Mitochondrial heteroplasmy in Australian native bee *Amphylaeus morosus* and its association with the parasite *Wolbachia*

Olivia K. Davies

Bachelor of Science (Honours)

Thesis
Submitted to Flinders University
for the degree of Doctor of Philosophy

College of Science and Engineering
Flinders University
Adelaide, Australia
27 August 2021

Table of Contents

Cover photo: female specimen of *Amphylaeus morosus* by James Dorey

Abstract

Mitochondrial heteroplasmy is the occurrence of more than one type of mitochondrial DNA (mtDNA) within a single cell or individual. It is reported haphazardly across animal taxa (and other multicellular life), but little is known about its origins or functions in an evolutionary context. Conducting a comprehensive literature review of mitochondrial heteroplasmy in animals revealed patterns in type and prevalence in some taxon groups. Ubiquitous heteroplasmy (where all individuals sampled were heteroplasmic) occurred most frequently in arthropods, but occurrences are documented across multiple phyla. I examined mitochondrial heteroplasmy in the mt-COI gene of the native bee species *Amphylaeus morosus* (Hymenoptera: Colletidae). Every individual sampled (n = 73) across its ~2,000 km range along Australia's eastern coast was heteroplasmic, with all individuals possessing the same two mitochondrial haplotypes. Next-generation sequencing results found variable nucleotide sites throughout the entire mitogenome, indicating the conserved presence of two distinct mitochondrial lineages (heteroplasmy). Furthermore, *A. morosus* was consistently infected with two supergroup A strains of the common insect endosymbiont, *Wolbachia* (Rickettsiales: Anaplasmataceae). I have proposed that consistent heteroplasmy in this species is maintained by a mutualism with co-inherited *Wolbachia* strains, as both types of genomes follow the same maternal-inheritance pathway. *Wolbachia* is known to cause selective sweeps on host mitogenomes, resulting in an overall lack of mtDNA variation in host populations. To explore this relationship further, I assessed the feasibility of traditional antibiotic treatment in this non-model host, to apply to investigations of the changes to mtDNA inheritance, post- *Wolbachia* infection. Although the infection was not completely removed within one generation, one of the *Wolbachia* strains showed consistent susceptibility to antibiotic treatment. Finally, to investigate whether the mtDNA uniformity across the *A. morosus* distribution was reflected in the nuclear DNA, genome-wide single nucleotide polymorphisms (SNPs) were generated for individuals across the bees' distribution using the DArTseq™ platform. Analyses produced strong evidence of genetic clustering between locations (and potential population structure), as opposed to a single, large population as would otherwise be suggested by the conserved mtDNA. This, in combination with habitat modelling showing well-connected habitats before European settlement, provide further evidence that the lack of mtDNA diversity is because of a past *Wolbachia*-induced selective sweep and not a population bottleneck. This analysis provides further, but indirect, evidence that *Wolbachia* has played a key role in the evolution of this host's unusual mtDNA traits.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university. I also certify that the thesis is an original piece of research and it has been written by me. No professional editing services were used during the construction or at the completion of this thesis. Any assistance that has been received for this work and the preparation of this thesis has been appropriately acknowledged. To the best of my knowledge and belief, this thesis does not contain any material previously published or written by another person except where due reference is made in text.

Olivia Kate Davies

List of Contributors

Recipient of RTP support and stipend from the Australian Government Research Training Program Scholarship and the Playford Trust PhD scholarship.

Research grants received for this project:

- Holsworth Wildlife Research Endowment, 2016 and 2018
- Linnean Society of New South Wales Joyce Vickery Fund, 2018
- Mark Mitchell Research Fund, 2016
- Ecological Society of Australia Student Research Award, 2016
- Royal Society of South Australia Small Research Grant, 2016
- Field Naturalists Society of South Australia Lirabenda Endowment Fund, 2016

Conference, development, and travel funding received for this project:

- Phil Carne Prize Finalist, Australian Entomological Society, 2019
- Society of Australian Systematic Biologists, Conference Student Travel Award, 2019
- Flinders College of Science and Engineering, Conference Travel Grant, 2019
- Flinders University Student Association (FUSA) Development Grant, 2016 and 2018
- Young Investigators Registration Reimbursement Award for Society for Molecular Biology and Evolution Conference, 2016

Supervisor	Michael P. Schwarz College of Science and Engineering Flinders University, Australia
Associated Supervisor	Mike G. Gardner College of Science and Engineering Flinders University, Australia
Adjunct Supervisor	Mark I. Stevens South Australian Museum Adelaide, Australia

Thesis Structure

This thesis consists of five chapters, each exploring a specific aspect of the complex mitochondrial-endosymbiont conundrum in *Amphylaeus morosus*. Each chapter is presented as an independently publishable unit, with appropriate introductory and supplementary material associated with each. At present, no chapters have been published. Because of the specificity of the literature associated with each research area addressed, the cited material is included in reference lists associated within its relevant chapter.

Co-Author Contributions

Division of labour in co-authored articles:

OKD – Olivia Davies, MPS – Michael Schwarz, MGG – Mike Gardner, MIS – Mark Stevens, JBD – James Dorey, TMB – Tessa Bradford, LRH – Lucas Hearn

Table 1: Co-author contributions break-down for all chapters presented in this thesis.

