. Tuberville, B.S. Metts, J.L. Greene, T. Mills, Y. Leiden, S. Poppy, C.T. Winne, The Global Decline of Reptiles, Deja Vu Amphibians, Bioscience. 50 (2000) 653–666.

[233] J.K. Webb, B.W. Brook, R. Shine, Collectors endanger Australia's most threatened snake, the broad-headed snake Hoplocephalus bungaroides, Oryx. 36 (2002) 170–181.

[234] R. Shine, The biology and management of the Diamond Python (Morelia spilota spilota) and Carpet Python (M. s. variegata) in NSW, Sydney, 1994.

[235] M. Auliya, Hot Trade in Cool Creatures: A Review of the Live Reptile Trade in the European Union in the 1990s with a focus on Germany, Brussels, Belgium, 2003.

[236] J.B. Slowinski, R. Lawson, Snake phylogeny: evidence from nuclear and mitochondrial genes, Mol. Phylogenet. Evol. 24 (2002) 194–202.

[237] A.-L. Ferchaud, A. Lyet, M. Cheylan, V. Arnal, J.-P. Baron, C. Montgelard, S. Ursenbacher, High Genetic Differentiation Among French Populations of the Orsini's Viper (Vipera ursinii ursinii) Based on Mitochondrial and Microsatellite Data: Implications for Conservation Management, J. Hered. 102 (2011) 67–78.

[238] T.A. Castoe, J.A.P. de Koning, K.T. Hall, K.D. Yokoyama, W. Gu, E.N. Smith, C. Feschotte, P. Uetz, D.A. Ray, J. Dobry, R. Bogden, S.P. Mackessy, A.M. Bronikowski, W.C. Warren, S.M. Secor, D.D. Pollock, Sequencing the genome of the Burmese python (Python molurus bivittatus) as a model for studying extreme adaptations in snakes., Genome Biol. 12 (2011) 406.

[239] F.J. Vonk, N.R. Casewell, C. V Henkel, A.M. Heimberg, H.J. Jansen, R.J.R. McCleary, H.M.E. Kerkkamp, R.A. Vos, I. Guerreiro, J.J. Calvete, W. Wüster, A.E. Woods, J.M. Logan, R.A. Harrison, T.A. Castoe, A.P.J. de Koning, D.D. Pollock, M. Yandell, D. Calderon, C. Renjifo, R.B. Currier, D. Salgado, D. Pla, L. Sanz, A.S. Hyder, J.M.C. Ribeiro, J.W. Arntzen, G.E.E.J.M. van den Thillart, M. Boetzer, W. Pirovano, R.P. Dirks, H.P. Spaink, D. Duboule, E. McGlinn, R.M. Kini, M.K. Richardson, The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system., Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 20651–6.

[240] T.A. Castoe, A.P.J. de Koning, K.T. Hall, D.C. Card, D.R. Schield, M.K. Fujita, R.P. Ruggiero, J.F.

Degner, J.M. Daza, W. Gu, J. Reyes-Velasco, K.J. Shaney, J.M. Castoe, S.E. Fox, A.W. Poole, D. Polanco, J. Dobry, M.W. Vandewege, Q. Li, R.K. Schott, A. Kapusta, P. Minx, C. Feschotte, P. Uetz, D.A. Ray, F.G. Hoffmann, R. Bogden, E.N. Smith, B.S.W. Chang, F.J. Vonk, N.R. Casewell, C. V. Henkel, M.K. Richardson, S.P. Mackessy, A.M. Bronikowski, M. Yandell, W.C. Warren, S.M. Secor, D.D. Pollock, The Burmese python genome reveals the molecular basis for extreme adaptation in snakes, Proc. Natl. Acad. Sci. 110 (2013).

[241] F.T. Burbrink, R. Lawson, J.B. Slowinski, Mitochonrial DNA Phylogeography of the Polytypic North American Rat Snake (Elaphe Obsoleta): A Critique of the Subspecies Concept, Evolution (N. Y). 54 (2000) 2107–2118.

[242] D.R. Frost, D.M. Hillis, Species in concept and practice: herpetological applications, Herepetologica. 46 (1990) 87–104.

[243] D.R. Frost, A.G. Kluge, A consideration of epistemology in systematic biology, with special references to species, Cladistics. 10 (1994) 259–294.

[244] E. Mayr, Of what use are subspecies?, Auk. 99 (1982) 593–595.

[245] M.C. McKitrick, R.M. Zink, Species concepts in ornithology, Condor. 90 (1988) 1–14.

[246] D. Taylor, Using DNA markers for wildlife management and protection: a study of the population structure and systematics of the Australian carpet pythons (Reptilia: Morelia spilota complex), Flinders University, SA, 2005.

[247] K. De Queiroz, Ernst Mayr and the modem concept of species, (2005). doi:10.1073/pnas.0502030102.

[248] G. Underwood, A.F. Stimson, A classification of pythons (Serpentes, Pythoninae), J. Zool. 221 (1990) 565–603.

[249] A.G. Kluge, Aspidites and the phylogeny of pythonine snakes, Rec. Aust. Museum. (1993) 0–1.

[250] T.P. Wilcox, D.J. Zwickl, T.A. Heath, D.M. Hillis, Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support, Mol. Phylogenet. Evol. 25 (2002) 361–371.

[251] R. Lawson, J.B. Slowinski, F.T. Burbrink, A molecular approach to discerning the phylogenetic placement of the enigmatic snake Xenophidion schaeferi among the Alethinophidia, J. Zool. Soc. London. 263 (2004) 285–294.

[252] N. Vidal, S.B. Hedges, Molecular evidence for a terrestrial origin of snakes, Proc. R. Soc. London B Ser. 271 (2004) S226–S229.

[253] B.P. Noonan, P.T. Chippindale, Dispersal and vicariance: The complex evolutionary history of boid snakes, Mol. Phylogenet. Evol. 40 (2006) 347–358.

[254] L.H. Rawlings, D.L. Rabosky, S.C. Donnellan, M.N. Hutchinson, Python phylogenetics: inference from morphology and mitochondrial DNA, Biol. J. Linn. Soc. 93 (2008) 603–619.

[255] W.D. Schleip, M.O. Shea, Annotated checklist of the recent and extinct pythons (Serpentes, Pythonidae), with notes on nomenclature, taxonomy, and distribution, Zookeys. 66 (2010) 29–79.

[256] G.F. Gow, Python bredli, new species from central Australia, Aust. J. Herpetol. 1 (1984) 29–34.

[257] D.G. Barker, T.M. Barker, Pythons of the World, in: Advanced Vivarium Systems Inc, California, 1994.

[258] C.B. Banks, Pythons of the Genus Morelia, ASRA. 1 (1979) 37–41.

[259] C.C. Austin, M. Spataro, S. Peterson, J. Jordan, J.D. McVay, Conservation genetics of Boelen's python (Morelia boeleni) from New Guinea: reduced genetic diversity and divergence of captive and wild animals, Conserv. Genet. 11 (2010) 889–896.

[260] J.A. Lyons, D.J.D. Natusch, Wildlife laundering through breeding farms: Illegal harvest, population declines and a means of regulating the trade of green pythons (Morelia viridis) from Indonesia, Biol. Conserv. 144 (2011) 3073–3081.

[261] L.H. Rawlings, S.C. Donnellan, Phylogeographic analysis of the green python, Morelia viridis, reveals cryptic diversity, Mol. Phylogenet. Evol. 27 (2003) 36–44.

[262] R.W. Wells, C.R. Wellington, A synopsis of the class Reptilia in Australia, Aust. J. Herpetol. 1 (1984) 73–129.

[263] R.W. Wells, C.R. Wellington, A classification of the Amphibia and Reptilia of Australia, Aust. J. Herpetol. Supplement (1985) 1–61.

[264] D.J. Slip, R. Shine, Thermoregulation of Free-ranging Diamond Pythons, Morelia spilota (Serpentes, Boidae), Copeia. 4 (1988) 984–995.

[265] D. Pearson, R. Shine, R. How, Sex-Specific niche partitioning and sexual size dimorphism in Australian pythons (Morelia spilota imbricata), Biol. J. Linn. Soc. 77 (2002) 113–125.

[266] R. Shine, M. Fitzgerald, Variation in mating systems and sexual size dimorphism between populations of the Australian python Morelia spilota (Sepentes: Pythonidae), Oecologia. 103 (1995) 490–498.

[267] D.J.D. Natusch, J.F. Carter, P.W. Aust, N. Van Tri, U. Tinggi, Mumpuni, A. Riyanto, J.A. Lyons, Serpent's source: Determining the source and geographic origin of traded python skins using isotopic and elemental markers, Biol. Conserv. 209 (2017) 406–414.

[268] W. Wuster, A.J. Dumbrell, C. Hay, C.E. Pook, D.J. Williams, B.G. Fry, Snakes across the Strait: trans-Torresian phylogeographic relationships in three genera of Australasian snakes (Serpentes: Elapidae: Acanthophis, Oxyuranus, and Pseudechis), Mol. Phylogenet. Evol. 34 (2005) 1–14.

[269] D.J. Williams, M.O. Shea, R.L. Daguerre, C.E. Pook, W. Wüster, C.J. Hayden, J.D. Mcvay, O. Paiva, T.-L. Matainaho, K.D. Winkel, C.C. Austin, Origin of the eastern brownsnake, Pseudonaja textilis (Dumeril, Bibron and Demeril) (Serpentes: Elapidae: Hydrophiinae) in New Guinea: evidence of multiple dispersals from Australia, and comments on the status of Pseudonaja textilis pughi Hoser 2003, Zootaxa. 1703 (2008) 47–61.

[270] B. Dubey, P.R. Meganathan, I. Haque, Molecular identification of three Indian snake species using simple PCR-RFLP method, J. Forensic Sci. 55 (2010) 1065–1067.

[271] B. Dubey, P.R. Meganathan, I. Haque, Multiplex PCR assay for rapid identification of three endangered snake species of India, Conserv. Genet. 10 (2009) 1861–1864.

[272] B. Dubey, P.R. Meganathan, I. Haque, DNA mini-barcoding: An approach for forensic identification of some endangered Indian snake species, Forensic Sci. Int. 5 (2011) 181–184.

[273] H.L. Gibbs, P.J. Weatherhead, Insights into Population Ecology and Sexual Selection in Snakes Through the Application of DNA-Based Genetic Markers, J. Hered. 92 (2001) 173–179.

[274] C. Crabtree, R. Murphy, Analysis of Maternal-Offspring Allozymes in Crotalus viridis, J. Herpetol. 18 (1984) 75–80.

[275] C.D. Anderson, Utility of a set of microsatellite primers developed for the massasauga rattlesnake (Sistrurus catenatus) for population genetic studies of the timber rattlesnake (Crotalus horridus), Mol. Ecol. Notes. 6 (2006) 514–517.

[276] X. Villarreal, J. Bricker, H.K. Reinert, L. Gelbert, L.M. Bushar, Isolation and Characterization of Microsatellite Loci for Use in Population Genetic Analysis in the Timber Rattlesnake , Crotalus horridus, J. Hered. 87 (1996) 152–155.

[277] L.M. Bushar, M. Maliga, H.K. Reinert, Cross-species amplification of Crotalus horridus microsatellites and their application in Phylogenetic Analysis, J. Herpetol. 35 (2001) 532–537.

[278] A.T. Holycross, M.E. Douglas, J.R. Higbee, R.H. Bogden, Isolation and characterization of microsatellite loci from a threatened rattlesnake (New Mexico Ridge-nosed Rattlesnake, Crotalus willardi obscurus), Mol. Ecol. Notes. 2 (2002) 537–539.

[279] C.S. Goldberg, T. Edwards, M.E. Kaplan, M. Goode, PCR primers for microsatellite loci in the tiger rattlesnake (Crotalus tigris, Viperidae), Mol. Ecol. Notes. 3 (2003) 539–541.

[280] M. Carlsson, M. Isaksson, M. Hoggren, H. Tegelstrom, Characterization of polymorphic microsatellite markers in the adder, Vipera berus, Mol. Ecol. Notes. 3 (2003) 73–75.

[281] I.A.W. Scott, C.M. Hayes, J.S. Keogh, J.K. Webb, Isolation and characterization of novel microsatellite markers from the Australian tiger snakes (Elapidae: *Notechis*) and amplification in the closely related genus *Hoplocephalus*, Mol. Ecol. Notes. 1 (2001) 117–119.

[282] J. Stapley, C.M. Hayes, J.K. Webb, J.S. Keogh, Novel microsatellite loci identified from the Australian eastern small-eyed snake (Elapidae: Rhinocephalus nigrescens) and cross species amplification in the related genus Suta, Mol. Ecol. Notes. 5 (2005) 54–56.

[283] P. Jordan, A. Goodman, S. Donnellan, Microsatellite primers for Australian and New Guinean pythons isolated with an efficient marker development method for related species, Mol. Ecol. Notes. 2 (2002) 78–82.

Appendix 1: Publication - Current Issues with the Investigation of Wildlife Crime in Australia: Problems and Opportunities for Improvement

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

CHAPTER 2

Forensic Species Identification Assay Development & Validation:

a bottom up approach demonstrated using serpents

Appendices included in Chapter 2:

Appendix 2.1 Publication - Sherryn Ciavaglia, Stephen Donnellan, Julianne Henry, Adrian Linacre, Species identification of protected carpet pythons suitable for degraded forensic samples, Forensic Sci., Med. Pathol. (2014) 10:295-305

Appendix 2.2 Publication - Sherryn Ciavaglia, Shanan Tobe, Stephen Donnellan, Julianne Henry, Adrian Linacre, Molecular identification of python species: Development and validation of a novel assay for forensic investigations, Forensic Sci. Int. Genet. (2015) 16:64-70

Appendix 2.3 Snake whole mitochondrial genome maps

Appendix 2.4 Mitochondrial genome duplications and translocations between snakes

Appendix 2.5 Sequence homology charts by gene

Appendix 2.6 Best nucleotide substitution model for each dataset

2.1 Preface

Species identification tests constitute the most commonly arising question in forensic casework. The tests to answer this question are also the simplest assays to develop and conduct due to the conservation of genes and therefore priming sites in the mitochondrial genome. A range of techniques can be developed, from broad-scale universal amplification across taxonomic orders to fine-scale genus- and species-specific tests. One must not be lured into thinking that species identification is a simple task of picking a marker with a well-represented reference dataset, sequencing the unknown sample, searching the sequence against the databank and reporting the closest match. An understanding of the target species taxonomy is essential, as well as the representation of this species and other closely related species on the particular databank, not to mention the ability of the locus in question to accurately reflect the species phylogenetic history within its genetic signal. The following chapter will discuss the development of a new species identification assay for Australasian snake species, describing a novel bottom up approach to ensure continued effectiveness of the assay. The technique is designed *in silico* and concludes with a practical validation examining amplified product to demonstrate the actual utility of the developed assay for degraded forensic remains.

2.2 Aim and objectives of this chapter

The aim of this chapter is to develop and validate a mitochondrial DNA based species identification assay for the carpet python that can separate this species from all other species within the genus and is specifically designed to be suitable for the analysis of samples commonly encountered in forensic casework.

My objectives to complete this aim are:

- to assess DNA sequence conservation at the mitochondrial loci across a range of snakes to determine the most appropriate locus for design of universal and species-specific snake PCR assays.

- to create a species-specific python test through the identification of appropriate primer sites in a conserved region for amplification of a highly variable fragment suitable for differentiating all species of the genus Morelia

- to perform a developmental validation of the developed assay to demonstrate its direct applicability for use in criminal investigations involving carpet pythons and closely related snake species.

2.3 Introduction

Snakes are a popular commodity in China as a source of food, leather and for their historically perceived benefits in traditional Chinese medicines (TCMs) [1]; their use in TCMs having been described as early as the first century A.D. [2]. Snake gall bladder is believed to provide medicinal relief for such ailments as cough and gastrointestinal disease [3], rheumatism and poor vision [4]. Their popularity in countries such as Taiwan elevates the price of snake gall bladders [5] and the bile can be sold for values 200-fold higher than those of cow and pig [3]. Snake exploitation is not restricted to East Asia; India is a major supplier of snake skin products [6] and traditional snake remedies are commonplace in Brazilian culture [7]. These products that make their way onto legal and illegal international markets occasionally contain internationally protected species, breaching wildlife trade legislation.

All python species are listed in Appendix 2 of the Convention on International Trade in Endangered Species of Flora and Fauna (CITES) and therefore subject to trade sanctions and protection by signatory countries (https://www.cites.org/eng/app/appendices.php, accessed 21/07/2018). Despite this protection, reptiles including pythons are regularly acquired in Australia and traded unlawfully [8]. Native and exotic snake trade is frequently associated with other organized criminal activities [9–13]. Variability in morphology and colour pattern makes Australasian pythons attractive to the national and international pet trade, while also complicating visual species identification for enforcement officers tasked to regulate the trade (S. James, OEH NSW, *pers. comm.*). Investigators locating animals during searches require reliable and definitive tools to identify whether the animals in question are smuggled exotics or local species. Identification of the species is essential for an evidentiary basis and has previously been inadequate to progress these cases to court (H. Dridan, DEWNR, *pers. comm.*).

The condition of evidentiary samples received at Flinders University DNA laboratory and SASA Wildlife DNA Forensic unit has varied greatly and many are received in a condition adverse to DNA preservation; the presence of mould can indicate that samples might have been stored in unfavourable conditions prior to submission for laboratory testing. The samples provided to the Flinders laboratory were sometimes scale clippings or poorly stored tissue samples. Swabs from an evidentiary item suspected to have come into contact with a snake (e.g. [14]) might also be submitted, with the instruction to identify the species present. From such samples there is an expectation based on previous experience that the quality of the DNA template could be highly compromised. Likewise, traditional medicines are often highly processed and do not retain any morphological resemblance to their source species.

Molecular techniques are necessary to determine the species from which products and samples derive. Genetic techniques can offer a means to identify to species and subspecies level [15]. There has been sustained debate over the adequacy of the popular species identification genes: cytochrome *b* (cyt *b* - the first gene chosen for species identification) and cytochrome oxidase c 1 (COI - the gene elected by the Barcode of Life consortium). Few studies have considered other candidate gene sequences in the mitochondrial genome for use in species identification [16–20]. While any of these genes might be equally as informative, their use for species identification hinges on a demonstration that the relationships between the mitochondrial DNA sequences for a particular locus reflect the evolutionary relationships among the individuals carrying the genes, i.e. the species tree relationships.

Careful primer design forms the foundation of successful PCR assays. Guidelines for primer design have been extensively published and refined. Primers that have higher complementarity, and therefore specificity, to themselves or each other than the target DNA will form secondary structures (e.g. hairpins) and primer dimer rather than amplifying the desired product [21–23]. Base compositions involving polynucleotide, di-nucleotide (some of which can form a triple helix with the DNA template) and other repetitive sequence tracts within a primer sequence should be avoided [24]. Note that section 3.5.6.7 and [25] provide an example of this scenario.

The length and composition of a primer dictate the optimal melting temperature (T_m, the temperature at which half of the target-primer duplex is dissociated). At this point, the primer is most efficient in amplifying the target product. Base composition dictates the strength of binding. Adenine and Thymine form double hydrogen bonds, whereas bases Guanine and Cytosine form treble hydrogen bonds, creating a stronger interaction [24,26]. A refined formula for melting temperature is based on nearest neighbour thermodynamic theory, involving the division of rates of change of enthalpy by entropy and incorporating salinity [27,28]. However, a simplified equation $T_m = 4(G + C) + 2(A + T)$ °C ([29], cited by [21]) can give a ball park estimate of the expected T_m.

Increased GC content increases the strength of primer-target binding, however high GC content has negative effects on overall binding specificity [24] and 40%-60% GC content is desirable [30]. Dieffenbach & colleagues [21] observe that primers 20 bases in length with 50% GC content generally have T_m values between 56-62 °C. The T_m of primers designed as a pair must be similar for the primer pair to create the desired product [30]. When the annealing temperature of the reaction exceeds the T_m of the primer pair, binding to the target product is reduced. The optimal annealing temperature (matching the T_m) will give the maximal yield of target product, whereas a reduced

annealing temperature will induce binding when non-specific base-pairs are present, increasing the incidence of the primers hybridising with non-specific products [24]. Melting temperatures are most often chosen between 45 and 65 °C [30], falling within the common range of enzymatic activation and deactivation in biological systems. Oligonucleotide primers are therefore most frequently designed to be between 18-30 base pairs to facilitate binding at these melting temperatures [21,31], but the base composition also influences the annealing strength.

The most crucial region of a primer for primer design is the terminal five bases of the 3' end, the region from which DNA extension is initiated. This region needs to be highly complementary to the target sequence for extension to commence [21]. The strength is dependent on the hydrogen bonding between complementary bases, as mentioned earlier; three fifths GC content within the terminal five nucleotides is desirable for a strong primer-target interaction.

Primer degeneracy, otherwise known as 'wobble' sites, can be incorporated into species specific primers. Primer degeneracy is most often discussed in the context of amino acid redundancy [23,31], where the unconstrained third base of the codon site mutates rapidly, so multiple primers with different bases at this site are included in the primer mix to amplify all forms of the codon. The theory translates in this context to mutations that occur between individuals within a species at the primer site (or between closely related species depending on the degree of sequence conservation in the target taxa and desired scope of the developed test). Incorporating multiple primer sequences varying at known crucial bases expands the capability of the test to capture all desired amplification products. Each wobble site adds 2-4 alternate primer sequences to the synthesised primer mix, so redundancy at only three sites in a primer creates a reaction mix of 6-12 primers in solution; this figure is doubled if used in forward and reverse primers. The danger is increased primer interactions, which will out-compete target hybridisation, so the use of primer degeneracy should be minimised.

Universal primers developed for cyt *b* have been applied to six snake species commonly consumed in Hong Kong for the purpose of species identification [32]. The data were combined with equivalent DNA sequence for another 84 snake species available on Genbank to create a searchable online snake cyt *b* database, but this resource has since been decommissioned (P.C. Shaw, *pers. comm.*). The DNA research group of the Central Forensic Science Laboratory in West Bengal, India have investigated the potential of the 16S ribosomal RNA and cyt *b* coding regions for snake species identification using a PCR-RFLP approach [6,33].

Forensic species identification tests should be designed to obtain results from remnant DNA fragments in case the forensic sample is in poor condition, which is not uncommon for biological exhibits; an example of such a test is nested PCR of the cyt *b* gene used to identify species of turtle from processed turtle shell [34]. 'Mini barcoding' of the COI gene has been used to interrogate very small fragments of the COI barcoding gene in eleven species of protected Indian snakes [35]. In this study, a figure of 98% similarity was claimed to be a species match, however it was based upon one sample per species with no apparent consideration of intraspecific variation. It should be acknowledged that the authors assessed their technique according to recommendations of the Scientific Working Group on DNA Analysis Methods (SWGDAM).

In most snakes, the cyt *b* gene is 1114 bp in length and encodes a protein of 371 amino acids and the ND6 gene encodes a protein of 171 amino acids in length and is 513 bp in length. The ND6 gene is the only protein-encoding gene coded by the light strand in snakes [36] and is adjacent to the cyt *b*

gene, separated by the tRNA-glutamine gene. ND6 is thus a good candidate gene to sequence and obtain a continuous contig. sequence of almost 1800 bp containing the complete ND6 and cyt *b* coding regions. Furthermore, alignment of existing GenBank data indicated that alongside cyt *b*, chosen both for its high variability and reputation in previous species identification research (e.g. [37,38], ND6 exhibited a high degree of interspecies variation (unpublished data). The ND6 gene was the second most variable of all genes in the mitochondrial genome (determined by methods described in [39]; the most variable mitochondrial gene locus, ATPase 8, was suspected by the authors to be too variable for species identification purposes as it is a very small gene (171 bp in *M. spilota*) and identification of PCR primer sites conserved across the snake species would likely be problematic

The following chapter examines the applicability of the different protein encoding regions in the mitochondrial genome for identifying to species level an unknown sample suspected to contain snake. While respective mutation rates have been published, this is first comparison of variability within each gene across all of the different mitochondrial protein coding regions within a Suborder, in this case *Serpentes*. A bottom up approach is suggested for developing species identification assays to improve marker specificity for very closely related species.

A mitochondrial DNA region is characterised that contains appropriate genetic signal for species identification of CITES Appendix 2 listed and legislatively protected members of the python genus *Morelia*. This genus, which comprises 11 species ([40] accessed 10/10/2013), is found in Australia, New Guinea and the eastern part of the Indonesian archipelago. Through phylogenetic reconstruction, the ability of the ND6 and cyt *b* genes to distinguish between the widespread and morphologically variable carpet python species and all other Australasian *Morelia* species is demonstrated. Within these genes, short segments can provide accurate species identification comparable to the entire respective gene region. Also presented is the identification of a small fragment of mitochondrial DNA with the ability to identify to species level scale clippings suspected to have been taken from a species of *Morelia*. This work was conducted in line with the recommendations of [41].

2.4 Methodology

2.4.1 Snake whole mitochondrial genome mapping

Thirty-nine whole mitochondrial genomes of species of the Suborder Serpentes were obtained from the online databank Genbank (http://www.ncbi.nlm.nih.gov/genbank/). Table 2.1 provides the details of the sequences examined. Schliep & O'Shea [42] was used to clarify snake taxonomic classifications. One whole mitochondrial genome (*Morelia spilota*) was obtained by 454 Next Generation Sequencing (Roche). Each genome was constructed into circular mitochondrial maps using code written in-house using R [43] to space the regions accordingly within the complete mitochondrial genome. Genes coded for by the heavy strand are shown on the outside of the map. A mitochondrial genome map was constructed for human to compare gene organisation within and between snakes.

2.4.2 Choice of mitochondrial gene region

The 40 snake sequences were broken down into fragments corresponding to coding regions. Each of the 13 protein coding genes was aligned by eye using MEGA5.0 [44] and Geneious v5.4.3 (Biomatters). The ATPase 8 gene could not be aligned across all species of snakes as sequences appeared too divergent. Similarly, bases 162-301 of ND6 could not confidently be aligned by the author due to sequence hyper-variability. The comparisons of these two regions across all snake species have not been included in this assessment. Alignments were exported into an Excel spreadsheet. Shannon (adapted by [45], cited in [39]) describes the R_s value as the measure of uncertainty

$$R_s = \log_2 N - (-\Sigma p \log_2 p)$$

The R_s values for each nucleotide position within a gene were charted to show the degree of sequence similarity within each gene for a certain set of individuals. Charts were constructed using alignments of all forty complete mitochondrial genomes obtained to demonstrate the degree of variation within a gene across all genomes available for the suborder Serpentes and alignments using seven species representing the superfamily Henophidia, to demonstrate the degree of variation at a lower taxonomic level. Complete Cytochrome *b* sequences were available for 25 species within Henophidia. The COI dataset could not be increased in this way because most gene sequences on Genbank were not complete. Moving averages were applied for each 20 base pair window to reflect size of primers for primer design. Both datasets were used to examine the best (most conserved) regions to place universal snake primers. The within Henophidia dataset was used to determine which genes exhibited the highest variability, as sequence that shows high variability between closely related species is most likely to contain candidate regions for species-specific primers.

2.4.3 Primer design

Primers were designed using the software Primer3 [46] as part of the Geneious software package (Geneious 5.4.3 created by Biomatters), which considers many common primer design parameters, in conjunction with the sequence similarity information to ensure that the annealing sites of primers are conserved across snake species and the primers amplify product containing high variability. Snake gene alignments from section 2.4.2 were used to design primers in locations conserved across

Table 2.1: Details of the 39 species for which whole mitochondrial genomes were obtained for gene alignments. Data obtained from reptiledatabase.org and eol.org, accessed 02/02/2018. This table gives an idea of the taxonomic breadth of sequences aligned in this study, - denotes a classification level that was not identified for the species.

Order	Suborder	Infraorder	Superfamily		Family	Subfamily	Genus	*Taxonomic name*	Genbank Accession #	Common name
Squamata	Serpentes	Alethinophidia	Caenophidia	-	Acrochordidae	-	Acrochordus	*Acrochordus granulatus*	AB177879	**Granulated File Snake**
				Colubroidea	Colubridae	Dipsadinae	Hypsiglena	*Hypsiglena chlorophaea chlorophaea*	EU728593	**Sonoran Nightsnake**
								Hypsiglena jani texana	EU728592	**Texas Nightsnake**
								Hypsiglena torquata	EU728591	**Night Snake**
								Hypsiglena ochrorhyncha klauberi	EU728589	**San Diego Nightsnake**
								Hypsiglena slevini	EU728584	**Baja California Nightsnake**
							Imantodes	*Imantodes cenchoa*	EU728586	**Blunthead Tree Snake**
							Leptodeira	*Leptodeira septentrionalis polysticta*	EU728590	**Cat-eyed Snake**
							Sibon	*Sibon nebulatus*	EU728583	**Cloudy Snail-Eating Snake**
							Pseudoleptodeira	*Pseudoleptodeira latifasciata*	EU728579	**False Cat-eyed Snake**
							Thermophis	*Thermophis zhaoermii*	GQ166168	**Sichuan Hot-spring Keel-back**
						Colubrinae	Dinodon	*Dinodon semicarinatus*	AB008539	**Loo-choo Big-tooth Snake**
							Pantherophis	*Pantherophis slowinskii*	DQ523162	**Slowinski's Corn Snake**
							Elaphe	*Elaphe poryphyracea*	GQ181130	**Black-banded Rat Snake**
					Xenodermatidae	-	Achalinus	*Achalinus meiguensis*	FJ424614	**Sichuan Odd-scaled Snake**
					Homalopsidae	Homalopsinae	Hypsiscopus	*Enhydris plumbea*	DQ343650	**Rice Paddy Snake**
					Viperidae	Crotalinae	Deinagkistrodon	*Deinagkistrodon acutus*	EU913476	**Chinese Moccasin**
							Agkistrodon	*Agkistrodon piscivorus*	DQ523161	**Cottonmouth**
							Crotalus	*Crotalus horridus*	HM641837	**Timber Rattlesnake**
							Ovophis	*Ovophis okinavensis*	AB175670	**Himehabu**
							Gloydius	*Gloydius blomhoffi brevicaudus*	EU913477	**Short tailed Mamushi**
							Trimeresurus	*Viridovipera stejnegeri stejnegeri*	FJ752492	**Chinese Green Tree Viper (labelled Taiwan GTV from Genbank)**
						Viperinae	Daboia	*Daboia russellii*	EU913478	**Russell's Viper**
						Causinae	Causus	*Causus defilippi*	GU045452	**Snouted Night Adder**
					Elapidae	-	Naja	*Naja atra*	EU921898	**Chinese Cobra**
							Naja	*Naja naja*	DQ343648	**Indian Cobra**
						-	Bungarus	*Bungarus fasciatus*	EU579523	**Banded Krait**
						-	Bungarus	*Bungarus multicinctus*	EU579522	**Many-Banded Krait**
						-	Ophiophagus	*Ophiophagus hannah*	EU921899	**King Cobra**
						-	Micrurus	*Micrurus fulvius*	GU045453	**Eastern Coral Snake**
			Henophidia	-	Aniliidae	-	Anilius	*Anilius scytale*	FJ755180	**False Coral Snake**
				-	Boidae	Boinae	Boa	*Boa constrictor*	AB177354	**Boa Constrictor**
				-	Cylindrophiidae	-	Cylindrophis	*Cylindrophis ruffus*	AB179619	**Red-tailed Pipe Snake**
				Pythonoidea	Pythonidae	-	Morelia	*Morelia spilota*	In-house NGS	**Carpet Python**
						-	Python	*Python regius*	AB177878	**Ball Python**
				-	Tropidophiidae	-	Tropidophis	*Tropidophis haetianus*	FJ755181	**Haitian Dwarf Boa**
				-	Xenopeltidae	-	Xenopeltis	*Xenopeltis unicolor*	AB179620	**Sunbeam Snake**
		Scolecophidia			Leptotyphlopidae	Epictinae	Leptotyphlops	*Leptotyphlops humilis* (synonym *Rena humilis*)	AB079597	**Western Threadsnake**
			-	-	Typhlopidae	Typhlopinae	Typhlops	*Typhlops reticulatus*	EU747730	**Reticulate Worm Snake**
							Ramphotyphlops	*Ramphotyphlops braminus*	DQ343649	**Brahminy Blind Snake**

all species. Primer sequences located in ND5 and tRNA-Glutamine were designed to amplify the entire ND6 gene. Cytochrome b is 1114bp in pythons, so two overlapping primer pairs were designed to amplify the entire gene. The first primer pair is situated at the 5' end of ND6 and amplifies tRNA-Glutamine and 471bp of the first half of cyt *b*. The second primer set amplifies bases 377-1114 of cyt b and 35 bp (5') of the *tRNA-Threonine* gene, creating a sequence overlap of 94 bp between the two amplification products. Primer pairs were searched using BLAST (Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify potential amplification in non-target species. Both primers would need to match to the same species within a region (~2 Kb) reasonable for PCR product amplification.

Primer sequences and locations are provided in figure 1 of the embedded manuscript [47] in Appendix 2.1. Primers with a redundant sequence were designed for cyt *b* to maintain universality. When the primer solution is prepared by the manufacturer, primers with alternatives of the redundant base are synthesised, so that products containing both alternatives will be sequenced. This means that stock solution of a primer with three bi-allelic redundant sites actually contains 16 unique primer sequences. For this reason, base redundancy should be kept to a minimum when designing primers.

2.4.4 Samples

Refer to the embedded manuscript [47], Appendix 2.1 for details of the samples used in this chapter.

2.4.5 DNA extraction and sequencing

Samples were removed from solution and placed on a heatblock at 56 °C for 10 minutes to remove ethanol traces or thawed. Tissue samples in good condition and expected to give a high DNA yield were extracted using the QIAamp DNeasy Blood and Tissue Kit (QIAGEN, Australia) following the manufacturer's protocol for the Purification of Total DNA from Animal Tissues, with the exception that the volume for the ATL buffer, Proteinase K and DTT was doubled for large skin samples. The QIAamp DNA Investigator kit (QIAGEN, Australia) was used following the "Isolation of Total DNA from Tissues" protocol for tissue samples suspected to be of poor condition and contain little DNA and the "Isolation of Total DNA from Nail Clippings and Hair" protocol for skin, scale and tail samples (volumes of ATL buffer, Proteinase K and DTT doubled for large skin samples). Elution volumes ranged from 40-600ul according to the expected quality of sample.

Previously extracted samples were purified using BIOLINE Isolate PCR and Gel Kit, following the "Isolation of PCR Products" protocol. Amplifications were performed using 2 µL extracted DNA, 200 nM dNTPs, 1 µM of each primer, 2 Units of DNA Polymerase (BIOLINE, Australia), in a total volume of 20 µL or 10µL when downstream sequencing was not required. PCR conditions for amplifying the ND6 gene were initial denaturation at 95 °C for three min, 35 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 30 s, with a final extension at 72 °C for 10 min. Amplification of cyt b was performed as above but for 30 cycles and with an annealing temperature of 56 °C. Product concentration was estimated by 2 % agarose gel electrophoresis and visualization by ethidium bromide using a Bio-Rad Gel Doc and Image Lab software (Bio-Rad, Australia). PCR products were prepared for sequencing using ExoSap (GE Healthcare, Australia), following the manufacturers protocol. DNA sequencing was performed by the Australian Genome Research Facility.

2.4.6 Sequence alignment and data analysis

DNA sequence data were edited and aligned using Geneious 5.4.3. Alignments of consensus sequences were imported into MEGA v5.0 for phylogenetic analysis. Sequences have been deposited on GenBank (accession numbers KJ666583-KJ666640). Phylogenetic trees were constructed using Neighbour-Joining and Maximum Likelihood algorithms using *Python regius* as an out-group. Neighbour-Joining trees were constructed using MEGA v5.0 with a p-distance model. Maximum Likelihood trees were constructed using RAxML Blackbox v7.6.3 (Cipres Science Gateway [48]) with partitioning strategy and nucleotide substitution models chosen using Partition Finder v1.1.1 [49]. Bootstrapping was used to provide statistical confidence for the tree topology. Bootstrap support for nodes using the Maximum Likelihood method is applied to the Neighbour Joining trees to compare concordance of the two construction types. In order to examine every member species of *Morelia*, a separate dataset comprising four 715 bp cyt *b* sequences of *M. clastolepis* retrieved from GenBank and the corresponding sequence data for our 53 cyt *b* reference samples was used to demonstrate the phylogenetic placement of *M. clastolepis* in relation to other *Morelia* species.

2.4.7 Genetic distance analysis

A dataset of 19 initial samples (shaded light blue in Table 1 of embedded manuscript [47], Appendix 2.1) sequenced for the entire ND6 and cyt b coding regions was used for this analysis. Pairwise distances were calculated for each dataset using p-distance, Jukes-Cantor, Kimura-2-Parameter and Tamura-Nei models. The MEGA 'find the best DNA model' was used to identify model with best fit to the data (Tamura-Nei model to be the best fit for all three data sets, using a Gamma Distribution parameter of 0.2) and pairwise distances were also calculated using these settings. Pairwise distances for the complete consensus sequence dataset were exported to an Excel spread sheet, arranged by increasing distance and values resulting from the p-distance and Tamura-Nei (G = 0.2) models charted.

2.4.8 Species testing using decreasing DNA fragment size

Often forensic samples that require DNA testing to determine species identification are processed or highly degraded, causing the DNA within to be fragmented and of poor quality. Amplification of intact gene sequences is unlikely. A continuous sequence containing the ND6 and cyt b genes from 48 reference samples including 32 Morelia spilota, one each of *M. amethistina* and *M. kinghorni* and two each of *M. boeleni*, *M. bredli*, *M. carinata*, *M. nauta*, *M. oenpelliensis*, *M. viridis* (northern population) and *M. viridis* (southern population) were divided into half and quarter fragments. Eighth fragments were also created for the larger cyt b gene.

Phylogenetic trees were constructed using the complete consensus sequences, the whole ND6 and Cytochrome b sequences, as well as the smaller fragments to test if any of the 6 smaller fragments from ND6 gene and 14 smaller fragments from cyt *b* gene could reproduce the ability of the larger gene fragments to separate individuals to species. Tree construction techniques used were Neighbour-Joining, Maximum Likelihood and Maximum Parsimony, with *Python regius* included as outgroup. The Neighbour-Joining trees made most intuitive sense in reflecting geographic relationships of samples and this method was chosen for further tree construction.

2.4.9 Simulated forensic samples

Five extra samples for which only the complete cyt b sequence could be obtained (due to poor sample quality) were included in the comparisons within cyt b fragments. These samples are indicated as “Reference (cyt b only)” in Table 1 of the embedded manuscript [47], Appendix 2.1. These samples increased the cyt b dataset to 53 individuals. Five blind trial or ‘test’ samples were used to test the ability of the best smaller gene fragment to identify species as per the larger gene fragment (samples marked Blind trial “test” in table 1 the embedded manuscript [47], Appendix 2.1). The species identity of these samples was checked in the ABTC records after testing.

2.4.10 Forensic case samples

The 278 bp cyt b segment was analyzed for each of the four samples of scale clippings (M001, M002, M003, and M004) believed to be taken from carpet pythons, using the MsCB2 primer pair (figure 1 of embedded manuscript [47], Appendix 2.1). These sequences were aligned against those already generated and a phylogenetic tree was constructed.

2.4.11 Design and validation of an assay suitable for degraded remains

The embedded manuscript [50], Appendix 2.2 describes the methods followed to design and validate an assay based on the amplification of smaller fragment sizes.

2.5 Results

2.5.1 Snake whole mitochondrial genome mapping

Appendix 2.3 shows the snake maps constructed. There are two distinct infraorders of snakes: Scolecophidia (thread snakes and blind snakes) and Alephinophidia (all other snakes). Appendix 2.4 illustrates differences seen between snake mitochondrial genomes. The most remarkable mitochondrial genome characteristic in snakes is a duplicate control region, a synapomorphy present in all Alephinophidia and none of the Scolecophidia. Many snake species that have a duplicate control region also show gene translocations so that the gene order around the second control region is now 16S rRNA, NADH dehydrogenase 1, Ile t-RNA molecule, control region 2, Leu tNRA molecule, Gln tRNA molecule then Met tRNA molecule.

A gene translocation of the Proline tRNA molecule from between the Threonine tRNA and Control Region 1 to between the Ile tRNA and Control Region 2 is present in all species of the Viperidae family. None of these snake species have a functional tRNA-Pro coding gene in the original position next to the Control Region (CR1). Ovophis has a tRNA-Pro pseudo gene in this position, but it partial and therefore not functional. Pantherophis and Dinodon, of the family Colubridae, have a pseudo gene for tRNA-Pro at the position of the translocation seen in Viperidae, but the coding region is only partial and not functional.

Isoleucine tRNA coding region appears to be duplicated in Ophiophagus genus. The coding region for tRNA-Gln swaps from the light strand to the heavy strand in Enhydris, Achilinus and Pantherophis, all members of the Colubrid family. It is possible that these taxa are closely related and have shared a translocation of the gene from one DNA strand to the complementary strand, but further

investigation into the relationships of these genera is beyond the scope of this thesis. Tropidophis exhibits a translocation of the Lysine tRNA from upstream to downstream of the ATP8 gene. Sibon nebulatus has a large repeat region extending the mt genome from the usual 16Kb to 21Kb.

2.5.2 Choice of mitochondrial gene region

The sequence similarity across the entirety of each gene were charted for the seven NADH coding regions, cytochrome b, the three cytochrome c oxidase regions and the two ATPase coding regions, both for the All snakes dataset and in Henophidia only (Appendix 2.5). The All Snakes dataset did not include the ATPase 8 gene or bases 162-301 of ND6. Moving averages of both datasets generally show the same trends towards variability or conservation of sequence at comparable sequence sites within each gene. Figure 2.1 provides an example comparison between the genes commonly used for species identity testing, cyt b and COI. The cyt b average reaches a minimum of 0.996 at one value then rises above 1.0 again (Figure 2.1a). The COI coding region shows the highest average conservation across the gene (Figure 2.1b). The degree of variability within the first 1436 base pairs is fairly maintained with the amplitude of the moving average staying within a 0.5 R_s range and only the last 166 base pairs exhibiting two regions of increased variability. The amplitude of the moving average over the length of the cytochrome b gene is larger, at some points changing by over a 0.7 R_s value. These characteristics provide multiple opportunities for conserved primers to be designed flanking a region of high variability, desirable traits for effective species identification tests. These data indicate the cyt b gene overall to be a better candidate for development of a species identification assay for snake species than the COI gene. As expected, higher sequence similarity is seen in the Henophidia dataset when comparing each gene chart for the two datasets.

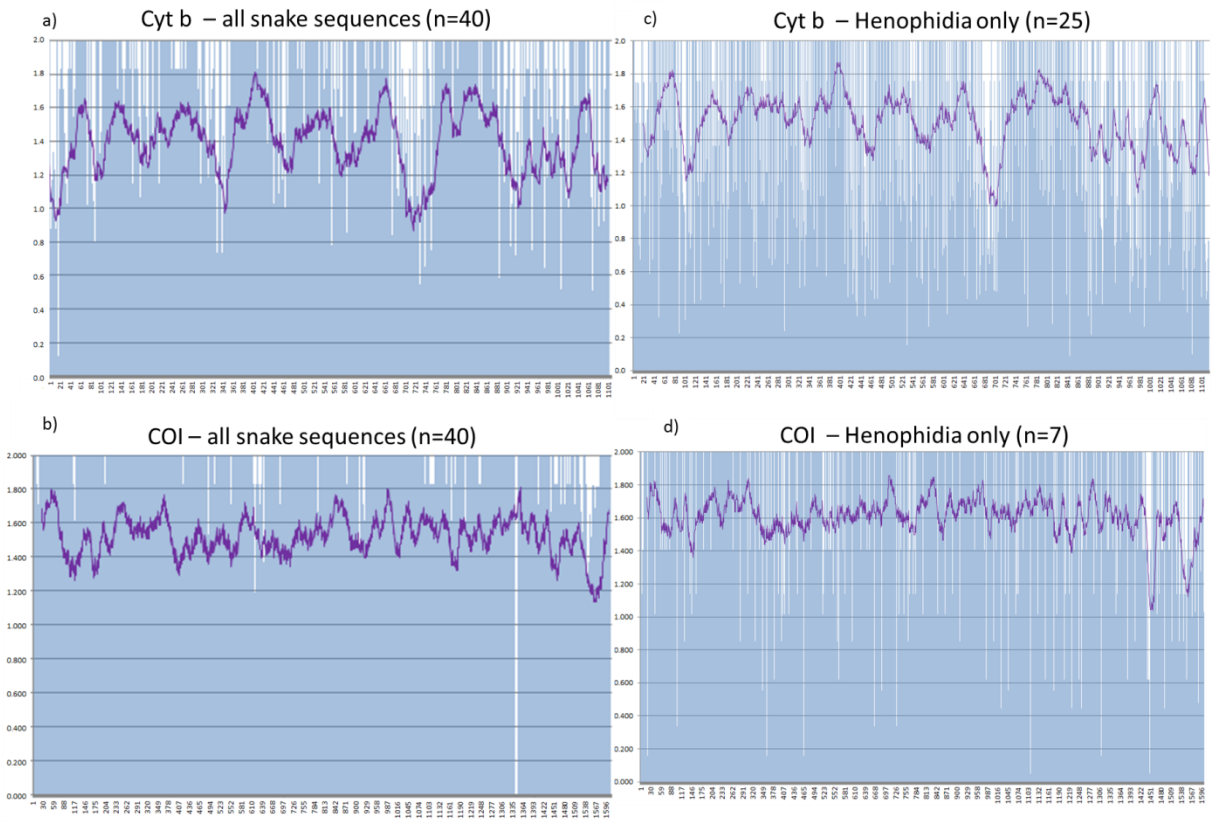


Figure 2.1: Examples of the sequence similarity across a mitochondrial gene compared between two coding regions, cyt *b* and COI using all available snake species, a) and b), and Henophidian species only, c) and d). The purple line represents a moving average over a 20 bp sliding window: a) and c) The average level of sequence similarity is lower over the entire cyt b gene region, with large alternating regions of sequence conservation and variability are seen, b) and d) Cytochrome Oxidase c I exhibits quite a high average sequence conservation across the gene and the amount of similarity at different gene regions varies little, with one region of increased sequence variability seen within the last 100 bp of the gene.

2.5.3 Within Henophidia variability

The average R_s value for each entire gene region are provided (table 2.2), however reference to the charts (Appendix 2.5) shows that sequence conservation can vary markedly within a coding region. The small 171 b.p. coding region for ATPase 8 shows the highest degree of variability across

Table 2.2: Average R_s values across each gene coding region. A higher R_s value denotes more sequence conservation within that gene across the 40 snake species analysed.

Coding region	Ave. R_s	Coding region	Ave. R_s
ND1	1.584	Cytochrome b (across species)	1.508
ND2	1.503	COI	1.608
ND3	1.472	COII	1.574
ND4	1.514	COIII	1.602
ND4L	1.504	ATPase 6	1.477
ND5	1.481	ATPase 8	1.201
ND6	1.380		

its entirety. Only the first 20 bases show enough conservation to give a moving average above 1.6. The entire gene exhibits an average R_s value of 1.2. The second most variable whole gene region is ND6. Whereas the average of most other gene charts tends to spend the majority of the time above an R_s value of 1.4, ND6 shows far less charted area greater than 1.4. The ND6 gene also exhibits a region of variability where the moving average drops very close to 0.8. The ND3 gene drops slightly below 1.0 over a 6 base region and ND5 also has four stretches below 1.0 within the first 100 base pairs of the open reading frame, the longest of these being for 12 base pairs.

2.5.4 Genetic Distance Analysis

Genetic distance increased for any particular comparison of two sequences when different models were applied in the order p-distance < Jukes Cantor < Kamura 2 Parameter < Tamura-Nei93 < Tamura-Nei93+G(0.2) (table 2.3).

Table 2.3: Genetic distance ranges of pairwise comparisons within and between species using five different nucleotide substitution models.

		p-dist.	JC	K2P	TN93	TN93+G(0.2)
ND6+cyt b (1792 bp)	Intra-species	0.000-0.066	0.000-0.070	0.000-0.071	0.000-0.072	0.000-0.110
	Inter-species	0.099-0.135	0.107-0.148	0.109-0.153	0.112-0.157	0.221-0.432
	Gap	*0.033*	*0.047*	*0.048*	*0.050*	*0.111*
ND6 (513 bp)	Intra-species	0.000-0.066	0.000-0.069	0.000-0.071	0.000-0.075	0.000-0.157
	Inter-species	0.078-0.133	0.082-0.146	0.084-0.149	0.087-0.158	0.161-0.533
	Gap	*0.012*	*0.013*	*0.013*	*0.012*	*0.004*
Cyt b (1114 bp)	Intra-species	0.000-0.072	0.000-0.075	0.000-0.077	0.000-0.077	0.000-0.117
	Inter-species	0.109-0.139	0.117-0.154	0.120-0.159	0.124-0.163	0.254-0.448
	Gap	*0.037*	*0.042*	*0.043*	*0.047*	*0.137*

Every model applied to the three datasets showed a distinct gap between intra-species and inter-species comparisons (see figure 2.2 for example). All intra-species comparisons within the included species (*M. spilota*, *M. bredli*, *M. viridis*, *M. Oenpelliensis* and *M. boeleni)* resulted in a smaller genetic distance value than any of the inter-species comparisons. Pairwise sequence comparisons between the *M. bredli* sequence and *M. spilota* individuals gave values within the range of intra-species comparison distances. The split between intra-species and inter-species comparisons was larger for the entire 1792bp sequence and whole cytochrome b sequence, but quite small for the whole ND6 sequence. In the ND6 dataset, the application of the correction for substitution rate at differing codon position decreased the value substantially, giving a very small gap between the largest intra-species value and smallest inter-species value. The small separation between intra-species pairwise distances and inter-species pairwise distances at ND6 suggests that this region not a good candidate for a species identification test.

Application of the Tamura-Nei93+G(0.2) model gave a much greater difference in the genetic distances than application of the other three models, likely due to the strong correction factor used for the nucleotide substitution rate at the different codon positions. The gap between intra-species and inter-species comparisons increased from 0.02 using p-distance to 0.1 when the Tamura-

Nei93+G(0.2) model was applied. The error rates were also seen to increase proportionate to the genetic distances when increasingly complex models were applied.

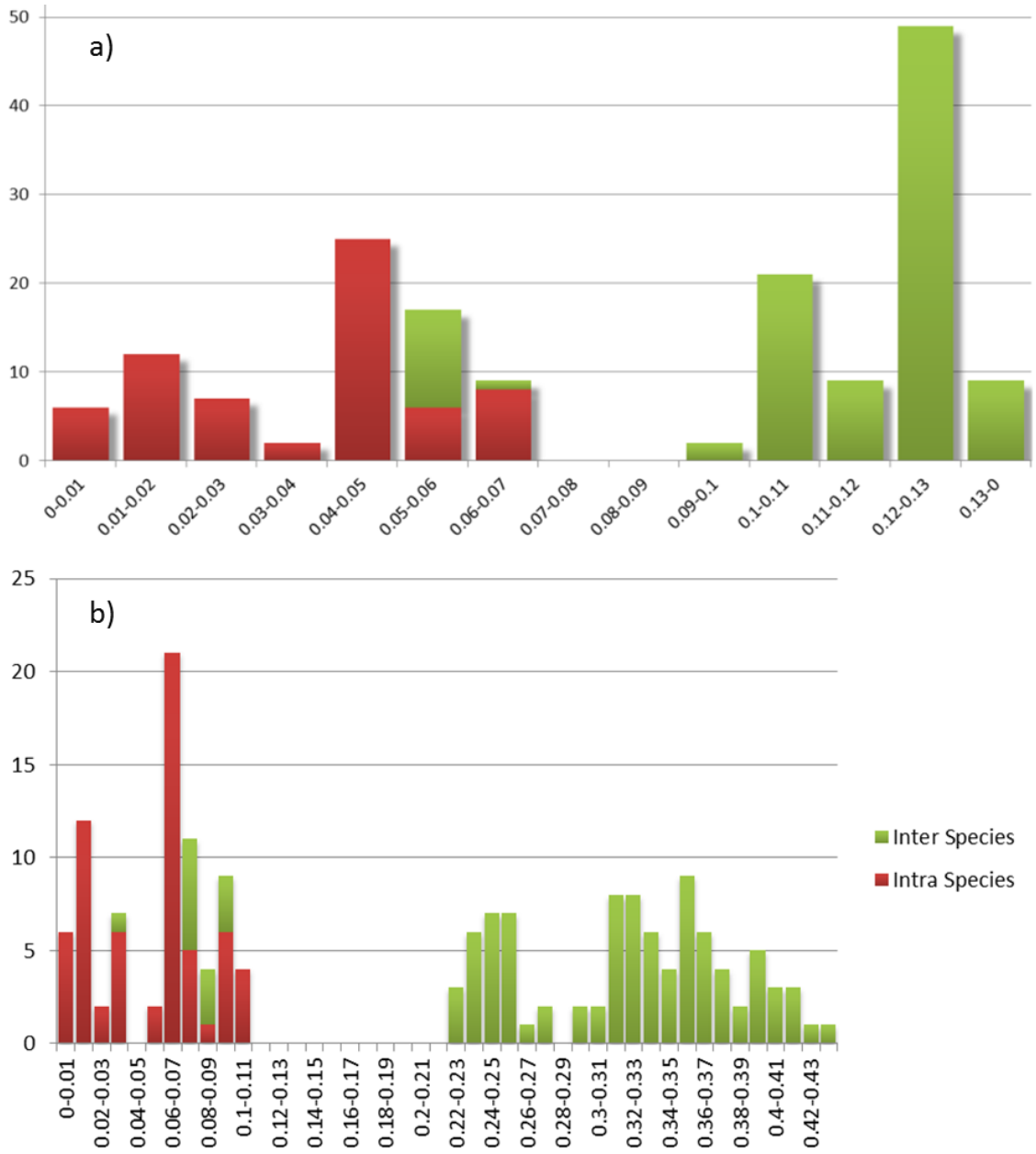


Figure 2.2: Genetic distances based on pairwise sequence comparisons: 2a) pairwise sequence comparisons using p-distance model. Intraspecies and interspecies comparisons for the most part fall into two distinct groupings. All intraspecies comparisons exhibit a genetic distance <0.07. All interspecies comparisons (excluding *Morelia spilota* - *Morelia bredli* comparisons) exhibit a genetic distance >0.09. The interspecies comparisons that fall below 0.07 relate to comparisons of *Morelia spilota* and *Morelia bredli*. 2b) the same pairwise sequence comparisons applying a Tamura-Nei 93 + Gamma (0.2) model. The split between the intraspecies and interspecies sequence comparisons is much larger when the more complex model is applied. All intraspecies comparisons exhibit a genetic distance <0.11 and all interspecies comparisons (excluding *Morelia spilota* - *Morelia bredli* comparisons) exhibit a genetic distance above 0.22.

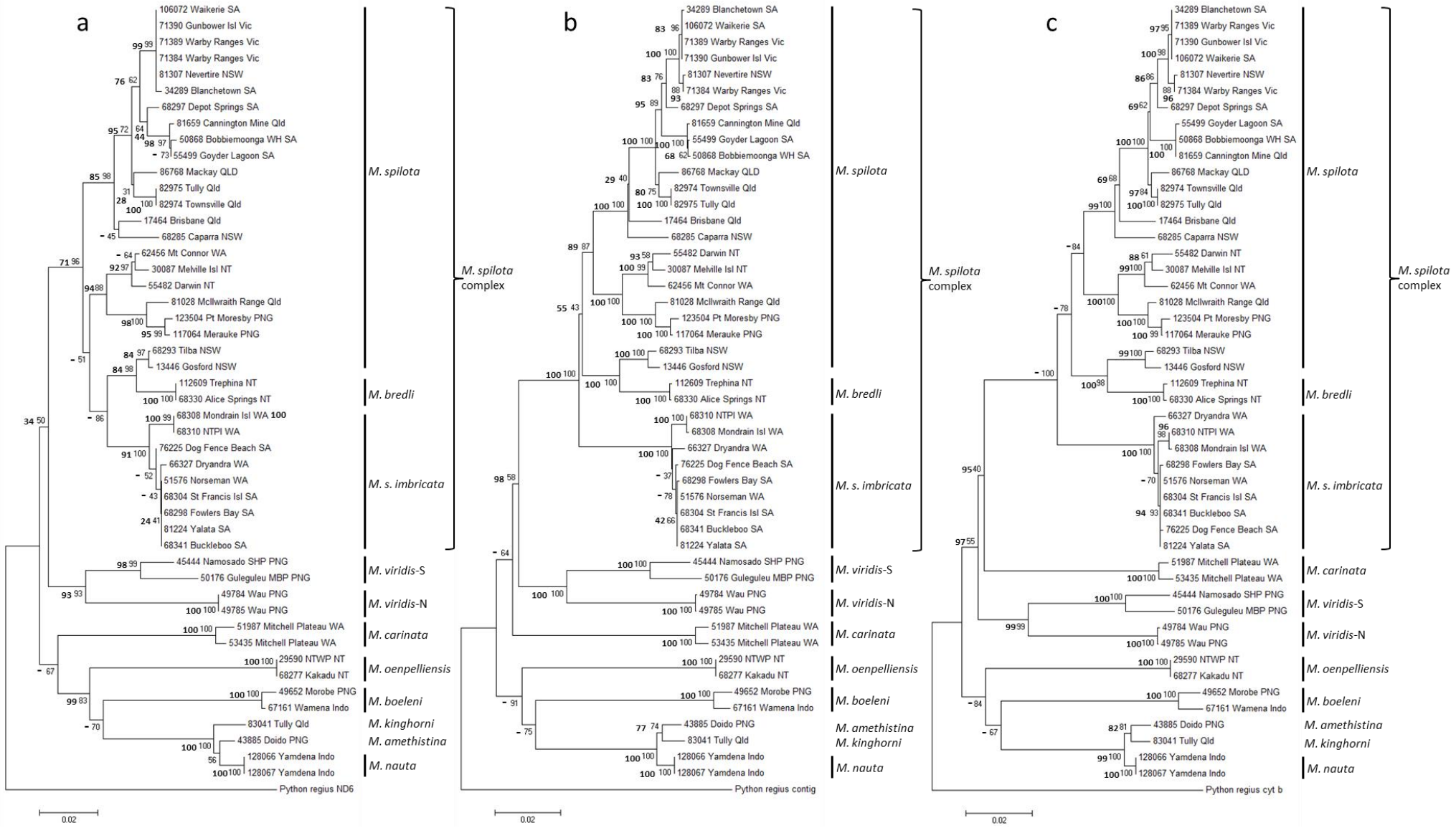


Figure 2.3: Neighbour Joining phylogenetic trees with p-distance model and 500 bootstrap replicates of 49 individuals including outgroup using a) the complete ND6 gene, b) a contig. of ND6/tRNA-Glutamine/cytochrome b genes, and c) the complete cytochrome b gene. Trees were rooted using *Python regius* as an outgroup. Maximum Likelihood bootstrap proportions are in bold to the left of Neighbour-Joining bootstrap proportions, - denotes a NJ node that was not present in the ML tree.

2.5.5 Phylogenetic tree reconstruction

Appendix 2.6 provides the best nucleotide substitution models using Partition Finder v1.1.1 for the various datasets used to construct phylogenetic trees during this study.

Phylogenetic trees constructed using the complete ND6 gene, the complete cytochrome b gene and the complete contiguous sequence of both ND6 and Cytochrome b are provided (figure 2.3). The partial cytochrome b tree successfully separates *Morelia clastolepis* as a clade from all other *Morelia* species (figure 2.4). This tree demonstrates the ability of the cytochrome b locus to identify to species ten of the eleven *Morelia* species. *Morelia bredli* is unable to be distinguished from *Morelia spilota*, indicating that at a genetic level these are not separate species. The mitochondrial genome is but a single locus and nuclear data would assist the resolution of the taxonomic classification of *M. bredli*.

See the embedded manuscript [47], Appendix 2.1 for further results and discussion about the phylogenetic trees.

2.5.6 Species testing using decreasing DNA fragment size

Results of the decreasing fragment size tests are presented and discussed in the embedded manuscript [47], Appendix 2.1.

2.5.7 Simulated forensic samples & forensic case samples

Results of the simulated and actual forensic case samples are presented and discussed in the embedded manuscript [47], Appendix 2.1.

2.5.8 Design and validation of an assay suitable for degraded remains

The embedded publication [50], Appendix 2.2 describes the methods followed to design and validate an assay based on the amplification of smaller fragment sizes.

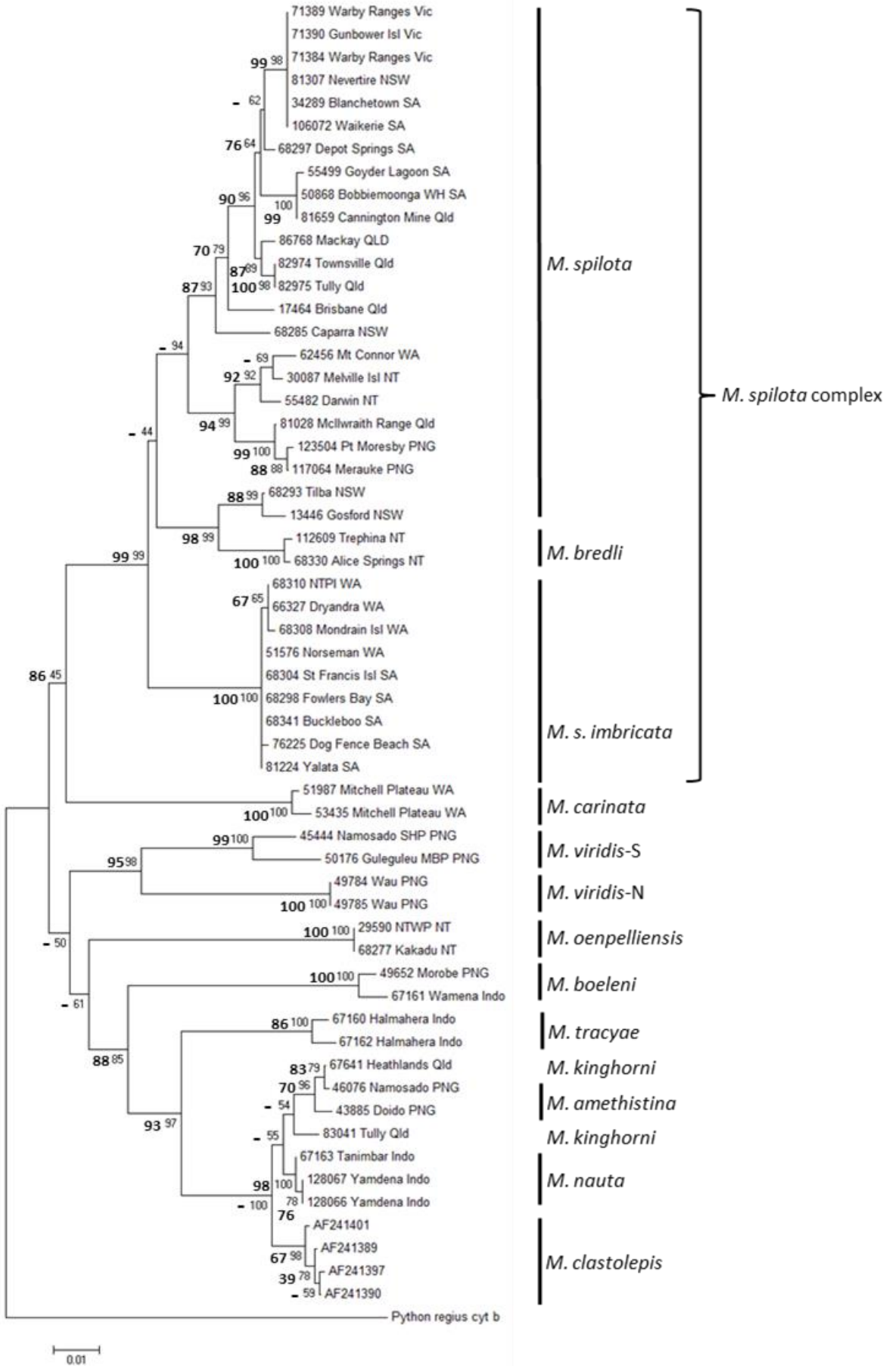


Figure 2.4: Phylogenetic tree constructed using 715 bp of cytochrome b from 53 *Morelia* reference samples used in this study and four sequences of *M. clastolepis* obtained from GenBank. Three constructed using Neighbour-Joining with p-distance model and 500 bootstrap replicates and rooted using *Python regius* as outgroup.

2.6 Further discussion

2.6.1 The Importance of well-designed primers

The Polymerase Chain Reaction forms the basis of most modern DNA techniques, enabling quantities of DNA product that can be visualised and further manipulated. Well-designed oligonucleotide probes or primers are crucial to successful PCR amplification. They prescribe the target region for the DNA replication process. This project focused on designing PCR based assays for species detection and identification, so well-designed primers were vital to a successful forensic assay that demonstrates fitness for purpose.

During primer design, theoretical calculations of a primer's T_m were made using various online software packages; however these rarely agreed with one another or the manufacturer's T_m supplied in the synthesised primer notes. A single primer sequence tested across various web calculation tools and the Geneious software gave the T_m values 55 °C, 59.7 °C, 60 °C, 61.4 °C, 62 °C, 69.7 °C, 73.9 °C and 74.1 °C. Empirical assessment by gradient PCR has always been necessary in my experience to determine the optimal annealing temperature for the designed primers, regardless of what the manufacturer claims the T_m to be.

For the species identification tests developed in this chapter, the optimal annealing temperature aspect of primer design can be used to the wildlife forensic biologist's advantage. Species-specific primers are designed with sequences highly specific to the target taxon and annealing at the optimal T_m produces products specific to this species and no other (provided the priming site is truly species specific). When the desired target is a range of closely related species, lowering the annealing temperature will relax the stringency of the binding, producing products from the target taxon and any other DNA sequences that are highly similar (as those of closely related species often are). Universal primers designed to target a broad range of species are best applied using low annealing temperatures, as this encourages the primers to amplify targets that are moderately similar and effectively extend from homologous priming sites that have varied base composition across wide ranging taxa. The danger is that analogous priming sequences from other origins in the genome will also amplify, increasing the incidence of non-specific binding products.

The species specific (MsND6F & -R, MsCB1F & -R, MsCB2F & -R) and forensic (MSFCBF & -R) primers developed in this chapter were all designed with the guidelines for primer design introduced in section 2.1 in mind. The ND6 primer pair often showed weaker amplification of the same DNA template. This primer pair did not have inbuilt primer degeneracy, which could be the cause of the lower efficiency of amplification in many individuals. No primer set had to be redesigned due to issues such as primer dimer formation. The MSFCB primer pair showed successful amplification across a wide range of python species and the difference in amplification results of the moderately closely related taxa (*Python reticulatus, P. curtus*) between the 64 °C and 67 °C annealing temperatures demonstrates the relationship between annealing temperature and primer stringency well. Section 3.5.7.7 and [25] in 3.5.13 of chapter 3 provide an example where poor primer design (or rather overlooked detail in a particular primer sequence) led to suboptimal amplification results failing quality requirements for a forensic assay, although the primer set could provide added utility in the context of certain population genetic questions.

2.6.2 Which locus is most suitable?

When designing a DNA test for species identification from an unknown sample for a particular group of taxa, snakes for example, the desired approach is to find a DNA region exhibiting conservation within one species and variability between that and all other species. Mitochondrial DNA is a preferable target for degraded samples. It contains 38 functional domains that exhibit varying mutation rates, which impact on the usefulness of the genetic sequence for species testing. The non-coding D-loop, for example, is generally not considered suitable for species identification assays as its increased mutation rate creates too much variability within a species. Protein coding sequences have highly conserved domains due to functional constraints imposed on the translated protein, yet can also contain highly variable stretches in areas where the sequence does not impact as rigorously on the protein structure. The degree of sequence conservation tends to oscillate across the gene, as can be seen in the results of this study. This oscillating pattern of variability is ideal for designing a species identification assay for a group of taxa because if the target region exhibiting high variability is flanked by conserved sequence, universal primers can be designed to amplify product across the various taxa, providing efficiency in assay design and allowing the test to be multiplexed in a single tube at a later stage.

Most often when species identification tests are applied, primers universal across all taxa (e.g. cytochrome b,[51]; COI,[52]) are applied to the unknown sample and the resulting sequence searched against an online database. Are the DNA region amplified by these commonly used primers really able to distinguish between every species in the world? Are there regions of mitochondrial DNA that are more suitable for particular taxa? Although this has been suggested in the Lepidopteran studies that led to adoption of COI by the Barcoding of Life consortium [53], there has been much debate around the subject as discussed in the introduction of this thesis. Sequence is conserved by functionality, so different taxa would be expected to exhibit similar degrees of conservation for a particular DNA sequence, reflective of the functional constraint on the translated protein. Yet, different taxa are at various stages of evolutionary divergence over differing temporal scales, so one gene is not going to reflect the different divergence histories and trajectories of all taxa equally. This has been demonstrated by phylogenetic reconstructions of various taxonomic groups using multiple genes (e.g. [39]), as discussed in the Introductory chapter.

This chapter examines the similarity of sequences for all protein coding genes in the mitochondrial genome. While this is useful specifically to identify the best region for snake species identification testing, one would expect similar trends across taxa so comparable studies in other taxonomic groups might show these regions to be superior for species testing across families. ND6 was chosen for further analysis due to its high variability and gene length amenable to standard sequencing reaction upper boundaries. Cytochrome b exhibited moderate levels of variability, but was chosen for further analysis because it is the most variable of the well characterised genes, allowing comparison to a greater number of existing sequences on Genbank. This gene pair has the added convenience that three overlapping primer pairs could be designed to amplify the entire sequence, creating redundancy in the overlaps if one sequencing reaction did not work as well.

One must also consider the availability of reliable reference sequences when conducting a species identification test. Cytochrome b and COI are popular choices because extensive sequence records are available online for sequence comparisons. However, if the genetic signal at these genes is

simply not suitable for differentiating between species, or the species of interest are not represented on these databanks, then obtaining and screening the required reference samples in-house might be a necessary option. Current database records from Genbank, BOLD, and indeed any resource that does not contain an underlying continuity record for sample identity and validity associated with the sequences, should always been treated with a degree of caution.

2.6.3 Design of species identification markers – a different approach

Previous species identification research has generally been dictated by casework or local enforcement needs to focus on a specific small number of diverse snake species (e.g. [32,54]). The method involves amplifying DNA from the species of interest, searching this sequence against an online database (e.g. Genbank) and reporting the closest matching results returned. Can the chosen gene region distinguish between very closely related species?

A previous study involving 11 species of snake reported that their primers successfully amplified in snake species across all major snake lineages (Typhlopidae, Boidae, Pythonidae, Colubridae, Elapidae, and Viperidae) and were therefore suitable for species identification testing [35]. The primers appear to distinguish between the chosen species well and identify a particular species to the exclusion of all others, but this could be symptomatic of an information deficit on the DNA database.

While the COI region is a popular choice due to the Barcode of Life (BOLD) project and extensively used for species characterisation, there are currently 224 species of snake on the searchable Genbank database listed as COI regions, assuming no typological errors (Genbank search by author, 03/11/12). There were estimated over 3300 species of snake as at 1 Feb 2012 (http://www.reptile-database.org/db-info/SpeciesStat.html, accessed 16/11/12), which suggests that less than 7% of snake species are represented at the COI gene on Genbank. As more species are added to the database and the presently missing species gaps are filled, I wonder whether the gene region being investigated would still adequately be able to separate individuals to species level. I expect that high match scores (e.g. >98%) would be obtained for very closely related species if the locus is not suitable for adequately separating sister species.

When comparing sequence similarity across all snakes at the cytochrome b and COI loci (refer to figure 1), cytochrome b exhibits greater levels of variability which I suggest is necessary for comparisons between very closely related species. The authors [35] have used a top down approach of starting with a very broad list of species to search on the databank that is presently relatively devoid of listings for all snake species known worldwide. I propose a bottom up approach to marker design for species identification. Legislation is based on species definitions, so markers for identity of one particular species must be able to differentiate between *all* closely related species.

It is beneficial to use two datasets of differing resolutions when designing a species identification assay. The broad comparison of gene regions across all snake lineages is important to identify priming sites that are conserved across snakes and act as universal snake primers. To be increasingly confident that a DNA sequence is suitable for species identification – that is identifying one species to the exclusion of all other species, no matter how closely or distantly related – the DNA sequence must be interrogated at a fine scale level, comparing the species of interest against its most closely

related species. I suggest congeneric and confamilial comparisons are appropriate to validate the ability of a sequence to identify to species level.

One would expect that genes with enough variability to distinguish between closely related species will also separate more distantly related species. As genetic databanks become increasingly populated with species, the markers will continue to distinguish between species because they were designed and validated using the most closely related species. Of course, this sort of marker development is time consuming and only possible when considerable time and resources can be devoted to building sample size of closely related samples – often not the situation when forensic casework on a seldom studied species is requested with a short timeframe.

2.6.4 Genetic Distance Analysis – Validation Versus Application

As sequences diverge from a common ancestor over time, multiple mutations accumulate at a single base site. States intermediate to the ancestral state are masked by subsequent mutations. For this reason, when a simple pairwise comparison (p-distance) is used to determine the genetic distance between two sequences, the resulting value is an underestimation of the true genetic distance between the two sequences. Phylogenetic models of nucleotide substitution are applied to correct for this underestimate [55].

A range of models have been developed, applying increasing numbers of parameters in order to attain a truer value of genetic distance, but in turn becoming increasingly complicated. I applied four time-homogeneous, time-continuous Markov models to look at how the genetic distances changed for intra-species and inter-species comparisons at our study locus. The rate of nucleotide substitution at each codon position was considered equal for these calculations. As expected, genetic distance increased as more complicated models were applied. The increase in the gap between intra-species sequence comparisons and inter-species sequence comparisons provides support that this gene region is appropriate for distinguishing between sequences originating from members of different species.

As models increase in complexity, the degree of variance in the final genetic distance value is a culmination of the variance associated with each additional parameter. An increasingly complicated model will result in more standard error in the final value. Error rates should be minimised when applying forensic tests to aid criminal investigations. The standard species identification test involves DNA extraction from the unknown source, amplification with published species identification primers, DNA sequencing and comparison of sequence homology against an online databank of genetic sequences. A database search, such as a BLAST search, essentially performs a p-distance comparison (simple % similarity base pair comparison) of the query sequence against the databank sequence records. This is essentially the most conservative approach to species identification with a set of published primers, shown during this study to give the least distance between intra-species and inter-species comparisons, as well as the lowest associated error.

Every species identification test using a new locus should be validated prior to application. Validation testing could include the application of increasingly complex phylogenetic models to confirm that the DNA region is appropriate for identification of the species in question, however in this study increasingly complex genetic models were seen to invariably extend the gap between inter-species

and intra-species homology. More case studies of this type would be useful to confirm that this is always the case, negating the need for this process in routine validation of new loci.

2.6.5 Forensic analysis of degraded remains requires short targets

While this study presents information using whole gene charts and averages across genes, the need for small fragment analysis in degraded remains means that charts are useful to hone in on small regions of variability flanked by conserved sequence appropriate for placing universal primers.

Reconstruction of trees using decreasing gene fragments demonstrated that the third quarter of Cyt b creates a more accurate tree than the last quarter of ND6. This indicates that this this region is more appropriate for species identification testing. This is fortuitous because the cyt b locus has a more extensive history of use in species identification tests and a wildlife forensic lab picking up this assay might prefer to use a cytochrome b marker than another locus, even though the reference sequences for all *Morelia* species have been made available for other researchers on Genbank. The species specificity test gave positive amplification product in Boids and Elapids. If a practitioner were to apply these tests to snakes in these families, then they would be much more likely to find online reference sequences at the cyt b locus than the ND6 locus.

The ability of the smaller assay to obtain an amplification product when a larger target amplicon could not be obtained was conclusively demonstrated in this study. The DNA sample of interest ABTC 12155 gave a DNA quantification of 1.72 ng/uL, which could be considered a reasonable amount for mitochondrial analysis if the DNA extract is in pristine condition. It is possible that the DNA in the sample either originated from bacterial contamination (not differentiated by Qubit fluorimetry), or is in poor condition and highly fragmented. Both of these scenarios are also very likely to arise in casework samples, where products such as processed snake derived leathers of protected species are seized by border control during their journey through illegal international trade [9,56–58]. The fragmented DNA remaining in these samples is likely too incomplete to be amplified by a standard universal species identification test (e.g. [52] COI primers amplify a 710 bp product). Assays targeting small yet discriminatory regions are required to obtain results from these poor quality samples. The smaller MSFCB primer pair was designed for this reason and successfully amplified a product from sample 12155 when three sets of primers targeting larger products at the same locus could not. This sample also failed to give any amplification product using the nuclear profiling markers developed in Chapter 3, confirming the poor quality of this sample.

2.6.6 What of *Morelia bredli*?

Two species examined in this study, *Morelia spilota* and *Morelia bredli*, were seen not to separate in any of the datasets examined. This demonstrates the importance of including every species within a taxonomic group. I deduce from the successful differentiation of all other congeneric species, as well as other literature [59,60], that this result is not an inaccuracy of the genetic test, but a point of contention in the Bredl Python's designation as a separate species. It is important that a forensic geneticist realises that a developed forensic genetic test can only identify species descriptions that are actually reflected in the DNA sequence. The distinction of *M. bredli* as a separate species to *M. spilota* does not appear to be supported by the mitochondrial locus studies herein, it was not supported by the nuclear data of Taylor [60] and it will be interesting to see whether the novel STR loci developed as part of this thesis also concur with this finding.

The unfortunate reality is that taxonomic descriptions can be based on morphological traits that are due to phenotypic plasticity rather than the underlying genetic code. The courts require forensic genetic tests to answer questions about legislation involving species definitions that are not necessarily reflected in the underlying genetic data. Validation is essential to determine whether the genetic test is fit for purpose in the species of interest.

2.6.7 Future directions and further species identification tests

The dataset presented in 2.5.2 provides the perfect foundation to develop a SNP multiplex assay for species identification to preclude the need for DNA sequencing. This type of assay has been developed at a broad level for UK [61] and New Zealand [62] mammals of forensic and conservation significance (these also represent common native or domestic species that might arise in forensic casework in each respective country). A similar species-specific multiplex assay has been developed targeting product only in tiger species and containing markers specific to tiger subspecies, acting as a diagnostic test for detection on both of these taxonomic levels [15]. My dataset provides a direct comparison of sequence homology and variability across different snake species at various taxonomic levels of genus, subfamily, family and superfamily (only three samples represent the infraorder Scolecophidia, so more sequences would be required to create tests to differentiate the two infraorders, however forensically there is no current need for this). Conserved regions of genes could be identified specific to any of these taxonomic levels.

As of the early 2000s, the species involved in the marker development part of this study, *Xenopeltis unicolor*, *Acrochrodus granulatus*, *Enhydris plumbea*, as well as species of the genera *Elaphe* and *Bungarus* were protected by Annex D of the EU Wildlife Trade Regulations (table 4 in [63]). Theoretically, using this dataset to develop Allele Specific Primer Extension assays, a forward primer universal to colubrids could be multiplexed with species-specific primers identified for *Elaphe* and *Enhydris*, which could be multiplexed with similar universal and species-specific primer pairs for elapids and the other taxa of relevance, to create a species identification SNP multiplex serving snakes listed in the EU Wildlife Trade Regulations. These Regulations have particularly strict import conditions [63] and the multiplex assay could potentially be used by border control to detect protected snake species in trade items. Single Base Extension primers could be designed to amplify tiny products containing SNPs diagnostic of target species to provide a positive/negative result. This type of test has potential to be automated into a portable screening device that could be used by border force. A positive result would prompt officials to confiscate the item and send to a forensic facility for more extensive species identification testing and reporting. The tiny amplicon size makes this test particularly suitable for the type of treated and processed products such as tanned skins and powdered tissues that are likely to be encountered in international trade.

2.7 Concluding remarks

Legislation pertaining to wildlife crimes is often based on species definitions, so species identification markers must be designed to differentiate between all closely related species. While the idea of a universal barcoding locus is nice, in practice not all loci are suitable for all taxa and the phylogenetic relationships of closely related species must be investigated at a locus before it is applied to casework involving that species. A bottom up approach for species identification assay

design is suggested instead of top down approach to avoid problems as reference databases become increasingly populated. The comparison of two datasets of different taxonomic resolutions is useful to identify conserved areas and design universal primers that can amplify product across broad taxonomic units, while also identifying amplified regions that are variable enough to be able to distinguish the desired targets to species level. The phylogenetic signal allowing a DNA sequence to reconstruct accurate phylogenetic relationships of species might be captured within a very small region of a gene, rather than the entire locus. This region offers great utility for examining samples of compromised DNA quality that often constitute evidentiary items in wildlife forensic investigations. In this chapter, I have demonstrated the design of a small species identification assay to identify python species from evidentiary items in forensic investigations using an approach that will ensure the ongoing integrity of test by demonstration of its utility in all species of the *Morelia* genus. Forensic validation testing has been conducted to ready this assay for direct application to judicial proceedings. All reference sequences have been uploaded to the online database Genbank so that any wildlife forensic laboratory can directly apply this test to forensic investigations involving *Morelia* pythons.

2.8 References

[1] Z.H. Zhou, Z.G. Jiang, International trade status and crisis for snake species in China, Conserv. Biol. 18 (2004) 1386–1394.

[2] S. Dharmananda, The medicinal use of snakes in China, 14th September 2013] (1997). http://www.itmonline.org/arts/snakes.htm.

[3] X. Qiao, M. Ye, D. Pan, W. Miao, C. Xiang, J. Han, D. Guo, Differentiation of various traditional Chinese medicines derived from animal bile and gallstone: Simultaneous determination of bile acids by liquid chromatography coupled with triple quadrupole mass spectrometry, J. Chromatogr. A. 1218 (2011) 107–117.

[4] T.C. Chao, M.L. Wu, W.J. Tsai, J. Ger, J.F. Deng, Acute hepatic injury and renal failure after ingestion of snake gallbladder, Clin. Toxicol. 44 (2006) 387–390.

[5] Y.-H. Yeh, D.-Y. Wang, M.-Y. Liau, M.-L. Wu, J.-F. Deng, T. Noguchi, D.-F. Hwang, Bile acid composition in snake bile juice and toxicity of snake bile acids to rats, Comp. Biochem. Physiol. Part C Toxicol. Pharmacol. 136 (2003) 277–284.

[6] B. Dubey, P.R. Meganathan, I. Haque, Multiplex PCR assay for rapid identification of three endangered snake species of India, Conserv. Genet. 10 (2009) 1861–1864.

[7] R.R. da Nobrega Alves, G.A. Pereira Filho, Commercialization and use of snakes in North and Northeastern Brazil: Implications for conservation and management, Biodivers. Conserv. 16 (2007) 969–985.

[8] Australian Crime Commission, A. Government, Organised crime in Australia 2011, (2011). http://www.crimecommission.gov.au/sites/default/files/files/OCA/2011/oca2011.pdf.

[9] J. Holden, By Hook or by Crook: A reference manual on illegal wildlife trade and prosecutions in the United Kingdom, 1998.

[10] S. Bricknell, Environmental crime in Australia, Res. Public Policy Ser. 109. (2010). http://www.aic.gov.au/documents/2/1/1/%7B211B5EB9-E888-4D26-AED4-1D4E76646E4B%7Drpp109.pdf.

[11] INTERPOL, INTERPOL co-ordinated operation targeting illegal trade in endangered reptiles leads to arrests and seizures worldwide, (2010). http://www.interpol.int/News-and-media/News/2010/PR089.

[12] C. Peddie, INTERPOL Organised crime behind illegal animal trade Wildlife terrorists, Advert. (2007) 3. http://search.proquest.com/docview/355049362?accountid=10910.

[13] J. Blindell, 21st Century Policing: The role of police in the detection, investigation and prosecution of environmental crime, ACPR Issues. 2 (2006) 1–8.

[14] R.N. Johnson, A.G. King, C. Vockler, G.M. Cooke, The use of DNA to reveal a potentially deadly package! Two case studies from the Australian Museum, in: Soc. Wildl. Forensic Sci. 1st Trienn. Meet., Jackson, Wyoming, 2012.

[15] T. Kitpipit, S.S. Tobe, A.C. Kitchener, P. Gill, A. Linacre, The development and validation of a single SNaPshot multiplex for tiger species and subspecies identification-Implications for forensic purposes, Forensic Sci. Int. 6 (2012) 250–257.

[16] F. Mayer, C. Dietz, A. Kiefer, Molecular species identification boosts bat diversity, Front. Zool. 4 (2007) 1–5.

[17] B. Balitzki-Korte, K. Anslinger, C. Bartsch, B. Rolf, Species identification by means of pyrosequencing the mitochondrial 12S rRNA gene, Int. J. Legal Med. 119 (2005) 291–294.

[18] T. Melton, C. Holland, Routine forensic use of the mitochondrial 12S ribosomal RNA gene for species identification, J. Forensic Sci. 52 (2007) 1305–1307.

[19] M. Vences, M. Thomas, R.M. Bonett, D.R. Vieites, Deciphering amphibian diversity through DNA barcoding: chances and challenges, Philos. Trans. R. Soc. B Biol. Sci. 360 (2005) 1859–1868. doi:10.1098/rstb.2005.1717.

[20] T. Mitani, A. Akane, T. Tokiyasu, S. Yoshimura, Y. Okii, M. Yoshida, Identification of animal species using the partial sequences in the mitochondrial 16S rRNA gene, Leg. Med. 11 (2009) S449–S450.

[21] C.W. Dieffenbach, T.M. Lowe, G.S. Dveksler, General concepts for PCR primer design., Genome Res. 3 (1993) S30–S37. doi:10.1101/gr.3.3.S30.

[22] V. Pelt-Verkuil, PCR Primers, Princ. Tech. Asp. PCR Amplifcation. (2008) 63–90.

[23] M. McPherson, S. Moller, PCR, Taylor & Francis, London, 2006.

[24] M. Mitsuhashi, Technical Report: Part 1. Basic Requirements for Designing Optimal Oligonucleotide Probe Sequences, J. Clin. Lab. Anal. 10 (1996) 277–284.

[25] S. Ciavaglia, S. Donnellan, S. Tobe, J. Henry, A. Linacre, A novel forensic DNA profiling technique for protected species, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e258–e260. doi:10.1016/j.fsigss.2015.09.103.

[26] D.P. Snustad, M.J. Simmons, Principles of Genetics, Wiley, 2000.

[27] W. Rychlik, W.J. Spencer, R.E. Rhoads, Optimization of the annealing temperature for DNA amplification *in vitro*, Nucleic Acids Res. 18 (1990) 6409–6412.

[28] J.J. SantaLucia, H.T. Allawi, P.A. Seneviratne, Improved nearest-neighbour parameters for predicting DNA duplex stability, Biochemistry. 35 (1996) 3555–3562.

[29] S. V. Suggs, T. Hirose, E.H. Myake, M.J. Kawashima, K.I. Johnson, R.B. Wallace, ICN-UCLA symposium for developmental biology using purified gene, in: D.D. Brown (Ed.), 1981: p. 23: 683-693.

[30] M. Mitsuhashi, Technical Report: Part 2. Basic Requirements for Designing Optimal PCR Primers, J. Clin. Lab. Anal. 10 (1996) 285–293.

[31] K. Hayashi, Manipulation of DNA by PCR, in: K.B. Mullis, F. Ferre, R.A. Gibbs (Eds.), Polym. Chain React., 1994: pp. 3–13.

[32] K.L. Wong, J. Wang, P.P.H. But, P.C. Shaw, Application of cytochrome b DNA sequences for the authentication of endangered snake species, Forensic Sci. Int. 139 (2004) 49–55.

[33] B. Dubey, P.R. Meganathan, I. Haque, Molecular identification of three Indian snake species using simple PCR-RFLP method, J. Forensic Sci. 55 (2010) 1065–1067.

[34] J.C.-I. Lee, L.-C. Tsai, S.-P. Liao, A. Linacre, H.-M. Hsieh, Species identification using the

cytochrome b gene of commercial turtle shells, Forensic Sci. Int. 3 (2009) 67–73.

[35] B. Dubey, P.R. Meganathan, I. Haque, DNA mini-barcoding: An approach for forensic identification of some endangered Indian snake species, Forensic Sci. Int. 5 (2011) 181–184.

[36] S. Dong, Y. Kumazawa, Complete mitochondrial DNA sequences of six snakes: Phylogenetic relationships and molecular evolution of genomic features, J. Mol. Evol. 61 (2005) 12–22.

[37] W. Parson, K. Pegoraro, H. Niederstatter, M. Foger, M. Steinlechner, Species identification by means of the cytochrome b gene, Int. J. Legal Med. 114 (2000) 23–28.

[38] H.M. Hsieh, H.L. Chiang, L.C. Tsai, S.Y. Lai, N.E. Huang, A. Linacre, J.C.I. Lee, Cytochrome b gene for species identification of the conservation animals, Forensic Sci. Int. 122 (2001) 7–18.

[39] S.S. Tobe, A.C. Kitchener, A.M.T. Linacre, Reconstructing mammalian phylogenies: A detailed comparison of the cytochrome b and cytochrome oxidase subunit i mitochondrial genes, PLoS One. 5 (2010). doi:10.1371/journal.pone.0014156.

[40] P. Uetz, The Reptile Database, October 10 (n.d.). http://www.reptile-database.org.

[41] A. Linacre, L. Gusmao, W. Hecht, A.P. Hellmann, W.R. Mayr, W. Parson, M. Prinz, P.M. Schneider, N. Morling, ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations, Forensic Sci. Int. Genet. 5 (2011) 501–505.

[42] W.D. Schleip, M.O. Shea, Annotated checklist of the recent and extinct pythons (Serpentes, Pythonidae), with notes on nomenclature, taxonomy, and distribution, Zookeys. 66 (2010) 29–79.

[43] R Core Team: A language and environment for statistical computing, (2013). http://www.r-project.org/.

[44] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.

[45] T.D. Schneider, R.M. Stephens, Sequence logos : a new way to display consensus sequences, 18 (1990) 6097–6100.

[46] S. Rosen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, in: Bioinforma. Methods Protoc., Humana Press, Totowa, NJ, 2000: pp. 365–386.

[47] S. Ciavaglia, S. Donnellan, J. Henry, A. Linacre, Species identification of protected carpet pythons suitable for degraded forensic samples, Forensic Sci. Med. Pathol. 10 (2014) 295–305.

[48] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: Gatew. Comput. Environ. Work. (GCE), 2010, 2010: pp. 1–8.

[49] R. Lanfear, B. Calcott, S.Y.W. Ho, S. Guindon, PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses, Mol. Biol. Evol. 29 (2012) 1695–1701.

[50] S.A. Ciavaglia, S.S. Tobe, S.C. Donnellan, J.M. Henry, A.M.T. Linacre, Molecular identification of python species: Development and validation of a novel assay for forensic investigations, Forensic Sci. Int. Genet. 16 (2015).

[51] T.D. Kocher, W.K. Thomas, A. Meyer, S. V. Edwards, S. Paabo, F.X. Villablanca, A.C. Wilson, Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers., Proc. Natl. Acad. Sci. 86 (1989) 6196–6200.

[52] O. Folmer, M. BLACK, W. HOEH, R. Lutz, R. Vrijenhoek, DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates, Mol. Mar. Biol. Biotechnol. 3 (1994) 294–299.

[53] B. Brown, R.M. Emberson, A.M. Paterson, Mitochondrial COI and II provide useful markers for *Weiseana* (Lepidoptera, Hepialidae) species identification, Bull. Entomol. Res. 89 (1999) 287–294.

[54] P.R. Meganathan, B. Dubey, K.N. Jogayya, I. Haque, Validation of a multiplex PCR assay for the forensic identification of Indian crocodiles, J. Forensic Sci. 56 (2011) 1241–1244.

[55] P. Lemey, M. Salemi, A.-M. Vandamme, The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing, 2nd edition, Cambridge University Press, 2009.

[56] M. Jenkins, S. Broad, International Trade in Reptile Skins: A Review and Analysis of the Main Consumer Markets, 1983-1991, 1994.

[57] J. Koegh, D. Barker, R. Shine, Exploited but poorly known: systematics and biogeography of commercially harvested python, Biol. J. Linn. Soc. (2001) 113–129.

[58] N.S. Sodhi, L.P. Koh, B.W. Brook, P.K.L. Ng, Southeast Asian biodiversity: an impending disaster, TRENDS Ecol. Evol. 19 (2004) 654–660.

[59] G.F. Gow, *Python bredli*, new species from central Australia, Aust. J. Herpetol. 1 (1984) 29–34.

[60] D. Taylor, Using DNA markers for wildlife management and protection: a study of the population structure and systematics of the Australian carpet pythons (Reptilia: *Morelia spilota* complex), Flinders University, SA, 2005.

[61] S.S. Tobe, A.M.T. Linacre, A multiplex assay to identify 18 European mammal species from mixtures using the mitochondrial cytochromeb gene, Electrophoresis. 29 (2008) 340–347.

[62] A. Ramón-Laca, A.M.T. Linacre, D.M. Gleeson, S.S. Tobe, Identification multiplex assay of 19 terrestrial mammal species present in New Zealand, Electrophoresis. 34 (2013) 3370–3376.

[63] M. Auliya, Hot Trade in Cool Creatures: A Review of the Live Reptile Trade in the European Union in the 1990s with a focus on Germany, Brussels, Belgium, 2003.

Appendix 2.1: Publication - Species identification of protected carpet pythons suitable for degraded forensic samples

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Appendix 2.2: Publication - Molecular identification of python species: Development and validation of a novel assay for forensic investigations

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

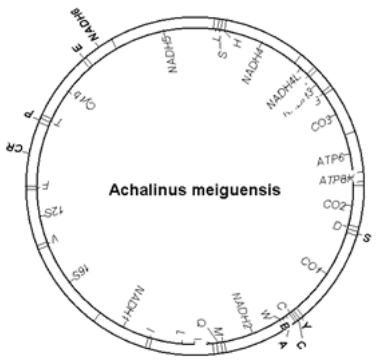
Publication has been removed due to copyright restrictions.

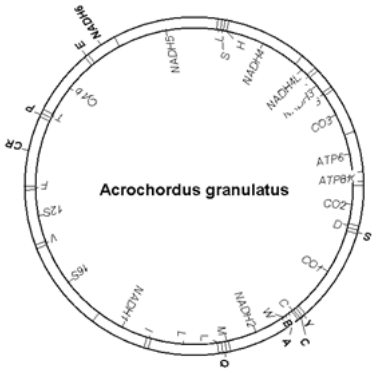
Publication has been removed due to copyright restrictions.

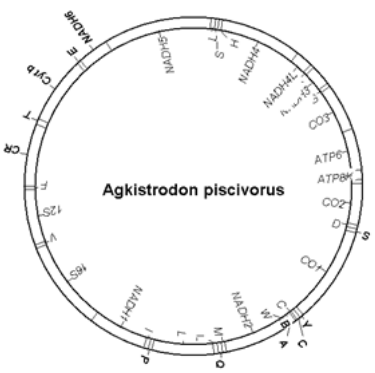
Publication has been removed due to copyright restrictions.

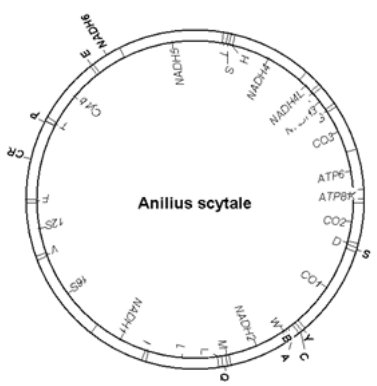
Publication has been removed due to copyright restrictions.

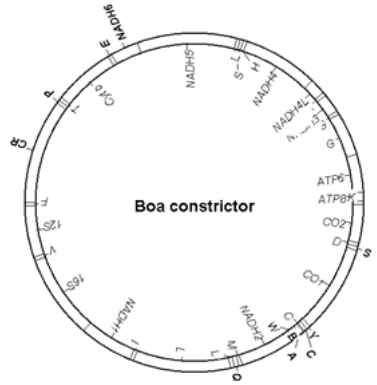
Appendix 2.3: Snake whole mitochondrial genome maps

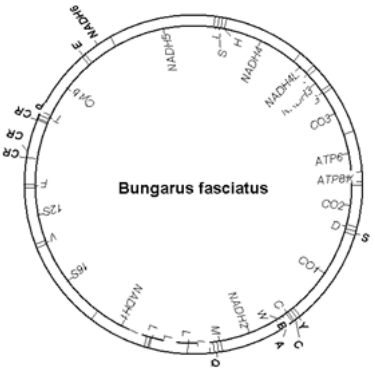


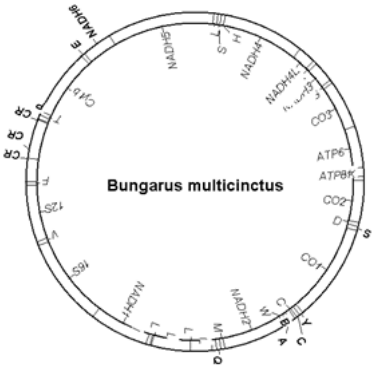


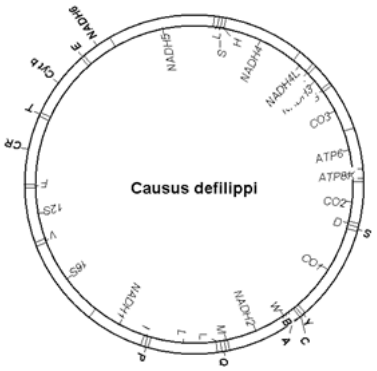


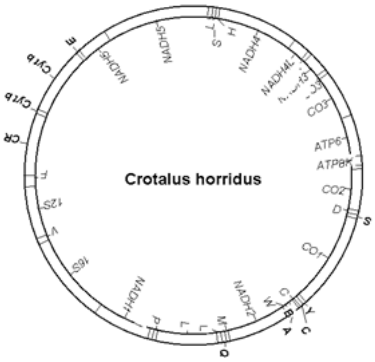


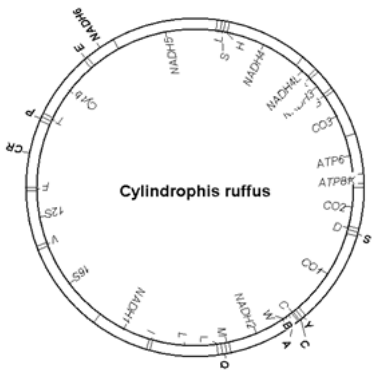


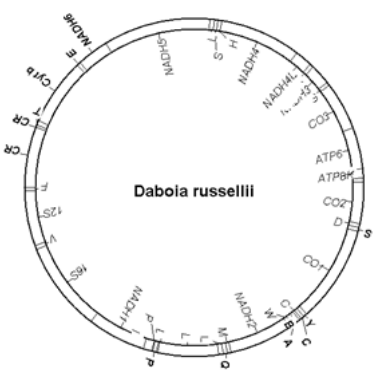


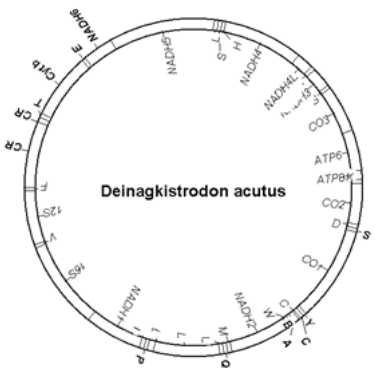


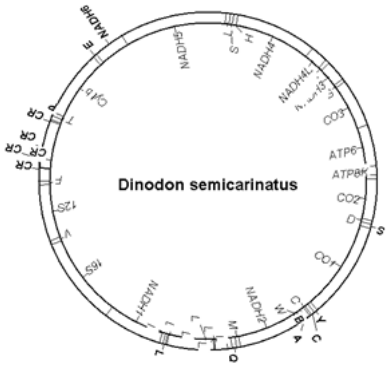


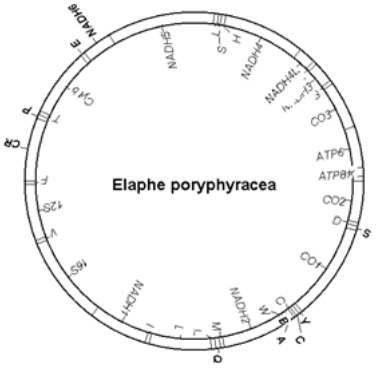


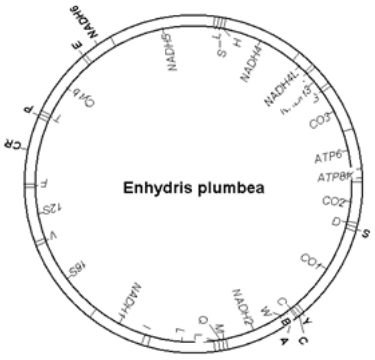


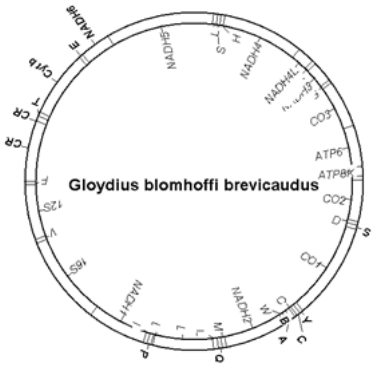


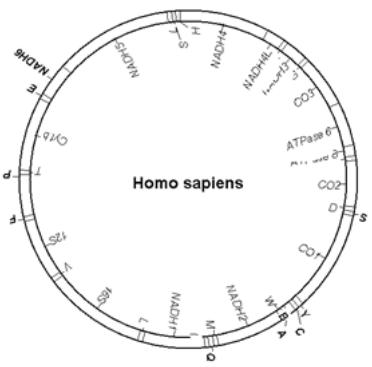


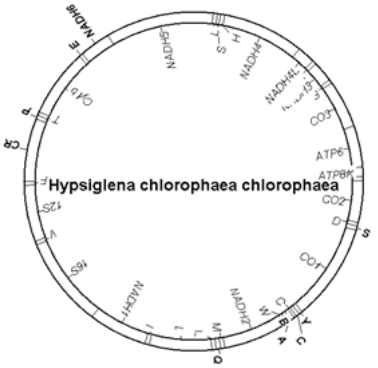

Dinodon semicarinatus

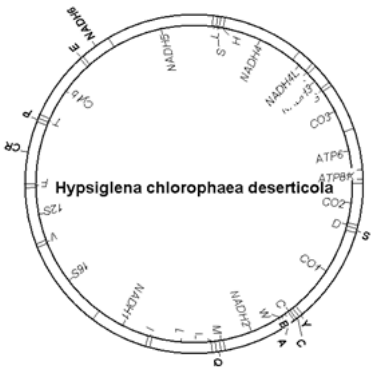

Elaphe poryphyracea

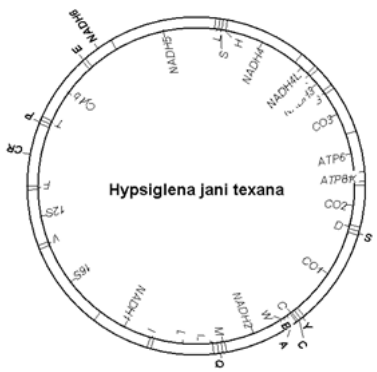

Enhydris plumbea

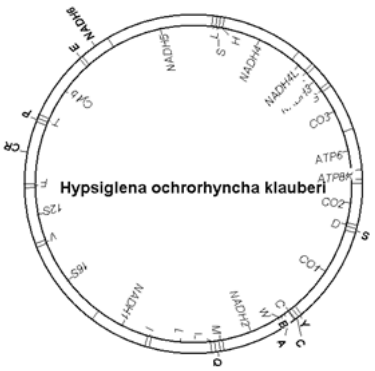

Gloydius blomhoffi brevicaudus

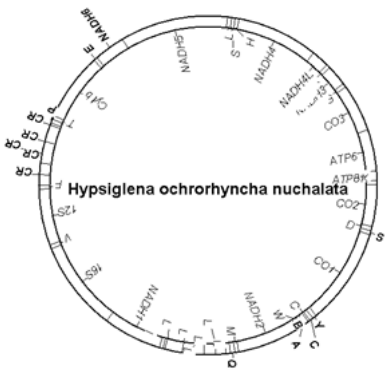

Homo sapiens


Hypsiglena chlorophaea chlorophaea

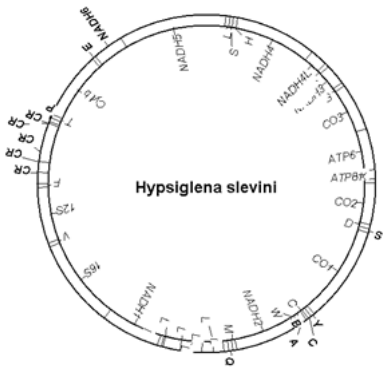

Hypsiglena chlorophaea deserticola

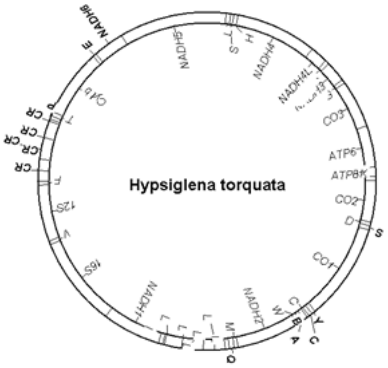

Hypsiglena jani texana

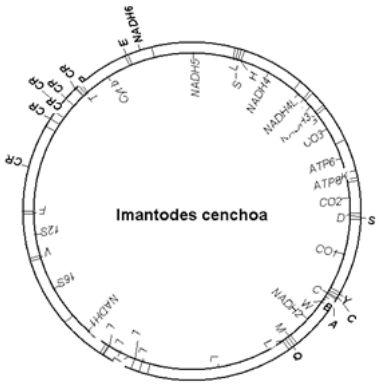

Hypsiglena ochrorhyncha klauberi

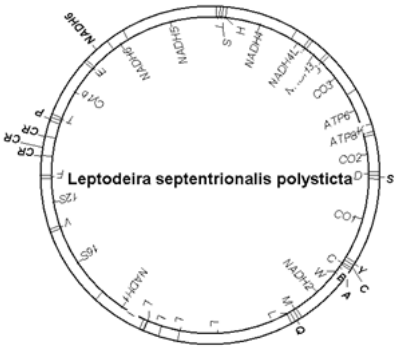

Hypsiglena ochrorhyncha nuchalata

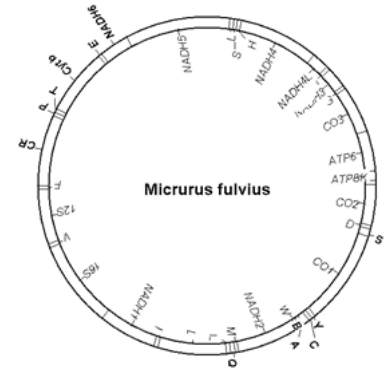

Hypsiglena ochrorhyncha ochrorhyncha

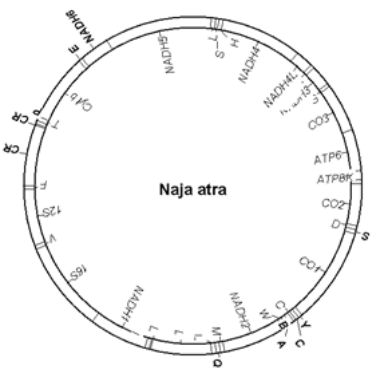

Hypsiglena slevini

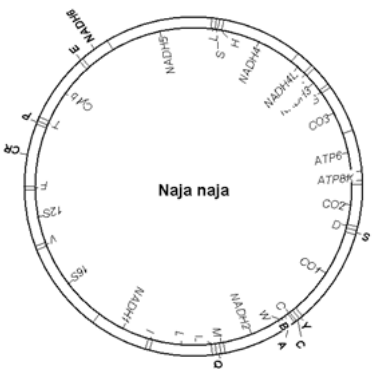

Hypsiglena torquata

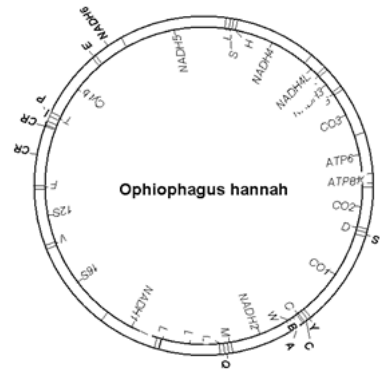

Imantodes cenchoa

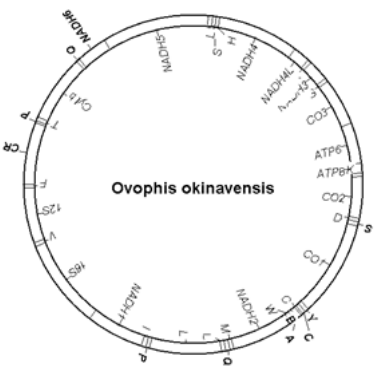

Leptodeira septentrionalis polysticta

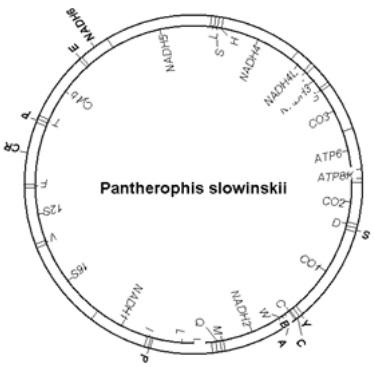

Micrurus fulvius

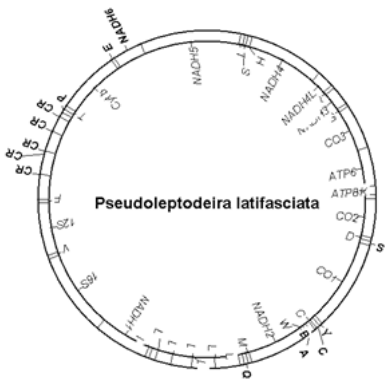

Naja atra

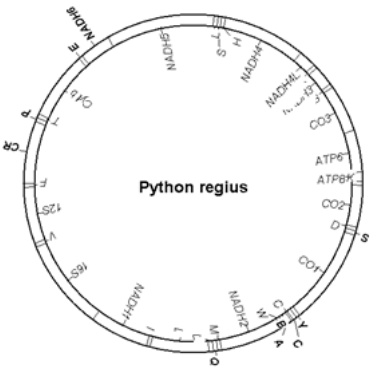

Naja naja

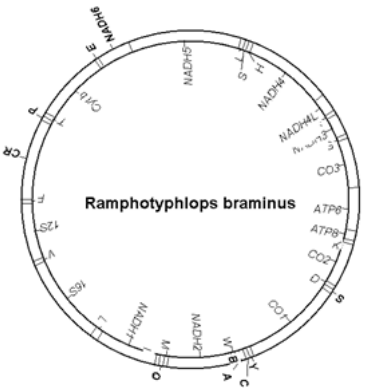

Ophiophagus hannah

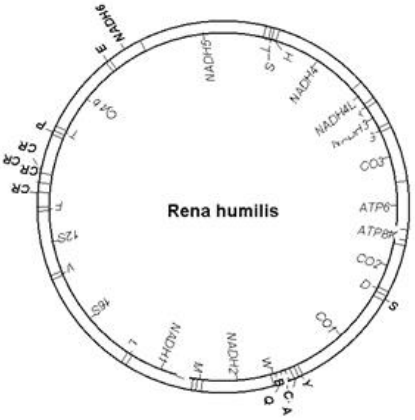

Ovophis okinavensis

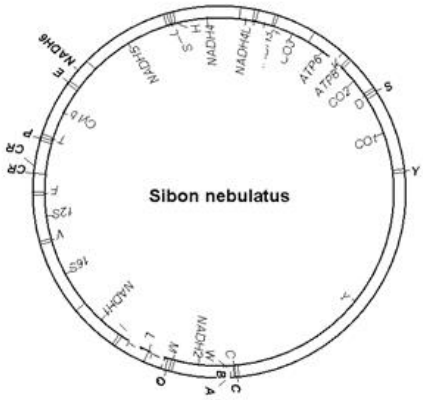

Pantherophis slowinskii

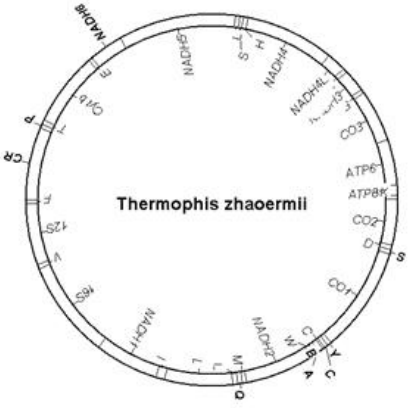

Pseudoleptodeira latifasciata

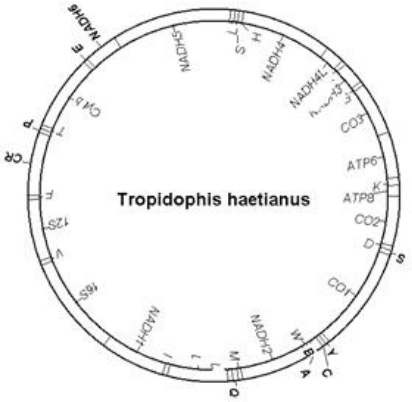

Python regius

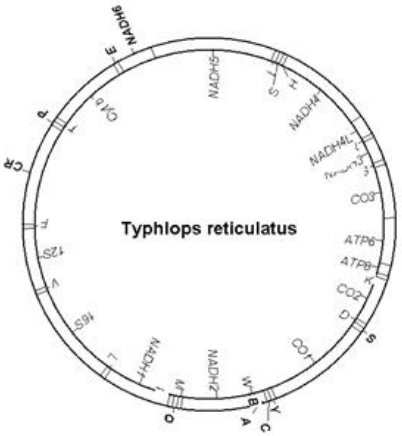

Ramphotyphlops braminus


Rena humilis

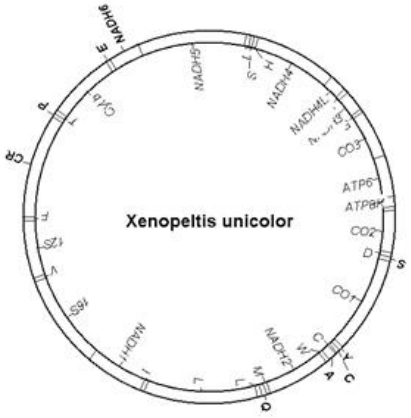

Sibon nebulatus


Thermophis zhaoermii

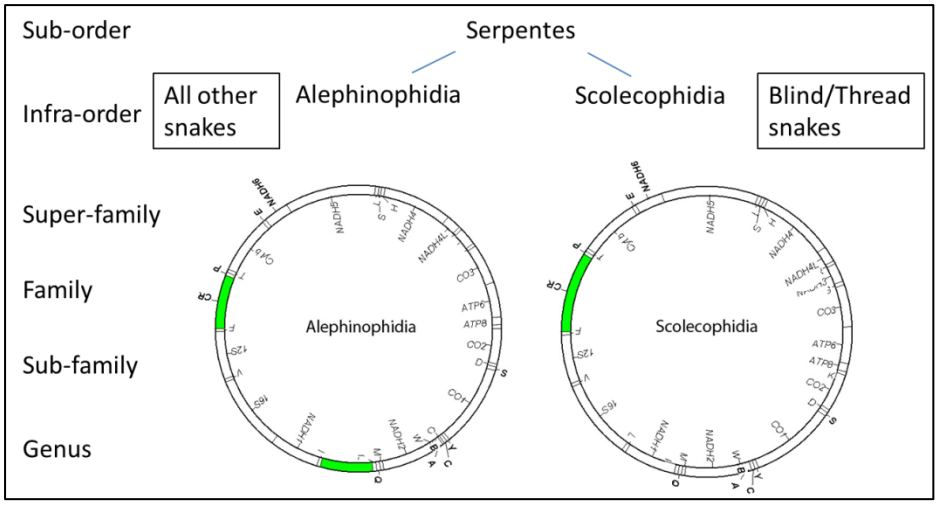

Tropidophis haetianus

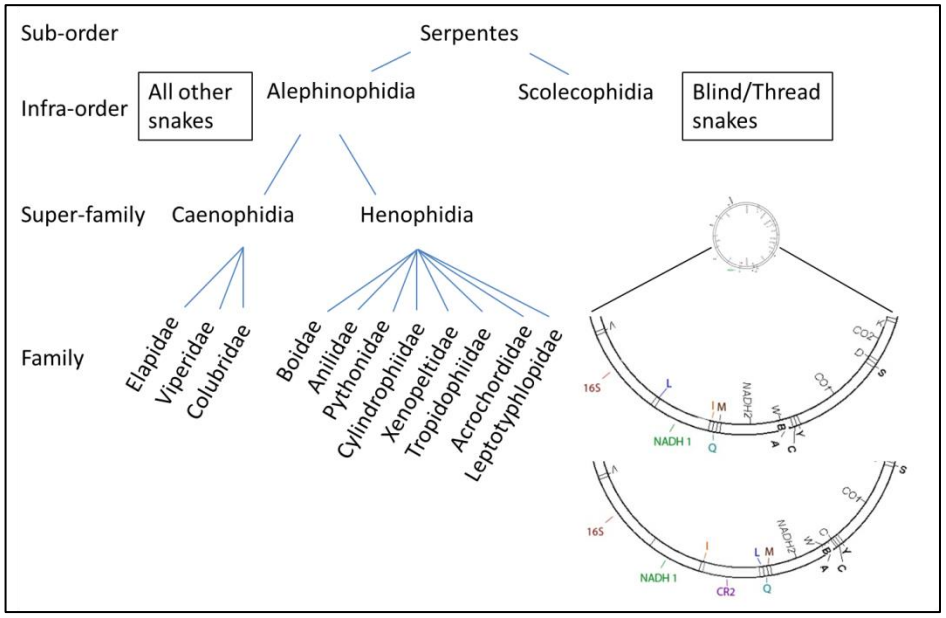

Typhlops reticulatus

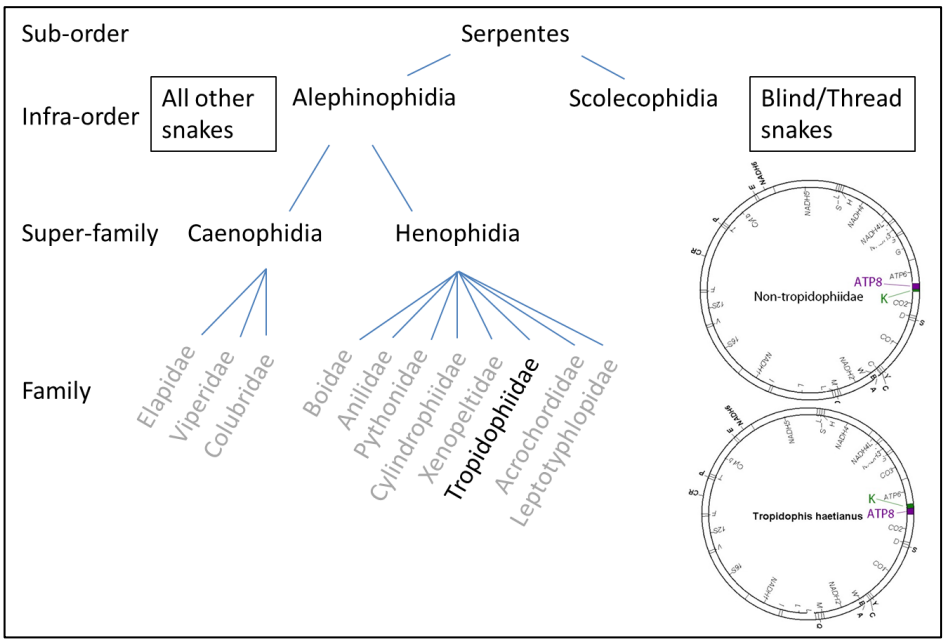

Viridovipera stejnegeri stejnegeri

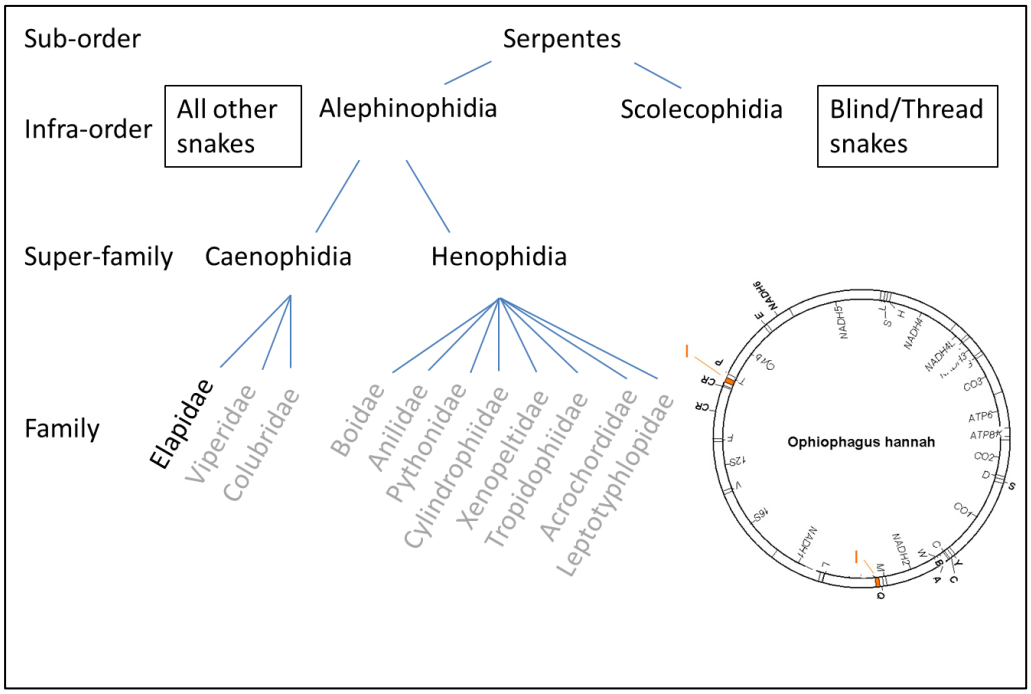

Xenopeltis unicolor

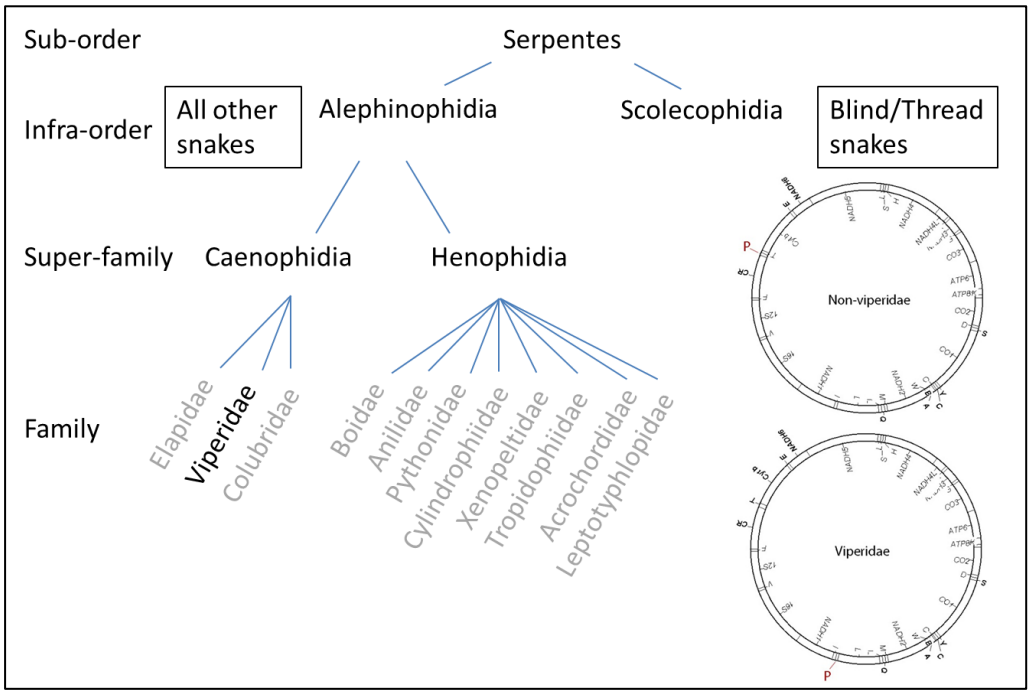
Appendix 2.4: Mitochondrial genome duplications and translocations between snakes

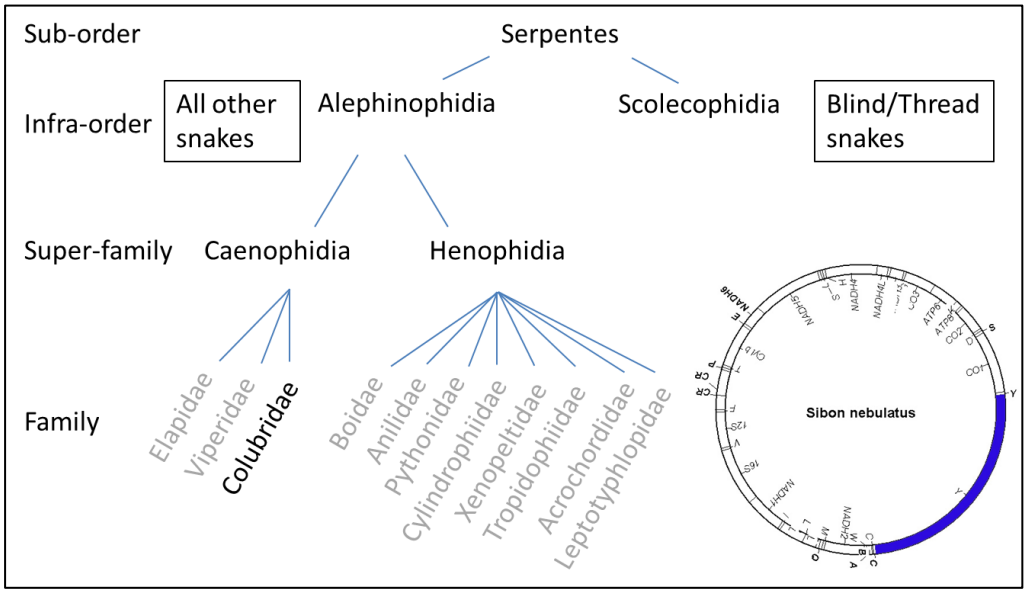




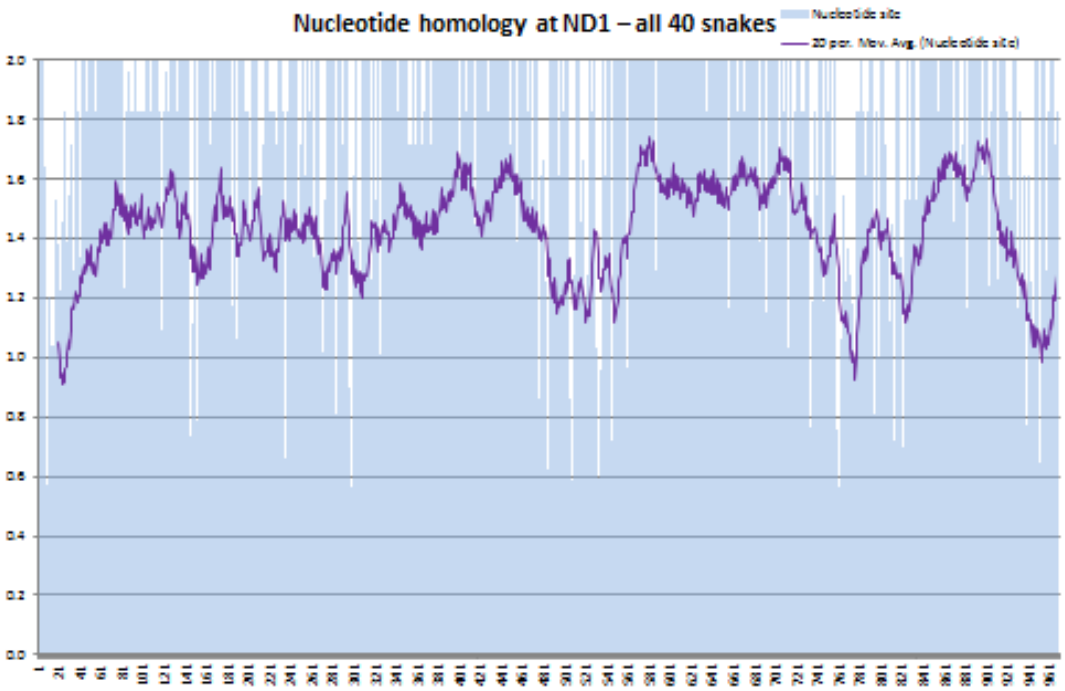


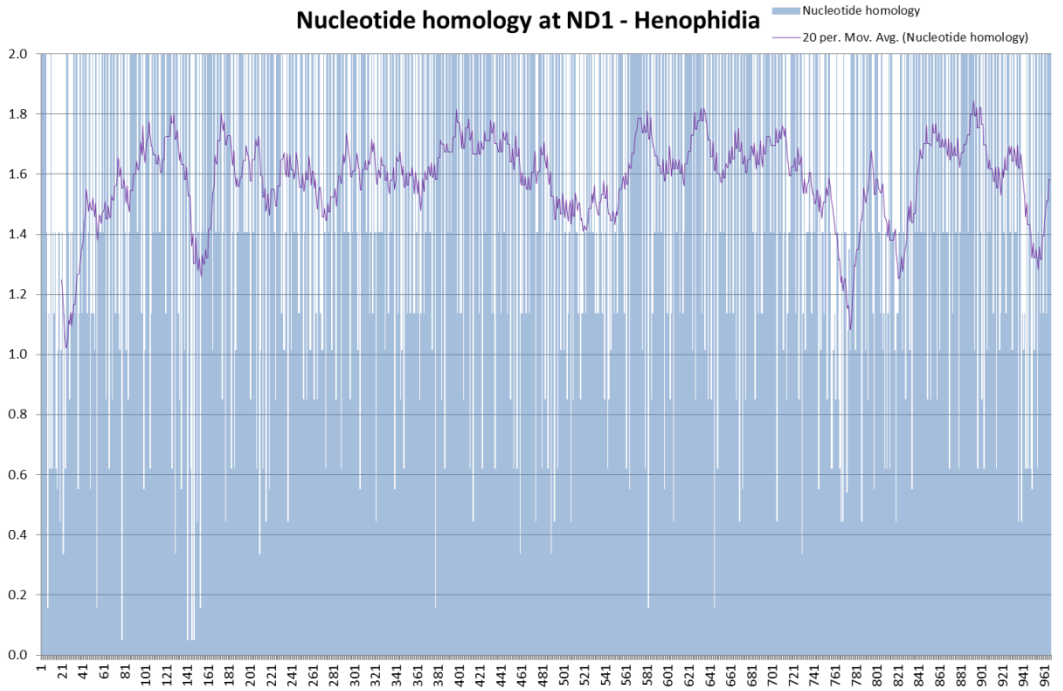

Sub-order
Serpentes
Infra-order
All other snakes
Alephinophidia
Scolecophidia
Blind/Thread snakes
Super-family
Caenophidia
Henophidia
Family
Elapidae
Viperidae
Colubridae
Boidae
Anilidae
Pythonidae
Cylindrophiidae
Xenopeltidae
Tropidophiidae
Acrochordidae
Leptotyphlopidae
Ophiophagus hannah

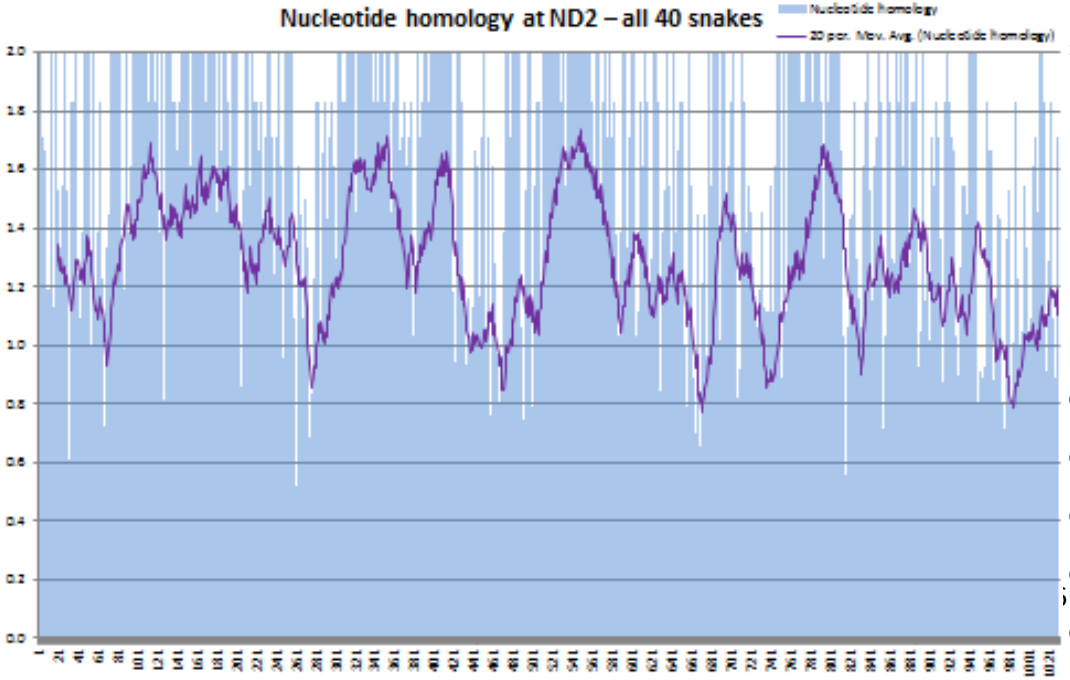

Sub-order
Serpentes
Infra-order
All other snakes
Alephinophidia
Scolecophidia
Blind/Thread snakes
Super-family
Caenophidia
Henophidia
Family
Elapidae
Viperidae
Colubridae
Boidae
Anilidae
Pythonidae
Cylindrophiidae
Xenopeltidae
Tropidophiidae
Acrochordidae
Leptotyphlopidae
Non-viperidae
Viperidae

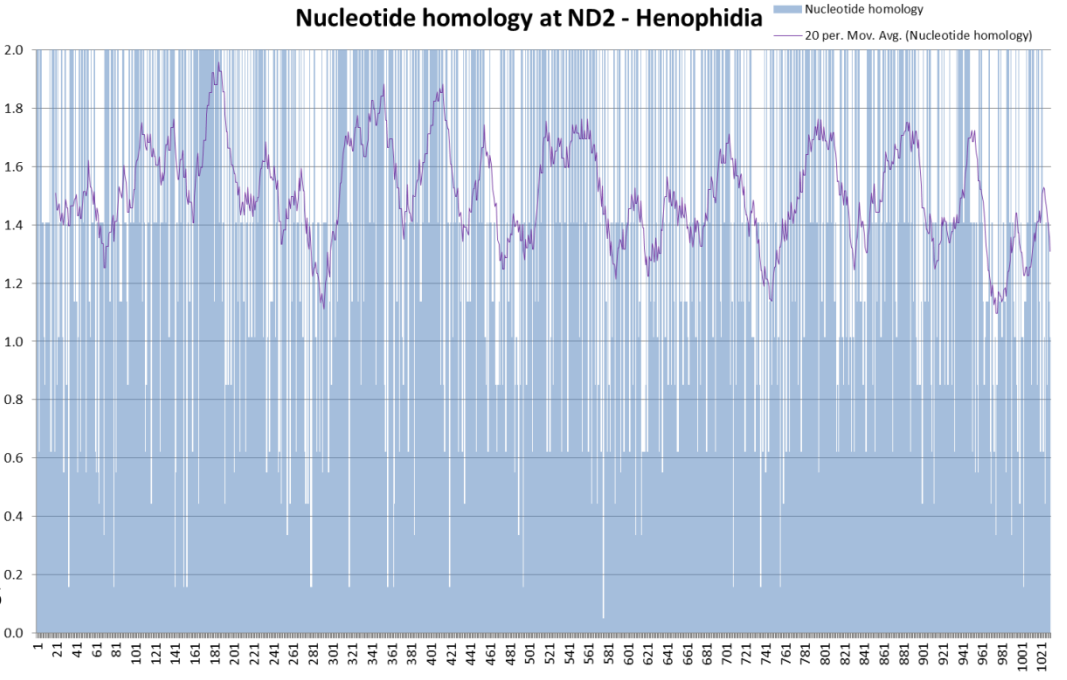

Sub-order
Serpentes
Infra-order
All other snakes
Alephinophidia
Scolecophidia
Blind/Thread snakes
Super-family
Caenophidia
Henophidia
Family
Elapidae
Viperidae
Colubridae
Boidae
Anilidae
Pythonidae
Cylindrophiidae
Xenopeltidae
Tropidophiidae
Acrochordidae
Leptotyphlopidae
Sibon nebulatus

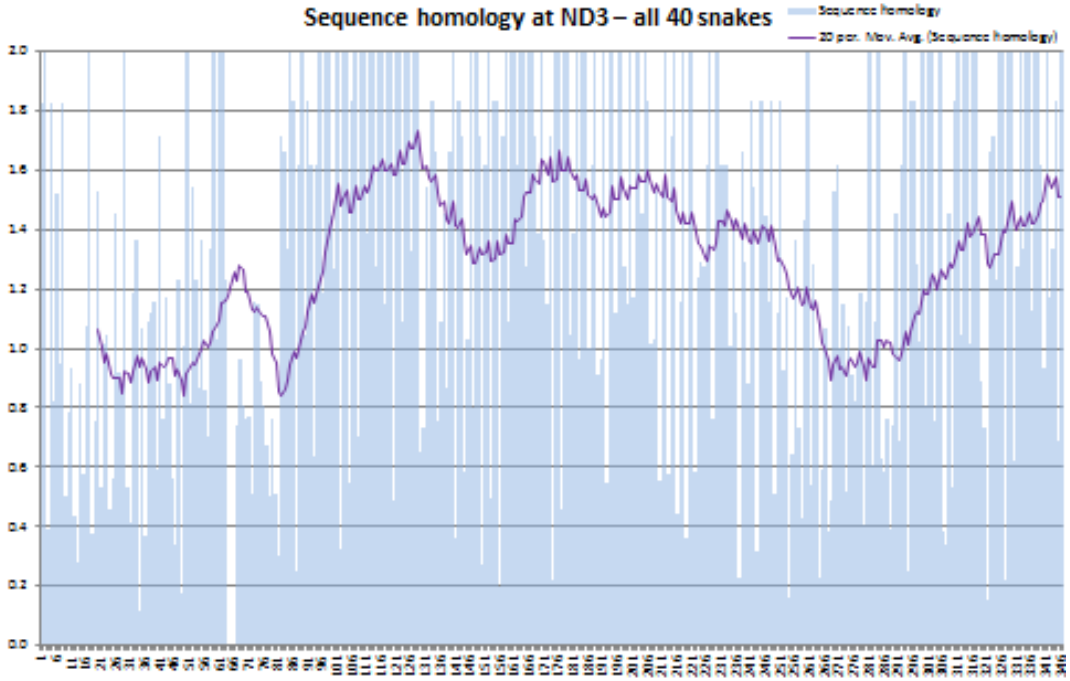
Appendix 2.5: Sequence homology charts by gene

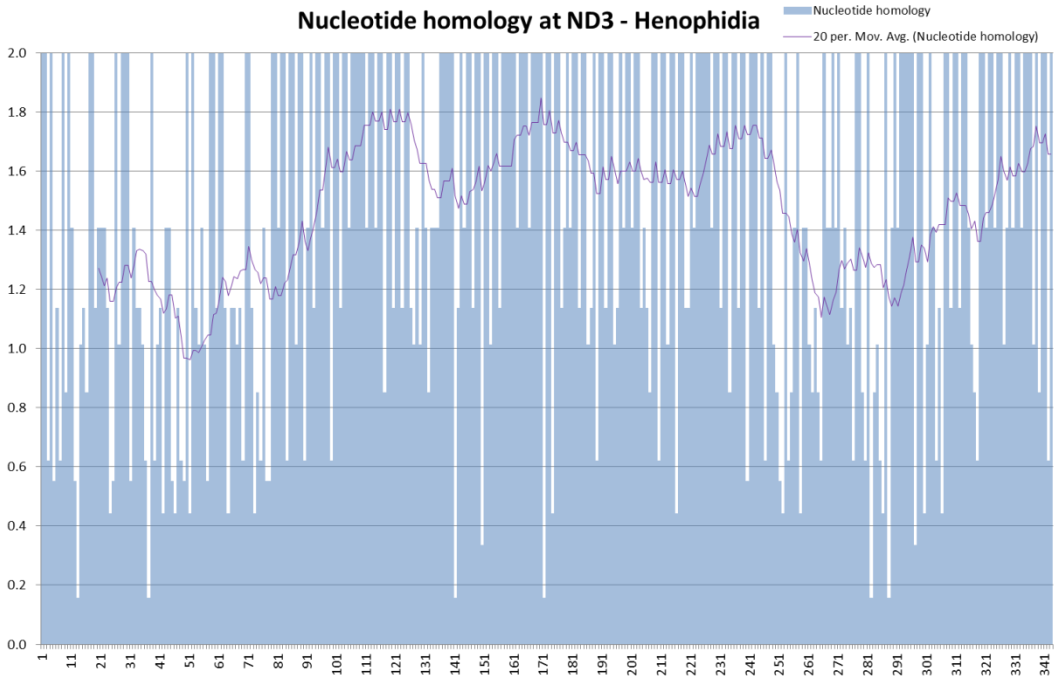


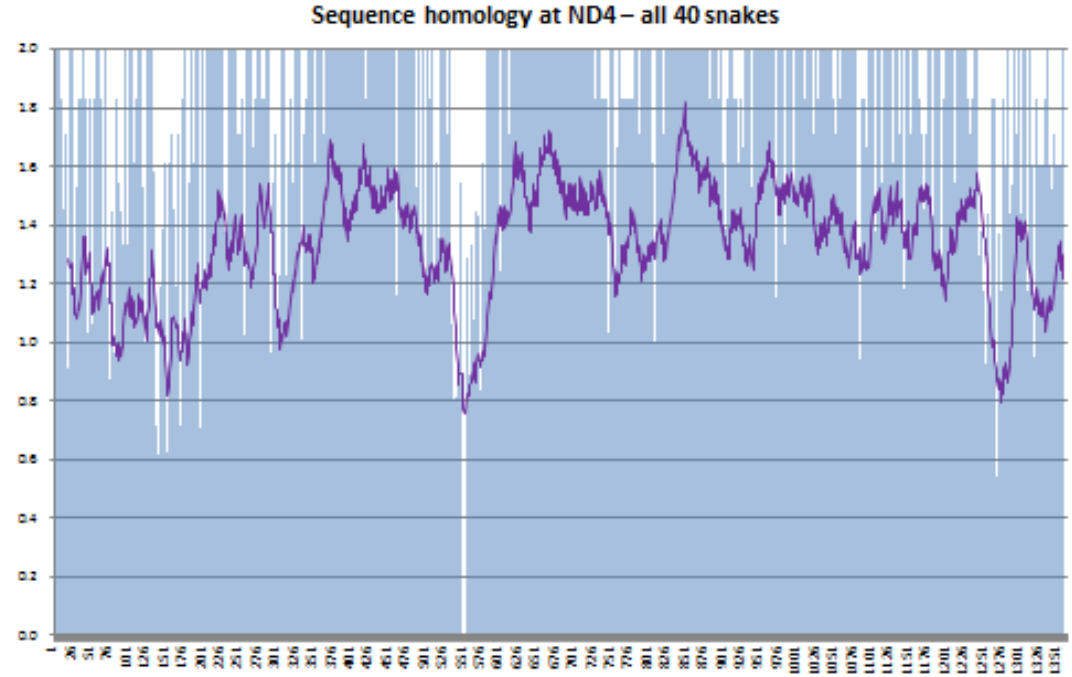


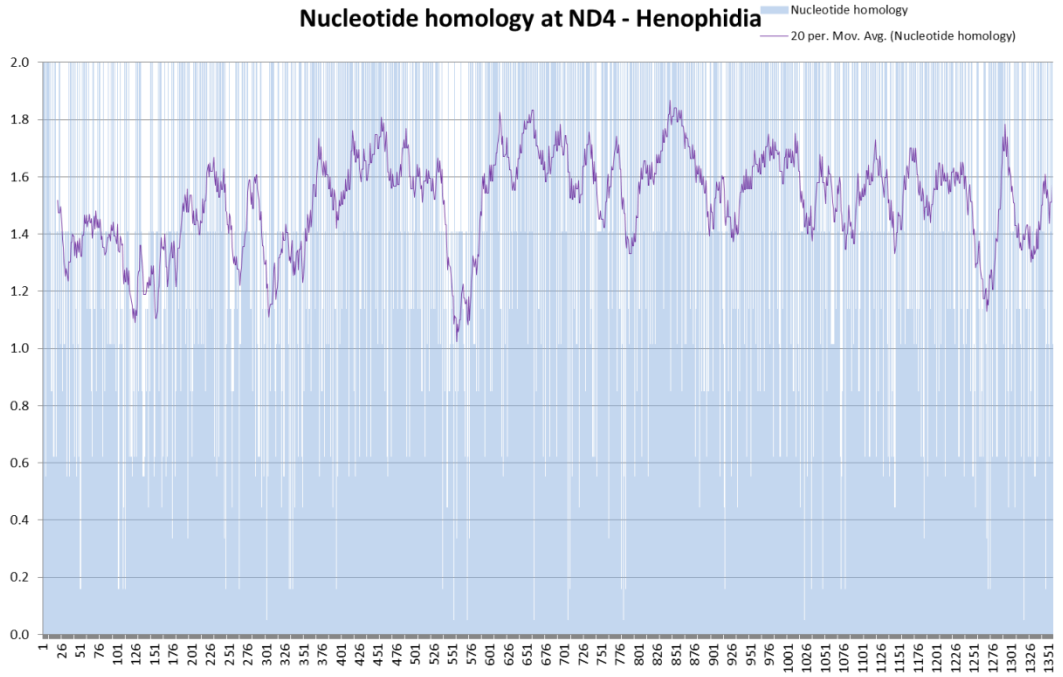


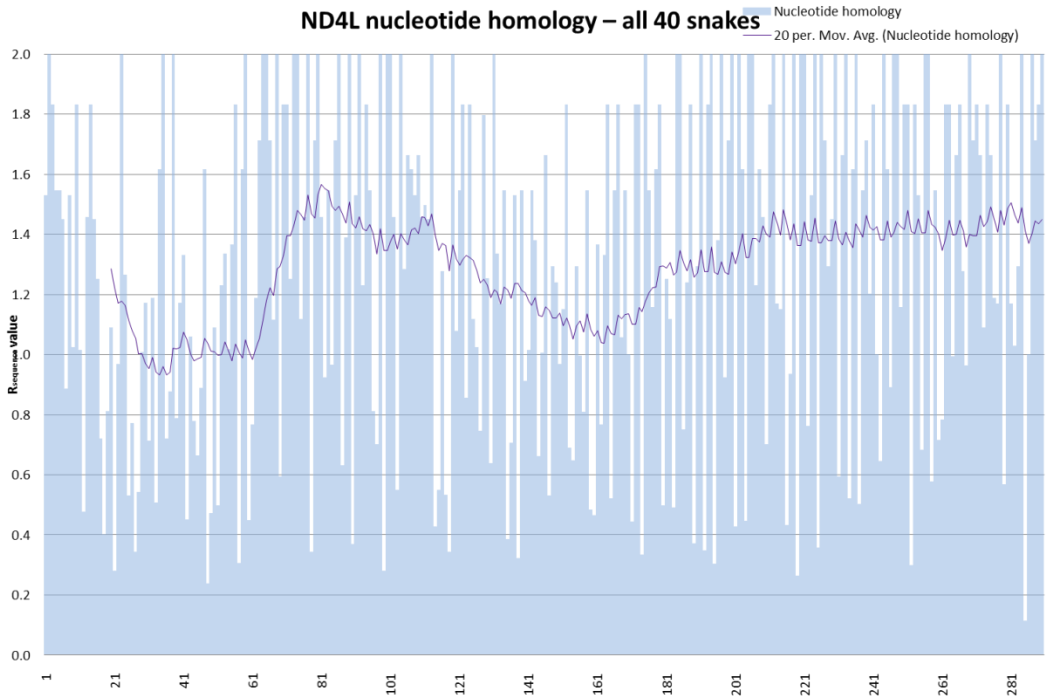


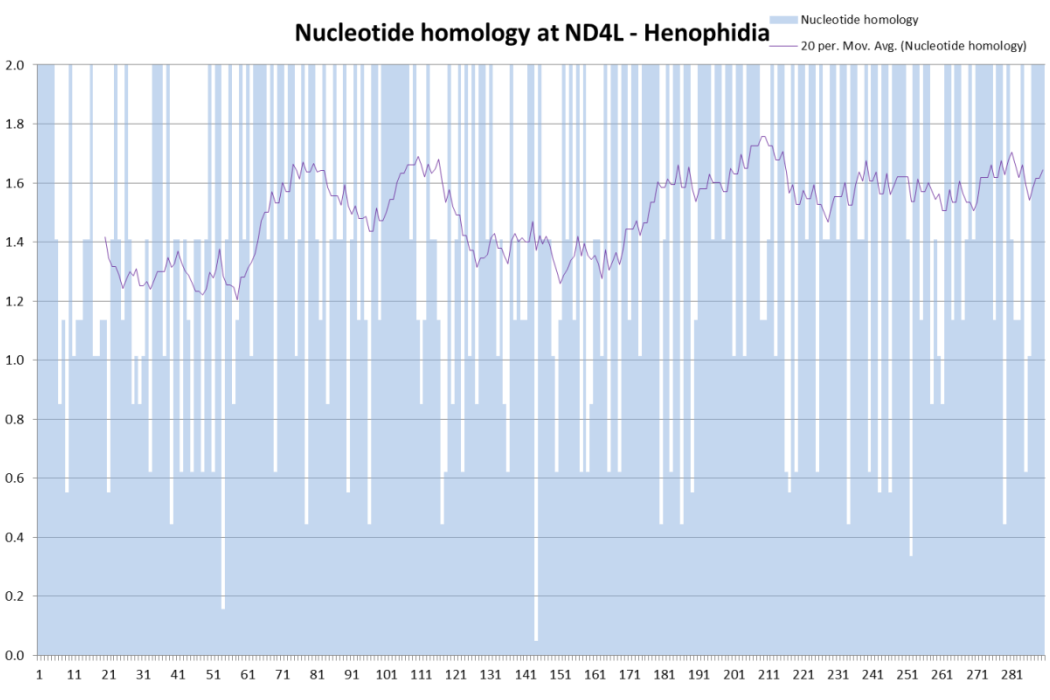

Sequence homology at ND3 – all 40 snakes
Sequence homology
20 per. Mov. Avg. (Sequence homology)

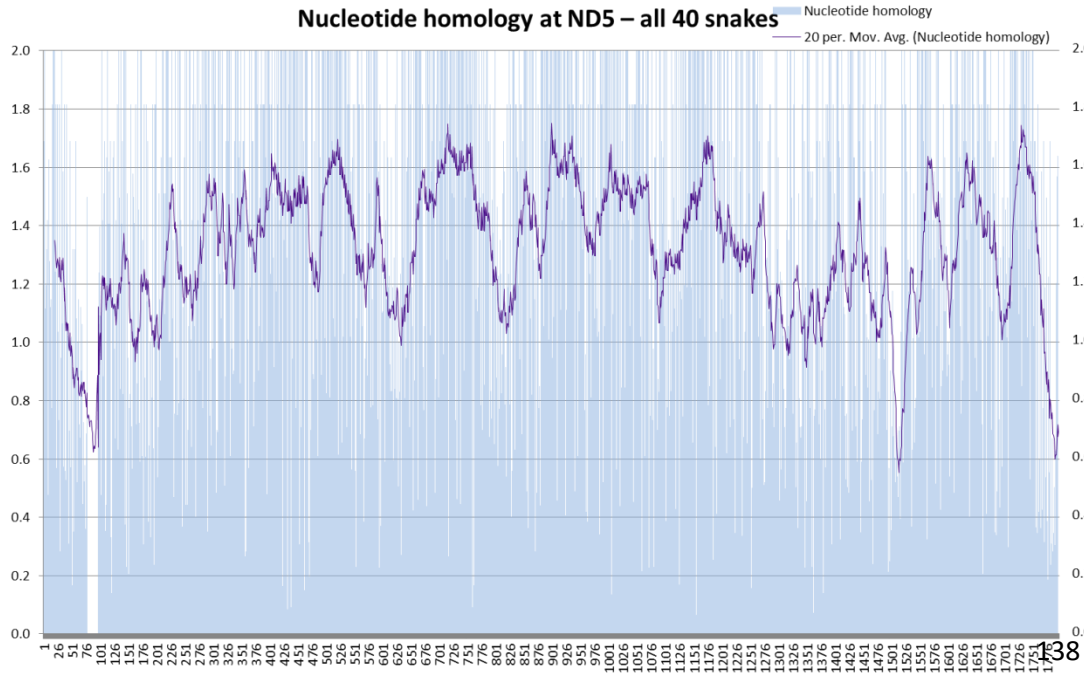

Nucleotide homology at ND3 - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)

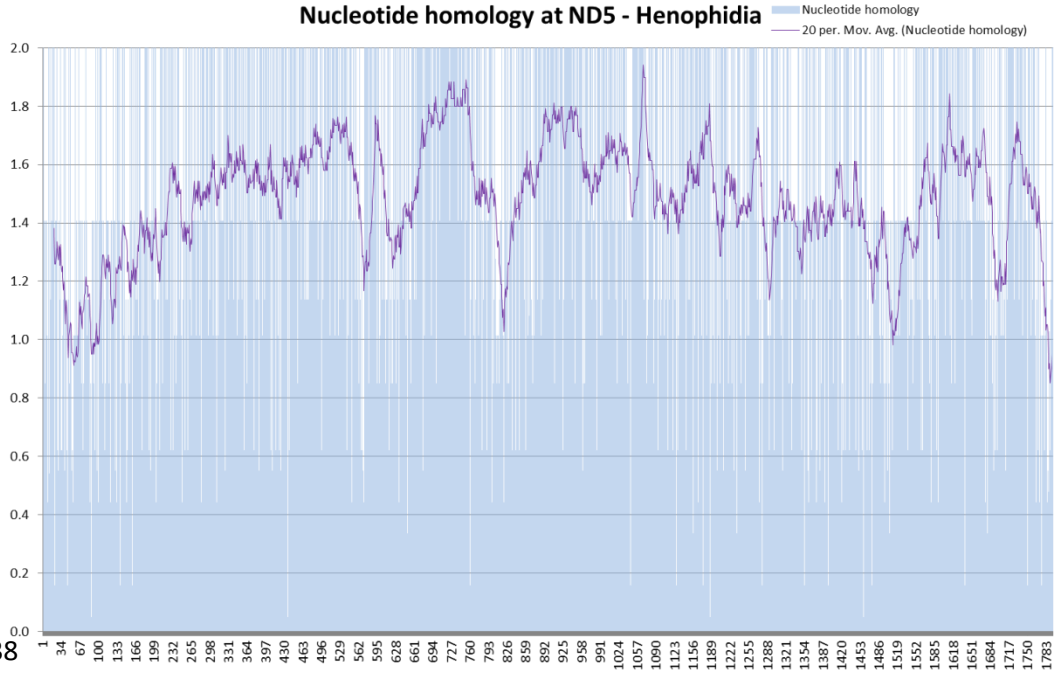

Sequence homology at ND4 – all 40 snakes

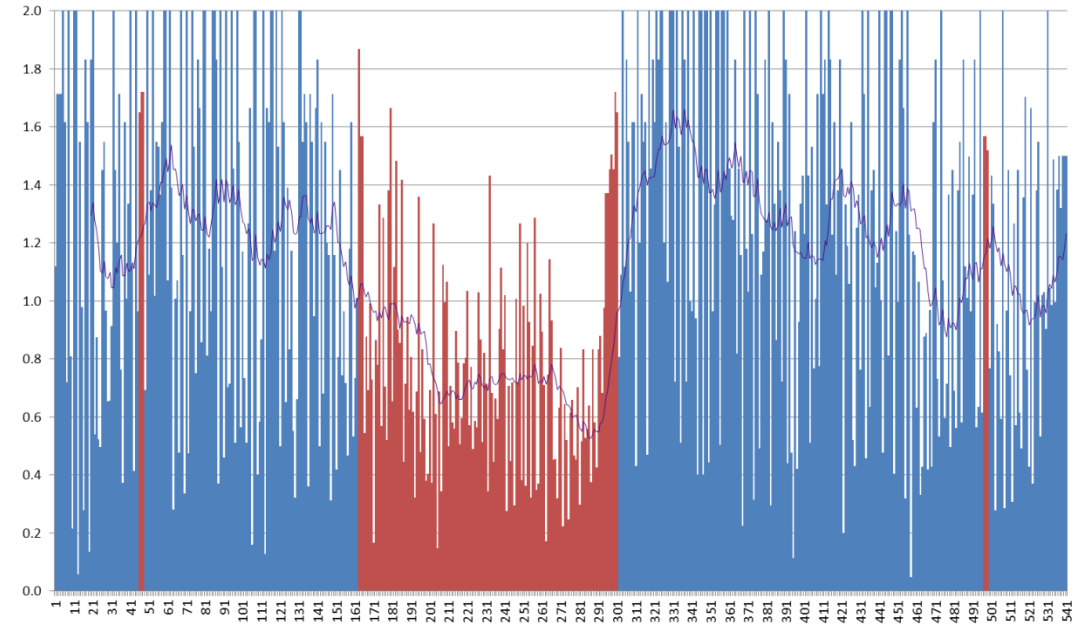

Nucleotide homology at ND4 - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)

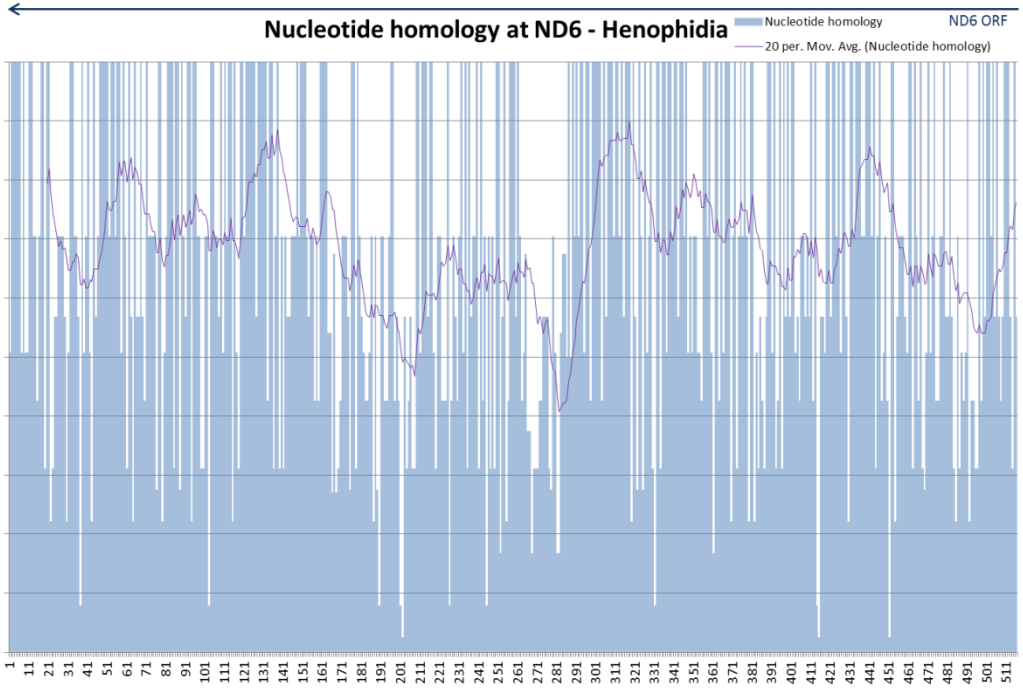

ND4L nucleotide homology – all 40 snakes
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)
R sequence value

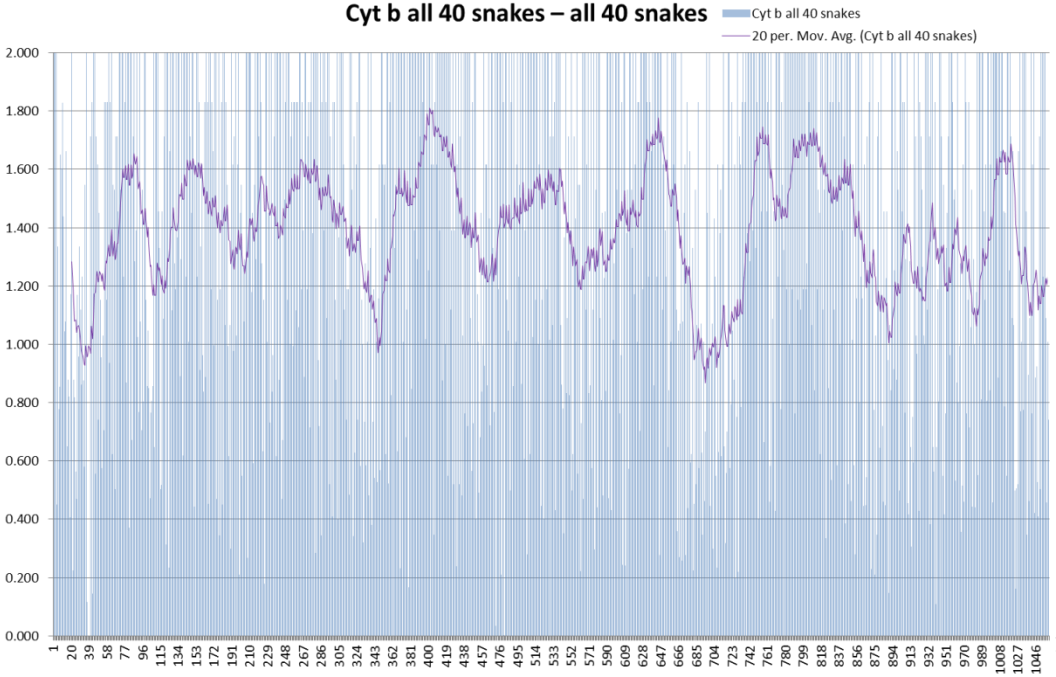

Nucleotide homology at ND4L - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)

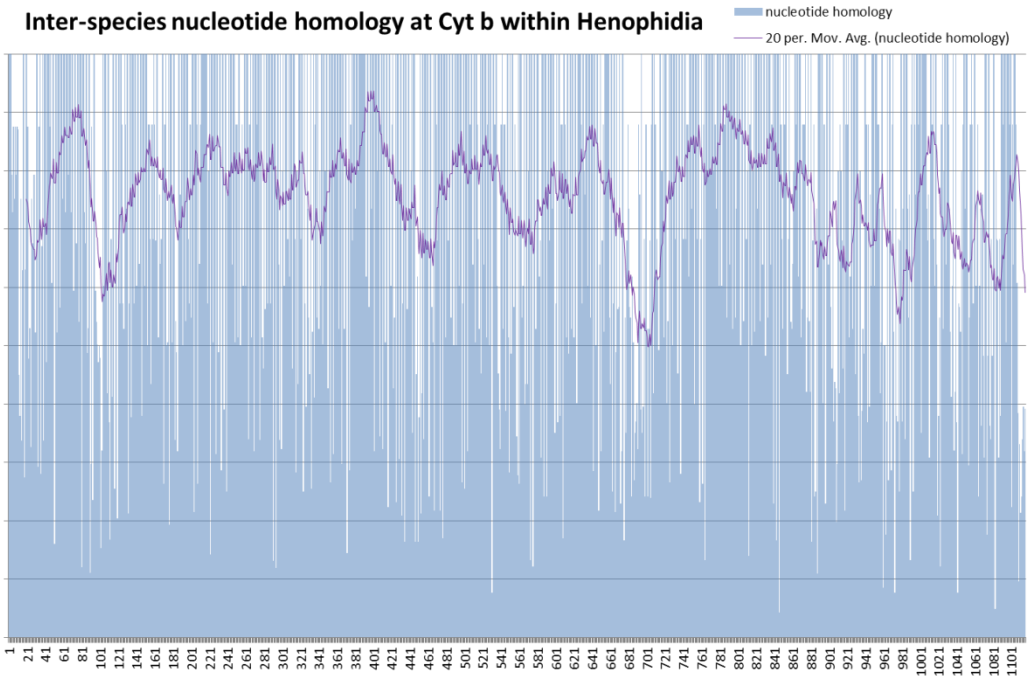

Nucleotide homology at ND5 – all 40 snakes
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)

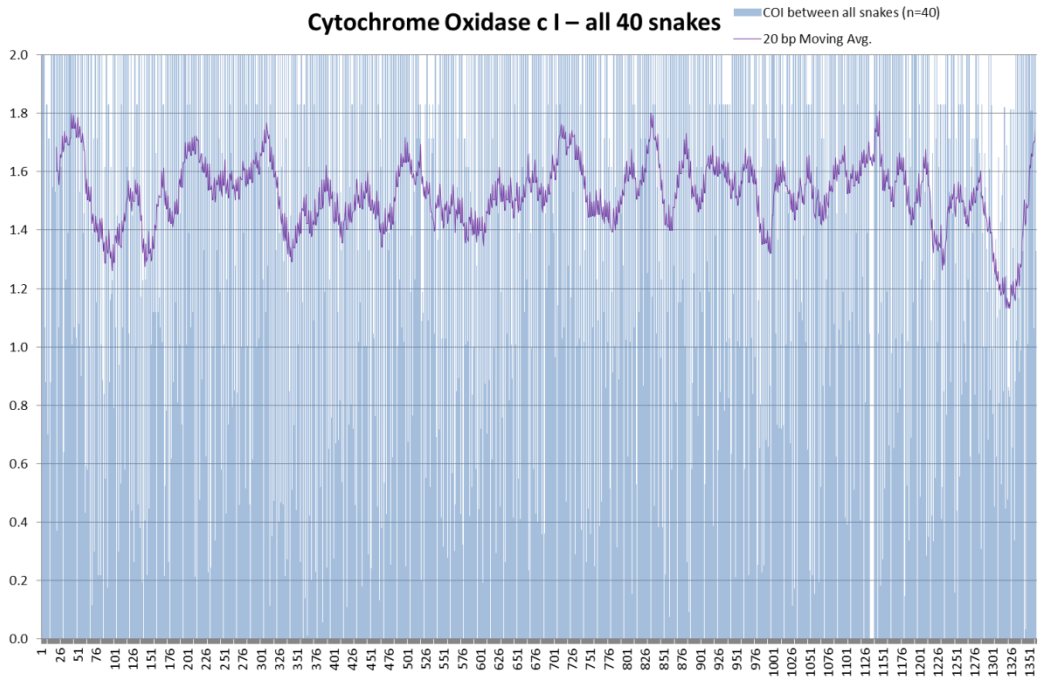

Nucleotide homology at ND5 - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)

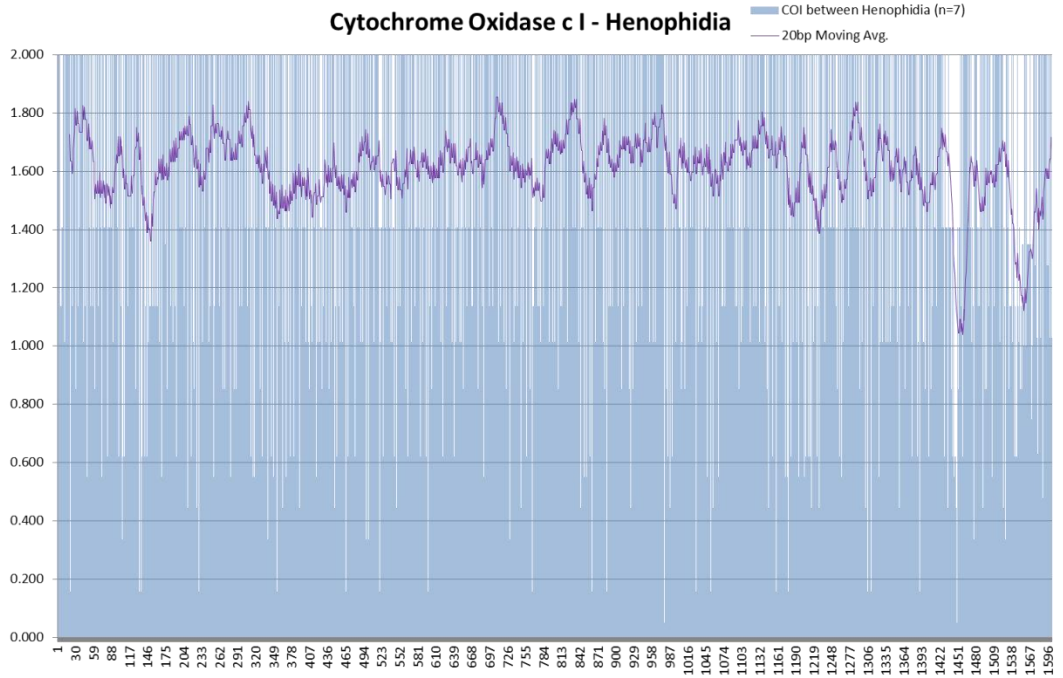

Nucleotide homology at ND6 – all 40 snakes

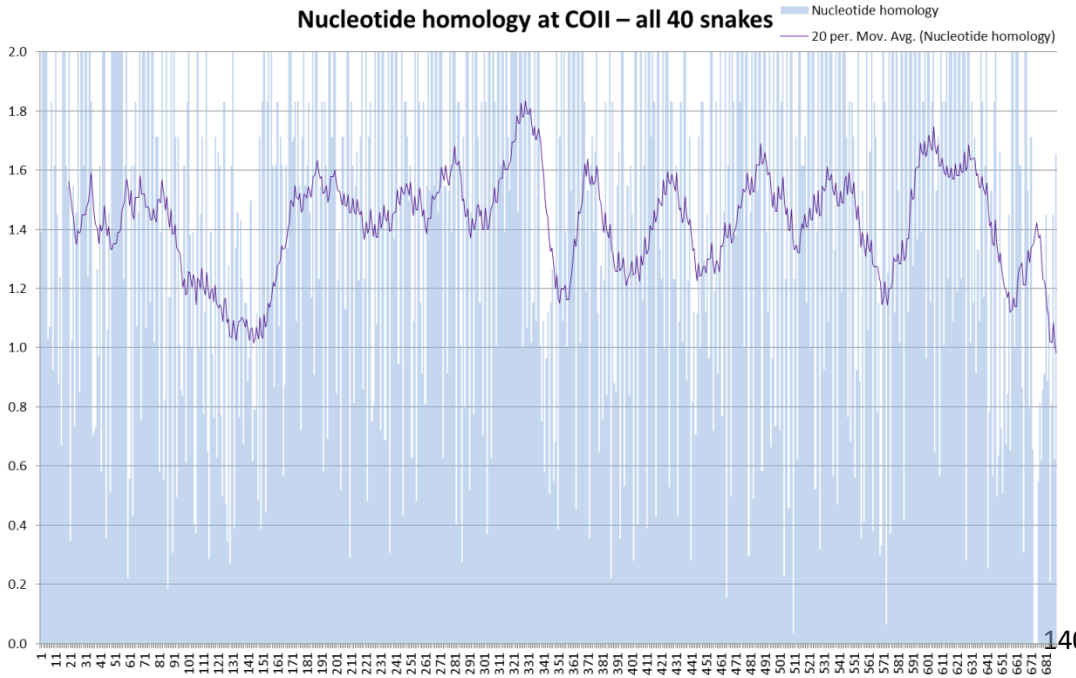

Nucleotide homology at ND6 - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)
ND6 ORF

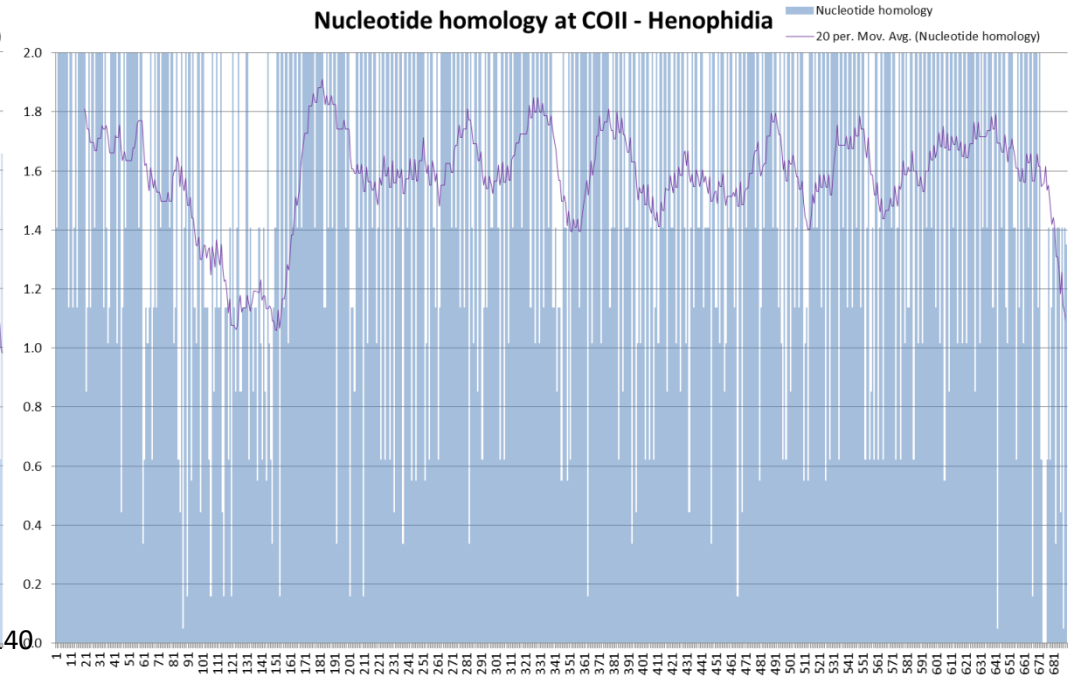

Cyt b all 40 snakes – all 40 snakes
Cyt b all 40 snakes
20 per. Mov. Avg. (Cyt b all 40 snakes)

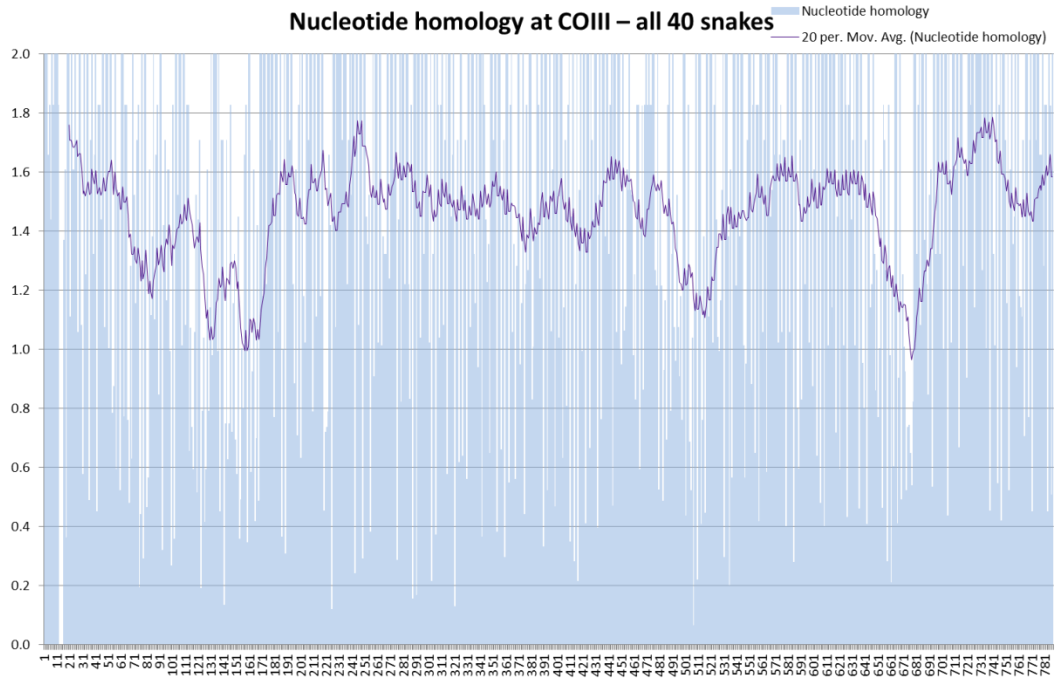

Inter-species nucleotide homology at Cyt b within Henophidia
nucleotide homology
20 per. Mov. Avg. (nucleotide homology)

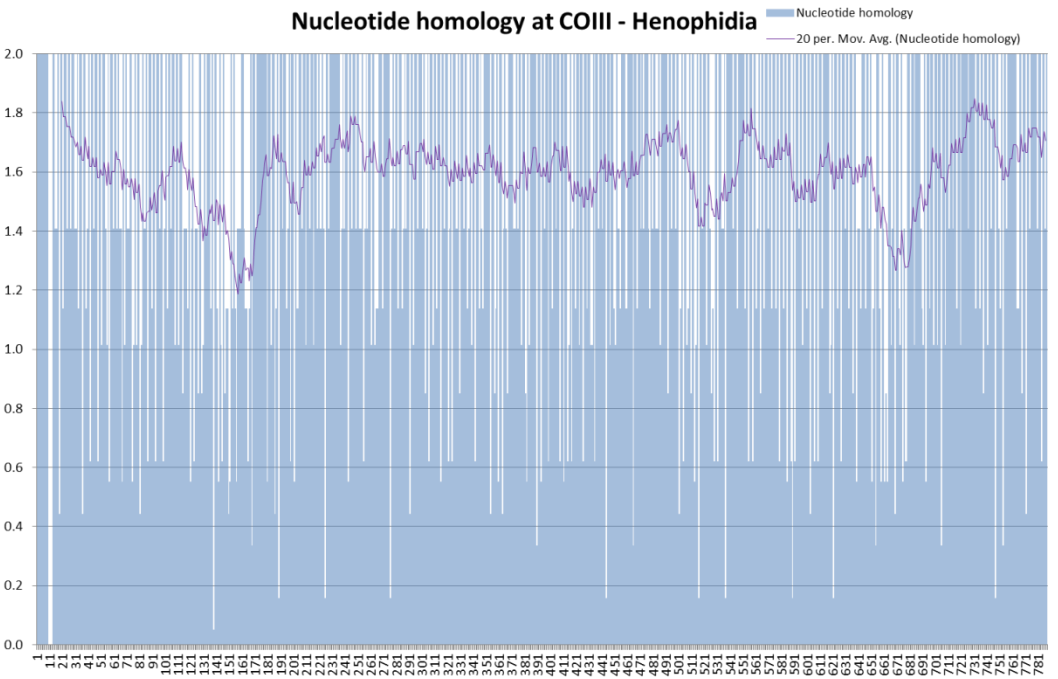

Cytochrome Oxidase c I – all 40 snakes
COI between all snakes (n=40)
20 bp Moving Avg.

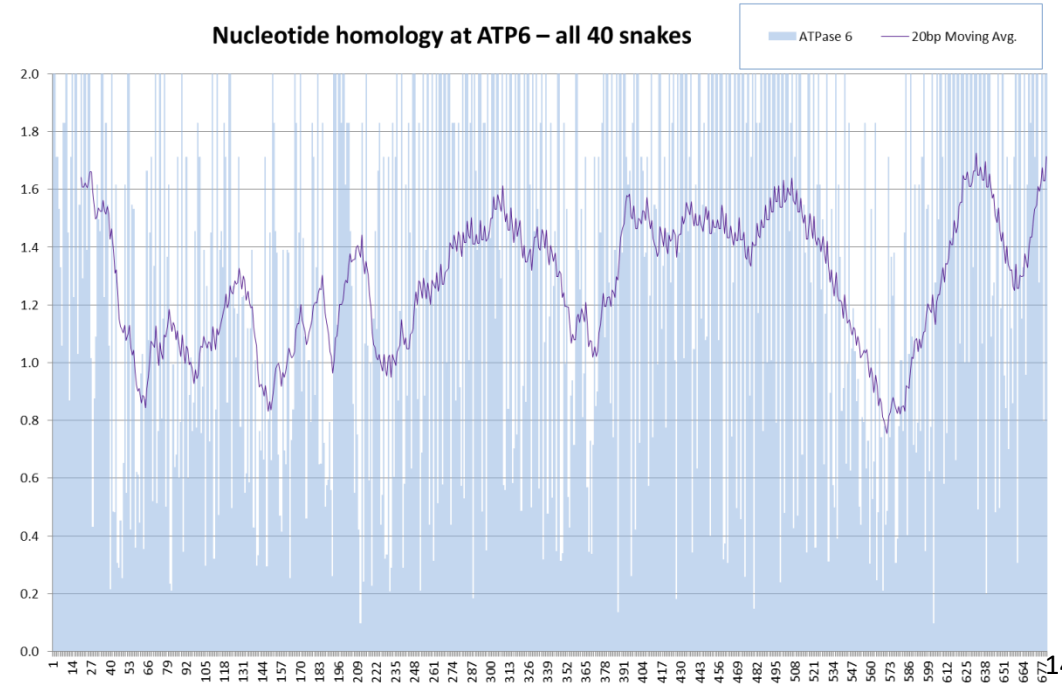

Cytochrome Oxidase c I - Henophidia
COI between Henophidia (n=7)
20bp Moving Avg.

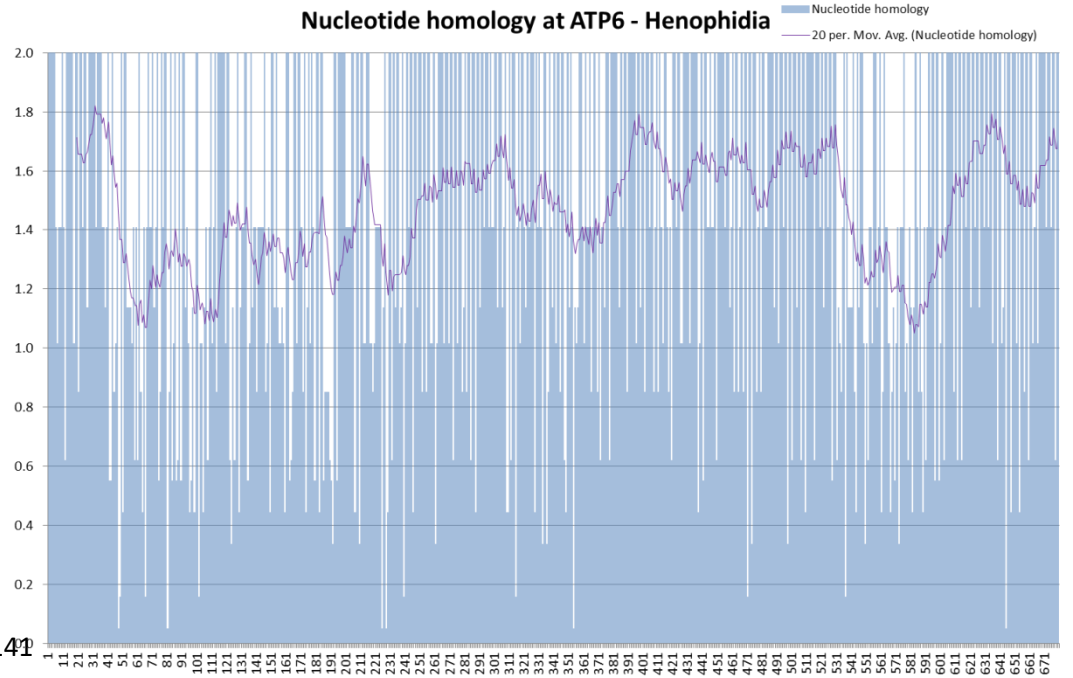

Nucleotide homology at COII – all 40 snakes
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)

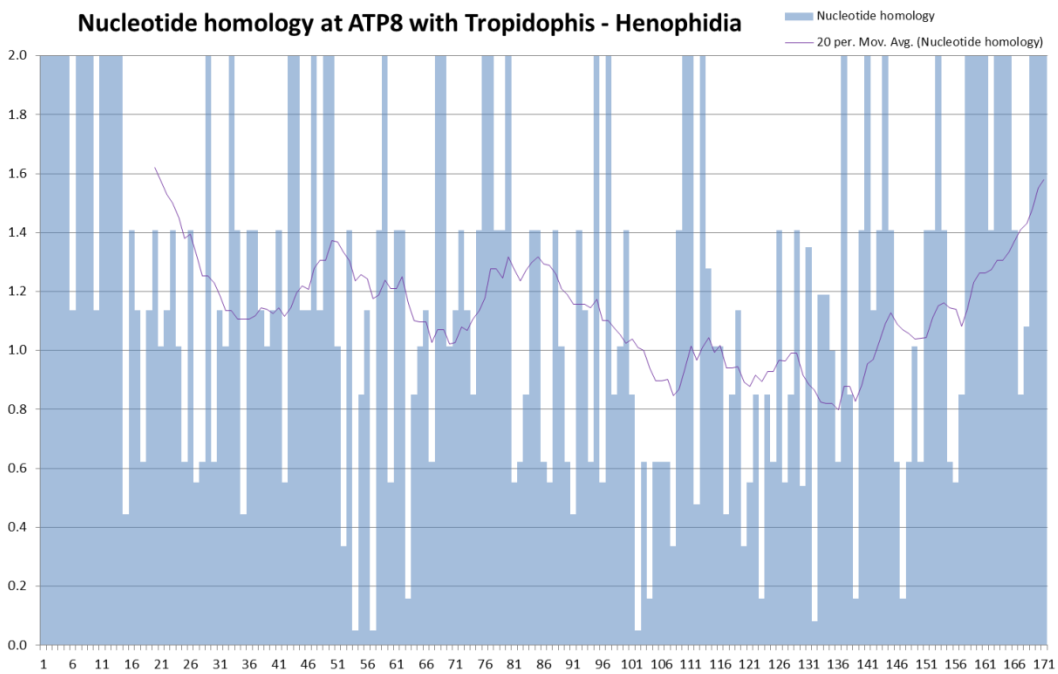

Nucleotide homology at COII - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)


Nucleotide homology at COIII – all 40 snakes
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)


Nucleotide homology at COIII - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)


Nucleotide homology at ATP6 – all 40 snakes
ATPase 6
20bp Moving Avg.


Nucleotide homology at ATP6 - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)


Nucleotide homology at ATP8 with Tropidophis - Henophidia
Nucleotide homology
20 per. Mov. Avg. (Nucleotide homology)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116 121 126 131 136 141 146 151 156 161 166 171

Appendix 2.6: Best nucleotide substitution model for each dataset

Analysis	*Best NSM*	*Subset sites*
ND6 whole	TIM	1st codon positions of ND6
(Figure 2.3a)	HKY+G	2nd codon positions of ND6
	TrN+G	3rd codon positions of ND6
Contig whole	TIM+I	1st codon positions of ND6
		3rd codon positions of Cyt *b*
(Figure 2.3b)	TrN+G	2nd codon positions of ND6 tRNA-Glu
		1st codon positions of Cyt *b*
	TrN+G	3rd codon positions of ND6
	HKY+I	2nd codon positions of Cyt *b*
Cyt b whole	HKY+G	1st codon positions of Cyt *b*
(Figure 2.3c)	HKY+I	2nd codon positions of Cyt *b*
	TIM+I	3rd codon positions of Cyt *b*
Cyt b n54	HKY+G	1st codon positions of Cyt *b*
(Appendix 2.1, Figure 3a)	HKY+I	2nd codon positions of Cyt *b*
	TIM+I	3rd codon positions of Cyt *b*
Cyt b n54 558-835	TrN	3rd codon positions of Cyt *b* within 278bp fragment
(Appendix 2.1, Figure 3b)	K80+I	1st codon positions of Cyt *b* within 278bp fragment
	HKY	2nd codon positions of Cyt *b* within 278bp fragment
Cyt b n59 5 test samples	TrN	3rd codon positions of Cyt *b* within 278bp fragment
(Appendix 2.1 Figure 4)	K80+I	1st codon positions of Cyt *b* within 278bp fragment
	HKY	2nd codon positions of Cyt *b* within 278bp fragment
Cyt b n63 4 casework	TrN	3rd codon positions of Cyt *b* within 278bp fragment
(Appendix 2.1, Figure 5)	K80+G	1st codon positions of Cyt *b* within 278bp fragment
	HKY	2nd codon positions of Cyt *b* within 278bp fragment
Cyt b *M. clastolepis*	TIM	3rd codon positions of Cyt *b* within 715bp fragment
(Figure 2.4)	TrNef+G	1st codon positions of Cyt *b* within 715bp fragment
	HKY+I	2nd codon positions of Cyt *b* within 715bp fragment

CHAPTER 3

STR Multiplex Assay Development and Validation

Appendices included in Chapter 3:

Appendix 3.1 Publication - Sherryn Ciavaglia, Julianne Henry, Adrian Linacre, Profiling pythons to combat common illegal wildlife activities, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e31–e32.

Appendix 3.2 Publication - Sherryn Ciavaglia, Stephen Donnellan, Shanan Tobe, Julianne Henry and Adrian Linacre, A novel forensic DNA profiling technique for protected species, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e258–e260

Appendix 3.3 Publication - Sherryn Ciavaglia, Hannah Dridan and Adrian Linacre, Getting more for less: can forensic tools for Australian wildlife enforcement support international compliance efforts? Australian Journal of Forensic Sciences, (2017) DOI: 10.1080/00450618.2017.1384060

Appendix 3.4 Publication - Sherryn Ciavaglia, Adrian Linacre, OzPythonPlex: An optimised and forensically validated STR multiplex assay set for the Australasian carpet python (*Morelia spilota*), Forensic Sci. Int. Genet. (2018) 34:231-248

Appendix 3.5 Samples used in Chapters 3 and 4

Appendix 3.6 First round optimisation results

Appendix 3.7 Sensitivity study Light multiplex results

Appendix 3.8 STR sequencing results

3.1 Preface

Certain wildlife crime scenarios arise that require techniques capable of individualising between members of a species. Investigative circumstances include the matching of a biological tissue such as a blood stain on a tool to a carcass and the determination of parentage where offspring from a breeder are suspected to have been taken from the wild rather than bred from the alleged parents. Although the same DNA technique well established in human forensic biology can be applied to non-human casework, the high specificity of the test that enables the separation of individuals precludes its application to all but the most closely related species. Novel markers must be identified and optimised for each species, rendering development for each new species time consuming and resource intensive. The following chapter describes the development of novel STR markers for the Australasian carpet python (*Morelia spilota*). The markers are combined into multiplex reactions to produce multi-locus genotypes from minimal DNA template addition. Validation testing is conducted to define the scope and establish the limitations of the assays. A mini reference DNA database is compiled from wild caught individuals representing the species' native range and comments made as to the applicability of this dataset for criminal investigations involving this species.

3.2 Aim and objectives of this chapter

The aim of this chapter is to design, optimise and validate a forensic DNA profiling assay specific to the carpet python following the example of human STR multiplex tests developed for forensic casework, with particular emphasis on forensic considerations such as incorporation of quality assurances measures and suitability to the small amounts of degraded remains frequently encountered in forensic cases.

My objectives are:

- To identify a large number (>15) of highly polymorphic tetra- , penta- and hexa-nucleotide STR loci specific to the carpet python.

- To combine the novel markers into a small number of multiplex assays and optimise the assays for reliable genotyping of all loci.

- To perform validation studies on the multiplex assays so as to prepare these tests for direct application to casework.

- To compile a DNA reference database from native individuals screened across the species range and subject this to statistical analyses in order to comment on direct applicability of the tests to casework.

3.3 Introduction

The extent of the global illegal wildlife trade is attracting increasing attention, evidenced in part by the recent publication of the inaugural United Nations Office on Drugs and Crime (UNODC) World Wildlife Crime report [1]. The growth of interest in wildlife forensic applications from within the forensic science community is also gaining momentum with the inclusion of a Wildlife Forensics Subcommittee in the Biology/DNA Committee of the Organisation of Scientific Area Committees (OSAC) for Forensic Science [2]. Iconic species such as elephant, rhino and tiger are the public face of wildlife crime, yet reptiles constitute the second largest taxonomic class among seizures worldwide [1], although seizures only represent detected, not total crime and reptiles are easily concealed [3,4]. Seizures indicate that reptiles dominate the live-trade international black market [5].

A prime example is the Australasian carpet python (*Morelia spilota*) which, although native to Australia, is highly sought after for legal breeding and trading both within Australia and abroad [6]. However, its unique colour and pattern-variations fetch high prices on the international black market [6], which creates high demand for breeding within and between closely related species to create new morphological varieties. Bredl's python (*Morelia bredli*), demonstrated by DNA to be part of an *M. spilota* species complex [7], is reportedly abundant in illegal international wildlife trade [8]. Other Australasian and Asian python species commonly targeted for the international commercial trade include the reticulated python (*Python reticulatus*), the Sumatran short-tailed python (*Python curtus*), the Indian python (*Python molurus*) [8] as well as Borneo short-tailed python (*P. breitensteini*) and the blood python (*P. brongersmai*) [9–11]. Overharvesting has caused declines in

population health of the ball python (*P. regius*) and pygmy pythons (*Antaresia* spp.) were among the species confiscated during a large reptile smuggling seizure [12].

Due to their compact size and low maintenance, carpet pythons have gained popularity in Australia as household pets and are commonly traded through pet shops, dedicated reptile expos, and online and often in offline reptile networks spread by word of mouth (H. Dridan, *pers. comm.*). The breeding and keeping of carpet pythons is regulated through various state and national legislation. Despite the removal of a native python from the wild without a permit being unlawful in Australia under the Commonwealth Environmental Protection and Biodiversity Conservation Act 1999, native populations provide an accessible source of fresh genetic variability to increase the genetic health of a small, continuously inbred captive collection. Uncontrolled collection of individuals from the wild has detrimental consequences on native populations; the inland subspecies of carpet python has been listed as endangered in the Australian state of Victoria since 1994 [13]. The increase of wildlife enforcement investigations involving Australian pythons across multiple Australian States has led to a direct request from enforcement authorities for the development of DNA tools to aid enforcement efforts involving carpet pythons.

Multiplex PCR assays have been optimised over many years to become routinely applied in human forensic DNA investigations. In fact, this form of analysis has been named the only forensic method to be “rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source” [14]. The development of such a single tube assay is not a trivial undertaking and the human forensic STR kits have undergone continual refinement and considerable investment from commercial companies over two decades to become the highly robust and information rich assays currently available. Fortunately, for those non-commercial entities undertaking this resource intensive task, some guidelines on identifying novel STR markers [15,16] and combining them into multiplex STR assays are available [17,18].

Due to their high cost of development, these species-specific nuclear Short Tandem Repeat (STR) DNA markers used for forensic individualisation and paternity testing have frequently been developed only for non-human species with high commercial returns, such as game animals e.g. deer [19–21] and wild boar [22] and prominent fisheries species e.g. salmon [23] and koi carp [24]. Animals often in conflict with humans have also been targeted, including those threatening human safety (e.g. bears [25], leopards [26], wolves [27]) as well as those considered a pest for their impact on human industry (e.g. badgers [28,29], foxes [30]). Profiling assays for companion animals have been used as independent evidence in crimes against humans [31–33] and increasingly tests are being developed for wildlife species targeted directly in criminal activities, such as birds of prey [34] and cockatoos [35]. The importance of extensive validation testing for direct forensic application is recognised [33,36–38]. Utility of STR marker systems has been further extended to provide intelligence about illegal poaching ‘hotspots’ through geographic assignment of poached individuals [39–41]. The application of eight previously characterised STR markers for carpet pythons [42] was found to provide inadequate genetic resolution to answer questions commonly occurring in wildlife investigations [43] and identified the need for development of a larger profiling marker set for this species.

The following chapter describes the development and partial validation of three novel 11-plex multiplex assays for forensic DNA profiling of the carpet python to assist in criminal investigations. The multiplex assays have been designed and developed following best practices evolved from the development and validation of commercial human forensic multiplex assays. The 24 independent loci comprise 22 tetra- repeat motifs and two penta-nucleotide motifs, known to cause less interpretation problems associated with stutter artefacts than the di- and tri-nucleotide repeat motifs traditionally used for non-forensic purposes [44]. Three markers have been included redundantly across the three multiplexes as quality markers to indicate sample continuity across the three independent assays. The three assays have been optimised for peak balance within and between loci across the profiles to maximise full profile amplification even from poor samples and to indicate when primer site mismatches result in reduced amplification product. The multiplex assays have been optimised for one nanogram of template DNA input in line with the amounts of DNA required by commercial human forensic profiling assays, enabling a complete 24 locus genotype from a mere 3 ng of template DNA.

Five dye technology (FAM, VIC, NED and PET fluorophore labelled primers and a LIZ labelled internal size standard) has been used to allow sufficient separation between locus allele ranges. Each multiplex assay has an accompanying allelic ladder, including a comprehensive range of alleles across the highly diverse loci, to facilitate continuity of accurate allele typing to single base pair resolution between laboratories. Associated bin sets have been created for the Genemapper ID software package (Applied Biosystems). A reference DNA database of allele frequencies is presented for 249 native individuals from across the species range of the carpet python.

All work has been conducted according to the ISFG recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations [44] and following the guidelines of the Scientific Working Group on DNA Analysis Methods [45]. Comprehensive validation of a forensic assay is well recognised as an important precursor to applications in casework (e.g. [46]). Validation testing involves a demonstration of the capabilities and limitations of technique under specific conditions to provide assurance in the reported results. The requirements of validation testing were described in section 1.1.6. Initial validation testing demonstrates the robustness of these multiplex assays to facilitate the application of the multiplex assay set to criminal investigations involving carpet pythons. A full validation is not presented within this chapter, however outstanding tests required in order to complete a comprehensive validation are discussed.

3.4 Materials and Methodology

3.4.1 Samples

Appendix 3.5 provides the details of the 304 samples included in this study. Google Earth was used to map the geographic locations of all samples, for which geographic coordinates or a specified locality were provided. Figure 3.1 shows geographic origin of all samples for which this could be determined. Bredl's python (*Morelia bredli*) were included in this study, as a previous phylogenetic analysis [7] suggests that this species is genetically less divergent than *M. spilota imbricata*. All reference samples were obtained from the Australian Biological Tissue Collection (ABTC) of the South Australian museum. Voucher specimens were included where available following best practice

[44]. DNA previously extracted from a separate project [43] was also obtained from the ABTC. Sample type is specified in Appendix 3.5 where known. Tissues were stored at -80°C or in ethanol at room temperature. Samples originated from liver, whole blood, skin, scale clippings and muscle.

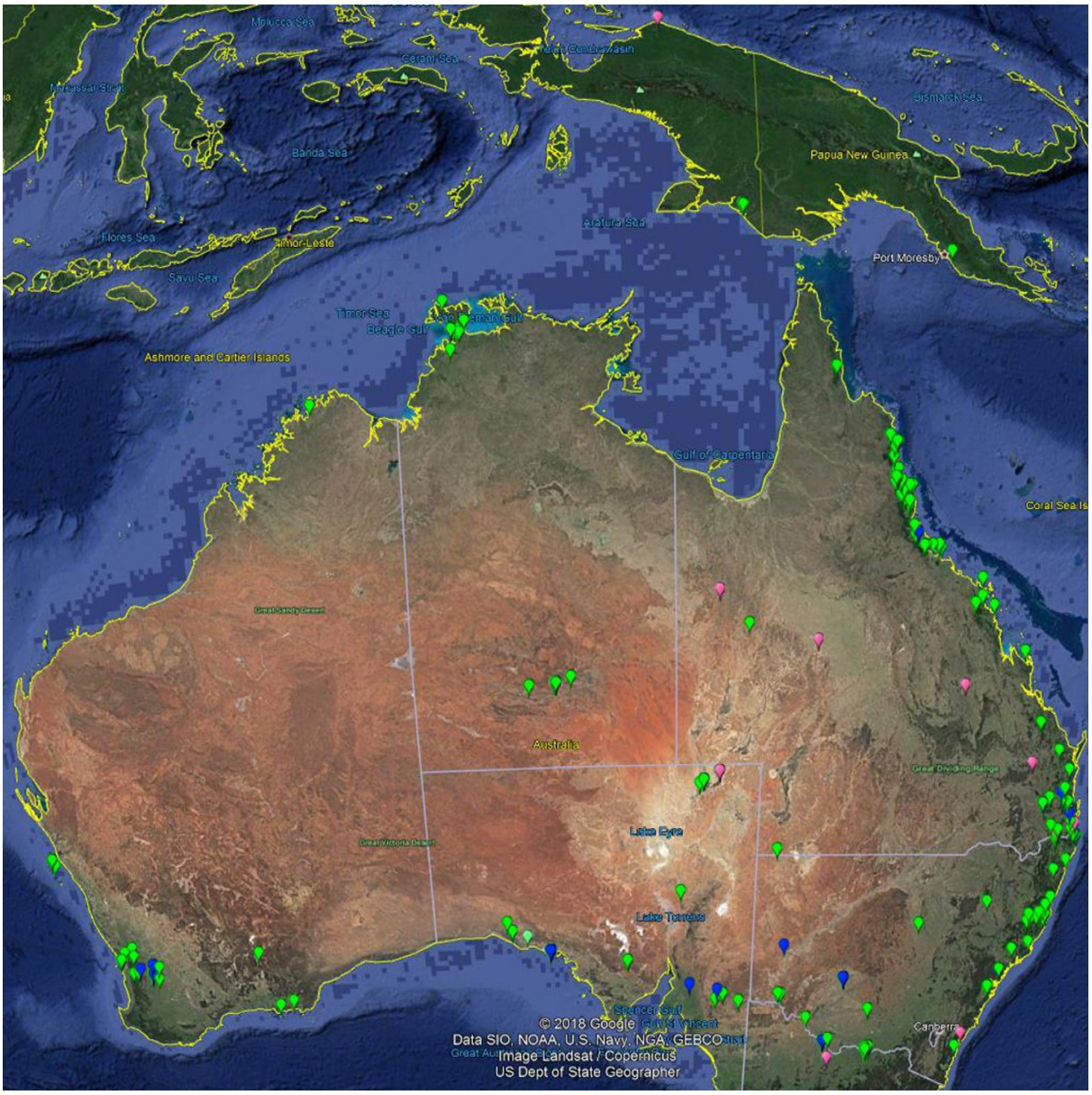


Figure 3.1: Geographic origins of samples used in Chapters 3 and 4. Where individuals are sampled from locations in close proximity, multiple individuals are not visibly distinguishable. This map illustrates the geographic coverage, but not the density of sampling. Pink samples were screened in this chapter but did not yield a useable profile. Green samples were used in the final reference database for chapter 4. Blue samples were unknowns used in chapter 4.

3.4.2 DNA extraction

The DNA extraction followed section 2.4.5 with a final elution volume of 100 µL. Each neat DNA extract (5 µL) was run on a 2% agarose gel stained with ethidium bromide to visualise quality and

quantity of the extracted DNA. DNA extracts were also quantified using a Qubit fluorometer (Life Technologies, NSW, Australia) and diluted to 1 ng/µL to ensure optimal and consistent sample DNA concentration for multiplexing. DNA samples used for multiplex optimisation steps were re-quantified following dilution using the Qubit High Sensitivity Assay Kit to ensure precise DNA template addition.

3.4.3 Characterisation of novel STR loci

3.4.3.1 Massively Parallel DNA Sequencing of the Carpet Python Genome

Characterisation of new STR markers for application to multiplexing was conducted following workflows suggested for identifying new STR markers using Mass Parallel Sequencing (MPS) [15,16].

A DNA extract (ABTC 30087) was purified and sent to the Australian Genome Research Facility (AGRF, Queensland, Australia) for MPS using 1/16 plate on a 454 Next Generation Sequencer (Roche). Mass Parallel Sequence (MPS) data were analysed using QDD [47].

3.4.3.2 Assortment of DNA Sequences

Sequences were sorted by sequence tag. Duplicate sequences in the dataset were removed or contigs created where sequence identity was >95%. Parameters were set to return sequences with a minimum of 80 base pairs (bp) containing simple and compound STR motifs of 2-6 bp repeats and ≥ 5 repeat units were selected for. The MPS run had yielded 167 putative STR loci, with the longest product length of 543 bp. Primer3 [48] was used to design primers to amplify putative STRs with a PCR product size range of 350-550 bp. Primer conditions were set to a minimum melting temperature (T_m) of 57 °C, a maximum T_m of 63 °C and an optimal T_m 60 °C to enable subsequent multiplexing (optimal amplification of all involved primer pairs within a tight temperature range). Three options each were returned for forward and reverse primer design. Primers were BLAST searched (www.ncbi.nlm.nih.gov/BLAST) to check that they would not target non-specific taxa.

Loci with successfully identified primer pairs were manually filtered further using Excel (Microsoft). The data were filtered and ranked using a hierarchical system proceeding with motif type (penta- and tetra-nucleotide) followed by PCR product size and repeat number. Putative loci were charted by size and coloured by repeat number (see figure 3.3 in results) to assist final marker choice. Simple sequences returned were of reasonably short length (<300 bp), so compound and complex motifs as well as sequences constructed from contigs were included. This increased the choice of high molecular weight makers available to utilise the entire 600 bp size migration range easily obtainable using a commercial capillary electrophoresis instrument with appropriate size standard (e.g. Liz600, Life Technologies).

3.4.3.3 Identifying putative STR loci using Multiplex Ready Technology

Multiplex Ready Technology [49] involving a nested PCR procedure was used to incorporate fluorescent dyes into PCR products for visualisation by capillary electrophoresis. Similarly to M13 incorporation procedures [50], this enabled testing of the polymorphic nature of 36 putative loci without expensive fluorescent labelling of every primer pair. Locus specific primer pairs were selected for the 36 loci exhibiting the most favourable characteristics. An MRT tag sequence was added to each primer sequence and primers were synthesised by Integrated DNA Technologies

(Iowa, USA). The amplification method is described in Ciavaglia, Dridan & Linacre [51]. The PCR product fragment analysis was conducted using a 3130xl (Applied Biosystems) with 36 cm capillary array and POP-4 polymer. Initial spectral calibration was conducted using the G5 dye set matrix standard (Applied Biosystems). One µL PCR product was added to 9.7 µL Hi-Di Formamide and 0.3 µL Liz500 internal size standard (Applied Biosystems).

Each locus was amplified in singleplex using six python DNA extracts from diverse localities across the species range to survey for initial locus polymorphism and heterozygosity. A subsequent singleplex amplification was performed using combined DNA templates from eight other geographically diverse samples to further indicate allelic diversity and peak morphology at each locus. Monomorphic loci and loci that showed no PCR amplification product or more than two PCR amplification products were discarded. Presence of stutter peaks aided the identification of true STR amplification products.

3.4.4 Multiplex Assay Design

Multiplex assay design was conducted following a protocol for constructing STR multiplex assays [18]. Initial screening for polymorphism yielded 28 promising loci, dictating a final set of two to three assays depending on locus size ranges. Loci were ranked according to allelic diversity (number of alleles seen in the initial 14 individuals), allele size range (facilitating multiplexing) and quality of peak morphology. The number of alleles seen in all markers was divided by the allele size range to provide an index of the most desirable markers (highly polymorphic with a small allelic window). To facilitate comparability between this study and previous studies using the loci developed by Jordan et al. 2005, available STR data was assessed from Duncan Taylor's doctoral thesis [43] and data provided by Nicole White (unpublished). The two respective investigations included eight and nine of the Jordan STR loci, four of which overlapped between these studies. These markers were similarly assessed for inclusion based on size and allelic diversity.

Newly developed multiplex loci (prefixed MsF for '*Morelia spilota* Forensic') and previously developed primers (prefixed MS for '*Morelia spilota*' by [42] were arranged in rank priority on a 2D plane according to obtained PCR product sizes to visualise multiplex layout (figure 3.2). The use of five dye technology (G5 dye set, Life Technologies) would allow inclusion of four dye channels plus an internal size standard (Liz fluorophore). The requirement for three separate assays was inferred from initial locus size ranges, to ensure that the size ranges of loci sharing a dye label within a multiplex assay did not overlap as sample genotyping was expanded. Loci were colour coded according to dye label and shaded light, medium or dark according to which of the final three assays they corresponded.

Three different multiplex assays were constructed, designated Light, Medium and Dark, incorporating 12, 11 and 12 loci respectively. Three loci were included in all of the multiplex assays as quality assurance markers for redundancy between assays. This three locus redundancy allowed reproducibility of allele calls to be assessed between multiplex assays and would flag potential sample mix up between assays. Primers were examined by eye and run through http://www.thermoscientificbio.com/webtools/multipleprimer/ to test for self-complementarity and complementarity between primers within the multiplex. A single guanine base was added to the 5' end of every reverse primer to facilitate complete adenylation of PCR products and accurate single base precision genotyping [52]. Locus specific primer pairs were ordered for the final three assay

primer set (fluorescently labelled primers from Applied Biosystems and unlabelled primers from Integrated DNA Technologies (IA, USA) and Geneworks (Adelaide, Australia)).

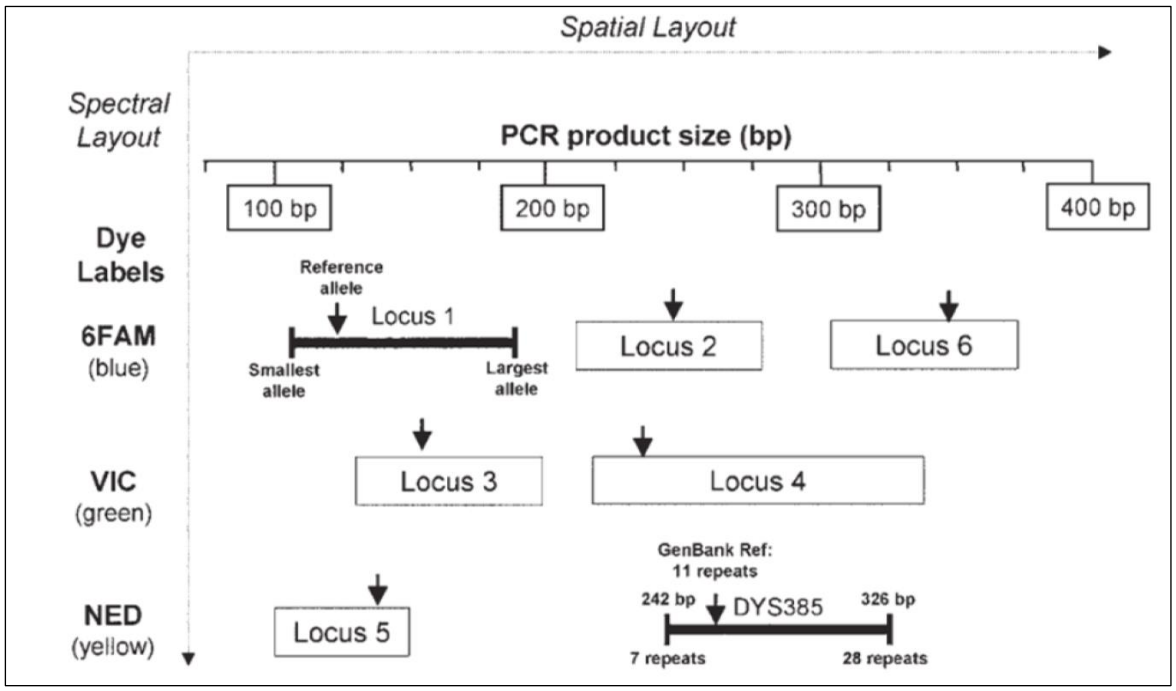


Figure 3.2: Example 2D layout for planning the locus layout and dye designations of a multiplex assay, reproduced from [18].

3.4.5 Multiplex Assay Optimisation

Initial multiplex amplifications were trialled using Amplitaq Gold II (Applied Biosystems) and Qiagen Multiplex Kit (Qiagen) at 27, 30 and 35 cycle numbers, with reaction volumes of 20 µL and 25 µL and annealing temperatures between 57 °C and 60 °C. Firstly, amplification of fluorescently labelled primer pairs was conducted using four quadruplex assays, each containing one locus for each fluorophore (16 primer pairs in total) at 2.5 µM concentration per primer for 35 cycles with annealing temperatures of 57 °C and 60 °C (optimal temperature should be around 58 °C). Primer concentration was adjusted to 250 nM and amplifications repeated. Primer pairs showing poor amplification were then diluted to a series of 250 nM, 100 nM, 50 nM and 25 nM and the best product identified. Following this, all 10 primer pairs chosen for the Light multiplex were amplified together and seven chosen for the Dark multiplex were also amplified in a single tube using Amplitaq Gold II™ polymerase (Life Technologies). The amplification results were of mixed success across loci.

Amplifications of the Light 10-plex and Dark 7-plex reactions were repeated using the QIAGEN Multiplex Mastermix containing primer concentrations of either 250 nM or 40 nM depending on the strength of previous amplifications. Cycling conditions followed the manufacturer's recommendations with an annealing temperature of 60 °C, 35 cycles of amplification and a final extension at 60 °C for 30 minutes. All remaining primer pairs that had not yet been amplified in multiplex were amplified at 250 nM in their respective allocated multiplexes to gauge strength of the products.

Complete primer mixes (11 or 12 loci) for the three respective multiplexes were made using each primer at 250 nM concentration, then each mix diluted in series to 125 nM, 62.5 nM and 31.25 nM and used to amplify 1 or 2 ng DNA template. These serial dilutions enabled relatively quick identification of the optimal primer concentration for each locus based on amplification efficacy (weak, optimal or overloaded product) at the different concentrations. These amplifications were also trialled at 27 PCR cycles.

Cycling conditions were altered to 29 cycles for subsequent optimisation reactions to mirror the commercially available Globalfiler (Life Technologies) STR profiling kit used in human forensic investigations. Subsequent multiplex assay optimisation proceeded in a 25µL reaction volume using 12.5µL Multiplex PCR Mastermix (Qiagen), 11.5 µL of primer mix (primers present in varying concentrations) and 1 ng template DNA. PCR conditions were initial activation at 95°C for 15 minutes, followed by 29 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 90 s, and extension at 72°C for 60 s, concluded with a final extension step of 60°C for 30 minutes to promote complete product adenylation [53].

Four different DNA templates diluted to 1 ng/µL were used and single primer pair concentrations were altered based on the amplification products obtained at previous concentrations and cycle numbers to strengthen or weaken products in order to achieve balance across the profile. Four optimisation rounds were performed tweaking primer concentrations towards balanced profiles. Different DNA templates were substituted in to expand initial screening results, but sample 62456 was always retained as a constant to compare inter-locus balance between optimisation rounds.

At the end of four optimisation rounds (when profiles were somewhat balanced), a screening amplification round was performed on a much larger number of individuals to gauge whether rearrangement of loci within and between assays was necessary due to emerging locus overlaps.

3.4.6 Multiplex assay redesign & optimisation

The multiplex assays were rearranged into three 11-plex assays exhibiting no overlapping allelic windows between loci. Some primers were redesigned to improve fit within new multiplex layout, to adjust T_m to improve multiplex amplification and to 'PIG-tail' one primer of the pair with either a single Guanine base or 'GTTTCTT' at the 5' end to promote complete adenylation [52]. The loci chosen to be the three quality markers were updated for improved peak morphology. Two of these quality markers were placed in the low molecular weight region of the multiplex to ensure that these quality indicators are still amplified in weak and degraded samples.

The reconfigured multiplex assays underwent three further rounds of optimisation. Optimised primer concentrations for the three finalised multiplexes are given in table 1 of [54] (embedded in appendix 3.1). Inter-locus and intra-locus peak balances of profiles resulting from the optimised assays are sufficient to indicate the presence of null alleles due to primer binding site mutations and also the presence of degraded or inhibited samples where a ski slope effect is exhibited by strong low molecular weight product tailing through to weak or non-existent high molecular weight product. An allelic ladder was developed corresponding to each multiplex assay to ensure accurate genotyping between different electrophoretic runs (detailed in the following section).

During profiling, inconsistent absolute base pair sizing (precluding allele calling to single base pair accuracy) was soon noted, affecting peaks between 200 – 300 bp in size (see discussion section 3.6.2.2). The Liz600 was substituted and the issue resolved. At least one allelic ladder was included per 31 samples; three allelic ladders were included on a 96 well plate to account for any potential migration shift across the course of the six injections. Data was analysed using GeneMapper ID v3.0 and applying panel and bin sets developed for each multiplex assay as outlined below. Samples displaying spectral saturation, resulting in pull up peaks in neighbouring dye sets were rerun at a dilution of either 1/50 or 1/100 based on degree of pull up. This was rarely necessary due to prior quantification of neat DNA extract and only occurred when the undiluted extract was very strong.

3.4.7 Construction of allelic ladders

The initial screening dataset (n43) was examined to identify all unique alleles at each locus. Heterozygous individuals were chosen that represented the majority of alleles present without duplication of an allele between two individuals. Singleplex PCR reactions were performed in 25µL reactions using 2x My Taq Hotstart Mix, 0.4 µM each primer and 0.5-3 µL additions of DNA template for each of the heterozygotes chosen for the specific locus (depending on initial amplification strength), resulting in a mixed DNA genotype per locus. A second round of singleplex reactions were performed including only DNA template from weak individuals and first and second round PCR products pooled to achieve greater balance in each single locus DNA mixture. Singleplex mixtures for each multiplex assay were pooled and visualised by fragment analysis. More singleplex product was added to strengthen weak loci in the resulting ladder profile. Subsequent alleles were added to ladders through the addition of mixed DNA singleplex template to the pooled ladder sample to improve the allelic representation of ladders as more alleles became known. Three allelic ladders were created, one for each respective multiplex containing all loci within that multiplex.

Panel and bin sets were created in Genemapper ID v3.0 (Applied Biosystems) for all three allelic ladders, based on the alleles used in each ladder. Stutter peaks were also used to designate allelic bins. Any micro-variant alleles (not complete motif repeat units) included in the allelic ladder were labelled with red bars to assist allele calling. Not every variant could be included in the ladder, either because the variant had not been seen in the dataset or due to limited resources. Virtual allele bins were created in Genemapper to assist calling of alleles that fell in gaps between ladder alleles. These were placed where a complete motif variant would theoretically lie (e.g. four base pairs away from a simple tetramer allele). These were applied with caution and only in close vicinity (up to 8 base pairs) from viewed alleles. Due to the nature of size migration, base pair size designations do not absolutely reflect underlying base pair number and drift is known to accumulate with distance from an allele. For this reason the allelic ladder bins need to be based on the sizes seen rather than those expected. Bin designations were based on the average size between ladders from the same plate run at both FSSA and Flinders. The Local Southern [55]) method of peak sizing against the internal size standard was chosen for sizing products.

An initial 43 individuals were screened with the original multiplex conformations prior to redesign. Following the above method, ladders were also constructed to accurately genotype these individuals incorporating the old multiplex layout. Panel and bin sets were not used for these ladders. Individuals were typed by hand with comparison to the peaks of known allele size included in the allelic ladder.

3.4.8 Ladder immortalisation

Light allelic ladder was diluted 1/1000 and 1/1000000 using PCR-grade H_2O. PCR amplification was performed in a 12 µL reaction including 6 µL Qiagen Multiplex mastermix, 5 µL Light primer mix and 1 µL diluted ladder template. Amplifications were performed using the cycling conditions described in 3.4.5 with cycle number adjusted to 15, 22 and 29 total amplification cycles. PCR products were visualised by 3130xl fragment analysis.

3.4.9 Validation

3.4.9.1 Repeatability

During the optimisation of the multiplex assays, sample ABTC 62456 was included on every run for comparative purposes. This included four optimisation runs on the 3130xl instrument at Flinders University, except for the Dark optimisation assay which was run twice. Light assay Optimisation 1 run was excluded from the repeatability study due to anomalous results (see results section 3.5.6.1). One PCR plate incorporating all three multiplex assays was amplified in the Flinders laboratory and then run twice on the 3130xl instrument at Forensic Science SA (FSSA). The same plate was subsequently rerun on the Flinders 3130xl instrument. The data on repeatability were collated from six runs for the Light assay, seven runs for the Medium assay and eight runs for Dark assay. Locus MS4h was added to the Medium multiplex assay during reconfiguration and was not run in enough replicates to be included in the repeatability study.

3.4.9.2 Reproducibility

As detailed above, one PCR plate was run on independent capillary electrophoresis instruments at FSSA and Flinders respectively with two weeks separation between runs. This does not constitute a true reproducibility experiment (see section 3.6.3.4 for further discussion).

3.4.9.3 Heterozygote balance

Profile data were collated from samples genotyped using the three optimised multiplex assays and validation tests. Only loci exhibiting heterozygote peaks >1000 RFU were included, as average peak height has a recognised inverse correlation with variability in heterozygote balance (H_b) [56]. This also avoided samples for which poor DNA quality led to a 'ski slope' effect of imbalance in alleles of large size differences within a locus. The heterozygote balance of each allele pair was calculated by:

$H_b = \phi HMW/\phi LMW$

where ϕHMW is the height in RFU of the high molecular weight product and ϕLMW is the weight of the low molecular weight product. The mean, variance and standard deviation of H_b were calculated per locus.

3.4.9.4 Species specificity

Cross-species amplification success was determined for all Australian python species and a number of less closely related Australasian python species of forensic significance, as well as human DNA template. PCR amplifications were performed using singleplex MRT reactions as per section 3.4.3.3.

Details of the species included are listed in table 1 of the embedded publication [54], appendix 3.3 and geographic ranges of the included species are shown in figure 1 of the same publication.

3.4.9.5 Assay sensitivity

The three multiplex assays were amplified using DNA template of one individual (ABTC 62456) diluted to decreasing concentrations to ascertain the lower limit for DNA profiling. Amplifications were performed as per section 3.4.5 with addition of the following concentrations of DNA template: 5 ng, 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.25 pg, 15.63 pg, 7.8 pg, 4 pg and 2 pg.

3.4.9.6 Selectivity

Appendix 3.5 provides details of the sample types used in this study, the range of which includes blood, liver, muscle, scale clipping, sloughed skin, tail and tissue. The effectiveness of each of these tissue types to provide a full DNA profile was reviewed.

3.4.9.7 Sanger sequencing

Sanger sequencing was performed to confirm the underlying repeat motif at each locus. Due to the high number of alleles obtained during this study, only a subset of alleles could be sequenced, however representative alleles from all loci were sequenced. Homozygote individuals were amplified in singleplex reactions using unlabelled primers. Amplification reactions contained 0.5-1 µL extracted DNA template (~1-5 ng), 200 nM dNTPs, 1.5 mM $MgCl_2$, 1.25 µM of each primer and 1 unit of Mango Taq DNA Polymerase (BIOLINE, Australia) in a total volume of 20 µL. Cycling conditions were 35 cycles of: denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 15 s. Successful product amplification was confirmed and DNA concentration estimated by 2% agarose gel electrophoresis and visualization by ethidium bromide (10 mg/mL) using a Bio-Rad Gel Doc and Image Lab Software (Bio-Rad Australia). PCR products were cleaned up in preparation for sequencing using ExoSap (GE Healthcare, Australia), following the manufacturer's protocol. DNA sequencing using either forward or reverse primer was performed by the Australian Genome Research Facility (AGRF) (Adelaide, Australia). Sequence alignment and editing was performed using Geneious v5.4.3 (Biomatters). The MS locus sequences were aligned with the existing Genbank records for each locus representing individual ABTC 55000: MS2, AF403194; MS3, AF403195; MS4, AF403196; MS9, AF403201; MS13, AF403205; MS16, AF403208; MS17, AF403209; MS25, AF403217; MS27, AF403219 (from [42]) and novel MsF primer sequences were aligned with sequence created from the initial MPS sequencing of individual ABTC 30087. Allele nomenclature was based on the underlying repeat motif number, as per [44,57].

3.4.10 Sample genotyping

The three multiplex primer assays were used to genotype native carpet pythons collected extensively from across the entire geographic range of the species. Two allelic ladders were included per plate, one at the start and one at the end, as ladder alleles were seen to migrate in an equivalent manner at these positions negating the need for extra ladder samples within the run plate. Samples were genotyped using allele calls corresponding to the underlying base motif number (in accordance with recommendation 7 of [44]).

3.4.11 Construction of an STR reference database and quality control

Genotypes obtained from all samples profiled using the three multiplex assays were collated in an Excel spreadsheet. All genotypes from any sample analysed on multiple occasions in separate runs were compared to ensure congruent results and any discrepancies were opened (with respective allele ladders) and compared directly in a Genemapper project. Where two individuals exhibited identical genotypes, one of these was removed from the dataset to avoid potential analytical difficulties caused by incidence of close relatives in the database or accidental repeated sampling of the same individual.

Any snakes with species name recorded as other than *Morelia spilota* or *M. bredli*, unspecified locality data, specified captive or hybrid, or exhibiting mixed profiles or profile ambiguities were removed from the dataset (although these samples can provide further utility as unknowns for the geographic assignment testing conducted in the following chapter). One individual exhibited a tri-allelic genotype at MsF24; this locus was omitted from analysis for this individual. Any homozygote loci that exhibited inter-locus peak balance patterns suggestive of presence of a null allele caused by primer binding site mutation were excluded from analyses. If any doubt was retained in an allele call within an individual, then this allele was specified as unknown. Any individual for which there was remaining doubt in the genotype was removed from the dataset. Any products suspected to be variant alleles due to anomalous size migration were designated the allele name with a * to differentiate them from other alleles of the same identity. Three of the 27 typed loci were removed from the final genotype dataset during quality assurance evaluation (see section 3.5.9 for further discussion about these observations).

Allele designations were converted directly to absolute base pair designations using an Excel Macro function for ease of downstream analysis in the software packages applied in this chapter. Each locus was converted individually. A colour based check was incorporated into the function to detect any unsuccessful conversions due to notation errors and all conversions were rechecked manually. As the largest allele identified was 525 bp, the four alleles designated * were allocated absolute base pair designations 600-603 to differentiate these as unique alleles.

3.4.12 STR database characteristics

GenAlEx v6.5 [58] was used to calculate population genetic parameters and estimate discrimination power. Estimates of the least and most common genotypes present in the population were calculated following [59]. Hardy-Weinberg's 2pq was applied to each locus using of the two most common allele frequencies and the results multiplied across all 24 loci to estimate the most common theoretical genotype frequency in the population [60]. The rarest allele frequency was calculated using the 5/2N minimum allele frequency rule (representing a minimum of five observations in the population) following [60,61].

Exact tests for deviations from Hardy Weinberg Equilibrium and examination for linkage disequilibrium between loci were conducted using Genepop on the Web (Genepop 4.2 ref). To detect equilibrium within loci, the probability option was run using the settings 1000 dememorization steps, 5000 batches and 10000 iterations per batch. Linkage testing was performed using the log likelihood ratio test (G-test) with Markov Chain Monte Carlo simulation settings of 1000 dememorization steps and 5000 batches, each at 10000 iterations.

3.5 Results and Discussion

3.5.1 Characterisation of Novel STR Loci

A total of 133,193 raw sequence reads were returned from MPS, yielding 7486 sequences containing unique STR loci. Primers were designed for 3225 loci. Sixty-five unique simple tetra- and penta-nucleotide repeat motifs were shortlisted. The maximum size of sequences containing simple motifs was 267 base pairs. Mutation rates have been demonstrated to increase with product size [62], so loci with longer amplicons were prioritised in the expectation that they would have higher mutation rates and therefore increased polymorphism and discrimination power. Loci with compound and complex motifs as well as sequences constructed of contigs were included to increase choices of amplicons of the high molecular weight range for multiplexing. Capillary electrophoresis instrumentation and available size standards easily allow fragment analysis of sequences up to 600 base pairs in length. A total of 167 putative STR loci were available for consideration with the longest amplification fragment size being 543 bp. Of these, 101 were shortlisted and ranked by repeat length and size to assess multiplexing options (figure 3.3). Thirty-six primer pairs were ordered with MRT tag sequences incorporated. Initial screening of fourteen individuals yielded eight loci that were discarded due to >2 amplification products, absence of amplification products or very poor peak morphology (figure 3.4). Of the remaining loci, allelic diversity ranged between 5-16 alleles. Allelic size ranges of loci ranged between 16-81 bp. Figure 3.5 shows the resulting 'ladder of alleles' from singleplex amplification of mixed DNA template from eight individuals, indicating a desirable allele diversity and range at this locus. See embedded publication [63] (appendix 3.2) for results of the MRT screening.

Taylor [43] screened 481 python samples with eight MS primers. Reported allele number ranged between 12-48 alleles per locus and allelic windows ranged between 48-156 bp. White (2011, unpublished) screened eight *Morelia spilota* samples with nine MS primers and obtained between four and nine alleles per locus with 22-84 bp allelic ranges. Nine MS primer pairs were chosen for inclusion in this study, prioritised by allelic diversity and size range, four also due to their overlap between the Taylor and White studies. Three loci were chosen due to interesting characteristics reported by Taylor [43]; MS3, MS13 and MS27 were all reported to exhibit indels, which could be of interest in indicating further population differentiation events masked by the homoplasy of rapidly mutating tandem repeats.

To avoid increased stutter artefacts associated with the smaller di- and tri-nucleotide repeat motifs, tetra-, penta- and hexa-motifs were targeted during initial locus selection. Tetra-nucleotide repeats are most commonly used in human forensic STR analysis [44,60], however the larger motif types were chosen also to examine whether these offered any potential advantage over four base pair repeats. The initial 36 loci included four loci representing simple penta-nucleotide repeat motifs. One of these loci (MsF6) failed to generate data from multiple attempts at Sanger sequencing, so was removed from the dataset. Another locus (MsF14) exhibited complications in consistent amplification, which was identified to be due to a repeat motif in the primer sequence that was originally overlooked (see embedded publication [64], appendix 3.4). The other two penta-nucleotide motif loci amplified consistent product and were included in the final dataset.