Chapter	**1**	**2**	**3**	**4**	**5**
Concept & design	OKD	OKD, MPS	OKD, JBD, MPS	OKD, MPS	OKD, MPS
Planning & implementation	OKD	OKD, MPS, TMB, MGG	OKD	OKD, MPS, MGG	OKD, LRH, MPS
Data collection	OKD	OKD, MPS, TMB	OKD, MPS, JBD	OKD	OKD, MPS, JBD, LRH
Analysis & interpretation	OKD, MPS	OKD, TMB, MPS, MGG, MIS	OKD, JBD	OKD	OKD, JBD, LRH, MPS
Writing the article	OKD, MPS	OKD, MPS	OKD, JBD, MPS, MIS	OKD, MPS, MIS	OKD, JBD, MPS
Overall responsibility	OKD	OKD	OKD	OKD	OKD

Chapter 1

Olivia K. Davies and Michael P. Schwarz

Manuscript: A review of mitochondrial heteroplasmy in animal systems

The candidate (OKD) was the primary author of the manuscript. OKD planned, collected, and analysed the data collated from the literature. Interpretation and critical appraisal of the data were performed by both authors. OKD wrote the manuscript which was critically reviewed by MPS.

Chapter 2

Olivia K. Davies, Michael G. Gardner, Mark I. Stevens, Tessa M. Bradford, and Michael P. Schwarz

Manuscript: Extensive and widespread mitochondrial heteroplasmy in a colletid bee, *Amphylaeus morosus*

The candidate (OKD) was the primary author of the manuscript. Specimens used for analyses were collected by OKD in conjunction with those collected by or with MPS and JBD (non-author). Planning of experimental approaches were developed by OKD with MPS, TMB, and MGG. All analyses were performed by OKD with advice from TMB. Interpretation and critical appraisal of these analyses were performed by all authors. The manuscript was written by OKD, and MPS and MIS critically reviewed the manuscript.

Chapter 3

Olivia K. Davies, James B. Dorey, Mark I. Stevens, and Michael P. Schwarz

Manuscript: Co-infection of *Wolbachia* in the Australian bee *Amphylaeus morosus* (Colletidae: Hylaeinae)

The candidate (OKD) was the primary author of the manuscript. Specimens used for analyses were collected by OKD in conjunction with those collected by or with MPS and JBD. All analyses were performed by OKD except for phylogenetic analyses performed by JBD. The manuscript was written by OKD, with phylogenetic protocols written by JBD. All authors critically reviewed the manuscript.

Chapter 4

Olivia K. Davies, Michael G. Gardner, Mark I. Stevens, and Michael P. Schwarz

Manuscript: Incomplete removal of the reproductive parasite *Wolbachia* from a non-model organism following antibiotic treatment

The candidate (OKD) was the primary author of the manuscript, conducting all specimen collections in association with MPS who collected additional control samples. Planning of the experiment was performed by OKD with MPS and MGG. Maintenance of the experiment, data collection and analyses were performed by OKD. The manuscript was written by OKD with critical review performed by MPS and MIS.

Chapter 5

Olivia K. Davies, James B. Dorey, Lucas R. Hearn, and Michael P. Schwarz

Manuscript: Population structure in a native bee species with no mitochondrial DNA variation?

The candidate (OKD) was the primary author of the manuscript. Specimens used for analyses were collected by OKD in conjunction with those collected by or with MPS, JBD, and LRH. Diversity Arrays Technology Ply Ltd was commissioned for data generation, and data analyses were performed by OKD, with advice from JBD. JBD performed the habitat modelling. The study was conceived and data interpreted by OKD, assisted by MPS, JBD, and LRH. The manuscript was written by OKD, with habitat modelling protocols written by JBD. The manuscript was critically reviewed by MPS and JBD.

List of Original Publications

Dorey, J. B., S. V. C. Groom, E. H. Freedman, C. S. Matthews, **O. K. Davies**, E. J. Deans, C. Rebola, M. I. Stevens, M. S. Y. Lee, and M. P. Schwarz. 2020. Radiation of tropical island bees and the role of phylogenetic niche conservatism as an important driver of biodiversity. Proceeding of the Royal Society B: Biological Sciences **287**:20200045.

Draper, J. T., T. Haigh, O. Atakan, D. T. Limgenco, T. Kearney, L. Taylor, J. Wong, E. Kalderovskis, M. Tuiwawa, **O. K. Davies**, M. I. Stevens, and M. P. Schwarz. 2021. Extreme host range in an insular bee supports the super-generalist hypothesis with implications for both weed invasion and crop pollination. Arthropod-Plant Interactions **15**:13–22.

Hearn, L. R., K. A. Williams, M. I. Stevens, M. P. Schwarz, **O. K. Davies**, and B. A. Parslow. 2019. Description and novel host records for a new species of Australian mutillid wasp (Hymenoptera: Mutillidae) from hylaeine bee nests (Hymenoptera: Colletidae). Austral Entomology **58**:524–532.

Acknowledgments

There are endless people who deserve recognition for supporting the production of this tome. Firstly, thank you to my supervisors M^3 (Mike Schwarz, Mark Stevens, and Mike Gardner). Particularly, thank you Mike S for your encouragement, experience, support, thoughtfulness, and all the dog cuddles. Thank you also to Tessa Bradford whose patience is endless. And to the people at Flinders. The brilliant staff who have guided me over years. My teaching friends and heroes. And especially the Office of Graduate Research and Tara Brabazon. Your belief in me made all the difference.

I extend a tremendous amount of appreciation to the societies and organizations who provided funding for this project. Thank you to the Playford Trust, Holsworth Wildlife Trust, Mark Mitchell Research Fund, Ecological Society of Australia, Royal Society of South Australia, Field Naturalists Society of South Australia, and the Linnean Society of New South Wales. Also thank you to the Australian Entomological Society, Society of Australian Systematic Biologists, and Genetics Society of Australasia for their continued interest in this research and support for conferences across our small continent. To the wonderful people involved in these incredible communities, who see the value in PhD research and meet the financial gap in ecological, evolutionary, and wildlife research, you are amazing and inspiring. Thank you also to the Australian Government for providing continuous and generous financial support for my PhD journey. This collective support and the opportunities they enabled were a once in lifetime experience. Thank you.

Sincerely, thank you to James Dorey, my partner in crime, triumphs, and tantrums. Of all the humans I have met, you are my favourite. Thank you to my main bee (and allies) crew, Ben Parslow, Lucas Hearn, Justin Holder, and Rebecca Dew (and James). Your passion, creativeness, enthusiasm, and insightfulness has kept me afloat. I will dearly miss our fieldwork adventures (and crises), coffee dates, tacos, horror movies, and office distractions. I wish you all the best things in life and will see you again soon.

Thank you to Robbie O'Reilly, Carmel Maher, Matt Elmer, Jess Buss, Jess Clayton, Celina Rebola, Stefan Caddy-Retalic, Lettee Dametto, Amy Slender, and Amy Butler. You top-notch humans are the best listeners, innovators, and wisemen/wisewomen a gal could ask for. Your interruptions were what I looked most forward to in the day. To the rest of the LEGS crew, thank you for your hard-work, chaos, and determination. You are the most wonderful group of people and I wish you the absolute best. Thank you also to the brilliant Adelaide University crew for adopting a stray like me and providing me a safe h(e)aven away from home.

And finally thank you to my family (both by birth lottery and financial limitations). Thank you to my amazing parents, Mandy and Martin. My siblings/sufferers, Bella and Joe. Thanks for giving me just enough "experience" to make me funny (well, sarcastic), but not enough to be totally ~~insane~~ — insane enough to do *waves vaguely*... this. Bless you all for weirdly just ~believing~ in me. I love you and thank you for always supporting my strange interests. To my housemates, I am so appreciative of our housemate dinners, Christmases, impromptu band nights, and ridiculous ghost/zombie/alien stories. I have had a blast… and Halsbury was undoubtedly haunted [RIP]. Finally, thank you to all my animal hostages, for keeping me feeling loved and purposeful. Mavrick and Yuri, you may not be able to read my gratitude's, but you can be unwillingly smothered with affection, and for me, that is so much better.

General Introduction

Animal mitogenomes are relatively small (~16–20 kb), circular molecules exhibiting little recombination, and are mostly inherited through maternal transmission (Boore 1999). Primarily associated with the production of ATP, multiple mitochondria, each with many copies of mitochondrial DNA (mtDNA) can occupy an animal cell — therefore, copy number of mtDNA per somatic cell can be ~50–3,500 compared to two copies of nuclear DNA (Satoh and Kuroiwa 1991, Cao et al. 2007). Animal mtDNA generally contains about 37 genes, including 13 protein-coding genes involved in oxidative phosphorylation, plus approximately two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes (Zhang and Broughton 2013). Additionally, these genomes lack DNA repair mechanisms and can accumulate mutations 10–50 times faster than nuclear DNA (nuDNA) (Lynch 2007, Martin 2011), but have generally be assumed to play limited roles in animal evolution (Barr et al. 2005).

Mitochondrial genetic variation within a species is generally reported between individuals, and instances of mtDNA variation *within individuals* (mitochondrial heteroplasmy) are reported relatively rarely. This characteristic of mtDNA is somewhat perplexing, because if copy numbers and mutation rates are high and repair mechanisms are negligible, then individuals might be expected to accumulate multiple different copies of mtDNA over time (Barr et al. 2005). Once thought to be uncommon, examples of mitochondrial heteroplasmy in animals are increasingly reported across the literature, as broad taxonomic sampling and sequencing technologies become more accessible and reliable. This might indicate that mitochondrial heteroplasmy is more prevalent than currently appreciated. Yet the drivers behind mitochondrial heteroplasmy, particularly pervasive forms where large numbers of individuals within populations appear heteroplasmic, are poorly described and even less understood.

Wolbachia (Rickettsiales: Anaplasmataceae) is a genus of intracellular bacterium that commonly infects terrestrial arthropods. *Wolbachia* is arguably one of the world's most successful reproductive parasites — estimated to infect up to 60% of all terrestrial insect species (Clark et al. 2003) and can have profound impacts on its hosts. For example, *Wolbachia* infections can induce selective sweeps on its host's co-occurring mitogenome, where the mtDNA 'hitchhikes' through reproductive events as it follows the same maternal-

inheritance pathway of the bacteria (Jiggins 2003, Schuler et al. 2016). Selective sweeps can occur when *Wolbachia* induces a bias for female host offspring through multiple types of reproductive phenotypic manipulations (Werren et al. 2008), which can be both strain- and host-specific (Charlat et al. 2003). Such selective sweeps result in *Wolbachia* proliferating itself, and its associated mtDNA, quickly through the host population. However, the mechanisms that dictate the interactions between *Wolbachia* and its host remain largely unknown and are difficult to examine. Additionally, despite its prevalence, *Wolbachia* infections often go unreported, despite being very common in insect taxa and often being accidently detected during standard sequencing of host DNA (Smith et al. 2012).

Little is known about the status of *Wolbachia* (and other bacterial endosymbiont) infections in Australia's bee fauna. The Australian bee fauna comprises an estimated 2,000–3,000 species and is dominated by the family Colletidae. One colletid subfamily, the Hylaeinae, is a hyper-diverse group of bees (Houston 1975) that originated in Australia during the early Eocene, where it has radiated extensively (Kayaalp et al. 2013). It has also spread from Australia to all other regions of the globe, except Antarctica, forming one of the most geographically successful bee groups in the World (Kayaalp et al. 2013). One such hylaeine species, *Amphylaeus morosus* (Smith, 1879), occurs widely along the eastern coast of Australia, extending from southern Queensland to south-western Victoria (Houston 1975). Despite being one of the most comprehensively studied colletid species in Australia, little is known about its overall biology and ecology. This thesis explores:

(i) the occurrence of conserved mitochondrial heteroplasmy in *A. morosus* and
(ii) the possible relationship between mitochondrial heteroplasmy in *A. morosus* and the bacterial endosymbiont, *Wolbachia*

Below, I briefly outline the contents of each chapter presented in this thesis:

Chapter 1: A review of mitochondrial heteroplasmy in animal systems

Mitochondrial heteroplasmy in animal taxa is haphazardly reported in the literature. However, the overall occurrence patterns within animals and the common evolutionary mechanisms that enable heteroplasmy have not been broadly explored. To address this, I conducted a comprehensive review of the published literature which recovered reports of mitochondrial heteroplasmy in animal species spanning multiple phyla. Heteroplasmy can be generally grouped into two types: length and site heteroplasmy (Barr et al. 2005). Length

heteroplasmy usually presents as a series of tandem repeats in hypervariable mtDNA regions. My analyses identified that length heteroplasmy is more often reported in chordates compared to other animal taxa. Site heteroplasmy is generally reported as variability at single nucleotide positions and occurs most often in arthropods — although it also occurs in diverse taxa across multiple phyla. The frequency at which heteroplasmy occurs within any particular species is variable. However, reports have increasingly identified animal populations where 100% of the individuals are heteroplasmic. The mechanisms driving the most prominent forms of heteroplasmy (length and site) are generally not well understood.

Chapter 2: Extensive and widespread mitochondrial heteroplasmy in a colletid bee, *Amphylaeus morosus*

Amphylaeus morosus occurs along the eastern coast of Australia, with a ~2,000 km geographical range. Genetic sequencing of the mtDNA cytochrome *c* oxidase subunit I (COI) gene indicated that not only does mitochondrial heteroplasmy occur in every individual sampled across this range (n = 73), but every individual also has the same two mitochondrial haplotypes. Ion Torrent shotgun sequencing demonstrated that this mitochondrial heteroplasmy (presenting as site variation) extends throughout the mitogenome, suggesting that an ultra-large nuclear pseudogene is an unlikely explanation for its genetic and geographic consistency. The combination of 100% frequency as well as conserved mtDNA haplotypes, particularly across such a large geographical area, has never been reported for any other animal species. The mechanisms that could enable widespread and consistent mitochondrial heteroplasmy in *A. morosus* remain, as of yet, uncertain.

Chapter 3: Co-infection of *Wolbachia* in the Australian bee *Amphylaeus morosus* (Colletidae: Hylaeinae)

In conjunction with mitochondrial heteroplasmy, *A. morosus* is consistently infected with two supergroup A strains of *Wolbachia* across its distribution. Phylogenetic analyses of these two *Wolbachia* strains from the northern and southern geographic limits of this bee host indicated that (i) these strains did not diverge from each other within this host, and (ii) they may have been introduced in separate events (i.e. these unrelated strains were not simultaneously introduced into *A. morosus* as an established co-infection). *Wolbachia* is known to cause selective sweeps of mtDNA in their hosts (Jiggins 2003, Cariou et al. 2017). I have proposed that the lack of mtDNA variation and the persistence of heteroplasmy could potentially be

explained by a *Wolbachia*-induced selective sweep and subsequent mutualism between the heteroplasmic mtDNA 'haplotypes' and one or both *Wolbachia* strains. This contrasts with the more common hypothesis that the persistence of extensive heteroplasmy is through an unrelated selective process such as balancing selection. However, further research is needed to confirm and describe the mechanism(s) enabling widespread mitochondrial heteroplasmy in *A. morosus*.

Chapter 4: Incomplete removal of the reproductive parasite *Wolbachia* from a non-model organism following antibiotic treatment

To explore the connection between heteroplasmic mtDNA inheritance and *Wolbachia,* I have investigated the utility of traditional antibiotic treatment methods for removing *Wolbachia* from this host. The removal of the bacterial infection from *A. morosus* with antibiotics provides an opportunity to assess the subsequent effect on the inheritance of host mtDNA, post infection. One *Wolbachia* strain showed consistent susceptibility to antibiotic treatment; however, the *Wolbachia* infection was not completely removed within one generation. Although antibiotic treatment is one the most effective avenues to investigate the effects of *Wolbachia* on this host, the challenges of maintaining *A. morosus* in long-term captivity makes its application to this host unfeasible at present.

Chapter 5: Population structure in a native bee species with no mitochondrial variation?

To investigate whether the uniformity of mtDNA across the *A. morosus* distribution is reflected in the nuDNA, I used genomic single nucleotide polymorphism (SNP) data generated using the DArTseq™ platform from individuals across six localities spanning the entire ~2,000 km geographical range. Analyses for inferring population structure produced strong evidence of genetic clustering between localities, rather than a single, large 'panmictic' population. This contrasts with the patterns observed in the mtDNA which supports the latter scenario. Furthermore, habitat modelling showed that before European settlement in Australia, *A. morosus* habitats were likely better connected than today. These two pieces of evidence suggests that a lack of mtDNA variation is more likely the result of a *Wolbachia*-induced selective sweep across the whole species' distribution in the pre-European past, rather than a recent population bottleneck. This provides further, but indirect,

evidence that *Wolbachia* has played a key role in the evolution of this host's unusual mtDNA traits.

References

Barr, C. M., M. Neiman, and D. R. Taylor. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist **168**:39–50.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Research **27**:1767–1780.

Cao, L., H. Shitara, T. Horii, Y. Nagao, H. Imai, K. Abe, T. Hara, J. Hayashi, and H. Yonekawa. 2007. The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. Nature Genetics **39**:386–390.

Cariou, M., L. Duret, and S. Charlat. 2017. The global impact of *Wolbachia* on mitochondrial diversity and evolution. Journal of Evolutionary Biology **30**:2204–2210.

Charlat, S., G. D. D. Hurst, and H. Merçot. 2003. Evolutionary consequences of *Wolbachia* infections. Trends in Genetics **19**:217–223.

Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.

Jiggins, F. M. 2003. Male-killing *Wolbachia* and mitochondrial DNA: sselective sweeps, hybrid introgression and parasite population dynamics. Genetics **164**:5–12.