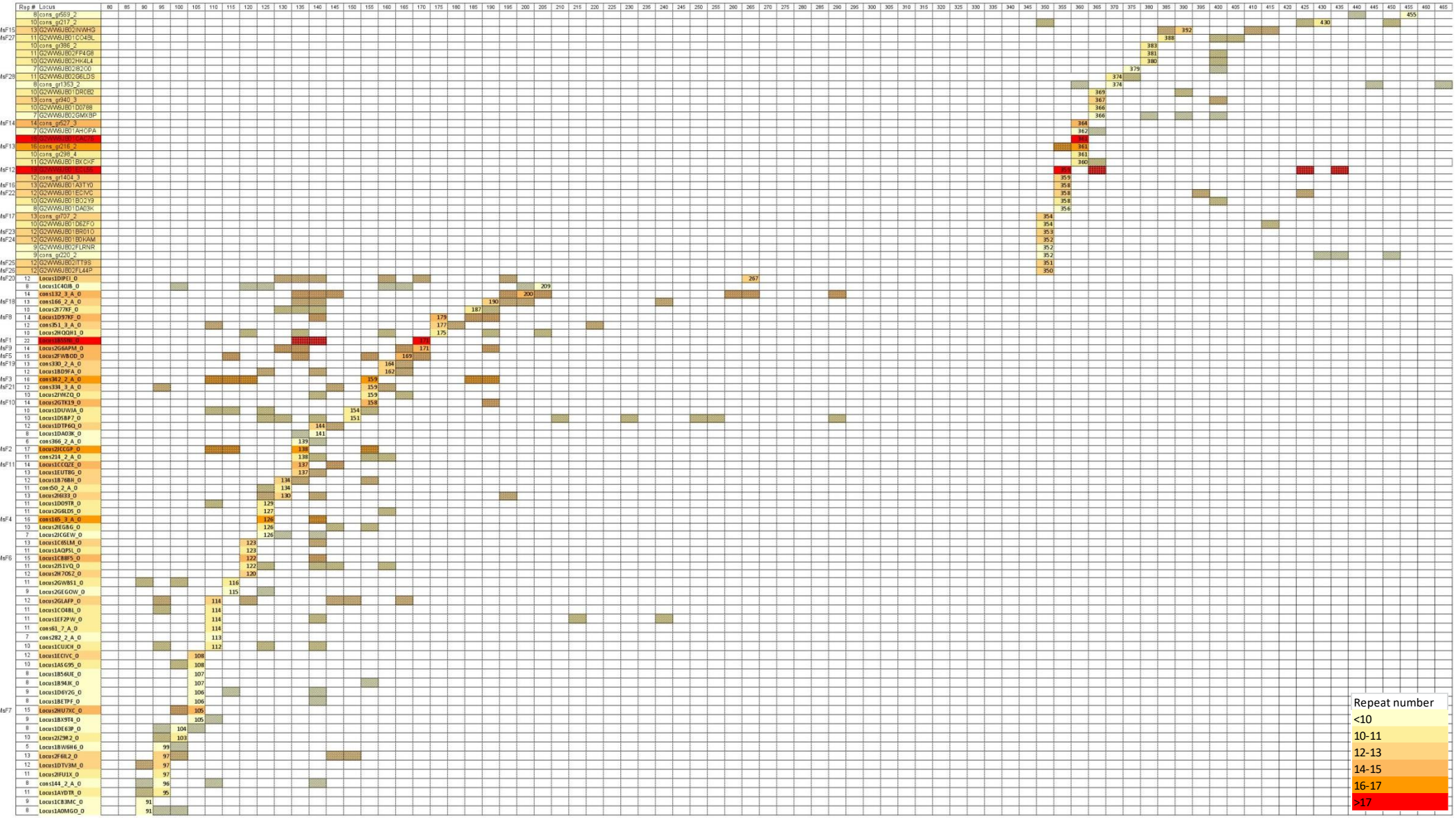


Figure 3.3: Planning chart of short-listed pure (<300 bp), compound and complex (>300 bp) STR loci. Colour heat map (bottom right) represents locus repeat motif number. Unshaded box contains product size (bp) using preferential primers, whereas other shaded boxes in row represent product size using alternative primers.

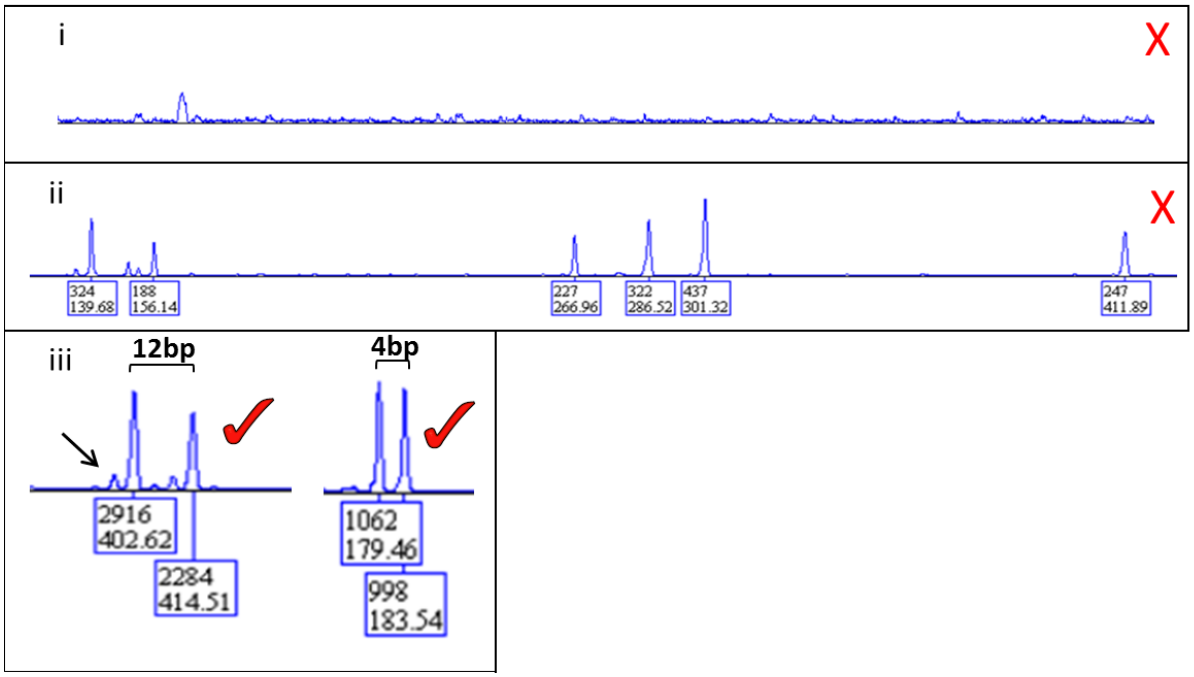


Figure 3.4: Examples of product scoring during MRT testing. Undesirable results are shown in i) no PCR product, and ii) more than two PCR products. iii) Two loci exhibit desirable products involving stutter peaks and the expected allele size separation reflecting the underlying motif type, thereby representing true STR amplification products.

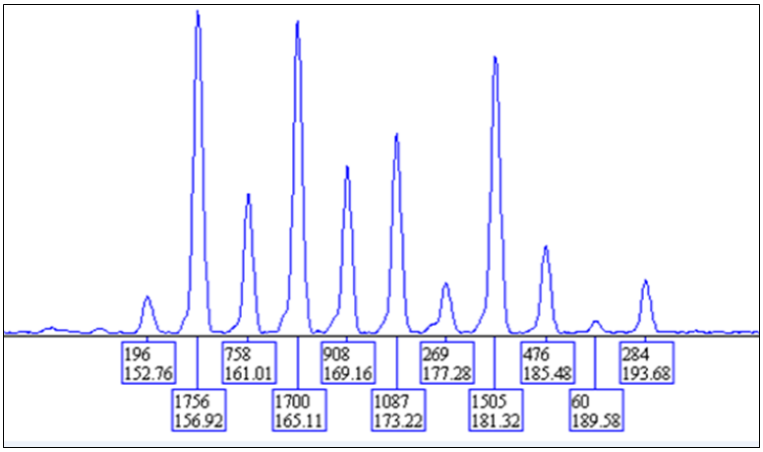


Figure 3.5: Amplification of a mixed DNA template of eight individuals using a single primer pair to determine the degree of allelic diversity at a locus.

A compound locus containing two penta-nucleotide motifs separated by a single base (MsF34) amplified what appeared to be genuine STR product, but was outranked by the loci chosen for the final multiplex marker set.

No preferable simple hexa-nucleotide repeat motifs were identified during *in silico* locus examination. One compound locus (MsF35) contained two tetra- motifs followed by a hexa-nucleotide motif. This locus failed to amplify any products during the MRT screening stage and was discarded. The initial 36 locus marker set was chosen to contain 29 simple repeat motifs and as well

as seven compound/complex motifs to see whether these brought any added advantage over simple repeat motifs (note that seven of the final MS markers chosen were compound or complex while only two were simple repeat types). However, interrupted repeat types have been observed to show decreased mutation rates [62], so the majority of motifs selected were simple. Certainly, the compound and complex markers displayed a much higher incidence of micro-variant alleles than the simple repeat motifs (see coloured bars on ladder diagrams), leading to much higher allele counts per locus. In the developmental stages of the project this was considered an advantage for maximising heterozygosity and therefore the discriminatory power of the assay. However, as will be discussed in chapter 4, high allelic diversity brought its own challenges, complicating the estimation of accurate allele frequencies from which to calculate discriminatory statistics.

3.5.2 Multiplex assay design & optimisation

There were a number of initial considerations for arranging the multiplex assays. A primary concern was the inclusion of all top ranked MsF and MS loci while avoiding allelic size range overlap within the same dye channel of the same multiplex assay. Having been screened in such a large number of samples, the MS loci exhibited a high number of alleles and a related large allelic range. This indicated that fitting more than one or two loci within a dye channel might be problematic once sample screening had been expanded – a consideration to be dealt with down the track. Translation of the MRT amplification products to a size range for STR multiplexing required a 30 bp decrease in product size with the removal of the MRT tag sequences from the primers. Any MS primers included in the study would either need to be tagged with the same fluorophore as used by previous studies, or an offset applied to the absolute base pair size of the fragment corresponding to the different fluorophore mobility, for cross-comparison between studies to be effective [65].

Qiagen Multiplex Kit results were far superior to Amplitaq Gold II results, the latter only partially amplifying the included loci. Cycling parameters were optimised to 29 amplification rounds and addition of 1 µL of ~1 ng DNA template in line with commercial STR kits (e.g. [73,74]). The use of 27 and 35 cycles created partial and grossly overloaded DNA profiles, respectively. A reaction volume of 20uL was used, but once 22 primer pairs were added, the reaction volume was adjusted to 25 µL to keep the reaction components in the required balance of concentrations whist allowing addition of many primers.

Figure 3.6 illustrates iterations of charts used to plan the layout of the three multiplex assays over the course of the design stage. This initial design based on MRT screening in up to 17 individuals allowed the three multiplex assays to contain either two or three loci per dye channel without allelic window overlap between loci in the same channel of the same assay. The three purple bordered allele ranges, MsF31, MS17 and MS2, represent those three loci included in all assays as redundant quality markers. These loci were chosen because their allele ranges did not overlap with allele ranges of any other loci labelled with the same fluorophore across all three assays. Ideally, at least one quality marker should be placed down the low molecular weight range so that there is still a flag for non-concordance between assays in degraded samples. It is a difficult trade-off because degraded samples might only amplify small fragments and including two extra unique loci rather than a redundant quality marker across the three multiplexes could add much needed discrimination

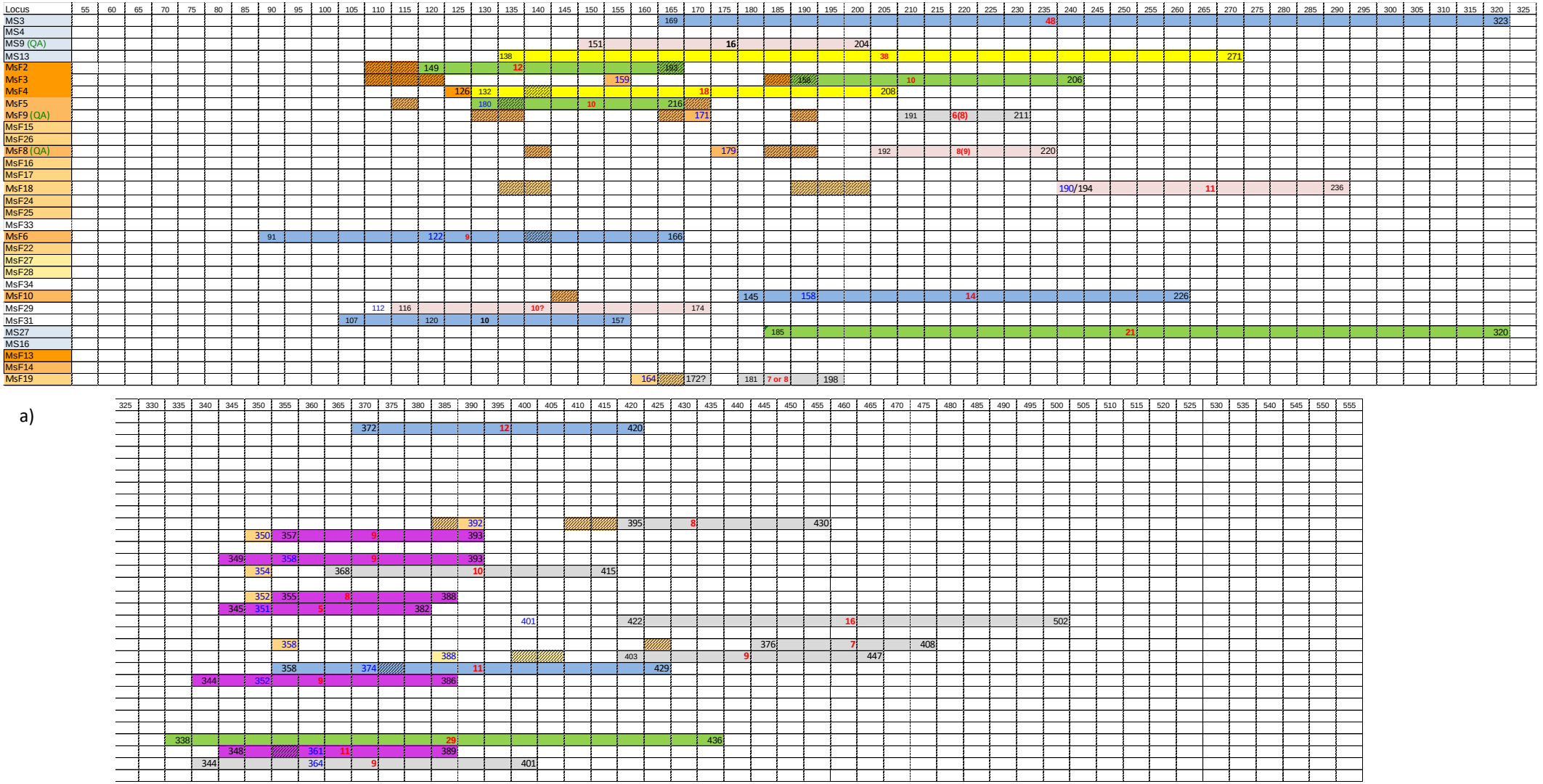
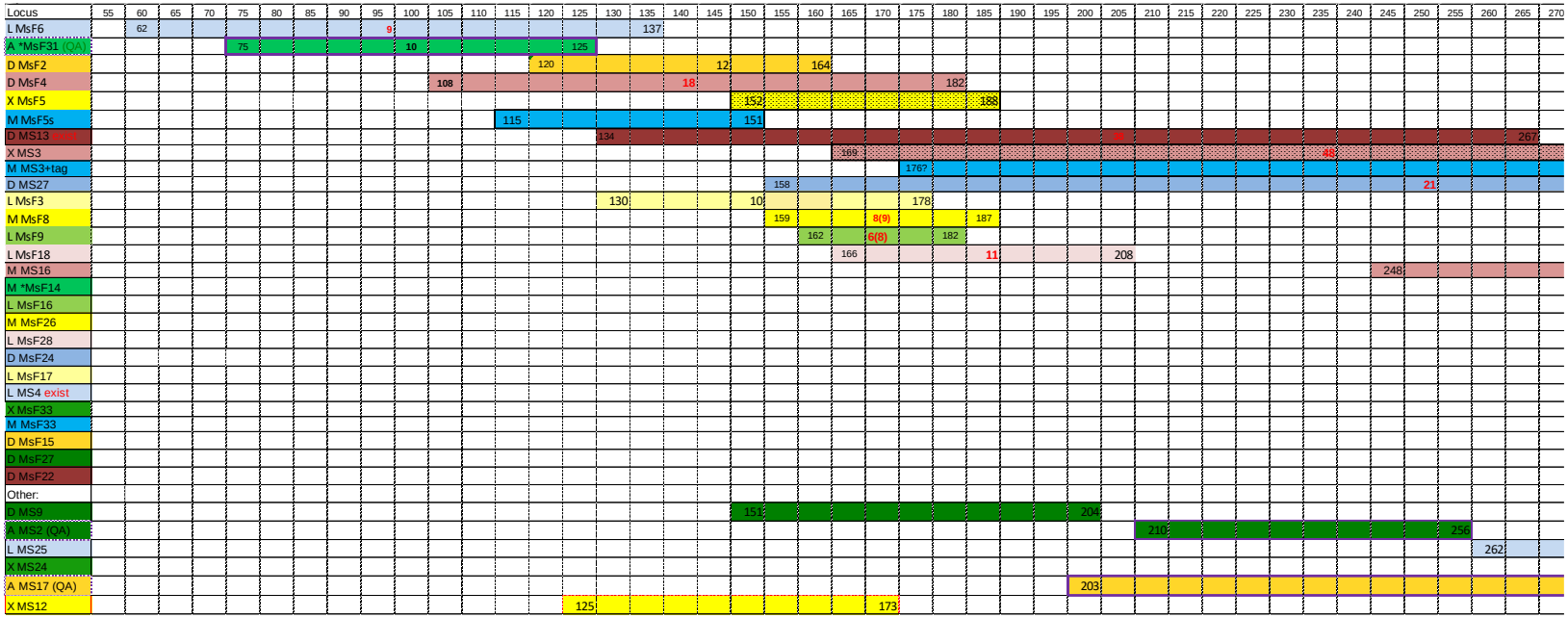


Figure 3.6: Iterations of multiplex planning. a) initial mapping, b) reordering of loci by product length and allocating fluorophore (colour) and multiplex (shade), c) locus sizes updated based on further screening, and d) redesigned multiplex assays.



b)

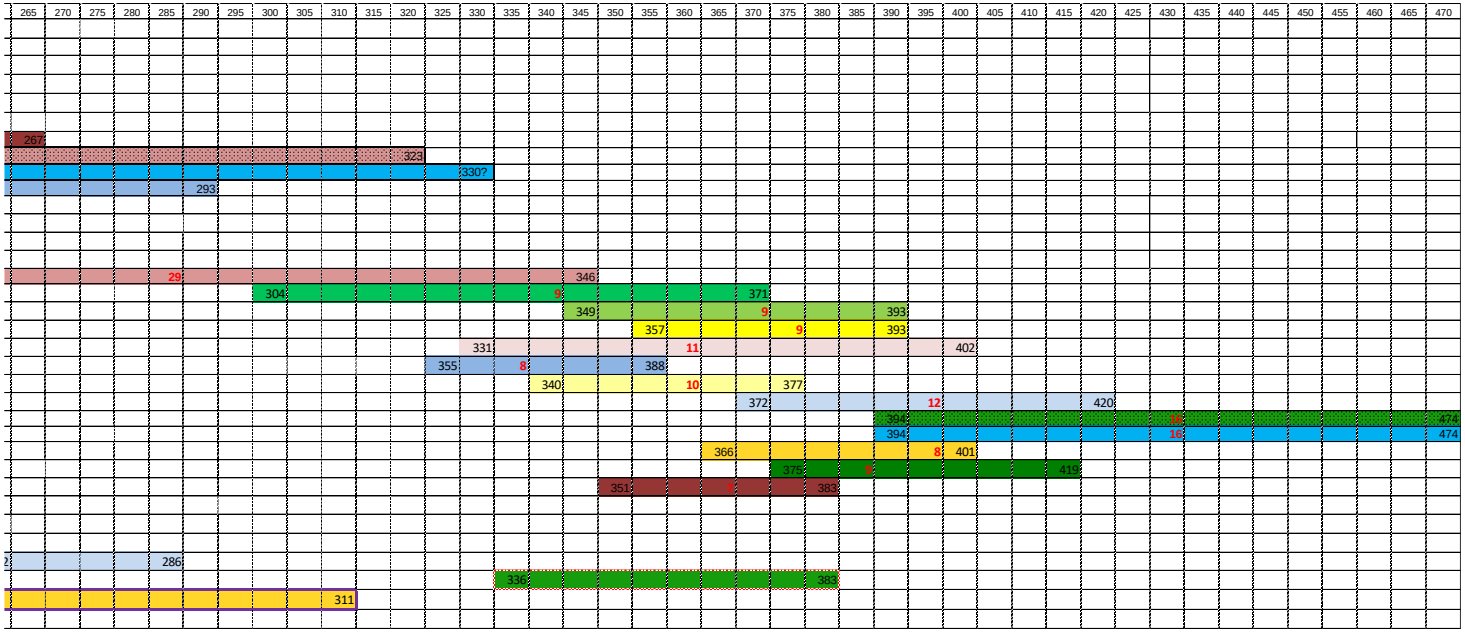
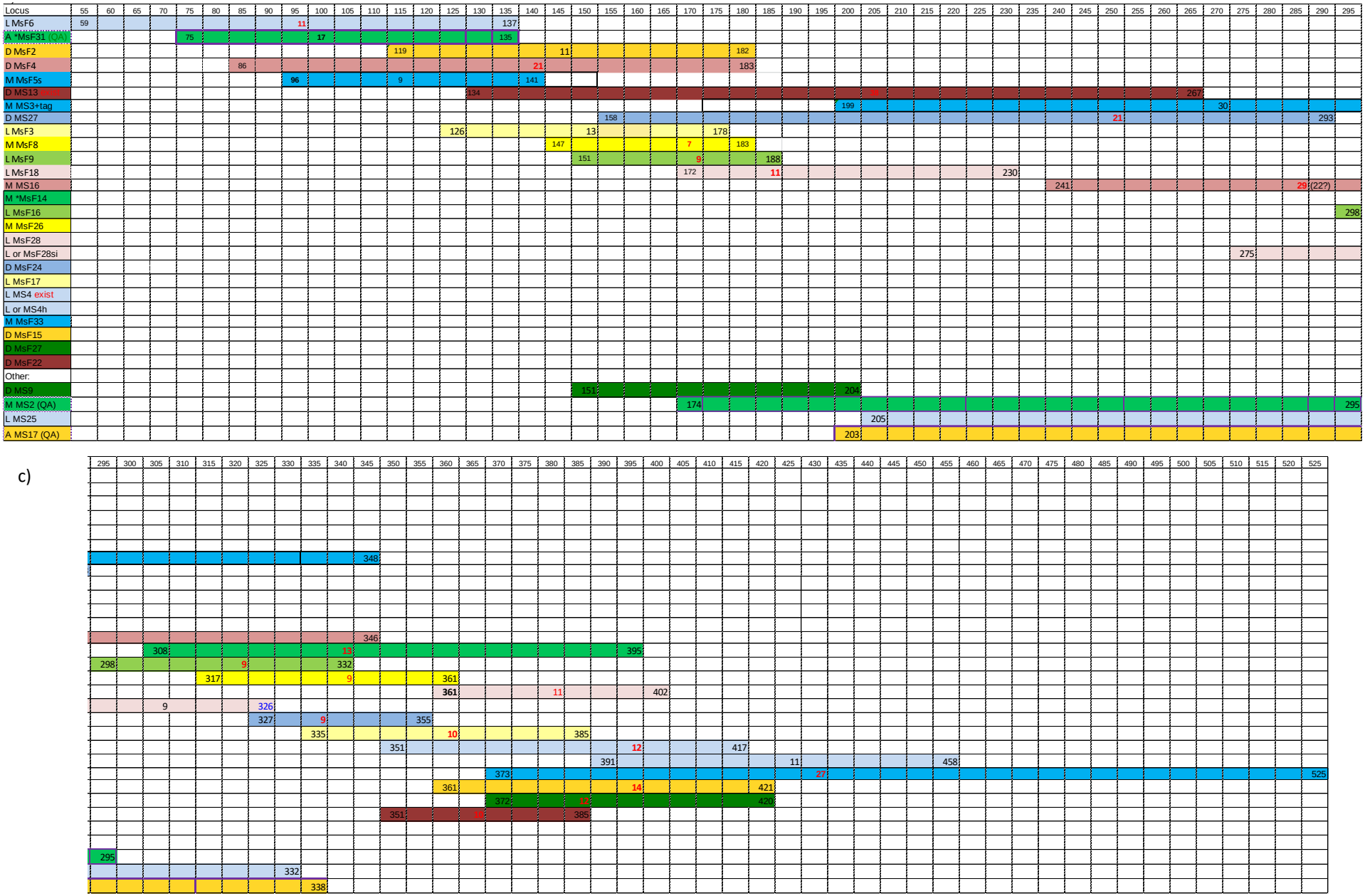
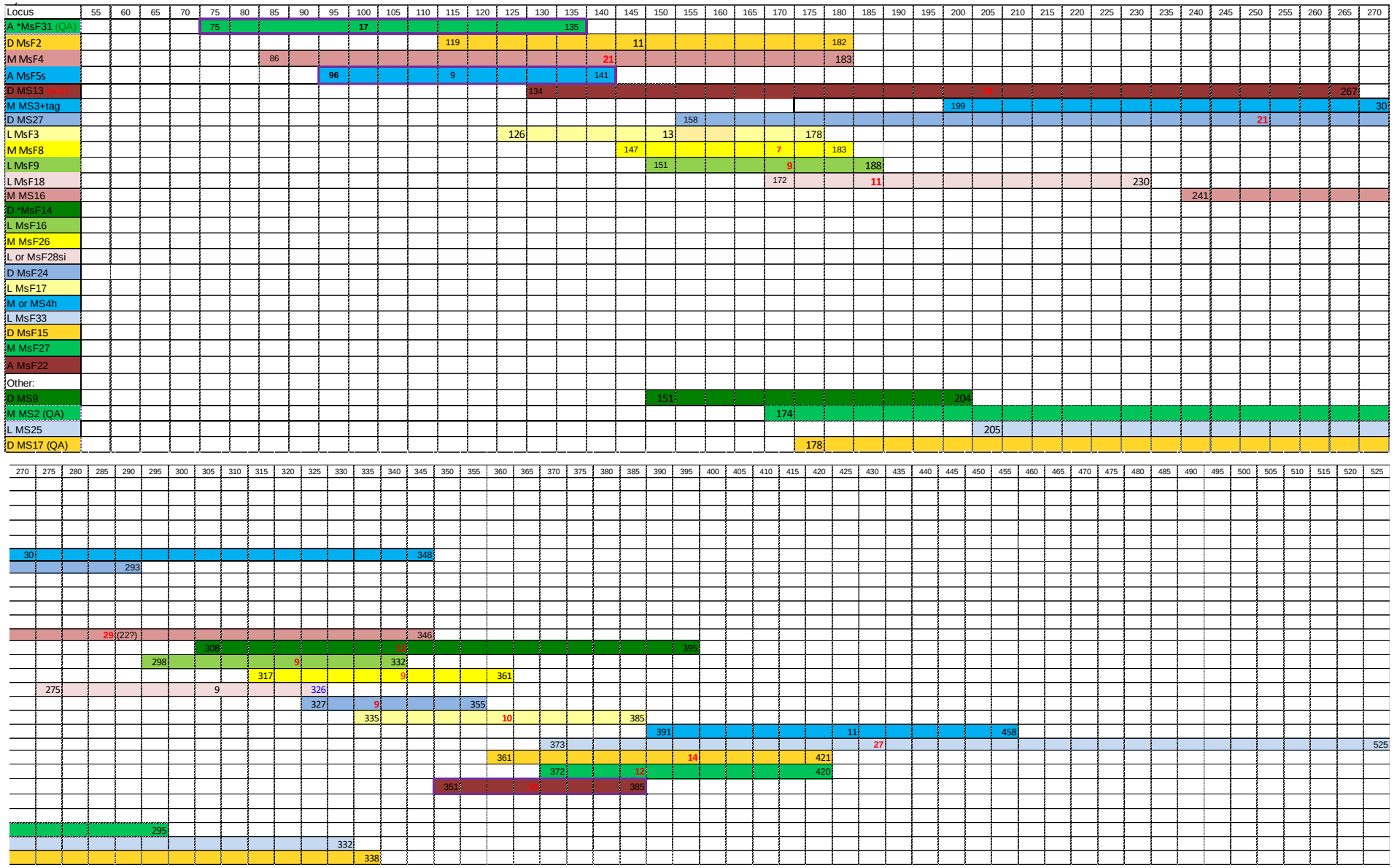


Figure 3.6 (cont.): Iterations of multiplex planning. a) initial mapping, b) reordering of loci by product length and allocating fluorophore (colour) and multiplex (shade), c) locus sizes updated based on further screening, and d) redesigned multiplex assays.



c)

Figure 3.6 (cont.): Iterations of multiplex planning. a) initial mapping, b) reordering of loci by product length and allocating fluorophore (colour) and multiplex (shade), c) locus sizes updated based on further screening, and d) redesigned multiplex assays.



d)

Figure 3.6 (cont.): Iterations of multiplex planning. a) initial mapping, b) reordering of loci by product length and allocating fluorophore (colour) and multiplex (shade), c) locus sizes updated based on further screening, and d) redesigned multiplex assays.

power to the final statistic. However, quality assurance is paramount in forensic assays and a highly discriminating statistic is perilous if the final profile has been unknowingly combined from two different individuals due to sample mix up between multiplex assays.

Primer design had to be well planned prior to synthesis, as the large fluorescently labelled marker set would be expensive to resynthesise. Primer design had been conducted using a standardised annealing temperature to facilitate multiplexing all primer pairs at the same optimal annealing temperature. The relatively high (60 °C) PCR annealing temperature maintained annealing stringency. Although some potential primer interactions were identified when the primers were analysed using the Multiple Primer web tool, these did not appear to affect the multiplex at the chosen melting temperature. Almost all loci performed well at this temperature; locus MS4 was redesigned due to poor amplification relating to low primer annealing temperature.

Primers were immediately redesigned for two loci (MsF16 and MsF26) prior to synthesis to yield shorter amplicons because they were noted to fall in a region that was highly congested with loci of overlapping product size (compare b) and c) of figure 3.6). This move attempted to avoid overlap in the later stages of database screening. A smaller product size is also advantageous for the higher likelihood of successful amplification with degraded or low concentration DNA template. Likewise the primers from locus MS27 from Taylor's [43] study were redesigned to give a smaller product and more evenly spread the FAM labelled products across the multiplex size range (compare a) and b) of figure 3.6). This was an important marker to include because Taylor had noted an indel in the flanking region captured by the primers that could distinguish population traits masked by homoplasy in the STR allelic variability itself. Both primers were redesigned to shorten the amplicon product as much as possible while conserving the indel within the flanking region (figure 3.7). Both original MS27 primers also showed potential complementarity with other locus primers, so redesign (new primers and locus suffixed with 's') served the dual purpose of shortening the amplicon size and overcoming potential primer-primer interactions during multiplexing.

As a polymerase replicates a DNA template, it will add a terminal nucleotide (adenosine in most cases) to the 3' end of the complementary strand, known as adenylation. Whether or not this has occurred is of no consequence for agarose gel electrophoresis visualisation. Inconsistent or incomplete adenylation does cause issues with genotyping by high resolution of capillary electrophoresis, which allows discrimination between single base pair products and for which micro-variant alleles are likely to be separated by single bases. The result is a shouldering of the target peak or a split peak.

Techniques can be employed to encourage complete adenylation. Addition of a final incubation step of 60 °C for 30 minutes can encourage complete adenylation of the complementary strand [53]. Another technique to promote complete adenylation is to commence the reverse primer sequence (seeing as the labelled complementary forward primer is the product visualised

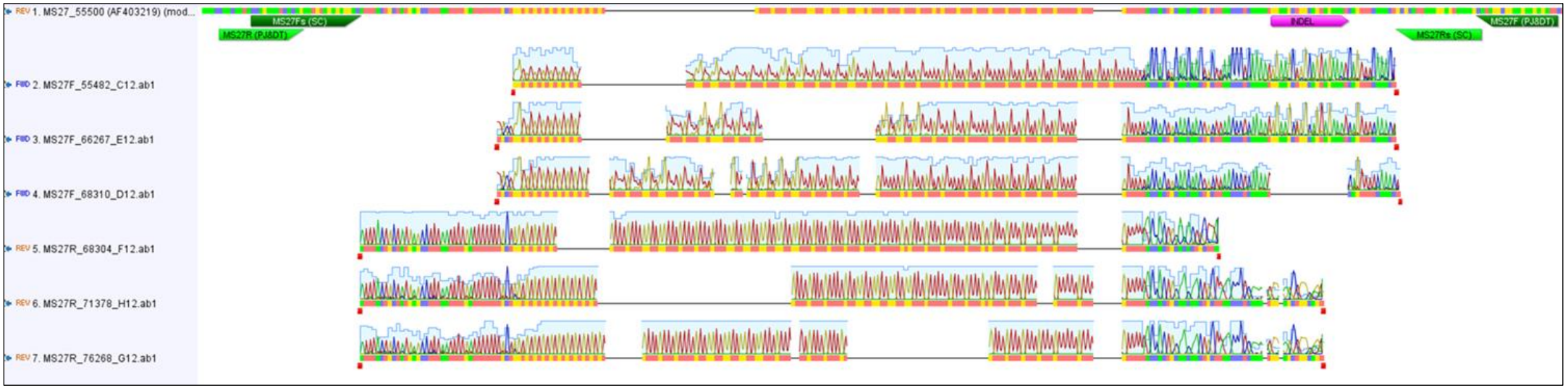


Figure 3.7: Positions of original primers designed by Phil Jordan and applied by Duncan Taylor at locus MS27, as well as slightly smaller primers designed during this project to shorten the product length while retaining the flanking region indel (shown in pink).

during electrophoresis) with a guanosine base [52]. This will ensure the visualised product ends in an adenosine nucleotide. During primer design, all initial reverse primers had a G nucleotide added to the 5' end to encourage complete adenylation. The MS primer set designed by Jordan and colleagues [42] was noted to give split peaks on occasion. A seven nucleotide tail was added to this primer pair, slightly modified from the 'GTTTCTT' pigtail reported by Brownstein and colleagues [52] to promote adenylation, to overcome the poor morphology initially seen. The new primer was named MS3Rt to denote inclusion of the tag sequence. This primer pair gave consistently good peak morphology with absent shouldering, as can be seen in the Medium ladder and multiplex profile results presented in the embedded publication [54], appendix 3.1.

Initial amplifications using the QIAGEN Multiplex Mastermix yielded amplification products at all loci, although some were quite weak and others were very overloaded. Primer concentrations were varied, which resulted in better balance between loci across the profile in subsequent optimisation rounds. Initial optimisation results obtained from the Light locus are shown in Appendix 3.6.

3.5.3 Multiplex assay redesign

As anticipated, screening of an initial 43 individuals resulted in an expansion of allele ranges that required a redesign of the multiplex assay to avoid allele range overlaps between loci. When rearranging the multiplex layout, if any locus required a fluorophore switch, then an allele base pair size correction corresponding to the shift in migration caused by the alternate dye label would need to be applied. This however was not an issue as all fluorescent labels were able to be retained for the final conformation.

The following changes were applied (figure 3.6, see b), c) and d)): locus MsF6 was discarded because all attempts to sequence the underlying repeat motif failed and was replaced by MsF5, which was designated a quality marker and added to all three multiplexes; MS2 and MS17 were removed as quality markers due to a large number of micro-variants causing extended size ranges and overlap in multiple assays, so these were retained in only the Medium and Dark multiplexes respectively; MsF22 replaced MS17 as a quality marker due to improved peak morphology and a smaller allele range, allowing it to fit easily within all three multiplex assays; MS4 was removed from the Light assay and replaced by MS4h in the Dark assay; MsF28 reverse primer was redesigned to capture an indel in the flanking region; MsF4 and MsF27 were moved from Dark to Medium and MsF14 was moved from Medium to Dark; MsF33 was moved from the Dark assay to the Light assay.

Four loci required primer redesign. Locus MS4 was redesigned due to consistently poor amplification. The T_m and GC content of Taylor's reverse primer were noted as considerably lower than all other primers. A new forward primer site was picked, adding an additional five base pairs to the primer length which raised the GC content and T_m to be uniform with the other primers. This primer was designated MS4Fh (h for hot). The resulting amplicons are 30 bp larger than MS4 amplicons, so the name of the new product MS4h was retained to differentiate the respective

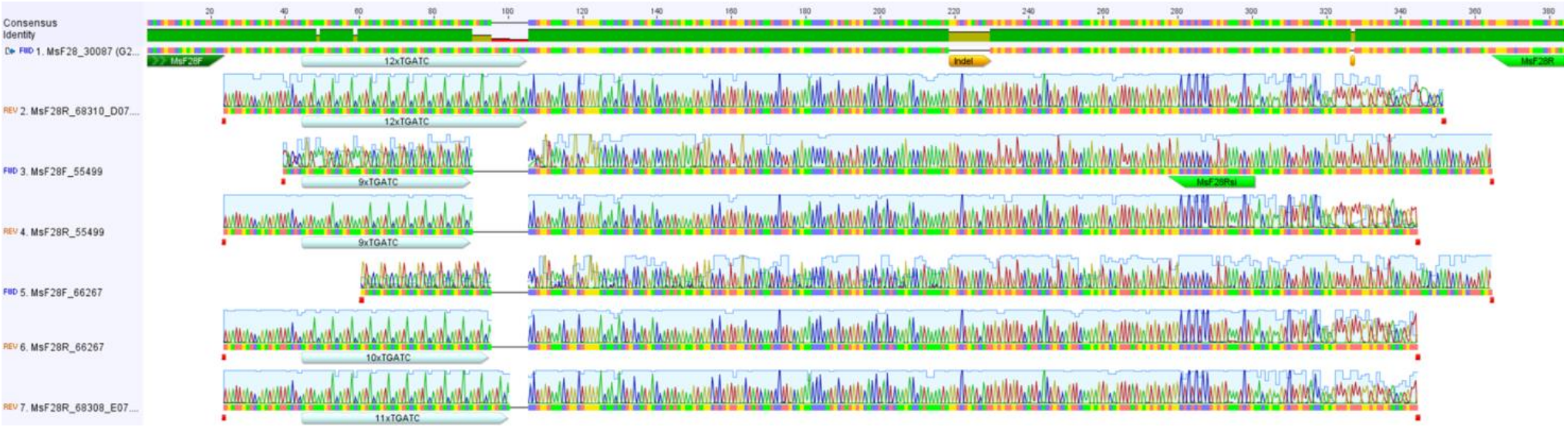


Figure 3.8: Locus MsF28 sequence alignments showing the position of the two indels within the flanking region of the MsF28 amplicon. These indels were evident in individual ABTC30087 from the Northern Territory, but none of the other sequenced individuals.

products from this and previous studies and a 30 bp sequence size increase was applied to all products successfully amplified previously using MS4. Primers MsF5R and MsF26R were redesigned to shorten the amplicon sizes and shift products from congested areas of the chart to lower molecular weight regions. This also enabled the new quality marker to better fit within all three multiplex assays.

Upon sequencing, locus MsF28 was found to have two indels in the flanking region, one an 11 bp deletion and the other a single base deletion. The reverse primer was shifted with the logic that removing the single base difference and shifting the sequence variation away from a factor of four would make individuals with the indel apparent as micro-variants by size migration without the need for sequencing (Figure 3.8). The combined 12 bp difference would otherwise make an individual with the two deletions indistinguishable from any individuals possessing a complete repeat motif three repeat units smaller. It was later realised that this locus is not in fact a tetra- motif, but one of the three penta-nucleotide repeat motifs, so individuals possessing the 12 bp deletion should present as micro-variants anyway. The reverse primer was redesigned 87 bp upstream to exclude the single base insertion and designated MsF28Rsi (s for shorter, i for indel), reducing the new amplicon size. This primer design was not completely in vain as the smaller PCR product is favourable for the multiplex layout. The smallest allele still sits 45 bp larger than the largest allele in the neighbouring PET labelled MsF18 locus and this downstream shift also allows the new MsF22 quality marker to fit upstream of the MsF28si allelic window.

The redesignation of the quality markers meant that these markers are spread across three different dye channels and also shifts two QA markers to the low molecular weight region of the profile and one to the high molecular weight (which importantly retains the quality feature at two loci in degraded or poorly amplifying samples).

Table 1 of [54] embedded in appendix 3.1 lists the final STR loci included in the three 11-plex profiling assays with corresponding primer sequences and optimised primer concentrations for each multiplex assay. Figure 1 of this publication provides an example of each multiplex assay.

3.5.4 Construction of three allelic ladders

Figure 2 of publication [54] shows the allelic ladder corresponding to each multiplex assay.

3.5.5 Ladder Immortalisation

The best allelic ladder reproduction resulted from a 1/1000 dilution amplified for fifteen cycles.

3.5.6 Validation testing

3.5.6.1 Repeatability

The sizing results for Light Optimisation run 1 exhibited a consistent approximate two base pair negative shift in the absolute base pair size of all alleles across the profiles. Although alleles within the profile sized correctly in comparison to each other (when comparing sample ABTC62456 alleles across optimisation runs), this run was excluded from the repeatability study as results were anomalous with the rest of the data. The anomalous migration of this run was initially thought to have been caused by either a run specific problem within the instrument, or the result of heavy

earth moving works being conducted nearby. However, external vibrations causing the instrumentation to knock out of alignment should not cause the observed outcome. The sizing is based on comparison to the internal size standard and this is run concurrently with the fragments being analysed within each capillary. A difference would not be expected to be seen between the migration of the size standard and the migration of the target fragments even with external movement - the advantage of including an internal size standard. The anomalous migration is believed to be caused by an internal run problem, although the exact nature of the cause is unknown.

Of the 24 loci chosen for the final database, 19 complete loci (34 alleles) exhibited less than 1.00 absolute base pair migration variability between all runs over all assays, obtained using different instruments in two independent laboratories. An average migration range of 0.61 bp was obtained over the 34 alleles. The homozygote allele at quality marker MsF31 exhibited 0.82 bp migration variability over 21 runs. Four loci (MS25, MS17, MsF16 and MsF33) exhibited at least one allele with variability greater than 1.00 bp. The alleles belonging to three of these four loci (MS25, MS17, MsF16) sized within the range of 272 bp to 319 bp and had been run using the Liz500 size standard, resulting in the inconsistent sizing of these alleles.

The fourth locus (MsF33) exhibited a large allele sizing at 435 bp which showed 2.24 base pairs variability. The two runs conducted on the different instrument at FSSA appear associated with this anomaly, both runs sizing below 434.6 absolute bp, whereas all five runs on the Flinders instrument sized >435.99 absolute bp. The reason for this anomaly cannot be identified, but this is the only instance in the repeatability dataset where an allele has migrated on the FSSA 3130xl instrument more than 1 bp different from the Flinders 3130xl. All other instances where the migration range exceeded 1 bp showed differentiation in migration size between runs on the same instrument rather than differentiation between the two instruments.

The results indicate the importance of including an allelic ladder with every run.

3.5.6.2 Reproducibility

Decreased fluorescence intensity was seen in the samples run at Flinders University following their visualisation at FSSA. Fluorophores are light sensitive and fluorescence intensity is known to decrease over time [66]. However, the decrease in signal intensity between the two electrophoresis events run only two weeks apart did not appear consistent with plate reruns at Flinders. The FSSA instrument, which is maintained regularly within an ongoing service contract with Applied Biosystems, is suspected to exhibit superior sensitivity to the instrument at the Flinders University lab.

3.5.6.3 Heterozygote balance

Figure 4 of embedded publication [54], appendix 3.1 provides the mean heterozygote balance seen per locus. The error bars represent 95% confidence intervals. Dashed lines represent 60% balance in either direction (ϕHMW peak larger than ϕLMW peak or vice versa), included to indicate the traditional recommended threshold of balance at a single source locus [67], prior to the development of continuous models. Only three loci exceed the balance threshold (2σ), but all exhibit H_b greater than 50%. All but one locus exhibit a mean balance favouring increased amplification of

the LMW product, demonstrating the recognised amplification efficiency bias towards smaller DNA templates and therefore alleles containing less repeat units [68].

3.5.6.4 Species specificity

The high resolving power that affords STR loci the ability to differentiate individuals within a species precludes the application of a developed set to all but the most closely related species of the developmental target taxon. Expansion of the species range for which forensic grade STR assays are available is hampered by the resource intensive development process (as evidenced in this chapter). A pilot study (see appendix 3.2 for embedded publication [63]) was conducted to assess the potential of the STR loci developed in the Australasian carpet python for application to other closely related pythons of forensic significance.

All Australian pythons showed successful amplification products at ten or more loci of the developed sites cross-amplifying in each species. Amplification success was limited in the more distantly related species of the *Python* genus. Human DNA failed to amplify any product; this is important to ensure that contaminating human DNA transfer at the time of collection or previous handling will not compromise the result. The pilot study demonstrates the potential of many of the STR markers developed for the Australian carpet python to cross-amplify in other Australasian python species of forensic significance. A subset of loci have been suggested for multiplex amplification, however locus size ranges can vary greatly between species [69], so initial screening using the MRT approach [49] is recommended for a small number of individuals of the target species to identify the most appropriate loci to pair in a fluorophore channel before the final fluorescently labelled markers are ordered for multiplex genotyping.

3.5.6.5 Assay sensitivity

Full profiles were obtained across all 24 loci from input DNA concentrations between 5 – 0.5 ng/µL template. Spectral saturation or pull-up was observed in all multiplexes at 5 ng/µL and 2 ng input volumes. Optimal results were obtained using 1 ng/µL with no pull-up observed and moderate to strong amplification of all alleles. At 250 picograms (pg) per µL, all loci showed amplification products in the Light and Dark assays, but locus MS4h was lost from the Medium assay. Addition of 125 pg/µL template results in homozygote peaks below 150 relative fluorescent units (RFU). Although a homozygote peak threshold has not yet been developed for this dataset, the low homozygote peak strength indicates potential for heterozygote allele dropout to occur. Partial amplification was seen down to 31.25 pg template and below this concentration no genuine locus amplification was detected in the Light and Medium assays. Five alleles were called at 15 pg/µL DNA input in the Dark assay, but below this concentration any potential allele peaks could not be confidently called as such. Determination of thresholds for allelic dropout of loci included in these assays is beyond the scope of this study, but should be conducted empirically by any laboratory planning to implement these profiling assays in casework. Appendix 3.7 shows the sensitivity results for the Light multiplex assay.

A 1 ng/µL template concentration is recommended for this multiplex, but full profiles have been obtained down to half a microliter for the Medium assay and 250 pg/µL for the other assays. The threshold for any amplification product has been demonstrated to around 30 pg/µL, except for the Dark assay which amplified product down to 15 pg/µL. This result cannot be compared to validation

results of one of the commercial kits (e.g. Globalfiler, Applied Biosystems) because the commercial test has been run on a 3500xl genetic analyser, known to be more sensitive than the earlier model 3130xl used in this study.

3.5.6.6 Selectivity

Full DNA profiles were obtained from all sample types extracted as part of this project: blood, liver, muscle, scale clipping, sloughed skin, tail and tissue. This indicates that all of these samples types are appropriate sources of DNA for carpet python DNA profiling.

3.5.6.7 Sanger sequencing

Sequencing the underlying STR motif of at least one allele at each locus is an important step of the validation process. It is crucial not only for verification that the expected underlying repeat motif is responsible for the alleles interpreted by size migration. Allele nomenclature is established by identifying the peak migration size that relates to a particular underlying repeat motif number and naming all other alleles in accordance.

Initial screening identified 368 homozygote individuals for potential sequencing. It is good practice to sequence a sample with both forward and reverse primers to check that both sequences agree. However, with limited resources the decision was made to sequence twice as many individuals in one direction to maximise the representation of sequenced alleles. Comparison between complementary sequences is particularly important when the sequences are of poor quality and there is ambiguity in the base designation at a site. Due to the high quality of genomic DNA available for most individuals, sequences returned were either of excellent quality or failed completely.

A total of 141 sequences were attained. Sequence was unable to be obtained from ten attempts (five attempts each with forward and reverse primers) at locus MsF6. This locus was subsequently removed from the study. Of the 27 loci included in this study, only MS16 did not give meaningful sequence that revealed the nature of the underlying repeat motif. This locus is not recommended for application to casework based on other genotyping issues observed [54]. All sequences containing a complete STR repeat region (121) have been deposited on Genbank under Accession Numbers MG548390-MG548510. See Appendix 3.8 for all sequence details.

Repeat motif types

Of the individuals sequenced, the following loci returned underlying sequences corresponding to simple repeat motifs: MS4, MS9, MsF2, MsF3, MsF4, MsF5, MsF8 MsF9, MsF14, MsF15, MsF16, MsF17, MsF18, MsF22, MsF24, MsF26, MsF27 and MsF28 (see figure 3.9 for examples). Note that many of these loci exhibited micro-variants. Those that show none are MsF3, MsF8 and MsF9, while loci MsF16 and MsF22 display a single micro-variant in almost 500 alleles screened (refer to table 2 in [54]. MsF24 displayed two micro-variants in the low molecular weight alleles (these are both confined to a single population as discussed further in chapter 4).

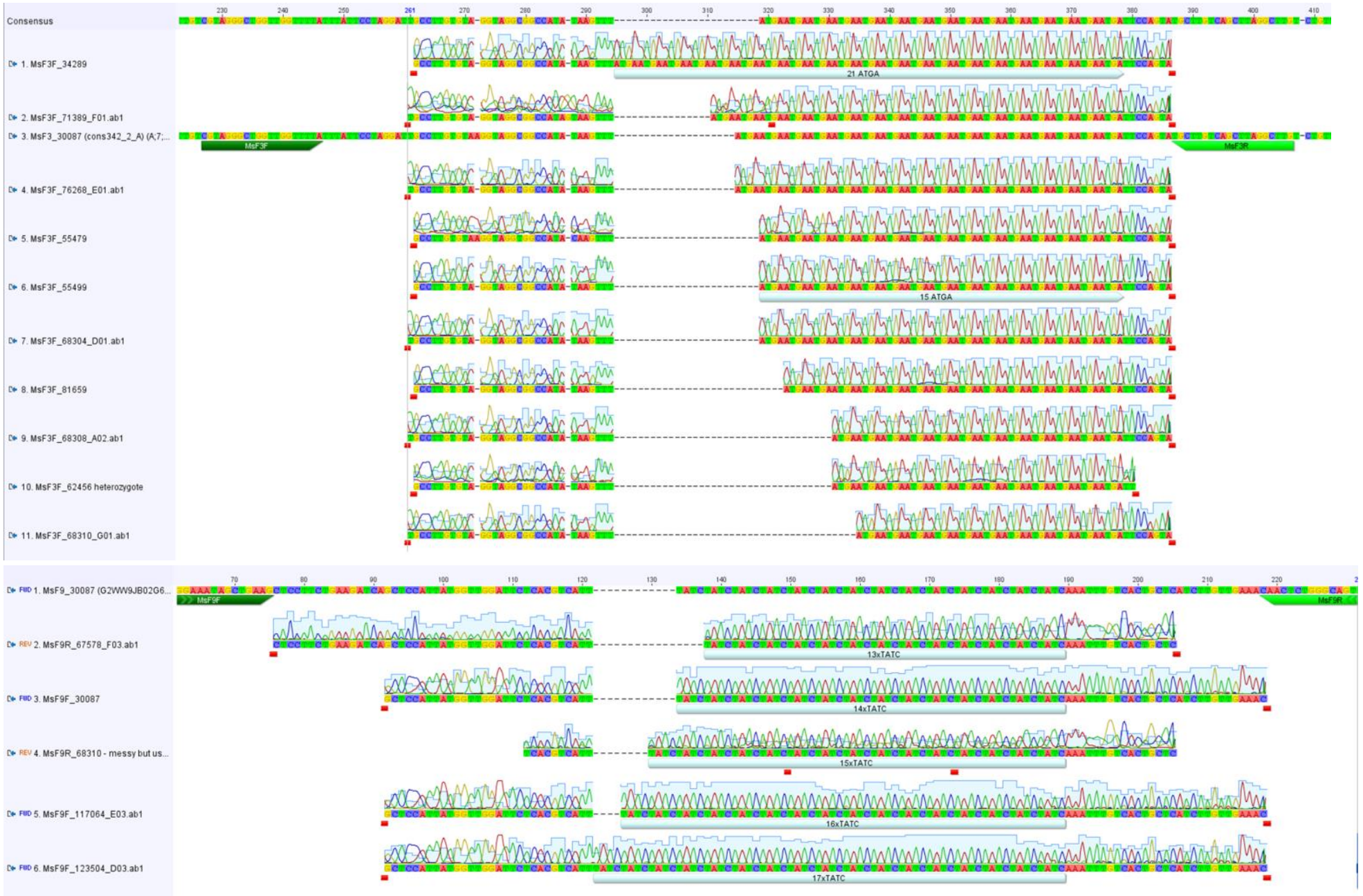


Figure 3.9: Example alignments of simple repeat motifs sequenced during this study. a) Locus MsF9, and b) Locus MsF3. Both loci nicely illustrate the stepwise whole motif increase of the underlying repeat unit. Five and seven different allelic polymorphisms can be seen based on the underlying complete tetra-nucleotide repeat units, 'TATC' and 'ATGA', respectively.

MsF18 showed a single micro-variant within the common range of alleles (9-21), but the two largest alleles sitting outside this allelic range were both micro-variants. It's possible that these are caused by a mutation causing a simple repeat unit to change motif and spiral out of control as seen at MS25 (figure 3.10), but without sequencing this is pure speculation.

MsF2 likewise has no micro-variants smaller than allele 20, but many micro-variant alleles larger than this size. All sequenced alleles are simple complete repeats except the largest allele, 18, which has a SNP present in the STR repeat region (figure 3.11). MsF4 is an example of a locus that displayed complete simple repeat motifs in all individuals sequenced, yet the allele frequency database shows a number of low frequency .2 micro-variant alleles. These could be due to an indel, but sequencing of individuals with these alleles is necessary to establish whether the variants share the same underlying cause.

Two compound loci are evident, MsF31 and MsF33, while it is debatable whether MS25 is compound or complex due to the mutation of a dinucleotide in one individual that is monomorphic in all other sequenced individuals (figure 3.10). The six remaining loci (MS2, MS3, MS13, MS16, MS17, MS27) are clearly complex, containing a mixture of polymorphic motif types (Appendix 3.8).

Indels

Loci MS3, MS13 and MS27 were included in this study because they were reported by Taylor [43] to contain indels characteristic to certain populations. A large $(AAGG)_{19}$ insertion in MS3 directly downstream of the STR repeat motif is reported as characteristic of Centralian carpet pythons and polymorphic in other populations [43]. This trait was not seen in the two individuals sequenced herein, neither of which originated from the Central Australian population.

An 'AA' insertion is present within the STR repeat unit of Locus MS13 and this was seen in two of the five individuals sequenced in this project (Figure 3.12). Taylor [43] also reported a 19 bp indel in locus MS27 downstream of the complex repeat motif. This indel was present in four of the five individuals successfully sequenced for the region and absent in the only individuals representing subspecies *Morelia spilota imbricata* (figure 3.7). Taylor [43] likewise reported the deletion state to be unique to individuals within this subspecies range and absent from those outside it. This indel has potential to act as a diagnostic assay for carpet pythons originating from this region and to investigate alleged hybridisation between this subspecies and others (a topic of interest for South Australian enforcement).

Some characteristics of interest were seen in the flanking regions of the newly characterised loci. Locus MsF3F had an 'A' insertion in the flanking region upstream of the STR motif. This is most likely an artefact of MPS, as the insertion is not seen in any of the eleven sequences obtained by Sanger sequencing and alleles of 30087 electrophoretically separate from the other sequenced individuals at this locus by degrees of whole repeat units (see figure 3.9b). The sequence peaks representing the double 'A' in 55479 have poor morphology; it is likely that this is a sequencing artefact at the position of a single 'A', rather than sequential 'A' bases.

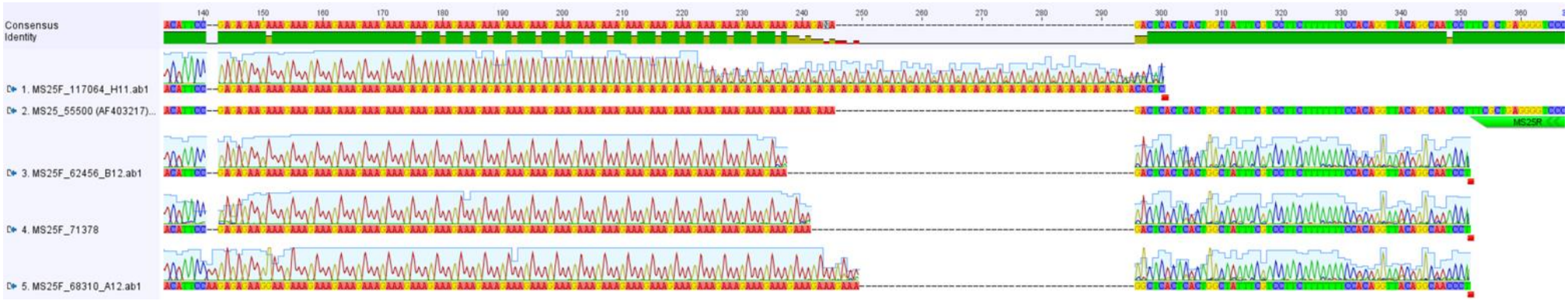


Figure 3.10: Alignment of locus MS25 sequences. While four individuals exhibit an MS25 sequence that behaves as a simple repeat unit, a fifth individual exhibits a sequence in which the simple repeat unit appears to have mutated into a di-nucleotide repeat unit.

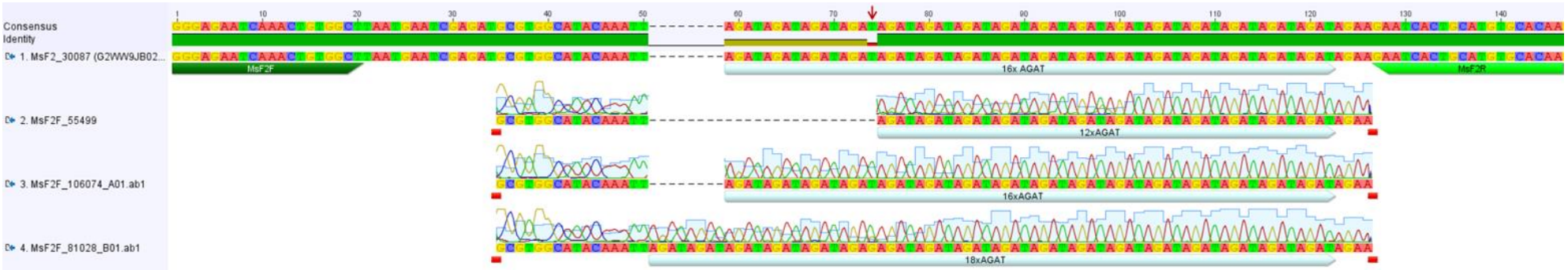


Figure 3.11: Locus MsF2 sequence alignment showing the SNP site in the STR repeat region. Allele 18 possessed by individual ABTC81028 from Queensland has a point mutation in the repeat motif, where a thymine base has mutated to a cytosine base. The nucleotide position is indicated by a red arrow above the consensus sequence. The other individuals sequenced are from Northern Territory and South Australia.

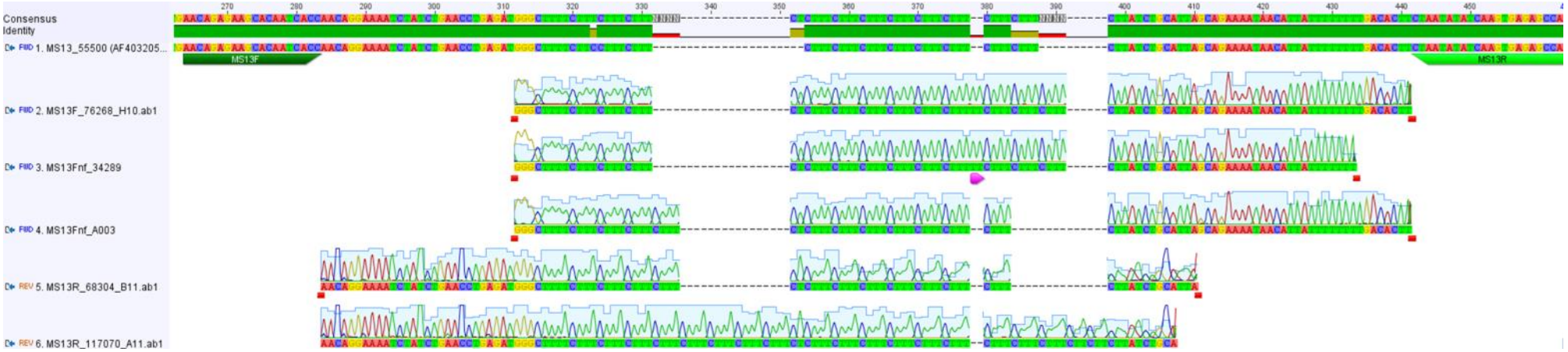


Figure 3.12: Sequence alignment at locus MS13 showing the indel present within the STR repeat region. The repeat region contains mixed repeat motifs, the majority of repeats being tetra-nucleotide in nature, interrupted by a single di-nucleotide motif, but individual 117070 also exhibits three terminal tetra-nucleotide repeat units.

As noted earlier, sequencing identified the initial MsF28 PCR product to contain two indels. An 11 bp indel occurs 114 bp downstream from the STR repeat region and 135 bp upstream from the original reverse primer. A second single adenine base deletion is also present 37 bp upstream from the reverse primer (see figure 3.8). One sequenced individual possessed both deletions, while all other individuals possessed neither. The reverse primer was redesigned to anneal 87 bp upstream, excluding the single base pair deletion. Any individual with the 11 bp deletion should appear as a .4 micro-variant.

Migration of the alleles at compound locus MsF31 was consistent with the underlying sequence repeat number, however inconsistent migration was observed in individual 66267 at MsF33. The repeat unit pattern suggests that this individual should exhibit a 25 allele peak, however the actual migration indicated a 21 allele was present. A large amount of flanking region between the STR region and the reverse primer has not sequenced and it's possible that an indel is present in this area within this individual to cause the migration discrepancy.

Homoplasy and Implications of genotyping by size migration vs sequencing

Due to their rapid mutation rate that makes them favourable for individualisation, STRs are recognised as subject to size homoplasy. Homoplasy is an evolutionary event (e.g. mutation, insertion, deletion) that is masked by a subsequent mutation changing the sequence back to its ancestral state. Therefore, two diverged sequences are identical by state, but not identical by descent, where one has undergone mutations and back mutations not present in the other [70]. The nature of STR mutations occurring under a Stepwise Mutation Model or an Infinite Allele Frequency Model is still debated; a Stepwise Mutation Model by nature would exhibit a higher incidence of homoplasy, but regardless of the model applied some level of homoplasy will inevitably plague rapidly mutating markers.

The size migration technique used to visualise alleles in forensic testing elevates the degree of masked divergence present within the results even further. Polymorphisms exist in underlying repeat motifs that are invisible to genotyping by size migration. Locus MsF16 illustrates this well (figure 3.13); all eleven sequenced alleles exhibit complete repeat motif patterns (and indeed of the 243 individuals genotyped in chapter 4, a single micro-variant allele is seen at 10.1), demonstrating that this marker mutates in a uniform manner and exhibits fairly low allelic diversity (13 alleles, 5th lowest of 24 loci) and polymorphism (H_o= 0.584, 5th lowest of 24 loci).

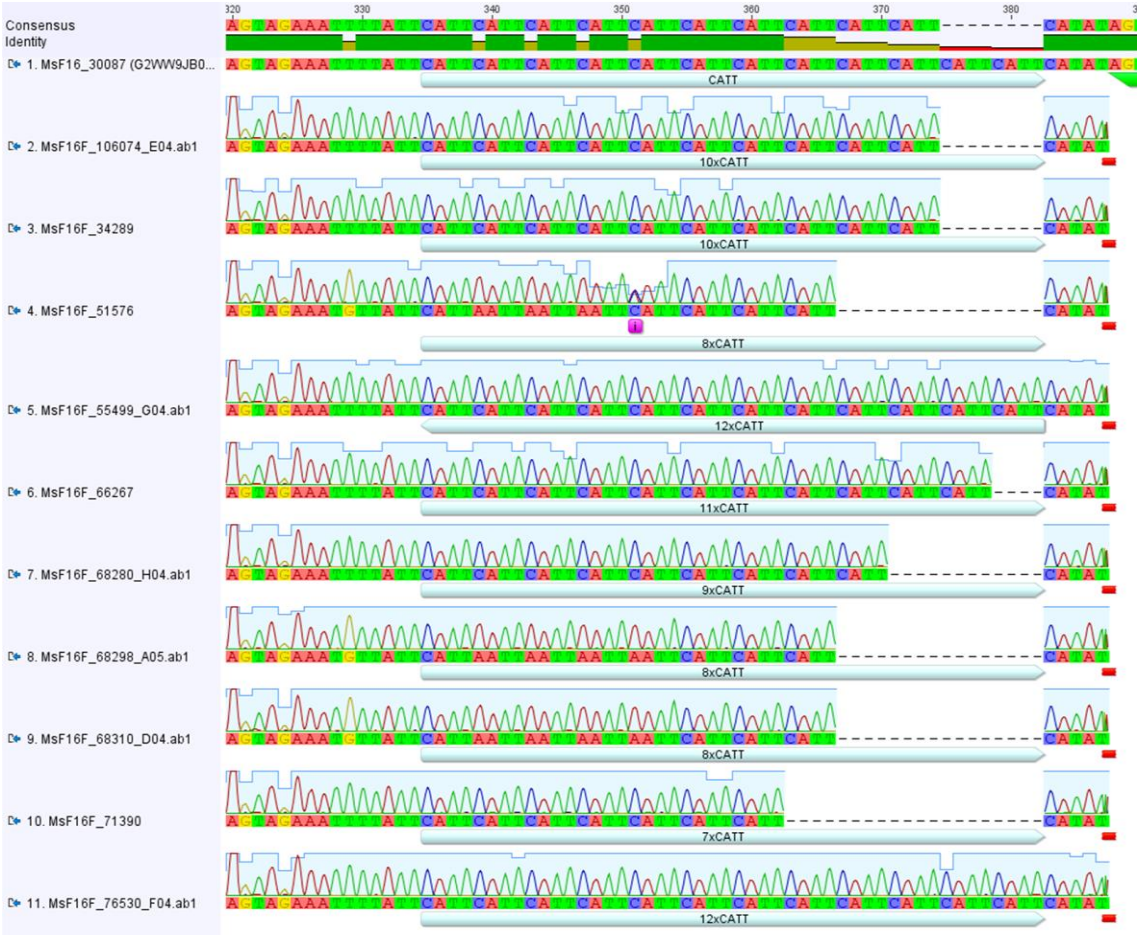


Figure 3.13: Most individuals contain a 'CATT' repeat motif for the entire repetitive tract. However, three individuals (ABTC51576, ABTC68298 and ABTC68310 contain an underlying mutation of three 'AATT' repeat units. These three individuals represent the only *Morelia spilota imbricata* to be sequenced at this locus.

Yet, sequencing reveals SNPs within the STR motifs in three individuals, hidden polymorphism that could potentially inform about population processes and assist assignment testing. This polymorphism appears from this initial sequencing to potentially be diagnostic of the *Morelia spilota imbricata* subspecies. Other polymorphisms undetected by size migration were evident within the STR repeat region (see figure 3.11; Figure 3.14) or the flanking region (Figure 3.15). These SNP sites do not appear to change the migration of the peak, but are potentially informative sequence characters lost through genotyping by size migration to collate genetic databases.

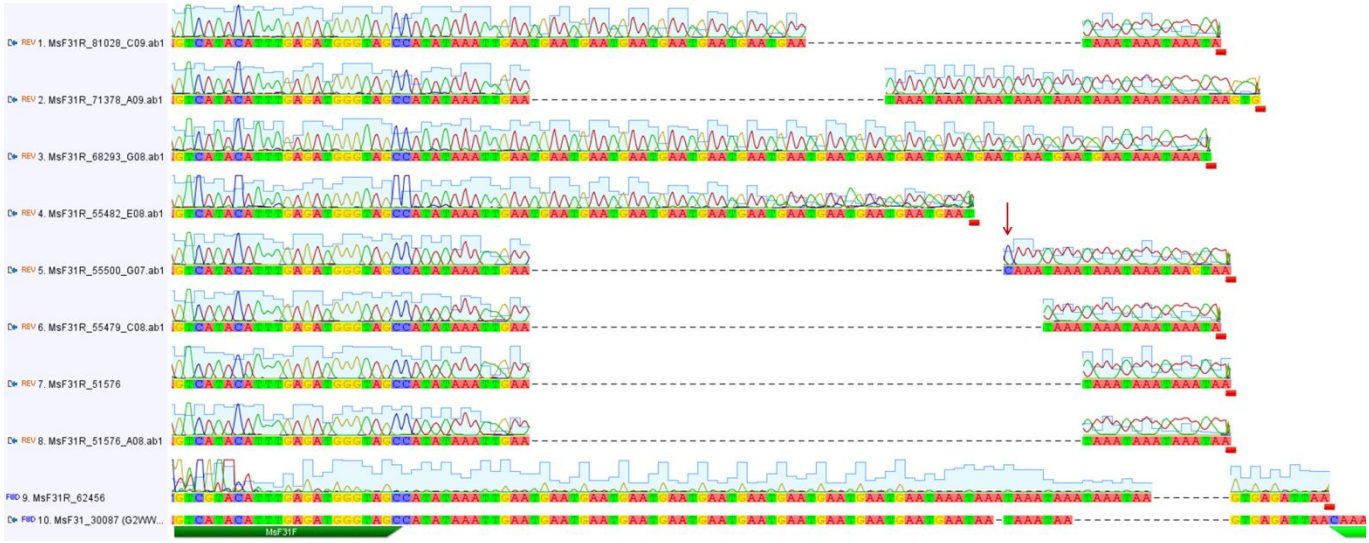


Figure 3.14: Compound STR locus MsF31 shows dual variability, as both of the underlying repeat motifs vary in number between individuals. One individual also possesses a point mutation, indicated by the red arrow.

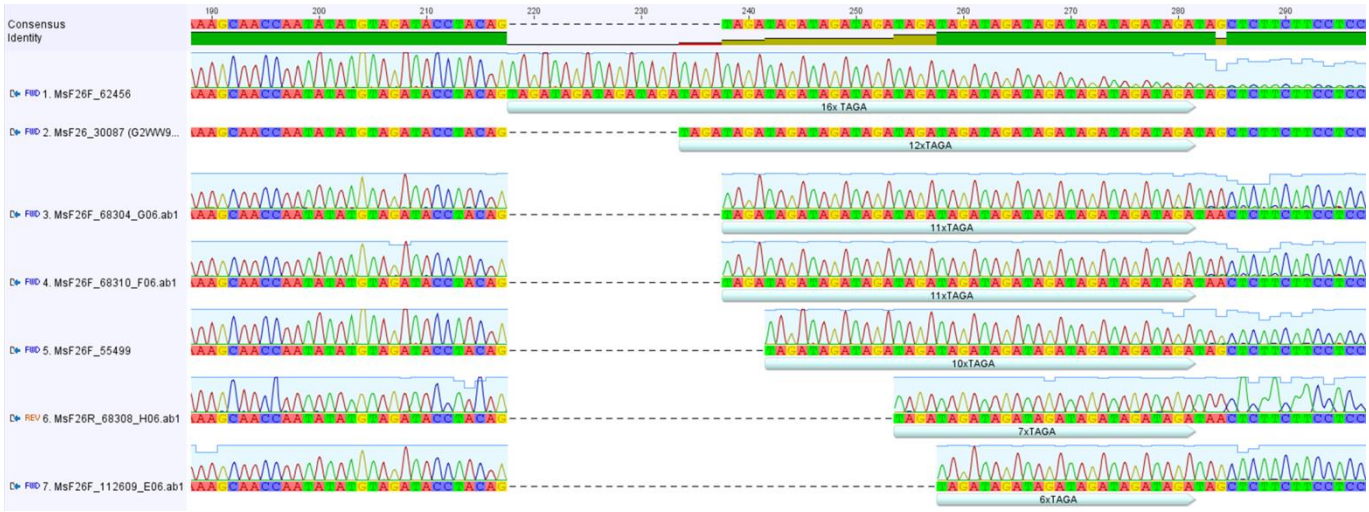


Figure 3.15: A polymorphism can be seen in the flanking region adjacent to the end of the repetitive tract. Three individuals possess 'AA', instead of 'AG'; these individuals are all *Morelia spilota imbricata* and this small number of sequences suggests that this could be a polymorphism specific to this subspecies.

Similar symptoms are seen within compound and complex STRs, where dual variable regions are likely to increase the prominence of hidden variability. A compound STR locus includes two hypervariable regions and this study has shown that both regions can exhibit size variability (e.g. MsF31). This dual polymorphism is demonstrated in the alignments of MsF31 and MsF33 sequences (figure 3.14, figure 3.16). When this is the case, there is actually much more variation existing at the sequence level than is captured by size migration alone. For instance, an MsF33 allele that has $(GACA)_8(TAGA)_{16}$ would be expected to migrate to the same absolute base pair size, and therefore exhibit the same allelic designation, as an allele with $(GACA)_6(TAGA)_{18}$, if the remainder of the sequence composition is identical between the two alleles. The implications of this hidden variability for forensic testing are addressed shortly.

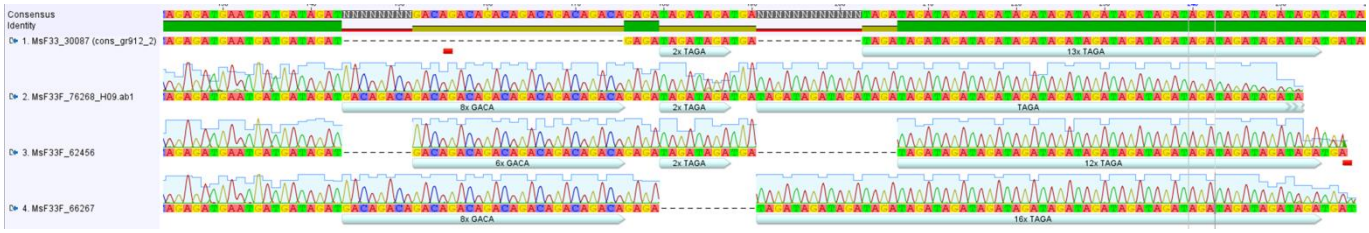


Figure 3.16: Similarly to MsF31, MsF33 also shows dual variability in the compound STR sites.

Underlying variability hinted by size migration

Sequence alignment is not required when genotyping via size migration. However, four instances were seen, all at loci with complex repeat motifs, where the size migration of peaks deviated from known alleles in the same size bin by fractions of a base pair. Over the entire dataset, these four peaks were seen to migrate inconsistently with all other alleles in relation to their placement within the allelic bins. The irregular migration is inferred to relate to difference in the underlying complex sequences affecting the rate of migration of these products. The products were considered unique to other alleles sizing within the same allele bin for the purpose of the reference database and given unique allelic designations for statistical analyses (see the "alternative alleles" in the final rows of table 2 in the embedded publication [54], appendix 3.1). None of these products were sequenced during this project to determine the cause of the anomalous migration, but this would be an interesting point to follow up. Three of the four fluorophores used are represented by these four loci, so the issue is not dye specific. Despite the inclusion of the snap freeze step and Hi-Di formamide to keep the genotyping products linear, it's possible that secondary structure formation based on the underlying repeat sequences has slightly inhibited the migration of these four products.

Other interesting polymorphisms

MsF14 – notes to supplement publication in section 3.5.13

During database compilation, locus MsF14 was observed to exhibit incomplete allelic amplification, correlated with population of origin. Closer examination revealed an STR motif to be present in the reverse primer (see embedded publication [64], appendix 3.4). Further characterisation of this locus is recommended prior to application to casework.

Although beyond the scope of this thesis, this locus would be very interesting to examine further in relation to population genetic processes. Complete fixation for allele 4 was exhibited by the Southwestern carpet python subspecies. This is a penta-nucleotide repeat motif and a .4 micro-variant was evident in many individuals throughout the allele range. Further investigation is warranted to assess whether this polymorphism is population correlated, in which case its inclusion could add strength to statistical calculations.

MsF31 sequencing anomaly

Eight individuals were sequenced in both forward and reverse direction for MsF31 and the addition of ABTC30087 from the MPS output totals sequence data for nine individuals at this locus. Sequence results from two individuals exhibited a very strange phenomenon. Complementary versions of the STR locus appear to have become concatenated during the sequencing reaction with a region of 13 unique bases linking them (figure 3.17). The two individuals for which this has occurred have a number of divergent polymorphisms within this region and one shares homologous sequences with other sequenced individuals, so it is unclear why it has occurred in these two individuals and none of the others. The small region flanking the duplicated version of the STR is identical between these two individuals and the compound repeat motif mirrors the complementary compound repeat motif in each.

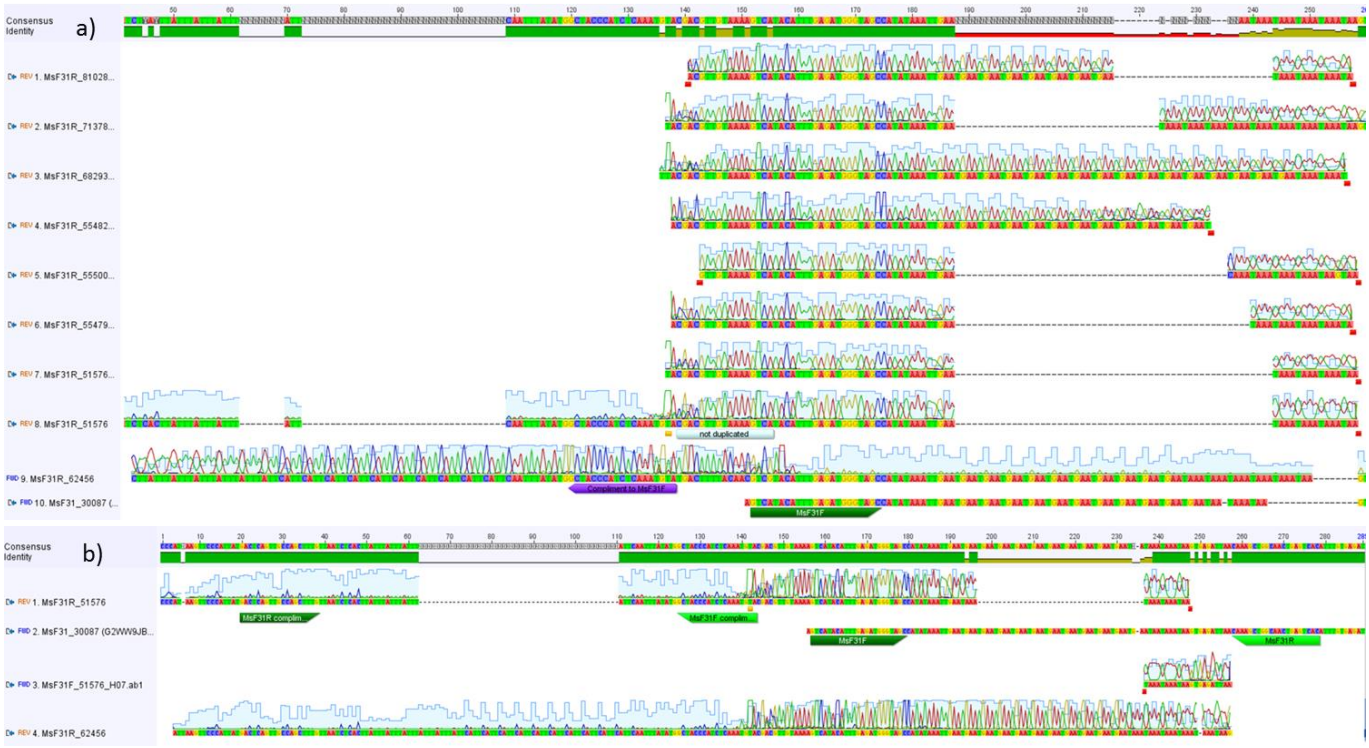


Figure 3.17: An anomaly seen in the sequencing of MsF31, the origins of which is unknown; a) two individuals show repetitive regions in sequence upstream of the target STR locus, b) closer inspection reveals that these two sequences are actually the complementary sequence of the repetitive region and a short length of flanking region from the same individual that have somehow become concatenated during sequencing. Note: MsF31R_51576 was trimmed of sequence GTGAGAT adjacent to the reverse primer sequence, so the terminal TAA shown does actually correspond with terminal region of STR repeat region.

The entire 133 bp fragment obtained from the MPS run is shown in this figure, so it's not possible to say whether there is an area partially complementary to forward primer slightly upstream of the priming site. This phenomenon could potentially be caused by a translocation of the 3' end of the forward priming site. However, if this were the case, then the other sequenced individuals could be expected to show the same duplication pattern in their sequence data, which they do not.

3.5.7 Sample genotyping

A total of 304 carpet pythons were profiled for 27 loci. Nine individuals exhibited no amplification product and 36 individuals exhibited alleles at 19 loci or less.

During electropherogram scrutiny, four individuals each exhibited one allele that displayed questionable migration patterns. A single allele at a single locus in each individual migrated noticeably closer to the boundary between two allele bins. Comparison of each allele peak and the associated allelic ladder to individuals displaying an identical allele peak with associated allelic ladder in a single Genemapper project showed obvious discrepancy between in allele migration patterns of alleles that should be designated the same call. All four alleles were unique to each other in allele call and locus, however all four loci represented complex repeat motifs. These abnormal migration patterns are believed to be caused by differences in the underlying sequence structure and these products are therefore treated as independent alleles in subsequent data analyses.

3.5.8 Reference database construction and quality control

Of the 27 loci genotyped during reference database construction, three loci MS2, MS16 and MsF14 were excluded from the reference database due quality assurance issues. Complex locus MS2 showed inconsistent migration patterns and very poor peak morphology, believed to be due to stretches of dinucleotide repeats in the underlying sequence of some alleles (see figure 3 in embedded publication [54], appendix 3.1). Many peaks exhibited split morphology. These characteristics precluded reliable allele calling to single base pair resolution. Complex locus MS16 displayed unacceptable amounts of missing data and due to unreliable typing success. Simple penta-nucleotide repeat motif locus MsF14 exhibited unusual population correlated amplification efficiency patterns, which was discovered to be due to an STR repeat motif in the reverse primer sequence (see [64], section 3.5.13 for further details).

A reference genotype database was constructed using 24 loci from 249 individual samples (see chapter 4 for further details). Fifteen individuals were excluded due to quality reasons and these are specified in Appendix 3.5. The choice of data included in the final reference dataset is a compromise between maximising the database size and minimising the amount of missing data present in partial profiles. A threshold of 22+ complete loci was chosen as the most appropriate final reference dataset. The database includes 52 and 12 individuals with incomplete data at one and two loci respectively. The resulting allele frequency database is presented in Table 2 of the embedded publication [54], appendix 3.1.

3.5.9 STR database characteristics

A total of 752 alleles were observed across 24 loci screened in 249 individuals. Allelic diversity was very high across all loci genotyped, ranging between 10 alleles and 88 alleles per locus (table 3.1). Observed heterozygosity ranged between 0.397 and 0.865. Mean observed heterozygosity was 0.671 (s.d. +/- 0.026) and allelic diversity was 0.865 (s.d. +/- 0.014), indicating the high potential of the marker set to be powerful markers for discrimination between individuals. The most common and rarest genotype estimates were 1 in 10^{29} and 1 in 10^{48}. The probability of obtaining an identical genotype from two individuals chosen at random from the population, assuming that the individuals

are siblings (Probability of Identity of Siblings, PI_{SIBS}) was 1 x 10^{-12}. These results indicate that the multiplex marker set exhibits a very high power of discrimination.

	Allele #:	10-19	20-29	30-39	40-49	50-59	60-69	>70	
Locus	Alleles	H_o	PIC	Motif type	Locus	Alleles	H_o	PIC	Motif type
MS17	77	0.846	0.9759	Complex	MsF31	28	0.506	0.8451	Compound
MS3	88	0.847	0.9704	Complex	MsF26	19	0.607	0.8427	Pure
MS25	67	0.855	0.9683	Complex	MS9	17	0.746	0.8309	Pure
MsF33	68	0.865	0.9553	Complex	MS4h	21	0.81	0.8257	Pure
MsF4	42	0.738	0.9258	Pure	MsF16	13	0.584	0.8169	Pure
MsF17	36	0.663	0.9012	Pure	MsF27	29	0.68	0.8083	Pure
MS13	39	0.61	0.8945	Complex	MsF28	22	0.49	0.8083	Pure
MS27s	33	0.542	0.8919	Complex	MsF9	11	0.703	0.8075	Pure
MsF3	14	0.586	0.8705	Pure	MsF5	21	0.618	0.7631	Pure
MsF15	29	0.682	0.8664	Pure	MsF22	12	0.609	0.7318	Pure
MsF2	24	0.759	0.8518	Pure	MsF8	10	0.597	0.7223	Pure
MsF18	21	0.771	0.8455	Pure	MsF24	11	0.397	0.6741	Pure

Table 3.1: Population genetic characteristics of the 24 loci included in the final reference database, ordered by Polymorphic Information Content. H_o = observed heterozygosity. The heat map indicates the loci with the highest numbers of alleles.

locus	P-val	S.E.
MsF5s	0	0
MS25	0	0
MsF33	0	0
MsF31	0	0
MsF9	0	0
MsF16	0	0
MsF3	0	0
MsF17	0	0
MsF18	0.0008	0.0001
MsF28si	0	0
MsF22	0	0
MS3	0	0
MS4h	0.0004	0.0001
MsF27s	0	0
MsF8	0	0
MsF26	0	0
MsF4	0	0
MS27	0	0
MsF24	0	0
MS9	0	0
MsF2	0	0
MS17	0	0
MsF15	0	0
MS13	0	0
Chi2	Infinity	
Df	48	
Prob	High. Sign.	

Table 3.2: Degree of deviation from Hardy-Weinberg Equilibrium is shown at each of the 24 STR loci calculated using the final reference database of 249 individuals. The 24 loci all showed highly significant deviation. P-value <0.05 equals significant deviation; S.E. = standard error; Df = degrees of freedom.

The entire 249 reference dataset exhibited highly significant deviation from Hardy-Weinberg equilibrium when considered as a single population. Hardy-Weinberg Equilibrium is a simplified principle based on the underlying assumptions of infinite population size, random mating, non-overlapping generations within the population, no migration, no mutation and no selection [71], all of which are violated by natural populations. For this reason, it is acceptable to see a low level of deviation from Hardy-Weinberg equilibrium in natural populations. This dataset however significantly deviates at every locus (table 3.2), indicating that there is potentially population sub-structure present. The implications of this for forensic testing will be discussed later. Accordingly, the dataset also did not display linkage equilibrium; loci that do not conform to HWE will not exhibit linkage equilibrium.

Sub-structuring must be accounted for with the application of a theta correction (θ) when calculating forensic statistics. The PI_{SIBS} is calculated here because the degree of population substructure present within this dataset is yet to be established. The kinship factor of 0.5 applied to a sibling relationship is a much more conservative figure than most co-ancestry coefficient corrections for substructure seen in wild populations (see table 4.1 in section 4.3.3). This calculation is expected to give a more conservative figure than a calculation involving θ based on empirically measured population substructure.

3.6 Further Discussion

Four objectives were specified at the outset of this chapter and further considerations within each of these areas are elaborated upon in this section.

3.6.1 To identify a large number (>15) of highly polymorphic tetra- , penta- and hexa-nucleotide STR loci specific to the carpet python.

The need for many highly polymorphic loci was evident from previous work conducted on the population genetics of carpet pythons [43]. The conduct to achieve this first aim was fairly straight forward and warrants no further discussion.

3.6.2 To combine the novel markers into a small number of multiplex assays and optimise the assays for reliable genotyping of all loci.

3.6.2.1 Assay design

Routine forensic STR profiling in human DNA investigations involves the application of commercial STR kits comprised of primer mixes and buffers that have been highly optimised through ongoing development over at least the last decade and a half. Wildlife forensic scientists rarely have the luxury to tapping into commercial endeavours in STR assay design, rather primer mixes have to be developed and optimised in house. The wildlife forensic scientist requires a much better appreciation of primer design and multiplex primer interactions than is ever warranted in the human forensic genetics field.

Simplex primer design was discussed in sections 2.1 and 2.6.1. This discussion builds on that section, with the added complexity that combining an increasing number of primers into a single reaction tube complicates the process of creating a complete, robust and easily interpretable result. The advantages are clear though, particularly when working with traces of potentially degraded forensic evidence. Multiplex analysis maximises the information that can be obtained with minimal DNA template input (in an application where little template is often available) and minimises the number of steps, correspondingly minimising the opportunity for sample mix up and exogenous DNA contamination. These outcomes are well worth the initial effort of careful design and comprehensive optimisation.

There are two important preparatory phases of multiplex assay design that must be carefully devised to allow interpretable results: multiplex primer design and the visual layout of the resulting DNA profile.

Multiplex PCR primer design

The consideration of primer T_m is fundamental to multiplex primer design. Chapter 2 introduced the concept that simplex primer pairs must possess the same requirements for PCR conditions in order create the target amplification product. Likewise, all primers in a multiplex reaction must have similar melting temperatures to produce results using the same reaction conditions. Identifying and designing primers manually using T_m hand calculations is a laborious and daunting task, but fortunately a suite of primer design software programs have been developed to assist this stage (perhaps the most well-known is Primer3,[48]). This function is built into the QDD software that was

utilised for this project, so that groups of candidate primers were returned, ranked by their utility in relation to multiplexing and the task of choosing the best primers at each locus to be combined into a single tube reaction was made somewhat easier.

The other essential consideration during multiplex primer design is the potential for interactions between primers; multiplexing primers in a single reaction greatly increases the risk of primer interactions outcompeting target amplification. This would either lead to formation of hairpins if primers annealed with themselves, or primer dimer between two primers containing complementary sequences (even if as small as a few base pairs), causing failed amplification for both loci involved. All possible combinations of primer complementarity must be considered during the multiplex planning stage [72]. The computer software http://www.thermoscientificbio.com/webtools/multipleprimer/ was very helpful for determining *in silico* interactions prior to primer synthesis.

Primer-primer interactions are favoured when numerous primers are present at high concentrations in a reaction. Chamberlain & Chamberlain [17] observe the increase of non-specific amplification artefacts as primer concentrations are increased. Reduction of primer concentrations was an important stage of multiplex optimisation rounds reported herein. The PCR annealing temperature was kept quite high (60 °C) in order to maintain annealing stringency; decreased annealing temperature allows non-specific binding to occur between sequences that contain mismatches in sequence, a desirable trait for universal primers, but not for a highly specific STR assay.

Other considerations discussed in 2.6.1 still stand but do not warrant further comment in the context of multiplex development.

Assay layout to achieve robust interpretation of results

John Butler [18] provides excellent preparatory guidelines for STR multiplex assay design, which are a must read for anyone undertaking the challenge. The visualisation of the resulting fragments is easiest considered on a two dimensional plane. This concept is touched on by Chamberlain & Chamberlain [17], but the Butler chapter provides a visual guide to planning the multiplex layout (figure 3.2) - which upon reflection is intrinsically obvious when looking at any multiplex profile. Use of multiple fluorophores allows multiplexes of loci exhibiting overlapping size ranges, however if loci display a narrow allelic window (more often seen in this project from simple sequence motifs), then multiple loci can be tagged using the same fluorophore and visualised in the same dye lane on the resulting electropherogram.

The process of planning the multiplex layout during this project was iterative (like many other aspects of the multiplex design and application). All primer options were mapped and visualised on a 2D plane representing size. Loci were prioritised by desirable attributes such as motif type and potential repeat length (colours). Once the primers were shortlisted, ordered with MRT tags and screened in a small number of individuals, visual charts of locus sizes were composed in order to assign loci with non-overlapping allele ranges the same fluorophore (figure 3.3). It became apparent quickly from the large allelic windows in carpet pythons that development of two to three multiplex assays would be required to interrogate the desired two dozen or so STR loci. Shades were assigned to loci to distinguish between the independent multiplex assays and alternative primers were chosen

in some loci to shift their allelic window and avoid overlap or increase the boundaries between locus windows in a dye channel.

Assay redesign

The likely need for assay redesign was realised early in the screening process, when a multitude of newly appearing alleles were expanding the range of each locus. The propensity of snakes to have highly polymorphic STR loci exhibiting large allele size ranges was gleaned through the work of Taylor [43] and the compound and complex loci developed in this project continued this trend. This also provided an opportunity to reassess and switch two of the chosen quality markers that were showing less than ideal characteristics. The frustrating aspect of the multiplex redesign was the allelic ladder that had already been designed for the original layout; three more allelic ladders needed to be developed based on the new layout.

The process of redesign can be seen by comparing charts b), c) and d) in figure 3.6. Section 3.5.3 presented the redesign steps between the first multiplex layout to be optimised and the new multiplex layout. Very little primer redesign was required at this point, due to the time invested during the initial primer design to identify and solve problems that might potentially arise. A redesign step also took place earlier in the project, prior to the initial optimisation. Loci MsF33F, MS3F and MsF5F were relabelled with FAM fluorophores to improve locus organisation with the three multiplex assays.

Another alternative approach for consideration in future studies, which might have proven more efficient than the redesign of the multiplex assays and associated allelic ladders, is the pooling of the 43 initially screened individuals into a single tube to be used as template for simplex reactions at each of the 27 loci, following the approach demonstrated in figure 3.5. While these samples could be pooled into a single run, I would argue that running them separately, or at least pooling only one locus per fluorescent label to cut the number of runs to a quarter, would avoid confusion of assigning alleles to loci where allelic windows of loci are seen to overlap.

However, one foreseeable issue to arise from including so many contributors in a single reaction is the imbalance in peak heights if allele frequencies vary dramatically at a locus, as was witnessed at many loci in this study. Common alleles would become incredibly overloaded, while rare alleles would drop out if the run were rerun at a dilution to overcome the gross overloading of fluorescent signal. Provided the rare alleles were not at the extremes of the allelic window (which in many cases alleles in this study were) then this approach might still give an indication of allelic window size and potential overlap between loci. If a study has access to ample amounts of DNA template from a large number of individuals, then this experiment might be a worthwhile alternative, precluding or minimising the need for extensive multiplex redesign due to lengthening allelic windows.

3.6.2.2 Incorrectly sizing peaks

Due to the high number of micro-variants present at many loci within the designed multiplex assays, the ability to genotype to single base pair resolution is crucial for robust allele calling. Initial runs exhibited inconsistent sizing of peaks occurring between 200-300 base pairs in size. This was believed to be caused by the sparse inclusion of sizing peaks representing this size range in the Liz500 size standard (peaks included only at 200, 250, 300 and 340 base pair sizes). The Liz500

internal size standard product insert was subsequently found to describe temperature related inconsistencies in the sizing of the 250 base pair peak. The Local Southern size calling method determines lines of best fit based on the two larger and one smaller size standard peaks in relation to each unknown allele followed by the two smaller and one larger size standard peaks and then takes the average of these to size the unknown peak. This means that the 250 base pair peak would be included in the sizing algorithm for any peak falling between 160-340 base pairs, potentially introducing inaccuracies into the sizing of these peaks across runs. Once this issue was identified, adoption of the Liz600 size standard enabled accurate sizing across runs.

3.6.2.3 Advantages and disadvantages of complex and compound primers

Higher allelic diversity was seen in the loci containing complex motifs than most of those of simple repeat motif type (see table 3.1). Of the six loci with complex motifs, four exhibited the highest allele counts and Polymorphic Information Content in the dataset and all six were in the top eight highest PIC scores. The lowest PIC value seen in a compound motif locus was higher than values at the majority of loci containing simple motifs. These loci are clearly useful for increasing the discrimination power of an STR assay.

However, complex STRs are not without their challenges. Two of the complex motif loci were removed from the final reference dataset due to quality issues relating to genotyping inconsistencies caused by unreliable morphology and migration. Underlying stretches of dinucleotide repeats were seen frequently at some of the loci, causing shouldered, inconsistent morphology of peaks and stutter patterns characteristic of dinucleotide motifs. Locus MS2 suffered poor peak morphology, evidently resulting from the di-nucleotide tracts at the sequence level (see figure 3 of embedded publication [54], appendix 3.1). This confounded the deliberate target of tetra-nucleotide repeat loci over di-nucleotide repeat loci to maintain the stringent quality standards required of forensic assays. The underlying sequences vary to the extent that attempting sequence alignments is at times futile. The alignments of MS17 (Figure 3.18) and MS27 (Figure 3.19) show that complex sites do not adhere to a single motif pattern, or even motif repeat size and sequence, between individuals at a locus.

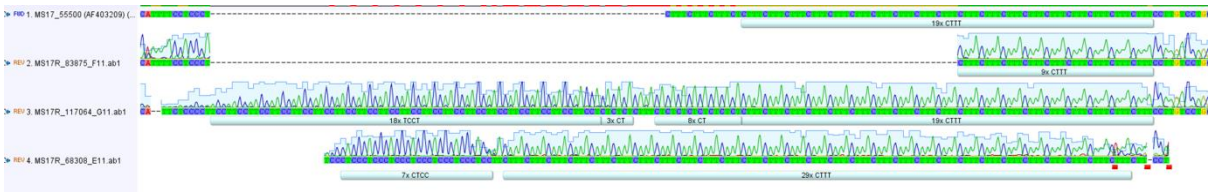


Figure 3.18: Homoplasy present in the repeat motif sequence of locus MS17. The underlying sequence displays at least three variations of tetra-nucleotide repeat, as well as tracts of di-nucleotide repeats in one of the individual sequenced.

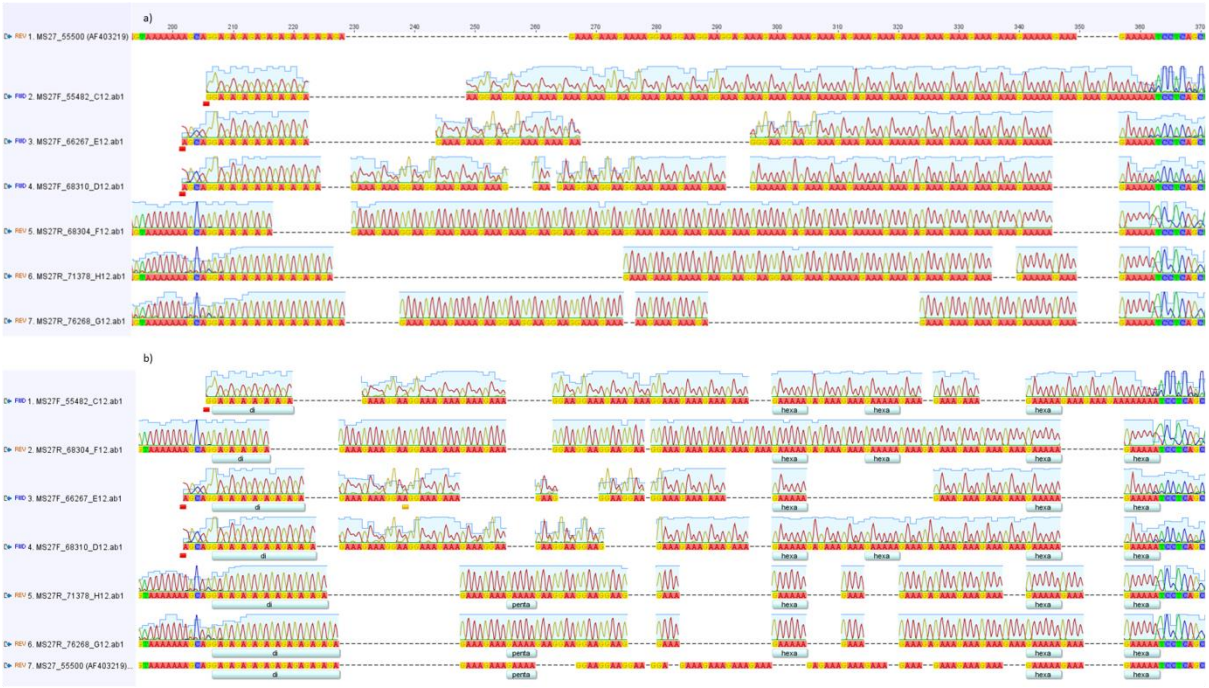


Figure 3.19: Complex underlying repeat regions at locus MS27. a) sequences upon initial alignment, and b) further partitioned to display a combination of di-, tetra-, penta- and hexa-nucleotide repeat motifs. Some of the hexa-nucleotide motifs are more likely to be tetra-nucleotide motifs with AA insertion at end, as these are not seen to repeat.

Allele nomenclature difficulties at complex loci

The process of naming alleles corresponding to the underlying repeat number was alluded to earlier. All new loci (MsF) adhered to the recommended nomenclature. However at the complex loci for which the underlying sequence information content diverged so much, the process is not as straight forward as at the MsF loci. Of the loci classed complex, only MS25 displayed behaviour of a simple STR motif in three of the four sequenced individuals and the fourth individual mutated into a long dinucleotide tract (figure 3.10). Underlying repeat patterns varied markedly and locus MS9 was the only pre-existing (MS) locus where the allele name represented the actual underlying repeat number. Allele designation at the other complex loci was based on an arbitrary fragment size rather than the actual number of underlying repeat motifs for the sake of database collation because it is clear that most alleles at the complex STR loci have very different underlying repeat patterns. Thirty-seven alleles were sequenced across six complex motif loci and not one allele was shared a similar underlying pattern within a locus. The consequence is inflated allele frequencies for the 'common loci' (common sizes of fragments), which is a conservative approach in favour of the defence.

Information loss from genotyping by size separation

Asides the high degree of homoplasy already present in the rapidly mutating STR loci, the results presented at simple STR loci (MsF2, MsF16, MsF26), compound loci (MsF31, MsF33) and all complex loci asides Ms4, Ms9 and Ms25 demonstrate that there is much polymorphism in the data that is lost through visualisation by size migration. The implication is that allele frequencies are likely overestimated in the database, artificially deflating the match probability, which does work in the defendant's favour. On the flipside, a defendant might actually be excluded as a contributor to a

DNA profile if sequencing of their underlying allele repeat motif reveals the DNA sequence to differ from the sequence of the evidentiary profile allele at that locus. The benefit of allele sequencing in criminal casework is clear and has recently become possible through the application of MPS. Implementation of this technology by forensic laboratories remains hampered by numerous issues, obvious hurdles being the prohibitive cost of current MPS technology, mass data storage requirements which add expense and the difficulty of compatibility with existing searchable genotype databases that comprise numerical allele names. The system of allele nomenclature, as discussed above, would need to be revised to allow for multiple sequence polymorphisms underlying the same fragment size, as well as additional information about SNPs and Indels, all the subject of recent discussion within the international forensic science community [73].

3.6.2.4 Quality aspects of forensic marker set must be maintained

From the outset of the project, the STR profiling kit was designed in a forensic context for ultimate application in criminal investigations involving carpet pythons. For this reason, building stringent quality controls into the test took priority. The large number of loci included in the test for the purpose of maximising discriminating power dictated that genotyping would have to be conducted using multiple assays. The amalgamation of multiple profiles into one final genotype for an individual was recognised as a possible point of quality failure, if through error the combined genotypes should happen not to originate from the same individual. Three redundant quality markers were built into each multiplex assay to flag non-concordance between the three assays for an individual. This should allow for any case of template addition from an incorrect individual to be identified and investigated.

The effectiveness of this quality control measure was witnessed firsthand during sample genotyping. Two screening plates of 94 samples plus two ladders each were screened using the optimised multiplex assays. The Light PCR plates were prepared first, followed by the Dark PCR plates and the Medium PCR plates prepared last. When merging the genotypes of samples included on these screening plates, four samples included on the Medium screening plate 1 showed discrepant genotypes at the QA markers from these sample samples genotyped using the Light and Dark multiplex assays. Samples in the same four wells of Medium screening plate 2 also contradicted genotypes obtained from each Light and Dark screening plate 2. However, the quality markers in the samples on Medium screening plate 2 matched the genotypes obtained by each Light and Dark screening plate 1. The quality marker genotypes of all other screened samples were checked and excluded as matching all six alleles in each of these four samples, demonstrating that the genotypes must have come from the individuals that were included in the same wells on screening plate 1 of the Light and Dark assays. Upon reflection, it was realised that when the Medium multiplex assays were set up, four individuals were substituted onto screening plate 1 and the four individuals replaced on screening plate 1 were shifted onto screening plate 2. Note of this was not made at the time (forgotten). It did however act as a useful demonstration of the incorporated quality markers to help identify and investigate an incident of sample mix up.

The quality markers also assisted investigation into an incident of sample contamination. A PCR plate was divided in thirds and set up to screen 30 samples and two ladders with each of the multiplex assays (columns 1-4 Light, 5-8 Medium and 9-12 Dark). Sample ABTC81320 in wells E4, E8 and E12 gave a mixed DNA profile resulting from two contributors from the Light and Medium assays, but a

single source profile from the Dark assay. The quality marker genotypes of all other screened samples were scrutinised and matching alleles found only in sample ABTC81320. This sample was designated in wells E3, E7 and E11 of the same plate. Well E11 gave a full profile, but wells E3 and E7 showed no amplification product, indicating that the DNA template has been added by accident to wells E4 and E8.

Without the information provided by the redundant quality markers, investigating the cause of these errors would have been much more difficult and in the case of the sample mix up, the incident might otherwise have been overlooked. They have been demonstrated to be a valuable component of the multiplex assays to assist in detection and investigation of potential errors during the genotyping process.

Another quality consideration worthy of brief discussion is the removal of a duplicate genotype from the database, mentioned in section 3.4.11. While two possible explanations are provided, the second scenario of sample mix-up during collection or curation prior to involvement in this project is far more likely given the origin of the samples used in this study. This is not intended in any way to imply incompetence in museum processes, but simply to point out that the forensic chain-of-evidence procedures are highly stringent to safeguard against the simple reality that is human error. The samples used in this project were not collected using a system of complete traceability, but rather opportunistically at times with no verification of collection records by a second person at the time of sampling. It is quite possible that mislabelling or sample mix up has occurred at some point and the two DNA profiles originate from the same snake (and I suggest more likely than identical twins). This demonstrates the importance of sourcing vouchered samples for forensic reference DNA databases whenever possible.

3.6.3 To perform validation studies on the multiplex assays so as to prepare these tests for direct application to casework.

Due to an employment opportunity arising in Scotland, the laboratory component of this project had to be concluded just after the time of screening for the reference database. Preliminary validation testing was conducted, but an ideal and thorough validation was not possible in the short timeframe. This section outlines the steps that would have been taken to conduct a much more comprehensive validation of this marker set.

Ideally, all stages of validation would have been conducted using the three optimised multiplexes as three primer mixes, whereas some stages had to include results from earlier simplex reactions. Data analysis of the validation tests also falls short of the ideal completion at times because all runs were saved as Genemapper ID .ser files. In Scotland, only Genemapper software is accessible, which does not recognise .ser files. It also does not recognise the allelic ladder designed for Genemapper ID (will not calibrate bins to centre of peaks and automatically assign alleles), so reanalysis of the raw data has been a further challenge.

3.6.3.1 Assay robustness

SWGDAM [45] section 3.9.2 specifies that PCR amplification conditions should be optimised, including primer concentrations, mastermix concentrations including DNA polymerase and magnesium chloride, and cycling parameters. In other sections of the PhD where simplex

amplifications were produced, the amount of DNA polymerase added, size of reaction volume and concentration of magnesium chloride had been altered for experimentation. In fact, all simplex reactions in chapter two were amplified in either 20 µL or 10 µL where template was not required for sequencing reactions, to maximise cost effectiveness. The polymerase enzyme manufacturer's routinely recommend that reactions are performed in 50 µL reactions, which is evidently excessive and unnecessary for a successful reaction. Where Mango Taq polymerase (Bioline) was used, addition of only 1.25 units was found to be sufficient for simplex amplification, despite the manufacturer recommending five units.

Different enzymes were trialled for the STR multiplexing. Amplitaq Gold showed only partial amplification of the included loci. For the purpose of multiplexing, the Qiagen Multiplex Kit was found to produce excellent multiplex amplification. This kit contains pre-mixed concentrations of enzyme and magnesium chloride that have been optimised by the manufacturer. Addition of further volumes of these components was not experimented with because the complex amplification was working well and alteration would likely lead to suboptimal results and waste expensive reagents. With infinite resources, subsequent volumes of polymerase and magnesium chloride would have been added to the mastermix. The range around the optimal concentrations, within which the reaction maintains robustness would be reported (see [74]ref for an example of this). Cycling parameters including annealing temperature and cycle number were optimised and primer concentrations followed to fine tune the multiplex balance once the overall cycling parameters had been decided.

3.6.3.2 Repeatability

This was conducted using sample 62456 from the first round and second round (new configuration multiplex) of optimisation screening tests. With more time, this study would have involved screening four individuals using the same PCR primer mix replicate times, over multiple days and including set up and genotyping by a different lab member. The PCR reactions would have been conducted on the different PCR platforms available in the Flinders laboratory. The anomalous run that was removed due to a suspected internal run problem demonstrates the importance of including an internal positive control with every amplification, the inclusion of which is best practice in all forensic casework. Should the same anomaly arise, this known genotype would exhibit incorrect allele designations, flagging the issue. Any lab applying this test to casework routinely include a positive control of known genotype.

3.6.3.3 Accuracy and Precision

The assessment of accuracy evaluates the ability of the test to provide a result close to the true result, whereas precision involves mutual agreement between replicates, even though the result might not be the true result [45]. The repeatability results assessed genotyping precision, but genotyping accuracy would also have been characterised if time had permitted. The inclusion of allelic ladders with each run and sequencing of alleles facilitate this exercise, but time and the issue involving .ser data file inaccessibility mentioned earlier preclude this from being completed within this study.

3.6.3.4 Reproducibility

The plate run at FSSA and rerun at Flinders University does not constitute an ideal reproducibility study because, despite being run on separate instruments in separate laboratories, it was still set up by the same operator. A more extensive reproducibility study would have involved asking a collaborating lab (the Australian Museum Genomics Facility, for instance) to genotype a small number of my python samples. I would have prepared aliquots of: multiplex mastermix, python DNA from half a dozen good quality individuals diluted to 1 ng/µL, the three allelic ladders and primer mixes including all eleven primer pairs in the appropriate concentrations for each of the three multiplex assays. If they use Genemapper analysis software, then the panel and bin files developed for this assay would also have been provided. These would have been mailed to the collaborating lab with the request that they genotype the samples using their own instrumentation and return the 24 locus genotypes obtained.

3.6.3.5 Sensitivity

This was one aspect of the validation testing that was able to be completed in the Flinders laboratory. All required concentrations were tested using all three multiplexes, allowing the limit of detection to be assessed for each assay. The only desired extension to this validation step would be to run the results using a 3500xl genetic analyser. This later model instrument is more sensitive and would therefore be expected to reveal a lower threshold of detection. Each laboratory implementing this multiplex assay would of course be required to conduct a verification study to assess the limit of detection applicable to their instrumentation.

3.6.3.6 Specificity

The specificity study was conducted using MRT reactions in singleplex. A more thorough specificity study would have involved amplification of each species using the three finalised multiplex assays. Amplification would ideally have been further expanded upon those included to the individuals representing the ten available species of the *Morelia* genus This was not possible, however given the nature of genetic markers demonstrated herein to decrease in efficacy in species of increasing genetic divergence, it can similarly be expected that markers will yield improved results for species more closely related to the target species of development. The broad headed snake (*Hoplocephalus bungaroides*) might also have been worthwhile including, as it is another snake species of interest to Australian law enforcers. There would be no utility to include further reptile species asides snakes if poor amplification results were seen in snake species, as the attribute of non-coding STR markers to be highly variable within a species also lends the counter attribute that taxa do not have to diverge very far in genetic relatedness before the nuclear complement of the STR primer sequences no longer exist.

3.6.3.7 Selectivity

This study obtained DNA profiles from a number of different sample origins, demonstrating that these should all be viable sample types for testing. Ideally, selectivity would also examine the ability of different sample types taken from the same individual to produce concordant DNA profiles and demonstrate that the DNA profile is not affected by sample source.

3.6.3.8 Characterisation of loci

Guideline 3.1 of SWGDAM [45] states that all loci included in an STR test should be characterised and documented. Characterisation includes mapping to determine the chromosomal location of each locus (which also aids determination of whether loci on the same chromosome might be linked), sequencing of alleles at each locus to demonstrate the nature of the underlying repeat motif and inheritance studies using known family groups to determine mode of inheritance and mutation rate. Genetic mapping was not within the scope of this project. Mapped genomes of closely related taxa are often available online, allowing chromosomal identify of loci to be assessed. However, no mapped reference genome was available for a snake species at the time of this study (Todd Castoe, *pers. comm.*). Assessment of linkage disequilibrium using the reference database genotypes is reported in the following chapter.

3.6.3.9 Stutter peak ratio

The stutter peak artefacts (n-1 peaks) routinely caused by strand slippage during DNA replication of repetitive units and therefore characteristic of STR loci were present at all loci included in this study. Indeed, stutter artefacts were an asset during locus identification, allowing a true STR product to be differentiated from non-target amplification. Characterisation of stutter peaks was not conducted as part of this validation, but involves evaluation of the expected proportion of stutter peak to associated allele at each locus by dividing the height of the stutter peak by the height of its parent peak. This percentage can subsequently be applied as a stutter filter, above which stutter peak labels will not automatically be removed by a genotype analysis software package such as Genemapper ID (Applied Biosystems).

Forward stutter (n+1) was very rarely seen during this study.

3.6.3.10 Other artefacts

Peak shouldering can occur due to incomplete Adenine addition. This has been discussed earlier in this chapter and measures were taken to minimise this, including addition of a ‘G’ base on reverse primers and incorporation of an elongated final extension step. As a result, split peaks and shouldering were not an issue in the profiles obtained from the optimised multiplexes. Two artefacts were seen during routine screening, one at the FAM labelled locus MS25 and the other at the VIC labelled locus MS2.

3.6.3.11 Blind trial tests

The final stages of the developmental validation process would ideally involve the reproducibility test outlined above to show replication of results between independent laboratories, as well as a blind trial using DNA extracts simulating case samples. The blind trial involves DNA extracts used in the study being selected at random by an independent person, their identities recorded and the tubes subsequently relabelled as unknowns. The unknowns are provided to the practitioner, who then obtains profiles from the unknowns and matches each profile back to an individual screened during the study. This demonstrates the robustness of the testing to identify the same individuals from independent tests. The inclusion both of samples included in the reference database to confirm concordance of full profiles and poor quality samples excluded from the reference database to demonstrate concordance in less than pristine samples would be of benefit.

While not specifically part of the validation process, the use of appropriate controls must be mentioned when applying this assay to casework. During development, positive and negative PCR controls were not applied to maximise samples amplified within the limited resources of the study. However, during the blind trial and any application to actual questions of identity positive and negative controls must be included during testing to identify false negatives due to a failed PCR and false positives due to DNA contamination, respectively.

3.6.3.12 Mixture studies and stability studies

These two studies are commonly included in developmental validation of commercial human STR profiling systems. However, their utility in non-human STR profiling is arguable, due to their rarity in the sample types encountered. Mixed DNA samples from two members of the same species are rarely likely to be encountered in investigations that involve python DNA profiling. Should this sort of case scenario arise, a validation study involving mixture analyses should be conducted prior to reporting of the case samples.

Stability studies involve examining the effect of inhibitors of PCR amplification of the profile. Common inhibitors are haem from blood, humic acid from soil and the dyes found in denim clothing. Common commercial forensic grade extraction kits such as the DNA Investigator kit used in this study are effective at removing haem from the extract, so inhibition caused by this molecule should not be an issue when dealing with extracts obtained using these kits. Samples originating from soil are likewise rarely encountered in case scenarios, along with DNA to be extracted from a denim substrate. Should these sample types arise, validation testing should be conducted to determine the potential effect of inhibition on the DNA profile before reporting of case results.