Kayaalp, P., M. P. Schwarz, and M. I. Stevens. 2013. Rapid diversification in Australia and two dispersals out of Australia in the globally distributed bee genus, *Hylaeus* (Colletidae: Hylaeinae). Molecular Phylogenetics and Evolution **66**:668–678.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates Sunderland.

Martin, S., A. 2011. Mitochondrial DNA Repair. *in* DNA Repair – On the Pathways to Fixing DNA Damage and Errors F. Storici, editor.

Satoh, M., and T. Kuroiwa. 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Experimental Cell Research **196**:137–140.

Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S. Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffer, W. Arthofer, and M. Riegler. 2016. The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Molecular Ecology **25**:1595–1609.

Smith, F. 1879. Descriptions of New Species of Hymenoptera in the Collection of the British Museum, London: British Museum.

Smith, M. A., C. Bertrand, K. Crosby, E. S. Eveleigh, J. Fernandez-Triana, B. L. Fisher, J. Gibbs, M. Hajibabaei, W. Hallwachs, K. Hind, J. Hrcek, D. Huang, M. Janda, D. H. Janzen, Y. Li, S. E. Miller, L. Packer, D. Quicke, S. Ratnasingham, J. Rodriguez, R. Rougerie, M. R. Shaw, C. Sheffield, J. K. Stahlhut, D. Steinke, J. Whitfield, M. Wood, and X. Zhou. 2012. *Wolbachia* and DNA barcoding insects: patterns, potential, and problems. PLOS ONE **7**:e36514.

Werren, J. H., L. Baldo, and M. E. Clark. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature Reviews Microbiology **6**:741–751.

Zhang, F., and R. E. Broughton. 2013. Mitochondrial–nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? Genome Biology and Evolution **5**:1781–1791.

Chapter 1: A review of mitochondrial heteroplasmy in animal systems

Olivia K. Davies[a*] and Michael P. Schwarz[a]

[a]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

*Corresponding author. Email: davies.oliviak@gmail.com

Keywords: mitochondrial heterogeneity, site polymorphism, tandem repeats, population fixation, biparental inheritance

1.1 Abstract

Mitochondrial heteroplasmy is the presence of more than one type of mitochondrial DNA (mtDNA) within a single cell or individual. Originally thought to occur rarely in healthy individuals, research now shows that mitochondrial heteroplasmy occurs frequently across animal species and populations. However, an understanding of the characteristics of heteroplasmy across animals and the mechanisms that particularly enable the most extensive forms (where as many as 100% of individuals within a population are heteroplasmic) remains poorly understood. We summarize the extent, characteristics, and proposed mechanisms of naturally occurring heteroplasmy in wild animal species collected from records in the broad literature. Mitochondrial heteroplasmy exhibits general trends in higher order taxon groups, with arthropods predominantly displaying site heteroplasmy (point mutations) and high rates of extensive heteroplasmy where entire populations can appear to be heteroplasmic. Conversely, chordates generally display length heteroplasmy (short tandem repeats) in the control region of the mitogenome and although extensive heteroplasmy occurs, it is less common than in arthropods. Other higher order taxon groups are significantly under-represented in the literature. Overall, heteroplasmy can be enabled by both taxon-specific and broadly shared mechanisms in animals, presenting many interesting questions surrounding unusual mitochondrial inheritance.

1.2 Introduction

1.2.1 Traditional views on mitochondria

Mitochondria are essential organelles and one of their key functions concerns the production of ATP. They are also associated with numerous other important cell functions, such as cell signalling, fertilization, developmental regulation, aging, and cell death (Green and Kroemer 2004, Balaban et al. 2005, Van Blerkom 2011, Chandel 2014). In animals, mitochondrial genomes are typically relatively small (~16–20 kb), circular, uniparentally inherited molecules that exhibit little recombination (Boore 1999). Hundreds to thousands of copies of mitochondrial DNA (mtDNA) can exist within a single animal cell, in contrast to nuclear DNA (nuDNA) with only two copies per cell in diploids (Pierce et al. 1990, Shay et al. 1990). Despite their fundamental role in energy production, mtDNA can accumulate mutations 10–50 times faster than nuclear DNA (nuDNA) (Lynch 2007, Martin 2011). But because of their generally constant gene composition, involving about 13 protein-coding genes used for the respiratory chain, plus approximately two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes (Zhang and Broughton 2013), mitochondria have traditionally been assumed to serve a very limited role in animal evolution.

1.2.2 Intra-individual mitochondrial variation

Mitochondrial genetic variation within animal species is generally reported to be limited to between individuals, and instances of genetic variation within individuals are reported relatively rarely. This characteristic of mtDNA has widely been considered counterintuitive because it would be expected that if copy numbers and mutation rates are high, and repair mechanisms are limited, then occurrences of multiple different copies of mtDNA should accumulate within an individual over time (Barr et al. 2005). Indeed, in humans (in which mtDNA abnormalities have been the most comprehensively categorised), intra-individual mtDNA variation occurs across tissues in most individuals to some extent — usually at low frequencies within a primarily homoplasmic composition (Melton 2004, Ye et al. 2014). Termed mitochondrial heteroplasmy (the existence of more than one type of mtDNA within a single cell or individual); it is sporadically reported in the literature, occurring across animals, fungi, and plants (the latter two reviewed by Barr et al. (2005), Kmiec et al. (2006) and Ramsey and Mandel (2019)). Here, we discuss the patterns of naturally occurring mitochondrial heteroplasmy in wild animals (excluding humans) across peer-reviewed, published literature including the (i) type, (ii) prevalence within species and/or populations,

and (iii) proposed mechanisms facilitating and maintaining mitochondrial heteroplasmy in natural systems.

1.2.3 Types of mitochondrial heteroplasmy and their origins

Mitochondrial heteroplasmy can be grouped into two general types: site heteroplasmy and length heteroplasmy (Barr et al. 2005). Each type can be naturally facilitated by several cellular processes, taking place stochastically or facilitated by segregation breakdowns during sexual reproduction. Site heteroplasmy, involving point mutations, results from a nucleotide substitution or point deletion. These single nucleotide polymorphisms (SNPs) can arise through replication errors and numerous mutagenic processes resulting in somatic mutations or oocyte heteroplasmy (Kvist et al. 2003, Aryaman et al. 2019), and these mutant mitogenomes can either be lost or perpetuated by random genetic drift (Elson et al. 2001).

Alternatively, length heteroplasmy (or length variation) is described as small tandemly repeated sections of a mtDNA sequence or large genomic duplications present in one mtDNA lineage (Brown et al. 1996). Length variation polymorphisms are thought to most commonly be the result of errors during mitochondrial replication, often in hypervariable mtDNA regions (Densmore et al. 1985, Kvist et al. 2003), but may also be the result of gene/fragment duplications. Extensive deletions of DNA fragments in the mitogenome may also occur; these are best understood in humans where they can cause mitochondrial diseases (DiMauro and Schon 2001). Recombination of mtDNA has been shown to be a source of heteroplasmic variation in plants (Barr et al. 2005), but evidence remains ambiguous as to its role in animals (Rokas et al. 2003, Piganeau et al. 2004).

Biparental inheritance, where both parents contribute their mitochondria to their offspring, is also a commonly reported source of mitochondrial heteroplasmy (mitogenomes with site and/or length variation). Although occasionally resulting in life-long mitochondrial heteroplasmy in the offspring, generally biparentally inherited mtDNA result in a temporary phase of heteroplasmy during early development, before one mtDNA haplotype is lost through genetic drift or selection, and a homoplasmic state is restored in the individual (Kondo et al. 1990, Barr et al. 2005, Nunes et al. 2013). Paternal leakage is the most commonly reported form of biparental inheritance, and occurs widely throughout animal taxa (Shitara et al. 1998, Kvist et al. 2003, Fontaine et al. 2007, Nunes et al. 2013). However, some animal systems have complex and perpetual forms of biparental inheritance (i.e. doubly uniparental inheritance (DUI)), discussed below. In all cases, these processes may result in

temporary phases of heteroplasmy restricted to individuals or may enable heritable forms of heteroplasmy where both mtDNA types are maintained together throughout subsequent generations.

1.2.4 Individual- versus population-wide mitochondrial heteroplasmy

Mitochondrial heteroplasmy is generally interpreted as an intermittent, transient condition in natural populations (Song et al. 2008). Occasionally, heteroplasmy is reported to be widespread in some taxa and poses problems for standard mitochondrial models. Here, we define two forms of mitochondrial heteroplasmy, where heteroplasmic haplotypes are maintained only at an individual level (*transient*) or at a population level (*heritable*).

Transient heteroplasmy: Humans are the most extensively studied taxon with regard to mtDNA abnormalities. As part of the normal aging process in humans, mtDNA has been shown to accumulate a low number of somatic mutations across tissues (Chinnery et al. 2002). However, within cells, the amount of mutant mtDNA has been shown to be very high and may cause age-related mitochondrial diseases. These mutant mtDNA are likely perpetuated by the same mechanisms that maintain normal cellular concentrations of mtDNA (Chinnery et al. 2002) and lineages of mtDNA can be lost or maintained by genetic drift or other selective forces (Sondheimer et al. 2011). Although not extensively demonstrated in other animals, this pattern is likely widespread, and any individual over their lifetime may be susceptible to age-related/mutagenic acquisition of somatic mitochondrial heteroplasmy, which we will define as *transient* heteroplasmy. *Transient* heteroplasmy may be inherited into the next generation if these mutations occur in gonadal tissues (through maternal or paternal transfer) but is not expected to persist due to shifts in mtDNA compositions by genetic drift or selective processes. Overall, this form of heteroplasmy is not maintained on a population level, and most likely presents as either a disadvantageous or neutral mutation.

Heritable heteroplasmy: In some animals (and other taxa hosting mitochondria), heteroplasmy appears to be highly prevalent in individuals within a population. In these cases, heteroplasmy appears heritable (e.g. Doublet et al. (2008), Van Leeuwen et al. (2008)), occurs throughout multiple tissues or shows high rates of tissue segregation suggesting functional properties (e.g. Jenuth et al. (1997), Magnacca and Brown (2010b)), and should be maintained through the organism's lifetime (although this hasn't been experimentally shown). The mtDNA haplotypes produced can be detected across multiple individuals of a population across generations (i.e. implying inheritance) and can be maintained as

heteroplasmic mitochondrial lineages with a population. This form we will define as *heritable* heteroplasmy and differs from *transient* as these mtDNA are likely biologically functional and/or are maintained by positive, balancing, or selfish selective forces. These heteroplasmic lineages must be maintained through inheritance, either paternal leakage or maternal transmission. A number of taxa appear to have developed *heritable* heteroplasmy, and these examples will be discussed in the following sections.

1.3 Methods

1.3.1 Literature search and data extraction

A list of research articles published between 1985 and 2020 was generated using the Web of Science and SCOPUS databases on 3rd July 2020. The following search string for words in the title, abstract or author provided key-words was used: '(heteroplasmy OR heteroplasmic) AND (mitochondria*) AND NOT (plasmid* OR fung* OR plant* OR chloroplast* OR phage*) AND NOT (human OR wom?n OR patient* OR disease* OR child* OR disorder* OR participant* OR donor* OR "stem cell*"), excluding reviews or book chapters (experimental, peer-reviewed articles only) and non-English language publications, and recovered 796 and 425 studies respectively, of which 328 overlapped between databases. This search string had been optimized to reduce once-off reports of heteroplasmic individuals with no population context, records of heteroplasmy from non-target taxon groups (non-animals and humans), and medical analyses of mitochondrial disorders, which are not within the scope of our assessment. A number of reviews addressing disease causing (Chinnery and Turnbull 1999, Wallace et al. 1999, Stewart and Chinnery 2015, Hatakeyama and Goto 2016, Stefano and Kream 2016) and the forensic implications (Melton 2004, Just et al. 2015) of mitochondrial heteroplasmy in humans are available, and were not within the scope of this study.