3.6.3.13 Determination of allele calling thresholds

There are two important thresholds not yet discussed that need to be determined in order to confidently and robustly interpret a DNA profile. The first is the analytical threshold, otherwise called the Limit of Detection or LOD. This is applied to differentiate peaks from background noise within the electropherogram. The noise in a capillary electrophoresis detection system is quantified through assessing the background noise present in control samples (extraction reagent blanks, negative and positive amplification controls) that should be included routinely in forensic casework. The LOD is calculated as three standard deviations of the mean of the background noise [75].

The other threshold is the stochastic threshold, or the Limit of Quantitation (LOQ). This is the level below which peak drop out can reasonably be expected to occur. Below this level, stochastic effects of amplification introduce difficulties in accurately quantifying an amount of product present. This is a particularly important confounding factor for mixture studies, which have traditionally manually deconvoluted contributing profiles by assessing the peak height ratios of the different contributors. The incidence of mixed samples involving carpet pythons in criminal casework is likely to be low. The LOQ is commonly calculated as ten standard deviations of the mean of background noise [75]. While heterozygote alleles at a locus of a known single source profile can be interpreted between the analytical threshold and the stochastic threshold, single peaks should not be assumed to be a homozygote as the probability of heterozygote peak dropout within this range is high.

These thresholds are specific to analytical conditions and therefore need to be assessed by every laboratory applying the multiplex assay as part of the internal validation.

3.6.3.14 Population data and forensic statistics

Population data and forensic statistics are the final aspects of the dataset that should be evaluated during a developmental validation, which leads into considerations of the final aim of the chapter.

3.6.4 To compile a DNA reference database from native individuals screened across the species range and subject this to statistical analyses in order to comment on direct applicability of the tests to casework.

The whole database statistics associated with this marker set that are presented in this chapter demonstrate a high power of discrimination, holding promise for an informative forensic individualisation and provenance test for carpet pythons. However, in its current form the reference database is not suitable for application to forensic casework. The dataset shows highly significant deviation from Hardy-Weinberg equilibrium at every locus. This can be indicative of Wahlund Effect - the pooling of sub-populations into one continuous population.

This result is unsurprising. A dataset covering such a large geographic range is extremely unlikely to be one interbreeding population unless the species in question exhibits very high dispersal rates. Many geographic barriers exist in the Australia landscape that are likely barriers to dispersal, one of the most recognisable being the Great Dividing Range mountain range that stretches down the southern east coast of the country, dividing the sea from the interior. The majority of the internal Australian landscape is arid desert and the inhospitable nature of this environment for carpet pythons is evident from their known geographic range [76], which excludes much of this area. Carpet pythons separated by thousands of kilometres of desert could not be expected to represent one interbreeding population.

3.7. Concluding remarks

The current reference dataset provides an overview of the genetic variability within carpet pythons across Australia, but is flawed for forensic application. Further population genetic analysis is required to determine the number of realistic sub-populations and degree of genetic structure present in the overall dataset. This will enable appropriate theta corrections to be established, a necessary component when reaching statistical conclusions about the strength of match probabilities associated with forensic DNA profiling evidence when population sub-structure is present. The degree of population genetic structuring present in this dataset and its capability to provide statistical support for the types of questions arising in forensic investigations of crimes involving carpet pythons is the subject of Chapter 4.

3.8 References

[1] UNODC, World Wildlife Crime Report: Trafficking in protected species, New York, 2016. http://www.unodc.org/unodc/en/data-and-analysis/wildlife.html.

[2] NIST, OSAC Organizational Structure, (2017). https://www.nist.gov/topics/forensic-science/osac-organizational-structure (accessed October 6, 2017).

[3] E. Alacs, A. Georges, Wildlife across our borders: a review of the illegal trade in Australia, Aust. J. Forensic Sci. 40 (2008) 147–160.

[4] J. Herbig, The illegal reptile trade as a form of conservation crime: a South African criminological investigation, in: R. White (Ed.), Glob. Environ. Harm Criminol. Perspect., Taylor & Francis, 2013: pp. 110–131.

[5] G.E. Rosen, K.F. Smith, Summarizing the evidence on the international trade in illegal wildlife, Ecohealth. 7 (2010) 24–32.

[6] C. Barry, Rogue Traders, Aust. Geogr. Mag. (2011) 124–130.

[7] S. Ciavaglia, S. Donnellan, J. Henry, A. Linacre, Species identification of protected carpet pythons suitable for degraded forensic samples, Forensic Sci. Med. Pathol. 10 (2014) 295–305.

[8] M. Auliya, Hot Trade in Cool Creatures: A Review of the Live Reptile Trade in the European Union in the 1990s with a focus on Germany, Brussels, Belgium, 2003.

[9] J. Holden, By Hook or by Crook: A reference manual on illegal wildlife trade and prosecutions in the United Kingdom, 1998.

[10] M. Jenkins, S. Broad, International Trade in Reptile Skins: A Review and Analysis of the Main Consumer Markets, 1983-1991, 1994.

[11] J.S. Keogh, D.G. Barker, R. Shine, Heavily exploited but poorly known: systematics and biogeography of commercially harvested pythons (Python curtus group) in Southeast Asia, Biol. J. Linn. Soc. 73 (2001) 113–129.

[12] M. Auliya, S. Altherr, D. Ariano-Sanchez, E.H. Baard, C. Brown, R.M. Brown, J.C. Cantu, G. Gentile, P. Gildenhuys, E. Henningheim, J. Hintzmann, K. Kanari, M. Krvavac, M. Lettink, J. Lippert, L. Luiselli, G. Nilson, T.Q. Nguyen, V. Nijman, J.F. Parham, S.A. Pasachnik, M. Pedrono, A. Rauhaus, D.R. Córdova, M.E. Sanchez, U. Schepp, M. van Schingen, N. Schneeweiss, G.H. Segniagbeto, R. Somaweera, E.Y. Sy, O. Türkozan, S. Vinke, T. Vinke, R. Vyas, S. Williamson, T. Ziegler, Trade in live reptiles, its impact on wild populations, and the role of the European market, Biol. Conserv. 204 (2016) 103–119.

[13] G.W. Heard, D. Black, P. Robertson, Habitat use by the inland carpet python (Morelia spilota metcalfei : Pythonidae): Seasonal relationships with habitat structure and prey distribution in a rural landscape, 1994 (2004) 446–460.

[14] National Research Council, Strengthening Forensic Science in the United States: A Path Forward, National Academies Press, 2009.

[15] M.G. Gardner, A.J. Fitch, T. Bertozzi, A.J. Lowe, Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development, Mol. Ecol. Resour. 11 (2011) 1093–1101.

[16] E. Guichoux, L. Lagache, S. Wagner, P. Chaumeil, P. Leger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin, R.J. Petit, Current trends in microsatellite genotyping, Mol. Ecol. Resour. 11 (2011) 591–611.

[17] J.S. Chamberlain, J.R. Chamberlain, Optimization of multiplex PCR, in: K.B. Muller, F. Ferre, R.A. Gibbs (Eds.), Polym. Chain React., 1994: pp. 38–46.

[18] J.M. Butler, Constructing STR multiplex assays., in: A. Carracedo (Ed.), Forensic DNA Typing Protoc., Springer Science & Business Media, 2005: pp. 53–66.

[19] M. Poetsch, S. Seefeldt, M. Maschke, E. Lignitz, Analysis of microsatellite polymorphism in red deer, roe deer, and fallow deer–possible employment in forensic applications, Forensic Sci. Int. 116 (2001) 1–8.

[20] R.M. Jobin, D. Patterson, Y. Zhang, DNA typing in populations of mule deer for forensic use in the Province of Alberta, Forensic Sci. Int. Genet. 2 (2008) 190–197.

[21] E. Socratous, E. a. M. Graham, G.N. Rutty, Forensic DNA profiling of Cervus elaphus species in the United Kingdom, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 281–282. doi:10.1016/j.fsigss.2009.08.127.

[22] S. Caratti, L. Rossi, B. Sona, S. Origlia, S. Viara, G. Martano, C. Torre, C. Robino, Analysis of 11 tetrameric STRs in wild boars for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 339–342.

[23] C.R. Primmer, M.T. Koskinen, J. Piironen, The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud., Proc. Biol. Sci. 267 (2000) 1699–1704.

[24] J.P. Grobler, A. Kotze, H. Swart, E.M. Hallerman, The application of microsatellite DNA markers for forensic analysis of koi carp (Cyprinus carpio), S. Afr. J. Sci. 101 (2005) 19–21.

[25] H.G. Eiken, R.J. Andreassen, A. Kopatz, S.G. Bjervamoen, I. Wartiainen, C. Tobiassen, P.M. Knappskog, P.E. Aspholm, M.E. Smith, J. Aspi, Population data for 12 STR loci in Northern European brown bear (Ursus arctos) and application of DNA profiles for forensic casework, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 273–274. doi:10.1016/j.fsigss.2009.07.007.

[26] S. Mondol, N. R, V. Athreya, K. Sunagar, V. Selvaraj, U. Ramakrishnan, A panel of microsatellites to individually identify leopards and its application to leopard monitoring in human dominated landscapes, BMC Genet. 10 (2009) 79.

[27] R. Caniglia, E. Fabbri, C. Greco, M. Galaverni, E. Randi, Forensic DNA against wildlife poaching: Identification of a serial wolf killing in Italy, Forensic Sci. Int. Genet. 4 (2010) 334–338.

[28] N. Dawnay, R. Ogden, R.S. Thorpe, L.C. Pope, D.A. Dawson, R. McEwing, A forensic STR profiling system for the Eurasian badger: A framework for developing profiling systems for wildlife species, Forensic Sci. Int. Genet. 2 (2008) 47–53.

[29] R. Ogden, N. Dawnay, R. McEwing, Development of STR profiling systems for individual identification in wildlife: A case study of the Eurasian badger, Meles meles, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 612–613. doi:10.1016/j.fsigss.2007.10.224.

[30] M. Wesselink, I. Kuiper, Individual identification of fox (Vulpes vulpes) in forensic wildlife investigations, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e214–e215. doi:10.1016/j.fsigss.2011.08.107.

[31] B. Van Asch, C. Alves, F. Pereira, L. Gusmão, A. Amorim, A new autosomal STR multiplex for canine genotyping, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 628–629. doi:10.1016/j.fsigss.2007.10.043.

[32] R. Ogden, R.J. Mellanby, D. Clements, A.G. Gow, R. Powell, R. McEwing, Genetic data from 15 STR loci for forensic individual identification and parentage analyses in UK domestic dogs (Canis lupus familiaris), Forensic Sci. Int. Genet. 6 (2012) e63–e65.

[33] E. Wictum, T. Kun, C. Lindquist, J. Malvick, D. Vankan, B. Sacks, Developmental validation of DogFiler, a novel multiplex for canine DNA profiling in forensic casework, Forensic Sci. Int. Genet. 7 (2013) 82–91.

[34] N. Dawnay, R. Ogden, J.H. Wetton, R.S. Thorpe, R. McEwing, Genetic data from 28 STR loci for forensic individual identification and parentage analyses in 6 bird of prey species, Forensic Sci. Int. Genet. 3 (2009).

[35] N.E. White, R. Dawson, M.L. Coghlan, S.R. Tridico, P.R. Mawson, J. Haile, M. Bunce, Application of STR markers in wildlife forensic casework involving Australian black-cockatoos (Calyptorhynchus spp.), Forensic Sci. Int. Genet. 6 (2012) 664–670.

[36] B. Berger, C. Berger, W. Hecht, A. Hellmann, U. Rohleder, U. Schleenbecker, W. Parson, Validation of two canine STR multiplex-assays following the ISFG recommendations for non-human DNA analysis, Forensic Sci. Int. Genet. 8 (2014) 90–100.

[37] M.J.C. Van Hoppe, M.A.V. Dy, M. Van Den Einden, A. Iyengar, SkydancerPlex: A novel STR multiplex validated for forensic use in the hen harrier (Circus cyaneus), Forensic Sci. Int. Genet. 22 (2016) 100–109.

[38] K.L. Dicks, L.M.I. Webster, I. McDowall, S.M. Muya, J. Hopper, P. O'Donoghue, Validation studies on dinucleotide STRs for forensic identification of black rhinoceros Diceros bicornis, Forensic Sci. Int. Genet. 26 (2017) e25–e27. doi:10.1016/j.fsigen.2016.10.016.

[39] S.K. Wasser, C. Mailand, R. Booth, B. Mutayoba, E. Kisamo, B. Clark, M. Stephens, Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 4228–4233.

[40] S.K. Wasser, W. Joseph Clark, O. Drori, E. Stephen Kisamo, C. Mailand, B. Mutayoba, M. Stephens, Combating the illegal trade in African elephant ivory with DNA forensics, Conserv. Biol. 22 (2008) 1065–1071.

[41] S.K. Wasser, A.M. Shedlock, K. Comstock, E. a Ostrander, B. Mutayoba, M. Stephens, Assigning African elephant DNA to geographic region of origin: applications to the ivory trade., Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 14847–14852.

[42] P. Jordan, A. Goodman, S. Donnellan, Microsatellite primers for Australian and New Guinean pythons isolated with an efficient marker development method for related species, Mol. Ecol. Notes. 2 (2002) 78–82.

[43] D. Taylor, Using DNA markers for wildlife management and protection: a study of the population structure and systematics of the Australian carpet pythons (Reptilia: Morelia spilota complex), Flinders University, SA, 2005.

[44] A. Linacre, L. Gusmao, W. Hecht, A.P. Hellmann, W.R. Mayr, W. Parson, M. Prinz, P.M. Schneider, N. Morling, ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations, Forensic Sci. Int. Genet. 5 (2011) 501–505.

[45] Scientific Working Group on DNA Analysis Methods (SWGDAM), Validation Guidelines for DNA Analysis Methods, (2012). https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf (accessed October 6, 2017).

[46] Forensic Science Regulator, Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System, London, UK, 2017. doi:978-1-84987-624-7.

[47] E. Meglécz, C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, J.F. Martin, QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects, Bioinformatics. 26 (2010) 403–404.

[48] S. Rosen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, in: Bioinforma. Methods Protoc., Humana Press, Totowa, NJ, 2000: pp. 365–386.

[49] M.J. Hayden, T.M. Nguyen, A. Waterman, K.J. Chalmers, Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping., BMC Genomics. 9 (2008) 80.

[50] M. Schuelke, An economic method for the fluorescent labeling of PCR fragments, Nat. Biotechnol. 18 (2000) 233–234.

[51] S. Ciavaglia, H. Dridan, A. Linacre, Getting more for less: can forensic tools for Australian wildlife enforcement support international compliance efforts?, Aust. J. Forensic Sci. (2017) 1–10.

[52] M.J. Brownstein, J.D. Carpten, J.R. Smith, Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping, Biotechniques. 20 (1996) 1004–1010.

[53] C.P. Kimpton, P. Gill, A. Walton, A. Urquhart, E.S. Millican, M. Adams, Automated DNA profiling employing multiplex amplification of short tandem repeat loci., Genome Res. 3 (1993) 13–22.

[54] S. Ciavaglia, A. Linacre, OzPythonPlex: An optimised forensic STR multiplex assay set for the Australasian carpet python (Morelia spilota), Forensic Sci. Int. Genet. 34 (2018).

[55] E. Southern, Measurement of DNA Length by Gel Electrophoresis, Anal. Biochem. 100 (1979) 319–323.

[56] J.A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Sci. Int. Genet. 7 (2013) 296–304.

[57] W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, W.R. Mayr, B. Olaisen, DNA recommendations. Further report of the DNA Commission of the ISFG regarding the use of short tandem repeat systems, Forensic Sci. Int. 87 (1997) 179–184.

[58] R. Peakall, P.E. Smouse, GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update, Bioinformatics. 28 (2012) 2537–2539.

[59] J.M. Butler, Advanced topics in forensic DNA typing: interpretation, Academic Press, 2014.

[60] J.M. Butler, Forensic DNA Typing: biology, technology and genetics of STR markers, Elsevier Academic Press, 2005.

[61] R.N. Johnson, L. Wilson-Wilde, A. Linacre, Current and future directions of DNA in wildlife forensic science, Forensic Sci. Int. Genet. 10 (2014) 1–11.

[62] J.M. Hancock, Microsatellites and other simple sequences: genomic context and mutational mechanisms, in: D.B. Goldstein, C. Schlotterer (Eds.), Microsatellites Evol. Appl., Oxford University Press, 1999.

[63] S.A. Ciavaglia, J.M. Henry, A.M.T. Linacre, Profiling pythons to combat common illegal wildlife activities, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e31–e32. doi:http://dx.doi.org/10.1016/j.fsigss.2013.10.015.

[64] S. Ciavaglia, S. Donnellan, S. Tobe, J. Henry, A. Linacre, A novel forensic DNA profiling technique for protected species, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e258–e260. doi:10.1016/j.fsigss.2015.09.103.

[65] J.T. Sutton, B.C. Robertson, I.G. Jamieson, Dye shift: a neglected source of genotyping error in molecular ecology, Mol. Ecol. Resour. 11 (2011) 514–520.

[66] H. Stan-Lotter, S. Leuko, A. Legat, S. Fendrihan, 24 The Assessment of the Viability of Halophilic Microorganisms in Natural Communities, Methods Microbiol. 35 (2006) 569–584.

[67] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures, Forensic Sci. Int. 160 (2006) 90–101.

[68] J.S. Buckleton, J.A. Bright, D. Taylor, Forensic DNA Evidence Interpretation, 2nd editio, CRC Press, Taylor & Francis Group, 2016.

[69] W. Amos, A comparative apporach to the study of microsatellite evolution, in: D.B. Goldstein, C. Schlotterer (Eds.), Microsatellites Evol. Appl., Oxford University Press, 1999.

[70] A. Estoup, J.-M. Cornuet, Microsatellite evolution: inferences from population data, in: D.B. Goldstein, C. Schlotterer (Eds.), Microsatellites Evol. Appl., Oxford University Press, 1999.

[71] Hartl, Primer of Population Genetics, 3rd editio, Sinauer Associates, Inc., 2000.

[72] C.W. Dieffenbach, T.M. Lowe, G.S. Dveksler, General concepts for PCR primer design., Genome Res. 3 (1993) S30–S37. doi:10.1101/gr.3.3.S30.

[73] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, L. Gusmão, D.R. Hares, J.A. Irwin, J.L. King, P. De Knijff, N. Morling, M. Prinz, P.M. Schneider, C. Van Neste, S. Willuweit, C. Phillips, Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements, Forensic Sci. Int. Genet. 22 (2016) 54–63.

[74] Applied Biosystems, GlobalFiler™ PCR Amplification Kit User Manual, Publication, ThermoFisher Scientific, 2015.

[75] J.R. Gilder, T.E. Doom, K. Inman, D.E. Krane, Run-specific limits of detection and quantitation for STR-based DNA testing, J. Forensic Sci. 52 (2007) 97–101.

[76] D.G. Barker, T.M. Barker, Pythons of the World, in: Advanced Vivarium Systems Inc, California, 1994.

Appendix 3.1: Publication - *OzPythonPlex*: An optimised forensic STR multiplex assay set for the Australasian carpet python (*Morelia spilota*)

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Appendix 3.2: Publication - Profiling pythons to combat common illegal wildlife activities

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Appendix 3.3: Publication - Getting more for less: can forensic tools for Australian wildlife enforcement support international compliance efforts?

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Appendix 3.4: Publication - A novel forensic DNA profiling technique for protected species

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Publication has been removed due to copyright restrictions.

Appendix 3.5: Samples used in chapters 3 and 4

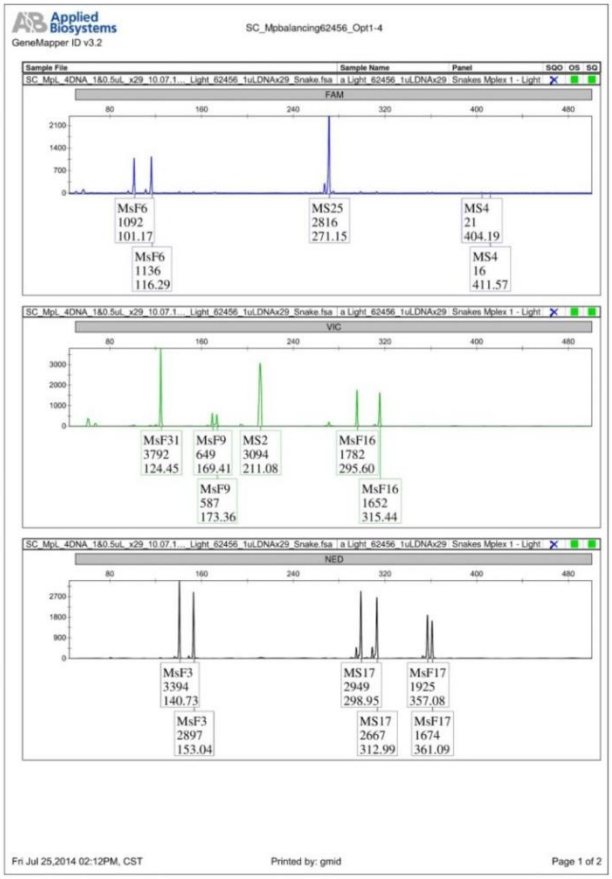
Complete loci typed	ABTC #	Sample Name	Provenance	Voucher #	Use	Sample type	Reference population
24	13446	*Morelia spilota*	Gosford, NSW		R24	Scale clipping	SEC
24	13449	*Morelia spilota variegata*	Malak, Darwin, NT		R24	DNA extract from DT (blood)	NEC
24	13455	*Morelia spilota variegata*	East Point, Darwin, NT		R24	DNA extract from DT (scale)	NEC
24	17464	*Morelia spilota*	Brisbane, Qld		R24	DNA extract from DT (blood)	SEC
24	17482	*Morelia spilota*	Buckleboo, SA		R24	DNA extract from DT (blood)	Msi ML
24	30087	*Morelia spilota*	Maxwell Creek, Melville Island, NT	SAMAR39733	R24	Tissue	NEC
24	32052	*Morelia spilota*	Mackay, Qld		R24	Tissue	SEC
24	34289	*Morelia spilota variegata*	2k SE Blanchetown, SA	SAMAR39733	R24	Liver	SIB
24	50866	*Morelia spilota*	St Lucia, Qld		R24	Unknown	SEC
24	51511	*Morelia spilota*	Parents from Tully		R24	DNA extract from DT (blood)	NEC
24	51574	*Morelia spilota imbricata*	Wooroloo, WA		R24	DNA extract from DT (blood)	Msi ML
24	55499	*Morelia spilota*	Goyder Lagoon, SA	SAMAR26877	R24	Tissue	SIB
24	55500	*Morelia spilota*	Goyder Lagoon, SA	SAMAR26878	R24	Tissue	SIB
24	55510	*Morelia spilota*	Goyder Lagoon, SA	SAMAR27496	R24	DNA extract from DT	SIB
24	59978	*Morelia spilota*	Brisbane, Qld		R24	DNA extract from DT (liver)	SEC
24	62456	*Morelia spilota*	Mt Connor, WA	WAMR96970	R24	Liver	NEC
24	66032	*Morelia spilota imbricata*	Dwellingup, WA		R24	DNA extract from DT	Msi ML
24	66062	*Morelia spilota*	Toowoomba, Qld		R24	DNA extract from DT	SEC
24	66063	*Morelia spilota*	Toowoomba, Qld		R24	DNA extract from DT	SEC
24	66068	*Morelia spilota spilota*	Bellangry via Wauchope, NSW		R24	DNA extract from DT	SEC
24	66106	*Morelia spilota*	Toowoomba, Qld		R24	DNA extract from DT	SEC
24	66138	*Morelia spilota*	Thompson Creek Causeway, Cape Tribulation, Qlc		R24	DNA extract from DT	NEC
24	66139	*Morelia spilota*	Alexandra Range, Daintree, Qlc		R24	DNA extract from DT	NEC
24	66140	*Morelia spilota*	Granite Gorge turnoff, Walkamin, Qld		R24	DNA extract from DT	NEC
24	66142	*Morelia spilota*	Cape Tribulation, Qlc		R24	DNA extract from DT	NEC
24	66152	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66153	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66154	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66156	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66158	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66159	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66160	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66161	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66162	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66163	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66164	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66165	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66167	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66168	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66169	*Morelia spilota imbricata*	St Francis Island, SA		R24	DNA extract from DT	Msi SFI
24	66267	*Morelia spilota spilota*	Seaham, NSW		R24	Tail	SEC
24	66297	*Morelia spilota*	N. Cape Tribulation village, Qlc		R24	Unknown	NEC
24	66298	*Morelia spilota*	Ravensbourne NP, Qlc		R24	DNA extract from DT	SEC
24	66322	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66323	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66325	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66326	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66329	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66330	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66332	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66333	*Morelia spilota imbricata*	Mundaring District, WA		R24	DNA extract from DT	Msi ML
24	66336	*Morelia spilota imbricata*	Boyogin Rock Nat Res, WA		R24	DNA extract from DT	Msi ML
24	66339	*Morelia spilota imbricata*	Boyogin Rock Nat Res, WA		R24	DNA extract from DT	Msi ML
24	66328	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	66341	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66342	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66343	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66344	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66345	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66346	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66349	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66350	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66352	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66354	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66355	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66356	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66357	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66361	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	66418	*Morelia spilota*	Mt Nebo, Qld		R24	DNA extract from DT	SEC
24	67126	*Morelia spilota*	Cape Tribulation, Qlc		R24	DNA extract from DT	NEC
24	67127	*Morelia spilota*	Atherton Tableland, Qlc		R24	DNA extract from DT	NEC
24	67128	*Morelia spilota*	Eungella, Qlc		R24	DNA extract from DT	SEC
24	67130	*Morelia spilota*	Eungella, Qlc		R24	DNA extract from DT	NEC
24	67131	*Morelia spilota imbricata*	Dwellingup, WA		R24	DNA extract from DT	Msi ML
24	67134	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT	Msi ML
24	67135	*Morelia spilota imbricata*	Dwellingup, WA		R24	DNA extract from DT	Msi ML
24	67138	*Morelia spilota imbricata*	Wickepine Reserve, Abrolhos Islands, WA		R24	DNA extract from DT	Msi ML
24	67139	*Morelia spilota imbricata*	West Wallabi Island, WA		R24	DNA extract from DT	Msi ML
24	67140	*Morelia spilota imbricata*	West Wallabi Island, WA		R24	DNA extract from DT	Msi ML
24	67141	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI
24	67142	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT	Msi GI

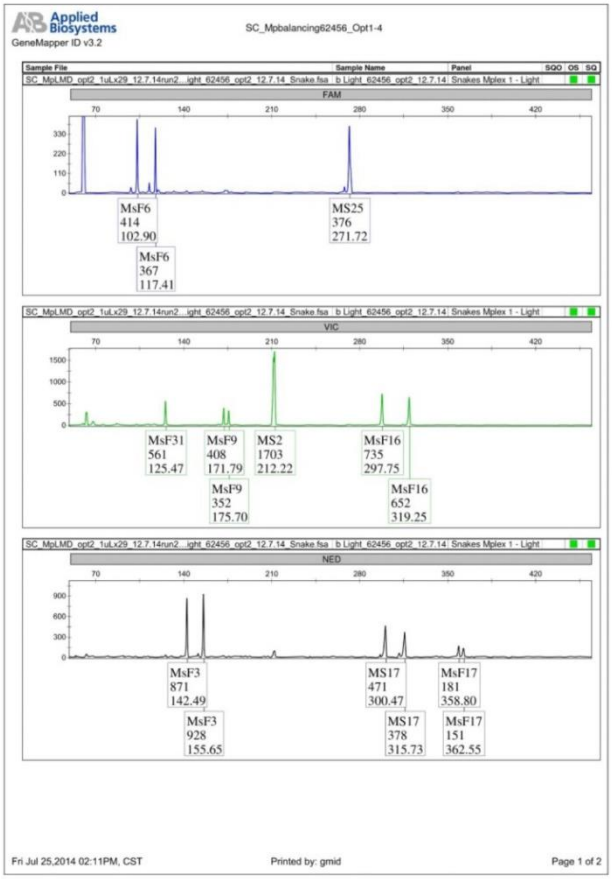
Complete loci typed	ABTC #	Sample Name	Provenance	Voucher #	Use	Sample type	Reference population
24	67143	*Morelia spilota imbricata*	Garden Island, WA		R24	Scale clipping	Msi GI
24	67144	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT (blood)	Msi GI
24	67145	*Morelia spilota imbricata*	Garden Island, WA		R24	DNA extract from DT (scale)	Msi GI
24	67146	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT (blood)	Msi ML
24	67147	*Morelia spilota imbricata*	Dryandra, WA		R24	DNA extract from DT (blood)	Msi ML
24	67149	*Morelia spilota imbricata*	Dryandra, WA		R24	Tissue	Msi ML
24	67164	*Morelia spilota*	Merauke, West Papua		R24	Tissue	NEC
24	67190	*Morelia spilota*	Merauke, West Papua		R24	Liver	NEC
24	67579	*Morelia spilota variegata*	Cape Tribulation, Qlc		R24	Unknown	NEC
24	67734	*Morelia spilota*	Broken Head, S Byron Bay, NSW		R24	DNA extract from DT (blood)	SEC
24	68284	*Morelia spilota*	Byabarra, NSW		R24	DNA extract from DT (blood)	SEC
24	68285	*Morelia spilota*	Caparra, NSW		R24	Tissue	SEC
24	68293	*Morelia spilota*	Tilba, NSW		R24	Tissue	SEC
24	68297	*Morelia spilota*	Depot Springs, 30k E Copley, SA		R24	DNA extract from DT	SIB
24	68298	*Morelia spilota*	Fowlers Bay, SA		R24	DNA extract from DT (liver)	Msi ML
24	68331	*Morelia spilota*	Coleambally, NSW		R24	Liver	SIB
24	68333	*Morelia spilota*	Lake Cathie/Bonny Hills, NSW		R24	DNA extract from DT	SEC
24	69338	*Morelia spilota*	Mareeba-Atherton Road, Qlc		R24	DNA extract from DT	SEC
24	69344	*Morelia spilota*	10k Sth Babinda, Qlc		R24	DNA extract from DT	NEC
24	69345	*Morelia spilota*	NW Cape Tribulation, Qlc		R24	DNA extract from DT	NEC
24	69346	*Morelia spilota*	Black Mountain Trevathan Range, Qlc		R24	DNA extract from DT	NEC
24	71371	*Morelia spilota*	Mt Hope, Vic		R24	DNA extract from DT	SIB
24	71372	*Morelia spilota*	Mt Hope, Vic		R24	DNA extract from DT	SIB
24	71374	*Morelia spilota*	Mt Hope, Vic		R24	DNA extract from DT	SIB
24	71375	*Morelia spilota*	Caanan, Mt Meg, Vic		R24	DNA extract from DT	SIB
24	71376	*Morelia spilota*	Mt Hope, Vic		R24	DNA extract from DT	SIB
24	71384	*Morelia spilota*	Warby Ranges, Vic		R24	DNA extract from DT	SIB
24	71388	*Morelia spilota*	Mildura, Vic		R24	DNA extract from DT	SIB
24	71390	*Morelia spilota*	Gunbower Island, Vic		R24	DNA extract from DT	SIB
24	71393	*Morelia spilota*	Nyah Forest, Vic		R24	DNA extract from DT	SIB
24	71395	*Morelia spilota*	Walpella Island, Vic		R24	DNA extract from DT	SIB
24	71396	*Morelia spilota*	Walpella Island, Vic		R24	DNA extract from DT	SIB
24	71398	*Morelia spilota*	Caanan, Mt Meg, Vic		R24	DNA extract from DT	SIB
24	71399	*Morelia spilota*	Piambie, Vic		R24	DNA extract from DT	SIB
24	71425	*Morelia spilota*	Chesney Vale, Vic		R24	DNA extract from DT	SIB
24	71426	*Morelia spilota*	Chesney Vale, Vic		R24	DNA extract from DT	SIB
24	71427	*Morelia spilota*	Chesney Vale, Vic		R24	DNA extract from DT	SIB
24	71428	*Morelia spilota*	Boneya, Vic		R24	DNA extract from DT	SIB
24	71429	*Morelia spilota*	Mt Bruno, Vic		R24	Tail	SIB
24	71430	*Morelia spilota*	Caanan, Mt Meg, Vic		R24	Unknown	SIB
24	71434	*Morelia spilota*	Bungeet, Vic		R24	DNA extract from DT	SIB
24	71435	*Morelia spilota*	Mt Bruno, Vic		R24	DNA extract from DT	SIB
24	71436	*Morelia spilota*	Boneya, Vic		R24	DNA extract from DT	SIB
24	71437	*Morelia spilota*	Chinaman's Creek, Mt Killawarra, Vic		R24	DNA extract from DT	SIB
24	71438	*Morelia spilota*	Mt Killawarra, Vic		R24	DNA extract from DT	SIB
24	71439	*Morelia spilota*	Caanan, Mt Meg, Vic		R24	DNA extract from DT	SIB
24	71441	*Morelia spilota*	Mt Killawarra, Vic		R24	DNA extract from DT	SIB
24	75732	*Morelia spilota*	Goondicum, Qld		R24	DNA extract from DT	NEC
24	76771	*Morelia spilota*	Tully Gorge Road, Qld		R24	DNA extract from DT	NEC
24	76772	*Morelia spilota*	Tully Gorge Road, Qld		R24	DNA extract from DT	SEC
24	76777	*Morelia spilota*	Trebonne Creek, Ingham, SA		R24	DNA extract from DT	NEC
24	76778	*Morelia spilota*	Bruce Highway between Cardwell and Tully, Qlc		R24	DNA extract from DT	NEC
24	79106	*Morelia spilota*	D'Aguilar, Mt Nebo, Qld		R24	DNA extract from DT	SEC
24	71442	*Morelia spilota*	Mt Bruno, Vic		R24	DNA extract from DT	SIB
24	79108	*Morelia spilota*	D'Aguilar, Mt Nebo, Qld		R24	DNA extract from DT	SEC
24	79109	*Morelia spilota*	Eungella, Qlc		R24	DNA extract from DT	SEC
24	79110	*Morelia spilota*	Atherton Tableland, Qlc		R24	DNA extract from DT	NEC
24	79113	*Morelia spilota*	61k S Childers, Qld		R24	DNA extract from DT	SEC
24	79115	*Morelia spilota*	3.6k N Cambridge Plateau, Richmond Range, NSW		R24	DNA extract from DT	SEC
24	79117	*Morelia spilota*	Tully, Qld		R24	DNA extract from DT	NEC
24	79118	*Morelia spilota*	Palmerston, Qld		R24	DNA extract from DT	NEC
24	79119	*Morelia spilota*	Palmerston, Qld		R24	DNA extract from DT	NEC
24	79120	*Morelia spilota*	Canyon Lookout, Springbrook N, Qlc		R24	DNA extract from DT	SEC
24	79121	*Morelia spilota*	Dome Mountain, Richmond Range, Qlc		R24	DNA extract from DT	SEC
24	80324	*Morelia spilota*	3.8k N Cambridge Plateau, Richmond Range, Qlc		R24	DNA extract from DT	SEC
24	80506	*Morelia spilota*	50 m from Wrights Creek Road, Qld		R24	DNA extract from DT	NEC
24	80511	*Morelia spilota*	Range View Housing Estate, Qlc		R24	DNA extract from DT	NEC
24	81029	*Morelia spilota*	Shoalwater Bay Army Training Reserve, Qlc	ANWCR05495	R24	DNA extract from DT	NEC
24	81030	*Morelia spilota*	60k East Wauchope, NSW	ANWCR06767	R24	DNA extract from DT	SEC
24	81239	*Morelia spilota*	Mountain Creek, Paluma, Qlc		R24	DNA extract from DT	NEC
24	81299	*Morelia spilota*	Mullumbimby, NSW	AMSR133491	R24	DNA extract from DT	SEC
24	81300	*Morelia spilota*	Atherton, Qld		R24	DNA extract from DT	NEC
24	81302	*Morelia spilota*	Spencers Cutting via Wauchope, NSW	AMSR152759	R24	DNA extract from DT	SEC
24	81303	*Morelia spilota*	Kempsey, NSW	AMSR152768	R24	DNA extract from DT	SEC
24	81304	*Morelia spilota*	Eden Creek Falls, Toonumbar NP, NSW	AMSR153858	R24	DNA extract from DT	SEC
24	81305	*Morelia spilota*	Acacia Plateau, NSW	AMSR153862	R24	DNA extract from DT	SEC
24	81308	*Morelia spilota*	Grassy Head, NSW	AMSR157290	R24	DNA extract from DT	SEC
24	81310	*Morelia spilota*	Coutts Crossing, NSW	AMSR157784	R24	DNA extract from DT	SEC
24	81312	*Morelia spilota*	Uki, NSW	AMSR157789	R24	DNA extract from DT	SEC
24	81313	*Morelia spilota*	Yarrawarrah, NSW	AMSR157793	R24	DNA extract from DT	SEC
24	81314	*Morelia spilota*	Angourie Village, NSW	AMSR158201	R24	DNA extract from DT	SEC

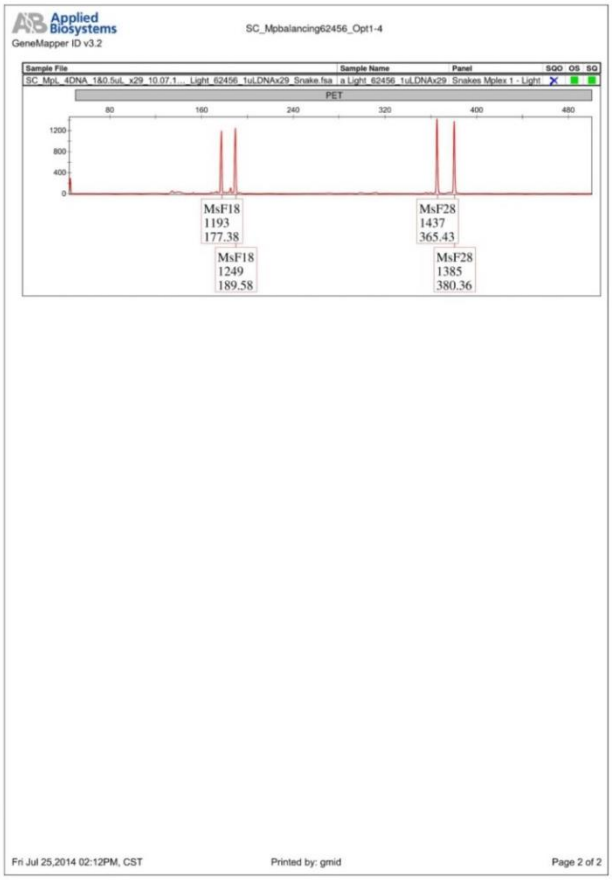
Complete loci typed	ABTC #	Sample Name	Provenance	Voucher #	Use	Sample type	Reference population
24	81315	*Morelia spilota*	Royal NP, NSW	AMSR158209	R24	DNA extract from DT	SEC
24	81316	*Morelia spilota*	Tolga, Qld	AMSR155064	R24	DNA extract from DT	NEC
24	81318	*Morelia spilota*	Dingo SF, NSW	AMSR148866	R24	DNA extract from DT	SEC
24	81319	*Morelia spilota*	Waterfall, NSW	AMSR151465	R24	DNA extract from DT	SEC
24	81321	*Morelia spilota*	Byron Bay, NSW	AMSR153022	R24	DNA extract from DT	SEC
24	81388	*Morelia spilota*	5k W Grevillea, NSW		R24	Tissue	SEC
24	81445	*Morelia bredli*	30k E Glen Helen Lodge, NT		R24	DNA extract from DT (skin)	SIB
24	81448	*Morelia bredli*	Alice Springs Desert Park, NT		R24	DNA extract from DT (skin)	SIB
24	81449	*Morelia bredli*	Alice Springs Desert Park, NT		R24	DNA extract from DT (skin)	SIB
24	82470	*Morelia spilota*	Myall Lakes, NSW		R24	Skin	SEC
24	82476	*Morelia spilota variegata*	Bulli Rock Hole, NT		R24	Skin	NEC
24	82967	*Morelia spilota*	Thuringowa, Qlc	QMJ82043	R24	Tail	NEC
24	82974	*Morelia spilota*	Townsville, Qld		R24	Tail	NEC
24	82986	*Morelia spilota*	Tully Gorge Road, Qld		R24	Tail	NEC
24	82987	*Morelia spilota*	AIMS Road, Qld		R24	Scale clipping	NEC
24	83022	*Morelia spilota*	Cungulla, Qlc	QMJ82038	R24	Tail	NEC
24	84767	*Morelia spilota metcalfei*	Cannonvale, Qlc		R24	Tail	NEC
24	86768	*Morelia spilota*	S. Kolijo, Qld		R24	Tail	NEC
24	86801	*Morelia spilota*	Paluma, Qld		R24	Tissue	NEC
24	106073	*Morelia spilota*	Karalia, E. Waikerie, SA		R24	Scale clipping	SIB
24	106075	*Morelia spilota*	Loxton, SA		R24	Skin	SIB
24	117064	*Morelia spilota*	Merauke, West Papua		R24	Scale clipping	NEC
24	118877	*Morelia spilota*	Marlow's Lagoon, NT		R24	Scale clipping	NEC
23	13382	*Morelia spilota variegata*	5k Nth Adelaide River, NT		R23	DNA extract from DT (scale)	NEC
23	13443	*Morelia spilota*	Humpty Doo HS, NT		R23	DNA extract from DT (skin)	NEC
23	17463	*Morelia spilota*	Brisbane, Qld		R23	DNA extract from DT (blood)	SEC
23	51575	*Morelia spilota imbricata*	Rivervale, WA		R23	DNA extract from DT (blood)	Msi ML
23	51576	*Morelia spilota imbricata*	Norseman, WA		R23	Blood	Msi ML
23	66061	*Morelia spilota*	Port Macquarie, NSW		R23	DNA extract from DT	SEC
23	66155	*Morelia spilota imbricata*	St Francis Island, SA		R23	DNA extract from DT	Msi SFI
23	66327	*Morelia spilota imbricata*	Dryandra, WA		R23	Scale clipping	Msi ML
23	67129	*Morelia spilota*	Tully Gorge Road, Qld		R23	DNA extract from DT	NEC
23	67137	*Morelia spilota imbricata*	West Wallabi Island, WA		R23	DNA extract from DT	Msi ML
23	67578	*Morelia spilota*	Cape Tribulation, Qlc		R23	Scale clipping	NEC
23	68279	*Morelia spilota*	Merauke, West Papua		R23	Skin	NEC
23	68280	*Morelia spilota*	Merauke, West Papua		R23	Skin	NEC
23	68308	*Morelia spilota*	Mondrain Island, WA		R23	Skin	Msi ML
23	68310	*Morelia spilota*	North Twin Peak Island, WA		R23	Skin	Msi ML
23	68334	*Morelia spilota*	Coleambally, NSW		R23	DNA extract from DT (skin)	SIB
23	68345	*Morelia spilota*	Port Macquarie, NSW		R23	Skin	SEC
23	71377	*Morelia spilota*	Piambie, Vic		R23	DNA extract from DT (skin)	SIB
23	71378	*Morelia spilota*	Piambie, Vic		R23	Skin	SIB
23	71389	*Morelia spilota*	Warby Ranges, Vic		R23	Skin	SIB
23	71392	*Morelia spilota*	Gol Gol, Vic		R23	DNA extract from DT	SIB
23	71402	*Morelia spilota*	Caanan, Mt Meg, Vic		R23	DNA extract from DT (skin)	SIB
23	71410	*Morelia spilota*	Piambie, Vic		R23	Skin	SIB
23	71433	*Morelia spilota*	Chesney Vale, Vic		R23	DNA extract from DT	SIB
23	76225	*Morelia spilota*	Dog Fence Beach, SA		R23	Scale clipping	Msi ML
23	76268	*Morelia spilota metcalfei*	Mt Wood HS, Sturt NP, NSW		R23	Scale clipping	SIB
23	76529	*Morelia spilota*	Ultoomurra WH, Goyder's Lagoon, SA		R23	Scale clipping	SIB
23	76530	*Morelia spilota*	Bobbiemoonga WH, Warburton Ra de, SA		R23	Scale clipping	SIB
23	79111	*Morelia spilota*	Atherton Tableland, Qlc		R23	DNA extract from DT	NEC
23	79122	*Morelia spilota*	36k from Yungaburra, Atherton Tableland, Qlc		R23	DNA extract from DT	NEC
23	71440	*Morelia spilota*	Caanan, Mt Meg, Vic		R23	DNA extract from DT	SIB
23	80507	*Morelia spilota*	Emerald Creek, Qld		R23	DNA extract from DT	NEC
23	80510	*Morelia spilota*	Gilles Highway, Qlc		R23	DNA extract from DT	NEC
23	81028	*Morelia spilota*	Eastern McIlwraith Range, Qlc	ANWCR05271	R23	Tissue	NEC
23	81224	*Morelia spilota imbricata*	11.8k W Yalata RH, SA		R23	Tail	Msi ML
23	81240	*Morelia spilota*	Paluma Peak, Qld		R23	Scale clipping	NEC
23	81311	*Morelia spilota*	Byron Bay, NSW	AMSR157785	R23	DNA extract from DT	SEC
23	81317	*Morelia spilota*	Tolga, Qld	AMSR155064	R23	DNA extract from DT	NEC
23	81320	*Morelia spilota*	10k from Gunnedah, NSW	AMSR151857	R23	Muscle	SEC
23	81447	*Morelia bredli*	Alice Springs Desert Park, NT		R23	DNA extract from DT (skin)	SIB
23	81659	*Morelia spilota*	Cannington Mine, Qlc		R23	Scale clipping	SIB
23	82975	*Morelia spilota*	Tully Gorge Road, Qld		R23	Tail	NEC
23	82996	*Morelia spilota*	AIMS Road, Qld	QMJ82121	R23	Tail	NEC
23	86774	*Morelia spilota*	Thuringowa, Qlc	NTMR27754	R23	Tail	NEC
23	106074	*Morelia spilota*	Karalia, E. Waikerie, SA		R23	Scale clipping	NEC
23	112609	*Morelia bredli*	Trephina, NT		R23	Unknown	SIB
23	117065	*Morelia spilota*	Merauke, West Papua		R23	Scale clipping	NEC
23	117068	*Morelia spilota*	Merauke, West Papua		R23	Scale clipping	NEC
23	117070	*Morelia spilota*	Merauke, West Papua		R23	Scale clipping	NEC
23	118879	*Morelia spilota*	Tiwi, NT		R23	Scale clipping	NEC
23	118880	*Morelia spilota*	Howard Springs, NT		R23	Scale clipping	NEC
23	123504	*Morelia spilota harrisoni*	Saroa-Pt Moresby, PNG		R23	Scale clipping	NEC
22	55482	*Morelia spilota variegata*	Darwin, NT	SAMAR21456	R22	Liver	NEC
22	66166	*Morelia spilota imbricata*	St Francis Island, SA		R22	DNA extract from DT	Msi SFI
22	66320	*Morelia spilota imbricata*	Dryandra, WA		R22	DNA extract from DT	Msi ML
22	66331	*Morelia spilota imbricata*	Dryandra, WA		R22	DNA extract from DT	Msi ML
22	66351	*Morelia spilota imbricata*	Garden Island, WA		R22	DNA extract from DT	Msi GI
22	67136	*Morelia spilota imbricata*	West Wallabi Island, WA		R22	DNA extract from DT	Msi ML

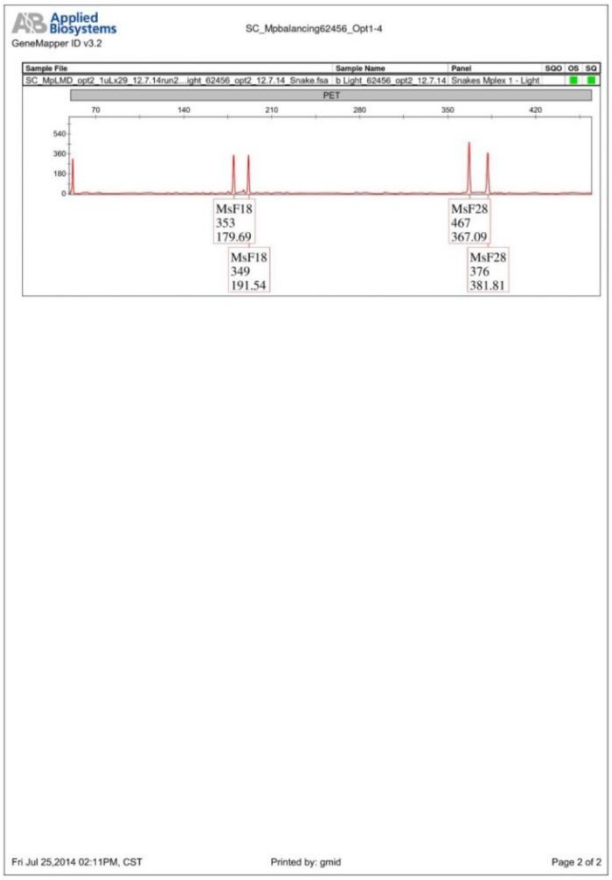
Complete loci typed	ABTC #	Sample Name	Provenance	Voucher #	Use	Sample type	Reference population
22	68304	*Morelia spilota*	St Francis Island, SA		R22	Skin	Msi SFI
22	68341	*Morelia spilota*	Buckleboo, SA		R22	Scale clipping	Msi ML
22	79114	*Morelia spilota*	3.6k N Cambridge Plateau, Richmond Range, NSW		R22	DNA extract from DT	SEC
22	79116	*Morelia spilota*	Lake Eacham, Atherton Tableland, Qld		R22	DNA extract from DT	NEC
22	81307	*Morelia spilota*	17k W Nevertire, NSW	AMSR155203	R22	Muscle	SIB
22	81446	*Morelia bredli*	Alice Springs, NT		R22	DNA extract from DT (skin)	SIB
21	66113	*Morelia spilota*	Black Mountain, Qld		Unkn	DNA extract from DT	
21	66171	*Morelia spilota imbricata*	St Francis Island, SA		Unkn	DNA extract from DT	
20	13429	*Morelia spilota variegata*	Fanny Bay, Darwin, NT		Unkn	DNA extract from DT (scale)	
20	68344	*Morelia spilota*	Byabarra, NSW		Unkn	DNA extract from DT (skin)	
19	66176	*Morelia spilota imbricata*	St Francis Island, SA		Unkn	DNA extract from DT	
19	67148	*Morelia spilota imbricata*	Dryandra, WA		Unkn	DNA extract from DT	
19	71370	*Morelia spilota*	Mt Hope, Vic		Unkn	DNA extract from DT (skin)	
19	71409	*Morelia spilota*	Piambie, Vic		Unkn	Skin	
19	83875	*Morelia spilota*	Kinchega NP, NSW		Unkn	Skin	
18	13432	*Morelia spilota variegata*	Botanic Gardents, Darwin, NT		Unkn	DNA extract from DT (skin)	
18	13459	*Morelia spilota variegata*	Darwin, NT		Unkn	DNA extract from DT (scale)	
17	13457	*Morelia spilota variegata*	Darwin, NT		Unkn	DNA extract from DT (scale)	
17	66031	*Morelia spilota*	Toowoomba, Qld		Unkn	DNA extract from DT	
17	79112	*Morelia spilota*	Brisbane, Somerset Dam, Qld		Unkn	DNA extract from DT	
16	17537	*Morelia spilota*	3k East of Morgan, SA		Unkn	DNA extract from DT (RP?)	
16	50868	*Morelia spilota*	Bobbie Moonga WH, SA		Unkn	Skin	
15	81309	*Morelia spilota*	Saratoga, NSW	AMSR157291	Unkn	DNA extract from DT	
13	66334	*Morelia spilota imbricata*	North Bannister, WA		Unkn	DNA extract from DT	
10	12150	*Morelia spilota*	Top end, NT		Unkn	DNA extract from DT (blood)	
10	66337	*Morelia spilota imbricata*	Boyogin Rock Nat Res, WA		Unkn	DNA extract from DT	
9	66317	*Morelia spilota*	Cedar Creek Falls, Mt Tamborine, Qlc		Unkn	Scale clipping	
9	68380	*Morelia spilota*	2k Nth Clare, SA	SAMAR52665	Unkn	DNA extract from DT (liver)	
8	12154	*Morelia bredli*	Todd River, Alice Springs, NT		Unkn	DNA extract from DT (blood)	
8	81322	*Morelia spilota*	Myocum, NSW	AMSR153024	Unkn	DNA extract from DT	
7	71415	*Morelia spilota*	Piambie, Vic		Unkn	DNA extract from DT (skin)	
5	13444	*Morelia spilota*	Fanny Bay, Darwin, NT		Unkn	DNA extract from DT (skin)	
5	13450	*Morelia spilota*	Botanic Gardents, Darwin, NT		Unkn	DNA extract from DT (scale)	
5	86801	*Morelia spilota*	Paluma, Qld.		Unkn	Tissue	
2	50867	*Morelia spilota*	Bobbie Moonga WH, SA		-	DNA extract from DT (skin)	
2	102726	*Morelia spilota*	Mount Isa, Qld		-	Skin	
1	68278	*Morelia spilota*	Merauke, West Papua		-	Skin	
1	71408	*Morelia spilota*	Mt Hope, Vic		-	Skin	
0	12143	*Morelia bredli*	Alice springs, NT		-	DNA extract from DT (blood)	
0	12155	*Morelia bredli*	Todd River, Alice Springs, NT		-	DNA extract from DT (blood)	
0	13454	*Morelia spilota variegata*	Darwin, NT		-	DNA extract from DT (scale)	
0	68283	*Morelia spilota*	Batemans Bay, Rosedale Beach, NSW		-	DNA extract from DT (skin)	
0	68286	*Morelia spilota*	Cobaki, via Tweed Heads, NSW		-	Skin	
0	69348	*Morelia spilota*	West Papua		-	DNA extract from DT	
0	81661	*Morelia spilota*	Boondoomba, Qld	QMJ45510	-	Scale clipping	
0	81662	*Morelia spilota*	Windemere Station, 10k W Winton, Qld	QMJ60089	-	Muscle	
0	81663	*Morelia spilota*	Dunlow, 32k SE Emerald, Qld	QMJ67356	-	Muscle	
24 - QA compromised	55457	*Morelia spilota*	St Francis Island, SA		Unkn	DNA extract from DT (blood)	
24 - no locality	68807	*Morelia bredli*	Locality unknown	SAMAR52604	-	Liver	
24 - no locality	71373	*Morelia spilota*	Vic		-	DNA extract from DT	
24 - no locality	75582	*Morelia spilota metcalfei*	Locality unknown		-	DNA extract from DT (skin)	
24 - no locality	80508	*Morelia spilota*	Qld		-	DNA extract from DT	
24 - no locality	80509	*Morelia spilota*	Qld		-	DNA extract from DT	
24 - captive	17552	*Morelia spilota cheynei*	Captive		-	DNA extract from DT (skin)	
24 - captive	68294	*Morelia spilota*	Townsville, captive bred		-	DNA extract from DT (skin)	
24 - captive	75756	*Morelia spilota spilota*	Captive bred		-	DNA extract from DT (skin)	
23 - no locality	121830	*Morelia spilota*	Locality unknown		-	Scale clipping	
24 - QA duplicate	71397	*Morelia spilota*	Walpella Island, Vic		-	DNA extract from DT	
24 - QA duplicate	71407	*Morelia spilota*	Mt Hope, Vic		-	DNA extract from DT	
24 - QA duplicate	81301	*Morelia spilota*	Atherton, Qld		-	DNA extract from DT	
24 - QA duplicate	81306	*Morelia spilota*	Toonumbar NP, NSW	AMSR153863	-	DNA extract from DT	
24 - QA duplicate	81450	*Morelia bredli*	Alice Springs Desert Park, NT		-	DNA extract from DT (skin)	

Appendix 3.6: First round optimisation results







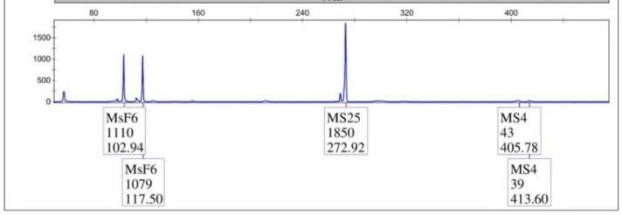


Applied Biosystems
GeneMapper ID v3.2

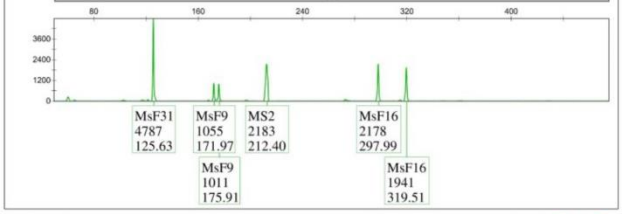
SC_Mpbalancing62456_Opt1-4

Sample File	Sample Name	Panel	SQO	OS	SQ
SC_MpLMD_Opt3_18.07.14_E10_Light_62456_Opt3_Snake.fsa	c Light_62456_Opt3	Snakes Mplex 1 - Light			

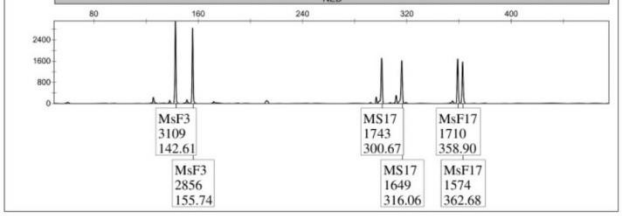
FAM



VIC



NED



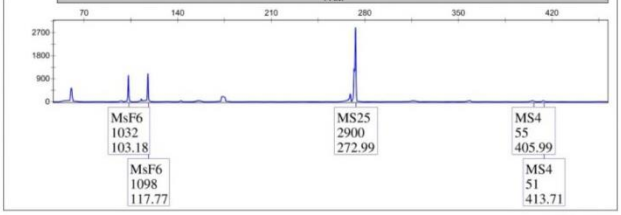
Fri Jul 25,2014 02:16PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

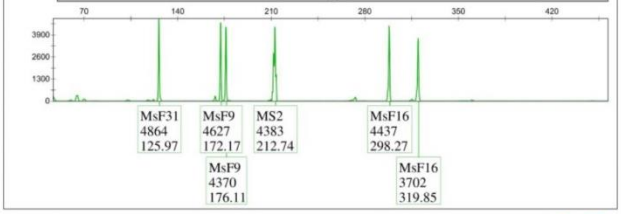
SC_Mpbalancing62456_Opt1-4

Sample File	Sample Name	Panel	SQO	OS	SQ
SC_MpLMD_Opt4_24.07.14_A01_Light_62456_Opt4_Snake.fsa	d Light_62456_Opt4	Snakes Mplex 1 - Light			

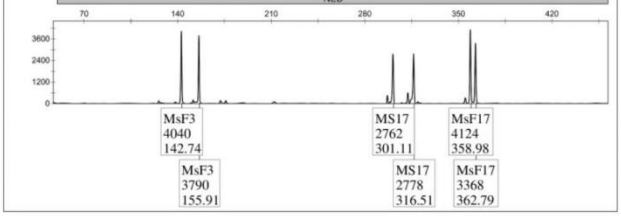
FAM



VIC



NED



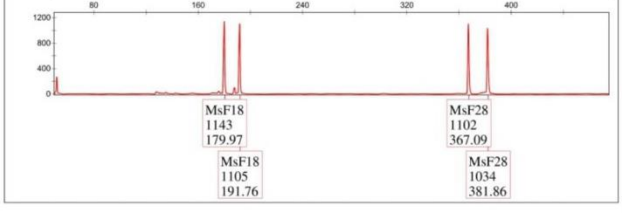
Fri Jul 25,2014 02:16PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

SC_Mpbalancing62456_Opt1-4

Sample File	Sample Name	Panel	SQO	OS	SQ
SC_MpLMD_Opt3_18.07.14_E10_Light_62456_Opt3_Snake.fsa	c Light_62456_Opt3	Snakes Mplex 1 - Light			

PET



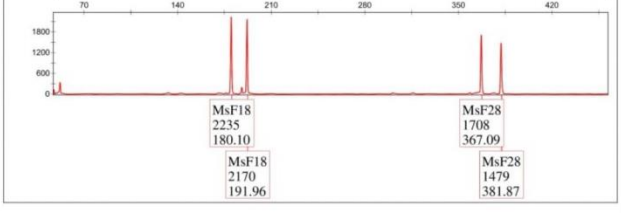
Fri Jul 25,2014 02:16PM, CST Printed by: gmid Page 2 of 2

Applied Biosystems
GeneMapper ID v3.2

SC_Mpbalancing62456_Opt1-4

Sample File	Sample Name	Panel	SQO	OS	SQ
SC_MpLMD_Opt4_24.07.14_A01_Light_62456_Opt4_Snake.fsa	d Light_62456_Opt4	Snakes Mplex 1 - Light			

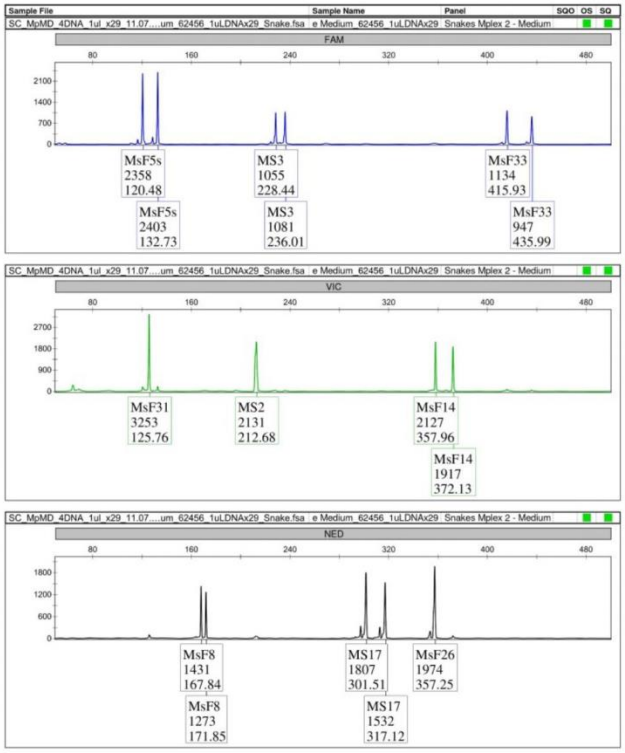
PET



Fri Jul 25,2014 02:16PM, CST Printed by: gmid Page 2 of 2

Applied Biosystems
GeneMapper ID v3.2

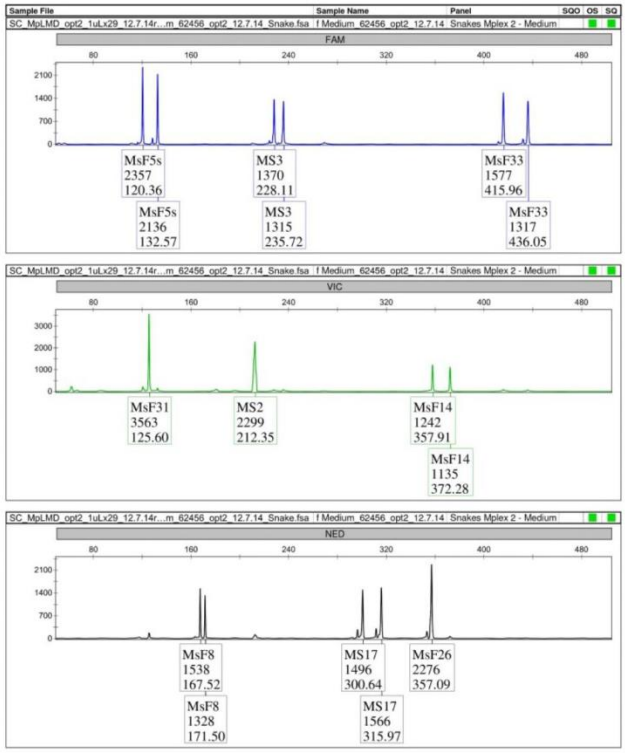
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:17PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

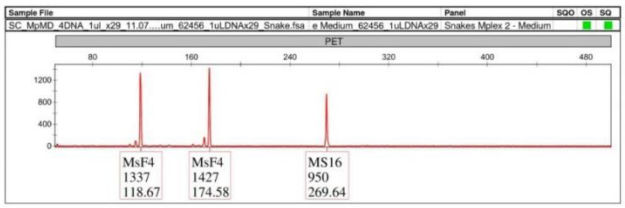
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:17PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

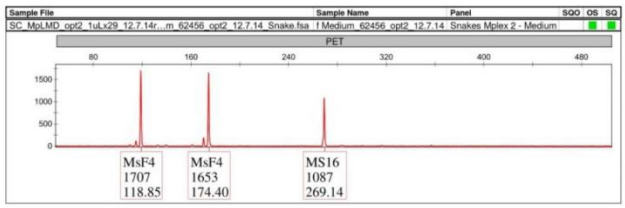
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:17PM, CST Printed by: gmid Page 2 of 2

Applied Biosystems
GeneMapper ID v3.2

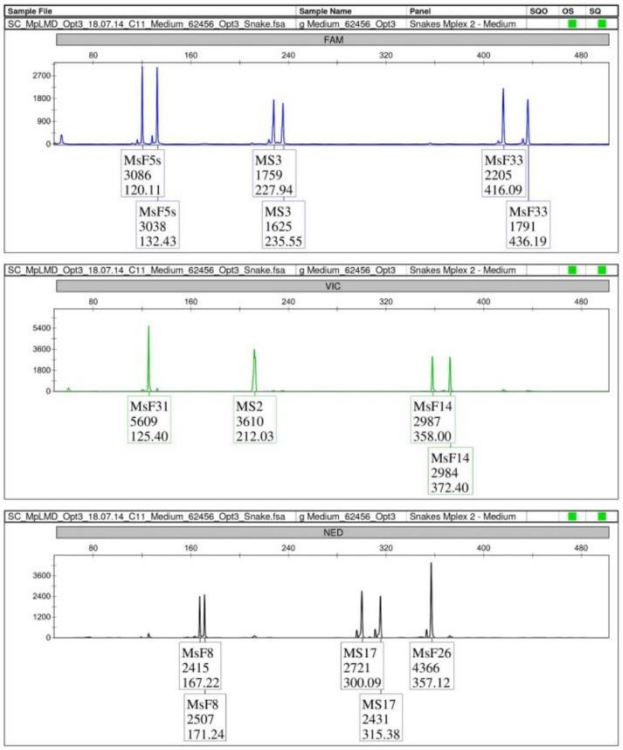
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:17PM, CST Printed by: gmid Page 2 of 2

Applied Biosystems
GeneMapper ID v3.2

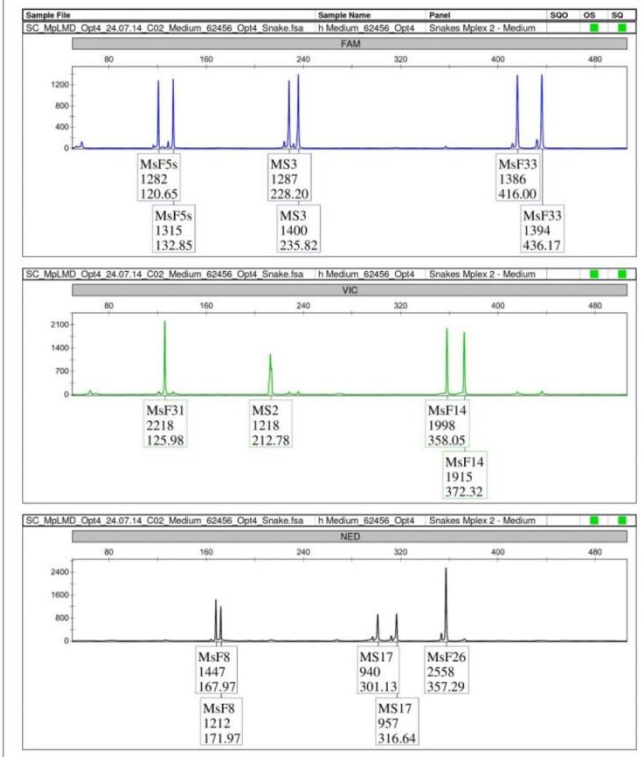
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:20PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

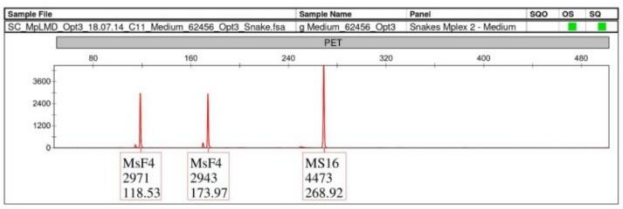
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:21PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

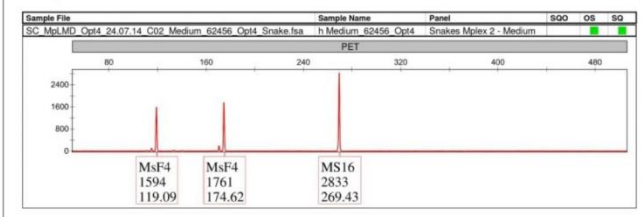
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:20PM, CST Printed by: gmid Page 2 of 2

Applied Biosystems
GeneMapper ID v3.2

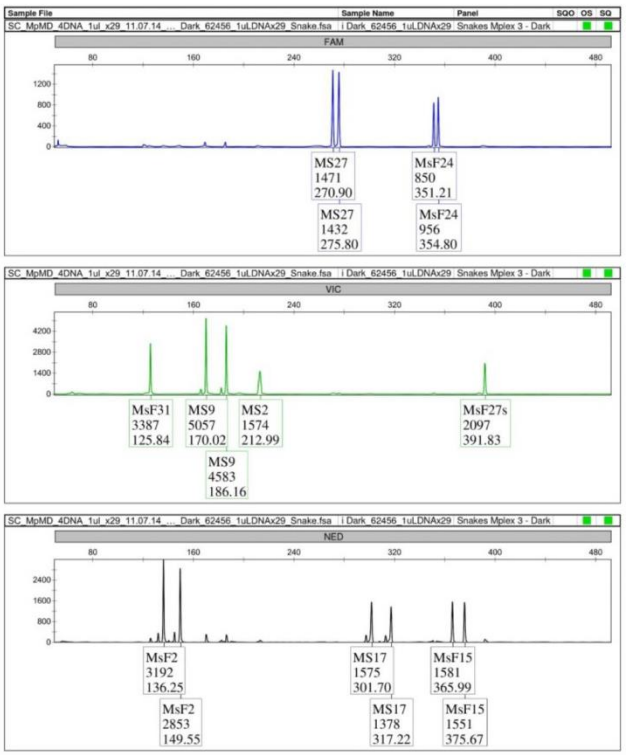
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:21PM, CST Printed by: gmid Page 2 of 2

Applied Biosystems
GeneMapper ID v3.2

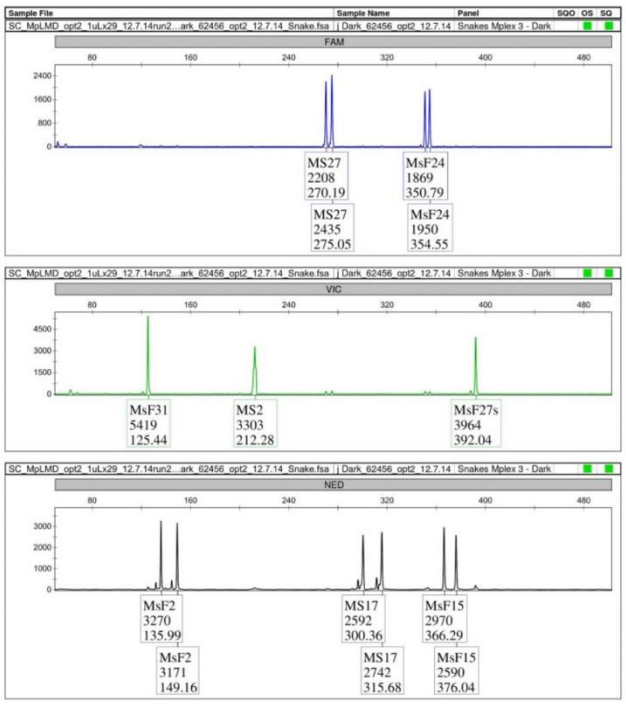
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:21PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

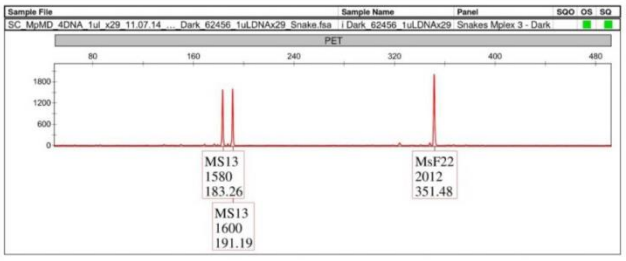
SC_Mpbalancing62456_Opt1-4



Fri Jul 25,2014 02:22PM, CST Printed by: gmid Page 1 of 2

Applied Biosystems
GeneMapper ID v3.2

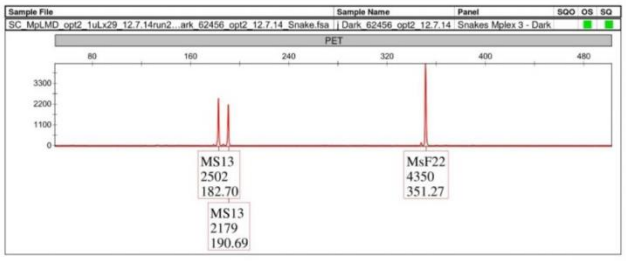
SC_Mpbalancing62456_Opt1-4



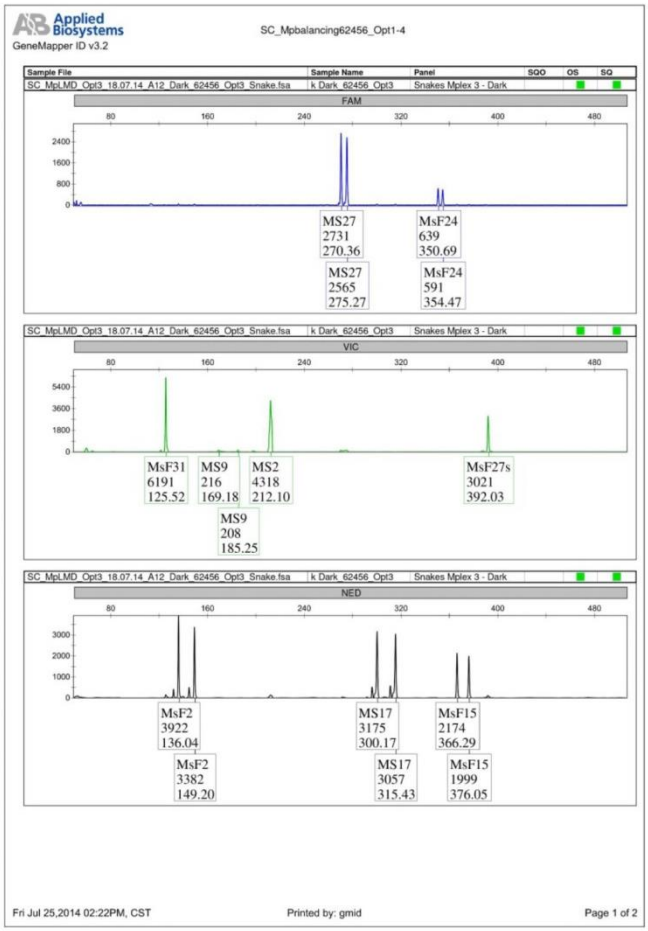
Fri Jul 25,2014 02:21PM, CST Printed by: gmid Page 2 of 2

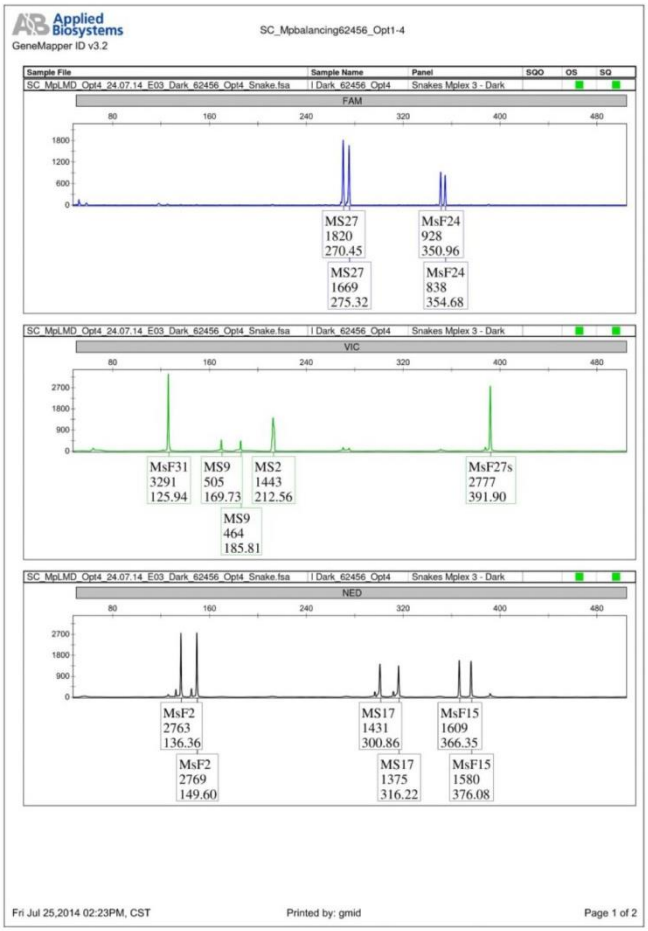
Applied Biosystems
GeneMapper ID v3.2

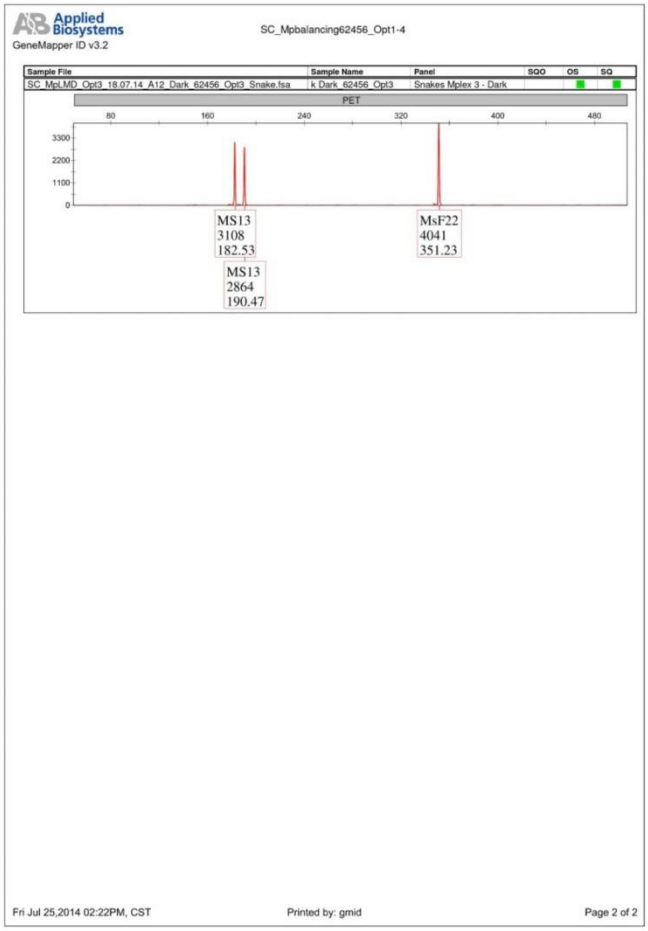
SC_Mpbalancing62456_Opt1-4

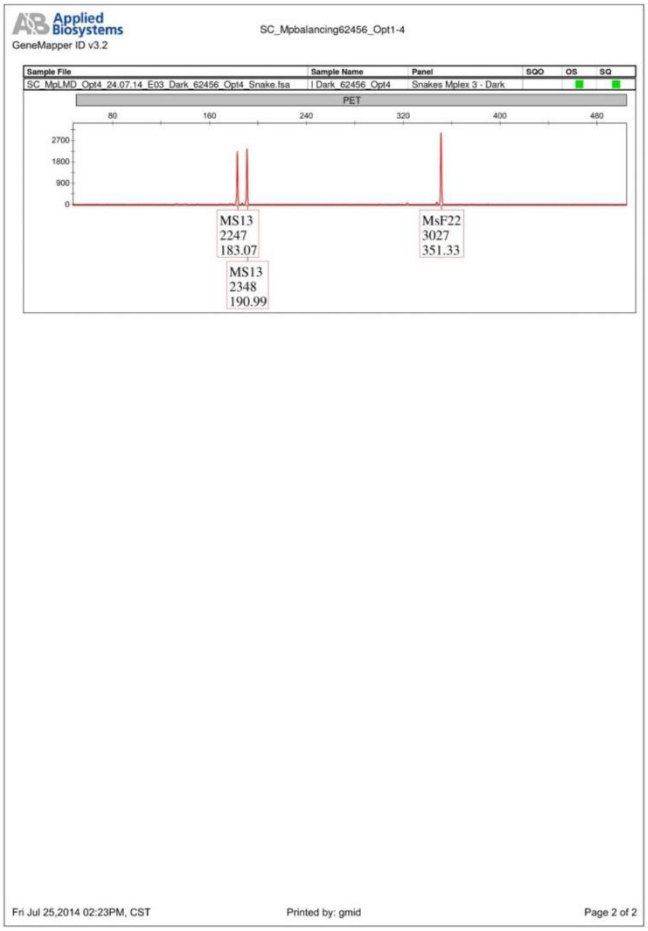


Fri Jul 25,2014 02:22PM, CST Printed by: gmid Page 2 of 2


Applied Biosystems
GeneMapper ID v3.2
SC_Mpbalancing62456_Opt1-4
Sample File
Sample Name
Panel
SQO
OS
SQ
SC_MpLMD_Opt3_18.07.14_A12_Dark_62456_Opt3_Snake.fsa
k Dark_62456_Opt3
Snakes Mplex 3 - Dark
FAM
MS27 2731 270.36
MsF24 639 350.69
MS27 2565 275.27
MsF24 591 354.47
VIC
MsF31 6191 125.52
MS9 216 169.18
MS2 4318 212.10
MsF27s 3021 392.03
MS9 208 185.25
NED
MsF2 3922 136.04
MS17 3175 300.17
MsF15 2174 366.29
MsF2 3382 149.20
MS17 3057 315.43
MsF15 1999 376.05
Fri Jul 25,2014 02:22PM, CST
Printed by: gmid
Page 1 of 2

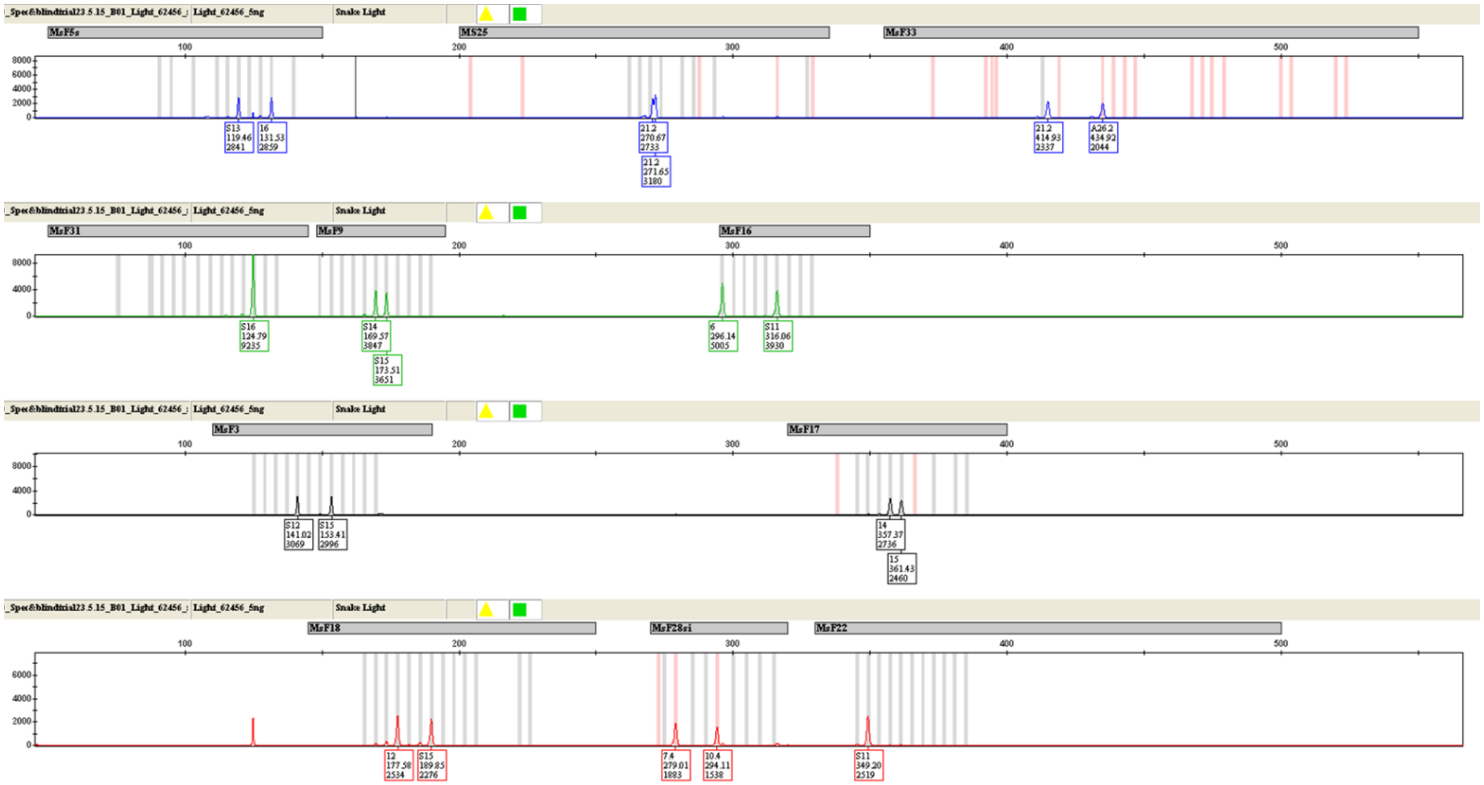

Applied Biosystems
GeneMapper ID v3.2
SC_Mpbalancing62456_Opt1-4
Sample File
Sample Name
Panel
SQO
OS
SQ
SC_MpLMD_Opt4_24.07.14_E03_Dark_62456_Opt4_Snake.fsa
l Dark_62456_Opt4
Snakes Mplex 3 - Dark
FAM
MS27 1820 270.45
MsF24 928 350.96
MS27 1669 275.32
MsF24 838 354.68
VIC
MsF31 3291 125.94
MS9 505 169.73
MS2 1443 212.56
MsF27s 2777 391.90
MS9 464 185.81
NED
MsF2 2763 136.36
MS17 1431 300.86
MsF15 1609 366.35
MsF2 2769 149.60
MS17 1375 316.22
MsF15 1580 376.08
Fri Jul 25,2014 02:23PM, CST
Printed by: gmid
Page 1 of 2


Applied Biosystems
GeneMapper ID v3.2
SC_Mpbalancing62456_Opt1-4
Sample File
Sample Name
Panel
SQO
OS
SQ
SC_MpLMD_Opt3_18.07.14_A12_Dark_62456_Opt3_Snake.fsa
k Dark_62456_Opt3
Snakes Mplex 3 - Dark
PET
MS13 3108 182.53
MsF22 4041 351.23
MS13 2864 190.47
Fri Jul 25,2014 02:22PM, CST
Printed by: gmid
Page 2 of 2

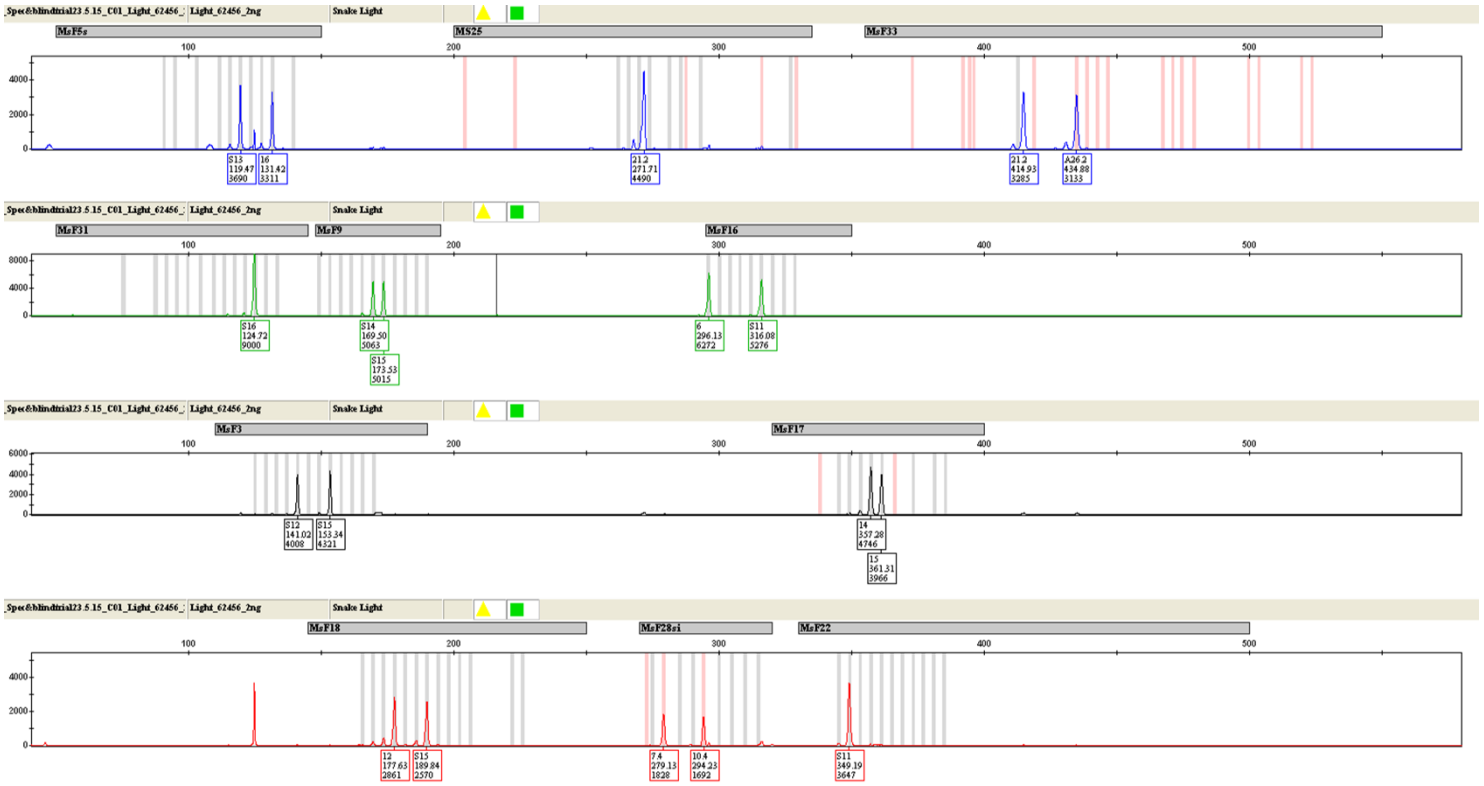

Applied Biosystems
GeneMapper ID v3.2
SC_Mpbalancing62456_Opt1-4
Sample File
Sample Name
Panel
SQO
OS
SQ
SC_MpLMD_Opt4_24.07.14_E03_Dark_62456_Opt4_Snake.fsa
l Dark_62456_Opt4
Snakes Mplex 3 - Dark
PET
MS13 2247 183.07
MsF22 3027 351.33
MS13 2348 190.99
Fri Jul 25,2014 02:23PM, CST
Printed by: gmid
Page 2 of 2

Appendix 3.7: Sensitivity study Light multiplex results

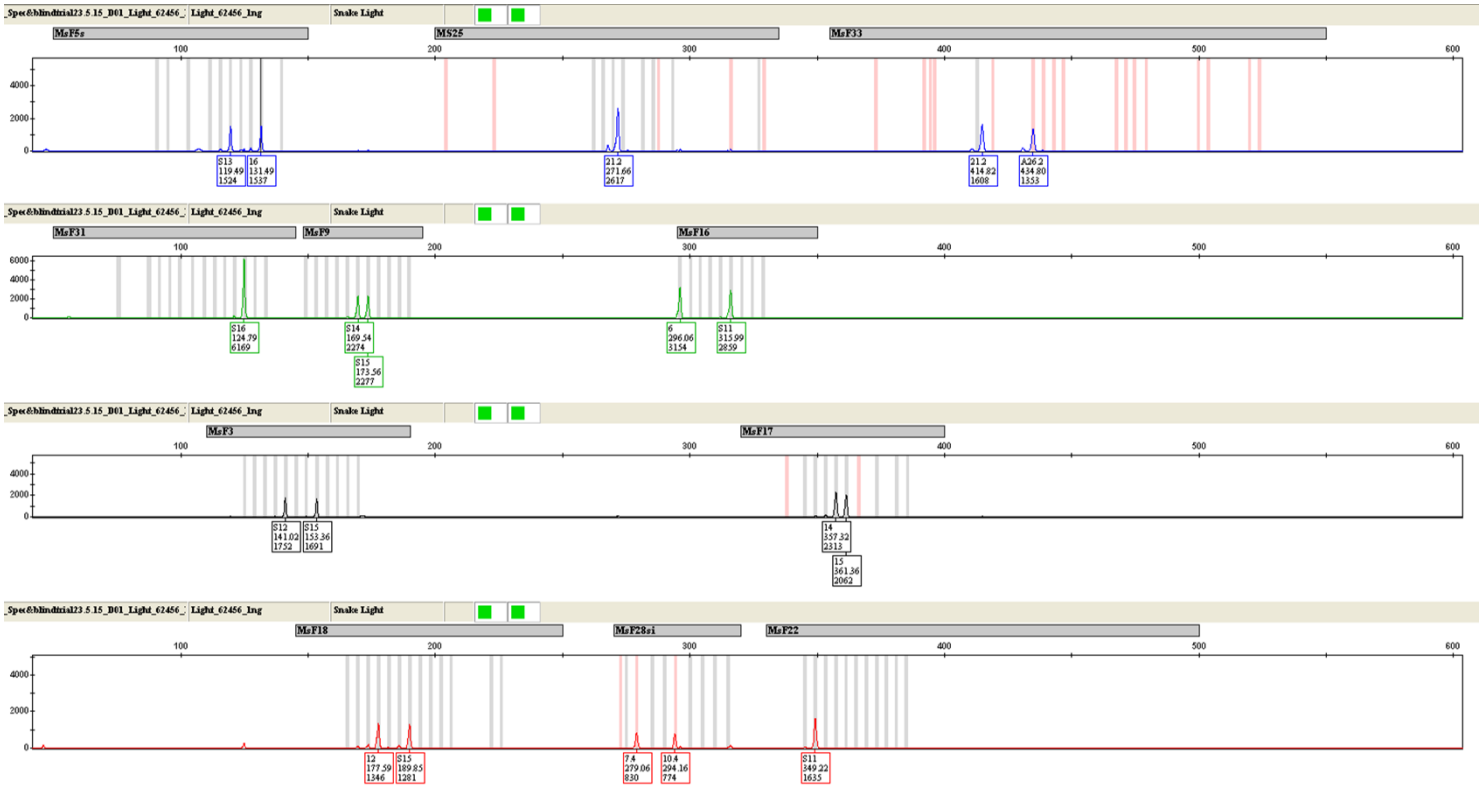
5 ng



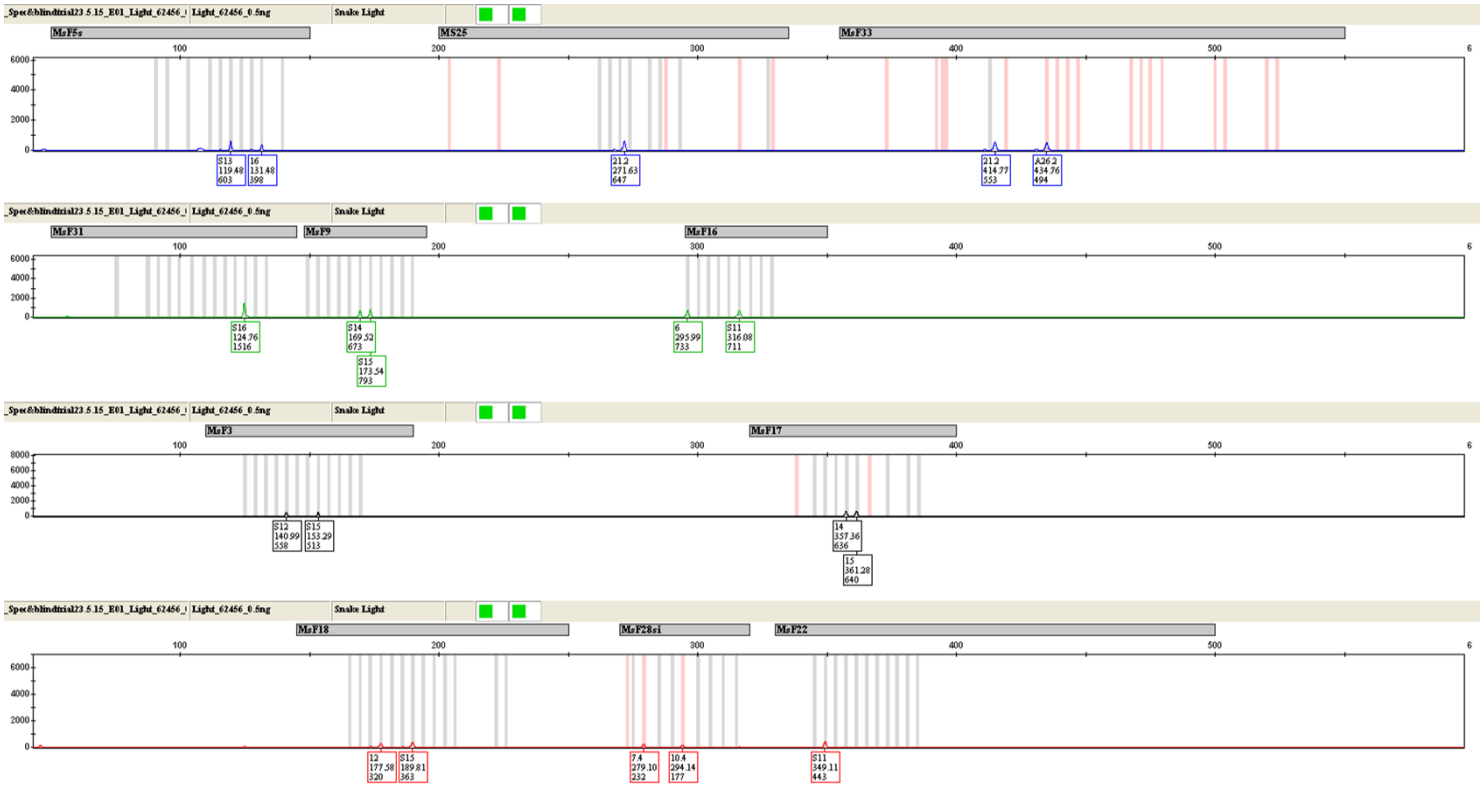
2 ng



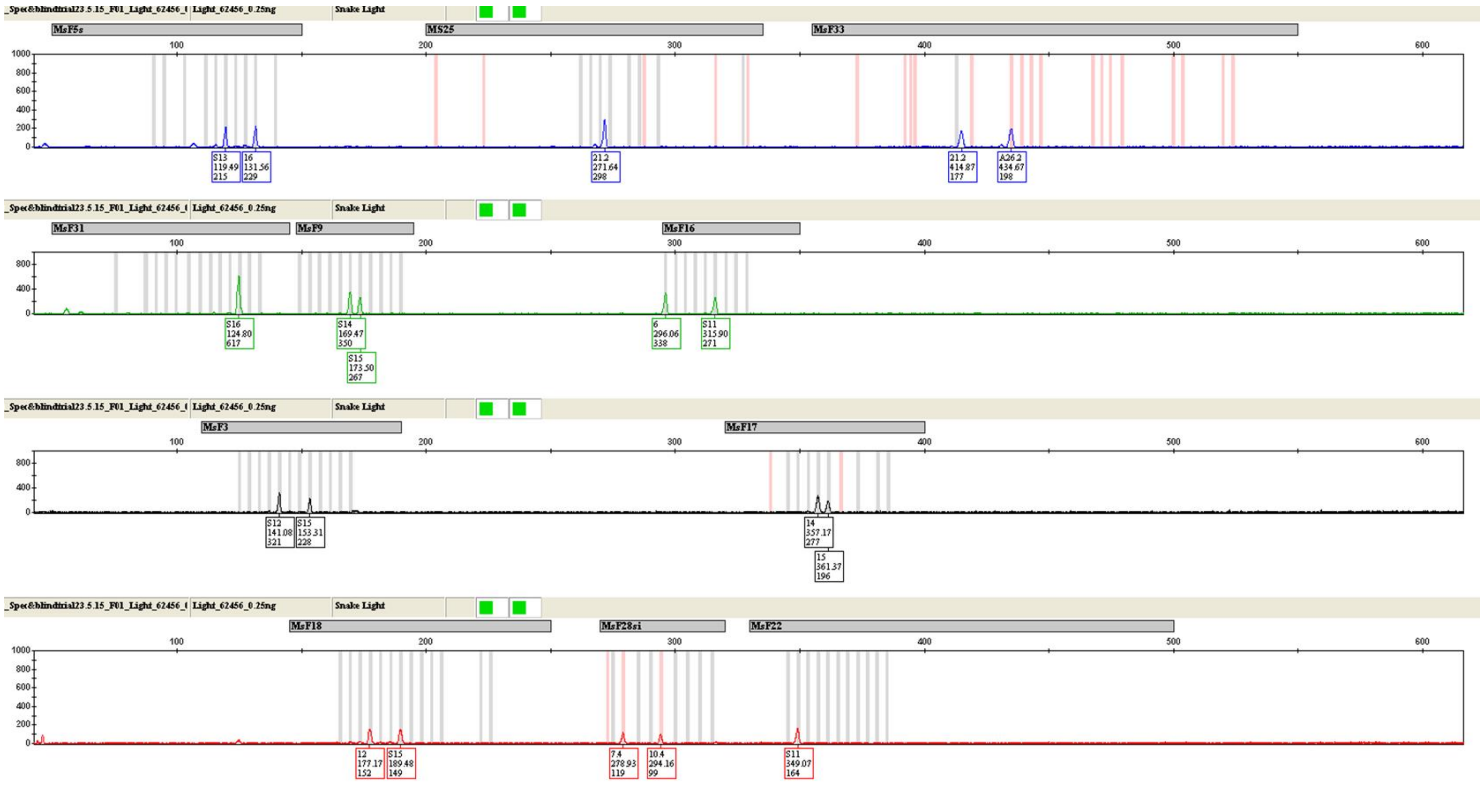
1 ng


MsF5s
MS25
MsF33
MsF31
MsF9
MsF16
MsF3
MsF17
MsF18
MsF28si
MsF22

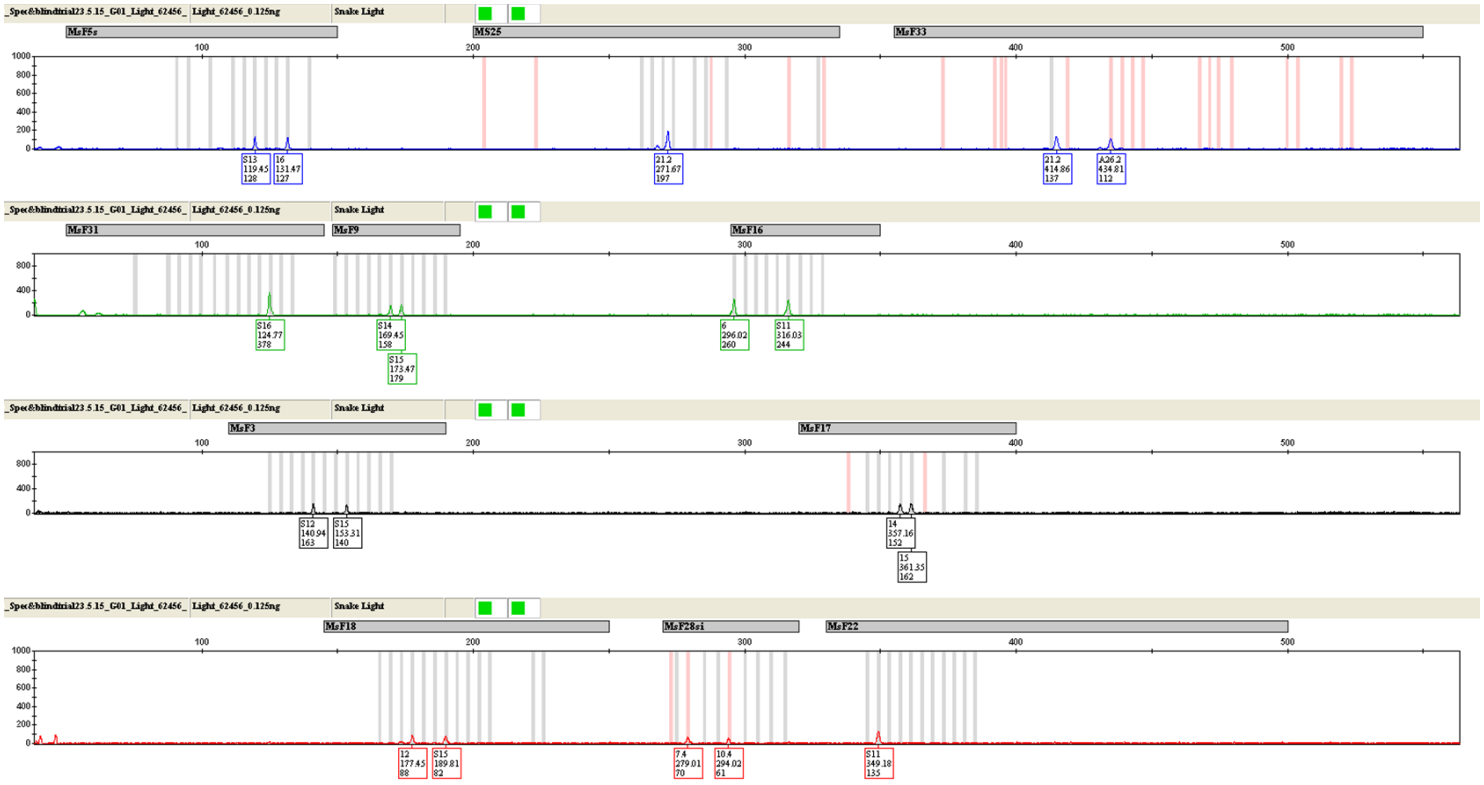
0.5 ng


MsF5s
MS25
MsF33
MsF31
MsF9
MsF16
MsF3
MsF17
MsF18
MsF28si
MsF22

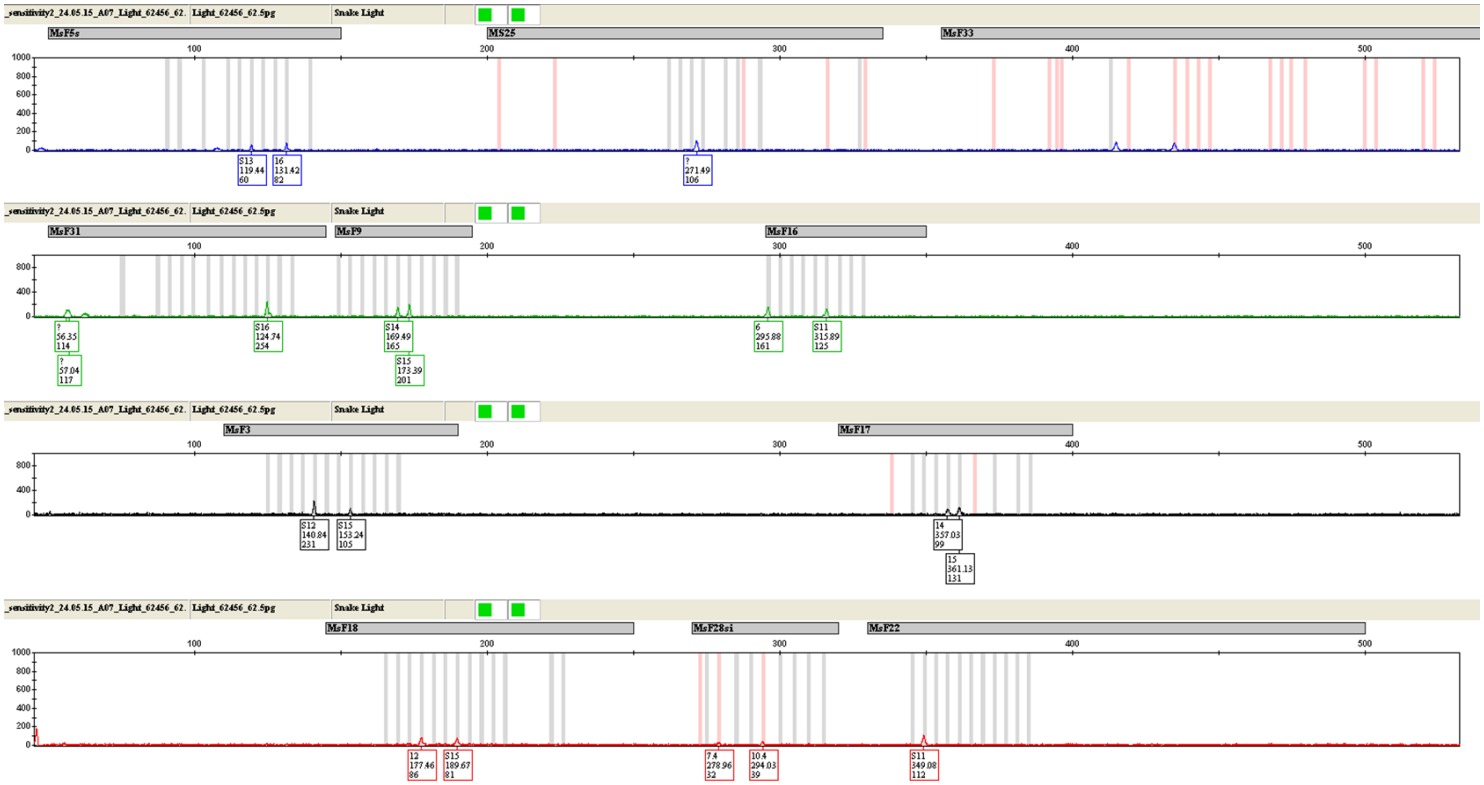
0.25 ng



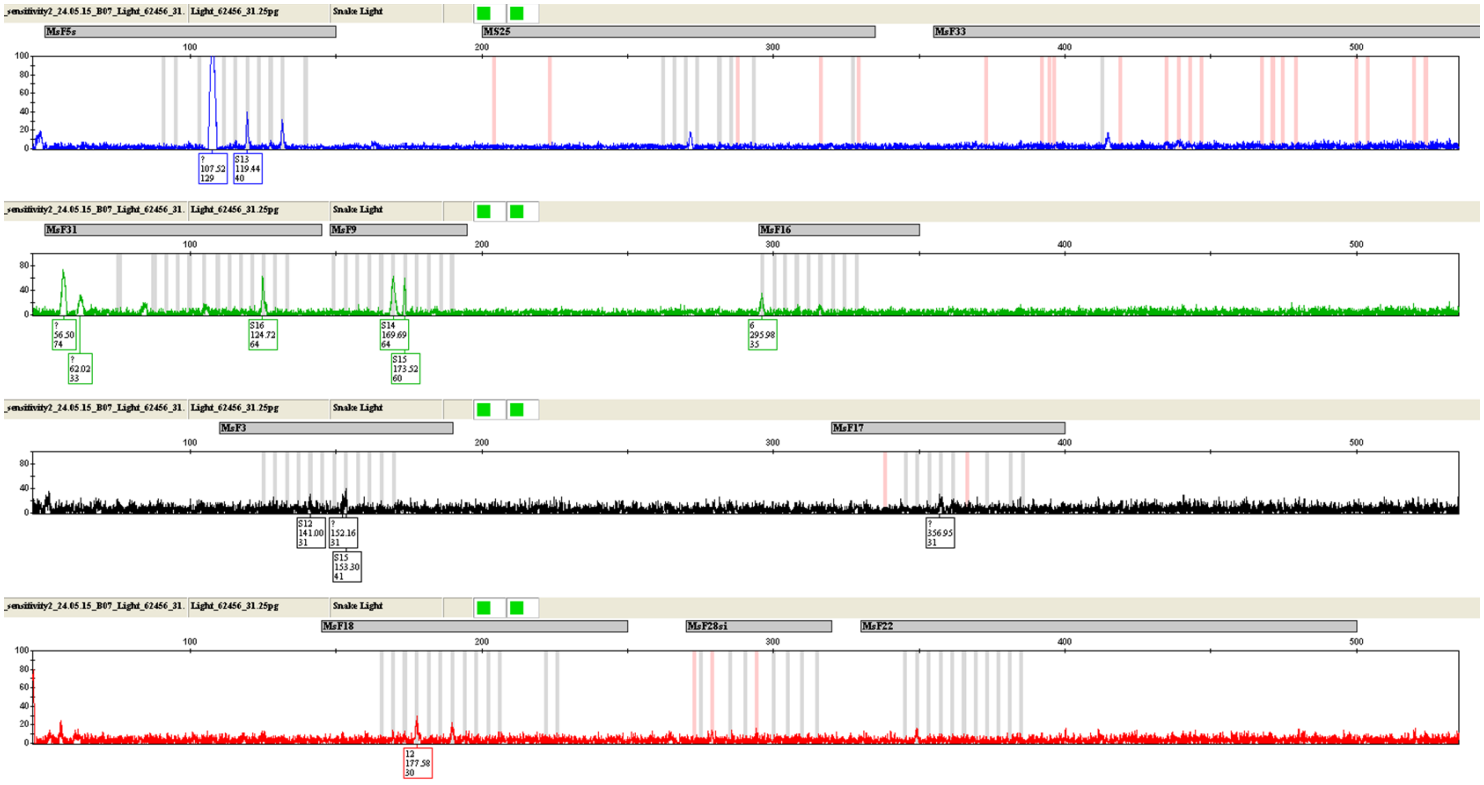
0.125 ng



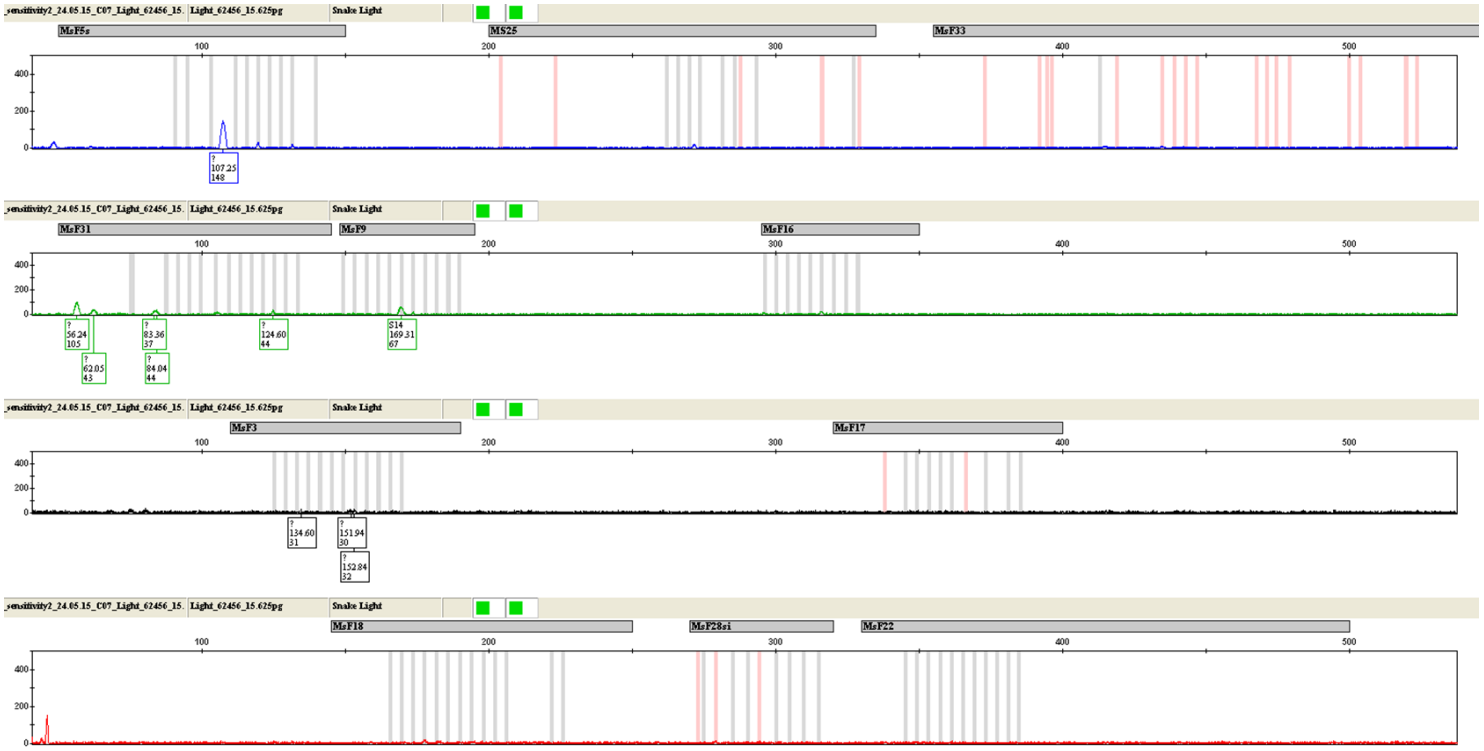
62.5 pg



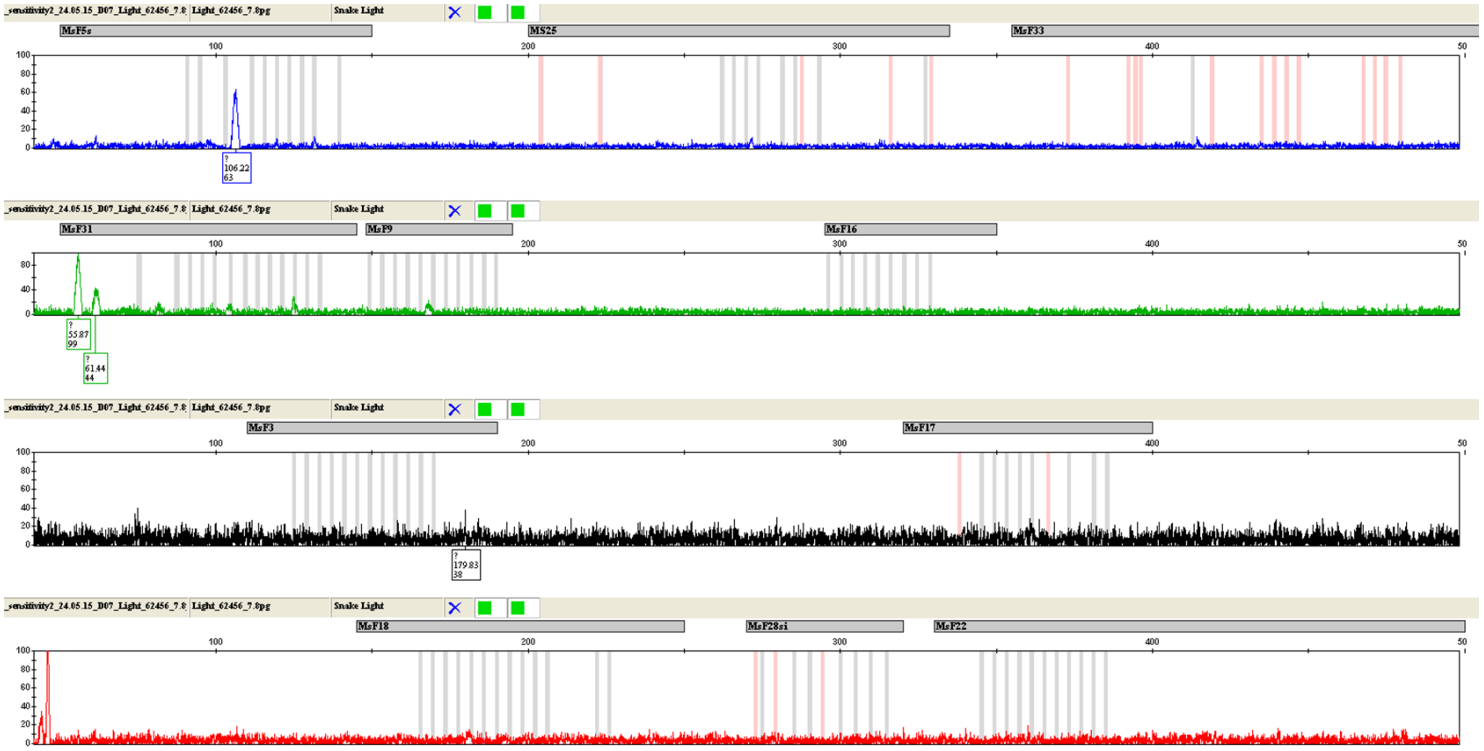
31.25 pg



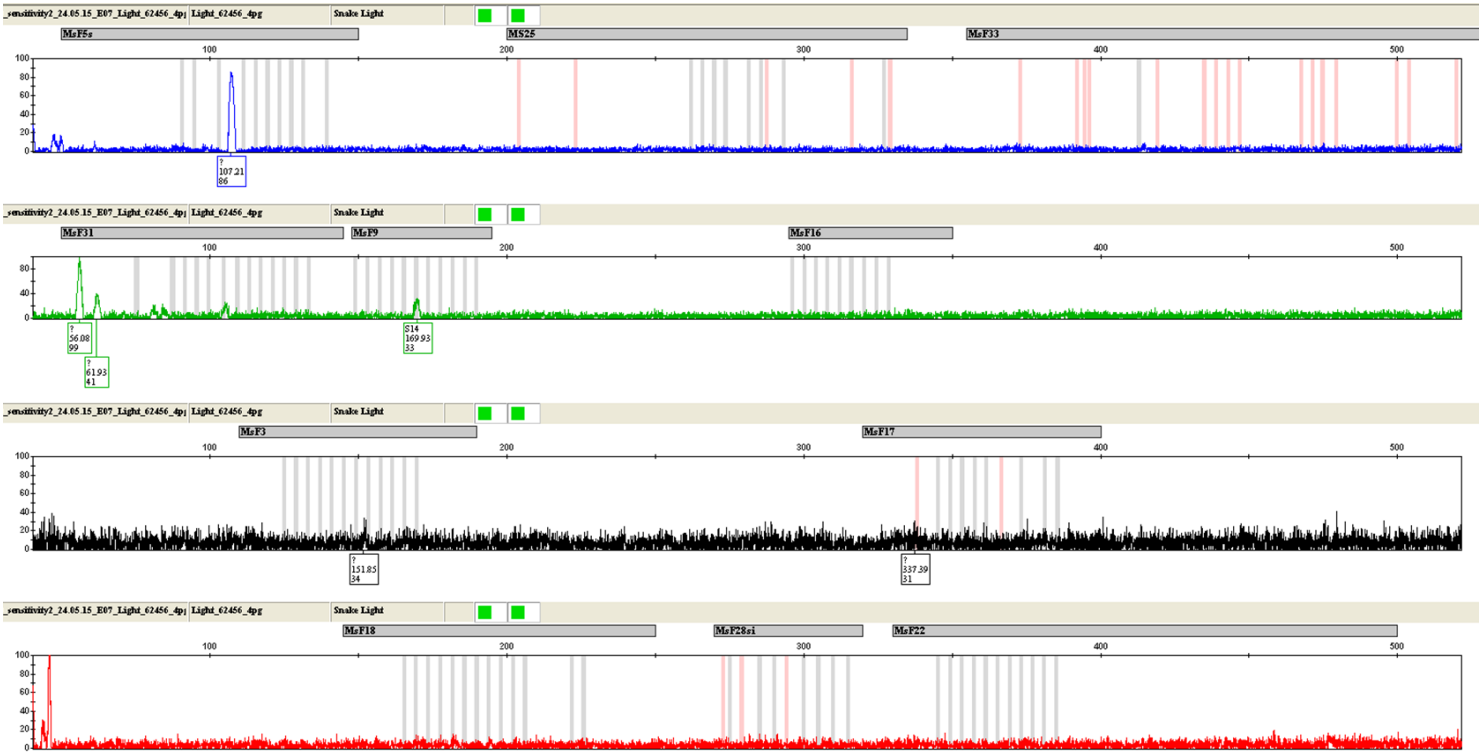
15.6 pg


MsF5s
MS25
MsF33
MsF31
MsF9
MsF16
MsF3
MsF17
MsF18
MsF28si
MsF22

7.8 pg


MsF5s
MS25
MsF33
MsF31
MsF9
MsF16
MsF3
MsF17
MsF18
MsF28si
MsF22

4 pg


_sensitivity2_24.05.15_E07_Light_62456_4pg
Light_62456_4pg
Snake Light
MsF5s
MS25
MsF33
MsF31
MsF9
MsF16
MsF3
MsF17
MsF18
MsF28si
MsF22

Appendix 3.8: STR sequence results

Locus	Accession #	Individual ID	Allele by migration	Locality	Sequence
MS2	AF403194	ABTC55500	*	Goyder Lagoon, SA	$(CTTT)(CTTTTT)(CTTT)_{18}$
Complex		ABTC34289	*	Blanchetown, SA	$(CTTT)(CTTTTT)(CTTT)_3(CT)_2(CTTT)_9(CT)(CTTT)_{12}(CT)_{13}$
		ABTC62456	*	Mount Connor, WA	$(CT)(CTTT)_9(CCTT)_5T(CT)_4(CCTT)(CT)_{12}$
		ABTC68304	*	St Francis Isl., SA	$(CT)_9(CTTT)(CT)_2(CTTT)(CT)_3(CTTT)_2(CT)_{14}$
		ABTC68310	*	NTP Isl., WA	$(CT)_3(CTTT)_3(CT)_2(CTTT)(CT)_{12}(CTTT)_3(CT)_{10}(CTTT)(CT)_3(CTT)$ $(CTTT)(CT)_{14}$
		ABTC117064	*	Merauke, West Papua, PNG	$(CT)_{36}(CTTT)$ $(CT)_{15}(CTTT)_4(CT)_{19}$
MS3		ABTC50868	17	Warburton Range, SA	$(GAAA)_{17}$
Complex	AF403195	ABTC55500	15	Goyder Lagoon, SA	$(GAAA)_{15}$
	MG548390	ABTC68341	19	Buckleboo, SA	$(GAAA)(GAA)(GAAA)_{18}$
		Unknown	#	Unknown	$(GAAA)_{18}$
		Unknown	#	Unknown	$(GAAA)_2(GAA)(GAAA)(GAA)_2(GAAA)_{12}$
MS4	AF403196	ABTC55500	*	Goyder Lagoon, SA	$(AAGA)_{19}$
$(AAGA)_n$	MG548391	A002	*		$(AAGA)_{20}$
Simple	MG548392	A003	*		$(AAGA)_{18}$
MS9	AF403201	ABTC55500	18	Goyder Lagoon, SA	$(TTTC)_{18}$
$(TTTC)_n$		ABTC34289	18	Blanchetown, SA	$....TC(TTTC)_{12}$
SImple		A003	Unknown		$....C(TTTC)_{12}$
	MG548393	ABTC76530	16	Warburton Range, SA	$(TTTC)_{16}$
	MG548394	ABTC117064	18	Merauke, West Papua, PNG	$(TTTC)_{18}$
MS13	MG548395	A003	*		$(CTTT)_4CT(CTTT)_7(CTT)$
Complex	MG548396	ABTC34289	*	Blanchetown, SA	$(CTTT)_3CT(CTTT)_5(CTTTTT)(CTTT)_3(CTT)$
	AF403205	ABTC55500	*	Goyder Lagoon, SA	$(CTTC)(CTTT)_{10}(CTT)$
		ABTC76268	*		$(CTTT)_3CT(CTTT)_5(CTTTTT)(CTTT)_3(CTT)$
		ABTC68304	*		$(CTTT)_4CT(CTTT)_7(CTT)$
		ABTC117070	*		$(CTTT)_8CT(CTTT)_9(CTT)_3$
MS16	AF403208	ABTC55500	*	Goyder Lagoon, SA	$(CAAT)_8CAAGAAAGAAAATAG(AAAG)_{11}$
Complex			*		
MS17	AF403209	ABTC55500	*	Goyder Lagoon, SA	$TTTTCCTCCCT(CTTT)_3CT(CTTT)_{19}$
Complex		ABTC68308	*	Mondrain Isl., WA	$....(CCCT)_7CCTT(CTTT)_{30}$
	MG548397	ABTC83875	*	Kinchega Nat. Park, NSW	$TTTTCCTCCCT(CTTT)_9$
	MG548398	ABTC117064	*	Merauke, West Papua, PNG	$TTCTCCCCT(TCCT)_{18}(CT)_3CTTT(CT)_8(CTTT)_{19}$

Locus	Accession #	Individual ID	Allele by migration	Locality	Sequence
MS25	AF403217	ABTC55500	*	Goyder Lagoon, SA	$(GA)_2(GAA)(GAAA)_{24}$
$(GAAA)_n(GA)_n$	MG548399	ABTC62456	*	Mount Connor, WA	$(GA)_2(GAA)(GAAA)_{22}$
Compound/Complex?	MG548400	ABTC68310	*	NTP Isl., WA	$(AA)(GA)_2(GAA)(GGAA)(GAAA)_{24}$
	MG548401	ABTC71378	*	Piambie, Victoria	$(GA)_2(GAA)(GAAA)_{23}$
	MG548402	ABTC117064	*	Merauke, West Papua, PNG	$(GA)_2(GAA)(GAAA)_6(GA)_{61}$
MS27	AF403219	ABTC55500	*	Goyder Lagoon, SA	$(GA)_{11}(GAAA)_2(GAAAA)(GGAA)_3(GGA)(GAAA)_4GA(GAAA)_7(GAAAAA)(GAAA)(GAAAAA)$
Complex	MG548403	ABTC55482	*	Darwin, NT	$(GA)_7(GAAA)(GGAA)_2A(GAAA)_3(GGAA)_2A(GAAA)_2(GGAA)A(GAAA)_3(GAAAAA)GA(GAAA)_2(GAAAAA)(GAAA)_3(GAAAAA)(GAAA)_2(GAAAAA)(AAA)$
	MG548404	ABTC66267	*	Seaham, NSW	$(GA)_8(GAAA)_2(GGAA)G(GAAA)_2(GAAG)(GGAA)_3A(GAAA)_2(GAAAAA)(GAAA)_4(GAAAAA)_2$
	MG548405	ABTC68304	*	St Francis Isl., SA	$(GA)_5(GAAA)_2(GGAA)_2A(GAAA)_3(GGAA)_5A(GAAA)_4(GAAAAA)GA(GAAA)_2(GAAAAA)(GAAA)(GA)(GAAA)_4(GAAAAA)_2$
	MG548406	ABTC68310	*	NTP Isl., WA	$(GA)_9(GAAA)_2(GGAA)G(GAAA)_3(GGAA)(GAAG)_3(GAAA)_4(GAAAAA)(GA)(GAAA)_2(GAAAAA)(GAAA)(GA)(GAAA)_4(GAAAAA)_2$
	MG548407	ABTC71378	*	Piambie, Victoria	$(GA)_{10}(GAAA)_3A(GAAG)_4(GAAA)(GAAAAA)(GAAA)_2(GA)(GAAA)_3(GAAAAA)(GAAA)(GAAAAA)$
	MG548408	ABTC76268	*	Sturt Nat. Park, NSW	$(GA)_{11}(GAAA)_3A(GAAG)_4(GAAA)(GAAAAA)(GAAA)_2(GA)(GAAA)_4(GAAAAA)(GAAA)(GAAAAA)$
MsF2	MG548409	ABTC30087	16	Melville Isl., NT	$(AGAT)_{16}$
$(AGAT)_n$	MG548410	ABTC55499	12	Goyder Lagoon, SA	$(AGAT)_{12}$
	MG548411	ABTC81028	18	East McIlwraith Range, Qld	$(AGAT)_7(AGAC)(AGAT)_{12}$
	MG548412	ABTC106074	16	East Waikerie, SA	$(AGAT)_{16}$
MsF3	MG548413	ABTC30087	16	Melville Isl., NT	$(ATGA)_{16}$
$(ATGA)_n$	MG548414	ABTC34289	21	Blanchetown, SA	$(ATGA)_{21}$
	MG548415	ABTC55479	15		$(ATGA)_{15}$
	MG548416	ABTC55499	15	Goyder Lagoon, SA	$(ATGA)_{15}$
	MG548417	ABTC62456	12	Mount Connor, WA	$(ATGA)_{12}$
	MG548418	ABTC68304	15	St Francis Isl., SA	$(ATGA)_{15}$
	MG548419	ABTC68308	12	Mondrain Isl., WA	$(ATGA)_{12}$
	MG548420	ABTC68310	11	NTP Isl., WA	$(ATGA)_{11}$
	MG548421	ABTC71389	17	Warby Ranges, Victoria	$(ATGA)_{17}$
	MG548422	ABTC76268	16	Sturt Nat. Park, NSW	$(ATGA)_{16}$
	MG548423	ABTC81659	14	Cannington Mine, Qld	$(ATGA)_{14}$

Locus	Accession #	Individual ID	Allele by migration	Locality	Sequence
MsF4	MG548424	A003	18		$(CTTT)_{18}$
$(CTTT)_n$	MG548425	ABTC30087	16	Melville Isl., NT	$(CTTT)_{16}$
	MG548426	ABTC55499	11	Goyder Lagoon, SA	$(CTTT)_{11}$
	MG548427	ABTC68280	7	Merauke, West Papua, PNG	$(CTTT)_{7}$
	MG548428	ABTC68310	12	NTP Isl., WA	$(CTTT)_{12}$
	MG548429	ABTC123504	21	Port Moresby, PNG	$(CTTT)_{21}$
MsF5	MG548430	ABTC30087	15	Melville Isl., NT	$(TAGA)_{15}$
$(TAGA)_n$	MG548431	ABTC50868	13	Warburton Range, SA	$(TAGA)_{13}$
	MG548432	ABTC51576	14	Norseman, WA	$(TAGA)_{14}$
	MG548433	ABTC68310	18	NTP Isl., WA	$(TAGA)_{18}$
MsF8	MG548434	ABTC30087	14	Melville Isl., NT	$(GATG)_{14}$
$(GATG)_n$	MG548435	ABTC117064	12	Merauke, West Papua, PNG	$(GATG)_{12}$
	MG548436	ABTC123504	7	Port Moresby, PNG	$(GATG)_{7}$
MsF9	MG548437	ABTC30087	14	Melville Isl., NT	$(TATC)_{14}$
$(TATC)_n$	MG548438	ABTC67578	13	Cape Tribulation, Qld	$(TATC)_{13}$
	MG548439	ABTC68310	15	NTP Isl., WA	$(TATC)_{15}$
	MG548440	ABTC117064	16	Merauke, West Papua, PNG	$(TATC)_{16}$
	MG548441	ABTC123504	17	Port Moresby, PNG	$(TATC)_{17}$
MsF14	MG548442	ABTC30087	13.4	Melville Isl., NT	$(AGTT)(AAGTT)_{13}$
$(AAGTT)_n$	MG548443	ABTC51576	4	Norseman, WA	$(AAGTT)_{4}$
	MG548444	ABTC66267	15	Seaham, NSW	$(AAGTT)_{15}$
	MG548445	ABTC68280	10.4	Merauke, West Papua, PNG	$(AGTT)(AAGTT)_{10}$
MsF15	MG548446	ABTC30087	13	Melville Isl., NT	$(TTCTA)_{13}$
$(TTCTA)_n$	MG548447	ABTC55499	15	Goyder Lagoon, SA	$(TTCTA)_{15}$
	MG548448	ABTC66267	9	Seaham, NSW	$(TTCTA)_{9}$
	MG548449	ABTC76268	12	Sturt Nat. Park, NSW	$(TTCTA)_{12}$
	MG548450	ABTC81307	19	West Nevertire, NSW	$(TTCTA)_{19}$
	MG548451	ABTC117064	8	Merauke, West Papua, PNG	$(TTCTA)_{8}$
MsF16	MG548452	ABTC30087	12	Melville Isl., NT	$(CATT)_{12}$
$(CATT)_n$	MG548453	ABTC34289	10	Blanchetown, SA	$(CATT)_{10}$
	MG548454	ABTC51576	8	Norseman, WA	$(CATT)_{8}$
	MG548455	ABTC55499	12	Goyder Lagoon, SA	$(CATT)_{12}$
	MG548456	ABTC66267	11	Seaham, NSW	$(CATT)_{11}$
	MG548457	ABTC68280	9	Merauke, West Papua, PNG	$(CATT)_{9}$
	MG548458	ABTC68298	8	Fowlers Bay, SA	$(CATT)_{8}$
	MG548459	ABTC68310	8	NTP Isl., WA	$(CATT)_{8}$
	MG548460	ABTC71390	7	Gunbower Isl., Victoria	$(CATT)_{7}$
	MG548461	ABTC76530	12	Warburton Range, SA	$(CATT)_{12}$
	MG548462	ABTC106074	10	East Waikerie, SA	$(CATT)_{10}$
MsF17	MG548463	ABTC30087	13	Melville Isl., NT	$(GATA)_{13}$
$(GATA)_n$	MG548464	ABTC68304	8	St Francis Isl., SA	$(GATA)_{8}$
	MG548465	ABTC76225	17	Dog Fence Beach, SA	$(GATA)_{17}$
MsF18	MG548466	ABTC34289	13	Blanchetown, SA	$(GATA)_{13}$
$(GATA)_n$	MG548467	ABTC55499	13	Goyder Lagoon, SA	$(GATA)_{13}$
	MG548468	ABTC68285	15	Caparra, NSW	$(GATA)_{15}$
	MG548469	ABTC68293	19	Tilba, NSW	$(GATA)_{19}$
	MG548470	ABTC68308	17	Mondrain Isl., WA	$(GATA)_{17}$
	MG548471	ABTC68310	18	NTP Isl., WA	$(GATA)_{18}$
MsF22	MG548472	ABTC30087	13	Melville Isl., NT	$(ATCC)_{13}$
$(ATCC)_n$	MG548473	ABTC34289	13	Blanchetown, SA	$(ATCC)_{13}$
	MG548474	ABTC55479	13		$(ATCC)_{13}$
	MG548475	ABTC55499	11	Goyder Lagoon, SA	$(ATCC)_{11}$
	MG548476	ABTC68310	20	NTP Isl., WA	$(ATCC)_{20}$

Locus	Accession #	Individual ID	Allele by migration	Locality	Sequence
MsF24	MG548477	ABTC30087	12	Melville Isl., NT	$(AACC)_{12}$
$(AACC)_n$	MG548478	ABTC34289	7	Blanchetown, SA	$(AACC)_7$
	MG548479	ABTC51576	7	Norseman, WA	$(AACC)_7$
	MG548480	ABTC68293	11	Tilba, NSW	$(AACC)_{11}$
	MG548481	ABTC68310	7	NTP Isl., WA	$(AACC)_7$
MsF26	MG548482	ABTC30087	12	Melville Isl., NT	$(TAGA)_{12}$
$(TAGA)_n$	MG548483	ABTC55499	10	Goyder Lagoon, SA	$(TAGA)_{10}$
	MG548484	ABTC62456	16	Mount Connor, WA	$(TAGA)_{16}$
	MG548485	ABTC68304	11	St Francis Isl., SA	$(TAGA)_{11}$
	MG548486	ABTC68308	7	Mondrain Isl., WA	$(TAGA)_7$
	MG548487	ABTC68310	11	NTP Isl., WA	$(TAGA)_{11}$
	MG548488	ABTC112609	6	Trephina, NT	$(TAGA)_6$
MsF27	MG548489	ABTC30087	11	Melville Isl., NT	$(ATCT)_{11}$
$(ATCT)_n$	MG548490	ABTC55482	12	Darwin, NT	$(ATCT)_{12}$
	MG548491	ABTC62456	12	Mount Connor, WA	$(ATCT)_{12}$
	MG548492	ABTC71378	13	Piambie, Victoria	$(ATCT)_{13}$
MsF28	MG548493	ABTC30087	9.4 indel	Melville Isl., NT	$(TGATT)(TGATC)_{11}$ + indel
$(TGATC)_n$	MG548494	ABTC55499	9	Goyder Lagoon, SA	$(TGATC)_9$
	MG548495	ABTC66267	10	Seaham, NSW	$(TGATC)_{10}$
	MG548496	ABTC68308	11	Mondrain Isl., WA	$(TGATC)_{11}$
	MG548497	ABTC68310	12	NTP Isl., WA	$(TGATC)_{12}$
MsF31	MG548498	ABTC30087	14	Melville Isl., NT	$(TGAA)_{12}(TAA)(TAAA)$
$(TGAA)_n(TAAA)_n$	MG548499	ABTC51576	4	Norseman, WA	$(TGAA)(TAAA)_3$
Compound	MG548500	ABTC55479	5		$(TGAA)(TAAA)_4$
	MG548501	ABTC55482	15	Darwin, NT	$(TGAA)_{12}T$….
	MG548502	ABTC55500	6	Goyder Lagoon, SA	$(TGAA)(CAAA)(TAAA)_3$
	MG548503	ABTC62456	16	Mount Connor, WA	$(TGAA)_{11}(TAAA)_5$
	MG548504	ABTC68293	18	Tilba, NSW	$(TGAA)_{16}(TAAA)_2$
	MG548505	ABTC71378	9	Piambie, Victoria	$(TGAA)(TAAA)_8$
	MG548506	ABTC81028	11	East McIlwraith Range, Qld	$(TGAA)_8(TAAA)_3$
MsF33	MG548507	ABTC30087	16.3	Melville Isl., NT	$(GAGA)(TAGA)_2(TGA)(TAGA)_{13}$
$(GACA)_n(GAGA)(TAGA)_n$	MG548508	ABTC62456	21.2	Mount Connor, WA	$(GACA)_6(GAGA)(TAGA)_2(TGA)(TAGA)_{12}$
Compound	MG548509	ABTC66267	21	Seaham, NSW	$(GACA)_8(GAGA)(TAGA)_{16}$
	MG548510	ABTC76268	25.3	Sturt Nat. Park, NSW	$(GACA)_8(GAGA)(TAGA)_2(TGA)(TAGA)_{15+...}$

CHAPTER 4

STR Database Validation and Application

Appendices included in Chapter 4:

4.1 Preface

Chapter 3 involved the genotyping of carpet python samples collected across the native range of the species in order to construct genotype and allele frequency databases representative of Australasian native carpet pythons. The previous chapter concluded that the reference database of 249 individuals was not yet suitable for application to criminal casework as deviations from Hardy-Weinberg equilibrium suggested the entire population exhibited Wahlund Effect - genetic structure caused by independently breeding subpopulations incorrectly grouped as one population.

The following chapter will use population genetic analysis software packages to explore the genetic structuring within the reference dataset and establish how many sub-populations are present. The presence of genetic structure and population subdivision within the dataset holds promise for applications to assignment of individuals back to population or geographic origin. Only when individuals and populations are sufficiently differentiated from one another can the dataset assist investigations regarding provenance. Should sufficient genetic structuring be identified, the ability of the dataset for answering questions about provenance of poached carpet pythons will also be explored in this chapter.

4.2 Aims and objectives of this chapter

The aims of this chapter are to determine the degree of population substructuring present within the database of native *Morelia spilota* constructed in Chapter 3 and to assess its appropriateness for application to forensic casework and intelligence gathering.

This will involve the following three objectives:

- To perform population genetic analyses on the reference DNA database presented in chapter 3 to establish the number of subpopulations present,

- To perform statistical analyses on the subpopulations present and comment on their utility for casework involving carpet pythons, including the strength of evidence involving match probability calculations,

- To examine the ability of the reference DNA dataset to establish the most likely geographic provenance of a poached native individual.

4.3 Introduction

Chapter 3 concluded that the reference database of 249 individuals exhibited characteristics that indicated it did not represent an interbreeding population, thereby violating the assumptions applied when calculating forensic statistics for criminal casework. A population refers to members of a species considered to reside in sufficiently restricted geographic proximity to form a randomly mating group [1]; a single population of carpet python spread over the entire continent of Australia plus surrounding islands could not be expected to represent such a group. Population genetic analyses are required to further delineate the overall population into smaller subpopulations that

exhibit behaviours more appropriate to the assumptions underlying the statistical calculations used in forensic applications.

4.3.1 Hardy-Weinberg Equilibrium

The Hardy-Weinberg principle is a concept central to population genetic analyses. It deals with the equilibrium of genotype frequencies in a natural population from one generation to the next based on the underlying allele frequencies. When populations are in Hardy-Weinberg Equilibrium (HWE), the genotype frequencies of two alleles at a locus are present according to the equation:

$p^2 + 2pq + q^2 = 1$

where p and q are the frequencies of the respective alleles in the previous generation. This principle dictates that from one generation to the next allele frequencies will be maintained at this ratio (hence at equilibrium), regardless of the frequency of each [2].

The principle of HWE is an idealised approach, as it involves a number of underlying assumptions, many of which are rarely adhered to in reality: diploid inheritance, sexual reproduction, infinite population size, no generational overlap, random mating, no mutation, no migration and no natural selection [1]. These assumptions are clearly violated by most if not all natural populations, however HWE can still provide a useful approximation of patterns that can be expected in more complex population scenarios.

The principle of HWE provides a useful model to simplify the complex scenarios encountered in natural populations enough to allow meaningful inference to be drawn though statistical analyses. For this reason, many population genetic software packages include the assumption that the populations within the dataset adhere to Hardy-Weinberg Equilibrium.

Large populations that are randomly mating should exhibit observed genotype frequencies that do not deviate significantly from those expected under HWE. Therefore, comparing observed genotype frequencies with expected genotype frequencies under HWE can be a useful indicator of whether an interbreeding population has been accurately delineated as such.

4.3.2 Wahlund Effect

Wahlund Effect [3] is caused by the pooling of multiple sub-populations into one population that is not in reality a single, randomly mating population. This leads to an excess of homozygote genotypes relative to the proportion expected with random mating [1]. Average allele frequencies will not reflect the allele frequencies within each subpopulation [4]. The use of allele frequencies that are not representative of the population in question will produce incorrect statistical conclusions. The grouping of individuals into subpopulations aims to partition like with like and separate increasingly genetically dissimilar groups from one another. The presence of Wahlund Effect can be indicated by deviation from HWE, due to decreased levels of heterozygosity. Certain computer software packages are available to explore the genetic partitioning within the dataset, allowing more appropriate subpopulations to be defined for separate downstream analyses.

4.3.3 Population substructure and theta correction

In reality, genetic substructure persists even within subpopulations. It is the nature of natural genetic variation that certain habitats or environmental influences cause genetic differentiation at fine geographic scales. Particularly when species have been imposed upon by human influences such as hunting or habitat destruction, few individuals might be spread over patchy habitat remnants and might be quite genetically dissimilar to one another. A theta correction (θ), otherwise known as the co-ancestry coefficient, can be applied to statistical analyses to account for the genetic sub-structuring known to occur within the population due to inbreeding or shared ancestry [5].

The degree of genetic sub-structuring present within a population must be determined before it can be corrected for. F-statistics are used for this purpose. Wright [6] proposed a fixation index, F, to assess the degree of inbreeding (caused by population substructure) within a population. This concept was subsequently expanded into F-statistics [7] to assess and compare the degree of inbreeding (genetic sub-structuring) present by examining the correlation of alleles at different levels of a population hierarchy: the individual (I), subpopulation (S) and total population (T). It follows that F_{IS} (*f*) examines the correlation of alleles picked at random in an individual compared to those within its subpopulation; F_{IT} (F), the correlation of alleles of an individual relative to those of the total population and F_{ST} (θ) examines the correlation between alleles from within a subpopulation to those of the total population [8]. The alternative terms in brackets can be used when populations are randomly mating.

This last metric, F_{ST}, is most informative for examinations of population substructure (and as an extension, for evaluations of allele frequencies within the subpopulation where a crime has occurred [9]). Values range between 0 and 1. A value of 0 indicates homogeneity between the subpopulations, whereas a value of 1 represents complete fixation of the subpopulations for discrete alleles. Thus, F_{ST} is directly indicative of Wahlund Effect – heterozygote deficiency caused by population subdivision [10]. It identifies the degree of population genetic sub-structuring, which can in turn be handled by either splitting the dataset into smaller allele frequency databases more representative of their respective subpopulations, or applying the theta value to statistical calculations to correct for the increased degree of homozygosity caused by genetic sub-structuring. A global FST value is an average of the pairwise FST values [11]. Wildlife species often display increased population structure to human populations, resulting in higher F_{ST} values [12]. Table 4.1 provides a few examples.

An equivalent measurement of genetic differentiation, R_{ST} [13] should be mentioned at this point, which has been developed specifically for tandem repeat markers. It extends the function of F_{ST} by incorporating variances based on allele size rather than frequency, making the measure more applicable to the assumption of the Stepwise Mutation Model that STR loci are believed to follow [10]. While the use of this statistic has been widely debated as more appropriate for application to microsatellites (STRs) in the conservation genetic field, it is not a statistic adopted by forensic practitioners for the purpose of establishing statistical theta corrections (likewise for G_{ST} [14]. In maintaining comparability with other human and non-human forensic studies, this project will consider only the population genetic parameter F_{ST} (refer to [8] for further information on the application of the statistical alternatives to F_{ST}).

Lastly, while the theta correction can be applied to adjust for known population substructure in a global population, it is preferable to use a local subpopulation database when population numbers suffice. The estimated allele frequencies will provide a more accurate representation of the actual frequencies at the local level and this avoids the increased uncertainty introduced by the estimation of theta, for which the true F_{ST} value of the population is unknown [15].

Table 4.1: Examples of F_{ST} values in different species.

Species	# Loci used	F_{ST}	Reference
Humans *Homo sapiens*	Numerous	0.01-0.03	[16]
Common shrew *Sorex araneus*	7	0.103	[17]
European harbour seal *Phoca vitulina vitulina*	7	0.172	[18]
Brush-tailed rock-wallaby *Petrogale penicillata*	12	0.169 to 0.230	[19]
Bighorn sheep *Ovis canadensis*	8	0.224	[20]
Brown trout *Salmo trutta*	6	0.263	[21]
Star ascidian *Botryllus schlosseri*	5	0.401 to 0.468	[22]

4.3.4 Linkage Disequilbrium

Linkage disequilibrium is the non-random association of alleles between genes, often but not always caused by chromosomal proximity of the genes. The presence of genetic linkage between loci can violate the assumptions of population genetic analysis software packages and cause spurious analysis results. If linkage is detected, then only one of linked loci should be included in any calculation. However, it is wise to repeat analyses of the dataset involving each locus separately to identify influence that either locus might have on any particular sub-population or the overall dataset. If loci used to calculate a match probability are independent (in linkage equilibrium), then the product rule can be used to multiply probabilities across loci and provide a total match probability of a multi-locus profile [1].

It should be noted that patterns suggestive of linkage disequilibrium can also be created by processes asides chromosomal proximity, such local population processes of inbreeding and hybridisation [23]. Admixture of subpopulations possessing differing allele frequencies - when an interbreeding population is at the interface of another population - can also cause apparent correlations in the data that could mistakenly be interpreted as locus interactions [1]. For this reason, linkage between loci should be examined across multiple populations exhibiting HWE. Exact tests for HWE are conducted to test for independence of alleles within a locus, whereas linkage disequilibrium tests are conducted to detect independence of alleles between loci [24].

4.3.5 Reporting the weight of DNA profile evidence

Ongoing debate has fuelled the evolution of the statistical procedures that are used to report the weight of DNA evidence to the justice system. This spectrum began with the simplest application of

Hardy-Weinberg proportions of allele frequencies (p^2 and $2pq$) of a single source profile with the product rule to give genotype frequencies. This ongoing evolution of techniques has reached the current day complexity of applying Bayesian statistics to DNA profiles of mixed template source to sample posterior values based on prior distributions, taking into account the many variables involved in obtaining a DNA profile, including degradation, locus amplification efficiency, replicate amplification variability and template amounts left by each contributor [24]. Expert systems such as STRmix™ are available for human forensic science, however due to the extensive and resource intensive validation that is required to apply the software to non-human species (as well as the price tag now attached), the method of calculating match probabilities for DNA evidence involving wildlife species still fall somewhere in the middle of this spectrum of development.

Match probability calculations in the context of wildlife forensic science follow the recommendations of the NRC II report , based on the equations first proposed by Balding and Nichols [9] for homozygote genotypes:

$$Pr(A_iA_i|A_iA_i) = \frac{[2\theta + (1-\theta)p_i][3\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)}$$

And heterozygote genotypes:

$$Pr(A_iA_j|A_iA_j) = \frac{2[\theta + (1-\theta)p_i][\theta + (1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

Where A_i is allele i, A_j is allele j, p_i is the frequency of allele i in the reference database, p_j is the frequency of allele j in the database and θ is the degree of substructure calculated within the overall population. These equations improve on the product rule approach by incorporating the co-ancestry coefficient to correct for population substructure; in other words, this theta correction accounts for the degree of uncertainty around the allele frequencies within the sampled population, which likely deviates from the allelic frequencies in the true total population due to population substructure. The equations further take into account the observation of identical alleles in the suspect as well as the offender (assuming innocence), leading to the conclusion of increased incidence of these alleles in the population [9]. If loci are deemed independent, then these equations are multiplied across loci in a multilocus genotype to give the overall probability attached to the DNA profile.

The likelihood ratio is the inverse of the match probability and places the result in the context of two competing hypotheses, which is preferable in the context of the adversarial legal system. In the context of a comparison of matching DNA profiles from two different sources, the likelihood ratio would be reported based on the probability that the DNA matched because they come from the subject of interest (the numerator) and this will be 1, compared to the probability that they come from two different individuals and match by chance (this is the denominator), which is based on the frequency of the combined genotypes in the local population.

In this chapter, various population genetic software packages are used to assess the dataset for population substructure. Resulting population subdivisions will be subjected to statistical analyses to assess their suitability as population reference databases. Their ability to provide statistical support

to hypotheses regarding matching and parentage testing is evaluated. Sufficient population substructure provides the genetic differentiation required to be able to assign individuals back to the most likely population of origin. The ability of the identified subpopulation datasets to assist this purpose is also explored.

4.4 Materials & Methodology

4.4.1 Samples

Chapter 3 provides details about collection of the data used in this chapter. Refer to Appendix 3.1 in chapter 3 for details of the samples used. Sampling locations of the 249 individuals used in final reference dataset is provided (figure 4.1).

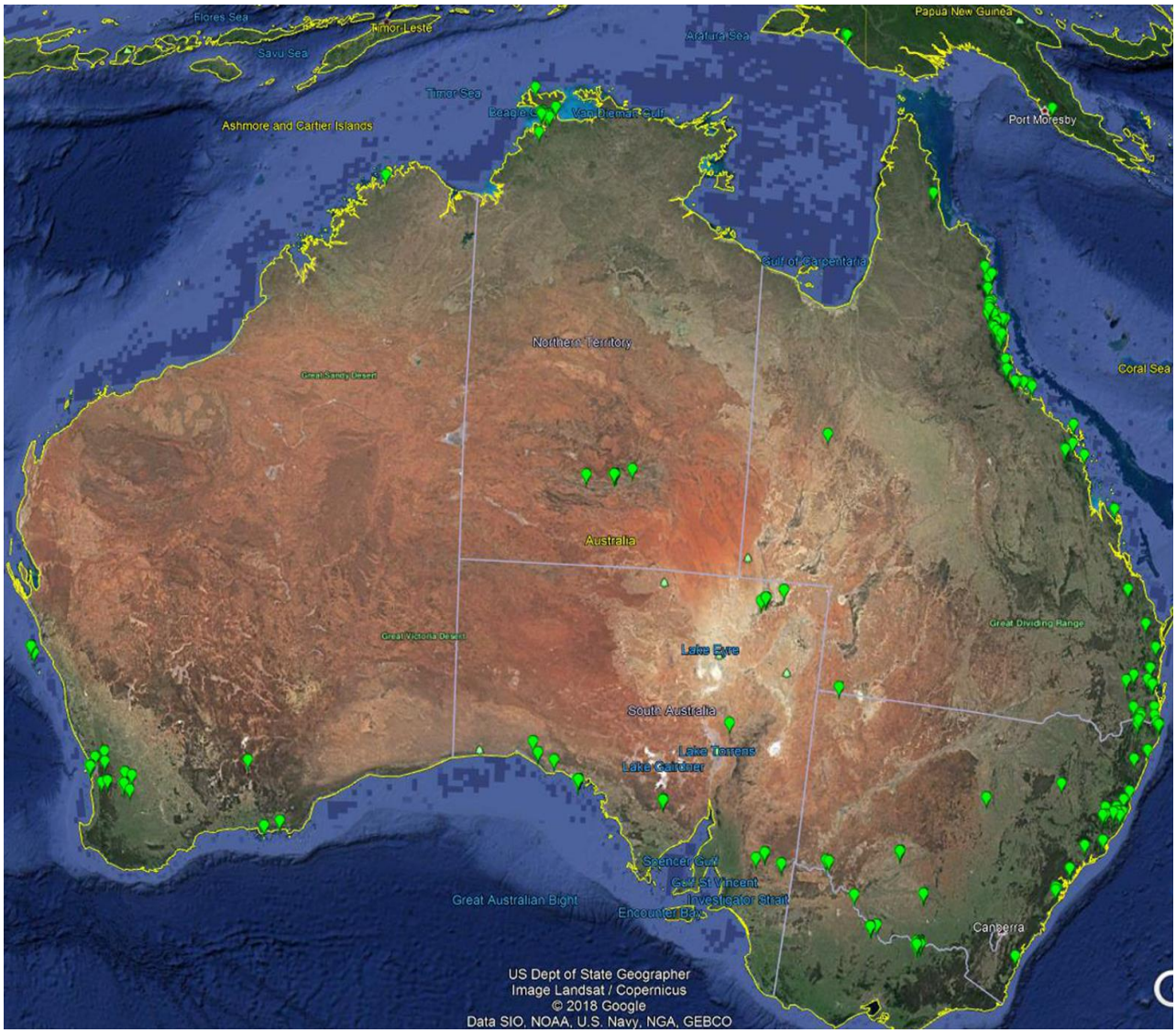


Figure 4.1: Geographic locations of the 249 individuals used in the final reference dataset, shown in higher resolution than figure 3.1.

4.4.2 Construction of a reference database

Chapter 3 described the sample genotyping, reference database construction and quality control checking involved in compiling the dataset analysed in this chapter.

4.4.3 STR database – Population genetic analyses

The number of individuals and complete loci included in the reference dataset for analysis required amendment based on the results obtained (n242, n254 and finally n249). Only the analysis results of

the final n249 dataset will receive more than a brief mention in this chapter, as the earlier datasets are considered training datasets that contributed to refinement of the final dataset. The decision making to arrive at the dataset considered most appropriate for application is elaborated on in discussion section 4.6.1.

The first step of analysis involved the use of clustering analysis software packages to explore the variability in the dataset and determine how many subpopulations the overall dataset most likely contains.

4.4.3.1 Model-based clustering analysis

The model-based clustering program STRUCTURE 2.3.4 [25] was used to partition the data into putative genetic populations. Details about the software methodology are provided in Appendix A4.1.1. As the systematics of the carpet python remains unresolved, relationships of all individuals in the reference dataset were considered unknown for analysis, rather than making any *a priori* assumptions about potential population clusters.

Initial STRUCTURE analyses run on the 23+_n242 dataset used the assumption of admixture due to the continuous spread of the data over large geographic ranges, without inclusion of sampling locations ('informative prior' - useful when populations are known to exhibit low variability). Five replicates each were run for K values one through seven, assuming both IAFM and CAFM. Burn-in and iterations were set to 50K each followed by 100K each to examine the effect on the output.

Genetic clustering of putative populations was repeated on the final 22+_n249 reference dataset, with parameters adjusted based on the results of the initial testing, which showed that extended conditions were required. Ten independent repetitions of K = 2-10 with 300 000 Markov Chain Monte Carlo (MCMC) iterations following a burn-in of 100 000 were performed. Extended iterations (200 000 burn-in and 1 000 000 iterations) were applied to some higher values of K where run convergence was not achieved. A model assuming correlated allele frequencies, non-informative priors and admixture was chosen. These settings were also applied to all subsequent STRUCTURE analyses for subpopulation differentiation.

The resulting STRUCTURE output files were exported into a Zip file and run through the online program STRUCTURE Harvester [26] to visualise the data and to assess the most likely K value of the dataset following [27]. The average log likelihoods were displayed. The analysis tool averages replicates from each run to give a single mean estimated Ln probability for each K, so the results did not provide information about any spurious run results that might be captured in the data and skew the averages.

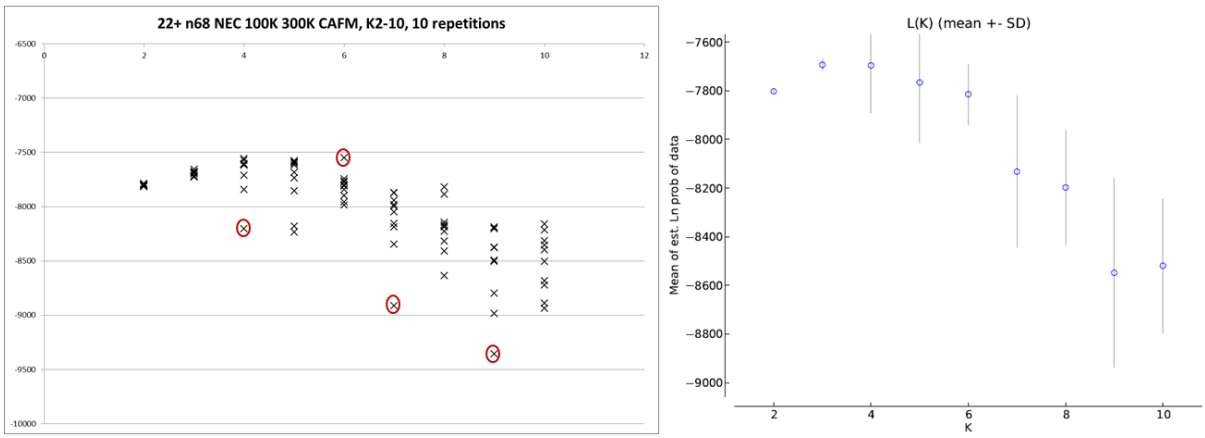


Figure 4.2: Manually plotted log likelihood values (left) versus Structure Harvester plot (right). Example outlier runs (circled) shown using manual plotting can be removed from subsequent analyses.

The Ln likelihood values of all runs for every K were also plotted manually in Excel [28] to evaluate the convergence of runs for each K value on a single result and identify and remove outlying runs (figure 4.2). Further inference about the most appropriate K value was made from the Delta K plot provided by STRUCTURE Harvester. Bar plots created from the STRUCTURE output file were used to visualise and assess the degree of genetic information within individuals attributed to the proposed populations. All representations of the data were compared to identify concordance about the most likely number of subpopulations present in the dataset (K value).

A more accurate representation of the component populations was gained by averaging the bar plots over the replicates of a K value. Population colours are not standardised across STRUCTURE bar plot outputs (making identification of replicate populations between bar charts difficult) and CLUMPP [29] was applied for this purpose. Averaging was eventually achieved by manually plotting the Q-matrix for each run belonging to a K value in Excel and then taking the average across runs. Each column of the Q-matrix represented a population. Column 1 was sorted for descending value. Individuals for which the majority of their genetic makeup was attributed to the same population were allocated a colour. All non-coloured individuals were sorted by descending value for column 2 and this process was repeated until the individuals had been sorted and coloured corresponding to the population to which most of their genetic information was attributed. This was done for every run of the same K value. By looking at the identity of individuals within each population, populations could be identified across runs. Columns were rearranged so that populations appeared in the same order across runs and the Excel average function was used to create a final population set showing average Q-values for each individual across all runs. Any run identified by the log likelihood plots as spurious (see figure 4.2 outliers) was removed from this analysis because the lack of convergence would yield erroneous Q-values that would skew the averages.

The above process was conducted for every K value obtained in a STRUCTURE run. The resulting average Q-matrix for each K value was compared to examine how attribution of individuals to populations changed with increasing population number (K). The populations defined by the Average Q values were compared to sample locations charted in Google Earth to assess whether the populations made biological sense.

Population partitions isolated from the first round of STRUCTURE analysis using the total dataset were then broken down and reanalysed to look for further subpopulation structure.

4.4.3.2 Principal Co-ordinate Analysis

Population and subpopulation partitions observed from STRUCTURE were further explored using Principal Co-ordinate Analysis (PCoA) in GenAlEx [30,31]. PCoA is another way of visualising variation and therefore population clusters within the dataset, although it does not rely on a stipulated biological model unlike STRUCTURE. Greater detail about the software methodology is provided in Appendix A4.1.2.The entire dataset as well as emerging clusters were explored using a Covariance-Standardised evaluation of the pairwise genetic distance matrix of the three axes to examine the degree of dissimilarity between individuals, and therefore population clustering, within the dataset. Emerging clusters were systematically removed or focused upon in further PCoA analyses to explore different levels of variability existing in the dataset.

4.4.3.3 Mantel test for Isolation by Distance

A Mantel test was used to detect pattens of Isolation by Distance using GenAlEx [30,31]. Geographic coordinates were exported from the Google Earth project (section 3.4.1) in Decimal Lat/Long format and added to the Codominant-Genotypic genetic information. Squared genetic distance (populations), Linear Genetic Distance (individuals) and Geographic Distance matrices were created in GenAlEx and Paired Mantel testing performed using 9999 permutations.

4.4.4 Assessing Hardy-Weinberg Equilibrium and Linkage Disequilibrium

Exact tests for deviations from Hardy-Weinberg Equilibrium were run on putative populations suggested by STRUCTURE and PCoA analyses using Genepop on the Web v4.2 [32,33]. The probability option was run using the settings 1000 dememorization steps, 5000 batches and 10000 iterations per batch.

All populations adhering to Hardy-Weinberg Equilibrium were examined for linkage disequilibrium to test for independence between loci using Genepop on the Web v4.2 [32,33]. Parameter values were increased until the Standard Error values were <0.01. The log likelihood ratio test (G-test) was performed using MCMC sampling with final settings of 1000 dememorization steps and 5000 batches, each at 10000 iterations.

Unfortunately, no mapped snake reference genome is available at the time of writing, so chromosomal locations of the STR loci used in this study are unknown and linkage patterns between loci must be determined by computational analyses alone.

4.4.5 F-statistics, summary statistics and forensic statistics

Subpopulation allele frequencies and summary statistics such as expected and observed heterozygosity, Polymorphic Information Content (PIC) and F-statistics were investigated using functions in GDA [34], GenAlEx v6.5 [31] and The Excel Microsatellite Toolkit [35]. A dog DNA profiling Excel spreadsheet (authored by R. Ogden and provided by L. Webster) was modified to calculate match probabilities and likelihood ratios for this study, using the overarching n249 and

each of the six sub-populations (refer to Appendix 4.5 for calculation spreadsheets). Spreadsheet rows P (het) and P (hom) contain the match probability equations given in section 4.3.4.

Estimates of the least and most common genotype frequencies present in the population were calculated following [15]. Hardy-Weinberg's 2pq was applied to each locus using of the two most common allele frequencies and the results multiplied across all 24 loci to give the most common genotype frequency. The rarest allele frequency was calculated using the 5/2N minimum allele frequency rule (representing a minimum of five observations in the population, [16]; this rule is applied to mitigate the sampling error inherent in the reference database, which represents a sample of the population of interest because sampling the total population is not possible. The minimum allele frequency rule was applied to the database allele frequencies, and any profile alleles not observed in the database, in all match probability calculation sheets.

4.4.6 Population Assignment

Geographic origin assignment testing was performed using six defined populations: the five populations reflected from the STRUCTURE K5 analysis excepting the two *Morelia spilota imbricata* island populations considered independent of one another (see section 4.5.2.2 for details). Individuals 81320 (Gunnedah, NSW) and 67128 (Eungella, Queensland) were removed from the assignment testing dataset (explained in section 4.5.7.1). The resulting n247 dataset was used to test assignment back to population of origin with Geneclass2 [36]. For each dataset, the following settings were applied: assign/exclude population as origin of individuals, applying Bayesian methods following [37], analysed both with no Monte-Carlo resampling and resampling of 10000 individuals using the [38] simulation algorithm.

The complete n247 dataset was tested for self-assignment. The self-assignment analysis was repeated using the frequency method [39] instead of the Bayesian method [37] to compare the outcome of these different approaches. Twenty-five individuals (10% of database) were removed from the database, selected by a semi-random process that ensured all populations were represented, and assignment tested against the remaining 222 reference samples. The same dataset was replicate tested five times.

Each population was systematically removed from the reference database and the removed individuals assigned back to the reference dataset excluding their population. Twenty-eight samples that were not included in the final reference dataset were treated as unknowns (geographic origin was recorded, but not revealed until after assignment was complete) and assigned back to the dataset of 249 reference samples; these samples did not meet the required minimum threshold of loci for the reference database. The twenty-eight individuals ranged between complete 21 locus genotypes and five complete loci. The accuracy of assigned population of origin was verified by examining whether the revealed geographic origin fell within the geographic distribution of members from the assigned sub-population. This could not be determined for three samples, as these individuals originated from close to the overlapping distributions of the NEC and SEC sub-populations.

4.4.7 Calculating match probabilities for ‘unknown’ samples

Using the results of the population assignment testing to dictate the appropriate sub-population, match probabilities and likelihood ratios were calculated for the 28 ‘unknown’ samples plus one quality compromised sample using the match probability calculation spreadsheets in Appendix 4.5. Where an individual was not excluded from >1 population, calculations were performed for each population. Where individuals were excluded from all populations, the overarching n249 allele frequency database was used.

If only one allele was recorded at a locus, the locus was excluded from the calculation. While one could consider the calculation incorporating a single allele at a locus to increase the information used, that approach was not applied here. Where required, theta corrections were included based on the results in table 4.11, section 4.5.6, however considerations surrounding the application of these are discussed in section 4.6.9. Calculations were conducted incorporating both the global theta value and the conservative upper 95th percentile theta value. Calculations involving the two island populations used a restricted set of loci, due to reasons presented in section 4.6.5. Excluded loci are shaded in grey in the allele frequency look up tables of the two respective calculation spreadsheets (sections A4.5.6 and A4.5.7) and calculations include eight loci for Garden Island individuals and ten loci for St. Francis Island individuals. The minimum allele frequency rule (5/2N) was applied to all databases and also to alleles of the ‘unknown’ individuals not observed in the database.

4.5 Results and Discussion

4.5.1 Reference Database Construction and Quality Control

As described in chapter 3, a total of 304 carpet pythons were genotyped at 27 loci, but only 24 loci were chosen for use in the final reference database (section 3.5.8). Nine individuals gave no amplification product, 27 individuals exhibited alleles at 19 or less loci and 190 individuals exhibited complete profile across the 24 loci included in the final database.

Selection of the final dataset to include in the reference database involved a trade-off between minimising the amount of missing data in the profile and maximising the number of individual profiles contributing to a small dataset covering a very large geographic area.

Table 4.2: Potential reference DNA database sizes using this dataset

Complete loci	24	23+	22+	21+	20+
Database size (n)	190	242	254	256	258

Table 4.2 provides the number of reference DNA profiles available for inclusion applying varying thresholds of missing data. Initially the 23+ dataset was analysed, however consideration of the extra 12 samples that had two missing loci from a complete profile identified that these individuals had diverse geographic origins and that their inclusion would improve the reference database for multiple potential final populations. Very few individuals would be gained from increasing the number of missing loci past this point and the amount of missing data included in analyses is best minimised, so a 22+ loci cut-off was chosen as the most appropriate final reference dataset.

The Multilocus ->Matches function in GenAlEx identified 5 pairs of individuals with duplicate genotypes: 71395/71397 (Walpella Island, Vic., SIB population), 71376/71407 (Mt Hope, Vic., SIB population), 81448/81450 (M. bredli, Northern Territory, SIB population), 81300/81301 (Atherton, Queensland, NC population), 81305/81306 (Acacia Plateau, NSW and Toonumbar NP, NSW respectively, NC population). All pairs were sampled from similar localities (although the latter two localities are 40 km apart) and many had close or successive ABTC numbers. It is unclear whether these duplicates have resulted from familial groups or potential human error (e.g. mislabelling, sample mix up) at time of sampling or since. Regardless of the cause, one sample of each pair of duplicate genotypes was removed from the reference dataset to overcome potential errors in analysis caused by their presence. The finalised reference dataset contained 249 individuals for subsequent analyses.

Genotyped samples not included in the final reference database can be used to test the assignment and statistical power of the dataset (see sections 4.5.7.4 and 4.5.8)

4.5.2 STR database – population genetic analysis

The complete n249 database was presented in Chapter 3, accompanied by the discussion that in this form it is inappropriate as a reference database to apply to forensic investigations involving this species.

Forensic validation of a human STR dataset might generally end at this point, as F_{ST} values tend to be quite small between human populations (ethnic groups) owing to their relatively recent divergence

from one another. Further, human reference DNA databases are usually constructed using a well sampled ethnic group, collected to be representative of the geographic area of interest. All of these factors are reflected in the frequent adherence of the STR loci to HWE, warranting no further investigation.

While the same type of STR markers are being applied in an intraspecies context, further analysis is potentially necessary when looking at non-human species due to elevated population structure typically evident in wildlife species [12]. The crucial factor to assess is whether or not the populations exhibit Hardy-Weinberg Equilibrium, as the calculations used in reporting forensic genetic statistics involve a number of underlying assumptions based on populations exhibiting adherence to Hardy-Weinberg Equilibrium. The n249 dataset significantly deviates from Hardy-Weinberg Equilibrium at every locus (section 3.5.9).

It comes as no surprise that HWE is violated, as this dataset relates to individuals spread over a continent, including areas separated by geographic barriers to geneflow. Allele frequencies averaged over every putative population are highly unlikely to be representative of all populations across such an expansive and ecologically diverse area. Population genetic analyses are required to determine if the dataset contains multiple heterogeneous populations and to partition the dataset into these sensible interbreeding populations.

4.5.2.1 Sub-populations

Model based clustering analysis

Initial STRUCTURE results of the entire n249 dataset identified either three or five sub-populations to which almost all individuals showed a strong degree of assignment. The data produced using five proposed sub-populations demonstrates robust enough groupings to be considered a realistic scenario (see Appendix 4.2 for the detailed analysis supporting this conclusion). Only three individuals exhibited an average assignment value of less than 0.666 to a single population. Average Q Values for K6 and above could not be calculated, as multimodality was seen between replicate runs (different combinations of individuals assigned to sub-populations between runs so the equivalent populations could not be identified and averaged across runs).

The geographic range of the five resulting sub-populations can be seen in figure 4.3 and these sub-populations are defined as: 1. *Morelia spilota imbricata* populations of St. Francis Island and Garden Island (Msi Islands), 2. *M. s. imbricata* mainland population (Msi ML – although four individuals are actually off-shore inhabitants of small islands they genetically identify most closely with the mainland population), 3. Southern, Internal & Bredli population (SIB), 4. South Eastern Coastal (SEC) and 5. Northern Eastern Coastal (NEC).

The results of the STRUCTURE runs at higher K values were examined in greater detail to establish robustness of population groupings, however this work is possibly reading too much into the data, so it has been placed into Appendix 4.2.2.

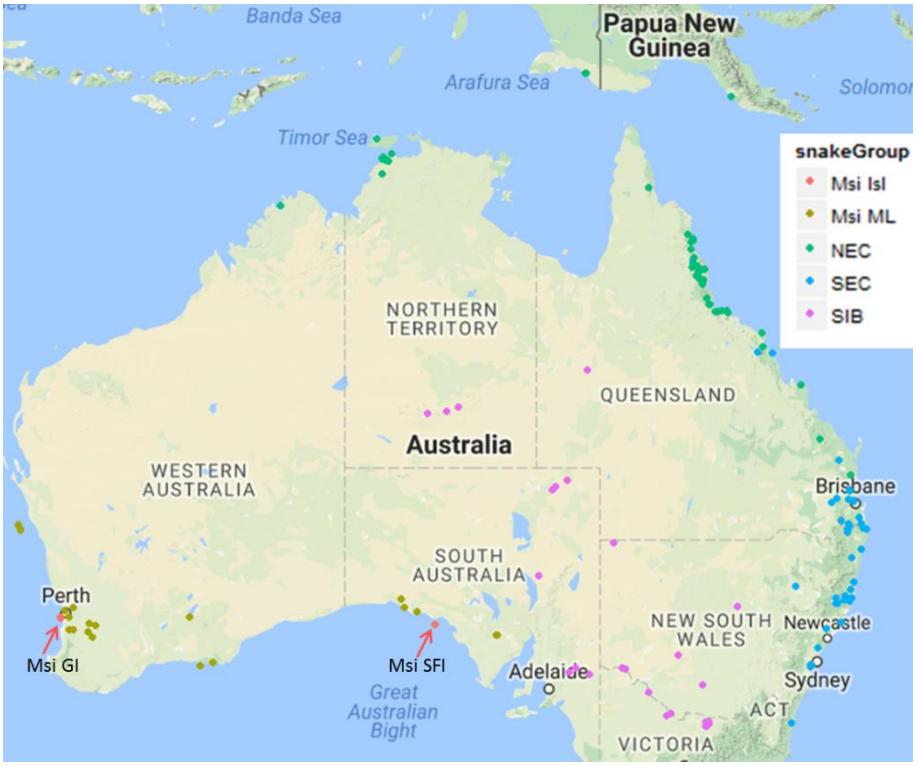


Figure 4.3: Locations of individuals representing the five overarching populations. The respective Msi Island populations are indicated by arrows.

Principal Co-ordinate Analysis

The PCoA clustering based on these five populations are shown in figure 4.4. These PCoA plots show the five population clusters visualised by all three paired axes. While the NEC and SEC clusters are located in close proximity from all three angles, the other three populations cluster discretely in at least one of the plots. Interestingly the Msi island population consistently forms two very discrete clusters, reaffirming the division of this group into two subsets mentioned above.

Figure 4.4: Principal Co-ordinate analysis plots of the five sub-populations viewed in all three planes of pairwise axial comparisons.

As the analysis does not apply a genetic model, the PCoA plots should not be given as much weight in assessing the genetic population structure as analyses use a genetic model such as Structure. Despite this, population clustering is evident in the three PCoA plots. Without population structure, points from all populations would be seen to overlap one another, scattering evenly around the axial intercept (see figure A4.10 for example). The Msi Mainland population clearly clusters separately to all other populations in the two charts displaying axis one. The SIB population also clusters separately from the two coastal populations in the two charts showing the second axis. Interestingly, the two Msi island populations that exhibited differentiation from one another in the previous section are also clearly quite divergent based on information from axes two and three.

All five overarching populations significantly deviate from Hardy-Weinberg Equilibrium (see section 4.5.3). This is not surprising as the vast geographic ranges of these populations, some including geographic features that would no doubt act as barriers to gene flow, would be expected to preclude random mating.

4.5.2.2 Further sub-population subdivision

Further genetic analysis of each of the putative sub-populations is presented in Appendix 4.3. Population divergence between the Msi island populations is supported by both the STRUCTURE and PCoA clustering. Examination of the allele frequencies exhibited by the two island populations suggests that they differ markedly (table 4.3) and should not be combined for the purpose of allele frequency estimates. These results justify the split of the Msi islands into two separate sub-

subpopulations – Garden Island (Msi GI) and Saint Francis Island (Msi SFI), creating the six overarching populations as referred to in subsequent analyses.

Table 4.3: Allele frequencies exhibited by the two *Morelia spilota imbricata* island populations calculated using the total populations of 18 and 20 individuals respectively. Loci fixed for a particular allele are highlighted in red.

Similar

Locus	Allele	GI	SFI
MS27	256	1.000	1.000
MsF24	329	1.000	1.000
MsF31	76	1.000	1.000
MsF8	165	1.000	0.806
	169	-	0.194
MsF16	300	-	0.029
	304	0.921	0.971
	324	0.079	-
MsF22	357	1.000	0.694
	361	-	0.194
	365	-	0.111
MS9	171	0.400	0.294
	175	0.100	0.029
	179	0.125	0.235
	183	0.375	0.353
	187	-	0.088
MsF5s	119	-	0.083
	123	0.900	0.833
	127	0.100	0.083
MsF15	370	0.026	-
	375	-	0.031
	380	0.789	0.656
	382	0.026	-
	385	0.158	0.313

Locus	Allele	GI	SFI
MsF27s	387	0.450	0.111
	388	0.125	-
	391	0.150	0.500
	392	0.025	-
	395	0.150	0.389
	396	0.100	-
MS25	254	-	0.028
	262	0.125	0.528
	266	0.250	0.278
	270	0.050	0.111
	274	0.100	-
	278	0.425	0.056
	282	0.050	-
MsF33	418	0.250	-
	421	-	0.412
	422	0.225	-
	425	-	0.147
	429	-	0.059
	433	0.200	0.088
	437	0.075	0.294
	445	0.175	-
	449	0.075	-
MsF9	157	-	0.028
	161	0.950	0.389
	165	-	0.444
	169	0.025	0.139
	173	0.025	-
MsF2	131	-	0.361
	135	0.150	0.556
	139	0.175	0.028
	143	0.550	-
	147	0.125	0.056
MsF18	186	0.675	0.111
	190	0.100	0.722
	194	-	0.167
	198	0.075	-
	206	0.150	-

Diverse

Locus	Allele	GI	SFI
MS4h	430	-	0.028
	434	-	0.306
	438	0.050	0.528
	442	0.600	0.056
	446	0.300	-
	450	-	0.083
	462	0.025	-
	466	0.025	-
MsF26	319	1.000	-
	334	-	0.056
	335	-	0.722
	339	-	0.222
MS13	162	1.000	-
	166	-	0.972
	183	-	0.028
MsF4	104	-	0.028
	105	-	0.972
	113	0.300	-
	117	0.175	-
	121	0.525	-
MsF17	333	-	1.000
	353	0.325	-
	357	0.675	-
MsF3	149	1.000	-
	153	-	0.917
	157	-	0.083

Locus	Allele	GI	SFI
MS17	228	0.225	-
	232	0.100	-
	240	0.025	-
	241	0.325	-
	245	0.125	-
	249	0.150	-
	253	0.050	-
	283	-	0.028
	287	-	0.028
	291	-	0.083
	295	-	0.167
	299	-	0.389
	303	-	0.222
	307	-	0.083
MS3	206	-	0.028
	210	-	0.361
	214	-	0.222
	218	-	0.361
	222	-	0.028
	226	0.125	-
	227	0.075	-
	230	0.350	-
	239	0.175	-
	243	0.225	-
	247	0.025	-
	251	0.025	-
MsF28si	295	1.000	-
	305	-	1.000

The updated overarching population contains six sub-populations (Figure 4.5).

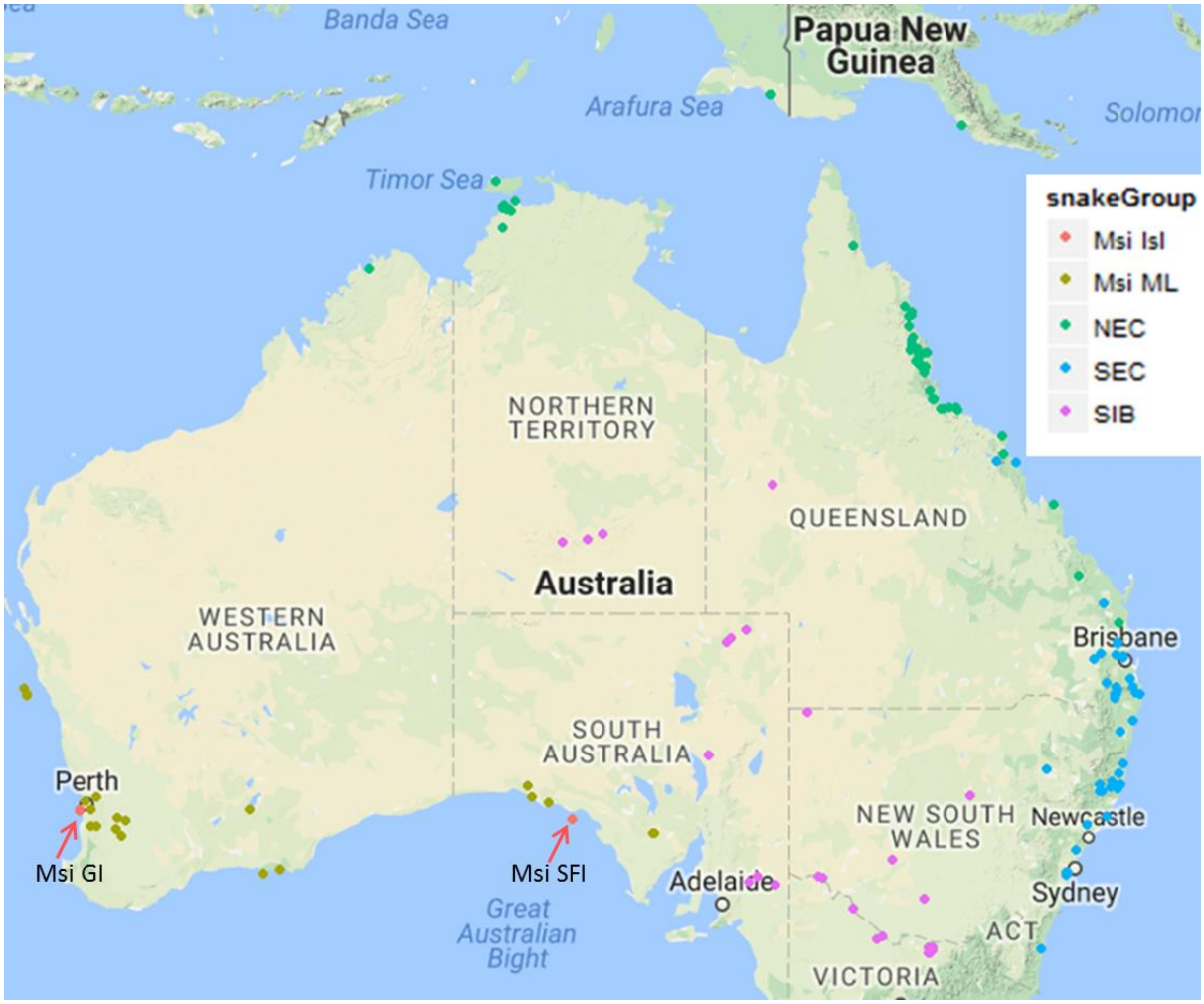


Figure 4.5: The final groupings of the n249 reference dataset into six overarching sub-populations. The two island populations are represented by salmon coloured dots and indicated by arrows.

Neither Msi island population demonstrates further genetic sub-structuring. This could be expected, given that both are very small islands with little to no cause for restricted interbreeding. Neither of the island populations require the application of a correction factor for substructure. However, the population sizes of the two island populations are arguably too small to be suitable for match probability calculations. This problem could potentially be overcome by the use of a reduced number of loci in statistical calculations, further discussed in section 4.6.5.

Somewhat more unexpectedly, the South Eastern Coastal population exhibit very slight population differentiation, surprising as it extends more than 2000 kilometres along the Australian coast and inland to the Great Dividing Range. Isolation by distance was not evident (R^2= 0.0103), but slight sub-structuring was evident when comparing the two overlapping populations each seen to be in Hardy-Weinberg Equilibrium, but not in HWE when combined (see section 4.5.6).

The findings indicate genetic sub-structuring in the other three populations. Both the Msi Mainland population and the SIB population show a similar pattern (figure 4.6). Just under half of the individuals have been sampled from a small geographic area (South West Western Australia and SE Vic/Warby Ranges, respectively). The members of these subpopulations show high genetic relatedness and adhere to HWE (enveloped by green ellipse in figure 4.6). Each dataset contains

another large membership cluster, however both of these clusters consist of single individuals spread over a large geographic area (salmon dots in 4.6a and blue dots in 4.6b).

Figure 4.6: Groupings of individuals based on average Q-matrices in a) the *Morelia spilota imbricata* mainland population, and b) Southern, Internal and Bredli population. Colouring of populations in a) correspond to K4 groupings (see figure A4.18). Two large, heavily sampled population exhibiting HWE are circled in green.

The remaining individuals form clusters of between 2- 6 members, but show a degree of genetic dissimilarity to the other subpopulations. The issue with these two populations is that, clearly genetic differentiation exists therein, but sample numbers are too low to be able to split these populations further and conduct meaningful genetic analyses. The Msi ML and SIB populations must each be treated as a single population with application of a theta value based on the global F_{ST} of the population to correct for the known genetic sub-structuring therein (section 4.6.9 further discusses the difficulties of this situation).

The final population NEC appears to split further into four subpopulations: Northern Territory individuals, Merauke Western Papua individuals, a Northern Queensland group including the New Guinean individual and a Southern Queensland group – however the latter two are possibly one continuous population judging by the degree of admixture of individuals around the boundary of these groups. While this information might be useful for geographic assignment testing, the population numbers are too small for statistical analysis. This population will likewise need to be kept as a single population and a theta correction applied to account for the substructure occurring within (see section 4.6.9 for further discussion).

4.5.3 Hardy-Weinberg Equilibrium and Linkage Disequilibrium

The overall reference database provided in Chapter 3 deviated significantly from Hardy-Weinberg Equilibrium at every locus. When split into the initial K5 populations discussed in section 4.5.2.1, no population exhibited HWE (table 4.4).

Table 4.4: Results of exact tests for Hardy-Weinberg Equilibrium in the five overarching populations.

n38 Msi Islands

Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0536	0.0001	0.1204	-0.0326
MS25	0.8918	0.0003	0.0421	0.0049
MsF33	0.0511	0.0003	0.1046	0.0716
MsF31	No information.			
MsF9	0.2868	0.0006	0.1921	0.0501
MsF16	1	0	-0.0332	-0.0155
MsF3	0	0	0.8555	0.4742
MsF17	0	0	0.6263	0.5073
MsF18	0.0083	0.0001	0.1409	0.026
MsF28si	0	0	1	1.027
MsF22	0.0416	0.0001	0.2962	0.2485
MS3	0	0	0.1937	0.1049
MS4h	0	0	0.1113	0.0412
MsF27s	0.0004	0	0.1241	0.1935
MsF8	0.2612	0.0001	0.2259	0.2296
MsF26	0	0	0.6534	0.5852
MsF4	0	0	0.4671	0.1933
MS27	No information.			
MsF24	No information.			
MS9	0.3225	0.0004	0.0189	0.0298
MsF2	0.0001	0	0.2062	0.2185
MS17	0.0001	0	0.0648	0.0524
MsF15	0.0053	0.0001	0.2587	0.0651
MS13	0	0	0.9498	0.5058

All (Fisher's method):

Chi2	Infinity
Df	42
Prob	Highly sign.

n36 Msi ML

Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	0.1962	0.0009	0.2099	0.2107
MS25	0.002	0.0002	0.1205	0.0578
MsF33	0.0046	0.0003	0.0573	0.0535
MsF31	No information.			
MsF9	0.1571	0.0005	0.0854	0.0504
MsF16	0.5513	0.001	-0.0568	-0.055
MsF3	0.0001	0	0.3636	0.274
MsF17	0.1867	0.0009	0.1387	0.1107
MsF18	0.5347	0.0009	0.0776	0.0567
MsF28si	0.0124	0.0002	0.3223	0.1677
MsF22	0.077	0.0005	0.18	0.1466
MS3	0.0006	0.0001	0.2019	0.1813
MS4h	0.3902	0.0013	0.0343	0.0079
MsF27s	0	0	0.3465	0.2259
MsF8	0.338	0.0004	0.0006	-0.0163
MsF26	0.0285	0.0003	0.1549	0.0659
MsF4	0.0254	0.0009	0.1788	0.0857
MS27	0	0	0.4979	0.6676
MsF24	0.0208	0	0.5354	0.5471
MS9	0.3651	0.0013	0.0581	0.1198
MsF2	0.2354	0.0007	0.0448	0.0232
MS17	0.4592	0.0032	0.0448	0.0907
MsF15	0.0351	0.0003	0.145	0.0888
MS13	0.0514	0.0013	0.1241	0.0625

All (Fisher's method):

Chi2	177.7786
Df	46
Prob	0

n68 NEC

Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0139	0.0005	0.1555	0.0894
MS25	0.018	0.0013	0.049	0.066
MsF33	0.046	0.0024	0.0495	0.0206
MsF31	0.0065	0.0004	0.1427	0.1124
MsF9	0.0052	0.0001	0.1281	0.1147
MsF16	0.3941	0.0009	0.0478	0.0171
MsF3	0.0239	0.0003	0.1775	0.1439
MsF17	0.1741	0.0012	0.0725	0.0125
MsF18	0.9401	0.0007	-0.026	-0.0031
MsF28si	0.0151	0.0007	0.1352	0.0986
MsF22	0.2105	0.0007	0.0969	0.0904
MS3	0.0011	0.0003	0.0953	0.0573
MS4h	0.0306	0.0004	-0.0173	0.0679
MsF27s	0.2536	0.003	0.0485	0.0387
MsF8	0.7955	0.0008	0.0799	0.0572
MsF26	0.3613	0.0015	0.1383	0.0733
MsF4	0.0287	0.0009	0.0997	0.0643
MS27	0.0151	0.0006	0.1455	0.1328
MsF24	0.0044	0.0001	0.1901	0.2907
MS9	0.0281	0.0004	0.1206	0.091
MsF2	0.0622	0.0013	0.0204	0.0228
MS17	0.0002	0.0001	0.1946	0.1285
MsF15	0.0064	0.0003	0.0666	0.0425
MS13	0.1882	0.0027	0.0173	0.0291

All (Fisher's method):

Chi2	162.7356
Df	48
Prob	0

n51 SEC

Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	0.9122	0.0009	0.0489	0.0088
MS25	0.9217	0.0026	-0.0113	-0.0094
MsF33	0.0512	0.002	0.0523	0.0227
MsF31	0.0362	0.0006	0.0552	0.0528
MsF9	0.8467	0.0006	-0.0541	-0.0128
MsF16	0.0841	0.0006	0.1416	0.0884
MsF3	0.0059	0.0001	0.2126	0.2926
MsF17	0.6684	0.0026	0.0151	0.0049
MsF18	0.3242	0.0016	0.103	0.0749
MsF28si	0.2163	0.0009	0.0928	0.0256
MsF22	0.0089	0.0001	0.0826	0.123
MS3	0.4812	0.0059	0.0093	0.0164
MS4h	0.5104	0.0015	-0.0547	-0.0279
MsF27s	0.2465	0.0023	0.029	0.001
MsF8	0.0136	0.0002	0.1803	0.1245
MsF26	0.9133	0.001	0.0132	0.0234
MsF4	0.8276	0.0028	0.0162	0.0041
MS27	0.0178	0.0007	0.0552	0.0124
MsF24	0.3248	0.0008	0.1148	0.0613
MS9	0.132	0.0008	0.1579	0.1109
MsF2	0.2758	0.0012	0.077	0.0393
MS17	0.749	0.005	-0.0243	-0.013
MsF15	0.6585	0.0027	0.0066	0.0254
MS13	0.0096	0.0008	0.1136	0.0793

All (Fisher's method):

Chi2	86.4598
Df	48
Prob	0.0006

n56 SIB

Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	0.1497	0.0007	0.1363	0.1384
MS25	0.0787	0.001	0.115	0.028
MsF33	0	0	0.1452	0.1172
MsF31	0	0	0.4278	0.4217
MsF9	0.5712	0.0007	0.0519	0.0626
MsF16	0	0	0.3518	0.4065
MsF3	0.0011	0	0.1824	0.2595
MsF17	0.0021	0.0002	0.2189	0.0913
MsF18	0.0096	0.0002	0.146	0.0891
MsF28si	0.2228	0.0002	0.1545	0.1106
MsF22	0.0062	0.0001	0.1756	0.2577
MS3	0.0065	0.0006	0.0643	0.0571
MS4h	0.0319	0.0004	0.1282	0.109
MsF27s	0	0	0.3237	0.1272
MsF8	0.047	0.0002	0.161	0.187
MsF26	0.0259	0.0002	0.2057	0.2227
MsF4	0.0029	0.0001	0.1524	0.1164
MS27	0	0	0.242	0.2259
MsF24	0.0067	0.0001	0.2455	0.3761
MS9	0.0133	0.0001	0.1025	0.1497
MsF2	0	0	0.198	0.2892
MS17	0.0921	0.0011	0.1521	0.0911
MsF15	0	0	0.3785	0.2915
MS13	0.0005	0	0.3451	0.2681

All (Fisher's method):

Chi2	Infinity
Df	48
Prob	Highly sign.

When the dataset was further subdivided into decreasing population sizes (section 4.5.2.2), seven populations were identified within the total dataset (representing a minimum of eighteen individuals) that displayed allelic segregation in line with Hardy-Weinberg Equilibrium (table 4.5). Population delineations that deviated from HWE are provided in Appendix A4.4. The deviation exhibited by these populations is likely due to the large geographic areas over which the respective constituents are spread, some involving dispersal barriers such as arid desert without estuarine corridors; individuals so geographically dispersed can't reasonably be expected to form a randomly mating population. However, interestingly the Queensland population that showed HWE is spread over a very large geographic area, particularly the outlier snake seen in the north of Western Australia (figure 4.7).

Table 4.5: Seven small sub-populations did not deviate significantly from Hardy-Weinberg Equilibrium.

n20 Msi GI
Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	1	0	-0.0857	-0.0877
MS25	0.9155	0.0002	-0.0071	-0.0244
MsF33	0.9271	0.0002	-0.0303	-0.0336
MsF31	No information.			
MsF9	1	0	-0.0133	-0.0007
MsF16	1	0	-0.0588	-0.0603
MsF3	No information.			
MsF17	1	0	0	0
MsF18	0.6497	0.0004	-0.1603	-0.1131
MsF28si	No information.			
MsF22	No information.			
MS3	0.2902	0.0006	0.1207	0.0643
MS4h	0.0757	0.0003	-0.1651	-0.0592
MsF27s	0.007	0.0001	0.1303	0.1198
MsF8	No information.			
MsF26	No information.			
MsF4	0.5816	0.0002	-0.0511	-0.1081
MS27	No information.			
MsF24	No information.			
MS9	0.7934	0.0002	-0.0878	-0.0934
MsF2	0.0546	0.0001	0.1504	0.2585
MS17	0.5656	0.0006	-0.174	-0.1184
MsF15	0.0191	0.0001	0.2742	0.0963
MS13	No information.			

All (Fisher's method):

Chi2	35.1636
Df	30
Prob	0.2367

n18 Msi SFI
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0379	0.0001	0.2649	-0.0049
MS25	1	0	-0.1218	-0.0681
MsF33	0.4654	0.0004	0.0376	0.0018
MsF31	No information.			
MsF9	0.9503	0.0001	-0.1162	-0.087
MsF16	No information.			
MsF3	1	0	-0.0625	-0.0642
MsF17	No information.			
MsF18	0.7648	0.0001	-0.2409	-0.1686
MsF28si	No information.			
MsF22	0.1705	0.0002	0.1959	0.1794
MS3	0.4839	0.0005	-0.0208	-0.03
MS4h	0.5699	0.0005	-0.1421	-0.0728
MsF27s	1	0	-0.1087	-0.1095
MsF8	0.5121	0.0001	0.1414	0.146
MsF26	0.0282	0.0001	-0.0149	0.3857
MsF4	No information.			
MS27	No information.			
MsF24	No information.			
MS9	0.0657	0.0002	0.1373	0.1043
MsF2	0.657	0.0003	-0.1691	-0.0709
MS17	0.2194	0.0006	0.0734	0.0886
MsF15	0.3313	0.0003	0.234	0.117
MS13	No information.			

All (Fisher's method):

Chi2	69.9958
Df	62
Prob	0.227

n21 SE Vics
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.5613	0.0007	0.0385	0.0586
MS25	0.0265	0.0004	0.0883	-0.013
MsF33	0.8752	0.0005	-0.0332	-0.0378
MsF31	0.6707	0.0003	0.0698	0.0193
MsF9	0.943	0.0001	0.0083	-0.0482
MsF16	0.9419	0.0002	-0.098	-0.0442
MsF3	0.5312	0.0006	-0.072	0.1031
MsF17	0.3382	0.0007	0.2218	0.1592
MsF18	0.3695	0.0003	-0.0191	-0.0368
MsF28si	1	0	-0.0526	-0.0538
MsF22	0.9112	0.0002	-0.115	-0.1092
MS3	0.4269	0.0011	-0.0127	0.0079
MS4h	0.4628	0.0007	0.1564	0.1161
MsF27s	0.7149	0.0009	0.0374	0.0475
MsF8	1	0	0.0909	0.0933
MsF26	0.9129	0.0002	-0.0476	-0.081
MsF4	0.5343	0.0006	-0.005	-0.025
MS27	1	0	-0.0323	-0.0428
MsF24	1	0	-0.0811	-0.0829
MS9	0.7116	0.0005	-0.1259	-0.0911
MsF2	0.6935	0.0003	0.0078	0.0085
MS17	0.844	0.0005	-0.0014	0.0001
MsF15	0.0001	0	0.575	0.346
MS13	0.2647	0.0004	0.04	-0.0238

All (Fisher's method):

Chi2	44.5469
Df	48
Prob	0.6151

n24 SW WA all individuals
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.3338	0.0009	0.192	0.0739
MS25	0.2355	0.0018	0.0407	0.0152
MsF33	0.2996	0.002	-0.0202	-0.0247
MsF31	No information.			
MsF9	0.3596	0.0006	0.0147	-0.0008
MsF16	0.8282	0.0005	-0.1471	-0.0893
MsF3	0.0788	0.0003	0.2896	0.1639
MsF17	0.7254	0.0007	0.0526	0.0654
MsF18	0.3891	0.0008	0.0374	0.0765
MsF28si	0.2232	0.0007	0.1288	0.0743
MsF22	0.316	0.0007	0.0521	0.0486
MS3	0.0698	0.001	0.064	0.0944
MS4h	0.6836	0.0009	0.0486	0.0319
MsF27s	0.0121	0.0003	0.1656	0.1091
MsF8	0.5224	0.0004	-0.0795	-0.0021
MsF26	0.4802	0.0008	0.0089	-0.0273
MsF4	0.8305	0.002	-0.0415	0.0274
MS27	0.8274	0.0001	0.0564	0.0292
MsF24	No information.			
MS9	0.7836	0.0006	0.0313	0.0096
MsF2	0.5153	0.0006	-0.0166	-0.0131
MS17	0.8267	0.0014	0.001	0.0047
MsF15	0.3783	0.0009	0.0016	-0.0259
MS13	0.8943	0.0007	0.0036	0.0233

All (Fisher's method):

Chi2	45.6274
Df	44
Prob	0.4043

n30 SEC smallest subset
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.5485	0.0013	0.0569	0.0147
MS25	1	0	-0.0314	-0.0178
MsF33	0.0792	0.0024	0.0744	0.0445
MsF31	0.2827	0.0014	0.0909	0.0855
MsF9	0.5037	0.0008	-0.0718	-0.0006
MsF16	0.2954	0.0009	0.1252	0.0736
MsF3	0.0626	0.0005	0.2132	0.2537
MsF17	0.4664	0.003	0.0776	0.0549
MsF18	0.2918	0.0012	0.0239	0.0214
MsF28si	0.315	0.0007	0.0882	0.0403
MsF22	0.1584	0.0007	-0.0294	0.0733
MS3	0.13	0.0033	0.0322	0.0308
MS4h	0.9433	0.0003	-0.0677	-0.0552
MsF27s	0.2693	0.0025	0.0838	0.0255
MsF8	0.0132	0.0001	0.2932	0.2488
MsF26	0.8256	0.0013	0.0034	0.0078
MsF4	0.8303	0.0028	0.0217	0.0059
MS27	0.6562	0.0015	0.0848	0.0499
MsF24	0.7422	0.0005	0.0555	0.0095
MS9	0.1833	0.0009	0.1379	0.128
MsF2	0.0156	0.0002	-0.0087	-0.0317
MS17	0.6664	0.0053	-0.0229	-0.0127
MsF15	0.5173	0.0023	0.0066	-0.001
MS13	0.0876	0.0016	0.0858	0.1083

All (Fisher's method):

Chi2	64.5244
Df	48
Prob	0.0558

n20 SEC other subset
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.7069	0.0015	0.041	0.0014
MS25	0.6376	0.0047	0.0204	0.0017
MsF33	0.7414	0.0034	0.0137	-0.0017
MsF31	0.2856	0.0022	0	-0.0122
MsF9	0.9381	0.0003	-0.0189	-0.0244
MsF16	0.1484	0.0007	0.1726	0.1783
MsF3	0.3024	0.0007	0.1667	0.1913
MsF17	0.4803	0.001	-0.0841	-0.0592
MsF18	0.1396	0.0012	0.2165	0.1717
MsF28si	0.4442	0.0013	0.0806	0.0339
MsF22	0.113	0.0006	0.1633	0.186
MS3	1	0	-0.0188	-0.0103
MS4h	0.0869	0.0007	-0.0217	0.0164
MsF27s	0.5637	0.0015	-0.0395	-0.0331
MsF8	0.5199	0.0007	-0.0133	-0.0372
MsF26	0.8337	0.0011	0.053	0.045
MsF4	0.95	0.0009	-0.0112	-0.0168
MS27	0.0002	0	-0.0212	-0.0103
MsF24	0.1854	0.0005	0.1433	0.1355
MS9	0.1845	0.0008	0.2176	0.1568
MsF2	0.7946	0.0007	0.2083	0.1482
MS17	0.6839	0.0037	-0.0354	-0.0203
MsF15	0.9482	0.0008	0.0172	0.0605
MS13	0.0276	0.0011	0.1461	0.034

All (Fisher's method):

Chi2	62.7671
Df	48
Prob	0.0746

n37 NEC larger subset
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.1888	0.0018	0.044	0.074
MS25	0.3877	0.0046	0.0266	0.0191
MsF33	0.4268	0.0054	0.0238	0.0239
MsF31	0.3216	0.0024	0.1266	0.098
MsF9	0.0894	0.0006	0.0188	0.0195
MsF16	0.8349	0.0005	0.0461	0.0112
MsF3	0.008	0.0001	0.2318	0.2138
MsF17	0.2006	0.0011	-0.0516	-0.0373
MsF18	0.7158	0.0017	-0.0224	0.0042
MsF28si	0.5201	0.0024	0.0775	0.0115
MsF22	0.776	0.0006	0.0584	0.0363
MS3	0.2689	0.005	0.0294	0.0084
MS4h	0.3705	0.001	-0.0312	0.0756
MsF27s	0.5421	0.0032	0.0009	-0.0061
MsF8	0.8697	0.0005	-0.1018	-0.0645
MsF26	0.4317	0.0011	0.1176	0.0622
MsF4	0.2834	0.0024	-0.0303	0.023
MS27	0.1798	0.0017	0.027	0.0435
MsF24	0.1075	0.0004	0.1872	0.1266
MS9	0.1355	0.0009	0.069	0.0246
MsF2	0.6521	0.0022	-0.0347	-0.0137
MS17	0.4039	0.0046	0.0748	0.036
MsF15	0.4165	0.0017	0.0154	-0.0015
MS13	0.4827	0.0032	0.0278	0.0408

All (Fisher's method):

Chi2	:	57.8233
Df	:	48
Prob	:	0.1566

Figure 4.7: Geographic spread of the Queensland sub-population in HWE.

Two overlapping subsets of the SEC sub-population were identified that adhered to HWE (figure A4.10b) but not when combined into a single sub-population.

When the two *Morelia spilota imbricata* island populations were split, both showed overall adherence to HWE, with just a couple of loci each deviating significantly (table 4.6).

Table 4.6: The two *Morelia spilota imbricata* island populations do not deviate significantly from HWE.

n20 Msi GI
Fis estimates

locus	P-val	S.E.	W&C	R&H
MsF5s	1	0	-0.0857	-0.0877
MS25	0.9155	0.0002	-0.0071	-0.0244
MsF33	0.9271	0.0002	-0.0303	-0.0336
MsF31	No information.			
MsF9	1	0	-0.0133	-0.0007
MsF16	1	0	-0.0588	-0.0603
MsF3	No information.			
MsF17	1	0	0	0
MsF18	0.6497	0.0004	-0.1603	-0.1131
MsF28si	No information.			
MsF22	No information.			
MS3	0.2902	0.0006	0.1207	0.0643
MS4h	0.0757	0.0003	-0.1651	-0.0592
MsF27s	0.007	0.0001	0.1303	0.1198
MsF8	No information.			
MsF26	No information.			
MsF4	0.5816	0.0002	-0.0511	-0.1081
MS27	No information.			
MsF24	No information.			
MS9	0.7934	0.0002	-0.0878	-0.0934
MsF2	0.0546	0.0001	0.1504	0.2585
MS17	0.5656	0.0006	-0.174	-0.1184
MsF15	0.0191	0.0001	0.2742	0.0963
MS13	No information.			

All (Fisher's method):

Chi2	35.1636
Df	30
Prob	0.2367

n18 Msi SFI
Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0379	0.0001	0.2649	-0.0049
MS25	1	0	-0.1218	-0.0681
MsF33	0.4654	0.0004	0.0376	0.0018
MsF31	No information.			
MsF9	0.9503	0.0001	-0.1162	-0.087
MsF16	No information.			
MsF3	1	0	-0.0625	-0.0642
MsF17	No information.			
MsF18	0.7648	0.0001	-0.2409	-0.1686
MsF28si	No information.			
MsF22	0.1705	0.0002	0.1959	0.1794
MS3	0.4839	0.0005	-0.0208	-0.03
MS4h	0.5699	0.0005	-0.1421	-0.0728
MsF27s	1	0	-0.1087	-0.1095
MsF8	0.5121	0.0001	0.1414	0.146
MsF26	0.0282	0.0001	-0.0149	0.3857
MsF4	No information.			
MS27	No information.			
MsF24	No information.			
MS9	0.0657	0.0002	0.1373	0.1043
MsF2	0.657	0.0003	-0.1691	-0.0709
MS17	0.2194	0.0006	0.0734	0.0886
MsF15	0.3313	0.0003	0.234	0.117
MS13	No information.			

All (Fisher's method):

Chi2	69.9958
Df	62
Prob	0.227

No pair of loci gave a significant p-value for linkage disequilibrium across all seven populations, suggesting that all 24 loci examined during this study are inherited independently of one another. This conclusion is in line with other research [12] that has observed single locus pairs displaying linkage disequlibrium within a population, but no consistent trend of disequilibrium across populations conforming to HWE.

4.5.4 Allele Frequency Database

Despite population genetic analysis revealing further genetic substructure within some of the sub-populations, none of these sub-populations demonstrated suitability to further dissection for the purpose of forensic statistical calculations. The allele frequency database split into the final six populations is provided in (Table 4.7).

Table 4.7: Allele frequencies of the six sub-populations. Allele number per locus, observed heterozygosity and Polymorphic Information Content (PIC) are provided at the bottom of the table.

	MsF24						MsF16						MsF8						MsF28						MsF26					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	36	67	51	55	20	18	34	67	51	55	20	18	36	67	51	56	20	18	35	66	51	53	20	18	36	67	51	55
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.2	-	-	-	-	-	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	0.049	0.073	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.014	-	-	-	-	-	-	-	-	-
3.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	-	-
5	-	-	-	0.015	-	-	-	-	-	-	0.020	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	0.010	-
5.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	-	-
6	-	-	-	-	0.010	0.100	-	-	-	0.007	0.020	-	-	-	-	-	-	-	-	-	-	0.008	0.010	-	-	-	-	-	0.010	0.045
6.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.3	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.015	-	-	-	-	-	-	-	-
7	1.000	1.000	0.903	0.112	0.029	0.791	-	0.029	-	-	-	0.364	-	-	-	0.037	-	-	-	-	0.029	0.182	0.010	-	1.000	-	-	0.236	-	0.039
7.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	-	-
8	-	-	0.097	0.269	0.245	0.082	0.921	0.971	0.603	0.067	0.147	0.027	-	-	-	0.007	-	-	-	-	-	0.008	0.059	-	-	-	-	0.028	-	-
8.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.015	-	-	-	-	-	-	-	-
9	-	-	-	0.254	0.078	-	-	-	0.103	0.201	0.108	0.091	-	-	-	-	-	-	-	-	0.057	0.258	0.059	0.849	-	-	-	0.015	0.049	-
9.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.014	-	-
9.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.045	-	-	-	-	-	-	-	-
10	-	-	-	0.164	0.265	-	-	-	0.029	0.351	0.167	0.236	-	-	0.042	0.075	0.029	0.063	1.000	-	0.329	0.045	0.176	0.038	-	-	-	0.167	0.015	0.029	0.045
10.1	-	-	-	-	-	-	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.056	-	-	-	-
10.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.030	-	-	-	-	-	-	-	-
11	-	-	-	0.134	0.245	-	-	-	0.088	0.179	0.225	0.055	-	-	0.153	0.157	0.069	0.089	-	-	0.300	0.091	0.196	0.113	-	0.722	0.403	0.052	0.039	0.545
11.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-
11.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.023	-	-	-	-	-	-	-	-
12	-	-	-	0.045	0.118	-	-	-	0.088	0.142	0.137	0.136	1.000	0.806	0.194	0.224	0.157	0.348	-	1.000	0.229	0.098	0.255	-	-	0.222	0.083	0.187	0.069	0.155
12.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.020	-
12.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	-	-
13	-	-	-	0.007	0.010	-	0.079	-	0.074	0.045	0.078	-	-	0.194	0.486	0.306	0.284	0.313	-	-	0.014	0.053	0.137	-	-	-	0.042	0.224	0.078	0.109
13.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	0.015	0.007	0.039	-	-	-	0.125	0.134	0.314	0.179	-	-	-	0.068	0.069	-	-	-	0.028	0.239	0.275	0.027
14.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.015	-	-
14.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.045	0.088	0.009	-	-	0.029	0.015	0.029	-	-	-	-	0.157	0.275	0.073
15.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.049	-	-	-	-	-	-	-	-	-	-	0.075	0.069	-
16.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-
17.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	0.020	-
Total alleles	1	1	2	8	8	5	2	2	7	8	11	8	1	2	5	10	8	6	1	1	8	21	10	3	1	3	8	12	15	7
Ho	0	0	0.0833	0.6567	0.7059	0.2727	0.1579	0.0588	0.6471	0.7463	0.7451	0.5091	0	0.2778	0.6944	0.7463	0.6471	0.625	0	0	0.5143	0.7576	0.7647	0.2264	0	0.4444	0.6389	0.7164	0.8235	0.5273
PIC	0	0	0.1601	0.7756	0.7565	0.3358	0.1349	0.0555	0.5793	0.7463	0.8426	0.7465	0	0.2642	0.6441	0.7786	0.7489	0.6935	0	0	0.7017	0.8569	0.8134	0.2443	0	0.3709	0.7085	0.8011	0.8075	0.6278

Table 4.7 (continued): Allele frequencies of the six sub-populations. Allele number per locus, observed heterozygosity and Polymorphic Information Content (PIC) are provided at the bottom of the table.

	MsF15						MsF5						MsF31						MsF9						MsF3						MsF17					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI		NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	29	68	50	51	20	18	36	68	51	56	20	18	36	68	51	56	20	18	36	68	51	56	20	18	36	68	51	56	20	18	36	68	51	56
4														1.000	1.000	0.074		0.125																		
4.2				0.022																																
4.3																																				
5																																				
5.2																																				
5.4																																				
6											0.049					0.022	0.039	0.089																		
6.1																																				
6.2																																				
6.3																																				
6.4																																				
7				0.007	0.060							0.107				0.029	0.029	0.250																		
7.2																																				
7.3																																				
7.4																																				
8			0.017	0.140	0.050					0.022						0.007	0.049	0.286				0.007					0.125					1.000	0.042			
8.1																																				
8.2																																	0.014			
8.3																																				
8.4				0.007																																
9	0.026		0.069	0.169	0.110					0.015	0.010					0.022		0.161									0.181	0.007					0.056		0.020	
9.1																0.015																				0.473
9.2				0.015													0.010																			
9.3				0.015	0.010																															
9.4																																				
10		0.031	0.207	0.176	0.070	0.010						0.027				0.110	0.029	0.071									0.319	0.044	0.078						0.010	
10.1										0.007						0.088	0.078																			0.009
10.2																																				
10.3				0.007												0.029	0.039																			
10.4																																				
11	0.789	0.656	0.328	0.081	0.130	0.127				0.014		0.010	0.018			0.081	0.020			0.028				0.009			0.069	0.037	0.059				0.007			
11.1					0.030											0.015	0.059																			
11.2	0.026				0.010											0.007		0.018																0.010		
11.3					0.030											0.022																				
11.4																																				
12	0.158	0.313	0.207	0.213	0.160	0.078				0.042	0.029	0.020	0.214			0.029	0.078		0.950	0.389	0.014	0.051	0.039	0.018			0.125	0.051	0.108	0.027			0.083	0.059		
12.1					0.010						0.007		0.027																							0.018
12.2																																			0.010	
12.3					0.030																															
12.4																																				
13			0.052	0.074	0.090	0.225			0.083	0.139	0.125	0.157	0.277			0.066	0.049			0.444	0.222	0.125	0.059	0.268			0.111	0.154	0.127	0.045	0.325		0.264	0.176	0.069	0.027
13.1											0.015	0.010																								0.009
13.2																																			0.010	
13.3					0.020																															
14			0.086	0.044	0.040	0.235	0.900	0.833	0.403	0.426	0.167	0.250				0.059	0.069		0.025	0.139	0.236	0.250	0.186	0.295	1.000		0.056	0.140	0.167	0.107	0.675		0.097	0.088	0.049	0.027
14.1										0.022																										0.018
14.2																																			0.010	
14.3					0.030												0.010																			
15			0.034	0.022	0.040	0.147	0.100	0.083	0.208	0.169	0.127	0.054				0.096	0.127		0.025		0.278	0.213	0.284	0.250		0.917		0.206	0.157	0.196			0.083	0.228	0.088	0.018
15.1										0.007																										0.054
15.2					0.010																														0.029	
15.3																0.029																	0.007		
16					0.040	0.078			0.069	0.103	0.284	0.027				0.081	0.108				0.194	0.206	0.225	0.098		0.083		0.162	0.088	0.179			0.264	0.279	0.186	0.018
16.1																																	0.007			0.116
16.2																																			0.039	
16.3					0.010											0.007																	0.007		
17					0.010	0.049			0.083	0.029	0.098					0.044	0.108				0.042	0.103	0.118	0.063			0.014	0.088	0.078	0.232			0.083	0.088	0.127	0.045
17.1											0.020																									0.071
17.2																																			0.020	
17.3						0.020										0.007																				
18						0.020			0.028		0.020					0.037	0.069				0.014	0.015	0.059				0.066	0.098	0.027				0.037	0.196	0.018
18.1																																				0.054
18.2																																				
18.3					0.010																															
19				0.007		0.010			0.014	0.015	0.010					0.007	0.029					0.029	0.020				0.037	0.020	0.125			0.014	0.015	0.049	0.018
19.1																																				0.009
19.2																																				
19.3																																				
20										0.007	0.020					0.007							0.010				0.007	0.020	0.045					0.049	
20.1																																				
20.2																																			0.010	
20.3																																				
21																													0.018						0.010	
21.1																																				
21.2																																			0.010	
16.3*																0.007																				
Total allele	4	3	8	15	22	11	2	3	9	15	14	9	0	1	1	26	18	7	3	4	7	9	9	7	1	2	8	12	11	10	2	1	10	12	20	17
Ho	0.2632	0.375	0.6897	0.8088	0.92	0.5294	0.2	0.2222	0.6111	0.6471	0.8039	0.6964	0	0	0	0.8088	0.8824	0.4643	0.1	0.7222	0.7222	0.7206	0.8627	0.7321	0	0.1667	0.5278	0.7206	0.7059	0.6964	0.45	0	0.7222	0.7647	0.8824	0.5893
PIC	0.31750	0.38560	0.76220	0.84440	0.91110	0.82120	0.1638	0.2723	0.7317	0.7363	0.8182	0.7694	0	0	0	0.9318	0.9192	0.773	0.094	0.5574	0.7413	0.7952	0.7856	0.7265	0	0.1411	0.7911	0.8547	0.8748	0.8239	0.3425	0	0.8046	0.7948	0.8772	0.7321

Table 4.7 (continued): Allele frequencies of the six sub-populations. Allele number per locus, observed heterozygosity and Polymorphic Information Content (PIC) are provided at the bottom of the table.

	MsF33						MsF18						MS3					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	36	67	50	55	20	18	36	68	51	56	20	18	36	67	51	56
9	-	-	-	-	-	-	-	-	-	0.037	-	-	-	-	-	-	-	-
9.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	0.014	0.015	-	0.018	-	-	-	-	-	-
10.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	0.007	0.010	0.009	-	-	-	-	-	0.009
11.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	0.051	0.010	-	-	-	0.028	-	-	0.009
12.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	0.042	0.066	0.069	0.116	-	-	0.083	-	-	0.098
13.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
13.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	0.675	0.111	0.222	0.110	0.108	0.063	-	0.028	0.028	-	-	0.054
14.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-
14.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14.3	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	0.100	0.722	0.125	0.147	0.127	0.232	-	0.361	-	-	-	0.080
15.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
15.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-
15.3	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	0.167	0.222	0.243	0.186	0.188	-	0.222	0.056	-	-	0.125
16.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
16.3	-	-	-	0.022	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	0.007	-	-	0.075	-	0.167	0.140	0.127	0.205	-	0.361	0.028	0.007	0.010	0.188
17.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.3	-	-	0.014	-	-	-	-	-	-	0.007	-	-	-	-	-	0.007	-	-
18	-	-	-	0.007	-	-	-	-	0.153	0.096	0.118	0.152	-	0.028	0.083	0.022	0.010	0.080
18.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18.3	-	-	0.014	0.015	-	-	-	-	-	-	-	-	-	-	-	0.037	0.020	-
19	-	-	-	0.007	-	-	0.150	-	-	0.037	0.098	0.009	0.125	-	0.167	-	-	0.107
19.1	-	-	-	-	0.020	-	-	-	-	-	-	-	0.075	-	0.069	-	-	-
19.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.015	-	-
19.3	-	-	0.083	0.015	0.020	-	-	-	-	-	-	-	-	-	-	0.015	0.069	-
20	-	-	-	0.022	0.030	-	-	-	0.042	-	0.069	-	0.350	-	0.056	0.022	0.010	0.018
20.1	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	0.069	0.007	-	0.009
20.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.037	0.020	-
20.3	-	-	0.069	-	0.020	-	-	-	-	-	-	-	-	-	-	-	0.039	-
21	-	-	-	0.007	0.050	-	-	-	0.014	-	0.029	0.009	-	-	0.056	0.045	0.029	0.009
21.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.042	0.007	-	0.036
21.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	0.014	0.030	-	-
21.3	-	-	0.056	0.045	-	-	-	-	-	-	-	-	-	-	-	0.060	0.069	-
22	0.250	-	0.014	0.015	-	-	-	-	-	-	-	-	-	-	0.056	-	-	0.009
22.1	-	-	-	-	0.010	-	-	-	-	-	-	-	0.175	-	0.069	-	-	-
22.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	0.045	0.010	-
22.3	-	0.412	0.014	0.045	0.020	0.036	-	-	-	-	-	-	-	-	-	0.045	0.069	-
23	0.225	-	-	0.022	0.030	-	-	-	-	-	-	-	-	-	-	-	-	0.036
23.1	-	-	-	-	0.010	-	-	-	-	-	-	-	0.225	-	0.042	-	-	-
23.2	-	-	-	0.030	-	-	-	-	-	-	-	-	-	-	-	0.030	-	-
23.3	-	0.147	-	0.030	0.020	0.009	-	-	-	-	-	-	-	-	-	0.052	0.088	-
24	-	-	-	0.007	0.040	0.009	-	-	-	0.007	0.020	-	-	-	0.028	-	0.010	0.018
24.1	-	-	-	-	-	-	-	-	-	-	-	-	0.025	-	0.028	-	-	-
24.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	0.037	-	-
24.3	-	0.059	0.125	0.052	0.050	0.127	-	-	-	-	-	-	-	-	-	0.022	0.069	-
25	-	-	-	0.015	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
25.1	-	-	-	-	0.020	-	-	-	-	-	-	-	0.025	-	-	0.007	-	-
25.2	-	-	-	0.045	0.010	-	-	-	-	-	-	-	-	-	-	0.067	0.010	-
25.3	0.200	0.088	0.083	0.060	0.050	0.155	-	-	-	-	-	-	-	-	-	0.007	0.020	-
26	-	-	-	0.007	0.030	-	-	-	-	-	-	-	-	-	-	-	-	-
26.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
26.2	-	-	-	0.060	0.030	-	-	-	-	-	-	-	-	-	-	0.007	-	-

	MsF33						MsF18						MS3					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	36	67	50	55	20	18	36	68	51	56	20	18	36	67	51	56
26.3	0.075	0.294	0.194	0.060	0.060	0.100	-	-	-	-	-	-	-	-	-	0.015	0.029	-
27	-	-	-	0.007	0.050	-	-	-	-	-	-	-	-	-	-	0.007	-	0.009
27.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27.2	-	-	-	0.052	0.020	0.009	-	-	-	-	-	-	-	-	-	0.052	-	0.009
27.3	-	-	0.111	0.007	0.010	0.082	-	-	-	-	-	-	-	-	-	-	0.010	-
28	-	-	-	-	0.050	-	-	-	-	-	-	-	-	-	-	-	-	-
28.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28.2	-	-	-	0.030	0.020	0.027	-	-	-	-	-	-	-	-	-	0.022	-	-
28.3	0.175	-	0.083	0.045	0.020	0.055	-	-	-	-	-	-	-	-	-	0.007	0.020	-
29	-	-	-	-	0.030	0.036	-	-	-	-	-	-	-	-	-	-	-	0.009
29.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
29.2	-	-	-	0.007	0.070	0.018	-	-	-	-	-	-	-	-	-	0.045	-	0.009
29.3	0.075	-	0.014	0.037	0.070	0.018	-	-	-	-	-	-	-	-	-	0.030	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30.2	-	-	-	0.015	0.040	0.027	-	-	-	-	-	-	-	-	-	0.037	0.010	-
30.3	-	-	0.042	0.015	-	0.027	-	-	-	-	-	-	-	-	-	-	0.039	-
31	-	-	-	-	0.020	-	-	-	-	-	0.010	-	-	-	-	-	-	-
31.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31.2	-	-	-	0.007	0.030	0.018	-	-	-	-	-	-	-	-	-	0.015	0.010	-
31.3	-	-	0.042	-	-	0.009	-	-	-	-	-	-	-	-	-	0.015	0.010	-
32	-	-	-	-	0.010	-	-	-	-	0.015	-	-	-	-	-	-	-	-
32.1	-	-	-	-	-	0.009	-	-	-	-	-	-	-	-	-	0.007	-	-
32.2	-	-	0.014	-	0.010	0.018	-	-	-	-	-	-	-	-	-	0.022	-	-
32.3	-	-	0.028	0.007	-	0.036	-	-	-	-	-	-	-	-	-	-	0.020	-
33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33.2	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	-	-	0.010	-
33.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	0.010	0.009
34.1	-	-	-	0.007	-	0.009	-	-	-	-	-	-	-	-	-	-	-	-
34.2	-	-	-	0.007	-	0.045	-	-	-	-	-	-	-	-	-	0.007	-	-
34.3	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-
35.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
35.2	-	-	-	-	-	0.055	-	-	-	-	-	-	-	-	-	-	0.010	-
36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.009
36.1	-	-	-	-	-	0.009	-	-	-	-	-	-	-	-	-	-	0.010	-
36.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	-
36.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.009
37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	0.018
37.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
37.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
37.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
38.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	0.009
38.2	-	-	-	0.015	-	0.018	-	-	-	-	-	-	-	-	-	0.007	0.039	-
39	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.018
39.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
39.2	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	0.029	-
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.009
40.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
40.2	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	-	-	-	-
41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41.1	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
41.3	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-
42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
42.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	-
42.2	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-	-
43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
43.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
44	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
44.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
45	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
45.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
46.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
46.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
48	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
48.2	-	-	-	0.030	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50.2	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
53	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	-
24.2*	-	-	-	-	-	0.018	-	-	-	-	-	-	-	-	-	-	-	-
Total allele	6	5	17	49	35	27	4	3	9	16	15	10	7	5	18	47	46	27
Ho	0.85	0.7059	0.8611	0.9254	0.92	0.8	0.6	0.5556	0.7778	0.8971	0.8039	0.7143	0.7	0.7222	0.75	0.8806	0.9608	0.8571
PIC	0.776	0.6644	0.8921	0.9648	0.9592	0.9217	0.4707	0.3957	0.8084	0.8553	0.8759	0.8048	0.7413	0.6275	0.9192	0.9642	0.9588	0.9006

Table 4.7 (continued): Allele frequencies of the six sub-populations. Allele number per locus, observed heterozygosity and Polymorphic Information Content (PIC) are provided at the bottom of the table.