Because of the nature of most heteroplasmy reports in natural animal populations (which often originate from peripheral observations associated with wider genetic/genomic questions), publications needed to be examined thoroughly beyond information provided in the abstract, to recover the targeted information (e.g. the sample size of the population assessed for mitochondrial heteroplasmy, detection method used, proposed mechanism(s) for the heteroplasmy observed). Literature that did not meet the following requirements were excluded from analyses. Specifically, for inclusion, studies had to meet the following criteria:

(i) experimentally investigate for the presence of mitochondrial heteroplasmy in an animal species,

(ii) not include species in captive populations assessing abnormal mitochondrial inheritance or genetically manipulated for other experimental questions,

(iii) assess samples that were not selected because of expectations of heteroplasmy,

(iv) present a clear sample size of both the number of individuals assessed *and* those that were identified as having mitochondrial heteroplasmy,

(v) assess individuals developed beyond cell-lines, gamete, or zygote stages, as individuals should be shown to be able to survive with heteroplasmy approaching and beyond reproductive age for it to not act as a lethal 'mutation' at a population level, and

(vi) analyse more than one individual within a taxon unless the publication was reporting multiple singletons of related taxa, or in-depth assessment of the entire mitogenome had been conducted for publications with one specimen.

In total, 234 publications (plus 17 additional publications where heteroplasmy results were clearly stated in the publications reviewed or were the original source for the data) of 893 met the criteria for inclusion.

1.3.2 Curation of taxon records

Table 1 summarizes reports of heteroplasmy across animals based on published literature recovered with our search parameters. Although thorough, it does not represent all studies on heteroplasmy in animals; but see rules above. Our list was compiled to demonstrate where frequencies of heteroplasmy within taxa were reported, and although it includes some examples of homoplasmic species that were recovered with our search parameters, these records of homoplasmy are likely massively under-represented. Taxa in Table 1 were sorted alphabetically within Phyla, Class, then Order, followed by species (mostly sorted alphabetically, but duplicated species were attempted to be grouped together for ease of comparison). All taxonomic names were assessed against the Global Biodiversity Information Facility (GBIF) (GBIF.org 2020) and (where applicable) the World Register of Marine Species (WoRMS) (Horton et al. 2020) databases to ensure currently accepted nomenclature was employed in this review. Additionally, all taxa were assessed against the

IUCN Red List of Threatened Species database (IUCN 2020), with taxa assessed as vulnerable to critically at risk of extinction denoted with a caret (^).

1.3.3 Grouping of taxa for statistical analyses

Published reports of heteroplasmy in animals have considerable taxon bias towards chordates, arthropods, and bivalves (molluscs) (Supp. Table 1). As such, taxon groups used in comparative statistical analysis included only these three phyla (Chordata, Arthropoda, and Mollusca; Figure 1A.). Representatives of seven other phyla were recovered in our search (Echinodermata, Nematoda, Annelida, Nemertea, Platyhelminthes, Cnidaria, and Porifera), however, these were excluded from statistical analyses due to low sample sizes of taxa (Figure 1B.).

1.3.4 Statistical analysis of mitochondrial heteroplasmy

Mitochondrial heteroplasmy was defined into three categories: “site”, “length”, or “both” (Figure 1). Instances of homoplasmic taxa (“none”) were not included in statistical analyses because these records were not targeted with our search parameters. Statistical analyses were performed in R version 3.6.3 (R Development Core Team 2020) using the packages DescTools version 0.99.40 (Signorell et al. 2021) and RVAideMemoire version 0.9-78 (Hervé 2020) in RStudio version 1.3.1056 (RStudio Team 2020). For statistical analyses, the frequency of taxa (grouped into Chordata, Arthropoda, and Mollusca) with each type of heteroplasmy was recorded. Fisher’s Exact tests with a Bonferroni corrected, simulated p-value based on 20,000 replications were performed on mitochondrial heteroplasmy defined as “site”, “length”, and “both”. Fisher’s Exact *post hoc* tests using a Bonferroni correct p-value (Supp. Table 2) were performed using R package RVAideMemoire to assess which categories significantly differed from the others.

Additionally, the distribution of the proportion of heteroplasmic individuals within each population for species in both Chordata and Arthropoda was produced (Figure 2). This was to explore possible differences in the prevalence of heteroplasmy within populations in both general taxon groups and whether one taxon group has comparatively more records of *extensive* heteroplasmy (Figure 2).

1.3.5 Methodology used for heteroplasmy characterization through time

We produced time series’ capturing the years in which heteroplasmic records (251 publications) were collected for this review (from 1985 to 2020); which summarised the (i)

types of heteroplasmy for each taxon published in each year (defined as “site”, “length”, “both”, and “none”; Figure 3), and (ii) laboratory protocols used to investigate heteroplasmy in each publication (Figure 4; Supp. Table 3). It is important to note that an included publication could be reporting on heteroplasmy (or homoplasmy) in multiple taxa. Additionally, multiple laboratory techniques were often applied to investigate heteroplasmy. Therefore, often the number of heteroplasmic records and applied techniques were higher than the number of publications for each year (Figures 3 and 4).