	MsF22						MS27						MS13						MS9					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	36	68	51	55	20	18	36	68	51	56	20	18	36	68	51	56	20	18	34	67	50	56
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	-	-	-	-	-	-	-
0.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.069	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.039	-	-	-	-	-	-	-
2.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.111	-	-	-	-	-	-	-	-	-
4.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.022	0.127	-	-	-	-	-	-	-
4.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.014	-	0.020	-	-	-	-	-	-	-
5.2	-	-	-	-	-	-	-	-	-	-	-	-	1.000	-	0.431	0.118	0.157	-	-	-	-	-	-	-
5.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.020	-	-	-	-	-	-	-
6.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.2	-	-	-	-	-	-	-	-	-	-	-	-	-	0.972	0.069	0.051	0.147	-	-	-	-	-	-	-
6.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.014	0.007	0.029	0.027	-	-	-	-	-	-
7.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.028	0.081	0.049	-	-	-	-	-	-	-
7.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.054	-	-	-	-	-	-
7.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.042	0.007	0.010	0.241	-	-	-	-	-	-
8.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.037	0.078	-	-	-	-	-	-	-
8.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	-	-	-	-
8.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.028	0.044	0.029	0.607	-	-	-	-	-	-
9.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.2	-	-	-	-	-	-	-	-	-	0.015	0.010	-	-	-	-	0.044	0.020	-	-	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	0.009	-	-	-	-	-	-
9.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.056	0.096	0.010	0.063	-	-	-	-	-	-
10.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.059	0.039	-	-	-	-	-	-	-
10.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.028	-	-	0.010	-	-	-	-	-	-	-
10.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	0.103	0.039	0.109	-	-	-	-	-	-	-	-	-	0.051	0.010	-	-	-	-	-	-	-
11.1	-	-	-	-	-	-	-	-	-	0.015	-	0.098	-	-	-	-	-	-	-	-	-	-	-	-
11.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	0.042	0.059	0.029	-	-	-	-	-	-	-
11.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-
11.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	0.014	0.191	0.108	0.036	-	-	-	-	0.176	-	-	-	0.014	0.022	0.020	-	-	-	0.029	0.037	-	-
12.1	-	-	-	-	-	-	-	-	-	-	-	0.491	-	-	-	-	-	-	-	-	-	-	-	-
12.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.014	0.029	0.020	-	-	-	-	-	-	-
12.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-
12.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	1.000	0.694	0.444	0.360	0.265	0.400	-	-	-	-	0.157	-	-	-	-	0.029	0.020	-	-	-	-	0.037	0.010	-
13.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13.2	-	-	-	-	-	-	-	-	-	0.015	-	-	-	-	0.014	0.029	0.010	-	-	-	-	-	-	-
13.3	-	-	-	-	-	0.018	-	-	-	-	0.078	0.339	-	-	-	-	-	-	-	-	-	-	-	-
14	-	0.194	0.153	0.191	0.167	0.209	-	-	-	-	0.010	-	-	-	-	0.007	-	-	-	-	0.015	0.037	0.040	0.054
14.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.056	0.110	0.020	-	-	-	0.015	-	-	-
14.3	-	-	-	-	-	-	-	-	-	0.037	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	0.111	0.056	0.096	0.176	0.045	-	-	-	0.051	0.010	0.071	-	-	-	0.022	0.020	-	-	-	0.103	0.090	0.130	0.107
15.1	-	-	-	-	-	-	-	-	0.111	0.015	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
15.2	-	-	-	-	-	-	-	-	0.028	0.022	0.029	-	-	-	-	0.015	0.010	-	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-	-	-	0.140	0.049	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	0.044	0.176	0.055	-	-	-	0.028	0.066	0.010	-	-	-	-	0.007	-	-	0.400	0.294	0.059	0.112	0.080	0.241
16.1	-	-	-	-	-	-	1.000	1.000	0.417	0.051	0.078	-	-	-	-	-	-	-	-	-	-	-	-	-
16.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	0.015	-	-	-	-	-	-	-	-
16.3	-	-	-	-	-	-	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	0.056	0.007	0.039	0.118	-	-	-	-	-	-	-	-	-	-	-	-	0.100	0.029	0.279	0.112	0.130	0.116
17.1	-	-	-	-	-	-	-	-	0.403	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.3	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	-	0.007	0.020	-	-	-	-	-	-	-	-	-	-	0.015	-	-	0.125	0.235	0.191	0.254	0.200	0.295
18.1	-	-	-	-	-	-	-	-	0.014	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	-	0.007	-	-
18.3	-	-	-	-	-	-	-	-	-	0.007	0.049	-	-	-	-	-	-	-	-	-	-	-	-	-
19	-	-	0.181	-	0.010	0.009	-	-	-	0.022	-	-	-	-	-	-	-	-	0.375	0.353	0.176	0.179	0.170	0.134
19.1	-	-	-	-	-	-	-	-	-	0.037	0.078	-	-	-	-	-	-	-	-	-	-	-	-	-
19.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
19.3	-	-	-	-	-	-	-	-	-	0.257	0.098	-	-	-	-	-	-	-	-	-	-	-	-	-
20	-	-	0.083	-	-	-	-	-	-	0.051	-	-	-	-	-	-	-	-	-	0.088	0.029	0.082	0.100	0.036
20.1	-	-	-	-	-	-	-	-	-	0.051	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.015	-	-
20.3	-	-	-	-	-	-	-	-	-	0.051	0.029	-	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	0.014	-	-	-	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	0.088	0.030	0.100	0.018
21.1	-	-	-	-	-	-	-	-	-	-	0.069	-	-	-	-	-	-	-	-	-	-	-	-	-
21.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21.3	-	-	-	-	-	-	-	-	-	0.066	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	0.015	-	0.020	-
22.1	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
22.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	0.010	-	-	-	-	-	-	-
22.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
23.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
Total allele	1	3	8	8	9	9	1	1	6	21	21	4	1	2	15	27	29	6	4	5	11	13	12	8
Ho	0	0.3889	0.6111	0.7059	0.7647	0.6364	0	0	0.3333	0.7647	0.8627	0.4821	0	0.0556	0.6944	0.9265	0.8235	0.375	0.75	0.6471	0.7941	0.7612	0.74	0.7321
PIC	0	0.4183	0.7024	0.7445	0.8017	0.7356	0	0	0.5842	0.8795	0.8962	0.5629	0	0.0526	0.7669	0.9322	0.9137	0.5163	0.613	0.6754	0.8097	0.8439	0.8546	0.7823

Table 4.7 (continued): Allele frequencies of the six sub-populations. Allele number per locus, observed heterozygosity and Polymorphic Information Content (PIC) are provided at the bottom of the table.

	MsF27s						MsF4						MS4h						MsF2					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	34	68	50	54	20	18	36	67	51	56	20	18	34	62	47	51	20	18	36	68	51	56
5	-	-	-	-	0.060	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	0.347	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.1	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	0.009	-	-	-	0.067	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	0.037	-	-	0.028	-	-	-	-	-	0.015	-	-	-	-	-	-	-	-	-
8.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	0.015	0.022	-	-	-	-	0.014	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	0.037	-	0.009	-	-	-	-	0.010	0.018	-	-	0.015	0.008	-	-	-	-	-	-	-	-
10.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.3	-	-	-	-	-	-	-	0.028	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	0.450	0.111	0.088	0.250	0.020	0.361	-	0.972	0.083	0.007	0.039	0.071	-	-	-	-	-	0.010	-	-	-	0.007	-	-
11.1	0.125	-	0.015	-	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.2	-	-	-	-	-	-	-	-	-	-	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-
11.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	0.150	0.500	0.279	0.338	0.160	0.213	-	-	0.083	-	0.020	0.018	-	-	-	-	0.011	0.039	-	-	-	-	-	0.080
12.1	0.025	-	0.044	0.007	0.020	0.037	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	0.150	0.389	0.309	0.184	0.230	0.139	0.300	-	0.042	0.007	0.020	0.036	-	-	0.029	-	0.011	0.029	-	-	-	0.007	-	0.036
13.1	0.100	-	0.015	0.007	0.030	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	0.118	0.066	0.130	0.130	0.175	-	0.153	0.060	0.039	0.054	-	-	0.103	0.048	0.043	0.059	-	-	0.014	0.037	-	0.054
14.1	-	-	-	-	0.030	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	0.010	-	-	-	-	-	-
14.3	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	0.008	-	-	-	-	-	-	-	-
15	-	-	0.088	0.015	0.120	0.019	0.525	-	0.042	0.060	0.049	0.045	-	0.028	0.162	0.137	0.085	0.078	-	0.361	0.167	0.059	0.078	0.170
15.1	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	0.007	0.070	0.019	-	-	0.042	0.164	0.069	0.134	-	0.306	0.191	0.161	0.181	0.127	0.150	0.556	0.194	0.125	0.108	0.250
16.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-
16.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	0.007	0.030	-	-	-	0.028	0.172	0.069	0.205	0.050	0.528	0.162	0.185	0.170	0.284	0.175	0.028	0.264	0.162	0.039	0.196
17.1	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17.2	-	-	-	-	-	-	-	-	0.014	0.007	0.029	-	-	-	-	-	-	-	-	-	-	-	-	-
17.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	0.015	0.007	0.010	-	-	-	0.028	0.067	0.069	0.241	0.600	0.056	0.132	0.250	0.213	0.245	0.550	-	0.222	0.213	0.147	0.143
18.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18.2	-	-	-	-	-	-	-	-	-	-	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-
18.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.011	-	-	-	-	-	-	-
19	-	-	-	-	0.020	0.009	-	-	0.042	0.082	0.108	0.116	0.300	-	0.044	0.145	0.096	0.078	0.125	0.056	0.111	0.096	0.196	0.036
19.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19.2	-	-	-	-	-	-	-	-	0.014	-	0.049	-	-	-	-	-	-	-	-	-	-	-	-	-
19.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	-	0.020	-	-	-	0.014	0.082	0.078	0.027	-	0.083	0.088	0.024	0.117	-	-	-	0.014	0.110	0.196	0.027
20.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20.2	-	-	-	-	-	-	-	-	-	0.007	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-
20.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
21	-	-	-	-	-	-	-	-	0.014	0.082	0.059	0.027	-	-	0.044	0.016	0.043	0.039	-	-	0.014	0.074	0.127	-
21.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21.2	-	-	-	-	-	-	-	-	-	-	0.059	-	-	-	-	-	-	-	-	-	-	-	-	-
21.3	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	0.015	0.029	0.009	-	-	0.015	-	0.011	-	-	-	-	0.037	0.059	-
22.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
22.2	-	-	-	-	-	-	-	-	0.014	0.015	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-
22.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	0.007	-	-	0.025	-	-	-	-	-	-	-	-	-	0.020	-
23.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
23.3	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	-	-	-	-	0.010	-	-	-	-	0.015	-	-	0.025	-	-	-	-	-	-	-	-	0.007	0.010	-
24.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.015	-	-
24.2	-	-	-	-	-	-	-	-	-	-	0.029	-	-	-	-	-	-	-	-	-	-	0.007	-	-
24.3	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	0.007	-	-	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
25.2	-	-	-	-	-	-	-	-	-	-	0.020	-	-	-	-	-	-	-	-	-	-	-	-	-
25.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	-	-	-	-	-	-	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	0.007	0.010	-
26.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.010	-
26.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
26.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
27.1	-	-	-	-	-	-	-	-	-	0.007	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
27.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-	-	0.037	-	-	-	-	-	-	-	-	-	-	-	0.007	-	-
28.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
28.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	-	-	-	-	-	-	-	-	0.015	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29.2	-	-	-	-	-	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
29.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	-	0.010	-	-	-	-	-	0.010	-	-	-	-	-	-	-	-	-	-	-	-	-
- -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
- -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
- -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.016	-	-	-	-	-	-	-	-
- -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
- -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
- -	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
38.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.009
Total allele	6	3	11	18	18	13	3	2	17	22	30	13	5	5	12	11	13	11	4	4	8	20	12	10
Ho	0.65	0.6667	0.5294	0.75	0.86	0.537	0.65	0.0556	0.6944	0.8209	0.9412	0.7321	0.65	0.7222	0.8529	0.8548	0.9149	0.7255	0.55	0.6667	0.7778	0.8676	0.8039	0.6786
PIC	0.6937	0.5009	0.7672	0.7528	0.8654	0.7575	0.5317	0.0526	0.8187	0.897	0.9446	0.8396	0.4781	0.5572	0.8561	0.8125	0.8431	0.8021	0.5841	0.4732	0.7731	0.8678	0.8466	0.8172

Table 4.7 (continued): Allele frequencies of the six sub-populations. Allele number per locus, observed heterozygosity and Polymorphic Information Content (PIC) are provided at the bottom of the table.

	MS25						MS17							MS25						MS17					
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB		Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
n	20	18	36	67	51	56	20	18	36	59	51	56		20	18	36	67	51	56	20	18	36	59	51	56
0	-	-	-	-	-	-	-	-	-	-	-	-	23.1	-	-	-	-	-	-	-	-	-	-	-	-
0.2	-	-	-	-	-	-	-	-	-	-	-	-	23.2	-	-	-	0.030	0.029	0.170	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-	-	-	-	-	23.3	-	-	0.014	-	0.010	0.018	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	24	0.050	-	0.125	0.022	0.069	-	-	-	0.153	-	-	-
2.3	-	-	-	-	-	-	-	-	-	-	-	0.009	24.1	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-	24.2	-	-	-	0.015	0.010	0.134	-	-	-	-	-	-
3.2	-	-	-	-	-	-	-	-	-	-	-	-	24.3	-	-	-	0.007	-	0.018	-	-	-	-	-	-
3.3	-	-	-	-	-	-	-	-	-	-	-	0.009	25	-	-	0.097	0.052	0.029	-	-	-	0.153	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	25.1	-	-	-	0.022	-	-	-	-	-	-	-	-
4.2	-	-	0.181	-	-	-	-	-	-	-	-	-	25.2	-	-	-	0.037	0.029	0.071	-	-	-	-	-	-
4.3	-	-	-	-	-	-	-	-	-	-	-	-	25.3	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-	26	-	-	0.014	0.112	0.029	-	-	-	0.069	-	-	-
5.2	-	-	-	-	-	-	-	-	-	-	-	-	26.1	-	-	-	0.007	-	-	-	-	-	-	-	-
5.4	-	-	-	-	-	-	-	-	-	-	-	-	26.2	-	-	-	0.060	0.049	0.018	-	-	-	0.008	-	-
6	-	-	-	-	-	-	-	-	-	-	-	-	26.3	-	-	-	-	-	-	-	-	-	-	-	-
6.1	-	-	-	-	-	-	-	-	-	-	-	-	27	-	-	-	0.127	0.029	-	-	-	0.056	-	-	-
6.2	-	-	-	-	-	-	-	-	-	-	-	-	27.1	-	-	-	-	-	-	-	-	-	-	-	-
6.3	-	-	-	-	-	-	-	-	-	-	-	-	27.2	-	-	-	0.030	0.020	-	-	-	-	0.017	-	-
6.4	-	-	-	-	-	-	-	-	-	-	-	-	27.3	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	28	-	-	-	0.030	0.078	-	-	-	0.014	-	0.010	-
7.2	-	-	-	-	-	-	-	-	-	-	-	-	28.1	-	-	-	-	0.010	-	-	-	-	-	-	-
7.3	-	-	-	-	-	-	-	-	-	-	-	-	28.2	-	-	-	0.022	0.039	0.009	-	-	-	0.008	-	-
7.4	-	-	-	-	-	-	-	-	-	-	-	-	28.3	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	29	-	-	-	0.037	0.049	-	-	-	0.014	-	0.010	-
8.1	-	-	0.097	-	-	-	-	-	-	-	-	-	29.1	-	-	-	-	0.010	-	-	-	-	-	-	-
8.2	-	-	-	-	-	-	-	-	-	-	-	-	29.2	-	-	-	0.030	0.010	0.018	-	0.028	-	0.008	-	-
8.3	-	-	-	-	-	-	-	-	-	-	-	-	29.3	-	-	-	0.007	0.059	-	-	-	-	-	-	-
8.4	-	-	-	-	-	-	-	-	-	-	-	-	30	-	-	-	0.022	0.010	-	-	-	-	0.025	0.010	-
9	-	-	-	-	-	-	-	-	-	-	-	0.009	30.1	-	-	-	0.022	0.078	-	-	-	-	-	-	-
9.1	-	-	0.056	-	-	-	-	-	-	-	-	-	30.2	-	-	-	-	0.010	-	-	0.028	-	-	0.010	-
9.2	-	-	-	-	-	-	-	-	-	-	-	-	30.3	-	-	-	0.015	0.039	-	-	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	-	-	-	31	-	-	-	-	0.010	-	-	-	-	-	-	-
9.4	-	-	-	-	-	-	-	-	-	-	-	-	31.1	-	-	-	0.022	0.039	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	0.018	31.2	-	-	-	0.007	0.010	-	-	0.083	-	0.025	-	-
10.1	-	-	0.014	-	-	-	-	-	-	-	-	-	31.3	-	-	-	0.015	-	-	-	-	-	-	-	-
10.2	-	-	-	-	-	-	-	-	-	-	-	-	32	-	-	-	-	-	-	-	-	-	0.008	0.010	-
10.3	-	-	-	-	-	-	-	-	-	-	-	-	32.1	-	-	-	-	0.029	0.009	-	-	-	-	-	-
10.4	-	-	-	-	-	-	-	-	-	-	-	-	32.2	-	-	-	-	-	-	-	0.167	0.014	0.008	0.029	-
11	-	-	-	-	-	-	-	-	-	-	0.010	0.063	32.3	-	-	-	0.030	0.020	-	-	-	-	-	-	-
11.1	-	-	0.083	-	-	-	-	-	-	-	-	-	33	-	-	-	-	-	-	-	-	0.014	0.017	0.039	-
11.2	-	-	-	-	-	-	-	-	-	-	-	-	33.1	-	-	-	-	0.020	-	-	-	-	-	-	-
11.3	-	-	-	-	-	-	-	-	0.028	-	-	-	33.2	-	-	-	-	-	-	-	0.389	-	0.025	-	-
11.4	-	-	-	-	-	-	-	-	-	-	-	-	33.3	-	-	-	0.015	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	0.010	0.080	34	-	-	-	-	-	-	-	-	-	0.051	0.039	-
12.1	-	-	0.014	-	-	-	-	-	-	-	-	-	34.1	-	-	-	-	-	-	-	-	-	-	-	-
12.2	-	-	-	-	-	-	-	-	-	-	-	-	34.2	-	-	-	-	0.010	-	-	0.222	-	0.008	-	-
12.3	-	-	-	-	-	-	-	-	-	-	-	-	34.3	-	-	-	-	-	-	-	-	-	0.017	-	-
12.4	-	-	-	-	-	-	-	-	-	-	-	-	35	-	-	-	-	-	-	-	-	0.028	0.034	0.059	-
13	-	-	-	-	-	-	-	-	-	-	0.049	0.143	35.1	-	-	-	-	0.029	-	-	-	-	-	-	-
13.1	-	-	0.028	-	-	-	-	-	-	-	-	-	35.2	-	-	-	0.015	-	-	-	0.083	-	0.076	0.049	-
13.2	-	-	-	-	-	-	-	-	-	-	-	-	36	-	-	-	-	0.010	-	-	-	-	0.042	0.029	-
13.3	-	-	-	-	-	-	-	-	-	-	-	-	36.1	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	0.010	0.188	36.2	-	-	-	-	0.010	-	-	-	-	0.034	0.029	-
14.1	-	-	-	-	-	-	-	-	-	-	-	-	36.3	-	-	-	-	-	-	-	-	-	-	-	-
14.2	-	-	-	-	-	-	-	-	-	0.008	-	-	37	-	-	-	-	-	-	-	-	-	0.042	0.039	-
14.3	-	-	-	-	-	-	-	-	0.014	-	-	-	37.1	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	0.042	0.010	0.179	37.2	-	-	-	-	0.010	-	-	-	-	0.034	0.039	-
15.1	-	-	-	-	-	-	-	-	-	-	-	-	37.3	-	-	-	-	0.010	-	-	-	-	-	-	-
15.2	-	-	-	-	-	-	-	-	-	-	-	-	38	-	-	-	0.015	0.020	-	-	-	-	0.025	0.010	-
15.3	-	-	-	-	-	-	0.225	-	0.056	-	-	-	38.1	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	0.017	0.010	0.143	38.2	-	-	-	-	-	-	-	-	-	0.025	0.039	-
16.1	-	-	-	0.007	-	-	-	-	-	-	-	-	39	-	-	-	-	-	-	-	-	-	0.085	-	-
16.2	-	-	-	-	-	-	-	-	-	-	-	0.036	39.1	-	-	-	-	-	-	-	-	-	-	-	-
16.3	-	-	-	-	-	0.018	0.100	-	0.014	-	-	-	39.2	-	-	-	-	-	-	-	-	-	0.076	0.020	-
17	-	0.028	-	-	-	-	-	-	-	0.042	0.010	0.054	40	-	-	-	-	-	-	-	-	-	0.034	0.069	-
17.1	-	-	-	-	-	-	-	-	-	-	-	-	40.1	-	-	-	-	0.010	-	-	-	-	-	-	-
17.2	-	-	-	-	-	-	-	-	-	0.025	-	0.027	40.2	-	-	-	-	-	-	-	-	0.014	0.008	0.029	-
17.3	-	-	-	-	-	-	-	-	0.097	-	-	-	41	-	-	-	-	-	-	-	-	-	0.025	0.020	-
18	-	-	0.014	-	-	-	-	-	-	-	0.010	-	41.1	-	-	-	-	0.010	-	-	-	-	-	-	-
18.1	-	-	-	-	-	-	-	-	-	-	-	-	41.2	-	-	-	-	-	-	-	-	-	0.017	0.039	-
18.2	-	-	-	0.007	-	0.027	-	-	-	-	-	0.009	41.3	-	-	-	-	-	-	-	-	-	-	-	-
18.3	-	-	-	-	-	-	0.025	-	0.028	-	-	-	42	-	-	-	-	-	-	-	-	-	-	0.020	-
19	0.125	0.528	0.014	-	-	-	0.325	-	0.014	0.008	0.010	0.009	42.1	-	-	-	-	-	-	-	-	-	-	-	-
19.1	-	-	-	-	-	-	-	-	-	-	0.010	-	42.2	-	-	-	-	-	-	-	-	0.028	-	0.020	-
19.2	-	-	-	0.015	-	0.071	-	-	-	-	-	-	43	-	-	-	-	-	-	-	-	-	0.008	0.010	-
19.3	-	-	-	-	-	-	-	-	-	-	-	-	43.2	-	-	-	-	-	-	-	-	-	-	0.029	-
20	0.250	0.278	0.042	0.007	-	-	0.125	-	0.069	-	-	-	44	-	-	-	-	-	-	-	-	-	-	0.029	-
20.1	-	-	-	-	-	-	-	-	-	-	0.010	0.027	44.2	-	-	-	-	-	-	-	-	-	0.042	0.039	-
20.2	-	-	-	0.015	0.010	0.107	-	-	-	-	-	-	45	-	-	-	-	-	-	-	-	-	-	0.010	-
20.3	-	-	0.014	0.015	-	-	-	-	-	-	0.010	-	45.2	-	-	-	-	-	-	-	-	-	0.008	0.010	-
21	0.050	0.111	0.028	-	0.020	-	0.150	-	0.014	-	-	-	46	-	-	-	-	-	-	-	-	-	-	0.020	-
21.1	-	-	-	-	-	-	-	-	-	-	-	-	46.1	-	-	-	-	-	-	-	-	-	-	-	-
21.2	-	-	-	0.037	0.020	0.107	-	-	-	-	-	-	46.2	-	-	-	-	-	-	-	-	-	-	0.010	-
21.3	-	-	0.014	0.007	-	-	-	-	-	-	-	-	47.2	-	-	-	-	-	-	-	-	-	-	0.010	-
22	0.100	-	0.028	0.022	-	-	0.050	-	0.056	-	-	-	48	-	-	-	-	-	-	-	-	-	-	-	-
22.1	-	-	-	0.007	-	-	-	-	-	-	0.010	-	48.2	-	-	-	-	-	-	-	-	-	-	-	-
22.2	-	-	-	0.022	-	0.196	-	-	-	-	-	-	50.2	-	-	-	-	-	-	-	-	-	-	-	-
22.3	-	-	-	0.007	0.010	0.009	-	-	-	-	-	-	53	-	-	-	-	-	-	-	-	-	-	-	-
23	0.425	0.056	0.125	-	-	-	-	-	0.056	-	-	-	20*	-	-	-	-	-	-	-	-	-	0.008	-	-
													Total allele	6	5	19	39	40	16	7	7	22	37	45	16
													Ho	0.75	0.7222	0.8056	0.9104	0.9804	0.7857	0.95	0.7222	0.8889	0.7797	1	0.75
													PIC	0.6882	0.5735	0.8937	0.9478	0.9586	0.8673	0.764	0.7229	0.9114	0.9566	0.9659	0.8631

4.5.5 Population specific STR attributes

4.5.5.1 Private alleles and monomorphic loci

Private alleles are alleles specific to a single population [23]. The incidence of private alleles increases the power to discriminate between populations when these alleles arise in the genotype of an individual of questioned origin. They cannot be diagnostic of a population because there is always a chance that they exist at a very low level in populations from which they have not yet been sampled. When applying match probabilities to a genotype that contains alleles not yet sampled in the population from which the allele frequency database arises, a minimum frequency should be applied to these alleles. This could be the 5/2N rule or another measure to account for their low level presence, such as adding two or four alleles to the total database in line with Balding and Nichols [9] correction factor.

Private alleles were seen in all populations in the following numbers (proportion of total alleles in population in brackets): 2 alleles (2.7%) in the *M. s. imbricata* Garden Island population, 3 alleles (4.1%) in the *M. s. imbricata* St. Francis Island population, 34 alleles (13.8%) in the mainland *M. s. imbricata* population, 129 alleles (27.1%) in the northern eastern coastal population, 124 alleles (26.3%) in the southern eastern coastal population and 45 alleles (17.3%) in the southern internal and bredli population.

Fixed loci are loci that exhibit no polymorphism and are of little value to discriminating between populations, unless the fixed allele is private. They can be indicators of founder effects or population bottlenecks [23] and in STRs can also be caused by ascertainment bias. Ascertainment bias is where STR loci have higher repeat numbers in the taxon from which they were developed than related taxa [40,41]. In extreme circumstances can be polymorphic in the species for which they were developed, but monomorphic in divergent species, such as could be expected in cross-amplification studies. Allele fixation was only seen within the *Morelia spilota imbricata* populations. Table 4.8 provides the fixed alleles observed.

Table 4.8: Fixed alleles are shown per locus in the three *M. s. imbricata* populations.

	MsF31	MS27	MsF24	MsF3	MsF28	MsF22	MsF8	MsF26	MS13	MsF17
Msi ML	4	-	-	-	-	-	-	-	-	-
Msi GI	4	16.1	7	14	10	13	12	7	5.2	-
Msi SFI	4	16.1	7	-	12	-	-	-	-	8

As this is a subspecies of the target species, the monomorphic nature of these markers is more likely due to a founder effect or bottleneck in population numbers, rather than ascertainment bias. Locus MsF31 is fixed for allele 4 in all *M. s. imbricata*, so it is beneficial that a second low molecular quality weight marker was included in the dataset for samples of this subspecies that might be degraded. Two loci were fixed for one allele across both island populations. Locus MsF28 was fixed for two different alleles in the two island populations. The Garden Island population shows a high number of fixed loci. This is likely due to both founder effect and genetic drift on this small island of1200 hectares or 12km^2 [42] that is home to a military base and airstrip. The smaller St. Francis Island (809ha or 8km^2, [42] looks to have less allelic fixation at face value, however the fixation indices,

based on heterozygosity ($F= 1 – H_o/H_e$), are very similar for the two islands: Garden island -0.036 +/- 0.026 and St. Francis Island -0.024 +/- 0.029, overlapping in their error ranges.

In many loci across all three *M. s. imbricata* populations, allele frequencies are imbalanced indicating that many loci could be close to fixation by genetic drift alone in these small populations. The use of the STR reference database for these populations will be revisited in Section 4.6.

4.5.5.2 Sequence level population specific characteristics

Section 3.5.6.7 described some sequence level characteristics of interest. It is worthwhile examining whether any of these might have utility in differentiating the six populations. All allele size ranges tend to overlap at a locus with the exception that St. Francis Island alleles are much larger than Garden Island and SIB population alleles at MS17 (table 4.7).

The majority of alleles at MsF28 are complete penta-nucleotide motif repeat units. Only one micro-variant (allele containing an incomplete underlying repeat unit) is seen at MsF28 outside of the NEC population. This is one *M. s. imbricata* individual from Dryandra, a heterozygote exhibiting a 3.2 allele. All other micro-variants occur in the NEC population. Alleles 6-13 have an associated .4 micro-variant. Sequenced individual 30087 that possessed the deletion discussed in section 3.5.6.7 sized as homozygote allele 9.4 and all other sequenced samples did not possess the deletion and sized as alleles representing complete repeat units. The .4 micro-variants almost certainly represent individuals that harbour the 11 bp deletion. This locus is likely to have had a considerable influence on the differentiation of the NEC population from the other subpopulations in the dataset. Examination of the .4 micro-variant represented deletion could prove a useful tool for examining individuals from northern Queensland and the Northern Territory in a forensic context.

Locus MS27 exhibits much higher allelic diversity in the two coastal populations than the SIB population, which is interesting as the population sizes are not very different. The large runs of single base pair micro-variants represent the very complicated nature of the underlying sequence (see figure 3.19 and section 3.5.6.7). This locus contains a 19 bp indel that Taylor [43] reported as specific to the *M. s. imbricata* subpopulations based on observing total adherence in 40 Southwestern carpet pythons and 65 carpet pythons of other subpopulations. Unfortunately, the allele frequency data identifies the two island populations to be fixed for an allele also present in every other population except SIB. The deletion does not cause an obvious distinction in allele size between Southwestern and other populations, as the other non *M. s. imbricata* populations all have alleles of the same size or smaller than the alleles in the *M. s. imbricata* mainland population. Two of the six alleles seen in this population are private. The indel does not have a substantial impact on differentiating this subspecies from others by allele size alone, although the private alleles would influence the statistical weighting, particularly as one has a frequency of 0.42. In conjunction with examining allele frequencies, an assay that specifically targets the indel can provide additional information towards investigations involving this subspecies.

Within MS13, a high proportion of .2 micro-variants were seen in all except the island populations. Taylor [43] discussed a 2 bp indel, which was seen throughout Centralian carpet pythons, but absent from *M. s. imbricata* and polymorphic in all other populations. Some .2 micro-variants occurred in the mainland *M. s. imbricata* population and it would be interesting to know if this is caused by the same underlying repeat region indel or perhaps another unrelated indel because over 40% of the

mainland population possesses an allele representing this micro-variant. Further sequencing would be required to assess the underlying nature of this sequence polymorphism.

Taylor [43] reported locus MS3 to contain a second variable repeat motif $(AAGG)_n$, not seen in any of the individuals sequenced in this study. From the sequencing results reported within that study, all individuals appeared to exhibit sequences composed of complete repeat units. Given the number of .1, .2 and .3 micro-variants seen in the allele frequency data herein, many incomplete tetra repeat units, or other complex motif types of other sizes must be present in individuals possessing higher molecular weight alleles. Allelic patterning at MS3 suggests that smaller alleles at this locus follow a simple motif repeat pattern, whereas larger products mutate into a high number of differentiated micro-variant alleles of complex motif types – likely based on mutation of the underlying repeat. This was not evident from the sequence results, but only individuals representing the smaller alleles 16-19 were sequenced. Alleles greater in length than 19 repeats exhibited many more micro-variants as seen by the larger migration products.

The sequencing results from MS25 illustrate the type of scenario that could be causing this pattern, where alleles possessing longer repeat regions contain the same simple repeat motif as shorter regions, but these have subsequently mutated into a more complex nature. Locus MS25 does show a similar pattern in the allelic database, where shorter alleles separate according to complete repeat motif sizes while larger alleles are more frequently present as micro-variants. MS17 also exhibits the pattern where shorter repeat tracts show simpler repeat patterns, but the larger repeat regions developed an increasingly complicated basis (figure 3.18).

Out of interest, the three MS25 individuals with complete repeat sequences sized as alleles 21.2 (62456), 22.2 (71378) and 25 (68310), while the sample with the dinuculotide tract sized as 35.2 (117064). Comparing these migration rates to the underlying sequence repeats shows that even when motifs appear simple in nature, migration rates to do not translate to underlying fragment size in a straightforward manner at complex STR loci. Sample 68310 (Msi ML population) possesses an AA insertion at the 5' end of the repeat unit. Migration of this sequence translates to allele 25. This indel is presumably the cause of some of the underlying .2 size difference from alleles seen in other populations, for instance SIB alleles of a similar size, but it is indistinguishable from the dinucleotide mutations that appear more common in high molecular weight products. If this indel is population specific, it is ambiguous from size migration alone. This population is offset from other populations at the low molecular weight alleles by 1 bp, but the underlying nature of this polymorphism is unclear without further sequencing.

Although complex in nature and displaying many micro-variants, locus MS17 repeats largely occur in 4bp extensions, mirroring a simple repeat motif expansion. More micro-variants based on incomplete motif expansions are seen in the high molecular weight alleles, as with many other loci. Smaller than allele 20, the *M. s. imbricata* population alleles are offset from the other populations by a single base, creating informative private alleles in these populations and suggesting an underlying indel might exist. None is seen in the sequencing results (figure 3.18), but only a limited amount of flanking region was obtained in the one individuals sequenced from this population. The island populations show very divergent patterns at this locus, particularly the St. Francis Island population. This island shows overlapping allele sizes with the NEC population, but it may be that this is

homoplasious in nature and sequencing would reveal much sequence diversity not captured by genotyping by size migration.

Alleles greater than 23 at MsF4 followed a simple repeat pattern in the NEC population, while all alleles in the SEC population followed a full repeat unit .2 micro-variant iteration until the ultimate allele 30. No sequence level basis for this is evident from the sequence data, however only individuals exhibiting whole units of 21 repeats or fewer were examined. Fewer than allele 23, most .2 micro-variants were seen in the SEC population, although there was some micro-variant overlap down to allele 17 with the NEC population and the Msi mainland population. The SIB population did not contain any micro-variant alleles at this locus. These patterns are quite influential on population differentiation.

4.5.5.3 Observations regarding repeat motif type

Upon scrutinising the patterns of micro-variant alleles in the allele frequency database, some of the complex loci appeared to be based on simple repeat patterns, which then spiral out of simple repeat control with micro-variants of all sizes appearing (e.g. MS17, MS25), while simple repeat loci exhibit more uniform patterns of micro-variant appearance. A question therefore is whether the repeat motif type appears to influence the formation of micro-variants. Table 4.9 summarises the repeat motif types seen across loci in this study. All loci discussed so far (in fact all MS loci plus MsF4) exhibit a CTTT/GGGA (or reverse complement) repeat type. Many of these have exhibited complex repeat types. The loci that do not, MS25, MsF4 and MS9, all demonstrate a tendency to form +/- .2 micro-variants. Except for one *M. s. imbricata* sample from South Australia, all .2 micro-variants at MS9 belonged to the NEC subpopulation, which might be useful for discriminating this population. MS4h provided contradictory data, exhibiting a simple repeat pattern, except for a couple of .1 micro-variants. There is a potentially polymorphic site involving three contiguous runs of polynucleotides 20 bp upstream from the reverse priming site; polynucleotide tracts can cause difficulties during the replication process. Only three individuals have been sequenced at this locus, but a single base pair deletion is seen in one of them within this feature.

Table 4.9: Repeat motif types grouped by similarity.

CTTT/GGGA		AGAT		Singletons	
MS4h	(AAGA)	**MsF5**	(TAGA)	**MsF3**	(ATGA)
MsF4	(CTTT)	**MsF9**	(TATC)	**MsF16**	(CATT)
MS9	(TTTC)	**MsF17**	(GATA)	**MsF24**	(AACC)
MS13	(CTTT)	**MsF18**	(GATA)	**Pentas**	
MS17	(CTTT)	**MsF26**	(TAGA)	**MsF14**	(AAGTT)
MS25	(GAAA)	**MsF27**	(ATCT)	**MsF15**	(TTCTA)
MS3	(GAAA)	**MsF2**	(AGAT)	**MsF28**	(TGATC)
MS27s	(GAAA)	**ATCC/ATGG**		**Mixed**	
MS2	(CTTT)	**MsF22**	(ATCC)	**MsF31**	(TGAA)(TAAA)
		MsF8	(GATG)	**MsF33**	(GACA)(GAGA)$(TAGA)_n$
				MS16	(CAAT)(AAAG)

Seven loci included in this study exhibit variations of TAGA and reverse complements. Locus MsF27 contained a considerable number of .1 micro-variants, but these do not appear to be population

specific. The six individuals sequenced for this locus represent alleles containing whole repeat units. MsF26 likewise shows a disposition for .1 micro-variants and these are similarly spread across populations, although not as commonly occurring as at MsF27. A couple of .3 micro-variants are also seen. Locus MsF5 has an identical motif type and shows the same increased incidence of .1 alleles. Their low incidence makes it unclear whether they are homologous or analogous alleles in terms of population indicators.

Interestingly, MsF2 shares this repeat motif type (TAGA), but the alleles observed retain a simple repeat pattern and .1 micro-variants was not seen until allele 21, which is much longer repeat region than where .1 micro-variants appear in the other loci. Nevertheless, this motif type appears to have a tendency to form +/- .1 micro-variants. MsF18 and MsF9 exhibit one .1 micro-variant and no micro-variants, respectively. Locus MsF17 attracts some interest in the allele types expressed in specific populations. Only complete repeat motif alleles are seen in *M. s. imbricata* populations, except for one 8.2 instant. The NEC and SIB populations exhibit complete repeat alleles or +/- .1 micro-variants. The SEC population however has a high incidence of .2 micro-variants. This unexpected pattern could be caused by an indel unique to this population. No representatives of this population were sequenced at this locus during this study and it would be interesting to perform further sequencing to confirm this. There is an additional adenine nucleotide seen to conclude the repeat region in one sequenced *Morelia bredli* individual that was not seen in any of the other sequenced samples (figure 4.8).

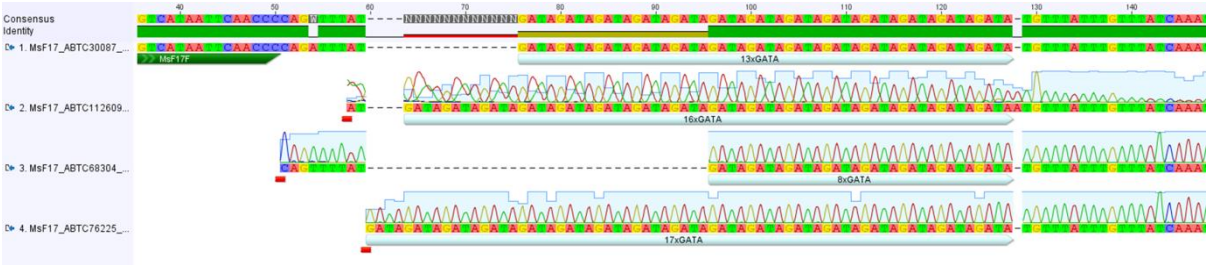


Figure 4.8: A single A base concludes the repeat region in one Morelia bredli individual that is not seen in the other individuals sequenced in this study.

The TAGA based repeat motif in most cases appears susceptible to +/- 1 base mutations, however further sequencing of the +/- .1 micro-variants is required to establish the underlying nature of this phenomenon. Most mutations in these loci are evenly spread across populations or show no correlation to any specific population, so the polymorphisms are unlikely to have an evolutionary basis. They could alternatively possibly have arisen due to multiple independent mutation events in these rapidly mutating STR markers.

Penta-nucelotide repeat locus MsF15 shows a much higher incidence of micro-variants in the two large coastal populations and interestingly micro-variants in these populations do not show much overlap; the lower molecular weight micro-variants occur in the NEC population and the higher molecular weight micro-variants in the SEC population.

Many of the results discussed in this section will have been captured in the population genetic analysis through differences in allele frequencies between populations and have shaped the differentiation into the final six populations. However, at least the MS27 indel and perhaps the MS3

indel also are unlikely to have influenced the population genetic analysis and their independent examination might prove a fruitful exercise when dealing with certain populations.

4.5.6 Summary Statistics, F-statistics and Forensic Statistics

Summary pairwise F_{ST} values for the six overarching populations are shown in table 4.10). Comparison of the two Eastern Coastal populations shows genetic structuring to be quite weak between them. Conversely, the island populations of *Morelia spilota imbricata* appear highly divergent from one another and the rest of the populations.

The global (average) F-statistics across and within these six populations are provided (table 4.11). If one were to use the overall allele frequency database of 249 individuals to calculate match statistics, a theta value of 0.154 would be applied, or the upper bound of 0.194 for an increasingly

Table 4.10: Pairwise F_{ST} values are shown for the six overarching populations.

Msi GI	0.000					
Msi SFI	0.311	0.000				
Msi ML	0.164	0.157	0.000			
NEC	0.206	0.189	0.059	0.000		
SEC	0.212	0.192	0.061	0.015	0.000	
SIB	0.246	0.207	0.079	0.052	0.062	0.000
	Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB

conservative approach. However, it is preferable to use local database as opposed to an all-inclusive database with application of an elevated correction factor (θ), as the result will be more accurate for the local population. Only F_{IS} can be calculated for the single island populations, but the software packages used require multiple populations to calculate F-statistics. These island populations are shown to lack population structure so no θ correction is required for the match probability calculation.

The rarest allele frequency for the overarching population and each of the sub-population databases, as well as the rarest and most common profile genotype frequencies, are provided (table: allele genotype frequency estimates). Using at least 22 loci of this profiling kit, the most common genotype seen using any of the databases has the verbal equivalent of 1 in ten billion, a number far larger than the total population of carpet pythons.

Match probability calculations of genotypes from snakes not included in the reference database are presented shortly.

Table 4.11: Global F-statistics for the sub-populations exhibiting genetic sub-structuring. The global value for each F-statistic is provided, with the 95% lower and upper bounds shown underneath.

Pop.	Definition	F_{IS} (*f*)		F_{IT} (F)		F_{ST} (θ)	
n249	Overarching population	0.111		0.248		0.154	
		0.090	0.132	0.200	0.296	0.115	0.194
Subpop.							
SEC	2 overlapping pops in HWE	0.055		0.063		0.008	
		0.029	0.084	0.036	0.090	0.000	0.019
Msi ML	As defined in K3 (fig. A4.19)	0.085		0.263		0.194	
		0.045	0.131	0.2	0.331	0.145	0.248
NEC	As defined in K4 (fig. A4.12), but two Qld pops merged	0.073		0.118		0.048	
		0.047	0.099	0.094	0.141	0.034	0.064
SIB	As defined in K5 (fig. A4.25), but excluding 71410	0.087		0.224		0.149	
		0.055	0.124	0.182	0.267	0.118	0.182

Table 4.12: Allele and genotype frequency estimates for each population

Population	Rarest allele frequency	Rarest genotype frequency	Most common genotype frequency
Total n249	0.010	1 in 1.22×10^{48}	1 in 7.36×10^{29}
NEC	0.037	1 in 8.56×10^{28}	1 in 1.65×10^{33}
SEC	0.049	1 in 1.37×10^{30}	1 in 1.07×10^{31}
SIB	0.045	1 in 1.55×10^{31}	1 in 4.14×10^{21}
Msi ML	0.069	1 in 1.6×10^{26}	1 in 6.57×10^{21}
Msi GI	0.125	1 in 3.69×10^{19}	1 in 7.46×10^{10}
Msi SFI	0.139	1 in 2.38×10^{18}	1 in 2.77×10^{10}

4.5.7 Geographic Assignment Testing

The Geneclass2 software allows for assignments to be made without and with the use of resampling (a probabilistic estimate of exclusion attached to each population). It is important to perform assignment testing using both approaches. When dealing with an unknown sample, assignment without resampling provides an idea about the population that best fits the individual's genotype - the population with which the individual is most genetically similar. However, it's particularly important to apply resampling as well because this step attaches a probability (or significance value) to their exclusion from each population. This can indicate that the actual population of origin is absent from the dataset if the individual is excluded from all candidate populations.

4.5.7.1 Test for self-assignment

Individuals 81320 (Gunnedah, NSW) and 67128 (Eungella, Queensland) both demonstrated admixture to two population during initial STRUCTURE analyses of the six populations, so these two individual were removed and the reference dataset reduced to 247 individuals for geographic assignment testing.

The assignment of an individual to multiple populations when their own genotype is included in one of the reference populations is a good initial step to perform as it indicates that the designated populations are too genetically similar to be differentiable using assignment methods. However, the Geneclass2 algorithm is designed to take a 'leave one out' approach of [44] cited by [36], so it is not possible to use this software to assign the individual back when its own genotype is present, unless every individual's genotype was duplicated and added by hand under a different identifier. The extra effort to do this was not justifiable, given the tests that follow.

Self-assignment was conducted testing the 247 reference samples against the rest of the dataset, applying the leave one out approach for each individual. All 247 individuals self-assigned to their known population of origin with a score of 100%, demonstrating the robustness of the six populations. When self-assignment was performed with probabilities attached, all Garden Island individuals assigned to Garden Island, but were also not significantly excluded from the Msi ML population, demonstrating the high degree of genetic relatedness between this island and the mainland. One sample is not excluded from the NEC population either. Only five of the 18 St. Francis Island individuals did not give a significant P value reflecting exclusion from the mainland as a population of origin. All SIB assigned to their own population exclusively, except one *Morelia bredli* that was not significant for the NEC population. Only three individuals from the SEC population were not excluded from the NEC population (all assigned to SEC) and a much higher proportion of NEC individuals (23) were not excluded from the SEC population. This is perhaps not surprising given the overlapping range of these two populations. This will undoubtedly cause difficulties in assignment testing of the NEC population.

One individual from the Msi ML population, 68310 gave a significant exclusion value from every population including their own. This result indicates that this individual is genetically distinct from every other individual included in the database. This raised the question of whether they are an appropriate individual to include in the reference database, as they are so genetically different from all other snakes. The decision was made to include the individual in the final reference database as a member of the mainland population because they represent natural variation seen within the population and, if an unknown individual was genetically similar to them, then it would assign to their population rather than no population at all. Individuals that represent the converse situation, not excluded from more than one population are trickier to deal with because they blur the boundary of population differentiation. In cases where they give a high probability of assignment to more than one population, they should be removed.

When the Bayesian analysis method [37] was substituted with the frequency analysis method [39], a higher number of individuals belonging to the NEC and SEC populations assigned incompletely to their population of origin. My findings support the findings of [45] that the Bayesian method employing likelihood probabilities and resampling performed better in assigning individuals completely to a single population of origin than the distance method.

In fact the results presented by Cornuet and colleagues [45] offer some solace regarding the parameters of this study. The authors report that assignment methods are more effective for populations exhibiting strong population structure when few individuals and many loci are used. Populations that are not so well differentiated give more efficient assignment performance when sample size rather than locus number is increased. My dataset better fits the former scenario than

the latter. The reduced sample size in the island populations in particular was concerning, but this publication encouragingly reports that where F_{ST} approaches 0.025, the optimal locus number is 15-20 with a sample size of 12-16 individuals per population. The island populations fit this profile well, so can be expected to perform effectively in assignment tests.

4.5.7.2 Remove 10% of dataset

The removal of 10% of the sample from the dataset was attempted; however this approach is potentially flawed when using such a small reference database. Removing 10% of the genetic information from each population will have a large effect on the resulting allele frequencies. Each subsample of the reference database is going to vary in allele frequency estimates, resulting in more variability in the posterior density sampling and subsequently be markedly less representative of the total allele frequencies in the population than they would if the principal dataset was much larger. An ideal situation would involve compilation of a much larger reference dataset, however this ideal is not going to be the reality for many species implicated in wildlife crimes.

To test the robustness of the analysis when 10% of reference samples are removed, the test was repeated four times removing the same 25 individuals and applying 1000 resampling steps. Table 4.13 shows the results, both without and with sampling. All individuals assigned to the correct population without resampling. With resampling, fifteen samples were non-significant for their source population, four were significant for their source population plus one other population and six individuals were excluded from all populations. Only one assignment changed between non-significant and significant, but this involved assigning an NEC individual to its source population.

4.5.7.3 Remove each population

To test the robustness of the population groupings, each population was removed from the dataset sequentially and assigned back to the dataset without that population present. This determines whether the population constituents are more likely to assign to another population or no population at all when their own population is not present. Twenty individuals from NEC were not excluded from the SEC population. Three individuals from the SEC population were not excluded from the NEC population. One SIB individual (an *M. bredli*) was not excluded from the NEC population. Four SFI individuals were not excluded from the Msi mainland population. All individuals in the above datasets were excluded from all other populations. All Garden Island snakes assigned to the Msi ML population and one individual also assigned to the NEC population. The results of both island populations further support their treatment as separate populations based on their genetic differentiation from one another. All mainland M. s. imbricata individuals were excluded from every other population, showing this population to be highly robust and these mainland individuals to be quite different from the other snakes on mainland Australia.

Table 4.13: Likelihood of assignment shown at left. At right, range of probabilities of exclusion from four replicate tests, non-significant exclusions highlighted in bold, Green = assigned to or not excluded from source population, orange = probability of exclusion from population changed between replicates, blue = not excluded from non-source population.

		Likelihood of assignment (no resampling)				Probability of exclusion using resampling (p-values)					
Sample	Population	1st population		2nd population		Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB
66344	GI	Msi GI	100	Msi ML	0	**0.507-0.538**	0	**0.222-0.240**	0.000-0.003	0.002-0.008	0
66156	SFI	Msi SFI	100	Msi ML	0	0	**0.522-0.553**	0.018-0.021	0.005-0.009	0.003-0.008	0
68298	ML	Msi ML	100	NEC	0	0	0	**0.063-0.070**	0.005-0.009	0.008-0.012	0
66336	ML	Msi ML	100	SIB	0	0	0	**0.416-0.438**	0	0	0
76225	ML	Msi ML	100	NEC	0	0	0	0.002-0.005	0.019-0.025	0.009-0.012	0
13382	NEC	NEC	100	SEC	0	0	0	0	**0.120-0.137**	0.009-0.013	0
66139	NEC	NEC	100	SEC	0	0	0	0	**0.092-0.097**	0.005-0.009	0
69345	NEC	NEC	94.837	SEC	5.163	0	0	0	0.037-**0.054**	**0.062-0.080**	0
82967	NEC	NEC	100	SEC	0	0	0	0	**0.491-0.522**	**0.156-0.172**	0
13446	SEC	SEC	100	NEC	0	0	0	0	0	0.008-0.011	0
32052	SEC	SEC	99.998	NEC	0.002	0	0	0	0.004-0.007	**0.090-0.110**	0
81308	SEC	SEC	100	NEC	0	0	0	0	0.017-0.024	**0.711-0.724**	0
81321	SEC	SEC	100	NEC	0	0	0	0	0	**0.285-0.315**	0
55500	SIB	SIB	100	SEC	0	0	0	0	0	0	0.002-0.006
71395	SIB	SIB	100	NEC	0	0	0	0	0.000-0.001	0	**0.168-0.188**
71430	SIB	SIB	100	NEC	0	0	0	0	0.000-0.001	0	**0.807-0.871**
71441	SIB	SIB	100	NEC	0	0	0	0	0.000-0.001	0	**0.603-0.659**
112609	SIB	SIB	100	NEC	0	0	0	0	0	0.000-0.001	0.004-0.008
76268	SIB	SIB	100	NEC	0	0	0	0	0.000-0.003	0.000-0.002	**0.073-0.088**
106075	SIB	SIB	100	Msi ML	0	0	0	0	0.000-0.001	0.000-0.001	**0.332-0.353**
34289	SIB	SIB	100	NEC	0	0	0	0	0		0.008-0.016
86774	NEC	NEC	99.984	SEC	0.016	0	0	0	0.035-0.047	0.030-0.035	0
66032	ML	Msi ML	100	NEC	0	0	0	**0.155-0.170**	0.000-0.002	0	0
66297	NEC	NEC	99.998	SEC	0.002	0	0	0	**0.392-0.417**	**0.228-0.262**	0
66106	SEC	SEC	99.998	NEC	0.002	0	0	0	0.002-0.005	**0.052-0.073**	0

4.5.7.4 Blind trial - assignment of 'unknown' samples

The results of the assignment testing on the individuals of 'unknown' origin are shown in table 4.14. All individuals for which the original population was known gave >99.4% assignment to the correct population and seventeen of these showed 100% assignment to the correct population. The three individuals for which population of origin could not be determined showed the highest degree of admixture between two populations; the two populations were accordingly the NEC and SEC populations.

Eighteen individuals were not excluded from the population of origin and significantly excluded from all other populations. Four individuals were excluded from all populations, indicating that they are genetically quite divergent from all individuals in the reference database. Only one of these individuals (83875, Kinchega, NSW) was recorded from a locality that could be considered quite remote from any other samples appearing in the reference database. Six individuals were not excluded from more than one population and the source population was included in every case. All cases included non-exclusion from both NEC and SEC; only one individual included non-exclusion to another population, SIB, which was identified as its true population and also the population it was assigned to without resampling. This individual possessed a genotype of just seven complete loci. Correct assignments and exclusions were observed down to the most partial genotype involving only five complete loci.

Table 4.14: Likelihood of assignment and probability of exclusion for twenty-eight samples with partial profiles and one quality compromised complete DNA profile. Green = Correct assignment or non-exclusion; Blue = excluded from all populations; Purple = not excluded from >1 population. *samples that fall in overlapping range of NEC and SEC, most likely population is speculated by proximity of sampling location to one sub-population or the other.

Likelihood of assignment (no resampling)							Probability of exclusion using resampling (p-values)								Provenance	
Sample	1st Population		2nd Population		3rd Population		Msi GI	Msi SFI	Msi ML	NEC	SEC	SIB	No. of loci	Population	Recorded locality	Population
66113	NEC	100	SEC	0	Msi ML	0	0	0	0	**0.5515**	0.0193	0	21	NEC	Black Mountain, Qld	NEC
66171	Msi SFI	100	Msi ML	0	NEC	0	0	**0.4223**	0.0076	0.0108	0.0001	0	21	Msi SFI	St Francis Island, SA	Msi SFI
13429	NEC	100	SEC	0	Msi ML	0	0	0	0	**0.0829**	0.0038	0	21	NEC	Fanny Bay, Darwin, NT	NEC
68344	SEC	100	NEC	0	Msi ML	0	0	0	0	0.0016	**0.8337**	0	20	SEC	Byabarra, NSW	SEC
66176	Msi SFI	100	Msi GI	0	Msi ML	0	0	**0.824**	0.0068	0.0036	0	0	19	Msi SFI	St Francis Island, SA	Msi SFI
67148	Msi ML	100	NEC	0	SIB	0	0	0	**0.8662**	0.0028	0.0001	0	19	Msi ML	Dryandra, WA	Msi Msi ML
71370	SIB	100	NEC	0	SEC	0	0	0	0	0.0002	0.0001	**0.7182**	19	SIB	Mt Hope, Vic	SIB
71409	SIB	100	NEC	0	Msi ML	0	0	0	0	0	0	**0.2282**	19	SIB	Piambie, Vic	SIB
83875	SIB	100	NEC	0	SEC	0	0	0	0	0	0	0.0063	19	Excluded from all	Kinchega NP, NSW	SIB
13432	NEC	100	SEC	0	Msi ML	0	0	0	0	**0.3931**	0.009	0	18	NEC	Botanic Gardents, Darwin, NT	NEC
13459	NEC	100	SEC	0	Msi ML	0	0	0	0	**0.206**	0.0013	0	18	NEC	Darwin, NT	NEC
13457	NEC	99.997	SEC	0.003	Msi ML	0	0	0	0	**0.2641**	**0.0756**	0	17	NEC, not excl from SEC	Darwin, NT	NEC
66031	NEC	63.371	SEC	36.629	Msi ML	0	0	0	0	0.016	0.0358	0	17	Excluded from all	Toowoomba, Qld	SEC?*
79112	SEC	88.65	NEC	11.35	Msi ML	0	0	0	0	0.0007	0.0015	0	17	Excluded from all	Brisbane, Somerset Dam, Qld	SEC?#
17537	SIB	100	NEC	0	Msi ML	0	0	0	0	0.0001	0	0.0457	16	Excluded from all	3k East of Morgan, SA	SIB
50868	SIB	100	NEC	0	SEC	0	0	0	0	0	0	**0.2083**	16	SIB	Bobbie Moonga WH, SA	SIB
81309	SEC	99.48	NEC	0.52	Msi ML	0	0	0	0	0.045	**0.0971**	0	15	SEC	Saratoga, NSW	SEC
66334	Msi ML	100	NEC	0	Msi SFI	0	0	0	**0.082**	0	0	0	12	Msi ML	North Bannister, WA	Msi Msi ML
12150	NEC	99.966	SEC	0.033	SIB	0.001	0	0	0	**0.3414**	**0.0742**	0.0007	10	NEC, not excl from SEC	Top end, NT	NEC
66337	Msi ML	100	NEC	0	SEC	0	0	0	**0.451**	0.0355	0.0155	0	10	Msi ML	Boyogin Rock Nat Res, WA	Msi Msi ML
66317	SEC	96.78	NEC	3.22	Msi ML	0	0	0	0	**0.2548**	**0.5696**	0	9	SEC or NEC	Cedar Creek Falls, Mt Tamborine, Qld	SEC?#
68380	SIB	100	NEC	0	SEC	0	0	0	0	0.024	0.004	**0.1058**	9	SIB	2k Nth Clare, SA	SIB
12154	SIB	99.991	NEC	0.009	SEC	0	0	0	0	0.0476	0.0055	**0.1177**	8	SIB	Todd River, Alice Springs, NT	SIB
81322	SEC	99.695	NEC	0.305	Msi ML	0	0	0	0.0011	**0.5505**	**0.9633**	0	8	SEC or NEC	Myocum, NSW	SEC
71415	SIB	99.589	NEC	0.22	SEC	0.191	0	0	0	**0.5971**	**0.6741**	**0.6223**	7	SEC or SIB or NEC	Piambie, Vic	SIB
13444	NEC	99.851	SEC	0.126	SIB	0.023	0	0	0	**0.4724**	0.0445	0.0006	5	NEC	Fanny Bay, Darwin, NT	NEC
13450	NEC	99.681	SEC	0.319	Msi ML	0	0	0	0	**0.8843**	**0.3768**	0	5	NEC or SEC	Botanic Gardents, Darwin, NT	NEC
55457	Msi SFI	100	Msi ML	0	NEC	0	0	**0.8105**	0.0131	0.0053	0.002	0	24	Msi SFI	St Francis Island, SA	Msi SFI

On closer examination of 71415 that was not excluded from three populations by its seven locus genotype, all of the alleles it possessed at six of the seven loci ranked within the four most common alleles at those loci and the alleles at the last locus ranked fifth and sixth most common of the 24 alleles seen at that locus. This genotype is particularly common within the dataset.

4.5.7.5 Summary of geographic assignment testing

Four of the six overarching populations have demonstrated highly robust population assignment. The two populations for which assignments were less certain, the NEC and SEC populations, possess a large area with overlapping ranges. It was clear from the population genetic analysis of these two populations that they would likely prove problematic for geographic assignment attempts. However, the assignments that showed poorer performance dealt with individuals nearing the interface of these two populations. Individuals belonging to these populations that placed geographically distantly from the interface showed reliable population assignment, suggesting that geographic assignment outside of the central coastal region of eastern Australia might still have merit for investigative purposes. These data have shown promising results that indicate that Australian carpet pythons taken from most areas of their native range can be predicted to have originated from one of the six overarching populations with a high degree of statistical support.

4.5.8 Match Statistics of 'Unknown' Samples

Match probabilities and associated likelihood ratios are provided for each of the 'unknown' individuals in the previous section, using the allele frequency database corresponding to the assigned population (table 4.15). In every case, the use of the conservative theta value only changes the resulting probability by one to two orders of magnitude. Given the magnitude of the results in most cases, this makes very little difference to the outcome. Many of the results calculated using the 95th percentile upper bound theta value exceed the verbal equivalent ceiling figure of one in one

billion that is sometimes adopted as a court reporting threshold (this value was used during my employment at Forensic Science SA, South Australia), or even one in one million, the ceiling probability reported by the Swedish National Laboratory of Forensic Sciences [46].

A complete 24 locus profile was obtained from sample 55457, however for reasons discussed shortly, only 10 loci were used to calculate probabilities. The resulting likelihood ratio is quite low, but accordingly this individual was noted to exhibit most of the common alleles in this island population.

Table 4.15: Forensic statistics calculated for each of the 'unknown' samples using the appropriate calculation sheet indicated by the assignment results (table 4.14). Assigned population specifies match probability calculation sheet used for each calculation. The overall n249 calculation sheet was used for individuals that were excluded for all populations. MP = Match Probability, the cumulative match probability over all loci present; Cons. θ = the global θ value from Table (global Fs) was replaced with the conservative upper 95th percentile θ value. Verbal equivalent is a conversion of conservative likelihood estimates less than 1 in 100 billion or 10^{11}. Individual * was included as a quality compromised sample not included in the reference database; this individual was not included in geographic assignment, but was known to originate from St. Francis Island.

Unknown sample	# loci	Assigned Population	Calc. sheet	MP	MP (Cons. Θ)	LR	LR (Cons. Θ)	verbal equivalent
55457	24	Msi SFI* (10 locus reduced set)	Msi SFI	0.00025815	-	**3873.733**	-	1 in 1 000
66113	21	NEC	NEC	1.18575E-28	4.36E-27	**8.43349E+27**	**2.29E+26**	
66171	21	Msi SFI (10 locus reduced set)	Msi SFI	7.83703E-06	-	**127599.4241**	-	1 in 100 000
13429	21	NEC	NEC	2.34846E-28	1.33497E-26	**4.25811E+27**	**7.4908E+25**	
68344	20	SEC	SEC	4.58675E-33	2.07354E-31	**2.18019E+32**	**4.82268E+30**	
66176	19	Msi SFI (10 locus reduced set)	Msi SFI	0.00051252	-	**1951.144913**	-	1 in 1 000
67148	19	Msi ML	Msi ML	2.13571E-14	2.30141E-13	**4.68229E+13**	**4.34516E+12**	
71370	19	SIB	SIB	1.20038E-16	1.27775E-15	**8.33071E+15**	**7.82624E+14**	
71409	19	SIB	SIB	1.24531E-15	1.43788E-14	**8.03012E+14**	**6.95467E+13**	
83875	19	Excluded from all	n249	4.82128E-21	3.78067E-19	**2.07414E+20**	**2.64503E+18**	
13432	18	NEC	NEC	6.6079E-25	1.73517E-23	**1.51334E+24**	**5.76312E+22**	
13459	18	NEC	NEC	8.24917E-25	3.09138E-23	**1.21224E+24**	**3.2348E+22**	
13457	17	NEC, not excluded from SEC	NEC	9.01568E-25	2.99466E-23	**1.10918E+24**	**3.33928E+22**	
			SEC	3.33963E-30	2.29388E-28	**2.99434E+29**	**4.35943E+27**	
66031	17	Excluded from all	n249	3.90813E-18	1.61686E-16	**2.55877E+17**	**6.18482E+15**	
79112	17	Excluded from all	n249	1.8302E-20	1.02279E-18	**5.46389E+19**	**9.77717E+17**	
17537	16	Excluded from all	n249	1.61598E-18	4.57101E-17	**6.1882E+17**	**2.1877E+16**	
50868	16	SIB	SIB	5.0927E-15	7.49398E-14	**1.96359E+14**	**1.33441E+13**	
81309	15	SEC	SEC	8.91938E-27	2.84542E-25	**1.12115E+26**	**3.51442E+24**	
66334	12	Msi ML	Msi ML	1.47396E-10	1.16456E-09	**6784437408**	**858695618.7**	1 in 100 million
12150	10	NEC, not excluded from SEC	NEC	1.00042E-14	6.47701E-14	**9.9958E+13**	**1.54392E+13**	
			SEC	9.01668E-18	9.76476E-17	**1.10906E+17**	**1.02409E+16**	
66337	10	Msi ML	Msi ML	1.24601E-08	7.02379E-08	**80255887.88**	**14237335.14**	1 in 10 million
66317	9	SEC or NEC	SEC	2.43409E-14	7.77789E-14	**4.10831E+13**	**1.2857E+13**	
			NEC	1.98571E-13	8.55568E-13	**5.03597E+12**	**1.16881E+12**	
68380	9	SIB	SIB	2.13445E-08	7.02393E-08	**46850425.87**	**14237036.23**	1 in 10 million
12154	8	SIB	SIB	1.0407E-08	2.88469E-08	**96089608.22**	**34665714.75**	1 in 10 million
81322	8	SEC or NEC	SEC	4.03694E-12	1.23316E-11	**2.47712E+11**	**81092593182**	1 in 10 billion
			NEC	8.53518E-12	3.55208E-11	**1.17162E+11**	**28152499567**	1 in 10 billion
71415	7	SIB, SEC or NEC	SIB	2.02305E-06	4.55559E-06	**494302.5782**	**219510.3678**	1 in 100 000
			SEC	2.92363E-12	1.46929E-11	**3.42041E+11**	**68059914294**	1 in 10 billion
			NEC	2.77069E-08	7.73724E-08	**36092073.57**	**12924504.64**	1 in 10 million
			n249	1.36724E-07	4.38982E-07	**7314003.541**	**2277997.055**	1 in 1 million
13444	5	NEC	NEC	5.29892E-07	1.20984E-06	**1887178.178**	**826555.8278**	1 in 100 000
13450	5	NEC or SEC	NEC	2.02747E-07	3.98203E-07	**4932244.954**	**2511281.961**	1 in 1 million
			SEC	4.64328E-10	1.7952E-09	**2153649675**	**557039628**	1 in 100 million
86801	5	NEC or SEC	NEC	1.06052E-07	2.67257E-07	**9429325.986**	**3741717.485**	1 in 1 million
			SEC	5.68373E-09	2.27222E-08	**175940725.2**	**44009748.24**	1 in 10 million

4.6 Further discussion – Considerations and Application

4.6.1 Choice of reference dataset

Choosing the number of individuals for inclusion in the reference database was an important consideration in this experimental design. A balance must be struck between maximising the sample size of the populations (particularly crucial in a dataset with such sparse sampling in large areas) and minimising the amount of missing data caused by incomplete loci. The sampling of populations proved an important factor in this decision. Individuals at the threshold of inclusion were assessed upon whether they would fill in any sampling areas or provide no added benefit to already heavily sampled regions.

Initially, a reference dataset involving 242 individuals with either complete profiles or only one missing locus was used. This dataset was subjected to population genetic analyses. By allowing a second locus with missing data, a reference dataset of individuals exhibiting 22+ complete loci – twelve extra samples could be added. Reference to the population analytical results showed that, not only did these extra samples provide important information for each dataset in the form of rare alleles, but their inclusion would improve coverage in poorly sampled areas. The reference dataset was amended to include these individuals, providing a total of 254 individuals exhibiting 22 or more complete loci.

Part way through reanalysis, this dataset was found to be inappropriate due to the presence of 5 pairs of duplicate genotypes detected using GenAlEx (Peakall & Smouse, 2006, 2012) Multilocus > Matches function. It is unclear whether these duplicate genotypes are due to accidental resampling of an individual during initial sampling, or perhaps mislabelling during post-processing. Quality controls used during testing indicate that these genotypes are the correct genotypes obtained from the extracts received at Flinders University (however only one sample 71407 was extracted at Flinders University, the rest being extracts obtained from Taylor's doctoral project). Only one individual was replicated in the results of that previous PhD project [43] and all genotypes were identical at the four loci overlapping between studies.

Pairs of snakes (one pair excluded) are recorded as sampled from the same locality indicating the possibility of recapture of the same individual or sampling of two related individuals. It is unlikely the results are caused by siblings with matching profiles at 24 loci as the PI_{SIBS} calculated in the previous chapter for this dataset was 1 in 10^{-12}. One pair originated from locations ~40 km apart, but the sampling dates are not available. Due to the ambiguity surrounding these duplicates, one of each pair was removed from the final dataset for analysis. The resulting reference dataset was comprised of 249 individuals exhibiting 22 or more complete loci.

The population genetic analysis to partition the overall reference population into sub-populations gave some unexpected results and raised some interesting population genetic questions and dilemmas for forensic application.

4.6.2 Homogeneity of Eastern Coastal populations

The first surprise encountered was the degree of homogeneity in the Eastern Coastal populations. The Eastern coastline of Australia inhabited by carpet pythons stretches roughly 3000 kilometres, adding a further 1500+ kilometres if the distance from the Queensland coast to Darwin overland is

included. While the snakes at either extreme of this range were quite genetically divergent, this area was partitioned into just two populations using STRUCTURE and showed weak isolation by distance. STRUCTURE has been noted to have difficulty with populations exhibiting isolation by distance [4] and it's possible that this coastal population comprises one continuous population, split into NEC and SEC as an artefact of the clustering analysis algorithm. Particularly confounding to the prospect of population assignment is the overlapping range of roughly 1300 kilometres between the northern-most SEC assigned snake and the southern-most NEC assigned snake (figure 4.9).

Figure 4.9: Map of the geographic distributions of samples belonging to NEC and SEC sub-populations.

The work of Taylor [43] also showed snakes from northern and eastern Australia to lacking the genetic structuring required for fine scale geographic assignment using eight STR loci. The homogeneity in these populations is indicated by their F_{ST} values. The SEC population showed the least genetic differentiation, with a global FST of 0.008, while the NEC population showed a larger but still relatively low sub-structuring of 0.048. The divergence in this figure is likely caused by the inclusion of the Northern Territory group and indeed the overseas individuals from Papua New Guinea.

Asides an absence through the west of Northern Queensland below the Gulf of Carpentaria that spreads into the north-east of the Northern Territory, snake populations (and their sampling for this study) are fairly continuous throughout the Australian mainland NEC and SEC populations. The genetic homogeneity indicates dispersal rates of snakes throughout this region to be adequate for gene-flow to prevent stronger patterns of isolation by distance (calculation of migration rates is beyond the scope of this study). Much genetic variability is seen in these populations, including the incidence of a large number of micro-variants across many alleles. For example, four loci in the SEC dataset displayed a highest allele frequency of <0.1, indicating these loci to be highly discriminative.

Yet, the population patterns are not desirable for assignment of individuals to a particular geographic location, one of the intended aims of this project. The hope was that an increased number of loci would mitigate the assignment issues encountered by Taylor, however even with

three times the number of loci, genetic differentiation of local populations within these sub-populations is still low.

4.6.3 Dilemmas raised by sparsely sampled, genetically divergent populations

In contrast to the above, the *Morelia spilota imbricata* mainland and Southern, Internal and Bredli populations showed a rather high degree of genetic structuring, further complicated by sparse sampling over large geographic distances. Correct implementation of the population analysis software was difficult in these respective populations because, due to these factors, it is possible that genetically dissimilar populations were represented by single individuals. Ideally, population genetic analysis should be conducted on multiple populations containing a minimum of two dozen or so individuals.

Each of these datasets contained a geographic locality that had been more intensively sampled (although sample numbers in the teens still falls short of idealised population numbers). It was difficult to find clusters of individuals that exhibited HWE outside of these small areas, unsurprising given the geographic spread of the snakes. So these datasets provided the dilemma that the populations are too small to provide accurate results from HWE calculations, but the larger populations are too diverse to pertain to HWE, violating assumptions of many population genetic analyses. I have been advised that violation of HWE is problematic, but can be dealt with so long as the reason for the deviation is known and can be explained (R. Ogden, *pers. comm.*).

Likewise, the overall allele frequencies of the total combined population would not accurately represent many of the genetically diverse individuals included within, but further subdivision would result in sample numbers too small to give accurate analytical outcomes. The best solution to the allele frequency dilemma lies in the application of the theta correction to account for the known elevated sub-structuring. However, even the calculation of the global theta value was problematic, being an average of the pairwise F_{ST} values between groups. How can these populations be further accurately subdivide into groups, when the genetically similar groups contain so few individuals that pairwise F_{ST}, and therefore theta, values would not be accurate?

Talking to some of the sharpest, most experienced minds in the application of wildlife and statistical forensic genetics resulted not in infallible solutions to this “pickle” (as one called it), but rather recommendations of pragmatism. I came to the conclusion that the best way forward is to acknowledge the limitations inherent in the dataset and lean towards the conservative approach in reporting. After all, in the field of wildlife forensic science, the species dealt with are sought after due to their rarity, and populations fragmented due to persecution. Lacking sample size is frequently an issue [12] and both deficient resources and sheer depletion of the species dictates that increasing this is often not an option. This example perhaps represents the reality for many populations encounter in wildlife forensic investigations. As wildlife forensic scientists, we need to be able to make some workable solution of these types of problematic dataset. It is better to be transparent about the limitations, utilise limited information and provide a conservative result, than to scrap an imperfect dataset and provide no assistance at all.

4.6.4 Choosing the right kind of markers for the job

From the outset, this part of the project aimed to answer two questions commonly encountered in forensic investigations: the question of individualisation and the question of geographic origin. Both of these questions can be informed through the application of highly variable STR markers. The work of Taylor [43] demonstrated that eight STR markers were insufficient for conducting geographic assignment, outside of differentiating the western imbricata sub-species from the rest of *Morelia spilota* (although the mitochondrial analyses of chapter 2 suggested the *M. s. imbricata* subspecies to be divergent enough to warrant elevation to a separate species of *Morelia*). The development of a much larger marker set was intended to resolve this problem. An equal expectation was that an increased marker set would likewise improve the discrimination power of matching statistics, benefiting cases involving individualisation.

Not only was the marker set increased in number, but compound and complex STR motif types were included, which provided a much larger incidence of micro-variants at each locus. The resulting 24 locus marker set, containing 752 alleles seen in 249 individuals, exceeded expectations in terms of increasing the discrimination power of the assay. High allele counts are excellent for discriminating closely related individuals in cases of parentage testing and determination of geographic origin. However they created unforeseen complications when applying allele frequency databases to answer questions of individualisation.

The high degree of genetic structure in this species creating these large allele counts also resulted in very small sample sizes when populations were partitioned by genetic similarity. The largest number of alleles seen at a particular locus was 88 at MS3, however sample sizes of the island populations were less than two dozen individuals, facilitating possible sampling of less than half the alleles known to exist at MS3, if every individual happened to exhibit a different allele. One could argue that the historical founder effect of this island population combined with the increased genetic drift likely to be strong in a small population would be likely to remove most of the rare alleles in the mainland populations from these group. However, with such small sample sizes it is still difficult to determine whether the sampled allele frequencies accurately represent the true allele frequencies of the entire population. The inclusion of a large number of highly polymorphic markers was gradually recognised to involve a trade-off between highly discriminating markers preferable for geographic assignment and an increased number of alleles confounding accurate allele frequency estimates in small populations.

4.6.5 Difficulty of accurate allele frequency estimates in small island populations

Back to the problem of how to treat the two *M. s. imbricata* island populations, Garden Island and St. Francis Island, which have reference datasets of only 20 and 18 individuals respectively. This is lower than the recommended population numbers for a reference database and one might be tempted to combine them. However, examination of respective allelic frequencies reveals that these two populations are very divergent at many loci, supporting the genetic differentiation of these islands seen from the population genetic analyses. Combining them into a single population would grossly misrepresent allele frequencies for both populations. These analyses have demonstrated that these populations are too genetically distinct to combine with each other or any mainland populations.

The ability to confidently estimate the allele frequencies at a locus exhibiting 88 alleles is confounded when only 36 alleles at that locus have been sampled within the target population. A workable solution to overcome this issue is to use only the most conserved loci, identified in the larger population, for the match probability calculation. Application of just the eight most conserved loci in the Garden Island population (MsF5s, MsF9, MsF16, MsF18, MS4h, MsF27s, MS9, MsF2) gives a Probability of Identity of 1 in 9523. While this figure sounds low in comparison to human populations, Garden Island is 1200 hectares or 12km^2 [42] and the population density would need to be ~600 snakes/ km^2 for the population to be this size. An estimate of a population size of 5000-10 000 snakes has been provided (David Pearson *pers. comm.*). In the context of a case, the population estimation might need to be revisited and a figure closer to the upper end of this initial estimate might warrant the addition of a couple of extra loci in the calculation.

A similarly small reference dataset is available for St. Francis Island. Using the eight most conserved loci in the St. Francis Island population (MsF5s, MsF9, MsF16, MsF3, MsF22, MsF8, MsF26, MS9) a probability of identity 1 in 1687 snakes is observed. If this figure is considered insufficient for adequate discrimination, adding two further markers MS4h and MsF18 gives a probability of identity of 1 in 22 800, which should be more than adequate for a population inhabiting an island 809ha or 8km^2 [42] in size. The highest allele count at any of these markers is 21 alleles.

The full 24 locus set is recommended for geographic assignment testing, which is not confounded by the small sample size and benefits from the increased information content.

4.6.6 Robustness of assignment testing

All individuals DNA profiled in this project that exhibited too few complete loci for inclusion in the reference database were used to test the robustness of the developed assay and subsequent statistical analyses. One further individual was screened using the wrong sample details. Although this sample could be traced back to the right individual of origin, its use in the reference DNA database was rejected and it was conversely treated as an unknown for robustness testing.

The assignment testing produced promising results both in terms of overall reference database robustness, as well as ability to assign DNA not contained within the dataset to the six putative populations or exclude from all populations. During initial robustness testing of the reference database using only the 249 included individuals, only two individuals showed admixture between multiple populations to the degree that they needed to be discarded as references for assignment purposes. The Msi Mainland population was shown to be most robust; all individuals were excluded from every other population. The NEC and SEC populations showed a degree of overlap problematic for assignment testing. This came as no surprise given the homogeneity in these populations discussed previously and the possibility that the population split could actually be an artefact of the cluster based analysis.

The 28 unknown samples gave pleasing assignment results. Assignment to a single population gave the correct known sampling origin, except in three cases involving individuals that occur around the southern boundary of the NEC range, well within the range of the SEC sub-population. In all three cases, the likelihood of assignment was shared between these two sub-populations. The probability of exclusion results showed four individuals to be excluded from all populations. This indicates that they are genetically divergent from the other individuals in all populations. Two of these were

individuals showing admixture in the assignment to a single individual. All other cases where individuals were not excluded from multiple populations involved the NEC and SEC populations. Of interest, one individual could not be excluded from SIB, as well as NEC or SEC, however this individual only exhibited seven complete loci.

Perhaps the most promising outcomes of this testing are the observations that: 1) only one individual with more than ten complete loci was excluded from all but a single population (discounting the four individuals excluded from all populations) and, 2) an individual containing only five complete loci was excluded from all but a single population. These data demonstrate four of the chosen datasets to be highly robust for geographic assignment testing, while the NEC and SEC populations show decreased robustness for some assignments.

4.6.7 Discriminating power of the dataset

The results presented in table 4.15 provide an idea of the discriminating power of the dataset when dealing with partial profiles not already represented on the database. The likelihood ratio incorporating a conservative theta value is most useful in the court reporting context. The results were generally very discriminatory, even when less than ten loci were used in calculations. The St. Francis Island population gave consistently low match probabilities, due to the restricted locus set used for calculations. Yet, depending on the population size inhabiting the island, these low figures could be sufficient. This is a point that should be reassessed if the need arises for a case involving this sub-population. More markers can be added to increase the discrimination power of the test.

Asides this population, all but two likelihood ratios gave a figure greater than 1 in 1 million. The individual exhibiting a result less than this only had seven complete loci. Four databases were used to calculate match statistics for this sample and all but one gave a result above this figure. It is interesting that the SIB sub-population database gave 1 in 100000, but the total n249 database gave a figure ten-fold the value. This demonstrates that using a smaller sub-population database is preferable in providing a potentially more accurate statistic, as this profile appears to be much more common within this sub-population. Even with the theta value, the statistic calculated from the total population is favourable to the prosecution.

4.6.8 Suggested application of the dataset

With all of the above considerations in mind, the following approach to applying this dataset for forensic casework involving carpet pythons is suggested. A DNA profile obtained from an evidentiary item should be assignment tested against the six putative sub-populations using both likelihood of assignment and probability of exclusion approaches to show which of the populations the DNA profile is most genetically similar, and also attach a significance value to its exclusion as originating from any or all of the putative sub-populations. The outcome should dictate the application of a particular reference database (or multiple databases) to calculate statistical probabilities for profile matching. Where an individual is excluded from all populations, the overarching n249 population reference database should be applied. Where an individual is not excluded from multiple sub-populations, each sub-population database should be used to calculate the likelihood ratio and the most conservative of these figures reported.

From the population genetic analyses, the two Msi island populations showed no further indication of internal population sub-structuring. A theta correction is not necessary when calculating a match probability using these sub-population databases. A conservative approach involves using reduced locus sets to calculate match probabilities using the very small island population reference databases. A theta corrected statistic using the overarching population can also be presented to account for the hypothesis that a snake did not originate from the island population, however this is a less likely scenario if the unknown is genetically very similar to the island source populations. The SEC population showed a very small degree of sub-structuring present, so this theta correction should be applied to match probability calculations. The two island sub-populations adhered to Hardy-Weinberg expectations and the SEC population contained two large overlapping populations exhibiting HWE (figure A4.10). The deviation of the total SEC sub-population, while significant, was not highly significant like the other larger sub-populations.

4.6.9 A conservative approach is required, but just how conservative?

The three remaining sub-populations are more problematic as they do not exhibit HWE, but cannot be split into smaller sub-population partitions, due to the high geographic structure seen between few individuals spread over large distances. Some of these individuals likely represent what should be an entire population in their genetic differentiation, yet accurate estimates of F-statistics and adherence to HWE require at least two dozen individuals per population. The proposed solution when populations containing known substructure cannot be broken down into more accurate sub-populations, is to use the broader population dataset and apply the co-ancestry coefficient to the match probability calculation to adjust for a higher incidence of allele sharing within the population.

This is a logical and accepted way forward, however this dataset has revealed problems with this approach. The theta correction applied is based on the global F_{ST} calculated for the overall population and this is the average F_{ST} of the populations within. The larger group still has to be delimited into internal populations in order to calculate the global theta value. Each of the three subpopulations examined in this project, NEC, SIB and Msi ML, contained very small groups of clearly genetically divergent individuals, but also contained a remaining subset of individuals that exhibit a spectrum of genetic differentiation, while also spread over a large distance.

The NEC population will be discussed first, as it has the mildest case of this spectrum of differentiation, caused to some extent by IBD. When comparing Q-values from K3 and K4 (Appendix 4.7), most of the Norther Territory and New Guinean samples form two distinct populations and most other samples cluster strongly as a third population. At K4, this third population is split into two populations, but they both exhibit individuals admixed between, suggesting these individuals straddle the two populations, representing the middle of the genetic cline. One approach is to calculate a global theta, either using the populations delineated at K3 or K4. At K3 the global theta value is $\theta = 0.048$ (95% upper: 0.064, lower: 0.034). It is known that a degree of variability exists within the dataset and the allele frequencies in the overall dataset do not fit the sub-populations as well as local population estimates would (had sample numbers sufficed). One could argue that the average theta value disfavours those sub-populations within that are more highly inbred. An added degree of conservativism would involve application of the upper bound of the global theta, which is 0.064 for this sub-population database.

The other two populations represent a more difficult scenario as the cline of genetic differentiation is more continuous. The group of 36 Msi ML individuals for example exhibit a small degree of isolation by distance (Appendix 4.3.4). When forced into three populations (K3), the farthest west WWI individuals and the farthest east Yorke Peninsula individuals break out into clear subpopulations and everything else assigns to the same middle group (figure A4.18). However, when four populations are specified (K4), the middle group is dissected; however the degree of admixture suggests that this grouping contains a spectrum of genetic variability. Even individual 68310 from Yorke Peninsula shows admixture suggesting them to be a continuation of the Coastal SA group. This middle group is shown to contain continual genetic variation to the extent that the two extremes are very different from one another, but there is no clear point of division into two populations for the purpose of calculating the global F_{ST}. Appendix 4.3.4 provides additional information about the geographic locations and genetic differentiation shown in PCoA plots that represent ~26% of the variability. The SIB sub-population demonstrates similar patterns.

All three populations show this same pattern, containing a number of small or very small genetically distinct groups and a remaining scatter of widely spread and genetically continuous individuals, suggestive of IBD (see Appendix 4.3.3 and 4.3.5). There is no clear solution to calculating a global F_{ST} for these populations. A number of approaches are available where populations or part-populations exhibit these patterns, but each has drawbacks:

1. *Sample more individuals to give higher population numbers*: this is a good solution if time and resources permit, but for rare and persecuted species the remaining individuals might be remnants spread over large areas and increased sampling is not an option.

2. *When theta is unknown, calculate the Probability of Identity based on sibling relationships*: The equation for the Probability of Identity between siblings (PI_{SIBS}) provides the upper limit of the range of PIs obtained from a population [47]. This was the approach taken in chapter three and is a conservative measure for accounting for increased relatedness in individuals at either ends of the genetic spectrum. However, it is extreme in its conservativeness and therefore loses a large degree of information, drastically reducing the final match probability. An STR marker set of this resolution somewhat mitigates the problem, as the resulting PI_{SIBS} from the overall dataset is still one in one trillionth, a very discriminating value.

3. *Report the statistic for most frequent genotype in the population rather than the genotype seen*: This involves using the two most frequent alleles where the individual is a heterozygote at a locus and the frequency of the most common allele in the population where that locus is a homozygote in the profile of interest. Similarly to the PI_{SIBS}, this will give a much more conservative figure than the actual match probability for the profile, but loses a great deal of information if the evidentiary DNA profile contains rare alleles. These results for each sub-population were provided in table (allele genotype frequency estimates) and the most common genotype estimate was concluded above to be far larger than the realistic number of extant snakes in this species. Note that this estimate did not take into account the restricted loci suggested for this sub-population database. The difference between probabilities of 10^3 using the restricted dataset and 10^{10} from the full STR complement is substantial and whether or not one needs to restrict the dataset to a third of the available loci in these island populations warrants further thought and discussion.

4. *Apply inflated allele frequencies*: Where allele frequencies at a locus sum to 1, their values can be inflated by 10% to sum to 1.1. Where the allele frequencies are suspected not to accurately represent all individuals in the population, this increases the commonality of the alleles, shifting the calculated statistic slightly more in favour of the defence. This is not extreme like the previous two approaches and favours conservativism without losing a large amount of discrimination power, however the decision of the value to inflate the allele frequencies is arbitrary and thus difficult to defend in court.

5. *Apply conservative theta values:* Probability calculations presented in section 4.5.8 incorporated a comparison between application of the normal global theta value and the 95th percentile upper bound theta value. This approach applies a small degree of conservativism, but retains much of the information captured in the profile. The probabilities observed changed by only one to two orders of magnitude when the conservative theta value was applied. This approach could be argued as not conservative enough, given the various discussed limitations within the dataset. An extension of this could be to apply the uppermost theta value witnessed within the dataset, which would be the pairwise F_{ST} of 0.311 seen between the two Msi island populations. This figure represents the most divergence that any two populations within the dataset display, so should represent the maximum degree of genetic sub-structure present within the overall population.

Applying this value of theta changes the likelihoods previously reported in table 4.15 to those reported in table 4.16. This reduces some of the reportable likelihood ratios drastically, particularly in the NEC and SEC populations where theta values were otherwise low, and could be considered too extreme and unnecessary an approach for these particular populations. Compare the difference this has on the likelihoods reported for individual 71415. Using the four different databases with associated allele frequencies, the results range between 1 in 100 000 and 1 in 10 million. However, when applying the maximum theta correction, all results except that using the overarching population database reduce down to 1 in 10 000.

These all represent possible options and the most appropriate choice of action needs to be assessed on a case by case basis.

Table 4.16: The forensic statistics presented in table 4.15 and recalculated using upper 95th percentile theta value and the revised results compared with the previous figures.

Unknown sample	LR (Cons. Θ)	verbal equivalent	LR (Max. Θ)	verbal equivalent	Unknown sample	LR (Cons. Θ)	verbal equivalent	LR (Max. Θ)	verbal equivalent
55457	-	1 in 1 000	-	-	66334	**858695618.7**	1 in 100 million	**127074982.3**	1 in 100 million
66113	**2.29E+26**		**2.08855E+15**		12150	**1.54392E+13**		23584452.86	1 in 10 million
66171	-	1 in 100 000	-	-		**1.02409E+16**		10863975.06	1 in 10 million
13429	**7.4908E+25**		7.91448E+13		66337	**14237335.14**	1 in 10 million	**2901985.338**	1 in 1 million
68344	**4.82268E+30**		**3.18841E+15**		66317	**1.2857E+13**		**28109397.66**	1 in 10 million
66176	-	1 in 1 000	-	-		**1.16881E+12**		**31842859.37**	1 in 10 million
67148	**4.34516E+12**		**4.71952E+11**		68380	**14237036.23**	1 in 10 million	**513961.2078**	1 in 100 000
71370	**7.82624E+14**		**1.01243E+12**		12154	**34665714.75**	1 in 10 million	2006864.972	1 in 1 million
71409	**6.95467E+13**		**82870702691**	1 in 10 billion	81322	**81092593182**	1 in 10 billion	**473324.9165**	1 in 100 000
83875	**2.64503E+18**		**4.03995E+14**			**28152499567**	1 in 10 billion	**990101.3657**	1 in 100 000
13432	**5.76312E+22**		4.99085E+12		71415	**219510.3678**	1 in 100 000	**23030.53029**	1 in 10 000
13459	**3.2348E+22**		**7.39054E+11**			**68059914294**	1 in 10 billion	**54044.88049**	1 in 10 000
13457	**3.33928E+22**		**1.37421E+12**			**12924504.64**	1 in 10 million	**10817.95857**	1 in 10 000
	4.35943E+27		**7.9567E+12**			**2277997.055**	1 in 1 million	**191400.8169**	1 in 100 000
66031	**6.18482E+15**		**3.39225E+12**		13444	**826555.8278**	1 in 100 000	**2017.191266**	1 in 1 000
79112	**9.77717E+17**		**2.95682E+14**		13450	**2511281.961**	1 in 1 million	**13232.33604**	1 in 10 000
17537	**2.1877E+16**		**2.43227E+13**			**557039628**	1 in 100 million	**10517.97694**	1 in 10 000
50868	**1.33441E+13**		**8620018876**	1 in 1 billion	86801	**3741717.485**	1 in 1 million	**6413.77075**	1 in 1 000
81309	**3.51442E+24**		**5.33096E+11**			**44009748.24**	1 in 10 million	**2031.997127**	1 in 1 000

4.6.10 Additional traits of interest

The population genetic analysis was based on allele frequency data. The model based partitioning software STRUCTURE uses genotypes and subsequent allele frequency calculations to partition individuals into the specified number of populations with the least deviation from equilibria of genetic models (Hardy-Weinberg and Linkage). All analyses relied on the input of genotype data based on allele, and therefore fragment, size rather than sequencing data. This means that, unless a sequence characteristic such as an indel created a unique allelic micro-variant, this character would be not be considered during analysis. Chapter 3 demonstrated how much underlying genetic variability is lost through genotyping by size migration. Some characteristics were observed in this study that might provide additional supporting information to genotyping by size migration in cases dealing with certain source populations.

A 19 bp indel occurs at locus MS27 in the *Morelia spilota imbricata* subspecies. Its presence is not evident by allele size alone, so sequencing this indel can add information to investigations suspected to involving this sub-species. Likewise, a 2 bp was observed in MS13, which is present in *Morelia bredli*, absent in *M. s. imbricata* and polymorphic in all other carpet pythons. Absence of this indel could exclude an individual as *M. bredli*, while presence could exclude an individual as *M. s. imbricata*.

Northern Queensland and Northern Territory populations exhibited a deletion at locus MsF28 represented by a .4 micro-variant, however this feature should be evident by routine genotyping. This indel could be of particular interest in an allegation of species or sub-species hybridisation allegedly involving a carpet python originating from northern Australia.

The marker set is highly discriminatory and in most cases genotyping by size migration should be sufficient to discriminate individuals, without the need to sequence alleles and identify underlying sequence polymorphisms.

4.7 Concluding remarks

This dataset demonstrates some issues not frequently seen in human DNA analysis, but more common to forensic investigations dealing with wildlife DNA. Sample size is often an issue. Wide scale sampling of a single species is resource intensive and for STR markers a reference genotype dataset needs to be compiled specific to each species and also from the target population relevant to the investigative question. While DNA testing costs are decreasing, the funds required to screen large numbers of individuals are often well beyond the scant amount available with which to conduct an entire investigation. Many of the species targeted in wildlife crimes are sought after because of their rarity. It might be that there simply aren't the wild populations to be surveyed for a DNA database. There may be no way to mitigate the problems of sample size, so the only solution is to deal with it as best one can.

Sample sizes created difficulties when conducting the population genetic analyses on sub-populations in this study, particularly those of the Msi Mainland and the Southern and Internal Bredli sub-populations. Both of these datasets exhibited one internal 'subpopulation' that contained few individuals and spanned a large geographic range. Knowing how to assess these individuals is

difficult because essentially one individual represents a population and individual genotype variability can't be separated from population variability. These areas are known to be sparsely inhabited by carpet pythons and heavier sampling to fill out population numbers is unlikely achievable. These populations do not show conformance to HWE, yet to reduce the number of individuals included to find genetic evidence of an interbreeding population would reduce population numbers below a minimum amount reasonable for calculations of F-statistics. The island populations brought their own difficulties, exhibiting adherence to HWE, but sample sizes that were problematic for forensic statistics. The eastern coastal populations were more heavily sampled, but provided further complications, being a continuous population displaying a degree of isolation by distance. These are all very realistic situations that wildlife forensic scientists are likely to encounter and need to be able to tackle and draw conclusions from in a defendable manner.

Geographic assignment analysis was highly robust in four of the sub-populations and is informative even when partial profiles of less than ten loci are obtained. Unfortunately, homogeneity of the northern and eastern populations, as well as sparse sampling of the south-western populations and a lack of distinct genetic population partitioning within all but the island populations, has confounded the ability of this multiplex assay and associated reference database to determine the most likely geographic origin of a poached individual to a finer resolution than assigning them to one of the six putative sub-populations. Calculations of forensic statistics showed this dataset to be highly discriminative within all of the sub-populations, even with partial DNA profiles of less than half the total STR assay. In summary, this study provided many difficult, but realistic scenarios and a great opportunity to learn how best to treat each different situation to provide support to the justice system in criminal investigations involving carpet pythons.

4.8 References

[1] D.L. Hartl, A.G. Clark, Principles of Population Genetics, 3rd edition, Sinauer Associates, Inc., Sunderland, Massachusetts, 1997.

[2] Hartl, Primer of Population Genetics, 3rd edition, Sinauer Associates, Inc., 2000.

[3] S.G.W. Wahlund, Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet (Composition of populations from the perspective of the theory of heredity), Hereditas. 11 (1928) 65–105.

[4] A.C. Frantz, J.T. Pourtois, M. Heuertz, L. Schley, M.C. Flamand, A. Krier, S. Bertouille, F. Chaumont, T. Burke, Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population, Mol. Ecol. 15 (2006) 3191–3203.

[5] I.W. Evett, B.S. Weir, Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists, Sinauer Associates, Inc., 1998.

[6] S. Wright, Systems of Mating, I-V, Genetics. 6 (1921) 111–178.

[7] S. Wright, The genetical structure of populations, Ann. Eugen. 15 (1951) 323–354.

[8] K.E. Holsinger, B.S. Weir, Genetics in geographically structured populations: defining, estimating and interpreting FST, Nat. Rev. Genet. 10 (2009) 639–650.

[9] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140.

[10] F. Balloux, N. Lugon-Moulin, The estimation of population differentiation with microsatellite markers, Mol. Ecol. 11 (2002) 155–165.

[11] J. Goudet, N. Perrin, P.M. Waser, Tests for sex-biased dispersal using bi-parentally inherited genetic markers, Mol. Ecol. 11 (2002) 1103–1114.

[12] N. Dawnay, R. Ogden, R.S. Thorpe, L.C. Pope, D.A. Dawson, R. McEwing, A forensic STR profiling system for the Eurasian badger: A framework for developing profiling systems for wildlife species, Forensic Sci. Int. Genet. 2 (2008) 47–53.

[13] M. Slatkin, A Measure of Population Subdivision Based on Microsatellite Allele Frequencies, Genetics. 139 (1995) 457–462.

[14] M. Nei, Analysis of Gene Diversity in Subdivided Populations, Proc. Natl. Acad. Sci. U. S. A. 70 (1973) 3321–3323.

[15] J.M. Butler, Forensic DNA Typing: biology, technology and genetics of STR markers, Elsevier Academic Press, 2005.

[16] National Research Council, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington D.C., 1996.

[17] N. Lugon-Moulin, H. Brunner, A. Wyttenbach, J. Hausser, J. Goudet, Hierarchical analyses of genetic differentiation in a hybrid zone of *Sorex araneus* (Insectivora: Soricidae), Mol. Ecol. 8 (1999) 419–431.

[18] S.J. Goodman, Patterns of Extensive Genetic Differentiation and Variation Among European

Harbor Seals (*Phoca vitulina vitulina*) Revealed Using Microsatellite DNA Polymorphisms, Mol. Biol. Evol. 15 (1998) 104–118.

[19] S.L. Hazlitt, A.W. Goldizen, M.D.B. Eldridge, Significant patterns of population genetic structure and limited gene flow in a threatened macropodid marsupial despite continuous habitat in southeast Queensland, Australia, Conserv. Genet. 7 (2006) 675–689.

[20] S.H. Forbes, J.T. Hogg, F.C. Buchanan, A.M. Crawford, F.W. Allendorf, Microsatellite evolution in congeneric mammals: Domestic and bighorn sheep, Mol. Biol. Evol. 12 (1995) 1106–1113.

[21] A. Estoup, F. Rousset, Y. Michalakis, J.M. Cornuet, M. Adriamanga, R. Guyomard, Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*), Mol. Ecol. 7 (1998) 339–353.

[22] R. Ben-Shlomo, E. Reem, J. Douek, B. Rinkevich, Population genetics of the invasive ascidian *Botryllus schlosseri* from South American coasts, Mar. Ecol. Prog. Ser. 412 (2010) 85–92.

[23] T. Beebee, G. Rowe, An Introduction to Molecular Ecology, Oxford University Press, 2004.

[24] J.S. Buckleton, J.A. Bright, D. Taylor, Forensic DNA Evidence Interpretation, 2nd edition, CRC Press, Taylor & Francis Group, 2016.

[25] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics. 155 (2000) 945–959.

[26] D.A. Earl, B.M. vonHoldt, STRUCTURE HARVESTER : a website and program for visualizing STRUCTURE output and implementing the Evanno method, Conserv. Genet. Resour. 4 (2012) 359–361.

[27] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study, Mol. Ecol. 14 (2005) 2611–2620.

[28] NERC Biomolecular Analysis Facility, Analysis of population genetic structure, Sheffield, UK, n.d.

[29] M. Jakobsson, N.A. Rosenberg, CLUMPP : a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, Bioinformatics. 23 (2007) 1801–1806.

[30] R. Peakall, P.E. Smouse, GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research, Mol. Ecol. Notes. 6 (2006) 288–295.

[31] R. Peakall, P.E. Smouse, GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update, Bioinformatics. 28 (2012) 2537–2539.

[32] M. Raymond, F. Rousset, GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism, J. Hered. 86 (1995) 248--249.

[33] F. Rousset, GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux, Mol. Ecol. Resour. 8 (2008) 103–106.

[34] P.O. Lewis, D. Zaykin, Genetic Data Analysis: Computer program for the analysis of allelic data, (2001).

[35] S.D.E. Park, The Excel Microsatellite Toolkit, (2008). http://animalgenomics.ucd.ie/sdepark/ms-toolkit/.

[36] S. Piry, A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin, A. Estoup, GENECLASS2 : A software for genetic assignment and first-generation migrant detection, J. Hered. 95 (2004) 536–539.

[37] B. Rannala, Mountain, Detecting immigration by using multilocus genotypes, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 9197–9201.

[38] D. Paetkau, R. Slade, M. Burden, E.A. Direct, Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power, Mol. Ecol. 13 (2004) 55–65.

[39] D. Paetkau, W. Calvert, I. Stirling, C. Strobeck, Microsatellite analysis of population structure in canadian polar bears, Mol. Ecol. 4 (1995) 347–354.

[40] G.K. Chambers, E.S. MacAvoy, Microsatellites: consensus and controversy, Comp. Biochem. Physiol. Part B. 126 (2000) 455–476.

[41] Y. Zhu, D.C. Queller, J.E. Strassmann, A Phylogenetic Perspective on Sequence Evoltion in Microsatellite Loci, J. Mol. Evol. 50 (2000) 324–338.

[42] D. Pearson, R. Shine, R. How, Sex-Specific niche partitioning and sexual size dimorphism in Australian pythons (*Morelia spilota imbricata*), Biol. J. Linn. Soc. 77 (2002) 113–125.

[43] D. Taylor, Using DNA markers for wildlife management and protection: a study of the population structure and systematics of the Australian carpet pythons (Reptilia: *Morelia spilota* complex), Flinders University, SA, 2005.

[44] B. Efron, Estimating the error rate of a prediction rule: improvement on cross-validation, J. Am. Stat. Assoc. 78 (1983) 316–331.

[45] J.-M. Cornuet, S. Piry, G. Luikart, A. Estoup, M. Solignac, New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals, Genetics. 153 (1999) 1989–2000.

[46] J.M. Butler, Advanced topics in forensic DNA typing: interpretation, Academic Press, 2014.

[47] P. Taberlet, G. Luikart, Non-invasive genetic sampling and individual identification, Biol. Journ. 68 (1999) 41–55.

[48] J.K. Pritchard, Q. Wen, D. Falush, Documentation for structure software: Version 2.3, (2010).

[49] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies, Genetics. 164 (2003) 1567–1587.

[50] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M. V. Lareu, An overview of STRUCTURE: Applications, parameter settings and supporting software, Front. Genet. 4 (2013) 1–13.

[51] A.C. Frantz, J.T. Pourtois, M. Heuertz, L. Schley, M.C. Flamand, A. Krier, S. Bertouille, F. Chaumont, T. Burke, Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population, Mol. Ecol. 15 (2006) 3191–3203.

Appendix 4.1 Analysis technique background information

A4.1.1 Model-based clustering analysis using STRUCTURE

This software package partitions the multi-locus genotype dataset into the most probable population, each characterised by particular allele frequencies, applying the assumptions of Hardy-Weinberg equilibrium (or least violation thereof) and linkage equilibrium [25,48]. The user specifies the number of populations (K) and associated parameters. The application of a genetic model facilitates a result that is much more relevant to biological systems than either clusters informed by subjective (and therefore possibly biased) assessments such as phenotypic characters or geographic boundaries, or a clustering algorithm that is based on all axes of variability within the dataset alone without biological context (such as Principal Co-ordinate Analyses or Multidimensional Scaling methods). As the algorithm requires no *a priori* assumptions about groupings of individuals within the dataset (although this can be specified for certain applications), the results reflect only the genetic variability given within the input dataset, which could include 'cryptic' diversity [25].

The STRUCTURE analysis software uses a Bayesian clustering approach following Bayes theorem. Bayes theorem is described by:

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$

Posterior probabilities are modelled on prior information to give a posterior probability distribution for a hypothesis, testing many scenarios at once to give probabilities for each, rather than culminating in a single P-value [23].

A Markov Chain Monte Carlo (MCMC) simulation is applied to explore the sample space (probability distribution based on priors) of the dataset and find the optimum sampling solution (partitioning of individuals into populations and associated allele frequencies) to fit the model. A random partitioning of individuals is selected by the software during iteration one and the strength of the evidence for this inferred sample set, and associated allele frequencies, is assessed. The hypothesis that the sample set chosen in each iteration (populations and allele frequencies) fits the evidence (genotype dataset) based on the model better or worse than that chosen in the previous run is used to accept or reject the sample set. Put simply, the sample set is accepted if the solution is a better fit to the data and rejected if it is a worse fit. The move to the new sampling space in the next iteration is inversely proportionate to the fit of the data; acceptance will cause the next sample set to be taken close to that space with similar composition whereas rejection will lead to sampling from elsewhere. A large iteration number allows this repetitive sampling process to converge on a posterior sampling space (allele frequencies per population and population membership) that best fits the model specified. Through the iterative process, the final posterior probability can be considered with confidence as meaningful.

Basic decisions to be made when initiating a new STRUCTURE run are: length of burn-in, number of iterations in a run, specification of K values, replicate runs per K value, ancestry model and allele frequency model. There are also advanced options to fine-tune parameters such as α and λ within these categories.

The burn-in period constitutes the initial sampling stage where the sample sets chosen do not fit the model well, so the sampling jumps around the sample space to find a region that fits better. Once better fitting sample sets are found, adjustments to the sample sets are made at a much finer scale. The burn-in period is therefore discarded from contributing to the final probability distribution from which sampling is conducted to draw inferences. Sufficient replicate runs should be conducted for each value of K to ensure that the results of different runs converge on the same value.

The number of iterations in a run is particular to the dataset and is adjusted based on the results of initial testing. A higher iteration number will provide more accurate results at the expense of computer processing power and time. The log likelihood of runs with the same K values should be assessed for convergence of results. A substantial degree of variability in the log likelihood between runs of the same K when compared to the variability in log likelihoods between runs of different K values indicates that the iteration number should be increased [48].

The ancestry model is user specified. Individuals in the dataset can be assumed to be derived from a single population (no admixture model) or contain alleles originating from multiple populations (admixture model). The latter assumption allows for proportions of an individual to be allocated to multiple populations. This scenario is more realistic to biological populations [48].

Two alternative allele frequency models are applied to a dataset: the Independent Allele Frequency Model (IAFM) and the Correlated Allele Frequency Model (CAFM). The IAFM specifies that populations are independent of one another and therefore exhibit quite different allele frequencies. The CAFM specifies that frequencies between populations are likely to be quite similar, due perhaps to migration or shared ancestry [49]. If no admixture is present, CAFM results will be very similar to IAFM results [50]. The CAFM can lead to an overestimation of populations [25]. Individuals that have very similar allele frequencies can be treated as originating from two different, but very similar populations, whereas they would be treated as members of one single population if IAFM were selected. For this reason, it is important particularly when using CAFM to choose the smallest value of K that maximises the likelihood without overestimating it (point of plateau, see figure A4.1 for a comparison these methods).

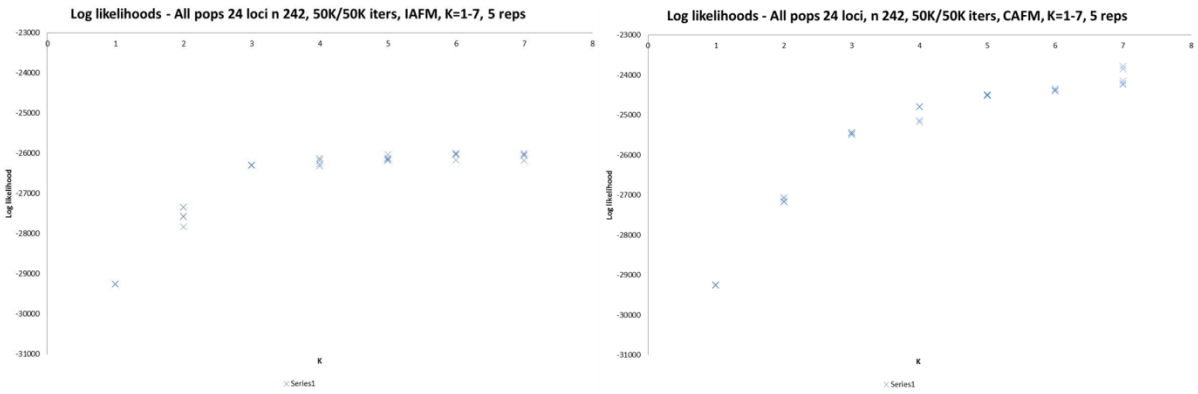


Figure A4.1: Results of identical analyses of a dataset using both IAFM and CAFM. IAFM shows a much more defined plateau, while CAFM log likelihoods continue to decrease at higher putative population numbers. This can lead to an overestimation of populations, so the largest change in log likelihood values between population numbers, indicated by the Delta K, should be used to inform likely population number.

Following Porras-Hurtado and colleagues [50], the alpha plot should be checked following a run to ensure that the alpha value has reached stability. This indicates that the burn-in length specified is sufficient to remove initial stochasticity from the probability distribution used in the final assessment. If the alpha value still shows fluctuations, then the burn-in stage should be increased. A stabilised alpha approaching 1 indicates that much admixture is present in the populations, whereas alpha approaching 0 is representative of independent populations.

Results given in the output file for each run of each K include: the estimated log likelihood of the data Pr(X|K) including mean and variance, mean α value, proportion of membership to the inferred clusters, allele frequency divergence between clusters, H_e between individuals in the same cluster, F_{ST} values, estimated allele frequencies in each cluster by locus and inferred ancestry of each individual. The latter values are given as proportions to each cluster (Q-values) when admixture is allowed and can be averaged across runs to obtain a much more accurate measure of inferred ancestry for each individual.

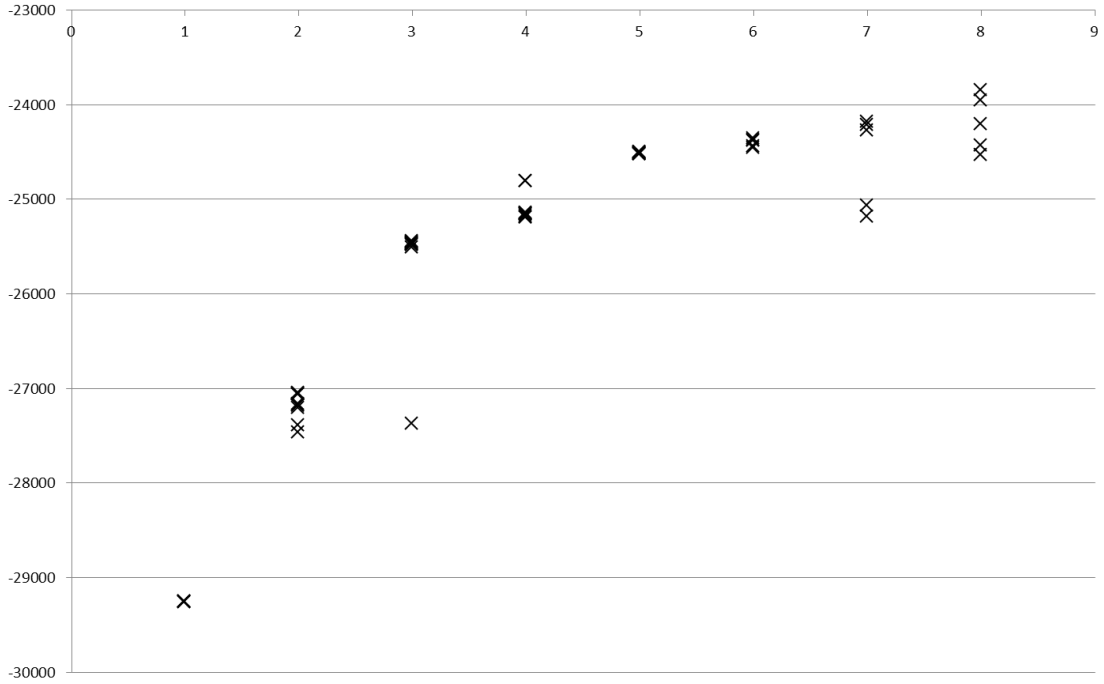


Figure A4.2: Example log likelihood plot showing log likelihood values represented by crosses for replicate runs at each K value from one to eight putative populations. Population numbers of 1, 5 and 6 show the most convergence between replicate runs, indicated by the close proximity of the results of replicates. K3 shows an outlying run that should be removed from any subsequent analyses. The plateau of this chart is around K3 and replicates are diverging after K6.

The most likely K value within a dataset is suggested to be the smallest K value for which log likelihoods cease to increase rapidly, evident as a plateau in the graph [27]. Convergence of log likelihood values of runs within a K value indicates the iteration number to be sufficient. Any outlier log likelihoods suggest the results of a particular run to be spurious and this run is excluded from subsequent averages of Q-values.

Further inference about the most appropriate K value can be made from the Delta K plot provided by STRUCTURE Harvester. This plot shows the second rate of change in log probability between

successive K values to identify the correct number of clusters in the data. The maximal value of Ln K is often indicative of the true number of populations in the dataset. However, Evanno et al. 2005 [27] note the drawback that the Delta K value only addresses the uppermost hierarchical level of genetic partitioning within the dataset, so the dataset should be sequentially divided into the resulting partitions and analysed again.

A bar plot visualisation of the data (see figure A4.5 below for examples) is a further component of the STRUCTURE output. This should be compared with the other data for concordance of the most appropriate K value. Individuals are represented by vertical lines. Colours represent populations, so an admixed individual has a multi-coloured bar showing the proportions of their genetic makeup corresponding to the respective populations. A K2 plot for a homogeneous interbreeding population would show every individual split between the two populations in equal proportions (see figure A4.9b for examples) because the software is trying to introduce genetic structure where none exists.

A more accurate representation of the component populations can be gained by averaging the bar plots over the replicates of a K value. Population colours are not standardised across replicates, so identifying replicate populations between bar charts can be challenging. The software package CLUMPP [29] is available to do this, however my attempts to run this software package were unsuccessful. I was able to create the same results by plotting the Q-matrix for each run belonging to a K value in Excel and then taking the average across runs. Each column of the Q-matrix represents a population. Column 1 was sorted for descending value. Individuals for which the majority of their genetic makeup was attributed to this population were allocated a colour. All non-coloured individuals were there sorted by descending value for column 2 and this process was repeated until the individuals had been sorted and coloured corresponding to the population to which most of their genetic information was attributed. This was done for every run of the same K value. By looking at the identity of individuals within each population, populations could be identified across runs. Columns were rearranged so that populations appeared in the same order across runs and the Excel average function was used to create a final population set showing average Q-values for each individual across all runs. Any run identified by the log likelihood plots as spurious was removed from this analysis because the lack of convergence would yield erroneous Q-values that would skew the averages.

A4.1.2 Multi-variate analysis (PCoA) using GenAlEx

Principal Co-ordinate Analysis, also known as Multidimensional Scaling is another way of visualising variation and therefore population clusters within the dataset, although it does not rely on a stipulated biological model as STRUCTURE does. PCoA allows compression of a complex multivariate dataset into a much simpler representation of the variability within the dataset based on the dissimilarity between points, demonstrating the major patterns of variability in a simplified manner.

A4.2 Detailed model based analysis of the overarching population

A4.2.1 Most likely population number

Log likelihood values and degree of run convergence at each K value appear equivalent using initial (100/300K) and extended (200K/1M) iteration analysis conditions (figure A4.3). Lack of run convergence at a K value can also indicate insufficient iteration number, so extended iterations were

applied at K6-9. A small log likelihood value is preferable. Yet, as higher population numbers are stipulated, the dataset is increasingly split into smaller groups beyond the realistic number of subpopulations present. The most likely K value within a dataset is therefore suggested to be the smallest K value for which log likelihoods cease to increase rapidly, evident as a plateau in the graph [27]. The log likelihood graph exhibits plateaus at three and five subpopulations. Convergence of log likelihood values of runs within a K value is also indicative of a likely K value once insufficient iteration length has been ruled out. Both the three and five population simulations exhibit a high degree of run convergence.

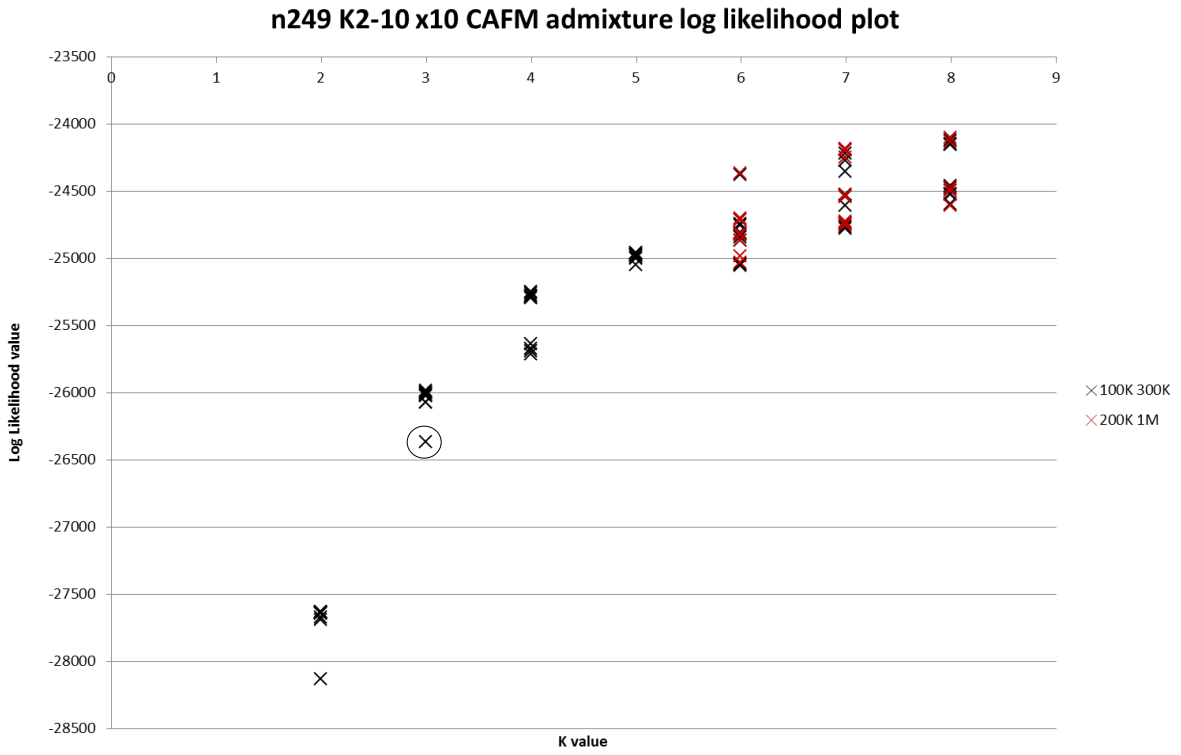
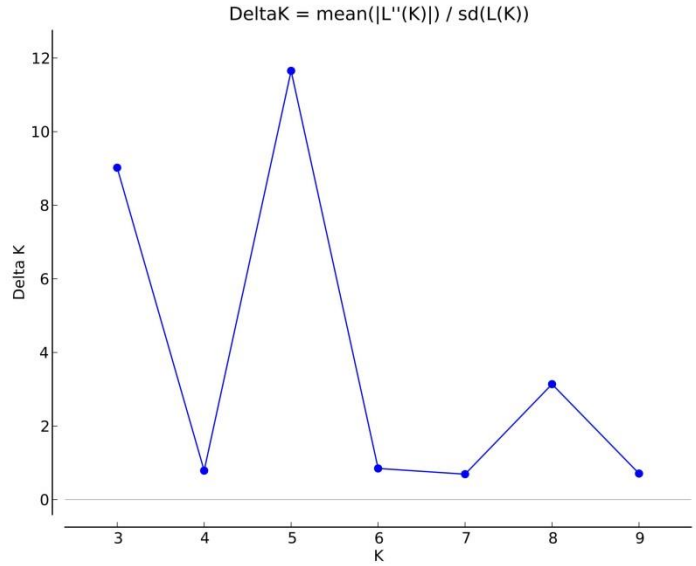


Figure A4.3: Log likelihood values for the 249 individual dataset using 10 replicates for each of two populations through ten populations scenarios. Replicates lacked convergence after five putative populations and comparison is made using normal and extended burn-in and run conditions for the larger population scenarios. Maximal run convergence is evident at K5. An outlier run at K3 is circled.

The Delta K analysis provides the second rate of change in log probability between successive K values to identify the correct number of clusters in the data. The Delta K plot for this data (figure A4.4) agrees with the log likelihood plot, showing three and five subpopulations to have the highest rate of change of all putative subpopulation compositions. The maximal value of Ln K is often indicative of the true number of subpopulations in the dataset [27]. While the Delta K value is noted only to address the uppermost hierarchical level of genetic partitioning within the dataset [27],

Figure A4.4: Second rate of change is shown between K values for the 249 individual dataset. The highest values are suggestive the most likely number of populations present in the dataset, in this case three populations or five populations.



my data appear to suggest the five subpopulation model to present a higher rate of change than the three subpopulation model. The CAFM analysis also comes with a caveat that the number of populations can often be overestimated, so the lowest population number for which a plateau and convergence are seen is likely to represent the realistic number of populations.

Any outlying run with a log likelihood value divergent from all other converged run likelihood values represents a spurious result (this could be caused by the analysis beginning in a less ideal sampling space – hence the necessity for replicate runs). An outlier run is evident in the K3 results (circled run in Figure A4.3), run 14 exhibiting a log likelihood of -26364 when the other nine runs converged around the value of -26018. This run was excluded from subsequent examination of average Q-values.

Figure A4.5 provides example bar plots for K values five through eight. Individuals are represented by vertical lines. Colours represent populations, so an admixed individual has a multi-coloured bar showing the proportions of their genetic makeup corresponding to the respective populations. Ideally, the highest number of populations with the least degree of admixture between populations is desirable to indicate genetically structured, independent populations within the dataset - although this is not always the reality of the populations and during their assessment of population genetic structure in wild boar Frantz et al. [51] highlight the inability of the STRUCTURE software package to deal with populations exhibiting isolation by distance. Interpretation of the bar charts obtained from the analysis of the n249 dataset is not clear cut and quite subjective, so examination of average Q-values was conducted to give a clearer representation of the population partitions reproduced consistently across the replicate runs.

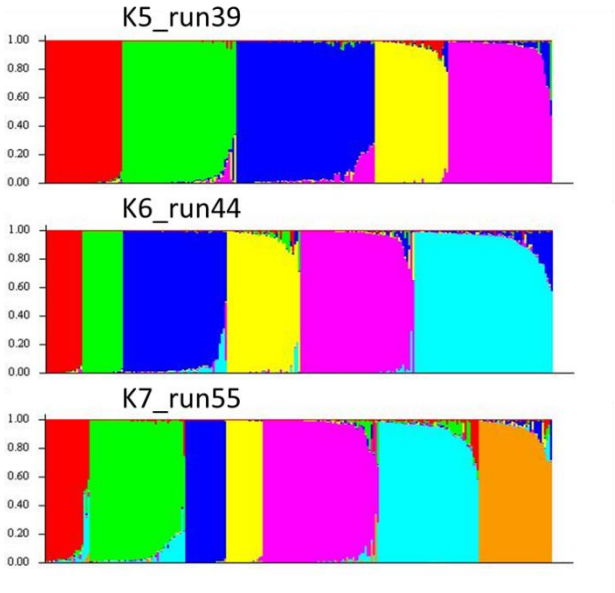
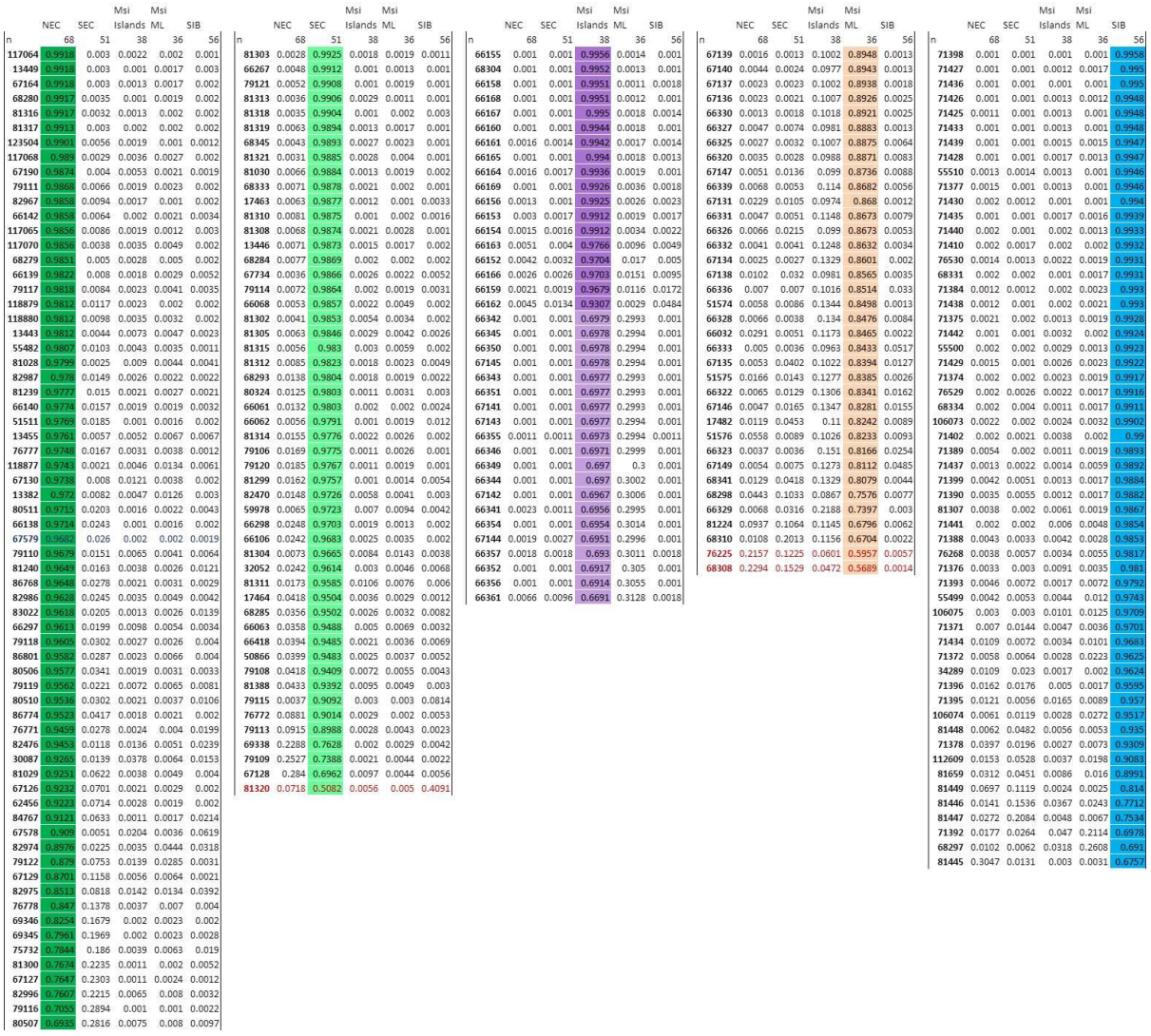


Figure A4.5: n249 bar charts: Examples box plots using the 249 individuals dataset with three population amounts specified. Individuals are represented by thin vertical lines. Different genetic populations are represented by colour. Multiple colours make up proportions of an admixed individual's vertical bar. This dataset shows quite genetically distinct populations with very little admixture, represented by the fairly clean colour groupings on the vertical axis.

In the K5 analysis, run 32 displayed a slightly increased log likelihood value compared to the other more highly convergent runs. Population compositions based on average Q-values were examined both including and excluding this population. No difference was seen in the citizenship of the five populations. The K5 average Q-values exhibit moderately clean partitioning of individuals into four of the populations; the fifth population shows half of the membership to attribute <0.7 of their genetic makeup to this population (figure A4.6). On closer examination, this population contains *Morelia spilota imbricata* subspecies representatives sampled from the two island populations (Garden Island and Saint Francis Island) and the split in genetic allocation corresponds with the memberships from the respective islands. This result indicates that the two islands are quite different and genetically divergent of one another. The dataset demonstrates a high degree of genetic structuring, indicated by relatively little admixture of each individual between multiple population clusters; only three individuals demonstrate less than two-thirds of their genetic information assigned to a single cluster.

The three individuals that assigned less than 0.65 in the K5 average Q-matrix are shown in figure A4.7. Individual 81320 assigned with highest uncertainty, showing 0.508 assignment to the South Eastern Coastal population and 0.409 assignment to the Southern, Internal and Bredli population. This individual continued to assign consistently weakly (<0.53) in all subsequent K6-8 runs examined, with admixture evident from SEC and SIB populations. These two populations are largely separated by a geographic barrier, the Great Dividing Range mountain range that runs up the South Eastern Coast of Australia. This range is likely to have created the barrier to geneflow that has resulted in the SIB populations showing slightly more differentiation with the SEC population (FST 0.052) than the NEC population (FST 0.062). However, this admixed individual comes from near Gunnedah in North Western NSW and could potentially represent a hybrid between the



	NEC	SEC	Msi Islands	Msi ML	SIB
n	68	51	38	36	56
117064	0.9918	0.003	0.0022	0.002	0.001
13449	0.9918	0.003	0.001	0.0017	0.003
67164	0.9918	0.003	0.0013	0.0017	0.002
68280	0.9917	0.0035	0.001	0.0019	0.002
81316	0.9917	0.0032	0.0013	0.002	0.002
81317	0.9913	0.003	0.002	0.002	0.002
123504	0.9901	0.0056	0.0019	0.001	0.0012
117068	0.989	0.0029	0.0036	0.0027	0.002
67190	0.9874	0.004	0.0053	0.0021	0.0019
79111	0.9868	0.0066	0.0019	0.0023	0.002
82967	0.9858	0.0094	0.0017	0.001	0.002
66142	0.9858	0.0064	0.002	0.0021	0.0034
117065	0.9856	0.0086	0.0019	0.0012	0.003
117070	0.9856	0.0038	0.0035	0.0049	0.002
68279	0.9851	0.005	0.0028	0.005	0.002
66139	0.9822	0.008	0.0018	0.0029	0.0052
79117	0.9818	0.0084	0.0023	0.0041	0.0035
118879	0.9812	0.0117	0.0023	0.002	0.002
118880	0.9812	0.0098	0.0035	0.0032	0.002
13443	0.9812	0.0044	0.0073	0.0047	0.0023
55482	0.9807	0.0103	0.0043	0.0035	0.0011
81028	0.9799	0.0025	0.009	0.0044	0.0041
82987	0.978	0.0149	0.0026	0.0022	0.0022
81239	0.9777	0.015	0.0021	0.0027	0.0021
66140	0.9774	0.0157	0.0019	0.0019	0.0032
51511	0.9769	0.0185	0.001	0.0016	0.002
13455	0.9761	0.0057	0.0052	0.0067	0.0067
76777	0.9748	0.0167	0.0031	0.0038	0.0012
118877	0.9743	0.0021	0.0046	0.0134	0.0061
67130	0.9738	0.008	0.0121	0.0038	0.002
13382	0.972	0.0082	0.0047	0.0126	0.003
80511	0.9715	0.0203	0.0016	0.0022	0.0043
66138	0.9714	0.0243	0.001	0.0016	0.002
67579	0.9682	0.026	0.002	0.002	0.0019
79110	0.9679	0.0151	0.0065	0.0041	0.0064
81240	0.9649	0.0163	0.0038	0.0026	0.0121
86768	0.9648	0.0278	0.0021	0.0031	0.0029
82986	0.9628	0.0245	0.0035	0.0049	0.0042
83022	0.9618	0.0205	0.0013	0.0026	0.0139
66297	0.9613	0.0199	0.0098	0.0054	0.0034
79118	0.9605	0.0302	0.0027	0.0026	0.004
86801	0.9582	0.0287	0.0023	0.0066	0.004
80506	0.9577	0.0341	0.0019	0.0031	0.0033
79119	0.9562	0.0221	0.0072	0.0065	0.0081
80510	0.9536	0.0302	0.0021	0.0037	0.0106
86774	0.9523	0.0417	0.0018	0.0021	0.002
76771	0.9459	0.0278	0.0024	0.004	0.0199
82476	0.9453	0.0118	0.0136	0.0051	0.0239
30087	0.9265	0.0139	0.0378	0.0064	0.0153
81029	0.9251	0.0622	0.0038	0.0049	0.004
67126	0.9232	0.0701	0.0021	0.0029	0.002
62456	0.9223	0.0714	0.0028	0.0019	0.002
84767	0.9121	0.0633	0.0011	0.0017	0.0214
67578	0.909	0.0051	0.0204	0.0036	0.0619
82974	0.8976	0.0225	0.0035	0.0444	0.0318
79122	0.879	0.0753	0.0139	0.0285	0.0031
67129	0.8701	0.1158	0.0056	0.0064	0.0021
82975	0.8513	0.0818	0.0142	0.0134	0.0392
76778	0.847	0.1378	0.0037	0.007	0.004
69346	0.8254	0.1679	0.002	0.0023	0.002
69345	0.7961	0.1969	0.002	0.0023	0.0028
75732	0.7844	0.186	0.0039	0.0063	0.019
81300	0.7674	0.2235	0.0011	0.002	0.0052
67127	0.7647	0.2303	0.0011	0.0024	0.0012
82996	0.7607	0.2215	0.0065	0.008	0.0032
79116	0.7055	0.2894	0.001	0.001	0.0022
80507	0.6935	0.2816	0.0075	0.008	0.0097

	NEC	SEC	Msi Islands	Msi ML	SIB
n	68	51	38	36	56
81303	0.0028	0.9925	0.0018	0.0019	0.0011
66267	0.0048	0.9912	0.001	0.0013	0.001
79121	0.0052	0.9908	0.001	0.0019	0.001
81313	0.0036	0.9906	0.0029	0.0011	0.001
81318	0.0035	0.9904	0.001	0.002	0.003
81319	0.0063	0.9894	0.0013	0.0017	0.001
68345	0.0043	0.9893	0.0027	0.0023	0.001
81321	0.0031	0.9885	0.0028	0.004	0.001
81030	0.0066	0.9884	0.0013	0.0019	0.002
68333	0.0071	0.9878	0.0021	0.002	0.001
17463	0.0063	0.9877	0.0012	0.001	0.0033
81310	0.0081	0.9875	0.001	0.002	0.0016
81308	0.0068	0.9874	0.0021	0.0028	0.001
13446	0.0071	0.9873	0.0015	0.0017	0.002
68284	0.0077	0.9869	0.002	0.002	0.002
67734	0.0036	0.9866	0.0026	0.0022	0.0052
79114	0.0072	0.9864	0.002	0.0019	0.0031
66068	0.0053	0.9857	0.0022	0.0049	0.002
81302	0.0041	0.9853	0.0054	0.0034	0.002
81305	0.0063	0.9846	0.0029	0.0042	0.0026
81315	0.0056	0.983	0.003	0.0059	0.002
81312	0.0085	0.9823	0.0018	0.0023	0.0049
68293	0.0138	0.9804	0.0018	0.0019	0.0022
80324	0.0125	0.9803	0.0011	0.0031	0.003
66061	0.0132	0.9803	0.002	0.002	0.0024
66062	0.0056	0.9791	0.001	0.0019	0.012
81314	0.0155	0.9776	0.0022	0.0026	0.002
79106	0.0169	0.9775	0.0011	0.0026	0.001
79120	0.0185	0.9767	0.0011	0.0019	0.001
81299	0.0162	0.9757	0.001	0.0014	0.0054
82470	0.0148	0.9726	0.0058	0.0041	0.003
59978	0.0065	0.9723	0.007	0.0094	0.0042
66298	0.0248	0.9703	0.0019	0.0013	0.002
66106	0.0242	0.9683	0.0025	0.0035	0.002
81304	0.0073	0.9665	0.0084	0.0143	0.0038
32052	0.0242	0.9614	0.003	0.0046	0.0068
81311	0.0173	0.9585	0.0106	0.0076	0.006
17464	0.0418	0.9504	0.0036	0.0029	0.0012
68285	0.0356	0.9502	0.0026	0.0032	0.0082
66063	0.0358	0.9488	0.005	0.0069	0.0032
66418	0.0394	0.9485	0.0021	0.0036	0.0069
50866	0.0399	0.9483	0.0025	0.0037	0.0052
79108	0.0418	0.9409	0.0072	0.0055	0.0043
81388	0.0433	0.9392	0.0095	0.0049	0.003
79115	0.0037	0.9092	0.003	0.003	0.0814
76772	0.0881	0.9014	0.0029	0.002	0.0053
79113	0.0915	0.8988	0.0028	0.0043	0.0023
69338	0.2288	0.7628	0.002	0.0029	0.0042
79109	0.2527	0.7388	0.0021	0.0044	0.0022
67128	0.284	0.6962	0.0097	0.0044	0.0056
81320	0.0718	0.5082	0.0056	0.005	0.4091

	NEC	SEC	Msi Islands	Msi ML	SIB
n	68	51	38	36	56
66155	0.001	0.001	0.9956	0.0014	0.001
68304	0.001	0.001	0.9952	0.0013	0.001
66158	0.001	0.001	0.9951	0.0011	0.0018
66168	0.001	0.001	0.9951	0.0012	0.001
66167	0.001	0.001	0.995	0.0018	0.0014
66160	0.001	0.001	0.9944	0.0018	0.001
66161	0.0016	0.0014	0.9942	0.0017	0.0014
66165	0.001	0.001	0.994	0.0018	0.0013
66164	0.0016	0.0017	0.9936	0.0019	0.001
66169	0.001	0.001	0.9926	0.0036	0.0018
66156	0.0013	0.001	0.9925	0.0026	0.0023
66153	0.003	0.0017	0.9912	0.0019	0.0017
66154	0.0015	0.0016	0.9912	0.0034	0.0022
66163	0.0051	0.004	0.9766	0.0096	0.0049
66152	0.0042	0.0032	0.9704	0.017	0.005
66166	0.0026	0.0026	0.9703	0.0151	0.0095
66159	0.0021	0.0019	0.9675	0.0116	0.0172
66162	0.0045	0.0134	0.9307	0.0029	0.0484
66342	0.001	0.001	0.6979	0.2993	0.001
66345	0.001	0.001	0.6978	0.2994	0.001
66350	0.001	0.001	0.6978	0.2994	0.001
67145	0.001	0.001	0.6978	0.2994	0.001
66343	0.001	0.001	0.6977	0.2993	0.001
66351	0.001	0.001	0.6977	0.2993	0.001
67141	0.001	0.001	0.6977	0.2993	0.001
67143	0.001	0.001	0.6977	0.2994	0.001
66355	0.0011	0.0011	0.6973	0.2994	0.0011
66346	0.001	0.001	0.6971	0.2999	0.001
66349	0.001	0.001	0.697	0.3	0.001
66344	0.001	0.001	0.697	0.3002	0.001
67142	0.001	0.001	0.6967	0.3006	0.001
66341	0.0023	0.0011	0.6956	0.2995	0.001
66354	0.001	0.001	0.6954	0.3014	0.001
67144	0.0019	0.0027	0.6951	0.2996	0.001
66357	0.0018	0.0018	0.693	0.3011	0.0018
66352	0.001	0.001	0.6917	0.305	0.001
66356	0.001	0.001	0.6914	0.3055	0.001
66361	0.0066	0.0096	0.6691	0.3128	0.0018

	NEC	SEC	Msi Islands	Msi ML	SIB
n	68	51	38	36	56
67139	0.0016	0.0013	0.1002	0.8948	0.0013
67140	0.0044	0.0024	0.0977	0.8943	0.0013
67137	0.0023	0.0023	0.1002	0.8938	0.0018
67136	0.0023	0.0021	0.1007	0.8926	0.0025
66330	0.0013	0.0018	0.1018	0.8921	0.0025
66327	0.0047	0.0074	0.0981	0.8883	0.0013
66325	0.0027	0.0032	0.1007	0.8875	0.0064
66320	0.0035	0.0028	0.0988	0.8871	0.0083
67147	0.0051	0.0136	0.099	0.8736	0.0088
66339	0.0068	0.0053	0.114	0.8682	0.0056
67131	0.0229	0.0105	0.0974	0.868	0.0012
66331	0.0047	0.0051	0.1148	0.8673	0.0079
66326	0.0066	0.0215	0.099	0.8673	0.0053
66332	0.0041	0.0041	0.1248	0.8632	0.0034
67134	0.0025	0.0027	0.1329	0.8601	0.002
67138	0.0102	0.032	0.0981	0.8565	0.0035
66336	0.007	0.007	0.1016	0.8514	0.033
51574	0.0058	0.0086	0.1344	0.8498	0.0013
66328	0.0066	0.0038	0.134	0.8476	0.0084
66032	0.0291	0.0051	0.1173	0.8465	0.0022
66333	0.005	0.0036	0.0963	0.8433	0.0517
67135	0.0053	0.0402	0.1022	0.8394	0.0127
51575	0.0166	0.0143	0.1277	0.8385	0.0026
66322	0.0065	0.0129	0.1306	0.8341	0.0162
67146	0.0047	0.0165	0.1347	0.8281	0.0155
17482	0.0119	0.0453	0.11	0.8242	0.0089
51576	0.0558	0.0089	0.1026	0.8233	0.0093
66323	0.0037	0.0036	0.151	0.8166	0.0254
67149	0.0054	0.0075	0.1273	0.8112	0.0485
68341	0.0129	0.0418	0.1329	0.8079	0.0044
68298	0.0443	0.1033	0.0867	0.7576	0.0077
66329	0.0068	0.0316	0.2188	0.7397	0.003
81224	0.0937	0.1064	0.1145	0.6796	0.0062
68310	0.0108	0.2013	0.1156	0.6704	0.0022
76225	0.2157	0.1225	0.0601	0.5957	0.0057
68308	0.2294	0.1529	0.0472	0.5689	0.0014

	NEC	SEC	Msi Islands	Msi ML	SIB
n	68	51	38	36	56
71398	0.001	0.001	0.001	0.001	0.9958
71427	0.001	0.001	0.0012	0.0017	0.995
71436	0.001	0.001	0.001	0.001	0.995
71426	0.001	0.001	0.0013	0.0012	0.9948
71425	0.0011	0.001	0.0013	0.001	0.9948
71433	0.001	0.001	0.0013	0.001	0.9948
71439	0.001	0.001	0.0015	0.0015	0.9947
71428	0.001	0.001	0.0017	0.0013	0.9947
55510	0.0013	0.0014	0.0013	0.001	0.9946
71377	0.0015	0.001	0.0013	0.001	0.9946
71430	0.002	0.0012	0.001	0.001	0.994
71435	0.001	0.001	0.0017	0.0016	0.9939
71440	0.002	0.001	0.002	0.0013	0.9933
71410	0.002	0.0017	0.002	0.002	0.9932
76530	0.0014	0.0013	0.0022	0.0019	0.9931
68331	0.002	0.002	0.001	0.0017	0.9931
71384	0.0012	0.0012	0.002	0.0023	0.993
71438	0.0012	0.001	0.002	0.0021	0.993
71375	0.0021	0.002	0.0013	0.0019	0.9928
71442	0.001	0.001	0.0032	0.002	0.9924
55500	0.002	0.002	0.0029	0.0013	0.9923
71429	0.0015	0.001	0.0026	0.0023	0.9922
71374	0.002	0.002	0.0023	0.0019	0.9917
76529	0.002	0.0026	0.0022	0.0017	0.9916
68334	0.002	0.004	0.0011	0.0017	0.9911
106073	0.0022	0.002	0.0024	0.0032	0.9902
71402	0.002	0.0021	0.0038	0.002	0.99
71389	0.0054	0.002	0.0011	0.0019	0.9893
71437	0.0013	0.0022	0.0014	0.0059	0.9892
71399	0.0042	0.0051	0.0013	0.0017	0.9884
71390	0.0035	0.0055	0.0012	0.0017	0.9882
81307	0.0038	0.002	0.0061	0.0019	0.9867
71441	0.002	0.002	0.006	0.0048	0.9854
71388	0.0043	0.0033	0.0042	0.0028	0.9853
76268	0.0038	0.0057	0.0034	0.0055	0.9817
71376	0.0033	0.003	0.0091	0.0035	0.981
71393	0.0046	0.0072	0.0017	0.0072	0.9792
55499	0.0042	0.0053	0.0044	0.012	0.9743
106075	0.003	0.003	0.0101	0.0125	0.9709
71371	0.007	0.0144	0.0047	0.0036	0.9701
71434	0.0109	0.0072	0.0034	0.0101	0.9683
71372	0.0058	0.0064	0.0028	0.0223	0.9625
34289	0.0109	0.023	0.0017	0.002	0.9624
71396	0.0162	0.0176	0.005	0.0017	0.9595
71395	0.0121	0.0056	0.0165	0.0089	0.957
106074	0.0061	0.0119	0.0028	0.0272	0.9517
81448	0.0062	0.0482	0.0056	0.0053	0.935
71378	0.0397	0.0196	0.0027	0.0073	0.9309
112609	0.0153	0.0528	0.0037	0.0198	0.9083
81659	0.0312	0.0451	0.0086	0.016	0.8991
81449	0.0697	0.1119	0.0024	0.0025	0.814
81446	0.0141	0.1536	0.0367	0.0243	0.7712
81447	0.0272	0.2084	0.0048	0.0067	0.7534
71392	0.0177	0.0264	0.047	0.2114	0.6978
68297	0.0102	0.0062	0.0318	0.2608	0.691
81445	0.3047	0.0131	0.003	0.0031	0.6757

Figure A4.6 n249 K5 aveQs: Average Q values of the n249 dataset tested for five populations. Four populations show clean assignment, whereas the fifth shows lower assignment of roughly half of the individuals, indicated by the two shades of lavender. Two admixed individuals highlighted red.

two sub-populations, an uncommon case of geneflow across a fairly inaccessible geographic barrier. This individual was excluded from subsequent subpopulation analyses.

The other two individuals that show admixture between the sub-populations are 76225 from Dog Fence Beach on Coastal SA and 68308 from Mondrain Island, off the south coast of WA. These individuals show roughly half assignment (0.596 and 0.569 respectively) to the Msi ML population, but show lesser proportions of assignment to NEC (0.216 and 0.23 respectively) and SEC (0.123 and 0.153 respectively) populations. -

The Q-values for K6 and above showed multimodality; citizenship of individuals within the populations was not consistent across runs. Average Q-values could not be calculated. This suggests that these K values are less likely to be the realistic representation of subpopulations. This result is supported by the lack of convergence in log likelihood values between the 10 replicates each of K6 and K7. Putative subpopulation values of K6-8 were reanalysed using extended parameter settings (200000 burn-in, 1000000 iterations, 10 replicates), but the log likelihoods of the ten replicates

failed to converge (figure A4.3). These STRUCTURE results suggest that the overall dataset is made up of five putative populations, but a further split of the two *Morelia spilota imbricata* island populations is justified by their genetic heterogeneity.

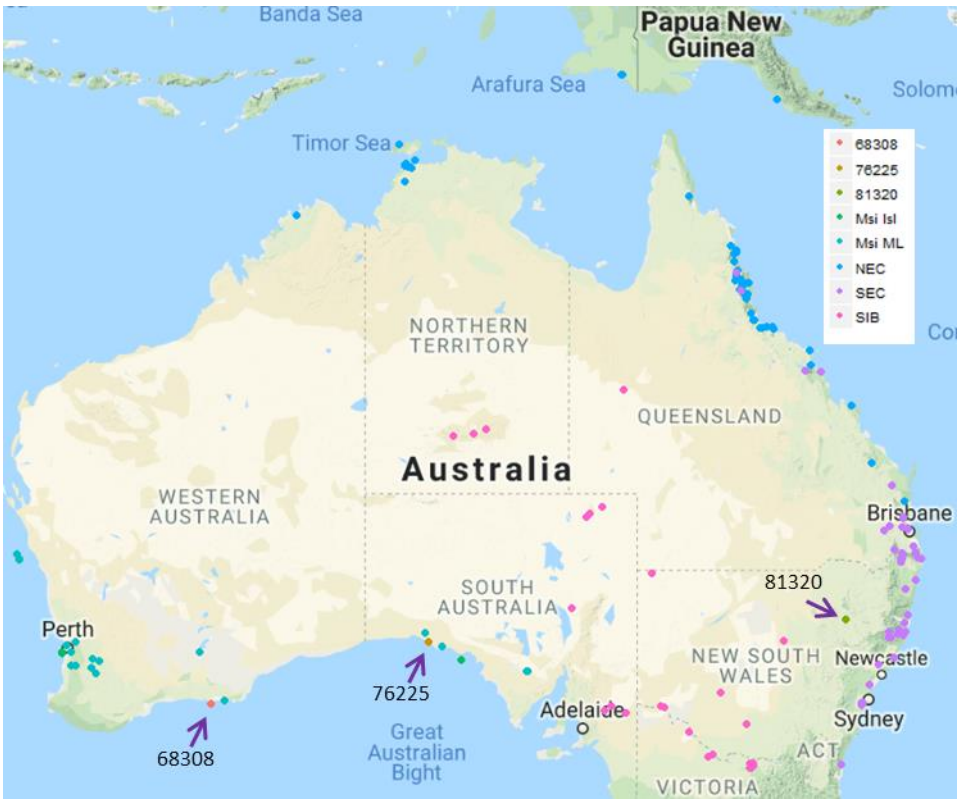


Figure A4.7: Map of n249 database showing five populations delimited by STRUCTURE K5 analysis (figure A4.6) with three admixed individuals identified. Individual 81320 that was admixed for SEC and SIB is from a location directly between these two populations, potentially a rare case of natural hybridisation between two populations separated by an effective geographic barrier.

A4.2.2 K greater than 5

The K6 STRUCTURE plots were revisited to see if further information could be gained about population substructure, despite the lack of run convergence in K values 6 onwards. Population partitioning within the five K6 replicates with the highest log likelihood values were examined for concordance.

The highest probability run resulted in identical population partitions to the K5 results, with a subsequent population created by the split of the two Msi island populations (figure A4.1). However, runs with 2nd and 4th highest likelihoods did not verify this split. A group of five individuals were split from the Southern, Internal and Bredli population, representing individuals from North-Eastern South Australia. The third and fifth highest runs gave quite different results, not replicated in any other runs. Run 44 from the K6 analysis run at lower settings gave the highest log probability, similar to the runs just discussed. The partitioning in this run was identical to run 10, splitting the two island groups to make six subpopulations.

All remaining runs (from n249 K plot) that exhibited a higher log probability score than the two highest K6 scores were examined for population partitioning concordance (table A4.2).

Table A4.1: K6 results using normal and extended run settings. Comparisons are made of the five runs with highest log probabilities with 200K burn-in and 1 million iteration run settings and single highest log probability run at 100K burn-in and 300K iterations.

Run number and settings	Log likelihood	Same as K5	Differences	Notes
Run 10 200K/1M	-24361.3	NEC, SIB, SEC, ML	Msi Islands split	
Run 3 200K/1M	-24696.8	NEC, SEC, ML, Msi Isls	SIB & NESA split	
Run 4 200K/1M	-24710.9	NEC & SFI	All other pops	Complicated
Run 1 200K/1M	-24711	NEC, SEC, ML, Msi Isls	SIB & NESA split	Identical to Run 3 above
Run 9 200K/1M	-24814	ML, Msi Isls	All other pops	
Run 44 K6 100K/300K	-24378.1	NEC, SIB, SEC, ML	Msi Islands split	Identical partitioning of individuals to run 10 above with very similar probability values. Make biological sense.

Table A4.2: Run results from K7 and K8 that gave a log probability greater than -24378.

Run number	Log likelihood	Same as K5	Differences	Notes
Run 21 K8	-24097	Only Msi Islands	All others	Result not supported by any other runs
Run 67 K8	-24118	SIB, ML, GI, SFI	Four new pops	
Run 22 K8	-24129	SIB, ML, GI, SFI	Four new pops	Identical to run 67 K8
Run 62 K8	-24143	SIB, ML, GI, SFI	Four new pops	Identical to run 67 K8
Run 61 K8	-24152	SIB, ML, GI, SFI	Four new pops	Two pops different, half identical to run 67 K8
Run 18 K7	-24181	SIB, ML, GI, SFI	Three new pops	
Run 55 K7	-24184	SIB, ML, GI, SFI	Three new pops	Identical to run 18 K7
Run 14 K7	-24196	SIB, ML, GI, SFI	Three new pops	Identical to run 18 K7

Multimodality was seen in the K8 runs, yet three runs produced Q-matrices with identical population partitioning and very similar assignment values. The run with the highest likelihood gave results not supported by any other runs. The next three highest runs replicated the K5 populations SIB and ML, while further splitting the island M. s. imbricata populations. Four new populations groupings were replicated in these runs. The fifth K8 run gave similar results, excepting two populations that had slightly different memberships. The three K7 runs exhibiting the highest log probabilities gave congruous results. They all separated the SIB and ML populations, likewise splitting the two Msi island populations, however three other populations were suggested, not replicated in any of the K6 or K8 runs.

The partitioning higher than K6 will not be considered due to lacking continuity.

A4.3 Further genetic analyses of the sub-populations

While all STRUCTURE runs do not support identical results, the groupings that are replicated across runs indicate that there could be underlying genetic structure at a finer scale within the populations that should be investigated. The SIB population appears quote robust, although a small group within might exhibit further structuring with continued investigation. The Msi mainland population also clusters robustly in almost all runs. Analyses focused within each of the five overarching populations revealed further subpopulations structure.

A4.3.1 Msi Island subpopulation

The two Msi island populations, Garden Island and St Francis Island, yielded the simplest further examination. The n38 dataset split clearly into two strongly assigned subpopulations, each representing individuals originating from a single island (figure A4.8). Higher values of K partitioned only small proportions of individuals to subsequent populations.

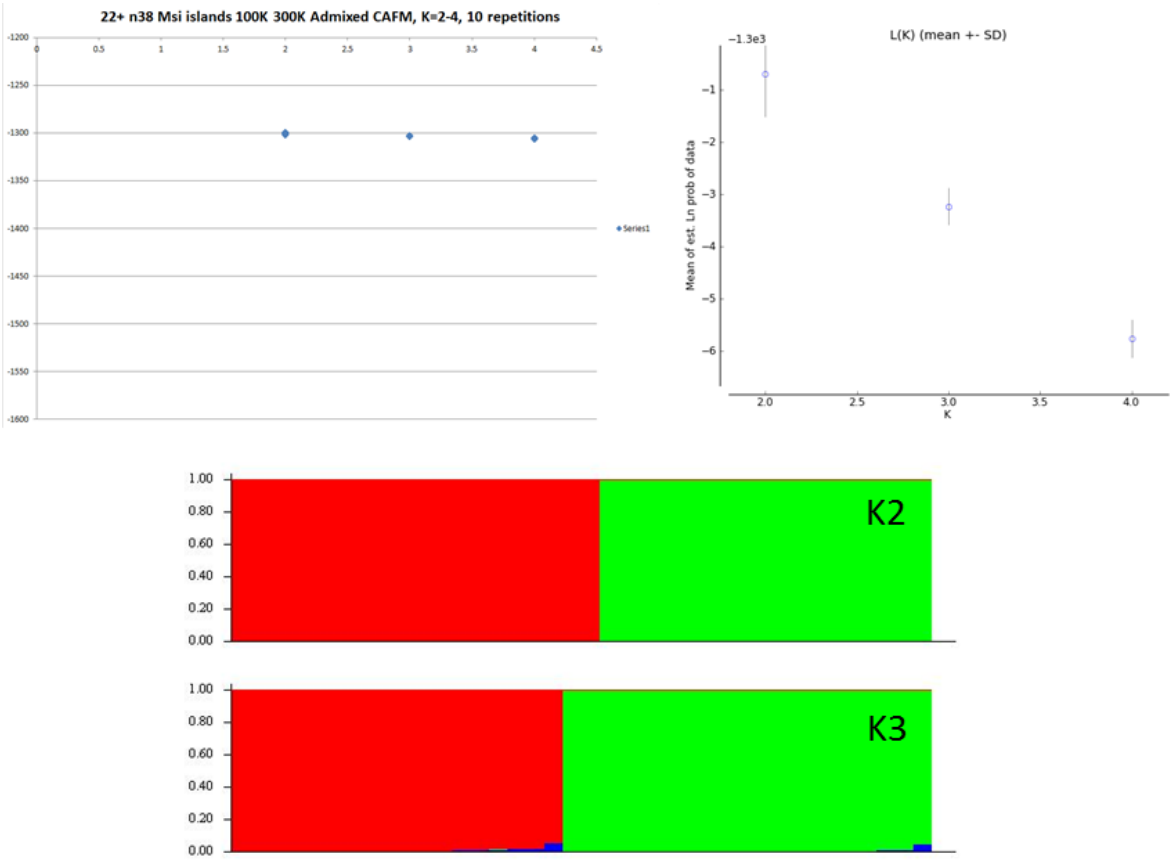


Figure A4.8: STRUCTURE analysis of Msi islands (n38) population splits these into two clear very divergent populations, indicated by the log likelihood plot (top left), Delta K plot (top right) and box plots for K2 and K3 (bottom).

A4.3.2 South Eastern Coastal subpopulation

Further investigation of the wide ranging (>2000km long) South Eastern Coastal population showed no further sign of population differentiation. All individuals were assigned with equal admixture to all putative populations for K2 and above (figure A4.9: n50 SEC Ks). This pattern of partitioning,

where every individual is split between the proposed populations in equal proportions, represents a homogeneous interbreeding population because the software is trying to introduce genetic structure where none exists.

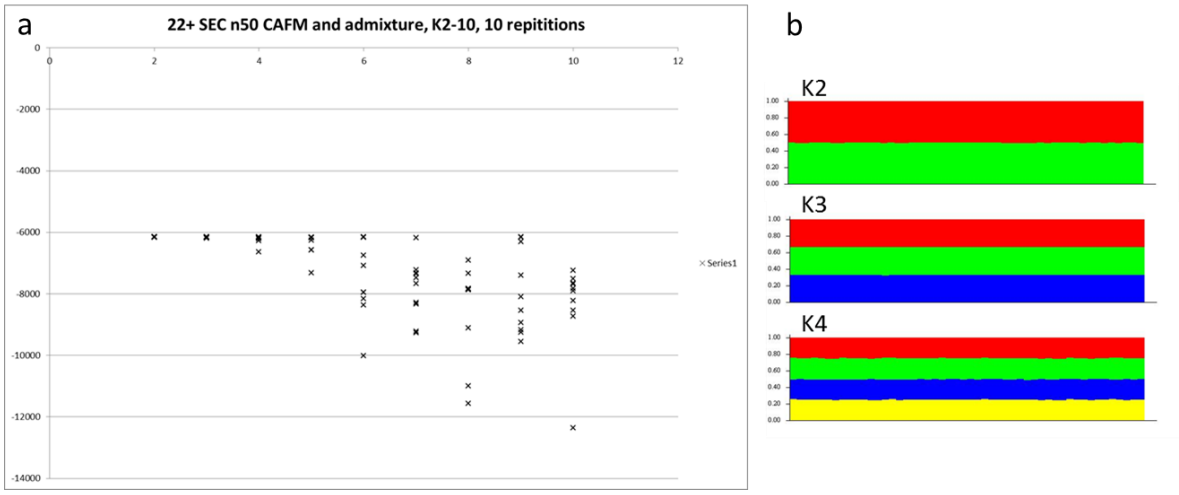


Figure A4.9: The SEC sub-population showed no further internal genetic differentiation, demonstrated by a) the log likelihood results and b) bar charts of K2-K4.

Principal Co-ordinate Analyses showed no clustering within the population (figure A4.10a). The whole population demonstrated no isolation by distance (R^2= 0.0103) however inclusion of only the extreme individuals gave a stronger regression value (R^2=0.48) indicating geographic structuring to be present over a large distance. Two sympatric populations containing in combination all SEC individuals were each found to be in allelic equilibrium (figure A4.10b), but n50 deviation from HWE is likely caused by geographic distance involved. The global FST value calculated using these two populations is 0.008 (95% upper: 0.019, lower: 0.0003).

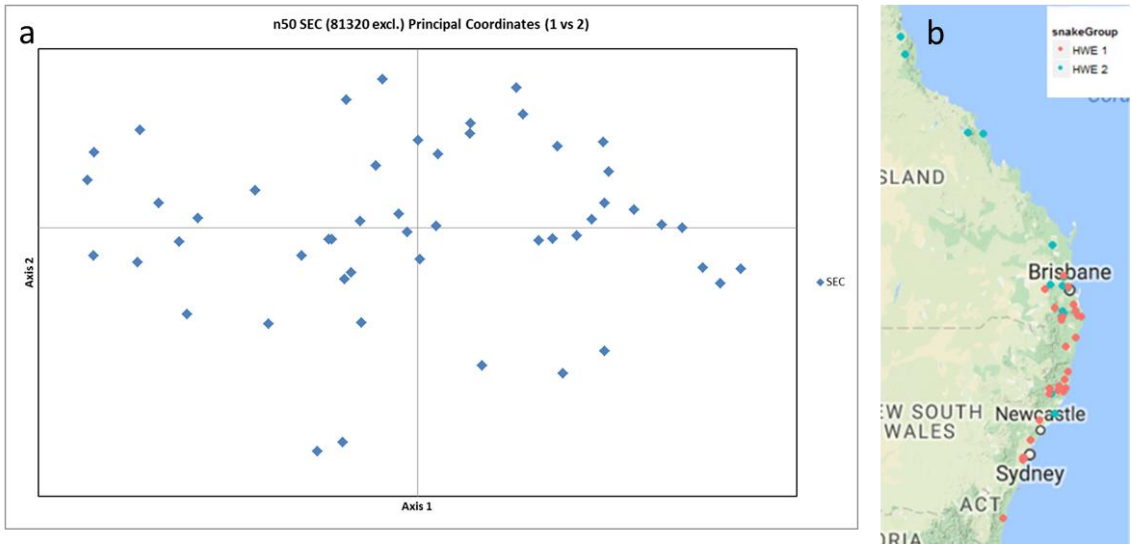


Figure A4.10: Analysis results of the n50 SEC population. a) PCoA plot of n50 population showing little to no genetic substructure and b) two internal populations conforming to HWE.

A4.3.3 Northern and Eastern Coastal subpopulation

The log likelihood and Delta K plots of further analyses for K2-10 within the Northern and Eastern Coastal subpopulation indicated three subpopulations as most likely. Average Q values are shown in figure (NEC aveQs K3-5) for K3-5. Average K3 data separated the dataset into one population from Western Papua, one population from the Northern Territory with the remaining individuals clustering into a third population. These K3 groupings did not agree with the geographic origins of the individuals, a couple of individuals each from Northern Territory and Coastal Queensland assigning unexpectedly to each other's populations. However, the K4 assignment results grouped all eight individuals from Merauke into one population with strong support and all ten individuals from the Northern Territory into a second population. The other two populations consisted of all individuals originating from Queensland, with the addition of the New Guinea snake and the snake from Northern WA. The results of this STRUCTURE analysis regarding the Northern Territory and West Papua populations make much more sense than the K3 results and were also recreated during the K5 analysis.

The two remaining populations from the K4 assignment groupings each include a number of individuals strongly assigned and also individuals that show admixture to both of these populations. This suggests that these two populations might represent one continuous population, however isolation by distance was not detected (R^2=0.0332). An additional population hypothesis (K5) caused two individuals from Tolga on the Queensland Coast to separate out into their own grouping (Figure A4.11). The two Queensland populations are seen to form Northern and Southern groupings with a large area of sympatric occupation at the interface (Figure A4.12). At K6, the three populations of Tolga, Northern Territory and West Papua remained unchanged and the remaining individuals split into three populations. Only a small number of individuals assigned strongly to any of these three populations, suggesting that the sixth artificially created population was not genetically justified.

K3 averages	Qld & NG	NT	Papua		K4 averages	Qld	Qld	NT	Papua		K5 averages	Qld	Qld	Qld	NT	Papua	
67126	0.8973	0.0362	0.0662		69344	0.97	0.0207	0.0041	0.0055		69344	0.951	0.002	0.0385	0.0032	0.0053	
84767	0.8941	0.0049	0.1011		67127	0.9678	0.021	0.0049	0.0065		67127	0.9498	0.001	0.0383	0.004	0.0066	
86801	0.893	0.0486	0.0584		67129	0.9631	0.0233	0.009	0.0045		67129	0.9465	0.0087	0.0329	0.0088	0.003	
69345	0.8928	0.0546	0.0527		79118	0.963	0.0231	0.0034	0.0101		81240	0.9452	0.003	0.034	0.0048	0.0131	
82975	0.8928	0.0111	0.0961		76778	0.9614	0.0242	0.0099	0.0048		79118	0.9424	0.004	0.0416	0.003	0.009	
76771	0.8914	0.0044	0.1042		81240	0.9606	0.0204	0.0055	0.0133		76778	0.9401	0.003	0.0436	0.0092	0.004	
69344	0.8909	0.0048	0.1044		76771	0.9601	0.0299	0.0034	0.0066		75732	0.9398	0.002	0.0394	0.0059	0.0132	
79116	0.8909	0.008	0.1009		75732	0.9594	0.021	0.0068	0.0128		80506	0.933	0.0023	0.0542	0.0065	0.0036	
86774	0.8907	0.0097	0.0999		80506	0.9593	0.0295	0.0069	0.0039		76771	0.929	0.003	0.0593	0.003	0.006	
80506	0.8899	0.0084	0.1019		82996	0.9507	0.0367	0.0048	0.0077		79122	0.9107	0.003	0.0318	0.008	0.0466	
86768	0.8894	0.0113	0.0991		84767	0.9478	0.0455	0.0039	0.0029		82986	0.9038	0.005	0.0772	0.0056	0.0084	
81239	0.8893	0.0254	0.0853		67579	0.9447	0.0402	0.0044	0.0104		80510	0.8977	0.0032	0.0832	0.0085	0.0072	
80507	0.8893	0.0372	0.0733		83022	0.9401	0.0472	0.0069	0.0056		51511	0.8965	0.001	0.0585	0.028	0.0158	
83022	0.8892	0.0077	0.103		80510	0.9391	0.0432	0.0097	0.008		79116	0.8937	0.0066	0.0924	0.004	0.003	
67127	0.8891	0.0051	0.1059		79116	0.939	0.0521	0.0047	0.0042		82996	0.8932	0.002	0.0928	0.004	0.0079	
82996	0.8885	0.0057	0.1057		82986	0.9326	0.0515	0.0061	0.0096		67579	0.8844	0.002	0.0995	0.004	0.0102	
79118	0.8875	0.0049	0.1076		86774	0.9292	0.0578	0.0097	0.0036		84767	0.8708	0.002	0.1211	0.003	0.003	
66297	0.8863	0.0793	0.0342		51511	0.9258	0.0283	0.031	0.0149		82975	0.8606	0.007	0.1251	0.0048	0.003	
67579	0.8859	0.0059	0.1081		79122	0.9224	0.0197	0.0111	0.0467		83022	0.8568	0.0016	0.1294	0.0066	0.0056	
67129	0.8855	0.0117	0.1028		82975	0.9031	0.0882	0.0051	0.0035		86774	0.8271	0.002	0.1574	0.0106	0.003	
79111	0.8852	0.0718	0.0433		82967	0.8821	0.0715	0.0426	0.0036		62456	0.8199	0.001	0.0226	0.1422	0.0139	
82986	0.8852	0.0117	0.103		67130	0.8789	0.0857	0.0046	0.0305		82967	0.7989	0.0029	0.1546	0.0409	0.003	
76778	0.8851	0.0122	0.1026		86768	0.871	0.1158	0.0097	0.0038		81300	0.7766	0.0139	0.1924	0.0104	0.0068	
79110	0.8835	0.0567	0.0597		81029	0.8429	0.1362	0.0058	0.015		79117	0.7631	0.0107	0.2087	0.0145	0.003	
80510	0.8832	0.0119	0.1049		82987	0.8312	0.1288	0.0095	0.0306		80511	0.7628	0.0862	0.0607	0.0813	0.0089	
81029	0.8829	0.0141	0.1032		62456	0.8268	0.0161	0.1441	0.013		67130	0.7384	0.004	0.222	0.004	0.0314	
66139	0.8827	0.0642	0.0532		79117	0.8154	0.1642	0.0155	0.0048		82974	0.7111	0.0036	0.0412	0.2389	0.0051	
81240	0.8812	0.0064	0.1123		81300	0.7626	0.2153	0.0127	0.0094		80507	0.6853	0.0205	0.2861	0.0051	0.003	
69346	0.881	0.0258	0.0929		82974	0.7523	0.0248	0.2172	0.0057		81029	0.6836	0.0094	0.2887	0.0049	0.0136	
75732	0.8807	0.0083	0.1108		81239	0.7418	0.2426	0.0077	0.008		86768	0.6354	0.002	0.3497	0.0096	0.0033	
79117	0.8795	0.0318	0.0886		69346	0.7308	0.2444	0.0045	0.0202		82987	0.6298	0.003	0.3299	0.0072	0.0299	
81300	0.878	0.0369	0.0852		80511	0.7143	0.1889	0.0705	0.0264		69346	0.6232	0.0095	0.3463	0.0039	0.0174	
79119	0.8758	0.0786	0.0458		80507	0.6869	0.3017	0.0058	0.0058		81239	0.62	0.0066	0.3598	0.0069	0.0067	
82967	0.8713	0.0293	0.0995		76777	0.5145	0.2582	0.03	0.1972		79110	0.549	0.0175	0.4211	0.0056	0.007	
67130	0.8675	0.0096	0.1229		79110	0.5027	0.4796	0.007	0.0108		81316	0.001	0.996	0.001	0.001	0.001	Tolga, Qld
82987	0.8666	0.0183	0.1149		86801	0.5	0.488	0.0042	0.0078		81317	0.001	0.996	0.001	0.001	0.001	Tolga, Qld
51511	0.8658	0.0233	0.1109		67126	0.4942	0.4951	0.0049	0.0056		66297	0.0297	0.0017	0.9624	0.004	0.0027	
66142	0.8643	0.0886	0.0473		66297	0.1888	0.8026	0.0047	0.0034		66138	0.0311	0.001	0.9535	0.0031	0.0108	
66138	0.8519	0.0775	0.0706		66140	0.1905	0.7922	0.0093	0.008		67578	0.0396	0.0076	0.9364	0.0097	0.0066	
79122	0.8345	0.0158	0.15		67578	0.1869	0.7915	0.0121	0.0096		66142	0.0469	0.0011	0.9281	0.0153	0.0081	
67578	0.8317	0.1278	0.0404		81316	0.0028	0.7888	0.0077	0.2009	tolga, qld	81028	0.0402	0.003	0.9275	0.017	0.0123	
66140	0.8306	0.1315	0.0379		81317	0.0029	0.7885	0.0084	0.2008	tolga, qld	66140	0.0543	0.0398	0.8947	0.0081	0.003	
62456	0.8287	0.0642	0.1066	N WA	66138	0.1878	0.7866	0.0042	0.0214		79111	0.1001	0.0073	0.88	0.0086	0.004	
81028	0.7891	0.1554	0.0553		81028	0.1875	0.7619	0.0316	0.0186		66139	0.1276	0.0099	0.8439	0.0073	0.0113	
76777	0.7166	0.066	0.2176		66142	0.2285	0.7312	0.0283	0.0119		79119	0.201	0.0261	0.7608	0.0061	0.0063	
82974	0.7116	0.1858	0.1028	QLD	79111	0.2597	0.7238	0.0109	0.0053		67126	0.2287	0.005	0.7574	0.004	0.0041	
13382	0.6081	0.2909	0.101	NT	79119	0.2831	0.6991	0.0073	0.0104		12350	0.0918	0.0039	0.7133	0.0089	0.1819	
12350	0.5865	0.1374	0.276	New Guin	66139	0.2919	0.6814	0.01	0.0165		69345	0.3215	0.0058	0.6627	0.006	0.004	
80511	0.586	0.3073	0.1071	QLD	69345	0.421	0.5664	0.0076	0.0051		86801	0.4662	0.0167	0.5085	0.0035	0.0051	
13449	0.0118	0.9239	0.0643	Darwin	12350	0.2037	0.5449	0.0297	0.2216		76777	0.3115	0.004	0.4682	0.0124	0.2043	
11887	0.0161	0.9101	0.0739	Darwin	13449	0.0044	0.0052	0.9868	0.0035	Darwin	13449	0.004	0.002	0.0049	0.986	0.003	
11888	0.0282	0.9041	0.0676	Darwin	11888	0.01	0.0069	0.9778	0.0054	Darwin	11887	0.0083	0.002	0.0068	0.9786	0.0043	
11887	0.0317	0.901	0.0671	Darwin	11887	0.0111	0.0065	0.9776	0.0047	Darwin	11888	0.0077	0.007	0.0045	0.9769	0.0041	
13455	0.0196	0.8309	0.1496	Darwin	11887	0.0065	0.0247	0.9522	0.0166	Darwin	13443	0.0157	0.001	0.0075	0.9705	0.005	
55482	0.1111	0.7975	0.0915	Darwin	13443	0.0415	0.0065	0.9465	0.0053	Darwin	30087	0.0276	0.002	0.0091	0.958	0.0038	
30087	0.1471	0.7759	0.0771	NT	30087	0.0599	0.0074	0.929	0.0039	NT	11887	0.0054	0.0354	0.0072	0.9411	0.0108	
13443	0.1607	0.7602	0.0791	Darwin	55482	0.0526	0.008	0.9133	0.0263	Darwin	82476	0.0624	0.0039	0.0142	0.9149	0.005	
82476	0.2767	0.6301	0.0931	NT	13455	0.0093	0.0079	0.8929	0.09	Darwin	55482	0.0502	0.002	0.0114	0.9105	0.0262	
81317	0.1166	0.6275	0.2559	tolga, qld	82476	0.1358	0.0132	0.8459	0.0054	NT	13455	0.008	0.003	0.0085	0.8809	0.0998	
81316	0.117	0.6267	0.2565	tolga, qld	13382	0.3094	0.0147	0.6713	0.0046	NT	13382	0.2147	0.001	0.0336	0.746	0.0047	
11706	0.0045	0.2018	0.794	Merauke	11706	0.004	0.0038	0.1021	0.8901	Merauke	11706	0.0032	0.004	0.003	0.003	0.9866	
67190	0.0101	0.1985	0.7916	Merauke	67190	0.0084	0.0053	0.0999	0.8866	Merauke	67190	0.0069	0.0058	0.0039	0.003	0.9808	
11707	0.0091	0.2027	0.7884	Merauke	11706	0.0081	0.0039	0.1079	0.88	Merauke	11706	0.0069	0.002	0.004	0.0086	0.9785	
11706	0.0083	0.2052	0.7864	Merauke	67164	0.011	0.0148	0.1026	0.8718	Merauke	67164	0.0094	0.002	0.0136	0.0061	0.969	
11706	0.0244	0.1909	0.7846	Merauke	11707	0.0068	0.0206	0.1034	0.8693	Merauke	11707	0.0061	0.0042	0.0182	0.0061	0.9653	
67164	0.0146	0.2015	0.7838	Merauke	11706	0.019	0.0203	0.0938	0.8668	Merauke	11706	0.0148	0.002	0.0199	0.003	0.9608	
68280	0.0377	0.2145	0.7477	Merauke	68280	0.0369	0.0222	0.1232	0.8179	Merauke	68280	0.0352	0.003	0.0311	0.0289	0.902	
68279	0.0913	0.2488	0.6603	Merauke	68279	0.0807	0.0386	0.1964	0.6841	Merauke	68279	0.0763	0.002	0.0626	0.1021	0.7564	

Figure A4.11: Average Q-value matrices for three, four and five putative NEC sub-populations. Individuals that swap primary assigned population between analyses are highlighted in red.

The whole dataset K7 and K8 (figure A4.13) runs did not yield any populations that were identically partitioned to these populations. One K8 population was very close, replicating the Merauke/Northern Territory grouping with the addition of 62456 from Northern WA. The K7 average produced this population with the addition of the New Guinean individual 123504 and two individuals from the Northern Queensland coast. In the K8 run, these three individuals grouped with a much larger contingent of samples from Northern Queensland. The composition of the NEC and SEC groupings also changed slightly in each run, showing replicated assignment at the extremes, but shifts of the central boundary zone. This result suggests that these samples cannot easily be

delimited as separate populations, but are more likely to represent one continuous coastal population.

Figure A4.12: Maps shows the geographic locations of the five NEC sub-populations delineated by the STRUCTURE K5 analysis.

Figure A4.13: Populations designated within K7 and K8 STRUCTURE analyses, but likely continuous populations.

The entire NEC/SEC population of 118 individuals exhibited only weak (R^2=0.1373, population by population; R^2=0.1347 individual by individual) signal for isolation by distance (figure A4.14a).

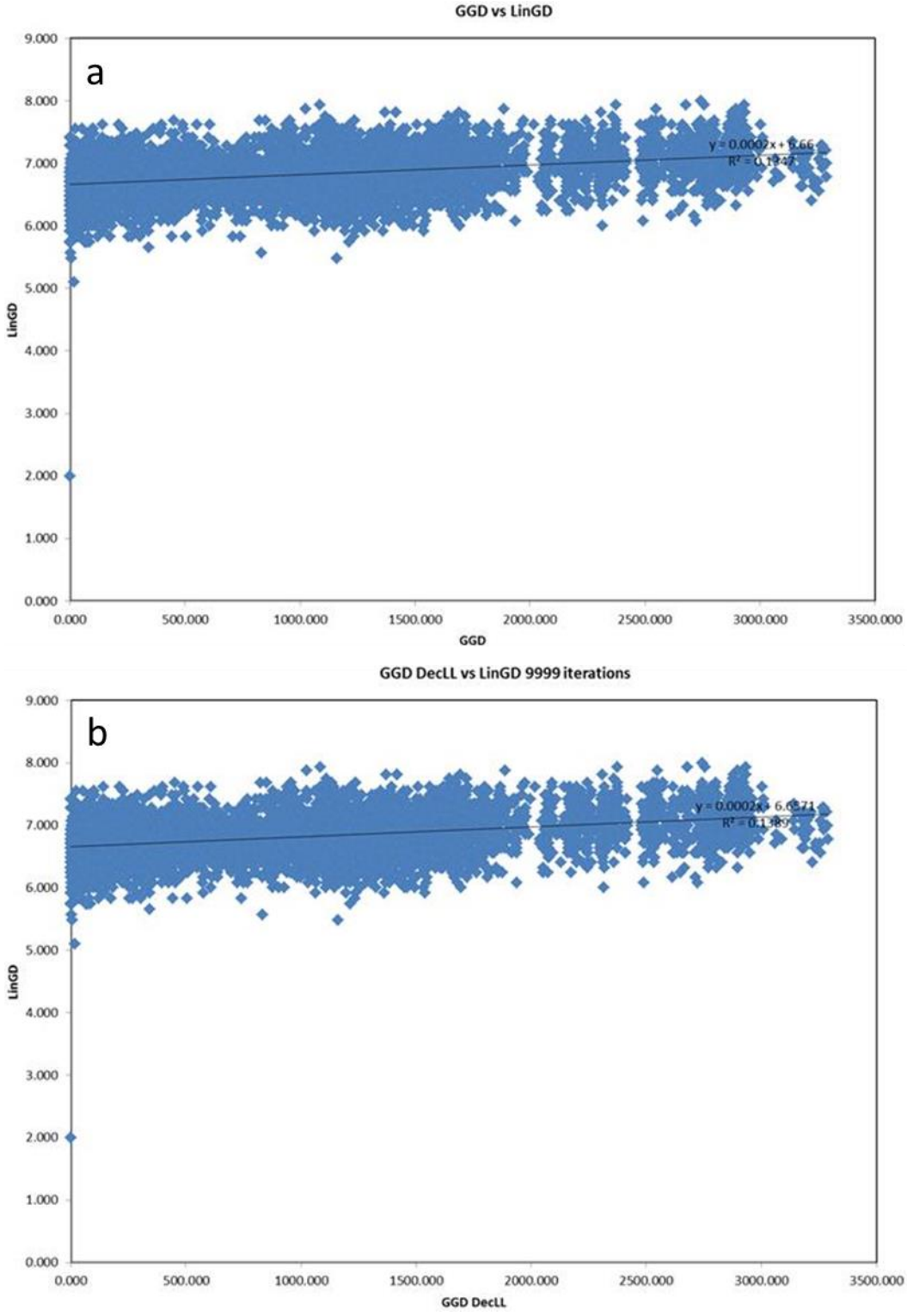


Figure A4.14 n118 IBD: a) Mantel test of n118 combination of NEC and SEC sub-populations showing weak isolation by distance, and b) n115 with three outlying individuals removed.

Alongside 81320, two individuals 68285 and 69346 were removed because they appeared potentially anomalous, remotely located from genetically similar individuals. The new n115 was

tested for IBD and gave a similarly low regression of R^2=0.1415, population by population or R^2=0.1389 individual by individual (figure A4.14b). The weak isolation by distance results concur with the clustering of large populations. The disequilibrium of alleles within loci could be caused by frequent migration, violating the HW assumption. Unfortunately, this scenario is problematic for geographic assignment analyses.

A population comprised of the Northern Territory and Western Papuan samples significantly deviated from Hardy-Weinberg equilibrium, including and excluding 62456. Neither the West Papua or Northern Territory populations alone were considered large enough for accurate Hardy-Weinberg analysis. Datasets of various sizes, based on the assignment scores of the remaining individuals, were tested for adherence to HWE. A grouping of 33 individuals displayed equilibrium, which was surprising as these individuals are spread over a very large geographic range (figure A4.15). Interestingly, the data suggest the New Guinean sample to be very closely related to the Northern Australian samples, while the Papuan samples from Merauke split into their own group.

Figure A4.15. Geographic locations of NEC individuals not significantly deviation from HWE.

The Principal Co-ordinate Analysis of the combined NEC and SEC populations was suggestive of a small degree of directionality in the clustering viewed from one or two planes (figure A4.16), however, the clusters are very spread out, indicating a lot of genetic variability among individuals within this dataset.

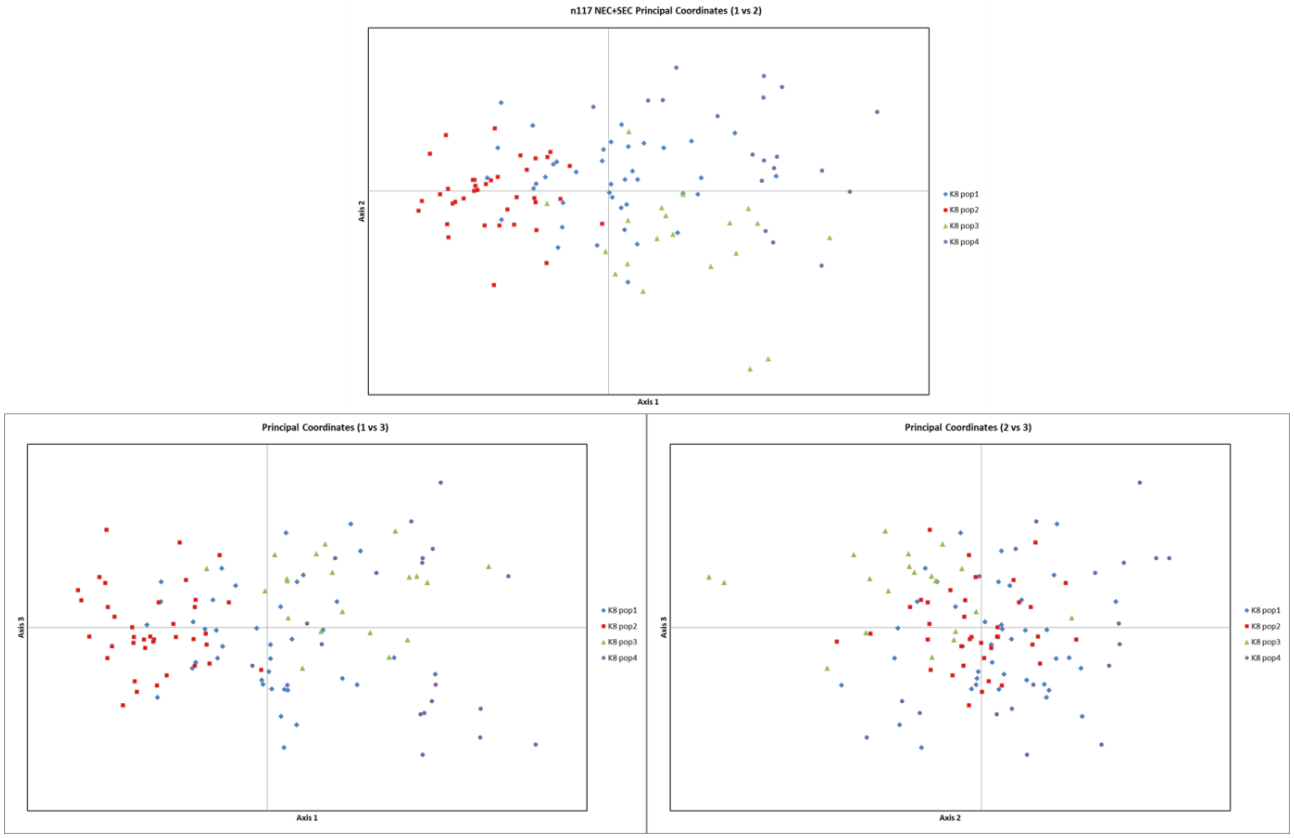


Figure A4.16: Principal Co-ordinate Analysis of the combined n118 NEC and SEC populations, viewed on three comparative planes.

A4.3.4 Msi Mainland subpopulation

Subsequent analysis of the Msi Mainland population (n36) suggested 3 or 5 further subpopulations (Figure A4.17). One outlying K5 run was excluded from calculations of Q value averages.

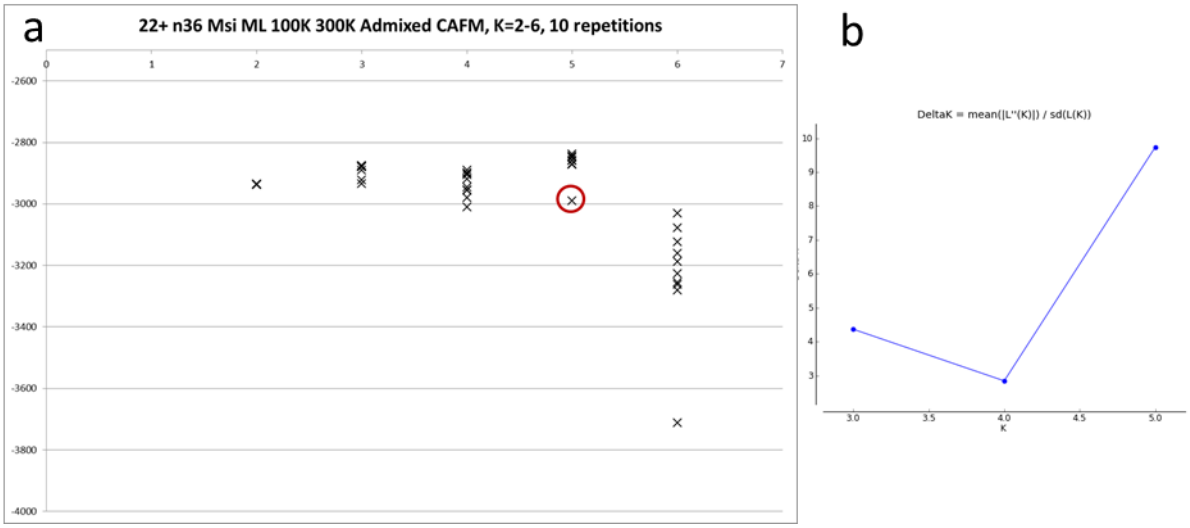


Figure A4.17: Log likelihood and Delta K plots of the n36 Msi ML subpopulation using K2-K6. The outlying run at K5 is circled. K3 and K5 are suggested, but K5 replicates cluster more tightly and are suggested as more likely by both plots.

At all three K values, two sub populations were consistently divergent: the West Wallabi Island population (n4) and the two individuals from Yorke Peninsula in South Australia (figure A4.18). These two populations represent the extremes of the M. s. imbricata range.

K=3	Msi WWI	Middle	York Pen.
67139	0.9979	0.001	0.001
67137	0.997	0.002	0.001
67140	0.997	0.002	0.001
67136	0.9941	0.004	0.0019
67131	0.0029	0.9958	0.0018
66333	0.002	0.9952	0.0024
66329	0.001	0.9951	0.0036
66032	0.003	0.9945	0.0026
66323	0.003	0.9944	0.0022
67146	0.003	0.9939	0.0034
66320	0.004	0.9938	0.0022
66327	0.004	0.9938	0.0024
66326	0.002	0.9938	0.0043
66339	0.0026	0.9935	0.004
67138	0.002	0.9919	0.0061
66336	0.0016	0.9919	0.0066
66325	0.002	0.9918	0.0067
67147	0.0053	0.9916	0.0032
67134	0.0038	0.9916	0.0048
66332	0.0049	0.9907	0.0043
66322	0.003	0.9905	0.0064
51575	0.002	0.9896	0.0087
66328	0.0032	0.9892	0.0073
66330	0.007	0.9876	0.0053
67149	0.001	0.9805	0.018
67135	0.0031	0.9688	0.028
51576	0.0152	0.9659	0.0188
51574	0.0045	0.9648	0.031
66331	0.0055	0.9612	0.0334
76225	0.003	0.9127	0.0843
68298	0.0025	0.9109	0.0863
81224	0.004	0.8853	0.1107
68308	0.003	0.8678	0.1291
17482	0.0013	0.0021	0.9968
68341	0.002	0.0022	0.9958
68310	0.001	0.1618	0.8372

K=4	Msi WWI	All SW WA	Coastal SA	York Pen.
67139	0.996	0.001	0.001	0.001
67137	0.995	0.002	0.002	0.001
67140	0.9949	0.002	0.002	0.001
67136	0.991	0.003	0.004	0.002
66325	0.002	0.9694	0.0218	0.007
66326	0.002	0.9447	0.0479	0.005
66328	0.0031	0.9431	0.0451	0.0087
66330	0.0082	0.9332	0.0535	0.005
67134	0.0044	0.9275	0.0638	0.0042
66329	0.001	0.9041	0.0915	0.0036
67149	0.001	0.8651	0.1172	0.0163
67131	0.003	0.8617	0.1339	0.0019
67147	0.007	0.8444	0.1455	0.003
66332	0.005	0.8366	0.1544	0.0042
67138	0.002	0.8365	0.1559	0.0059
66327	0.0046	0.8323	0.1609	0.0022
66339	0.003	0.829	0.1642	0.004
66336	0.002	0.7322	0.2595	0.007
66333	0.0028	0.7149	0.2801	0.0022
66323	0.003	0.6676	0.3265	0.0023
67135	0.0039	0.597	0.3686	0.0307
66320	0.004	0.573	0.4202	0.0028
51575	0.002	0.5108	0.482	0.0055
66331	0.006	0.4984	0.4228	0.0729
76225	0.003	0.0091	0.981	0.0071
68308	0.003	0.0099	0.9768	0.0105
68298	0.0021	0.0164	0.974	0.007
81224	0.0033	0.0092	0.9516	0.0358
51574	0.0042	0.1109	0.8752	0.0094
51576	0.0148	0.1938	0.7869	0.0043
67146	0.0033	0.3603	0.6334	0.003
66322	0.0036	0.4733	0.5185	0.0045
66032	0.003	0.4963	0.4987	0.002
68310	0.001	0.0118	0.4602	0.527
17482	0.001	0.002	0.002	0.995
68341	0.002	0.002	0.002	0.9949

>0.35 admixture

Average values	Msi WWI	SW WA	SE WA	Coastal SA +	York Pen. SA
67139	0.995111111	0.001	0.001	0.001	0.001
67137	0.994	0.001889	0.001	0.002	0.001
67140	0.994	0.002	0.001	0.002	0.001
67136	0.989888889	0.003	0.002	0.003888889	0.002
66325	0.001444444	0.861778	0.002222222	0.127333333	0.006888889
66328	0.003	0.828444	0.002777778	0.156333333	0.009555556
66326	0.002	0.827	0.003	0.163111111	0.004888889
66330	0.007333333	0.818222	0.005333333	0.165555556	0.003111111
67134	0.004	0.81	0.006	0.176888889	0.003111111
66329	0.001	0.797	0.002888889	0.196111111	0.003
67149	0.001	0.757111	0.006777778	0.226	0.008888889
67131	0.003	0.744889	0.002111111	0.248444444	0.001111111
67138	0.002	0.730778	0.008	0.255444444	0.004111111
67147	0.006111111	0.712778	0.002111111	0.276222222	0.002888889
66332	0.004222222	0.704889	0.002222222	0.284555556	0.003888889
66339	0.002777778	0.692778	0.004888889	0.296555556	0.003111111
66327	0.004	0.680222	0.002777778	0.310666667	0.002
66336	0.001888889	0.611	0.004111111	0.378222222	0.005333333
66333	0.002	0.578889	0.002111111	0.414444444	0.002
66323	0.003	0.534111	0.002111111	0.458333333	0.002
68308	0.002	0.003	0.953555556	0.005666667	0.035222222
68310	0.001	0.003444	0.939555556	0.005444444	0.051111111
76225	0.003	0.104889	0.011888889	0.872444444	0.008111111
68298	0.002	0.115889	0.014666667	0.857222222	0.009666667
81224	0.003111111	0.100889	0.013777778	0.825333333	0.056555556
51574	0.004	0.167333	0.006111111	0.816666667	0.005777778
67146	0.003	0.261667	0.003111111	0.729333333	0.003
51576	0.014888889	0.271333	0.014111111	0.695666667	0.004111111
66032	0.002888889	0.371778	0.003111111	0.620888889	0.002
66322	0.003	0.394889	0.006888889	0.592	0.003111111
66320	0.004	0.414444	0.002	0.577444444	0.002
66331	0.005888889	0.382222	0.006	0.520222222	0.086111111
51575	0.002	0.466	0.007555556	0.519666667	0.005
67135	0.003111111	0.482111	0.006666667	0.480111111	0.027555556
17482	0.001	0.001	0.035333333	0.001111111	0.960666667
68341	0.001	0.001222	0.035888889	0.002	0.959777778

Colours correspond to K5 PCoA plot

Figure A4.18: n36 Msi ML K3-5 Q-matrices: Average Q-matrices for K3, K4 and K5. West Wallabi Island individuals are shown in blue and Yorke Peninsula individuals in yellow.

The geographic dispersal of the K5 populations are shown (figure A4.19).

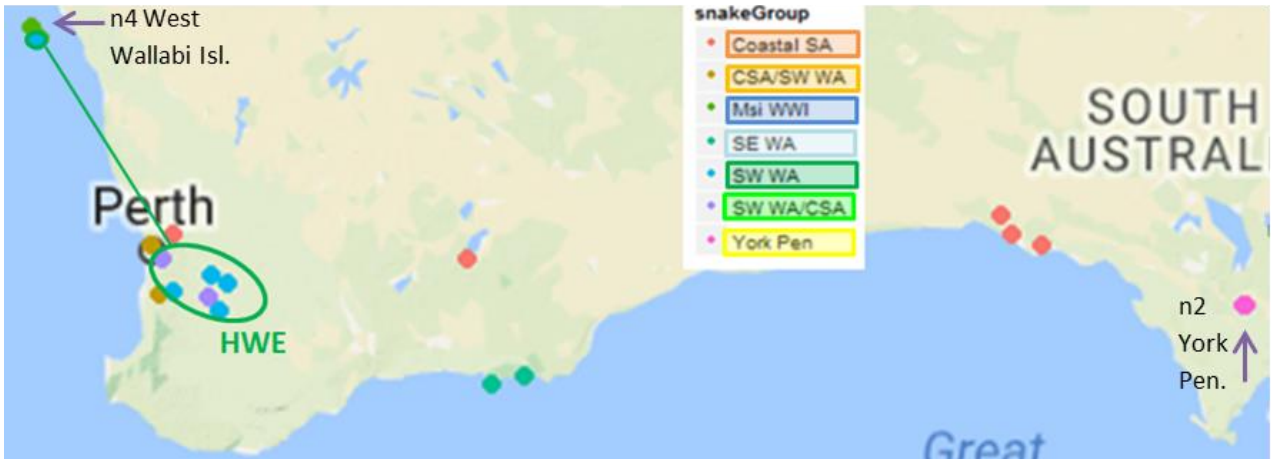


Figure A4.19: Geographic locations of populations from the K5 analysis. The coloured boxes in the key relate to the coloured designations in figure A4.18).

K=3	Msi WWI	Middle	York Pen.
67139	0.9979	0.001	0.001
67137	0.997	0.002	0.001
67140	0.997	0.002	0.001
67136	0.9941	0.004	0.0019
67131	0.0029	0.9958	0.0018
66333	0.002	0.9952	0.0024
66329	0.001	0.9951	0.0036
66032	0.003	0.9945	0.0026
66323	0.003	0.9944	0.0022
67146	0.003	0.9939	0.0034
66320	0.004	0.9938	0.0022
66327	0.004	0.9938	0.0024
66326	0.002	0.9938	0.0043
66339	0.0026	0.9935	0.004
67138	0.002	0.9919	0.0061
66336	0.0016	0.9919	0.0066
66325	0.002	0.9918	0.0067
67147	0.0053	0.9916	0.0032
67134	0.0038	0.9916	0.0048
66332	0.0049	0.9907	0.0043
66322	0.003	0.9905	0.0064
51575	0.002	0.9896	0.0087
66328	0.0032	0.9892	0.0073
66330	0.007	0.9876	0.0053
67149	0.001	0.9805	0.018
67135	0.0031	0.9688	0.028
51576	0.0152	0.9659	0.0188
51574	0.0045	0.9648	0.031
66331	0.0055	0.9612	0.0334
76225	0.003	0.9127	0.0843
68298	0.0025	0.9109	0.0863
81224	0.004	0.8853	0.1107
68308	0.003	0.8678	0.1291
17482	0.0013	0.0021	0.9968
68341	0.002	0.0022	0.9958
68310	0.001	0.1618	0.8372

K=4	Msi WWI	All SW WA	Coastal SA	York Pen.
67139	0.996	0.001	0.001	0.001
67137	0.995	0.002	0.002	0.001
67140	0.9949	0.002	0.002	0.001
67136	0.991	0.003	0.004	0.002
66325	0.002	0.9694	0.0218	0.007
66326	0.002	0.9447	0.0479	0.005
66328	0.0031	0.9431	0.0451	0.0087
66330	0.0082	0.9332	0.0535	0.005
67134	0.0044	0.9275	0.0638	0.0042
66329	0.001	0.9041	0.0915	0.0036
67149	0.001	0.8651	0.1172	0.0163
67131	0.003	0.8617	0.1339	0.0019
67147	0.007	0.8444	0.1455	0.003
66332	0.005	0.8366	0.1544	0.0042
67138	0.002	0.8365	0.1559	0.0059
66327	0.0046	0.8323	0.1609	0.0022
66339	0.003	0.829	0.1642	0.004
66336	0.002	0.7322	0.2595	0.007
66333	0.0028	0.7149	0.2801	0.0022
66323	0.003	0.6676	0.3265	0.0023
67135	0.0039	0.597	0.3686	0.0307
66320	0.004	0.573	0.4202	0.0028
51575	0.002	0.5108	0.482	0.0055
66331	0.006	0.4984	0.4228	0.0729
66032	0.003	0.4963	0.4987	0.002
66322	0.0036	0.4733	0.5185	0.0045
67146	0.0033	0.3603	0.6334	0.003
51576	0.0148	0.1938	0.7869	0.0043
51574	0.0042	0.1109	0.8752	0.0094
81224	0.0033	0.0092	0.9516	0.0358
68298	0.0021	0.0164	0.974	0.007
68308	0.003	0.0099	0.9768	0.0105
76225	0.003	0.0091	0.981	0.0071
68310	0.001	0.0118	0.4602	0.527
17482	0.001	0.002	0.002	0.995
68341	0.002	0.002	0.002	0.9949

Figure A4.20: At K4, the intermediate K3 population (green on left) is split, and exhibits a continuum of admixture, between the two resulting populations, suggesting that this might be a single population exhibiting isolation by distance at the extremes. Colouring designates 10% increments of admixture between the two populations. When the membership is forced into three populations, they appear quite defined. When opened up to four populations, even individual 68310 on the Yorke Peninsula appears to be a continuation of the Coastal SA group.

A Mantel test involving all 36 individuals showed weak evidence of IBD (R^2 = 0.1688). At K3, the remainder of the Msi ML population formed one subpopulation. However this population did not adhere to Hardy Weinberg equilibrium, unsurprisingly given that it spreads over such a large geographic range. At K4, this mainland population split into two separate populations, yet these two populations exhibit a continuum of admixture between the resulting populations (figure A4.20) possibly suggestive of isolation by distance. At K5, two individuals formed a subsequent population and the same two K4 populations exhibited roughly the same citizenship.

Figure A4.21 shows the clustering of individuals at K3 and demarcation of individuals that showed admixture at K5. Note that the K3 plots only capture a quarter of the variability represented within the genotypic dataset. Between K4 and K5, some of the admixed individuals switched their primary assignment value between subpopulation, although still exhibited a similar degree of admixture between the two subpopulations. This pattern is potentially suggestive of an isolation-by-distance relationship, where individuals at the extremes of the populations are assigning strongly to either respectively, but intermediate individuals assign to both. Mantel testing conducted on 27 individuals (no WWI, Yorke Pen or SE WA; 67138 removed as geographically remote) was more strongly supportive of IBD (R^2 = 0.3008).

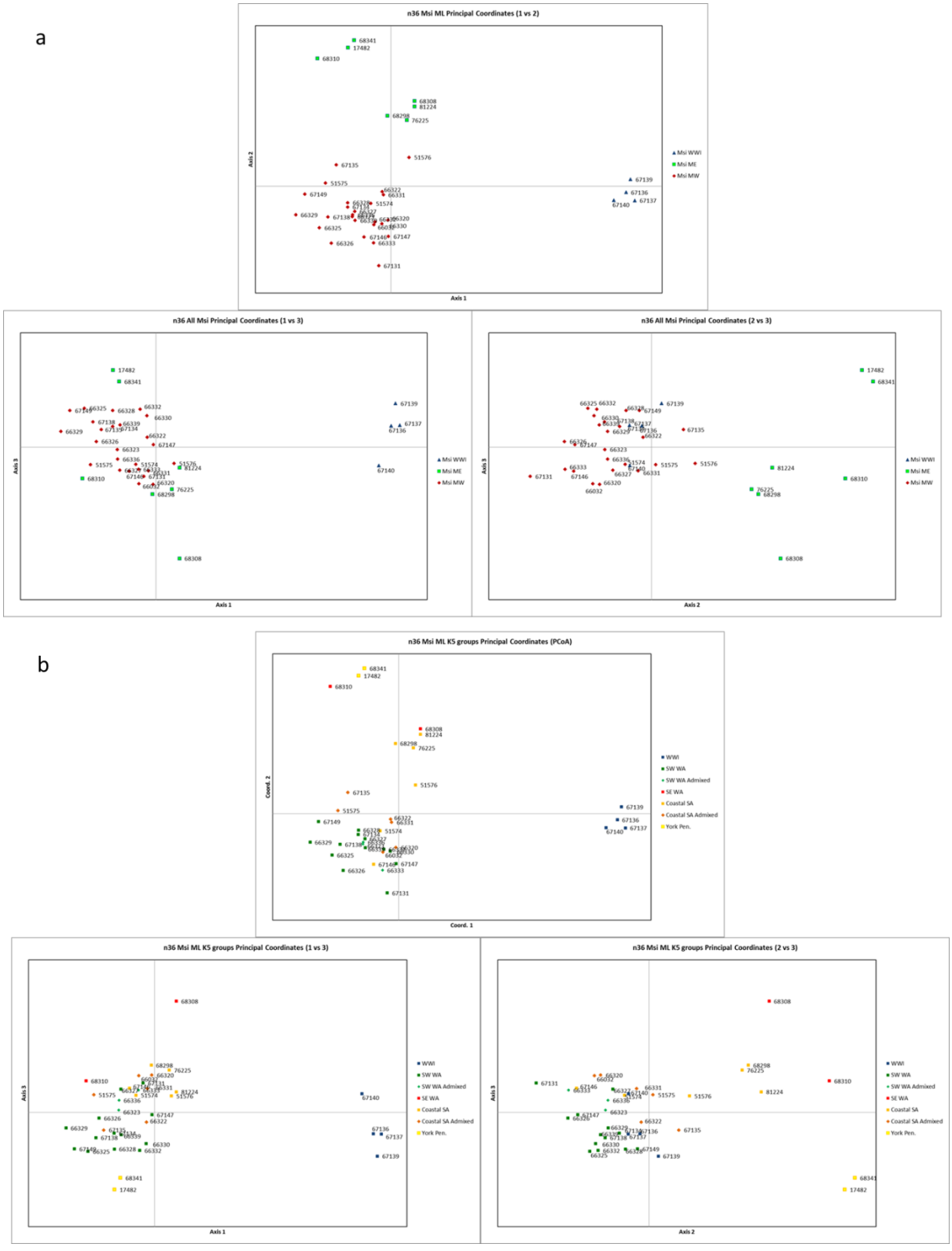


Figure A4.21: PCoA plots of the Msi ML sub-population using a) K3 and b) K5 sub-population partitions. The K5 sub-populations are split according to the colouring given in the figure A4.18 K5 graph and also seen in the (K5 map). Note that the a) PCoA plot only represents 26.29% of the total variability in the dataset.

The 16 individuals that assigned more strongly to the Western extreme of the population (green in figure A4.18 and figure A4.19) were located within a relatively small geographic range. Allelic patterns within this subpopulation adhered to Hardy-Weinberg equilibrium, suggesting this to represent one interbreeding population (South West Western Australia, SW WA). The Q-matrices and principal co-ordinate analyses (Figure A4.21a) showed the two island individuals 68308 and 68310 who formed their own population at K5 to be more closely genetically related to the South Australian population, particularly 68310 who grouped genetically with the Yorke Peninsula population in K3 even though geographically removed from these two individuals.

The grouping of the two Southern Island WA individuals (68308 and 68310) at K5 is interesting given that the overarching five population assignment of the total dataset identified 68308 and 76225 to similarly be showing admixture between Msi and the Eastern Coastal populations. Given that 68308 grouped closely with the Coastal SA populations at K4, this subsequent level of structure might be an overestimation of population number.

In summary, this population appears to contain two small subsets that are quite genetically different from the rest of this population: the four West Wallabi Island individuals (67136, 67137, 67139 and 67140) and the two Yorke Peninsula individuals (17482 and 68341). There is a cluster of genetically similar individuals in SW WA that exhibits HWE. The remaining individuals appear to be one continuous population involving isolation by distance, however the limited sample number spanning a large geographic range has led to the consequence that single individuals are divergent enough that they may well be representative of separate putative genetic subpopulations. This scenario is problematic for routine population genetic analyses that are used to determine how best to apply forensic genetic statistics to the dataset. The dataset is particularly difficult to define due to the limited sampling (indeed limited habitation) of snakes along southern coastal South Australia and Western Australia and the identification of a subpopulation as only represented by a very small number of individuals.

A4.3.5 Southern, Internal and Bredli population

Analysis of the n56 Southern, Internal and Bredli population revealed the most likely number of subpopulations to be somewhere between 3 and 5, while the delta K plot suggested three subpopulations (Figure A4.22). Advice for using STRUCTURE with correlated allele frequencies is that this can cause an overestimation of the number of populations when assessing the log likelihood results and the fewest number of populations that make biological and geographical sense should be chosen. However K4 points were well converged and the box plots still exhibited relatively little admixture. Examination of the geographic groupings (figure A4.23) indicated by average Q values for K3-6 (figure A4.24) suggested three small subpopulations to be quite genetically different from the rest of this dataset. K3 saw the separation of 21 individuals within the Warby Ranges of Victoria into a subpopulation, as well as another subpopulation of 5 individuals in Goyder Lagoon in the North East corner of South Australia. This is the same NE SA grouping seen to split from the SIB population in some of the initial n249 K6 STRUCTURE runs. Only the Warby Ranges population (SE Vic) is in Hardy-Weinberg equilibrium, the Goyder Lagoon population (NE SA) is considered too small to estimate.

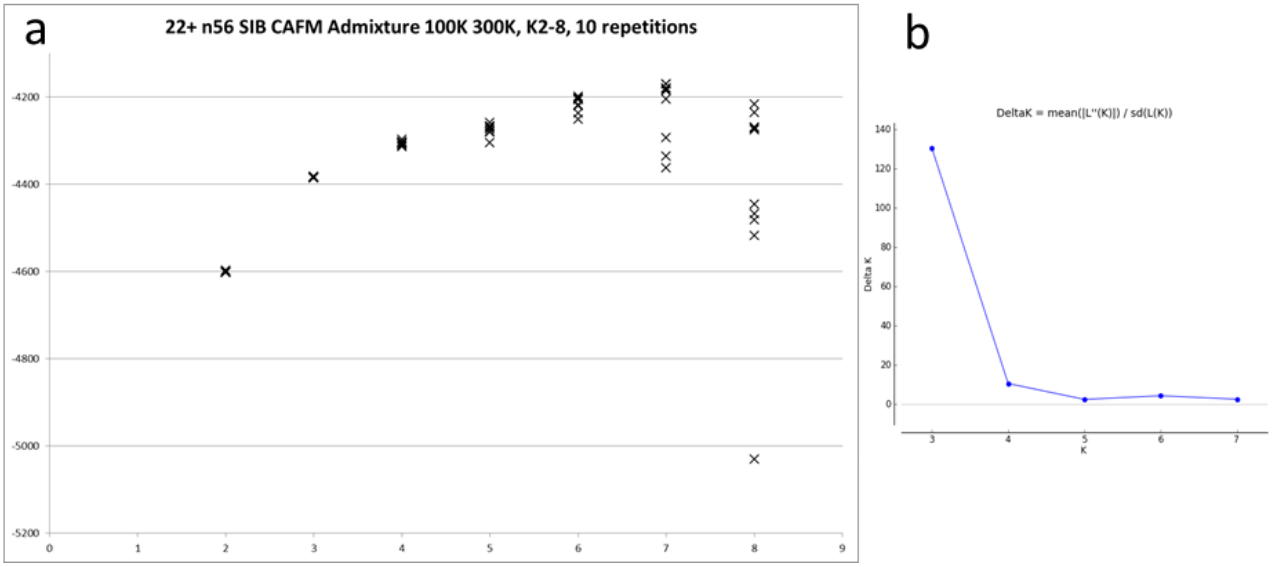
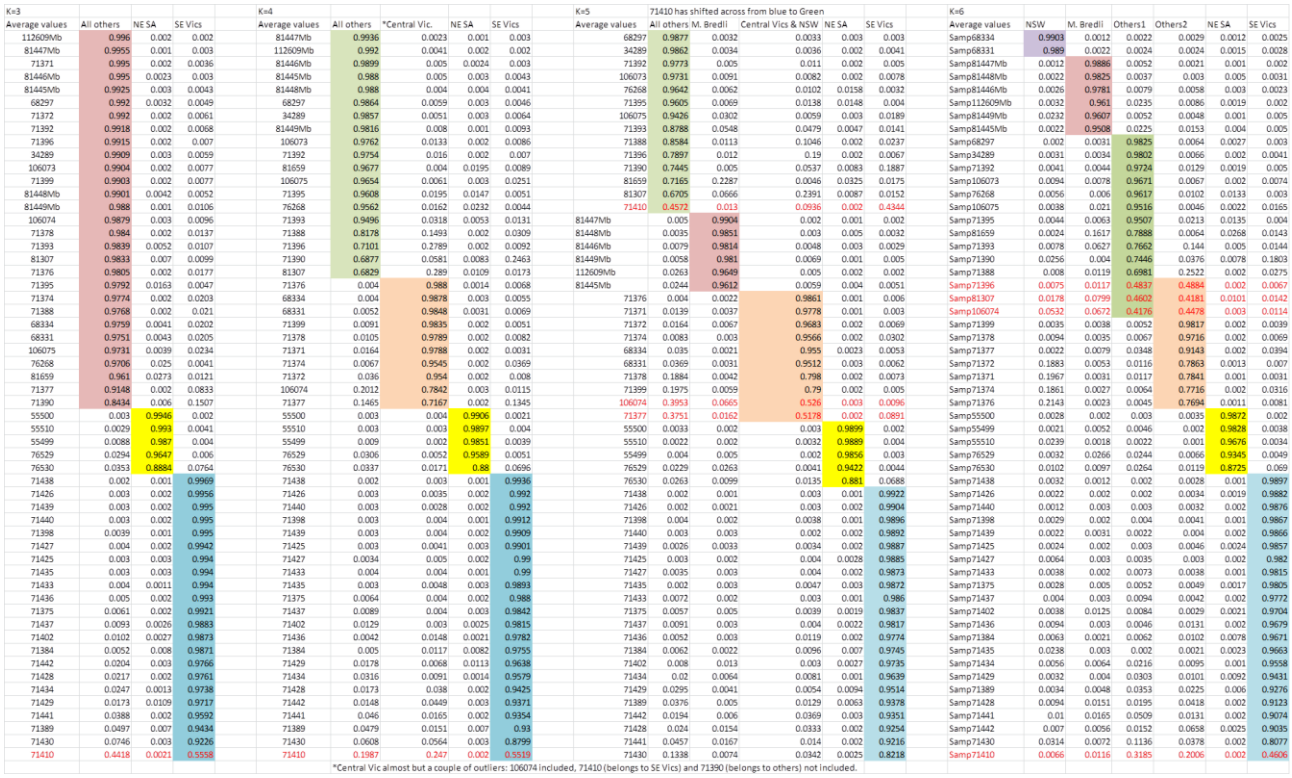


Figure A4.22: Log likelihood and Delta K plots of the n56 Southern, Internal and Bredli subpopulation using K2-K8.

Figure A4.23: Geographic locations of the populations seen from the K6 analysis (figure A4.24) with the two other populations grouped. The SE Vic population exhibiting HWE is circled.



K=3				K=4					K=5	71410 has shifted across from blue to Green					K=6						
Average values	All others	NE SA	SE Vics	Average values	All others	*Central Vic.	NE SA	SE Vics	Average values	All others	M. Bredli	Central Vics & NSW	NE SA	SE Vics	Average values	NSW	M. Bredli	Others1	Others2	NE SA	SE Vics
112609Mb	0.996	0.002	0.002	81447Mb	0.9936	0.0023	0.001	0.003	68297	0.9877	0.0032	0.0033	0.003	0.003	Samp68334	0.9903	0.0012	0.0022	0.0029	0.0012	0.0025
81447Mb	0.9955	0.001	0.003	112609Mb	0.992	0.0041	0.002	0.002	34289	0.9862	0.0034	0.0036	0.002	0.0041	Samp68331	0.989	0.0022	0.0024	0.0024	0.0015	0.0028
71371	0.995	0.002	0.0036	81446Mb	0.9899	0.005	0.0024	0.003	71392	0.9773	0.005	0.011	0.002	0.005	Samp81447Mb	0.0012	0.9886	0.0052	0.0021	0.001	0.002
81446Mb	0.995	0.0023	0.003	81445Mb	0.988	0.005	0.003	0.0043	106073	0.9731	0.0091	0.0082	0.002	0.0078	Samp81448Mb	0.0022	0.9825	0.0037	0.003	0.005	0.0031
81445Mb	0.9925	0.003	0.0043	81448Mb	0.988	0.004	0.004	0.0041	76268	0.9642	0.0062	0.0102	0.0158	0.0032	Samp81446Mb	0.0026	0.9781	0.0079	0.0058	0.003	0.0023
68297	0.992	0.0032	0.0049	68297	0.9864	0.0059	0.003	0.0046	71395	0.9605	0.0069	0.0138	0.0148	0.004	Samp112609Mb	0.0032	0.961	0.0235	0.0086	0.0019	0.002
71372	0.992	0.002	0.0061	34289	0.9857	0.0051	0.003	0.0064	106075	0.9426	0.0302	0.0059	0.003	0.0189	Samp81449Mb	0.0232	0.9607	0.0052	0.0048	0.001	0.005
71392	0.9918	0.002	0.0068	81449Mb	0.9816	0.008	0.001	0.0093	71393	0.8788	0.0548	0.0479	0.0047	0.0141	Samp81445Mb	0.0022	0.9508	0.0225	0.0153	0.004	0.005
71396	0.9915	0.002	0.007	106073	0.9762	0.0133	0.002	0.0086	71388	0.8584	0.0113	0.1046	0.002	0.0237	Samp68297	0.002	0.0031	0.9825	0.0064	0.0027	0.003
34289	0.9909	0.003	0.0059	71392	0.9754	0.016	0.002	0.007	71396	0.7897	0.012	0.19	0.002	0.0067	Samp34289	0.0031	0.0034	0.9802	0.0066	0.002	0.0041
106073	0.9904	0.002	0.0077	81659	0.9677	0.004	0.0195	0.0089	71390	0.7445	0.005	0.0537	0.0083	0.1887	Samp71392	0.0041	0.0044	0.9724	0.0129	0.0019	0.005
71399	0.9903	0.002	0.0077	106075	0.9654	0.0061	0.003	0.0251	81659	0.7165	0.2287	0.0046	0.0325	0.0175	Samp106073	0.0094	0.0078	0.9671	0.0067	0.002	0.0074
81448Mb	0.9901	0.0042	0.0052	71395	0.9608	0.0195	0.0147	0.0051	81307	0.6705	0.0666	0.2391	0.0087	0.0152	Samp76268	0.0056	0.006	0.9617	0.0102	0.0133	0.003
81449Mb	0.988	0.001	0.0106	76268	0.9562	0.0162	0.0232	0.0044	71410	0.4572	0.013	0.0936	0.002	0.4344	Samp106075	0.0038	0.021	0.9516	0.0046	0.0022	0.0165
106074	0.9879	0.003	0.0096	71393	0.9496	0.0318	0.0053	0.0131	81447Mb	0.005	0.9904	0.002	0.001	0.002	Samp71395	0.0044	0.0063	0.9507	0.0213	0.0135	0.004
71378	0.984	0.002	0.0137	71388	0.8178	0.1493	0.002	0.0309	81448Mb	0.0035	0.9851	0.003	0.005	0.0032	Samp81659	0.0024	0.1617	0.7888	0.0064	0.0268	0.0143
71393	0.9839	0.0052	0.0107	71396	0.7101	0.2789	0.002	0.0092	81446Mb	0.0079	0.9814	0.0048	0.003	0.0029	Samp71393	0.0078	0.0627	0.7662	0.144	0.005	0.0144
81307	0.9833	0.007	0.0099	71390	0.6877	0.0581	0.0083	0.2463	81449Mb	0.0058	0.981	0.0069	0.001	0.005	Samp71390	0.0256	0.004	0.7446	0.0376	0.0078	0.1803
71376	0.9805	0.002	0.0177	81307	0.6829	0.289	0.0109	0.0173	112609Mb	0.0263	0.9649	0.005	0.002	0.002	Samp71388	0.008	0.0119	0.6981	0.2522	0.002	0.0275
71395	0.9792	0.0163	0.0047	71376	0.004	0.988	0.0014	0.0068	81445Mb	0.0244	0.9612	0.0059	0.004	0.0051	Samp71396	0.0075	0.0117	0.4837	0.4884	0.002	0.0067
71374	0.9774	0.002	0.0203	68334	0.004	0.9878	0.003	0.0055	71376	0.004	0.0022	0.9861	0.001	0.006	Samp81307	0.0178	0.0799	0.4602	0.4181	0.0101	0.0142
71388	0.9768	0.002	0.021	68331	0.0052	0.9848	0.0031	0.0069	71371	0.0139	0.0037	0.9778	0.001	0.003	Samp106074	0.0532	0.0672	0.4176	0.4478	0.003	0.0114
68334	0.9759	0.0041	0.0202	71399	0.0091	0.9835	0.002	0.0051	71372	0.0164	0.0067	0.9683	0.002	0.0069	Samp71399	0.0035	0.0038	0.0052	0.9817	0.002	0.0039
68331	0.9751	0.0043	0.0205	71378	0.0105	0.9789	0.002	0.0082	71374	0.0083	0.003	0.9566	0.002	0.0302	Samp71378	0.0094	0.0035	0.0067	0.9716	0.002	0.0069
106075	0.9731	0.0039	0.0234	71371	0.0164	0.9788	0.002	0.0031	68334	0.035	0.0021	0.955	0.0023	0.0053	Samp71377	0.0022	0.0079	0.0348	0.9143	0.002	0.0394
76268	0.9706	0.025	0.0041	71374	0.0067	0.9545	0.002	0.0369	68331	0.0369	0.0031	0.9512	0.003	0.0062	Samp71372	0.1883	0.0053	0.0116	0.7863	0.0013	0.007
81659	0.961	0.0273	0.0121	71372	0.036	0.954	0.002	0.008	71378	0.1884	0.0042	0.798	0.002	0.0073	Samp71371	0.1967	0.0031	0.0117	0.7841	0.001	0.0031
71377	0.9148	0.002	0.0833	106074	0.2012	0.7842	0.003	0.0115	71399	0.1975	0.0059	0.79	0.002	0.005	Samp71374	0.1861	0.0027	0.0064	0.7716	0.002	0.0316
71390	0.8434	0.006	0.1507	71377	0.1465	0.7167	0.002	0.1345	106074	0.3953	0.0665	0.526	0.003	0.0096	Samp71376	0.2143	0.0023	0.0045	0.7694	0.0011	0.0081
55500	0.003	0.9946	0.002	55500	0.003	0.004	0.9906	0.0021	71377	0.3751	0.0162	0.5178	0.002	0.0891	Samp55500	0.0028	0.002	0.003	0.0035	0.9872	0.002
55510	0.0029	0.993	0.0041	55510	0.003	0.003	0.9897	0.004	55500	0.0033	0.002	0.003	0.9899	0.002	Samp55499	0.0021	0.0052	0.0046	0.002	0.9828	0.0038
55499	0.0088	0.987	0.004	55499	0.009	0.002	0.9851	0.0039	55510	0.0022	0.002	0.0032	0.9889	0.004	Samp55510	0.0239	0.0018	0.0022	0.001	0.9676	0.0034
76529	0.0294	0.9647	0.006	76529	0.0306	0.0052	0.9589	0.0051	55499	0.004	0.005	0.002	0.9856	0.003	Samp76529	0.0032	0.0266	0.0244	0.0066	0.9345	0.0049
76530	0.0353	0.8884	0.0764	76530	0.0337	0.0171	0.88	0.0696	76529	0.0229	0.0263	0.0041	0.9422	0.0044	Samp76530	0.0102	0.0097	0.0264	0.0119	0.8725	0.069
71438	0.002	0.001	0.9969	71438	0.002	0.003	0.001	0.9936	76530	0.0263	0.0099	0.0135	0.881	0.0688	Samp71438	0.0032	0.0012	0.002	0.0028	0.001	0.9897
71426	0.003	0.002	0.9956	71426	0.003	0.0035	0.002	0.992	71438	0.002	0.001	0.003	0.001	0.9922	Samp71426	0.0022	0.002	0.002	0.0034	0.0019	0.9882
71439	0.003	0.002	0.995	71440	0.003	0.0028	0.002	0.992	71426	0.002	0.0021	0.003	0.002	0.9904	Samp71440	0.0012	0.003	0.003	0.0032	0.002	0.9876
71440	0.003	0.002	0.995	71398	0.003	0.004	0.001	0.9912	71398	0.004	0.002	0.0038	0.001	0.9896	Samp71398	0.0029	0.002	0.004	0.0041	0.001	0.9867
71398	0.0039	0.001	0.995	71439	0.003	0.004	0.002	0.9909	71440	0.003	0.003	0.002	0.002	0.9892	Samp71439	0.0022	0.0031	0.0022	0.004	0.002	0.9866
71427	0.004	0.002	0.9942	71425	0.003	0.0041	0.003	0.9901	71439	0.0026	0.0033	0.0034	0.002	0.9887	Samp71425	0.0024	0.002	0.003	0.0046	0.0024	0.9857
71425	0.003	0.003	0.994	71427	0.0034	0.005	0.002	0.99	71425	0.003	0.002	0.004	0.0028	0.9885	Samp71427	0.0064	0.003	0.0035	0.003	0.002	0.982
71435	0.003	0.003	0.994	71433	0.004	0.004	0.001	0.99	71427	0.0035	0.003	0.004	0.002	0.9873	Samp71433	0.0038	0.002	0.0073	0.0038	0.001	0.9815
71433	0.004	0.0011	0.994	71435	0.003	0.0048	0.003	0.9893	71435	0.002	0.003	0.0047	0.003	0.9872	Samp71375	0.0028	0.005	0.0052	0.0049	0.0017	0.9805
71436	0.005	0.002	0.993	71375	0.0064	0.004	0.002	0.988	71433	0.0072	0.002	0.003	0.001	0.986	Samp71437	0.004	0.003	0.0094	0.0042	0.002	0.9772
71375	0.0061	0.002	0.9921	71437	0.0089	0.004	0.003	0.9842	71375	0.0057	0.005	0.0039	0.0019	0.9837	Samp71402	0.0038	0.0125	0.0084	0.0029	0.0021	0.9704
71437	0.0093	0.0026	0.9883	71402	0.0129	0.003	0.0025	0.9815	71437	0.0091	0.003	0.004	0.0022	0.9817	Samp71436	0.0094	0.003	0.0046	0.0131	0.002	0.9679
71402	0.0102	0.0027	0.9873	71436	0.0042	0.0148	0.0021	0.9782	71436	0.0052	0.003	0.0119	0.002	0.9774	Samp71384	0.0063	0.0021	0.0062	0.0102	0.0078	0.9671
71384	0.0052	0.008	0.9871	71384	0.005	0.0117	0.0082	0.9755	71384	0.0062	0.0022	0.0096	0.007	0.9745	Samp71435	0.0238	0.003	0.002	0.0021	0.0023	0.9663
71442	0.0204	0.003	0.9766	71429	0.0178	0.0068	0.0113	0.9638	71402	0.008	0.013	0.003	0.0027	0.9735	Samp71434	0.0056	0.0064	0.0216	0.0095	0.001	0.9558
71428	0.0217	0.002	0.9761	71434	0.0316	0.0091	0.0014	0.9579	71434	0.02	0.0064	0.0081	0.001	0.9639	Samp71429	0.0032	0.004	0.0303	0.0101	0.0092	0.9431
71434	0.0247	0.0013	0.9738	71428	0.0173	0.038	0.002	0.9425	71429	0.0295	0.0041	0.0054	0.0094	0.9514	Samp71389	0.0034	0.0048	0.0353	0.0225	0.006	0.9276
71429	0.0173	0.0109	0.9717	71442	0.0148	0.0449	0.003	0.9371	71389	0.0376	0.005	0.0129	0.0063	0.9378	Samp71428	0.0094	0.0151	0.0195	0.0418	0.002	0.9123
71441	0.0388	0.002	0.9592	71441	0.046	0.0165	0.002	0.9354	71442	0.0194	0.006	0.0369	0.003	0.9351	Samp71441	0.01	0.0165	0.0509	0.0131	0.002	0.9074
71389	0.0497	0.007	0.9434	71389	0.0479	0.0151	0.007	0.93	71428	0.024	0.0154	0.0333	0.002	0.9254	Samp71442	0.007	0.0056	0.0152	0.0658	0.0025	0.9035
71430	0.0746	0.003	0.9226	71430	0.0608	0.0564	0.003	0.8799	71441	0.0457	0.0167	0.014	0.002	0.9216	Samp71430	0.0314	0.0072	0.1136	0.0378	0.002	0.8077
71410	0.4418	0.0021	0.5558	71410	0.1987	0.247	0.002	0.5519	71430	0.1338	0.0074	0.0342	0.0025	0.8218	Samp71410	0.0066	0.0116	0.3185	0.2006	0.002	0.4606

*Central Vic almost but a couple of outliers: 106074 included, 71410 (belongs to SE Vics) and 71390 (belongs to others) not included.

Figure A4.24: Average Q-value matrices of the SIB population for K3 to K6.

At K4, the remaining population from K3 was split in two. This population showed similar patterns to the Msi Mainland population, where most individuals assigned to either population extremes, but a few individuals showed a continuum of admixed assignment, suggesting they are intermediates possibly exhibiting isolation by distance. Again, the patterns in this population are very similar to the Msi Mainland where a couple of small and one larger localised population separate out from the main dataset, however the remaining individuals are few in number and spread over a large geographic area, creating difficulties of assigning this group as a population when they do not conform to HWE due to limited samples per locality and encompassing a large geographic range. K5 separates the *Morelia bredli* of central Australia into a separate group of six individuals. At K6, two individuals from the NSW Riverina separate further from the main grouping and some of the less strongly assigned individuals switch between the two main populations, but show continued admixture to both. The two K6 populations that are potentially one continuous population involving isolation by distance gave a positive correlation (R^2 = 0.1918), suggesting the presence of weak to moderate distance related geographic structuring.

Like the Msi Mainland population, this population contained a couple of very small yet genetically distinct populations (too small to give accurate pairwise FST calculations), a larger and geographically confined population which adhered to HWE and a remaining continuous population with few individuals spread over large geographic area.

A4.4 Populations deviating from HWE

Table A4.3: Sub-populations deviating significantly from Hardy-Weinberg Equilibrium expectations

n50 SI — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.1744	0.0007	0.1093	0.1285
MS25	0.092	0.001	0.093	0.016
MsF33	0.0002	0.0001	0.1583	0.1442
MsF31	0	0	0.4503	0.5166
MsF9	0.6029	0.0007	0.0779	0.0732
MsF16	0.0005	0	0.3471	0.3928
MsF3	0.0015	0	0.2103	0.3093
MsF17	0.0101	0.0006	0.2074	0.0965
MsF18	0.0071	0.0001	0.1622	0.1167
MsF28si	0.2566	0.0003	0.2159	0.1232
MsF22	0.0076	0.0002	0.1074	0.2165
MS3	0.1123	0.0014	0.0388	0.0252
MS4h	0.0293	0.0004	0.1097	0.1479
MsF27s	0	0	0.3358	0.1307
MsF8	0.0388	0.0002	0.1958	0.2257
MsF26	0.3127	0.0007	0.1568	0.1202
MsF4	0.0005	0	0.1822	0.1419
MS27	0.3658	0.0003	0.1707	0.0832
MsF24	0.0601	0.0003	0.1165	0.2656
MS9	0.0358	0.0002	0.1434	0.1671
MsF2	0	0	0.2008	0.2977
MS17	0.0424	0.0006	0.1439	0.0992
MsF15	0	0	0.4297	0.3598
MS13	0.0028	0.0001	0.2869	0.2407

All (Fisher's method):

Chi2	Infinity	
Df	48	
Prob	High. Sign.	

n29 SIB other — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.5463	0.0011	0.0939	0.1405
MS25	0.1204	0.0014	0.1137	0.0264
MsF33	0	0	0.1812	0.1196
MsF31	0.0007	0	0.3865	0.2881
MsF9	0.0927	0.0004	-0.048	0.0337
MsF16	0.0004	0	0.4026	0.4324
MsF3	0.0086	0.0001	0.1527	0.1873
MsF17	0.4532	0.0028	0.0686	0.0539
MsF18	0.0033	0.0001	0.1276	0.0615
MsF28si	0.2575	0.0002	0.1706	0.1383
MsF22	0.2527	0.0008	0.1511	0.0847
MS3	0.0001	0	0.1123	0.0883
MS4h	0.6055	0.0012	-0.0125	-0.0207
MsF27s	0.0001	0	0.2961	0.1291
MsF8	0.302	0.0005	0.0937	0.0391
MsF26	0.0047	0.0001	0.3458	0.248
MsF4	0.4491	0.0013	0.0628	0.0491
MS27	0.0004	0	0.2827	0.2276
MsF24	0.3308	0.0004	0.2131	0.1077
MS9	0.0457	0.0003	0.1216	0.1456
MsF2	0.0017	0	0.1648	0.2295
MS17	0.0281	0.0003	0.2096	0.1134
MsF15	0.0456	0.0004	0.1369	0.0779
MS13	0.0021	0.0001	0.3364	0.2741

All (Fisher's method):

Chi2	209.4686	
Df	48	(Central Vic, NSW & Bredli)
Prob	High. Sign	

n23 SI other — Not to include

Fis — Still significant

locus	P-val	S.E.	W&C	R&H
MsF5s	0.5909	0.0012	0.0323	0.129
MS25	0.2156	0.0018	0.0671	0.0014
MsF33	0.0002	0	0.2273	0.1671
MsF31	0.0003	0	0.4344	0.375
MsF9	0.1301	0.0005	-0.0299	0.043
MsF16	0.0048	0.0001	0.4307	0.4589
MsF3	0.0046	0.0001	0.2226	0.2367
MsF17	0.545	0.0027	0.0402	0.0532
MsF18	0.0078	0.0001	0.1518	0.084
MsF28si	0.3633	0.0003	0.2653	0.161
MsF22	0.4319	0.0011	0.0059	0.0203
MS3	0.0085	0.0002	0.0742	0.0648
MS4h	0.4041	0.001	-0.0825	-0.0421
MsF27s	0.0006	0.0001	0.2927	0.1022
MsF8	0.2986	0.0006	0.1492	0.067
MsF26	0.0914	0.0003	0.3136	0.1504
MsF4	0.2043	0.0011	0.101	0.1008
MS27	0.7201	0.0002	0.1619	0.0832
MsF24	1	0	-0.1	-0.0385
MS9	0.0928	0.0004	0.2103	0.2061
MsF2	0.002	0.0001	0.1669	0.2304
MS17	0.022	0.0002	0.2072	0.1402
MsF15	0.0307	0.0003	0.1655	0.0841
MS13	0.0076	0.0001	0.3173	0.2604

All (Fisher's method):

Chi2	156.7708	
Df	48	(Central Vic & NSW)
Prob	0	

n50 NEC all Qld — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0631	0.0011	0.0659	0.0793
MS25	0.2157	0.0039	-0.0004	0.0144
MsF33	0.202	0.0045	0.0337	0.0203
MsF31	0.0302	0.0008	0.1651	0.1175
MsF9	0.0526	0.0004	0.0837	0.0598
MsF16	0.5886	0.0008	0.0699	0.0229
MsF3	0.0283	0.0003	0.1685	0.1654
MsF17	0.211	0.0012	-0.0313	-0.033
MsF18	0.8465	0.0013	-0.0216	-0.0029
MsF28si	0.1743	0.002	0.1222	0.0179
MsF22	0.1427	0.0006	0.1073	0.0773
MS3	0.0424	0.0022	0.0549	0.0261
MS4h	0.0259	0.0003	-0.0222	0.0753
MsF27s	0.3901	0.0033	-0.0433	-0.0133
MsF8	0.762	0.0009	0.0301	0.0404
MsF26	0.3599	0.0014	0.1023	0.042
MsF4	0.0835	0.0015	0.0258	0.0687
MS27	0.2585	0.0019	0.0196	0.0243
MsF24	0.011	0.0001	0.187	0.2444
MS9	0.0085	0.0002	0.1538	0.1118
MsF2	0.2941	0.0026	-0.0213	-0.0104
MS17	0.0128	0.0009	0.1348	0.0889
MsF15	0.3903	0.0017	0.0189	0.0222
MS13	0.6136	0.0031	0.006	0.0327

All (Fisher's method):

Chi2	102.0333	
Df	48	
Prob	0	

n44 SEC - 6 northerners — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.7929	0.0012	0.0611	0.0208
MS25	0.7716	0.0043	-0.006	-0.0066
MsF33	0.0514	0.002	0.0656	0.0313
MsF31	0.0239	0.0005	0.0444	0.0497
MsF9	0.7959	0.0007	-0.0368	0.006
MsF16	0.0462	0.0005	0.1587	0.0815
MsF3	0.05	0.0004	0.1829	0.2747
MsF17	0.4949	0.0029	0.021	0.0118
MsF18	0.1572	0.001	0.1147	0.0941
MsF28si	0.1367	0.0005	0.0796	0.0274
MsF22	0.1321	0.0006	0.0517	0.1236
MS3	0.3361	0.0055	0.0147	0.0185
MS4h	0.7087	0.001	-0.0415	-0.0305
MsF27s	0.3711	0.0026	0.0327	0.0053
MsF8	0.0139	0.0002	0.2145	0.1689
MsF26	0.7711	0.0015	0.0072	0.0299
MsF4	0.8687	0.0025	0.0055	-0.0023
MS27	0.0288	0.0007	0.0622	0.0255
MsF24	0.3224	0.0006	0.1169	0.0666
MS9	0.0228	0.0003	0.2085	0.1539
MsF2	0.0575	0.0004	0.0787	0.0514
MS17	0.6531	0.0053	-0.0255	-0.0143
MsF15	0.2663	0.0025	0.0241	0.0303
MS13	0.0212	0.001	0.1098	0.0892

All (Fisher's method):

Chi2 :	88.1651	
Df :	48	
Prob :	0.0004	

n50 SEC — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.9165	0.0008	0.0529	0.0111
MS25	0.907	0.0029	-0.0109	-0.0092
MsF33	0.0454	0.0018	0.0537	0.0229
MsF31	0.0262	0.0005	0.0579	0.0537
MsF9	0.736	0.0007	-0.0522	-0.0117
MsF16	0.0734	0.0006	0.1452	0.0896
MsF3	0.0171	0.0002	0.196	0.2828
MsF17	0.6185	0.0028	0.0198	0.0061
MsF18	0.3675	0.0017	0.1081	0.077
MsF28si	0.2141	0.0009	0.0968	0.0287
MsF22	0.0083	0.0001	0.0904	0.1263
MS3	0.5204	0.0059	0.0093	0.0168
MS4h	0.5111	0.0015	-0.0547	-0.0279
MsF27s	0.188	0.002	0.0324	0.0024
MsF8	0.0209	0.0002	0.1665	0.1192
MsF26	0.8888	0.0012	0.0174	0.0278
MsF4	0.9066	0.002	0.0181	0.0043
MS27	0.0173	0.0006	0.056	0.0127
MsF24	0.4424	0.0009	0.0984	0.055
MS9	0.1659	0.0009	0.1634	0.1132
MsF2	0.3404	0.0012	0.0805	0.0437
MS17	0.722	0.0051	-0.0242	-0.0128
MsF15	0.5943	0.0029	0.0085	0.0256
MS13	0.0111	0.0008	0.115	0.0821

All (Fisher's method):

Chi2	83.7976	
Df	48	
Prob	0.0011	

n18 NEC other subset — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.3435	0.001	0.0134	-0.0064
MS25	0.7454	0.0019	-0.1107	-0.0441
MsF33	0.147	0.0022	0.0103	0.0103
MsF31	0.0001	0	0.2892	0.1691
MsF9	0.0776	0.0003	0.1451	0.0641
MsF16	0.7091	0.0002	0.0391	-0.0174
MsF3	0.0389	0.0002	0.0998	0.0912
MsF17	0.1832	0.0003	-0.0303	-0.0696
MsF18	0.5901	0.001	-0.0815	-0.0564
MsF28si	0.2684	0.0012	0.1614	0.0294
MsF22	0.0847	0.0002	0.0943	0.1281
MS3	0.023	0.001	0.1327	0.0691
MS4h	0.6552	0.0004	-0.1748	-0.1122
MsF27s	0.48	0.0008	-0.1079	-0.0556
MsF8	0.157	0.0005	0.184	0.109
MsF26	0.4223	0.0009	-0.0205	-0.0249
MsF4	0.0051	0.0002	0.0778	0.0637
MS27	0.1249	0.0012	0.0811	0.0628
MsF24	0.1447	0.0004	0.2225	0.2843
MS9	0.0019	0.0001	0.3009	0.2663
MsF2	0.7583	0.0009	0.023	-0.0011
MS17	0.0064	0.0003	0.3057	0.2617
MsF15	0.6835	0.0006	0.0192	0.0126
MS13	0.2183	0.0022	0.0034	0.0294

All (Fisher's method):

Chi2	111.4409	
Df	48	
Prob	0	

n30 SEC 2nd diff. N subset — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.7878	0.0015	0.0176	-0.0017
MS25	0.5139	0.0053	0.0024	-0.0055
MsF33	0.4041	0.0044	0.0299	0.0054
MsF31	0.4313	0.002	0.0104	-0.004
MsF9	0.854	0.0005	-0.0407	-0.0366
MsF16	0.081	0.0006	0.1714	0.1141
MsF3	0.0675	0.0004	0.2291	0.2363
MsF17	0.3744	0.0017	-0.0149	-0.0169
MsF18	0.5686	0.002	0.1008	0.0848
MsF28si	0.2817	0.001	0.0459	0.0077
MsF22	0.0125	0.0001	0.08	0.1424
MS3	0.7706	0.0046	0.0077	-0.0027
MS4h	0.1318	0.0008	0.0038	0.0057
MsF27s	0.3728	0.0016	0.0127	-0.0118
MsF8	0.3607	0.0007	0.0787	-0.0002
MsF26	0.7402	0.0015	0.0143	0.0161
MsF4	0.9821	0.0006	-0.0163	-0.0174
MS27	0.0014	0.0001	-0.0087	-0.01
MsF24	0.1398	0.0005	0.2043	0.1461
MS9	0.3482	0.0011	0.1856	0.139
MsF2	0.5341	0.0012	0.1902	0.1071
MS17	0.7492	0.0045	-0.0259	-0.0145
MsF15	0.8563	0.0015	0.0243	0.0383
MS13	0.0196	0.001	0.1262	0.018

All (Fisher's method):

Chi2	68.8394	
Df	48	
Prob	0.0259	

n16 Msi SW WA — Adheres but include larger subset

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.8445	0.0005	-0.0465	-0.0528
MS25	0.0925	0.0007	0.0557	0.041
MsF33	0.3998	0.0014	-0.0613	-0.0405
MsF31	No	information.		
MsF9	0.7368	0.0004	-0.0909	-0.0631
MsF16	0.4876	0.0007	-0.2129	-0.1124
MsF3	0.0981	0.0004	0.1768	0.0759
MsF17	0.8975	0.0003	0.0501	0.0477
MsF18	0.4154	0.0008	0.0955	0.0875
MsF28si	0.1185	0.0004	0.1228	0.0936
MsF22	0.3068	0.0006	0.1329	0.1263
MS3	0.2972	0.0018	0.0646	0.0863
MS4h	0.9372	0.0003	-0.1139	-0.0815
MsF27s	0.0107	0.0002	0.2308	0.2056
MsF8	0.9283	0.0002	-0.1573	-0.1217
MsF26	0.8275	0.0005	0.0826	0.0171
MsF4	0.7684	0.0015	-0.0274	0.0979
MS27	1	0	0.0875	0.0506
MsF24	No	information.		
MS9	0.7011	0.0009	-0.0526	-0.0489
MsF2	0.4049	0.0005	-0.1811	-0.132
MS17	0.9381	0.0007	-0.0441	-0.0312
MsF15	0.5067	0.0007	-0.0694	-0.061
MS13	0.4452	0.0009	0.0265	-0.0036

All (Fisher's method):

Chi2	40.5407	
Df	44	
Prob	0.6207	

Table A4.3 (cont.): Sub-populations deviating significantly from Hardy-Weinberg Equilibrium expectations

n36 Msi ML — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0019	0.0002	0.1205	0.0578
MS25	0.0042	0.0002	0.0573	0.0535
MsF33	No information.			
MsF31	0.1567	0.0005	0.0854	0.0504
MsF9	0.5511	0.001	-0.0568	-0.055
MsF16	0.0001	0	0.3636	0.274
MsF3	0.1852	0.0009	0.1387	0.1107
MsF17	0.5358	0.0009	0.0776	0.0567
MsF18	0.0121	0.0002	0.3223	0.1677
MsF28si	0.0772	0.0005	0.18	0.1466
MsF22	0.0005	0.0001	0.2019	0.1813
MS3	0.3901	0.0013	0.0343	0.0079
MS4h	0	0	0.3465	0.2259
MsF27s	0.3378	0.0004	0.0006	-0.0163
MsF8	0.0287	0.0003	0.1549	0.0659
MsF26	0.0238	0.0009	0.1788	0.0857
MsF4	0	0	0.4979	0.6676
MS27	0.0208	0	0.5354	0.5471
MsF24	0.364	0.0013	0.0581	0.1198
MS9	0.2344	0.0007	0.0448	0.0232
MsF2	0.4642	0.0033	0.0448	0.0907
MS17	0.0351	0.0003	0.145	0.0888
MsF15	0.0517	0.0013	0.1241	0.0625
MS13				

All (Fisher's method):

Chi2	Infinity
Df	46
Prob	High. Sign.

n29 Msi ML excl extremes — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.2221	0.0008	0.2241	0.0767
MS25	0.1659	0.0019	0.0318	0.0064
MsF33	0.07	0.001	-0.0385	-0.0306
MsF31	No information.			
MsF9	0.28	0.0006	0.0405	0.0086
MsF16	0.681	0.0009	-0.0646	-0.0593
MsF3	0.0069	0.0001	0.3333	0.1845
MsF17	0.5469	0.0009	0.0874	0.1024
MsF18	0.4842	0.0009	0.0633	0.0479
MsF28si	0.1244	0.0006	0.1805	0.1026
MsF22	0.2589	0.0008	0.0629	0.0671
MS3	0.0372	0.0007	0.1193	0.1513
MS4h	0.7588	0.0007	0.0204	0.0095
MsF27s	0.0101	0.0003	0.1702	0.1055
MsF8	0.5317	0.0004	-0.0612	-0.0176
MsF26	0.215	0.0006	0.0924	0.0385
MsF4	0.5812	0.0039	0.0085	0.0221
MS27	0.0163	0.0001	0.2607	0.4111
MsF24	0.0048	0	0.7879	0.8129
MS9	0.5414	0.0012	0.1005	0.1276
MsF2	0.3448	0.0008	-0.0471	-0.0257
MS17	0.5432	0.0028	0.0421	0.0527
MsF15	0.3089	0.0008	-0.027	-0.0384
MS13	0.3331	0.0022	0.0674	0.0742

All (Fisher's method):

Chi2	84.4894
Df	46
Prob	0.0005

n12 Msi ML CSA/WA — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0183	0.0002	0.4527	0.2267
MS25	0.4156	0.0019	0.0163	0.0164
MsF33	0.1852	0.0013	0.0041	-0.0116
MsF31	No information.			
MsF9	0.2164	0.0005	0.1538	0.0843
MsF16	0.1846	0.0005	0.0435	-0.0126
MsF3	0.0341	0.0002	0.4107	0.2902
MsF17	0.176	0.0008	0.0476	0.0392
MsF18	0.8609	0.0004	-0.105	-0.079
MsF28si	0.1245	0.0005	0.1659	0.0786
MsF22	0.7115	0.0006	-0.1469	-0.0852
MS3	0.269	0.001	0.0947	0.1229
MS4h	0.5065	0.0008	0.1837	0.135
MsF27s	0.7717	0.0008	0.0336	0.0082
MsF8	0.2674	0.0004	0.0388	0.1035
MsF26	0.0245	0.0002	0	0.0164
MsF4	1	0	-0.082	-0.0417
MS27	0.2894	0.0001	0.3529	0.374
MsF24	0.031	0	0.766	0.8268
MS9	1	0	0.1127	0.0737
MsF2	0.2313	0.0004	0.1161	0.0594
MS17	0.6816	0.0017	0.032	0.0164
MsF15	0.9206	0.0004	0	-0.0134
MS13	0.7889	0.0011	-0.0342	0.0106

All (Fisher's method):

Chi2	63.2681
Df	46
Prob	0.0463

n47 SEC subset — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.7891	0.0012	0.07	0.0207
MS25	0.8195	0.004	-0.0077	-0.0074
MsF33	0.0605	0.0022	0.0616	0.0288
MsF31	0.0172	0.0004	0.044	0.0469
MsF9	0.7799	0.0007	-0.0406	-0.0052
MsF16	0.0486	0.0004	0.1462	0.0778
MsF3	0.0868	0.0005	0.1721	0.2514
MsF17	0.3548	0.0029	0.0349	0.0161
MsF18	0.2095	0.0011	0.1015	0.092
MsF28si	0.1394	0.0005	0.093	0.0348
MsF22	0.0831	0.0005	0.0395	0.118
MS3	0.3504	0.0055	0.0142	0.018
MS4h	0.6718	0.0012	-0.0456	-0.0288
MsF27s	0.2852	0.0025	0.0435	0.0065
MsF8	0.0155	0.0002	0.1979	0.1483
MsF26	0.6813	0.0018	0.0136	0.0325
MsF4	0.906	0.002	0.0033	-0.0028
MS27	0.0207	0.0006	0.059	0.0194
MsF24	0.3289	0.0006	0.0964	0.0575
MS9	0.0506	0.0004	0.1908	0.1423
MsF2	0.108	0.0007	0.0724	0.0456
MS17	0.6806	0.0052	-0.0245	-0.0142
MsF15	0.2755	0.0025	0.0188	0.0298
MS13	0.0186	0.001	0.1061	0.0827

All (Fisher's method):

Chi2	86.573
Df	48
Prob	0.0005

n39 SEC smaller subset — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.6065	0.0014	0.088	0.0326
MS25	0.691	0.0048	-0.0035	-0.0042
MsF33	0.0703	0.0024	0.0784	0.0441
MsF31	0.0983	0.0011	0.0593	0.0596
MsF9	0.728	0.0008	-0.0753	-0.024
MsF16	0.0822	0.0006	0.1507	0.0802
MsF3	0.036	0.0003	0.1632	0.2396
MsF17	0.4844	0.003	0.039	0.0291
MsF18	0.245	0.0012	0.1159	0.0926
MsF28si	0.3645	0.0007	0.0788	0.0312
MsF22	0.0976	0.0005	-0.0008	0.0997
MS3	0.2622	0.005	0.0188	0.0208
MS4h	0.8973	0.0005	-0.0308	-0.026
MsF27s	0.0711	0.0014	0.0754	0.0194
MsF8	0.0228	0.0002	0.2491	0.1674
MsF26	0.4954	0.0022	0.0477	0.0481
MsF4	0.7274	0.0036	0.0123	0.0023
MS27	0.5219	0.0023	0.0819	0.0248
MsF24	0.4946	0.0007	0.1301	0.0763
MS9	0.0926	0.0006	0.1969	0.1459
MsF2	0.046	0.0004	0.0768	0.0438
MS17	0.8391	0.0039	-0.0245	-0.0142
MsF15	0.1588	0.0019	0.0423	0.0372
MS13	0.0181	0.0009	0.1316	0.1016

All (Fisher's method):

Chi2	78.3715
Df	48
Prob	0.0037

n35 SEC different N subset — Not to include

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.8714	0.0011	0.063	0.0183
MS25	0.6156	0.005	-0.0009	-0.0073
MsF33	0.5622	0.0045	0.0209	0.0018
MsF31	0.2446	0.0016	-0.0022	-0.0095
MsF9	0.8331	0.0006	-0.0679	-0.0492
MsF16	0.0817	0.0007	0.1325	0.0915
MsF3	0.0996	0.0005	0.1792	0.1954
MsF17	0.6362	0.0016	-0.0048	-0.0034
MsF18	0.5776	0.0018	0.0768	0.0491
MsF28si	0.3245	0.001	0.0499	-0.0031
MsF22	0.0127	0.0001	0.0585	0.1161
MS3	0.8492	0.0039	0.0017	-0.0036
MS4h	0.3313	0.0012	-0.0277	-0.0122
MsF27s	0.1037	0.001	0.0395	-0.0061
MsF8	0.0351	0.0003	0.1591	0.0724
MsF26	0.5296	0.0021	0.0115	0.0156
MsF4	0.8742	0.0018	0.0075	-0.0014
MS27	0.005	0.0003	0.0401	0.0143
MsF24	0.3502	0.0008	0.137	0.0865
MS9	0.5896	0.0013	0.1641	0.1144
MsF2	0.3884	0.0013	0.1492	0.0858
MS17	0.8367	0.0038	-0.0272	-0.0151
MsF15	0.8128	0.0018	0.0032	0.0326
MS13	0.034	0.0014	0.0946	0.0102

All (Fisher's method):

Chi2	66.6696
Df	48
Prob	0.0385

n33 NEC large subset — Adheres but include larger subset

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.0803	0.0012	0.0805	0.0829
MS25	0.2953	0.0042	0.0337	0.0216
MsF33	0.635	0.0048	0.0307	0.0292
MsF31	0.4045	0.0019	0.1125	0.0523
MsF9	0.0499	0.0004	0.0054	0.0247
MsF16	0.6622	0.0008	0.0783	0.0417
MsF3	0.099	0.0005	0.1834	0.1892
MsF17	0.0375	0.0005	-0.0486	-0.0408
MsF18	0.7696	0.0016	-0.0054	0.0085
MsF28si	0.5566	0.0018	0.0758	0.0134
MsF22	0.8073	0.0006	0.0879	0.0546
MS3	0.8263	0.0042	0.0031	-0.004
MS4h	0.4318	0.001	-0.0107	0.0916
MsF27s	0.6811	0.0024	-0.0183	-0.0079
MsF8	0.9493	0.0002	-0.0756	-0.0647
MsF26	0.3664	0.0011	0.1674	0.0829
MsF4	0.3118	0.0026	-0.0242	0.0248
MS27	0.4306	0.0021	-0.0393	-0.0205
MsF24	0.2427	0.0006	0.1382	0.0993
MS9	0.2504	0.0012	0.0659	0.0146
MsF2	0.4244	0.0024	-0.0207	-0.0107
MS17	0.7394	0.0038	0.0267	0.0065
MsF15	0.3934	0.0018	0.0016	-0.0065
MS13	0.5966	0.0032	0.002	0.031

All (Fisher's method):

Chi2	50.1127
Df	48
Prob	0.3896

n21 NEC subset — Not signifcant but larger pop seen

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.4957	0.0017	0.0503	0.0847
MS25	0.4928	0.0039	0.0649	0.0348
MsF33	1	0	-0.0298	-0.0183
MsF31	0.7978	0.0014	0.0734	0.0372
MsF9	0.2882	0.0007	-0.1128	-0.0553
MsF16	0.4898	0.0008	0.0789	0.1123
MsF3	0.0536	0.0005	0.2443	0.2128
MsF17	0.0572	0.0004	-0.0271	-0.0391
MsF18	0.6023	0.0017	0.0257	0.0165
MsF28si	0.9402	0.0004	-0.0099	-0.0212
MsF22	0.4011	0.0007	0.0433	0.025
MS3	0.8046	0.0038	0.0069	-0.0019
MS4h	0.4295	0.0012	-0.0502	0.0523
MsF27s	0.9136	0.001	0.0476	0.0127
MsF8	0.9612	0.0001	-0.1014	-0.1006
MsF26	0.3598	0.0008	0.2561	0.184
MsF4	0.5607	0.0027	-0.0418	-0.0236
MS27	0.5938	0.0016	-0.0106	-0.0092
MsF24	0.4783	0.0006	0.0698	0.0416
MS9	0.0927	0.0006	0.1409	0.0593
MsF2	0.4776	0.0021	-0.0471	-0.0189
MS17	0.415	0.0039	0.0606	0.025
MsF15	0.1164	0.001	0.0204	0.0123
MS13	0.597	0.003	-0.0243	0.025

All (Fisher's method):

Chi2	43.1586
Df	48
Prob	0.6712

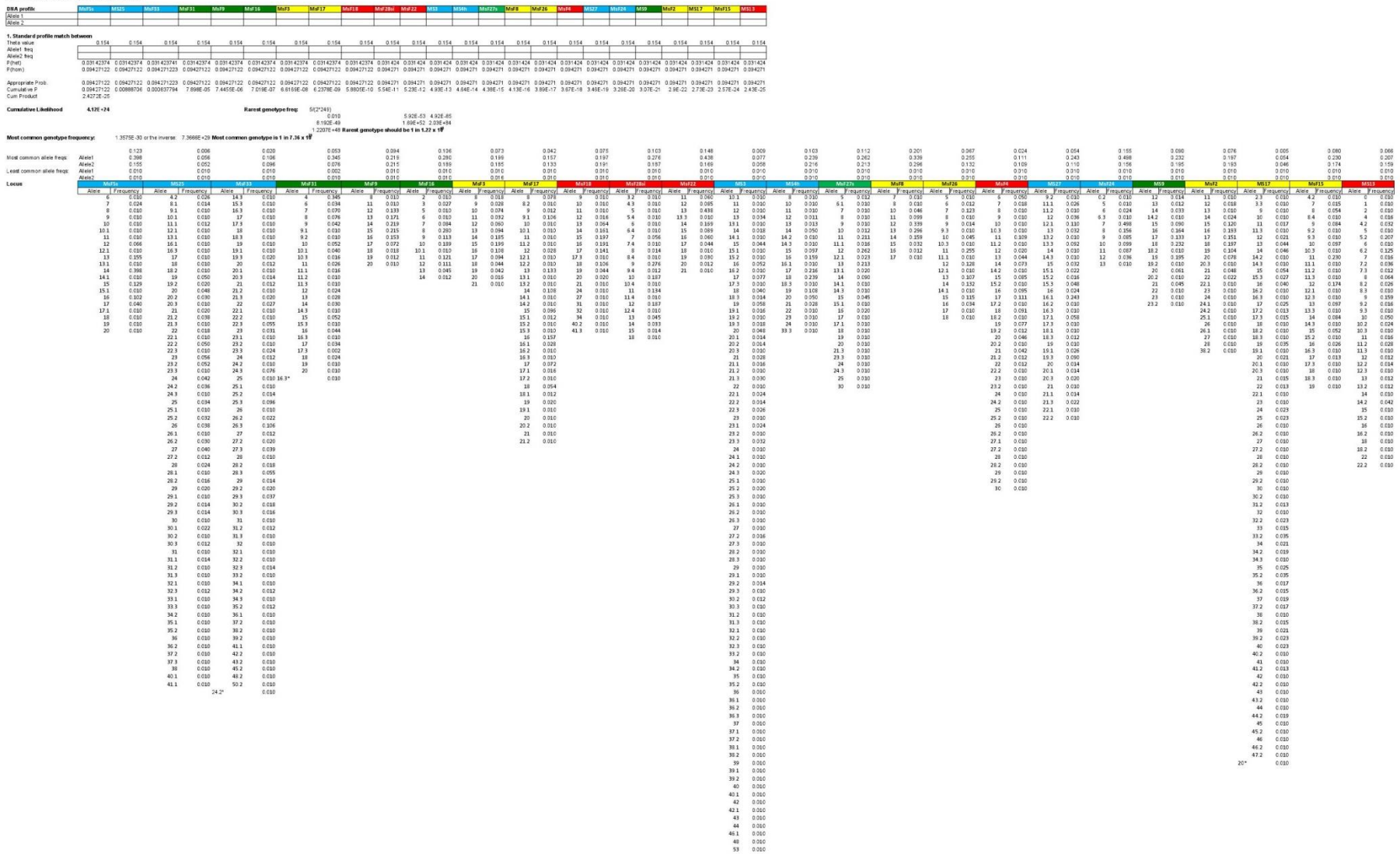
n27 NEC larger subset — Not signifcant but larger pop seen

Fis

locus	P-val	S.E.	W&C	R&H
MsF5s	0.2227	0.0018	0.0824	0.0817
MS25	0.3162	0.0043	0.0488	0.0263
MsF33	0.3749	0.0047	0.0499	0.0346
MsF31	0.6025	0.0018	0.1034	0.0417
MsF9	0.1294	0.0008	-0.0853	-0.039
MsF16	0.4789	0.0008	0.0948	0.0668
MsF3	0.057	0.0004	0.2115	0.2174
MsF17	0.0131	0.0002	-0.0205	-0.0374
MsF18	0.6263	0.0018	0.0164	0.0172
MsF28si	0.8381	0.001	0.0066	-0.0071
MsF22	0.5245	0.0008	0.0933	0.064
MS3	0.8124	0.0042	0.0048	-0.0028
MS4h	0.3508	0.001	-0.0136	0.0974
MsF27s	0.8511	0.0015	-0.0547	-0.0127
MsF8	0.9274	0.0002	-0.0506	-0.0617
MsF26	0.475	0.0009	0.2054	0.116
MsF4	0.4247	0.0026	-0.0647	-0.0331
MS27	0.8136	0.0011	-0.0492	-0.0253
MsF24	0.2238	0.0005	0.0964	0.0702
MS9	0.1663	0.0009	0.0722	0.0251
MsF2	0.421	0.0022	-0.026	-0.0169
MS17	0.5998	0.0043	0.0385	0.0088
MsF15	0.2098	0.0012	0.0229	0.0071
MS13	0.5044	0.0032	-0.0242	0.0192

All (Fisher's method):

Chi2	51.2534
Df	48
Prob	0.3473



Overarching population Match Probability Calc. Sheet

DNA profile	MsF5s	MS25	MsF33	MsF31	MsF9	MsF16	MsF3	MsF17	MsF18	MsF28d	MsF22	MS3	MS4h	MsF27s	MsF8	MsF26	MsF24	MS9	MsF2	MS17	MsF15	MS13
Allele 1																						
Allele 2																						

1. Standard profile match between

Theta value	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154	0.154
Allele1 freq																								
Allele2 freq																								
P(het)	0.03142374	0.03142374	0.03142374	0.03142374	0.03142374	0.03142374	0.03142374	0.03142374	0.03142374	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424	0.031424
P(hom)	0.09427122	0.09427122	0.09427123	0.09427122	0.09427122	0.09427122	0.09427122	0.09427122	0.09427122	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271
Appropriate Prob.	0.09427122	0.09427122	0.09427123	0.09427122	0.09427122	0.09427122	0.09427122	0.09427122	0.09427122	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271	0.094271
Cumulative P	0.09427122	0.00888706	0.000837794	7.898E-05	7.445E-06	7.019E-07	6.616E-08	6.237E-09	5.880E-10	5.54E-11	5.23E-12	4.93E-13	4.64E-14	4.38E-15	4.13E-16	3.89E-17	3.67E-18	3.46E-19	3.26E-20	3.07E-21	2.9E-22	2.73E-23	2.57E-24	2.43E-25
Cum Product	2.427E-25																							

Cumulative Likelihood **4.12E+24** **Rarest genotype freq:** 5/(2*249) = 0.010 — 5.92E-53, 4.92E-85, 8.192E-49, 1.69E+52, 2.03E+84

1.2207E+48 Rarest genotype should be 1 in 1.22 x 10^48

Most common genotype frequency: 1.3575E-30 or the inverse: 7.3666E+29 **Most common genotype is 1 in 7.36 x 10^29**

		0.123		0.006		0.020		0.053		0.094		0.136		0.073		0.042		0.075		0.103		0.148		0.009	
Most common allele freqs:	Allele1	0.398		0.056		0.106		0.345		0.218		0.280		0.189		0.157		0.197		0.276		0.438		0.077	
	Allele2	0.155		0.052		0.096		0.076		0.215		0.189		0.185		0.133		0.191		0.197		0.166		0.068	
Least common allele freqs:	Allele1	0.010		0.010		0.010		0.002		0.010		0.010		0.010		0.010		0.010		0.010		0.010		0.010	
	Allele2	0.010		0.010		0.010		0.010		0.010		0.010		0.016		0.010		0.010		0.010		0.010		0.010	

Locus	MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28d		MsF22		MS3		MS4h		MsF27s		MsF8		MsF26		MsF4		MsF27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	6	0.010	4.2	0.026	14.3	0.010	4	0.345	8	0.010	2	0.010	8	0.018	8	0.078	9	0.010	9.2	0.010	11	0.060	10.1	0.010	8	0.010	5	0.012	7	0.010	5	0.010	5	0.050	9.2	0.010	6.2	0.010	13	0.014	11	0.010	2.3	0.010	4.2	0.010	6	0.010
	7	0.024	8.1	0.014	15.3	0.010	6	0.034	11	0.010	3	0.027	9	0.028	8.2	0.010	10	0.010	4.3	0.010	12	0.085	11	0.010	10	0.010	6.1	0.010	8	0.010	6	0.012	7	0.018	11.1	0.026	5	0.010	13	0.012	12	0.018	3.3	0.010	7	0.015	1	0.010
	8	0.010	9.1	0.010	16.3	0.010	7	0.070	12	0.133	5	0.010	10	0.074	9	0.012	11	0.010	5	0.010	13	0.438	12	0.010	11	0.010	7	0.010	10	0.046	7	0.123	8	0.030	11.2	0.010	6	0.024	14	0.033	13	0.010	5	0.010	8	0.554	2	0.010
	9	0.010	10.1	0.010	17	0.010	8	0.076	13	0.171	6	0.010	11	0.092	9.1	0.106	12	0.016	5.4	0.010	13.3	0.010	13	0.034	12	0.011	8	0.010	11	0.099	8	0.010	9	0.030	12	0.036	6.3	0.010	14.2	0.010	14	0.024	10	0.010	8.4	0.010	4	0.036
	10	0.010	11.1	0.012	17.3	0.010	9	0.042	14	0.219	7	0.094	12	0.060	10	0.010	13	0.064	6	0.010	14	0.169	13.1	0.010	13	0.013	9	0.010	12	0.339	9	0.014	10	0.010	12.1	0.110	7	0.498	15	0.090	15	0.120	11	0.017	9	0.084	4.2	0.032
	10.1	0.010	12.1	0.010	18	0.010	9.1	0.010	15	0.215	8	0.280	13	0.094	10.1	0.010	14	0.161	6.4	0.010	15	0.089	14	0.018	14	0.050	10	0.032	13	0.296	9.3	0.010	10.3	0.030	13	0.032	8	0.156	16	0.364	16	0.193	11.3	0.010	9.2	0.010	5	0.010
	11	0.010	13.1	0.010	18.3	0.010	9.2	0.010	16	0.153	9	0.113	14	0.185	11	0.010	15	0.197	7	0.056	16	0.060	14.1	0.010	14.2	0.010	11	0.211	14	0.159	10	0.045	11	0.109	13.2	0.010	9	0.085	17	0.133	17	0.151	12	0.021	9.3	0.010	5.2	0.207
	12	0.066	16.1	0.010	19	0.010	10	0.052	17	0.072	10	0.189	15	0.189	11.2	0.010	16	0.191	7.4	0.010	17	0.044	15	0.044	14.3	0.010	11.1	0.016	15	0.032	10.3	0.030	11.2	0.030	13.3	0.092	10	0.099	18	0.232	18	0.197	13	0.044	10	0.097	6	0.010
	12.1	0.010	16.3	0.010	19.1	0.010	10.1	0.040	18	0.018	10.1	0.010	16	0.108	12	0.028	17	0.141	8	0.014	18	0.010	15.1	0.010	15	0.097	12	0.262	16	0.032	11	0.255	12	0.020	14	0.010	11	0.087	18.2	0.010	19	0.104	14	0.046	10.3	0.010	6.2	0.125
	13	0.155	17	0.010	19.3	0.020	10.3	0.016	19	0.012	11	0.121	17	0.094	12.1	0.010	17.3	0.010	8.4	0.010	19	0.090	15.2	0.010	16	0.159	12.1	0.023	17	0.010	11.1	0.010	13	0.044	14.3	0.010	12	0.036	19	0.395	20	0.078	14.2	0.010	11	0.230	7	0.036
	13.1	0.010	18	0.010	20	0.012	11	0.026	20	0.010	12	0.111	18	0.046	12.2	0.010	18	0.106	9	0.276	20	0.012	16	0.052	16.1	0.010	12	0.128	14	0.073	15	0.032	13	0.010	14.3	0.010	20.3	0.010	14.3	0.030	11.1	0.010	7.2	0.036
	14	0.398	18.2	0.010	20.1	0.010	11.1	0.016			13	0.045	19	0.042	13	0.133	19	0.044	9.4	0.012	21	0.010	16.2	0.010	17	0.216	13.1	0.020					12.1	0.010	14.2	0.010	15.1	0.022			20	0.061	21	0.048	15	0.054	11.2	0.010	7.3	0.012
	14.1	0.010	19	0.050	20.3	0.014	11.2	0.010			14	0.012	20	0.010	13.1	0.010	20	0.020	10	0.187			17	0.077	18	0.239	14	0.090					13	0.307	15	0.085	15.2	0.016			20.2	0.010	22	0.022	15.3	0.027	11.3	0.010	8	0.064
	15	0.329	19.2	0.020	21	0.012	11.3	0.010					21	0.010	14.2	0.010	21	0.010	10.4	0.010			18.3	0.010	19	0.108	14.3	0.030					14	0.332	15.2	0.030	15.3	0.048			21	0.045	22.1	0.010	16	0.040	12	0.174	8.2	0.026
	15.1	0.010	20	0.048	21.2	0.010	12	0.024							14	0.108	24	0.010	11	0.134			18.3	0.014	20	0.950	15	0.045					14.1	0.010	16	0.095	16	0.024			22	0.010	23	0.010	16.2	0.010	12.1	0.010	8.3	0.030
	16	0.102	20.2	0.030	21.3	0.020	13	0.028							14.1	0.010	27	0.010	11.4	0.010			19	0.058	21	0.028	15.1	0.030					15	0.135	17	0.111	16.1	0.243			23	0.010	24	0.010	16.3	0.010	12.3	0.030	9	0.159
	17	0.040	20.3	0.010	22	0.027	14	0.030							14.2	0.010	31	0.010	12	0.187			19.1	0.016	22	0.030	16	0.020					16	0.034	17.2	0.010	16.2	0.010			23.2	0.010	24.1	0.010	17	0.025	13	0.097	9.2	0.036
	17.1	0.010	21	0.020	22.1	0.010	14.3	0.010							15	0.096	32	0.010	12.4	0.010			19.2	0.010	23	0.010	17	0.030					17	0.010	18	0.051	16.3	0.010					24.2	0.010	17.2	0.013	13.3	0.030	9.3	0.010
	18	0.010	21.2	0.036	22.2	0.010	15	0.052							15.1	0.012	34	0.010	13	0.045			19.3	0.018	24	0.010	17.1	0.010					18	0.010	18.2	0.030	17.1	0.058					25.1	0.010	17.3	0.015	14	0.084	10	0.050
	19	0.010	21.3	0.010	22.3	0.055	15.3	0.010							15.2	0.010	40.2	0.010	14	0.093			20	0.048	33.3	0.030	18	0.010							19	0.077	17.3	0.010					26	0.010	18	0.010	14.3	0.010	10.2	0.024
	20	0.010	22	0.018	23	0.031	16	0.044							15.3	0.010	41.3	0.010	15	0.014			20.1	0.014			19	0.010							19.2	0.012	18.1	0.010					26.1	0.010	18.2	0.010	15	0.052	10.3	0.010
			22.1	0.010	23.1	0.010	16.3	0.010							16	0.157			18	0.010			20.2	0.014			20	0.046	18.3	0.012					20.2	0.030	19	0.010					27	0.010	18.3	0.010	15.2	0.010	11	0.016
			22.2	0.050	23.2	0.010	17	0.034							16.1	0.028							20.3	0.010			20	0.030	21.3	0.030					21	0.042	19.1	0.026					28	0.010	19	0.035	16	0.026	11.2	0.028
			22.3	0.010	23.3	0.024	17.3	0.002							16.2	0.010							21	0.028			21	0.010	23.3	0.030					21.2	0.032	19.3	0.090							20.1	0.010	17.3	0.010	16.3	0.010	11.3	0.010
			23	0.056	24	0.012	18	0.024							16.3	0.010							21.1	0.018			22	0.012	24	0.030					22	0.012	20	0.014							20.3	0.010	18	0.010	12.3	0.010	12	0.012
			23.2	0.052	24.2	0.010	19	0.010							17	0.072							21.2	0.010			22.2	0.010	24.3	0.030					22.2	0.010	20.1	0.026							21	0.035	18.3	0.010	13.2	0.014
			23.3	0.010	24.3	0.076	20	0.010							17.1	0.016							21.3	0.090			23	0.010	25	0.030					23	0.010	20.2	0.020							22	0.013	19	0.010	13.2	0.012
			24	0.042	25	0.010 16.3*		0.010							17.2	0.010							22	0.010			23.2	0.030	30	0.030					23.2	0.030	21	0.010							22.1	0.010			13.3	0.010
			24.2	0.036	25.1	0.010									18	0.054							22	0.010			23.3	0.010							24	0.010	21.1	0.010							23	0.010			14	0.010
			24.3	0.010	25.2	0.014									18.1	0.012							22.1	0.024			24	0.010							24.2	0.010	21.3	0.022							23.1	0.010			14.2	0.042
			25	0.034	25.3	0.096									19	0.020							22.2	0.014			25	0.010							25	0.010	21.3	0.010							24	0.023			15	0.010
			25.1	0.010	26	0.010									19.1	0.010							22.3	0.026			25.2	0.030							25.2	0.030	22.2	0.010							25	0.023			15.2	0.010
			25.2	0.032	26.2	0.022									20	0.010							23	0.030			26	0.030							25.2	0.030									26	0.010			16	0.010
			26	0.036	26.3	0.106									20.2	0.010							23.1	0.024			26.2	0.030							26	0.030									26.2	0.010			16.2	0.030
			26.1	0.010	27	0.012									21	0.010							23.2	0.032			27.1	0.030							27	0.010									27	0.010			18	0.010
			26.2	0.030	27.2	0.020									21.2	0.010							23.3	0.032			27.2	0.030							28	0.030									27.2	0.010			18.2	0.030
			27	0.040	27.3	0.039																	24	0.010			28	0.030							28.2	0.030									28	0.010			22	0.010
			27.2	0.012	28	0.010																	24.1	0.010											28.2	0.030									28.2	0.010			22.2	0.030
			28	0.024	28.2	0.018																	24.2	0.010											29	0.030									29	0.010				
			28.1	0.010	28.3	0.055																	24.3	0.020											29.2	0.030									29.2	0.010				
			28.2	0.016	29	0.014																	25.1	0.010																					30	0.010				
			29	0.020	29.2	0.020																	25.2	0.020																					30.2	0.010				
			29.1	0.010	29.3	0.037																	25.3	0.010																					31.2	0.010				
			29.2	0.014	30.2	0.018																	26.1	0.010																					31.2	0.010				
			29.3	0.014	30.3	0.016																	26.2	0.010																					32	0.010				
			30	0.010	31	0.010																	26.3	0.010																					32.2	0.023				
			30.1	0.022	31.2	0.012																	27	0.010																					33	0.025				
			30.2	0.010	31.3	0.010																	27.2	0.016																					33.2	0.035				
			30.3	0.012	32	0.010																	27.3	0.010																					34	0.021				
			31	0.010	32.1	0.010																	28.2	0.010																					34.2	0.019				
			31.1	0.024	32.2	0.010																	28.3	0.010																					34.3	0.030				
			31.2	0.010	32.3	0.014																	29	0.010																					35	0.025				
			31.3	0.010	33.2	0.010																	29.1	0.010																					35.2	0.035				
			32.1	0.010	34.1	0.010																	29.2	0.014																					36	0.017				
			32.3	0.012	34.2	0.012																	29.3	0.010																					36.2	0.015				
			33.1	0.010	34.3	0.010																	30.2	0.012																					37	0.019				
			33.3	0.010	35.2	0.012																	30.3	0.010																					37.2	0.017				
			34.2	0.010	36.1	0.010																	31.2	0.010																					38	0.010				
			35.1	0.010	37.2	0.010																	31.3	0.010																					38.2	0.015				
			35.2	0.010	38.2	0.010																	32.1	0.010																					39	0.021				
			36	0.010	39.2	0.010																	32.2	0.010																					39.2	0.023				
			36.2	0.010	41.1	0.010																	32.3	0.010																					40	0.023				
			37.2	0.010	42.2	0.010																	33.2	0.010																					40.2	0.010				
			37.3	0.010	43.2	0.010																	34	0.010																					41	0.010				
			38	0.010	45.2	0.010																	34.2	0.010																					41.2	0.013				
			40.1	0.010	48.2	0.010																	35	0.010																					42	0.010				
			41.1	0.010	50.2	0.010																	35.2	0.010																					42.2	0.010				
				24.2*		0.010																	36	0.010																					43	0.010				
																							36.1	0.010																					43.2	0.010				
																							36.2	0.010																					44	0.010				
																							36.3	0.010																					44.2	0.019				
																							37	0.010																					45	0.010				
																							37.1	0.010																					45.2	0.010				
																							37.2	0.010																					46	0.010				
																							38.1	0.010																					46.2	0.010				
																							38.2	0.010																					47.2	0.010				
																							39	0.010																					20*		0.010			
																							39.1	0.010																										
																							39.2	0.010																										
																							40	0.010																										
																							40.1	0.010																										
																							42	0.010																										
																							42.1	0.010																										
																							43	0.010																										
																							44	0.010																										
																							46.1	0.010																										
																							48	0.010																										
																							53	0.010																										

A4.5.2 Sub-population NEC

NEC population Match Probability Calc. Sheet

DNA profile	MsF5s	MS25	MsF33	MsF31	MsF9	MsF16	MsF3	MsF17	MsF18	MsF28si	MsF22	MS3	MS4h	MsF27s	MsF8	MsF26	MsF4	MS27	MsF24	MS9	MsF2	MS17	MsF15	MS13
Allele 1																								
Allele 2																								

1. Standard profile match between

Theta value	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048
Allele1																								
Allele2																								
P(hat)	0.004011811	0.004011811	0.004011813	0.004011811	0.004011811	0.004011811	0.004011811	0.004011811	0.004011811	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012	0.004012
P(hom)	0.012035444	0.012035444	0.012035438	0.012035444	0.012035444	0.012035444	0.012035444	0.012035444	0.012035444	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035
Appropriate Prob:	0.012035444	0.012035444	0.012035438	0.012035444	0.012035444	0.012035444	0.012035444	0.012035444	0.012035444	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035	0.012035
Cumulative P	0.012035444	0.00014485	1.74335E-06	2.0982E-08	2.5253E-10	3.0393E-12	3.6579E-14	4.4026E-16	5.2985E-18	6.38E-20	7.68E-22	9.24E-24	1.11E-25	1.34E-27	1.61E-29	1.94E-31	2.33E-33	2.81E-35	3.38E-37	4.07E-39	4.89E-41	5.89E-43	7.09E-45	8.53E-47
Cum Product	8.5327E-47																							

Cumulative Likelihood **1.172E+46**

Rarest genotype freq: 5/(2*67)
0.037
6.0518E-34
1.5524E+33 **Rarest genotype should be 1 in 1.65 x 10^{33}**

Most common genotype frequency: 1.1686E-29 or the inverse: 8.5574E+28 **Most common genotype is 1 in 8.56 x 10^{28}**

		0.144		0.028		0.007		0.021		0.107		0.141		0.067		0.127		0.071		0.094		0.138		0.008
Most common allele freqs:	Allele1	0.426		0.127		0.060		0.110		0.250		0.351		0.206		0.279		0.243		0.258		0.360		0.067
	Allele2	0.169		0.112		0.060		0.096		0.213		0.201		0.162		0.228		0.147		0.182		0.191		0.060
Least common allele freqs:	Allele1	0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037
	Allele2	0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037

Allele Freq Look Up Table

MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
8	0.037	16.1	0.037	14.3	0.037	4	0.074	12	0.051	6	0.037	9	0.037	11	0.037	9	0.037	4.3	0.037	11	0.103	10.1	0.037
9	0.037	18.2	0.037	15.3	0.037	6	0.037	13	0.125	8	0.067	10	0.044	12	0.059	10	0.037	5	0.037	12	0.191	14.1	0.037
10.1	0.037	19.2	0.037	16.3	0.037	7	0.037	14	0.250	9	0.201	11	0.037	13	0.176	11	0.037	5.4	0.037	13	0.360	15.2	0.037
12	0.037	20	0.037	17	0.037	8	0.037	15	0.213	10	0.351	12	0.051	14	0.088	12	0.051	6	0.037	14	0.191	16.2	0.037
12.1	0.037	20.2	0.037	18	0.037	9	0.037	16	0.206	11	0.179	13	0.154	15	0.228	13	0.066	6.4	0.037	15	0.096	17	0.037
13	0.125	20.3	0.037	18.3	0.037	9.1	0.037	17	0.103	12	0.142	14	0.140	15.3	0.037	14	0.110	7	0.182	16	0.044	17.3	0.037
13.1	0.037	21.2	0.037	19	0.037	10	0.110	18	0.037	13	0.045	15	0.206	16	0.279	15	0.147	7.4	0.037	17	0.037	18	0.037
14	0.426	21.3	0.037	19.3	0.037	10.1	0.088	19	0.037	14	0.037	16	0.162	16.1	0.037	16	0.243	8	0.037	18	0.037	18.3	0.037
14.1	0.037	22	0.037	20	0.037	10.3	0.037					17	0.088	16.3	0.037	17	0.140	8.4	0.037			19.2	0.037
15	0.169	22.1	0.037	21	0.037	11	0.081					18	0.066	17	0.088	17.3	0.037	9	0.258			19.3	0.037
15.1	0.037	22.2	0.037	21.2	0.037	11.1	0.037					19	0.037	18	0.037	18	0.096	9.4	0.045			20	0.037
16	0.103	22.3	0.037	21.3	0.045	11.2	0.037					20	0.037	19	0.037	19	0.037	10	0.045			20.1	0.037
17	0.037	23.2	0.037	22	0.037	11.3	0.037									24	0.037	10.4	0.037			20.2	0.037
19	0.037	24	0.037	22.2	0.037	12	0.037									27	0.037	11	0.091			21	0.045
20	0.037	24.2	0.037	22.3	0.045	13	0.066									32	0.037	11.4	0.037			21.1	0.037
		24.3	0.037	23	0.037	14	0.059									40.2	0.037	12	0.098			21.2	0.037
		25	0.052	23.2	0.037	15	0.096											12.4	0.037			21.3	0.060
		25.1	0.037	23.3	0.037	15.3	0.037											13	0.053			22.2	0.045
		25.2	0.037	24	0.037	16	0.081											14	0.068			22.3	0.045
		26	0.112	24.2	0.037	16.3	0.037											15	0.037			23.2	0.037
		26.1	0.037	24.3	0.052	17	0.044											18	0.037			23.3	0.052
		26.2	0.060	25	0.037	17.3	0.037															24.2	0.037
		27	0.127	25.2	0.045	18	0.037															24.3	0.037
		27.2	0.037	25.3	0.060	19	0.037															25.1	0.037
		28	0.037	26	0.037	20	0.037															25.2	0.067
		28.2	0.037	26.2	0.060 16.3*		0.037															25.3	0.037
		29	0.037	26.3	0.060																	26.1	0.037
		29.2	0.037	27	0.037																	26.2	0.037
		29.3	0.037	27.2	0.052																	26.3	0.037
		30	0.037	27.3	0.037																	27	0.037
		30.1	0.037	28.2	0.037																	27.2	0.052
		30.3	0.037	28.3	0.045																	28.2	0.037
		31.1	0.037	29.2	0.037																	28.3	0.037
		31.2	0.037	29.3	0.037																	29.1	0.037
		31.3	0.037	30.2	0.037																	29.2	0.045
		32.3	0.037	30.3	0.037																	29.3	0.037
		33.3	0.037	31.2	0.037																	30.2	0.037
		35.2	0.037	32.3	0.037																	31.2	0.037
		38	0.037	34.1	0.037																	31.3	0.037
				34.2	0.037																	32.1	0.037
				37.2	0.037																	32.2	0.037
				38.2	0.037																	34.2	0.037
				39.2	0.037																	35	0.037
				41.1	0.037																	38.1	0.037
				42.2	0.037																	38.2	0.037
				43.2	0.037																	43	0.037
				45.2	0.037																	44	0.037
				48.2	0.037																		
				50.2	0.037																		

		0.093		0.169		0.137		0.107		0.056		0.072		0.136		0.091		0.069		0.013		0.075		0.026
		0.250		0.338		0.306		0.239		0.172		0.257		0.269		0.254		0.213		0.085		0.213		0.118
		0.185		0.250		0.224		0.224		0.164		0.140		0.254		0.179		0.162		0.076		0.176		0.110
		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037
		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037		0.037

MS4h		MsF27s		MsF8		MsF26		MsF4		MS27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
10	0.037	6.1	0.037	7	0.037	9	0.037	7	0.067	9.2	0.037	5	0.037	12	0.037	11	0.037	14.2	0.037	4.2	0.037	4.2	0.037
14	0.048	9	0.037	8	0.037	10	0.037	11	0.037	11.1	0.037	7	0.112	13	0.037	13	0.037	15	0.042	7	0.037	5.2	0.118
14.3	0.037	10	0.037	10	0.075	11	0.052	13	0.037	13.2	0.037	8	0.269	14	0.037	14	0.037	16	0.037	8	0.140	6.2	0.051
15	0.137	11	0.250	11	0.157	11.1	0.037	14	0.060	14.3	0.037	9	0.254	15	0.090	15	0.059	17	0.042	8.4	0.037	7	0.037
16	0.161	12	0.338	12	0.224	12	0.187	15	0.060	15	0.051	10	0.164	16	0.112	16	0.125	17.2	0.037	9	0.169	7.2	0.081
17	0.185	12.1	0.037	13	0.306	12.1	0.037	16	0.164	15.1	0.037	11	0.134	17	0.112	17	0.162	19	0.037	9.2	0.037	8	0.037
18	0.250	13	0.184	14	0.134	13	0.224	17	0.172	15.2	0.037	12	0.045	18	0.254	18	0.213	26.2	0.037	9.3	0.037	8.2	0.037
19	0.145	13.1	0.037	15	0.045	14	0.239	17.2	0.037	15.3	0.140	13	0.037	18.2	0.037	19	0.096	27.2	0.037	10	0.176	8.3	0.037
20	0.037	14	0.066	16	0.037	14.1	0.037	18	0.067	16	0.066			19	0.179	20	0.110	28.2	0.037	10.3	0.037	9	0.044
21	0.037	14.3	0.037	17	0.037	15	0.157	19	0.082	16.1	0.051			19.2	0.037	20.3	0.037	29.2	0.037	11	0.081	9.2	0.044
33.3	0.037	15	0.037			16	0.075	20	0.082	16.3	0.037			20	0.082	21	0.074	30	0.037	12	0.213	10	0.096
		16	0.037			17	0.037	20.2	0.037	17.3	0.037			20.2	0.037	22	0.037	31.2	0.037	13	0.074	10.2	0.059
		17	0.037					21	0.082	18.3	0.037			21	0.037	22.1	0.037	32	0.037	14	0.044	11	0.051
		18	0.037					22	0.037	19	0.037					24	0.037	32.2	0.037	15	0.037	11.2	0.059
		21.3	0.037					22.2	0.037	19.1	0.037					24.1	0.037	33	0.037	19	0.037	12	0.037
		23.3	0.037					23	0.037	19.3	0.257					24.2	0.037	33.2	0.037			12.2	0.037
		24.3	0.037					24	0.037	20	0.051					25.1	0.037	34	0.051			13	0.037
		25	0.037					25	0.037	20.1	0.051					26	0.037	34.2	0.037			13.2	0.037
								26	0.037	20.3	0.051					27	0.037	34.3	0.037			14	0.037
								27.1	0.037	21	0.037					28	0.037	35	0.037			14.2	0.110
								28	0.037	21.3	0.066							35.2	0.076			15	0.037
								29	0.037									36	0.042			15.2	0.037
																		36.2	0.037			16	0.037
																		37	0.042			16.2	0.037
																		37.2	0.037			18	0.037
																		38	0.037			18.2	0.037
																		38.2	0.037			22	0.037
																		39	0.085				
																		39.2	0.076				
																		40	0.037				
																		40.2	0.037				
																		41	0.037				
																		41.2	0.037				
																		43	0.037				
																		44.2	0.042				
																		45.2	0.037				
																		20*	0.037				

A4.5.3 Sub-population SEC

SEC population Match Probability Calc. Sheet

DNA profile	MsF5s	MS25	MsF33	MsF31	MsF9	MsF16	MsF3	MsF17	MsF18	MsF28si	MsF22	MS3	MS4h	MsF27s	MsF8	MsF26	MsF4	MS27	MsF24	MS9	MsF2	MS17	MsF15	MS13
Allele 1																								
Allele 2																								

1. Standard profile match between

Theta value	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
Allele1																								
Allele2																								
P(het)	0.00012498	0.00012498	0.00012498	0.00012498	0.00012498	0.00012498	0.00012498	0.00012498	0.00012498	0.00012498	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125	0.000125
P(hom)	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375
Appropriate Prob.	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.00037495	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375	0.000375
Cumulative P	0.00037495	1.4059E-07	5.27146E-11	1.9766E-14	7.4111E-18	2.7788E-21	1.0419E-24	3.9068E-28	1.4649E-31	5.49E-35	2.06E-38	7.72E-42	2.9E-45	1.09E-48	4.07E-52	1.53E-55	5.72E-59	2.15E-62	8.05E-66	3.02E-69	1.13E-72	4.24E-76	1.59E-79	5.96E-83
Cum Product	5.9628E-83																							

Cumulative Likelihood **1.6771E+82**

Rarest genotype freq: 5/(2*51)
0.049
7.2947E-31
1.3709E+30 **Rarest genotype should be 1 in 1.37 x 10^{30}**

Most common genotype frequency: 9.3102E-32 or the inverse: 1.0741E+31 **Most common genotype is 1 in 1.07 x 10^{31}**

		0.095		0.012		0.010		0.027		0.128		0.075		0.052		0.073		0.047		0.100		0.093		0.012
Most common allele freqs:	Allele1	0.284		0.078		0.070		0.127		0.284		0.225		0.167		0.196		0.186		0.255		0.265		0.088
	Allele2	0.167		0.078		0.070		0.108		0.225		0.167		0.157		0.186		0.127		0.196		0.176		0.069
Least common allele freqs:	Allele1	0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049
	Allele2	0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049

Locus

MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
6	0.049	20.2	0.049	19.1	0.049	6	0.049	12	0.049	3	0.049	10	0.078	9	0.049	11	0.049	6	0.049	11	0.049	13.1	0.049
9	0.049	21	0.049	19.3	0.049	7	0.049	13	0.059	5	0.049	11	0.059	10	0.049	12	0.049	7	0.049	12	0.108	14.1	0.049
11	0.049	21.2	0.049	20	0.049	8	0.049	14	0.186	6	0.049	12	0.108	11.2	0.049	13	0.069	8	0.059	13	0.265	15.1	0.049
12	0.049	22.3	0.049	20.1	0.049	9.2	0.049	15	0.284	8	0.147	13	0.127	12.2	0.049	14	0.108	9	0.059	14	0.167	15.2	0.049
13	0.157	23.2	0.049	20.3	0.049	10	0.049	16	0.225	9	0.108	14	0.167	13	0.069	15	0.127	10	0.176	15	0.176	17	0.049
13.1	0.049	23.3	0.049	21	0.050	10.1	0.078	17	0.118	10	0.167	15	0.157	13.2	0.049	16	0.186	11	0.196	16	0.176	18	0.049
14	0.167	24	0.069	22.1	0.049	10.3	0.049	18	0.059	11	0.225	16	0.088	14	0.049	17	0.127	12	0.255	17	0.049	18.3	0.049
15	0.127	24.2	0.049	22.3	0.049	11	0.049	19	0.049	12	0.137	17	0.078	14.2	0.049	18	0.118	13	0.137	18	0.049	19.3	0.069
16	0.284	25	0.049	23	0.049	11.1	0.059	20	0.049	13	0.078	18	0.098	15	0.088	19	0.098	14	0.069	19	0.049	20	0.049
17	0.098	25.2	0.049	23.1	0.049	12	0.078			14	0.049	19	0.049	15.2	0.049	20	0.069	15	0.049			20.2	0.049
17.1	0.049	26	0.049	23.3	0.049	13	0.049					20	0.049	16	0.186	21	0.049					20.3	0.049
18	0.049	26.2	0.049	24	0.049	14	0.069							16.2	0.049	24	0.049					21	0.049
19	0.049	27	0.049	24.3	0.050	14.3	0.049							17	0.127	31	0.049					21.3	0.069
20	0.049	27.2	0.049	25	0.049	15	0.127							17.2	0.049	34	0.049					22.2	0.049
		28	0.078	25.1	0.049	16	0.108							18	0.196	41.3	0.049					22.3	0.069
		28.1	0.049	25.2	0.049	17	0.108							19	0.049							23.3	0.088
		28.2	0.049	25.3	0.050	18	0.069							20	0.049							24	0.049
		29	0.049	26	0.049	19	0.049							20.2	0.049							24.3	0.069
		29.1	0.049	26.2	0.049									21	0.049							25.2	0.049
		29.2	0.049	26.3	0.060									21.2	0.049							25.3	0.049
		29.3	0.059	27	0.050																	26.3	0.049
		30	0.049	27.2	0.049																	27.3	0.049
		30.1	0.078	27.3	0.049																	28.3	0.049
		30.2	0.049	28	0.050																	30.2	0.049
		30.3	0.049	28.2	0.049																	30.3	0.049
		31	0.049	28.3	0.049																	31.2	0.049
		31.1	0.049	29	0.049																	31.3	0.049
		31.2	0.049	29.2	0.070																	32.3	0.049
		32.1	0.049	29.3	0.070																	33.2	0.049
		32.3	0.049	30.2	0.049																	34	0.049
		33.1	0.049	31	0.049																	35	0.049
		34.2	0.049	31.2	0.049																	35.2	0.049
		35.1	0.049	32	0.049																	36.1	0.049
		36	0.049	32.2	0.049																	36.2	0.049
		36.2	0.049	34.3	0.049																	37	0.049
		37.2	0.049																			37.1	0.049
		37.3	0.049																			37.2	0.049
		38	0.049																			38.2	0.049
		40.1	0.049																			39.1	0.049
		41.1	0.049																			39.2	0.049
																						40.1	0.049
																						42	0.049
																						42.1	0.049
																						46.1	0.049
																						48	0.049
																						53	0.049

		0.077		0.074		0.178		0.151		0.017		0.055		0.130		0.068		0.077		0.008		0.042		0.046
		0.213		0.230		0.314		0.275		0.108		0.176		0.265		0.200		0.196		0.069		0.160		0.157
		0.181		0.160		0.284		0.275		0.078		0.157		0.245		0.170		0.196		0.069		0.130		0.147
		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049
		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049		0.049

MS4h		MsF27s		MsF8		MsF26		MsF4		MS27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
12	0.049	5	0.060	10	0.049	5	0.049	10	0.049	9.2	0.049	6	0.049	13	0.049	15	0.078	11	0.049	7	0.060	0	0.049
13	0.049	11	0.049	11	0.069	6	0.049	11	0.049	11.2	0.049	7	0.049	14	0.049	16	0.108	12	0.049	8	0.050	2	0.049
14	0.049	11.1	0.049	12	0.157	7	0.049	11.2	0.049	12	0.176	8	0.245	15	0.130	17	0.049	13	0.049	9	0.110	4.2	0.127
15	0.085	12	0.160	13	0.284	9	0.049	12	0.049	13	0.157	9	0.078	16	0.080	18	0.147	14	0.049	9.3	0.049	5	0.049
16	0.181	12.1	0.049	14	0.314	10	0.049	13	0.049	13.3	0.078	10	0.265	17	0.130	19	0.196	15	0.049	10	0.070	5.2	0.157
16.1	0.049	13	0.230	15	0.088	11	0.049	14	0.049	14	0.049	11	0.245	18	0.200	20	0.196	16	0.049	11	0.130	6	0.049
17	0.170	13.1	0.049	16	0.049	11.1	0.049	14.2	0.049	15	0.049	12	0.118	19	0.170	21	0.127	17	0.049	11.1	0.049	6.2	0.147
18	0.213	14	0.130	17	0.049	12	0.069	15	0.049	15.1	0.049	13	0.049	20	0.100	22	0.059	18	0.049	11.2	0.049	7	0.049
18.3	0.049	14.1	0.049			12.1	0.049	15.2	0.049	15.2	0.049			21	0.100	23	0.049	19	0.049	11.3	0.049	7.2	0.049
19	0.096	15	0.120			13	0.078	16	0.069	15.3	0.049			22	0.049	24	0.049	19.1	0.049	12	0.160	8	0.049
20	0.117	16	0.070			14	0.275	17	0.069	16	0.049			23	0.049	26	0.049	20.1	0.049	12.1	0.049	8.2	0.078
21	0.049	17	0.049			15	0.275	17.2	0.049	16.1	0.078			23.2	0.049	26.1	0.049	20.3	0.049	12.3	0.049	9	0.049
22	0.049	17.1	0.049			16	0.069	18	0.069	16.2	0.049							22.1	0.049	13	0.090	9.2	0.049
		18	0.049			17	0.049	18.2	0.049	18.3	0.049							28	0.049	13.3	0.049	9.3	0.049
		19	0.049			18	0.049	19	0.108	19.1	0.078							29	0.049	14	0.049	10	0.049
		20	0.049					19.2	0.049	19.3	0.098							30	0.049	14.3	0.049	10.2	0.049
		24	0.049					20	0.078	20.3	0.049							30.2	0.049	15	0.049	10.3	0.049
		30	0.049					20.2	0.049	21.1	0.069							32	0.049	15.2	0.049	11	0.049
								21	0.059	21.3	0.049							32.2	0.049	16	0.049	11.2	0.049
								21.2	0.059	22.1	0.049							33	0.049	16.3	0.049	11.3	0.049
								22	0.049	22.2	0.049							34	0.049	17	0.049	12	0.049
								22.2	0.049									35	0.059	18.3	0.049	12.2	0.049
								23.2	0.049									35.2	0.049			12.3	0.049
								24.2	0.049									36	0.049			13	0.049
								25.2	0.049									36.2	0.049			13.2	0.049
								26.2	0.049									37	0.049			14.2	0.049
								27.2	0.049									37.2	0.049			15	0.049
								28.2	0.049									38	0.049			15.2	0.049
								29.2	0.049									38.2	0.049			22.2	0.049
								30	0.049									39.2	0.049				
																		40	0.069				
																		40.2	0.049				
																		41	0.049				
																		41.2	0.049				
																		42	0.049				
																		42.2	0.049				
																		43	0.049				
																		43.2	0.049				
																		44	0.049				
																		44.2	0.049				
																		45	0.049				
																		45.2	0.049				
																		46	0.049				
																		46.2	0.049				
																		47.2	0.049				

A4.5.4 Sub-population SIB

SIB population Match Probability Calc. Sheet

DNA profile	MsF5s	MS25	MsF33	MsF31	MsF9	MsF16	MsF3	MsF17	MsF18	MsF28si	MsF22	MS3	MS4h	MsF27s	MsF8	MsF26	MsF4	MS27	MsF24	MS9	MsF2	MS17	MsF15	MS13
Allele 1	13	22.2	-	8	14	7	16	16.1	16	-	-	19	18	12	12	-	16	12.1	-	18	13	15	11	8
Allele 2	13	22.2	-	9	15	8	19	17.1	16	-	-	21	21	13	13	-	19	12.1	-	18	16	15	16	9

1. Standard profile match between

Theta value	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149
Allele1	0.277	0.196	-	0.286	0.295	0.364	0.179	0.116	0.188	-	-	0.107	0.245	0.213	0.348	-	0.134	0.491	-	0.295	0.045	0.179	0.127	0.241
Allele2	0.277	0.196	-	0.161	0.250	0.045	0.125	0.071	0.188	-	-	0.045	0.045	0.139	0.313	-	0.116	0.491	-	0.295	0.250	0.179	0.078	0.607
P(het)	0.19830654	0.13404617	#VALUE!	0.15027384	0.19391881	0.11514913	0.10306872	0.06970632	0.12767961	#VALUE!	#VALUE!	0.060324	0.089812	0.118326	0.247797	#VALUE!	0.087379	0.430972	#VALUE!	0.214283	0.090859	0.121468	0.074488	0.316145
P(hom)	0.2441813	0.19155449	#VALUE!	0.25040961	0.2567223	0.30811007	0.18071	0.14519922	0.18609352	#VALUE!	#VALUE!	0.140435	0.222679	0.201867	0.296217	#VALUE!	0.15496	0.415168	#VALUE!	0.256722	0.109429	0.18071	0.151383	0.220014
Appropriate Prob.	0.2441813	0.19155449	1	0.15027384	0.19391881	0.11514913	0.10306872	0.06970632	0.18609352	1	1	0.060324	0.089812	0.118326	0.247797	1	0.087379	0.415168	1	0.256722	0.090859	0.18071	0.074488	0.316145
Cumulative P	0.2441813	0.04677402	0.046774023	0.00702891	0.00136304	0.00015695	1.6177E-05	1.1276E-06	2.0985E-07	2.1E-07	2.1E-07	1.27E-08	1.14E-09	1.35E-10	3.33E-11	3.33E-11	2.91E-12	1.21E-12	1.21E-12	3.1E-13	2.82E-14	5.1E-15	3.8E-16	1.2E-16
Cum Product	1.2004E-16																							

Cumulative Likelihood **8.3307E+15**

Rarest genotype freq: 5/(2*56)
0.045
6.4112E-32
1.5598E+31 **Rarest genotype should be 1 in 1.55 x 10^{31}**

Most common genotype frequency: 2.4122E-22 or the inverse: 4.1455E+21 **Most common genotype is 1 in 4.14 x 10^{21}**

		0.138		0.067		0.039		0.143		0.158		0.172		0.091		0.110		0.095		0.192		0.167		0.047
Most common allele freqs:	Allele1	0.277		0.196		0.155		0.286		0.295		0.364		0.232		0.473		0.232		0.849		0.400		0.188
	Allele2	0.250		0.170		0.127		0.250		0.268		0.236		0.196		0.116		0.205		0.113		0.209		0.125
Least common allele freqs:	Allele1	0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045
	Allele2	0.045		0.045		0.045		0.071		0.045		0.045		0.045		0.045		0.045		0.113		0.045		0.045

Locus

MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
7	0.107	16.3	0.045	22.3	0.045	4	0.125	11	0.045	3	0.073	12	0.045	9.1	0.473	10	0.045	9	0.849	11	0.109	11	0.045
10	0.045	18.2	0.045	23.3	0.045	6	0.089	12	0.045	7	0.364	13	0.045	10.1	0.045	11	0.045	10	0.045	12	0.045	12	0.045
11	0.045	19.2	0.071	24	0.045	7	0.250	13	0.268	8	0.045	14	0.107	12.1	0.045	13	0.116	11	0.113	13	0.400	13	0.098
12	0.214	20.2	0.107	24.3	0.127	8	0.286	14	0.295	9	0.091	15	0.196	13	0.045	14	0.063			13.3	0.045	14	0.054
12.1	0.045	21.2	0.107	25.3	0.155	9	0.161	15	0.250	10	0.236	16	0.179	13.1	0.045	15	0.232			14	0.209	15	0.080
13	0.277	22.2	0.196	26.3	0.100	10	0.071	16	0.098	10.1	0.045	17	0.232	14	0.045	16	0.188			15	0.045	16	0.125
14	0.250	22.3	0.045	27.2	0.045	11.2	0.045	17	0.063	11	0.055	18	0.045	14.1	0.045	17	0.205			16	0.055	17	0.188
15	0.054	23.2	0.170	27.3	0.082					12	0.136	19	0.125	15	0.045	18	0.152			17	0.118	18	0.080
16	0.045	23.3	0.045	28.2	0.045							20	0.045	15.1	0.054	19	0.045			19	0.045	19	0.107
		24.2	0.134	28.3	0.055							21	0.045	16	0.045	21	0.045					20	0.045
		24.3	0.045	29	0.045									16.1	0.116							20.1	0.045
		25.2	0.071	29.2	0.045									17	0.045							21	0.045
		26.2	0.045	29.3	0.045									17.1	0.071							21.1	0.045
		28.2	0.045	30.2	0.045									18	0.045							22	0.045
		29.2	0.045	30.3	0.045									18.1	0.054							23	0.045
		32.1	0.045	31.2	0.045									19	0.045							24	0.045
				31.3	0.045										0.045							27	0.045
				32.1	0.045																	27.2	0.045
				32.2	0.045																	29	0.045
				32.3	0.045																	29.2	0.045
				33.2	0.045																	34	0.045
				34.1	0.045																	36	0.045
				34.2	0.045																	36.3	0.045
				35.2	0.055																	37	0.045
				36.1	0.045																	38.1	0.045
				38.2	0.045																	39	0.045
			24.2*		0.045																	40	0.045
	0.139		0.154		0.218		0.169		0.099		0.333		0.158		0.142		0.098		0.067		0.106		0.293
	0.284		0.361		0.348		0.545		0.241		0.491		0.791		0.295		0.250		0.188		0.235		0.607
	0.245		0.213		0.313		0.155		0.205		0.339		0.100		0.241		0.196		0.179		0.225		0.241
	0.045		0.045		0.045		0.045		0.045		0.071		0.045		0.045		0.045		0.045		0.045		0.045
	0.045		0.045		0.063		0.045		0.045		0.098		0.045		0.045		0.045		0.045		0.045		0.045

MS4h		MsF27s		MsF8		MsF26		MsF4		MS27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
11	0.045	7	0.045	10	0.063	6	0.045	10	0.045	11.1	0.098	0.2	0.045	14	0.054	12	0.080	2.3	0.045	10	0.045	7	0.045
12	0.045	8	0.045	11	0.089	10	0.045	11	0.071	12.1	0.491	6	0.100	15	0.107	13	0.045	3.3	0.045	11	0.127	7.3	0.054
13	0.045	10	0.045	12	0.348	11	0.545	12	0.045	13.3	0.339	6.3	0.045	16	0.241	14	0.054	9	0.045	12	0.078	8	0.241
14	0.059	11	0.361	13	0.313	12	0.155	13	0.045	15	0.071	7	0.791	17	0.116	15	0.170	10	0.045	13	0.225	9	0.607
14.2	0.045	12	0.213	14	0.179	13	0.109	14	0.054			8	0.082	18	0.295	16	0.250	11	0.063	14	0.235	9.3	0.045
15	0.078	12.1	0.045	15	0.045	14	0.045	15	0.045					19	0.134	17	0.196	12	0.080	15	0.147	10	0.063
16	0.127	13	0.139			15	0.073	16	0.134					20	0.045	18	0.143	13	0.143	16	0.078		
17	0.284	13.1	0.045					17	0.205					21	0.045	19	0.045	14	0.188	17	0.049		
18	0.245	14	0.130					18	0.241							20	0.045	15	0.179	17.3	0.045		
19	0.078	14.1	0.045					19	0.116							38.2	0.045	16	0.143	18	0.045		
21	0.045	15	0.045					20	0.045									16.2	0.045	19	0.045		
		16	0.045					21	0.045									17	0.054				
		19	0.045					22	0.045									17.2	0.045				
																		18.2	0.045				
																		19	0.045				
																		20.1	0.045				

A4.5.5 Sub-population Msi ML

Msi ML population Match Probability Calc. Sheet

DNA profile	MsF5s	MS25	MsF33	MsF31	MsF9	MsF16	MsF3	MsF17	MsF18	MsF28si	MsF22	MS3	MS4h	MsF27s	MsF8	MsF26	MsF4	MS27	MsF24	MS9	MsF2	MS17	MsF15	MS13
Allele 1																								
Allele 2																								

1. Standard profile match between

Theta value	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194	0.194
Allele1																								
Allele2																								
P(het)	0.04541922	0.04541922	0.045419219	0.04541922	0.04541922	0.04541922	0.04541922	0.04541922	0.04541922	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419	0.045419
P(hom)	0.13625766	0.13625766	0.136257657	0.13625766	0.13625766	0.13625766	0.13625766	0.13625766	0.13625766	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258
Appropriate Prob.	0.13625766	0.13625766	0.136257657	0.13625766	0.13625766	0.13625766	0.13625766	0.13625766	0.13625766	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258	0.136258
Cumulative P	0.13625766	0.01856615	0.00252978	0.0003447	4.6968E-05	6.3998E-06	8.7202E-07	1.1882E-07	1.619E-08	2.21E-09	3.01E-10	4.1E-11	5.58E-12	7.6E-13	1.04E-13	1.41E-14	1.92E-15	2.62E-16	3.57E-17	4.87E-18	6.63E-19	9.04E-20	1.23E-20	1.68E-21
Cum Product	1.6775E-21																							

Cumulative Likelihood **5.9613E+20**

Rarest genotype freq: 5/(2*36)
0.069
6.2512E-27
1.5997E+26 **Rarest genotype should be 1 in 1.6 x 10^{26}**

Most common genotype frequency: 1.522E-22 or the inverse: 6.5704E+21 **Most common genotype is 1 in 6.57 x 10^{21}**

		0.168		0.045		0.049		#NUM!		0.131		0.124		0.115		0.139		0.099		0.197		0.160		0.028
Most common allele freqs:	Allele1	0.403		0.181		0.194		1.000		0.278		0.603		0.319		0.264		0.222		0.329		0.444		0.167
	Allele2	0.208		0.125		0.125		#NUM!		0.236		0.103		0.181		0.264		0.222		0.300		0.181		0.083
Least common allele freqs:	Allele1	0.069		0.069		0.069		1.000		0.069		0.069		0.069		0.069		0.069		0.069		0.069		0.069
	Allele2	0.069		0.069		0.069		#NUM!		0.069		0.069		0.069		0.069		0.069		0.069		0.069		0.069

Locus	MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	11	0.069	4.2	0.181	17.3	0.069	4	1.000	12	0.069	8	0.603	8	0.125	8	0.069	10	0.069	3.2	0.069	12	0.069	12	0.069
	12	0.069	8.1	0.097	18.3	0.069			13	0.222	9	0.103	9	0.181	8.2	0.069	13	0.069	7	0.069	13	0.444	13	0.083
	13	0.139	9.1	0.069	19.3	0.083			14	0.236	10	0.069	10	0.319	9	0.069	14	0.222	9	0.069	14	0.153	14	0.069
	14	0.403	10.1	0.069	20.3	0.069			15	0.278	11	0.088	11	0.069	12	0.083	15	0.125	10	0.329	15	0.069	16	0.069
	15	0.208	11.1	0.083	21.3	0.069			16	0.194	12	0.088	12	0.125	13	0.264	16	0.222	11	0.300	17	0.069	17	0.069
	16	0.069	12.1	0.069	22	0.069			17	0.069	13	0.074	13	0.111	14	0.097	17	0.167	12	0.229	19	0.181	18	0.083
	17	0.083	13.1	0.069	22.3	0.069			18	0.069	14	0.069	14	0.069	15	0.083	18	0.153	13	0.069	20	0.083	19	0.167
	18	0.069	18	0.069	24.3	0.125							17	0.069	16	0.264	20	0.069	15	0.069	21	0.069	19.1	0.069
	19	0.069	19	0.069	25.3	0.083									17	0.083	21	0.069					20	0.069
			20	0.069	26.3	0.194									19	0.069							20.1	0.069
			20.3	0.069	27.3	0.111																	21	0.069
			21	0.069	28.3	0.083																	21.1	0.069
			21.3	0.069	29.3	0.069																	21.2	0.069
			22	0.069	30.3	0.069																	22	0.069
			23	0.125	31.3	0.069																	22.1	0.069
			23.3	0.069	32.2	0.069																	23.1	0.069
			24	0.125	32.3	0.069																	24	0.069
			25	0.097																			24.1	0.069
			26	0.069																				

		0.062		0.173		0.189		0.190		0.106		0.336		0.176		0.107		0.117		0.047		0.136		0.096
		0.191		0.309		0.486		0.403		0.347		0.417		0.903		0.279		0.264		0.153		0.328		0.431
		0.162		0.279		0.194		0.236		0.153		0.403		0.097		0.191		0.222		0.153		0.207		0.111
		0.069		0.069		0.069		0.069		0.069		0.069		0.097		0.069		0.069		0.069		0.069		0.069
		0.069		0.069		0.125		0.069		0.069		0.069		0.903		0.069		0.069		0.069		0.069		0.069

	MS4h		MsF27s		MsF8		MsF26		MsF4		MS27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	8	0.069	9	0.069	10	0.069	7	0.236	6	0.347	15.1	0.111	7	0.903	12	0.069	14	0.069	11.3	0.069	8	0.069	1	0.069
	10	0.069	11	0.088	11	0.153	8	0.069	8	0.069	15.2	0.069	8	0.097	14	0.069	15	0.167	14.3	0.069	9	0.069	4	0.111
	13	0.069	11.1	0.069	12	0.194	9.3	0.069	9	0.069	16	0.069			14.2	0.069	16	0.194	15.3	0.069	10	0.207	5	0.069
	14	0.103	12	0.279	13	0.486	10	0.167	11	0.083	16.1	0.417			15	0.103	17	0.264	16.3	0.069	11	0.328	5.2	0.431
	15	0.162	12.1	0.069	14	0.125	11	0.403	12	0.083	17.1	0.403			16	0.069	18	0.222	17.3	0.097	12	0.207	6.2	0.069
	16	0.191	13	0.309			12	0.083	13	0.069	18.1	0.069			17	0.279	19	0.111	18.3	0.069	13	0.069	7	0.069
	17	0.162	13.1	0.069			13	0.069	14	0.153					18	0.191	20	0.069	19	0.069	14	0.086	7.2	0.069
	18	0.132	14	0.118			14	0.069	15	0.069					19	0.176	21	0.069	20	0.069	15	0.069	8	0.069
	19	0.069	15	0.088					16	0.069					20	0.069			21	0.069			9	0.069
	20	0.088	15.1	0.069					17	0.069					21	0.088			22	0.069			10	0.069
	21	0.069	18	0.069					17.2	0.069					22	0.069			23	0.069			11.2	0.069
	22	0.069							18	0.069									24	0.153			12	0.069
									19	0.069									25	0.153			12.2	0.069
									19.2	0.069									26	0.069			13.2	0.069
									20	0.069									27	0.069			14.2	0.069
									21	0.069									28	0.069				
									22.2	0.069									29	0.069				
																			32.2	0.069				
																			33	0.069				
																			35	0.069				
																			40.2	0.069				
																			42.2	0.069				

A4.5.6 Sub-population Msi GI

Msi GI population Match Probability Calc. Sheet

DNA profile	MsF5s	MsF9	MsF16	MsF18	MS4h	MsF27s	MS9	MsF2
Allele 1								
Allele 2								

1. Standard profile match between

Theta value	0	0	0	0	0	0	0	0
Allele1								
Allele2								
P(het)	0	0	0	0	0	0	0	0
P(hom)	0	0	0	0	0	0	0	0
Appropriate Prob.	0	0	0	0	0	0	0	0
Cumulative P	0	0	0	0	0	0	0	0
Cum Product	0							

Cumulative Likelihood #DIV/0! **Rarest genotype freq:** 5/(2*20)
0.125
2.7105E-20
3.6893E+19 **Rarest genotype should be 1 in 3.69 x 10^{19}**

Most common genotype frequency: 1.3268E-10 or the inverse: 7536705446 **Most common genotype is 1 in 7.46 x 10^{9}**

		0.225		0.213		0.113		#NUM!		0.238		0.230		#NUM!		0.439		0.203		#NUM!		#NUM!		0.158
Most common allele freqs:	Allele1	0.900		0.425		0.250		1.000		0.950		0.921		1.000		0.675		0.675		1.000		1.000		0.350
	Allele2	0.125		0.250		0.225		#NUM!		0.125		0.125		#NUM!		0.325		0.150		#NUM!		#NUM!		0.225
Least common allele freqs:	Allele1	0.125		0.125		0.125		1.000		0.125		0.125		1.000		0.325		0.125		1.000		1.000		0.125
	Allele2	0.900		0.125		0.125		#NUM!		0.125		0.921		#NUM!		0.675		0.125		#NUM!		#NUM!		0.125

Locus	MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	14	0.900	19	0.125	22	0.250	4	1.000	12	0.950	8	0.921	14	1.000	13	0.325	14	0.675	10	1.000	13	1.000	19	0.125
	15	0.125	20	0.250	23	0.225			14	0.125	13	0.125			14	0.675	15	0.125					19.1	0.125
			21	0.125	25.3	0.200			15	0.125							17	0.125					20	0.350
			22	0.125	26.3	0.125											19	0.150					22.1	0.175
			23	0.425	28.3	0.175																	23.1	0.225
			24	0.125	29.3	0.125																	24.1	0.125
																							25.1	0.125

	0.360		0.135		#NUM!		#NUM!		0.315		#NUM!		#NUM!		0.300		0.193		0.146		0.249		#NUM!	
	0.600		0.450		1.000		1.000		0.525		1.000		1.000		0.400		0.550		0.325		0.789		1.000	
	0.300		0.150		#NUM!		#NUM!		0.300		#NUM!		#NUM!		0.375		0.175		0.225		0.158		#NUM!	
	0.125		0.100		1.000		1.000		0.175		1.000		1.000		0.125		0.125		0.125		0.125		1.000	
	0.125		0.125		#NUM!		#NUM!		0.300		#NUM!		#NUM!		0.125		0.150		0.125		0.125		#NUM!	

	MS4h		MsF27s		MsF8		MsF26		MsF4		MsF27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	17	0.125	11	0.450	12	1.000	7	1.000	13	0.300	16.1	1.000	7	1.000	16	0.400	16	0.150	15.3	0.225	9	0.125	5.2	1.000
	18	0.600	11.1	0.125					14	0.175					17	0.125	17	0.175	16.3	0.125	11	0.789		
	19	0.300	12	0.150					15	0.525					18	0.125	18	0.550	18.3	0.125	11.2	0.125		
	23	0.125	12.1	0.125											19	0.375	19	0.125	19	0.325	12	0.158		
	24	0.125	13	0.150															20	0.125				
			13.1	0.100															21	0.150				
																			22	0.125				

A4.5.7 Sub-population Msi SFI

Msi SFI population Match Probability Calc. Sheet

DNA profile	MsF5s	MsF9	MsF16	MsF3	MsF18	MsF22	MS4h	MsF8	MsF26	MS9
Allele 1										
Allele 2										

1. Standard profile match between

Theta value	0	0	0	0	0	0	0	0	0	0
Allele1										
Allele2										
P(het)	0	0	0	0	0	0	0	0	0	0
P(hom)	0	0	0	0	0	0	0	0	0	0
Appropriate Prob.	0	0	0	0	0	0	1	1	1	1
Cumulative P	0	0	0	0	0	0	0	0	0	0
Cum Product	0									

Cumulative Likelihood #DIV/0!

Rarest genotype freq: 5/(2*18)
0.139
4.1951E-19
2.3837E+18 **Rarest genotype should be 1 in 2.38 x 10^{18}**

Most common genotype frequency: 3.6077E-11 or the inverse: 2.7718E+10 **Most common genotype is 1 in 2.77 x 10^{10}**

		0.232		0.293		0.242		#NUM!		0.346		0.270		0.255		#NUM!		0.241		#NUM!		0.270		0.261
Most common allele freqs:	Allele1	0.833		0.528		0.412		1.000		0.444		0.971		0.917		1.000		0.722		1.000		0.694		0.361
	Allele2	0.139		0.278		0.294		#NUM!		0.389		0.139		0.139		#NUM!		0.167		#NUM!		0.194		0.361
Least common allele freqs:	Allele1	0.139		0.139		0.139		1.000		0.139		0.139		0.139		1.000		0.139		1.000		0.139		0.139
	Allele2	0.139		0.139		0.139		#NUM!		0.139		0.971		0.917		#NUM!		0.167		#NUM!		0.194		0.139

Locus	MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	13	0.139	17	0.139	22.3	0.412	4	1.000	11	0.139	7	0.139	15	0.917	8	1.000	14	0.139	12	1.000	13	0.694	14	0.139
	14	0.833	19	0.528	23.3	0.147			12	0.389	8	0.971	16	0.139			15	0.722			14	0.194	15	0.361
	15	0.139	20	0.278	24.3	0.139			13	0.444							16	0.167			15	0.139	16	0.222
			21	0.139	25.3	0.139			14	0.139													17	0.361
			23	0.139	26.3	0.294																	18	0.139

	0.323		0.389		0.313		0.321		0.270		#NUM!		#NUM!		0.208		0.401		0.173		0.410		0.270	
	0.528		0.500		0.806		0.722		0.972		1.000		1.000		0.353		0.556		0.389		0.656		0.972	
	0.306		0.389		0.194		0.222		0.139		#NUM!		#NUM!		0.294		0.361		0.222		0.313		0.139	
	0.139		0.139		0.194		0.139		0.139		1.000		1.000		0.139		0.139		0.139		0.139		0.139	
	0.139		0.389		0.806		0.222		0.972		#NUM!		#NUM!		0.139		0.139		0.139		0.313		0.972	

	MS4h		MsF27s		MsF8		MsF26		MsF4		MS27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
	15	0.139	11	0.139	12	0.806	10.3	0.139	10.3	0.139	16.1	1.000	7	1.000	16	0.294	15	0.361	29.2	0.139	10	0.139	6.2	0.972
	16	0.306	12	0.500	13	0.194	11	0.722	11	0.972					17	0.139	16	0.556	30.2	0.139	11	0.656	10.3	0.139
	17	0.528	13	0.389			12	0.222							18	0.235	17	0.139	31.2	0.139	12	0.313		
	18	0.139													19	0.353	19	0.139	32.2	0.167				
	20	0.139													20	0.139			33.2	0.389				
																			34.2	0.222				

A4.5.8 Example calculation – SIB individual 71370

SIB population Match Probability Calc. Sheet

DNA profile	MsF5s	MS25	MsF33	MsF31	MsF9	MsF16	MsF3	MsF17	MsF18	MsF28si	MsF22	MS3	MS4h	MsF27s	MsF8	MsF26	MsF4	MS27	MsF24	MS9	MsF2	MS17	MsF15	MS13
Allele 1	13	22.2	-	8	14	7	16	16.1	18	-	-	19	18	12	12	-	18	12.1	-	18	13	15	11	8
Allele 2	13	22.2	-	9	15	8	19	17.1	18	-	-	21	21	13	13	-	19	12.1	-	18	16	15	16	9

1. Standard profile match between

Theta value	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149	0.149
Allele1	0.277	0.196	-	0.286	0.295	0.364	0.179	0.116	0.188	-	-	0.107	0.245	0.213	0.348	-	0.134	0.491	-	0.295	0.045	0.179	0.127	0.241
Allele2	0.277	0.196	-	0.161	0.250	0.045	0.125	0.071	0.188	-	-	0.045	0.045	0.139	0.313	-	0.116	0.491	-	0.295	0.250	0.179	0.078	0.607
P(het)	0.19830654	0.13404617	#VALUE!	0.15027384	0.19391881	0.11514913	0.10306872	0.06970632	0.12767961	#VALUE!	#VALUE!	0.060324	0.089812	0.118326	0.247797	#VALUE!	0.087379	0.430972	#VALUE!	0.214283	0.090859	0.121468	0.074488	0.316145
P(hom)	0.2441813	0.19155449	#VALUE!	0.25040961	0.2567223	0.30811007	0.18071	0.14519922	0.18609352	#VALUE!	#VALUE!	0.140435	0.222679	0.201867	0.296217	#VALUE!	0.15496	0.415168	#VALUE!	0.256722	0.109429	0.18071	0.151383	0.220014
Appropriate Prob.	0.2441813	0.19155449	1	0.15027384	0.19391881	0.11514913	0.10306872	0.06970632	0.18609352	1	1	0.060324	0.089812	0.118326	0.247797	1	0.087379	0.415168	1	0.256722	0.090859	0.18071	0.074488	0.316145
Cumulative P	0.2441813	0.04677402	0.046774023	0.00702891	0.00136304	0.00015695	1.6177E-05	1.1276E-06	2.0985E-07	2.1E-07	2.1E-07	1.27E-08	1.14E-09	1.35E-10	3.33E-11	3.33E-11	2.91E-12	1.21E-12	1.21E-12	3.1E-13	2.82E-14	5.1E-15	3.8E-16	1.2E-16
Cum Product	1.2004E-16																							

Cumulative Likelihood **8.3307E+15** **Rarest genotype freq:** 5/(2*56)

0.045

6.4112E-32

1.5598E+31 **Rarest genotype should be 1 in 1.55 x 10^{31}**

Most common genotype frequency: 2.4122E-22 or the inverse: 4.1455E+21 **Most common genotype is 1 in 4.14 x 10^{21}**

Most common allele freqs:	Allele1	0.138		0.067		0.039		0.143		0.158		0.172		0.091		0.110		0.095		0.192		0.167		0.047
	Allele1	0.277		0.196		0.155		0.286		0.295		0.364		0.232		0.473		0.232		0.849		0.400		0.188
	Allele2	0.250		0.170		0.127		0.250		0.268		0.236		0.196		0.116		0.205		0.113		0.209		0.125
Least common allele freqs:	Allele1	0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045		0.045
	Allele2	0.045		0.045		0.045		0.071		0.045		0.045		0.045		0.045		0.045		0.113		0.045		0.045

Locus

MsF5s		MS25		MsF33		MsF31		MsF9		MsF16		MsF3		MsF17		MsF18		MsF28si		MsF22		MS3	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
7	0.107	16.3	0.045	22.3	0.045	4	0.125	11	0.045	3	0.073	12	0.045	9.1	0.473	10	0.045	9	0.849	11	0.109	11	0.045
10	0.045	18.2	0.045	23.3	0.045	6	0.089	12	0.045	7	0.364	13	0.045	10.1	0.045	11	0.045	10	0.045	12	0.045	12	0.045
11	0.045	19.2	0.071	24	0.045	7	0.250	13	0.268	8	0.045	14	0.107	12.1	0.045	13	0.116	11	0.113	13	0.400	13	0.098
12	0.214	20.2	0.107	24.3	0.127	8	0.286	14	0.295	9	0.091	15	0.196	13	0.045	14	0.063			13.3	0.045	14	0.054
12.1	0.045	21.2	0.107	25.3	0.155	9	0.161	15	0.250	10	0.236	16	0.179	13.1	0.045	15	0.232			14	0.209	15	0.080
13	0.277	22.2	0.196	26.3	0.100	10	0.071	16	0.098	10.1	0.045	17	0.232	14	0.045	16	0.188			15	0.045	16	0.125
14	0.250	22.3	0.045	27.2	0.045	11.2	0.045	17	0.063	11	0.055	18	0.045	14.1	0.045	17	0.205			16	0.055	17	0.188
15	0.054	23.2	0.170	27.3	0.082					12	0.136	19	0.125	15	0.045	18	0.152			17	0.118	18	0.080
16	0.045	23.3	0.045	28.2	0.045							20	0.045	15.1	0.054	19	0.045			19	0.045	19	0.107
		24.2	0.134	28.3	0.055							21	0.045	16	0.045	21	0.045					20	0.045
		24.3	0.045	29	0.045									16.1	0.116							20.1	0.045
		25.2	0.071	29.2	0.045									17	0.045							21	0.045
		26.2	0.045	29.3	0.045									17.1	0.071							21.1	0.045
		28.2	0.045	30.2	0.045									18	0.045							22	0.045
		29.2	0.045	30.3	0.045									18.1	0.054							23	0.045
		32.1	0.045	31.2	0.045									19	0.045							24	0.045
				31.3	0.045										0.045							27	0.045
				32.1	0.045																	27.2	0.045
				32.2	0.045																	29	0.045
				32.3	0.045																	29.2	0.045
				33.2	0.045																	34	0.045
				34.1	0.045																	36	0.045
				34.2	0.045																	36.3	0.045
				35.2	0.055																	37	0.045
				36.1	0.045																	38.1	0.045
				38.2	0.045																	39	0.045
			24.2*		0.045																	40	0.045
	0.139		0.154		0.218		0.169		0.099		0.333		0.158		0.142		0.098		0.067		0.106		0.293
	0.284		0.361		0.348		0.545		0.241		0.491		0.791		0.295		0.250		0.188		0.235		0.607
	0.245		0.213		0.313		0.155		0.205		0.339		0.100		0.241		0.196		0.179		0.225		0.241
	0.045		0.045		0.045		0.045		0.045		0.071		0.045		0.045		0.045		0.045		0.045		0.045
	0.045		0.045		0.063		0.045		0.045		0.098		0.045		0.045		0.045		0.045		0.045		0.045

MS4h		MsF27s		MsF8		MsF26		MsF4		MS27		MsF24		MS9		MsF2		MS17		MsF15		MS13	
Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency	Allele	Frequency
11	0.045	7	0.045	10	0.063	6	0.045	10	0.045	11.1	0.098	0.2	0.045	14	0.054	12	0.080	2.3	0.045	10	0.045	7	0.045
12	0.045	8	0.045	11	0.089	10	0.045	11	0.071	12.1	0.491	6	0.100	15	0.107	13	0.045	3.3	0.045	11	0.127	7.3	0.054
13	0.045	10	0.045	12	0.348	11	0.545	12	0.045	13.3	0.339	6.3	0.045	16	0.241	14	0.054	9	0.045	12	0.078	8	0.241
14	0.059	11	0.361	13	0.313	12	0.155	13	0.045	15	0.071	7	0.791	17	0.116	15	0.170	10	0.045	13	0.225	9	0.607
14.2	0.045	12	0.213	14	0.179	13	0.109	14	0.054			8	0.082	18	0.295	16	0.250	11	0.063	14	0.235	9.3	0.045
15	0.078	12.1	0.045	15	0.045	14	0.045	15	0.045					19	0.134	17	0.196	12	0.080	15	0.147	10	0.063
16	0.127	13	0.139			15	0.073	16	0.134					20	0.045	18	0.143	13	0.143	16	0.078		
17	0.284	13.1	0.045					17	0.205					21	0.045	19	0.045	14	0.188	17	0.049		
18	0.245	14	0.130					18	0.241							20	0.045	15	0.179	17.3	0.045		
19	0.078	14.1	0.045					19	0.116							38.2	0.045	16	0.143	18	0.045		
21	0.045	15	0.045					20	0.045									16.2	0.045	19	0.045		
		16	0.045					21	0.045									17	0.054				
		19	0.045					22	0.045									17.2	0.045				
																		18.2	0.045				
																		19	0.045				
																		20.1	0.045				

CHAPTER 5

Concluding Remarks

5.1 What the PhD has achieved

This PhD project set out to design forensic DNA assays for an Australian species of national enforcement significance to be applied directly to criminal casework. The choice of species was dictated by the needs of local wildlife enforcement officers and their counterparts in other Australian states, and also championed by the Australasian Environmental Law Enforcers and Regulators Network (AELERT). This work encompassed the two techniques most frequently used by forensic biologists to answer the questions commonly encountered during investigations of crimes involving wildlife species.

5.1.1 Mitochondrial DNA assay for species identification

Due to the simplicity of the design and application of species identification tests, as well as legislation often dealing with the *category* of species, this question most commonly arises in routine wildlife DNA forensic casework. A mitochondrial species identification test has been developed that allows the amplification and differentiation of every species of python in the genus *Morelia*. Often forensic cases are dealing with trace amounts of DNA and this is no different in the case of snakes (e.g. Australian Museum case involving DNA identification of a Common Death Adder from a swab of a cardboard box, R. Johnson, *pers. comm.*). The developed species identification test for *Morelia* pythons targets a short region of DNA, ideal for obtaining a result from tiny amounts of degraded or fragmented DNA. The validation study has shown this test to be highly sensitive, amplifying product from as little as 30 femtograms of DNA template. A published developmental validation permits this test to be applied by wildlife forensic scientists directly to criminal investigations with verification required to demonstrate the boundaries of the test within their own laboratory.

At the outset, this test involved a familiarisation with the structure of the mitochondrial genome and the different gene orders that can arise in quite closely related taxa. The crux of a well-designed species identification test is that the locus should be able to group all members of a species to the exclusion of the members of all other species. In terms of sequence similarity, when pairwise comparisons of the target region are made an ideal test should provide a reasonable gap in percent homology between the comparison of the two least similar members of the same species and the comparison of a member of the target species with that of the most closely related sister species. The rather arbitrary definition of 98% similarity has been adopted as the recognised threshold for sequence comparisons (as applied by BOLD, boldsystems.org, accessed 22/07/2018) – this figure can provide accurate species differentiation for many, but certainly not all taxa. A thorough validation should demonstrate this gap and include all species within a genus where possible, to give the reporting officer confidence to report the identification of a species to the exclusion of all other possible species, based on the degree of homology of their unknown sequence with the most closely matching species on the database. The long term advantages of designing a test from the bottom up and including all relevant species, as opposed to the historical and quick top down approach of including just the case relevant species when a need arises were discussed.

The design of the species identification test requires careful consideration. Is the test designed for universal amplification, where sequence analysis and phylogenetic reconstruction is used to reveal the species, or is it intended as a presence/absence assay? A universal test requires that primers are placed in regions highly conserved across widely ranging taxa and that the area bridging the primers is highly variable between all individuals except those of the same species. Yet, some loci are too

rapidly mutating, causing loss of the genetic signal that reveals the phylogenetic relationships. As phylogenetic relationships of species differ temporally, there is no one panacea or magical locus to suit all species (as much as the Barcode of Life initiative would like this to be the case, discussed in section 2.6.2).

Chapter two examined the degree of homology in each mitochondrial gene across a diverse range (all that were available at the time) of species of the suborder Serpentes. This allowed a gene region suitable for differentiating python species to be identified; a region that contains priming sites highly conserved within a chosen group of taxa (the family Pythonidae, for instance), but high levels of differentiation between species in the sequence that the priming sites flank. This facilitated the development of a species identification test to approximately amplify members of Pythonidae family to the exclusion of members of other families, as demonstrated by the specificity study within the validation testing. I use the term approximately because it is too idealistic to propose that a genetic test will be able to amplify every species within a family to the exclusion of every species falling outside of this family. Even within a species, it's possible that a small number of members will have mutations at the primer binding sites (for instance, falling on codon 'wobble' positions that happen to be in the 3' clamp of the primer) that preclude their amplification by that specific test. Species might also hold primer site mutations that preclude their amplification by the assay.

The assay demonstrated incredibly high sensitivity, successfully amplifying product from DNA template amounts in the tens of femtograms. The small forensic fragment also demonstrated successful amplification of difficult samples when other larger amplicons could not be obtained. This assay shows promise for analysis of degraded forensic samples such as trace DNA remains. The phylogenetic reconstruction from this tiny fragment remained true to the separation of all species with *Morelia* and grouping of conspecifics within each. The required gap was evident between the smallest intra-species pairwise comparisons and the largest inter-species comparisons for genetic distances of *Morelia* species, even at the most conservative p-distance model of nucleotide substitution (which is comparable to a simple nucleotide BLAST search).

Had the project had more time to focus on the mitochondrial assay, this work could be extended to identifying key SNP sites diagnostic of particular species. These SNPs could then be incorporated into a species-specific assay that indicates the presence or absence of a species. These tests are particularly useful for analysing samples likely to be involved a mixture of DNA templates, such as traditional medicines. If a single base extension test is designed, then it is well suited to fragmented and degraded remains. This aspect of the mitochondrial species identification development offers a logical future direction for this project (see section 2.6.7 for more discussion on this topic).

5.1.2 Development, optimisation and validation of three STR multiplex assays

Chapters three and four dealt with this main focus of the PhD project, the development, validation and potential for application of STR multiplexes to generate DNA profiles to answer questions regarding individualisation, paternity and ultimately geographic provenance of a poached sample. The degree of time and resource required to develop this type of test for a wildlife species (given that each species but the most closely related requires a unique test developed) dictated that this task would consume the bulk of time spent on this doctoral project. However, these tests are still relatively rare and having the skills to develop and apply such a test for wildlife species will be a crucial asset for the wildlife forensic biologist for many years to come.

The DNA profiling assay developed in this study was a result of a need for forensic tests to aid investigations of a species commonly involved in wildlife crimes, the Australasian Carpet Python (*Morelia spilota*). Previous work had suggested that this species formed a fairly continuous distribution along the Northern and Eastern coast of Australia. Eight STR markers previously developed could not provide the genetic resolution to allow identification of a putative population of origin of an individual suspected to have been taken from the wild. At the outset of this work, the research intent was a forensic assay designed to the high standards required for the scrutiny of the judicial system. A compliment of 24 STR markers was designed to profile the carpet python; a marker set equal in potential discriminatory power to current commercial human DNA profiling kits. Design of the markers included in the kit was directed by the need to have higher discriminatory power in carpet pythons for the intent of differentiating different populations across small geographic areas. Indeed, markers with compound and complex underlying repeat motifs were intentionally included to increase the allelic diversity of the dataset. This was intended to improve the capability of the test for application to statistical analyses regarding increased discrimination between individuals (as required for paternity analyses), as well as better determination of population of origin to within smaller geographic ranges. This led to an unanticipated result of the large sets of alleles at some loci confounding the accurate estimation of alleles in the small island populations.

Putative marker identification has become much less resource intensive with the developments in whole genome sequencing technology, making this development of this type of technique much more accessible. The random acquirement of fragments through shotgun sequencing also yields a much better variety of putative repeat motif types and lengths. Traditional target enrichment and cloning allowed identification of loci containing a single motif, whereas a variety of motifs have been obtained during this project and over a much shorter timeframe. This has allowed time to be spent selecting the loci that appear most suitable for forensic application, such as tetra- and penta-nucleotide motifs and the inclusion of some complex loci allowing greater numbers of alleles to be obtained at a locus. As with the mitochondrial work, having the time to carefully plan the multiplex assays to overcome potential issues of incompatibility or cross-compatibility of primers and overlap of loci once more widespread screening expanded the locus allele ranges, has allowed for robust assays to be developed with minimal resources spent on redesigning and reordering expensive fluorescently labelled primers.

Highly optimised commercially available polymerases and buffers enabled large multiplexes to amplify a full complement of loci relatively quickly and only a small number of optimisation rounds were needed in order to reliably obtain fairly balanced DNA profiles. As the conditions of the multiplexes were modelled closely on those of human forensic kits (e.g. annealing temperature, cycle number, starting DNA template amount) the results are relatively comparable to the quality of profiles obtained from similar amounts of starting DNA in human forensic casework, using the same equipment.

Ladder creation for each multiplex was a laborious step of identifying heterozygote individuals containing unique and where possible non-overlapping alleles, combining similar DNA template concentrations of these individuals and amplifying by multiplex reaction, reamplifying individuals that gave weak results and pooling the products with the original amplification product, etc. until each locus was well represented with strong products representing the most commonly seen alleles.

This was nonetheless a crucial exercise to be able to offer reproducibility of genotyping results to single base pair resolution between independent laboratories.

Sequencing to obtain the underlying nature of the alleles not only confirmed different alleles believed to originate from the same locus to do so, but also revealed interesting patterns of homoplasy that, upon development of mass parallel sequencing facilitated genotyping by sequencing, has the promise to allow increased differentiation between individual genotypes even though alleles appear identical by size migration. The improvement of the multiplex assays into optimised primer mixes facilitated straightforward and quick screening and genotyping of large numbers of individuals to construct the genotype database that forms the basis of chapter four. The validation testing that was performed also suggested the multiplex assays to be quite robust and reproducible, offering an excellent level of sensitivity and some possibility for cross-species amplification in other Australasian python species of interest to enforcement. The concluding section of chapter three demonstrated that the resulting reference database in its entirety was unsuitable to direct application to forensic casework.

5.1.3 Population genetic analysis of carpet pythons and validation of the reference database for forensic application

Chapter four was the crucial part of the investigation where the dataset started to prove problematic for direct application to casework. It raised some interesting yet difficult questions to answer regarding the best use of less than ideal results and whether presenting information with identifiable caveats (flaws even) is better than presenting nothing at all.

Population genetic analysis was required in order to break the dataset down into more meaningful and useful populations. Interpreting the results of the analyses was not as straightforward as originally anticipated. The various populations involved offered different challenges to working with the dataset. A predominant difficulty with this dataset was the sample size of roughly 300 individuals, spread over a very large geographic areas with many physical barriers asides distance to increase genetic structure. Outside of the east coast, many individuals likely represented whole populations in their genetic divergence, confounding population genetic analyses. The groups of genetically similar individuals seen were often very small. However, this complication is a reality faced when dealing with the rare species often targeted in wildlife crime.

The large number of highly polymorphic loci chosen exhibiting large numbers of micro-variant alleles offers the advantage of high discrimination power required for fine scale population assignment. Yet, an abundance of rare alleles created unanticipated problems when attempting to accurately estimate allele frequencies in small populations, where the number of sampled alleles for the reference database was far less than the alleles known to exist at a particular locus. One could argue that the actual number of alleles at this locus is far less in small populations and so the rare alleles are insignificant. To take a conservative approach, the number of loci used was restricted to only the most conserved loci when calculating forensic statistics for the island populations.

To return to the issue of sample numbers, the high geographic structuring seen in the dataset caused difficulties in defining interbreeding populations with confidence. Large populations would include individuals so genetically divergent that the population would not adhere to Hardy-Weinberg Equilibrium, an important assumption for the population genetic analyses and forensic statistical

conclusions that are the critical endpoint of the work. Allele frequencies would be accurate for only a small subset of the chosen population. Yet, identifying populations that adhered to HWE resulted in population numbers too small for accurate population genetic analyses and F-statistics to be calculated, likewise confounding the statistical calculations reported to the courts. The various issues with the dataset were discussed and a range of approaches offered for reporting forensic statistics in a case to mitigate the problems encountered. In the field of wildlife forensic science, where datasets are likely to contain these problems, it is desirable to present a conservative approach accompanied by caveats and treated with transparency, rather than to despair at the difficult issues encountered and present no evidence at all.

Asides the main objective of DNA profiling by size migration, the sequencing of alleles identified some interesting sequence polymorphisms that could provide added utility in casework involving carpet pythons from particular populations. The *Morelia spilota imbricata* subspecies contained an indel at locus MS27 not seen in any other population. Locus MS13 exhibited another indel specific to Morelia bredli and absent from M. s. imbricata. These loci could provide extra exclusionary information about these species and potentially assist questions of hybridisation involving individuals of these populations, an issue that is difficult to investigate by genotyping alone. Locus MsF28 could also provide extra benefit to hybridisation investigations, as discussed in section 4.6.10. Lastly, the STR in the primer sequence at locus MsF14 was not examined in detail in this study, but provides additional prospects for population discrimination and provides an avenue of further investigation in future studies of this species.

This PhD project has successfully developed and optimised a highly discriminating DNA profiling set for the carpet python to a quality standard rarely seen in forensic STR development for non-human species. Three allelic ladders are included to aid reliable genotyping to single base pair resolution. The population genetic makeup of the target species had presented various difficulties in partitioning the reference dataset into appropriate sub-population databases for calculating forensic statistics. Various approaches have been discussed for drawing statistical conclusions from difficult populations that have small sample sizes and high genetic structure.

Validation studies have shown the assays to be robust and specific to the target species with some scope for extension to other closely related Australasian pythons of forensic significance. The assay is sensitive, requiring only a small amount of starting DNA template in line with commercial forensic DNA profiling kits. A highlight of the results has been the ability of geographic assignment tests to show great power in accurate assignment to the correct source population or exclusion from non-source populations. Six well defined populations were identified within the carpet python's native range and only one overlapping interface between two of these caused an issue for robust assignment. Assignment was seen to be robust for some populations down to only five loci, offering much potential for investigations involving degraded remains. The resulting forensic statistics gave likelihood ratios valuable for drawing strong conclusions about the forensic evidence, most examples having an associated statistic of 1 in 1 million or higher.

5.2 Future applications of the developed tests

By their nature, the transferability of mitochondrial tests for species identification is straightforward. The primers developed in the above study are widely accessible through publication, the reference DNA sequences for the samples examined herein are available for download from Genbank

(Accession Numbers MG548390-MG548510) and the sequencing results are comparable between laboratories without the need for an internal size standard. International inter-laboratory validation of a mitochondrial test has been conducted elsewhere [1] and it would be a simple matter of communication and collaboration between laboratories to make this happen again. For an inter-laboratory comparison of results, snake extracts have been retained at the Flinders University Forensic DNA Laboratory and these could easily be provided to other laboratories within Australia, such as the Australian Museum's Australian Centre for Wildlife Genomics, to reproduce the results independently.

The inter-laboratory validation of multiplex STR tests is not quite so straightforward; more co-ordination is required as a larger amount of materials needs to be shared. Chapter 3, section 3.6.3.4 discussed the methodology to conduct a small reproducibility study with an independent laboratory, while Chapter 3, section 3.6.3.11 and the manuscript Ciavaglia & Linacre 2018, Section 3.8 built on this by describing what would be required to expand the undertaking into an inter-laboratory collaborative exercise. All of the fluorescently labelled primer stocks have been retained at Flinders University and these can be supplied to collaborating laboratories, along with aliquots of each of the three allelic ladders. Snake extracts can be supplied as positive controls and for blind-trial testing to ensure consistent genotyping between laboratories. The allele frequency data has been published, but if collaborating labs wanted to reanalyse this themselves then the genotype data for the snakes genotyped during this project could also be shared. In order to apply this assay to casework, the genetic database would benefit from addition sampling to fill geographic areas that carpet pythons occupy, but for which few individuals were represented. Any lab intending to apply this assay in house would need to conduct their own internal verification and establish the lab specific limitations of the test, including the analytical threshold specific to their instrumentation.

Replicating the truly international breadth of Ewart and colleagues [1] would be more difficult for this study because all pythons are CITES appendix 2 listed. This would cause issues with cross-border sharing of blind-trial extracts and CITES import permits would be required for this to happen. Ewart and colleagues ingeniously created a synthetic DNA product for rhino to be used as a positive control for the mitochondrial species identification testing. An equivalent for STR testing would be very expensive, as amplification products at every locus of a snake would have to be synthesised – a prohibitively expensive undertaking. There could be scope for parentage testing of carpet pythons outside of Australia, if they are cross-bred in contravention of the local laws, warranting international standardisation of the test. However, the geographic provenance aspect of this test is only relevant New Guinea, as this is the only range state of *Morelia spilota* outside of Australia.

In terms of facilitating collaboration between laboratories internationally, the Society of Wildlife Forensic Science is instrumental to this cause. The Society is the premier means of contact between the international wildlife forensic science community and the bi-annual conferences enable regular discussions between laboratories that would be likely to use these tests. I am already familiar with international validation testing as a collaborator of the rhino study (Ewart *et al.*) and my involvement in the Society has made organisation of this type of multi-country undertaking much easier to conduct. While it is not within the scope of my current role to conduct an international inter-laboratory validation of the *OzPythonPlex* multiplex, I would certainly be keen to facilitate this if other laboratories had a need for applying the test to casework and expressed an interest to do this sort of testing.

5.3 My reflections

Overall, the project has presented some themes that overarch the development of forensic DNA techniques. The importance of quality assurance at every stage of development cannot be underemphasised and quality measures can be built into the design of tests to detect non-conformances such as sample mix up when applied to forensic casework. Tests need to be well designed at the outset and investment in time for careful planning prior to development assists final test to be of the highest standard and appropriate for application to the justice system. Validation testing is a crucial part of the conclusion of test development, prior to application to casework. The project has shown that excellent STR assay design is not everything and less than ideal situations can arise depending on the species and their population history. The validation process helps to demonstrate the limitations of a test. Discussions with colleagues with experience in wildlife forensic science, human forensic DNA analysis and just as importantly the presentation of forensic statistics involved with DNA evidence, have been crucial in weighing up the different options and assessing the best way to deal with a given scenario.

There is admittedly a small degree of disappointment having invested so much time and resources in thoroughly developing a genotyping test for a species in need, only to find the resulting reference database to be less than ideal for application to criminal casework due to the limited sample numbers and high structuring of this species. Not all questions asked can be easily answered. Yet, the solace of this situation is that this scenario is likely to be true of many species targeted in wildlife crimes. It is the rarest species that are the most valuable to poachers and populations are likely to be highly fragmented and sparsely sampled. This project studying the carpet python has been ideal in providing a difficult scenario, which may in fact be the reality for other species encountered in casework. It has raised difficult questions of how to proceed and how much a test can offer the courts. These issues are not going to be unique to the carpet python. A study species offering the ideal population genetic outcome would have made for an easy PhD project, but far fewer lessons learnt.

In closing, I am the first to acknowledge that this dissertation is far more extensive than that required to gain a doctorate. Its conduct has now covered a fifth of my own lifetime and has sapped a lot of my energy. I do feel a different and much tireder person than when I began (although hopefully some of that energy will return after some time away from the computer screen). If I had the choice to go back and complete it full time in three years, I would not. If I had the choice to do anything differently, I would not. This project was never about becoming a Doctor; it was about gaining the necessary skills and knowledge to found a career in a specialist field of forensic science for which I am passionate.

The career arrived more quickly than the doctorate. However, there were many concepts about DNA interpretation and statistical reporting that I still did not understand with ten years of experience in the field, and only in the last couple of months of writing, have I reached a far better understanding of the results that we report and why we report them in the way that we do. I do not feel an expert in all areas of the project. Yet, I do believe that this project has left me far better equipped to tackle many aspects of wildlife forensic DNA test development, results analysis and reporting that I will encounter as my career progresses in this young and exciting field that is so important to the ongoing survival of many species with which we co-habit our planet.

5.4 References

[1] K.M. Ewart, G.J. Frankham, R. McEwing, L.M.I. Webster, S.A. Ciavaglia, A.M.T. Linacre, D.T. The, K. Ovouthan, R.N. Johnson, An internationally standardized species identification test for use on suspected seized rhinoceros horn in the illegal wildlife trade, Forensic Sci. Int. Genet. 32 (2018) 33–39.