1.4 Results and Discussion

1.4.1 Patterns of heteroplasmy across animals

There was a significant association between the taxon group and the type of mitochondrial heteroplasmy presented (Fisher’s Exact test, two-sided, P-value ~2.2×10^{-16}). We observed that length heteroplasmy is most common in Chordata and Mollusca and site heteroplasmy was most common in Arthropoda (Figure 1A.), which was supported by our Fisher’s Exact *post hoc* results (Supp. Table 2). This finding is concordant with other studies (Hale and Singh 1986, Moritz and Brown 1987, Barr et al. 2005, Robison et al. 2015) suggesting that certain broad taxon groups differentially exhibit types of mitochondrial heteroplasmy, although this has not been quantified until now. Statistical analyses could not be performed on other taxon groups due to low sample sizes (Figure 1B).

The distribution of mitochondrial heteroplasmy prevalence across populations (Figure 2) was similar for chordates and arthropods. Records of sampled populations with 100% heteroplasmic individuals are frequent (Figure 2). Although in both groups, a large portion of these records were due to assessments of a single individual (singletons) representing a population (Figure 2; Table 1). For both taxon groups, low levels of heteroplasmy — heteroplasmic individuals accounting for less than 50% and particularly less than 20% of the sample population — are the next most common (Figure 2). High levels of heteroplasmy (greater than 50%) appear less frequent, until heteroplasmy reaches fixation within a population (Figure 2). This poses questions as to the mechanisms that drive mitochondrial heteroplasmy beyond relatively low frequencies within a population (<50%), to near or complete fixation. Sample sizes for the other phyla were too low to assess the distribution of heteroplasmy prevalence across populations. In the case of molluscs, sex-biased heteroplasmy in Bivalvia generates records of heteroplasmy skewed towards males within populations, therefore it was not assessed. We will discuss patterns of heteroplasmy within

taxon groups and explore whether there are taxon specific mechanisms that may lead to high rates of heteroplasmy.

1.4.2 Length heteroplasmy in the mitochondrial control region of chordates

In chordates, mitochondrial heteroplasmy is most commonly the result of length variation, particularly associated with control region (and the D-loop and adjacent RNA regions) of the mitogenome (Table 1). Certain clades are well represented in the literature with this type heteroplasmy, with multiple species assessed over several independent studies. Heteroplasmy in bony fishes (Actinopterygii) is particularly well represented, and sturgeons (Acipenseriformes) are comprehensively assessed in the literature (Table 1). Sturgeons present a good example where multiple studies independently assessed numerous species, demonstrating the variation in rates of heteroplasmy identified within a species (Table 1). For example, in populations of the endangered Russian sturgeon, *Acipenser gueldenstaedtii*, two studies (Ludwig et al. 2000, Çiftci et al. 2013) reported relatively low rates of heteroplasmy (between 8% (n = 24) and 14% (n = 98), respectively) compared to a third study reporting 100% heteroplasmy among 145 individuals (Pourkazemi et al. 1999). These three studies proposed varying mechanisms to explain the observed heteroplasmy; namely, intramolecular recombination, high mutation rates via insertions and deletions, and a combination of paternal leakage and stochastic mutation, respectively. The Pourkazemi et al. (1999) study did assess site heteroplasmy as well as the length heteroplasmy in the control region. Nonetheless, these studies report markedly different conclusions of the frequency and sources of heteroplasmy for the same species.

Length heteroplasmy in the mitochondrial control region also occurs commonly in numerous amphibians, birds, mammals, and some reptiles (turtles (testudines) and crocodilians), but interestingly not in squamates (Table 1). There are also some examples of site heteroplasmy being identified in the control region of chordates (e.g. Pourkazemi et al. (1999), Maté et al. (2007), Edwards et al. (2010), Vollmer et al. (2011), Rollins et al. (2016)), but these reports are much less common (Table 1). Many vertebrate studies target the control region as it is the main non-coding region of the mitogenome, and it often has a higher substitution rate (Parsons et al. 1997), useful for phylogenetic and population studies. Overall, this region functions to control DNA and RNA synthesised however, hypervariable regions where these tandem repeats commonly occur have no known function (Kozhukhar et al. 2020).

The mechanisms that underpin this form of intra-individual mtDNA variation are not yet well understood, and numerous processes have been proposed. Slipped strand-mispairing, also called replication slippage, (a similar process to that which occurs in nuDNA) is the most commonly proposed source of tandem repeat variation in the mtDNA control region (e.g. Berg et al. (1995), Savolainen et al. (2000), Vollmer et al. (2011), He et al. (2013)), although this has not been shown experimentally outside of model organisms. Alternatively, this variation is attributed to recombination (both between and within molecules) (e.g. Hoarau et al. (2002)) or displacement/misalignment of heavy-strand and the light-strand or D-loop strand (Buroker et al. 1990, Hayasaka et al. 1991). These molecular mechanisms have generally maintained their plausibility since being summarised by Brown et al. (1996), and few other molecular processes proposed for wild animal populations (Table 1) have gathered traction with researchers.

The above molecular processes could also facilitate control region length variation at a population level, existing *within individuals* as homoplasmy (i.e. the length of tandem repeats vary between individuals, rather than within individuals). Following breakdowns in 'typical' inheritance patterns, paternal leakage has been described as a source of mtDNA length variation (e.g. Pourkazemi et al. (1999), Hoarau et al. (2002), Maté et al. (2007)). However, in these systems it is likely that this is a temporary heteroplasmic phase before one mtDNA haplotype is lost. These mechanisms might explain how heteroplasmy in the control region arises, however there are scarce reports of possible mechanisms that maintain this form of heteroplasmy across generations (heritable heteroplasmic tandem repeats). Occasionally, this form of heteroplasmy is shown to be maintained through maternal inheritance (Nesbø et al. 1998, Mayer and Kerth 2005, McLeod and White 2010). Considering the high frequency at which it occurs in some species of chordates (Figure 2A., Table 1), it is likely that there are inheritance mechanisms maintaining heteroplasmy in some taxa, although it is unclear what these mechanisms might be.

1.4.3 Other types of heteroplasmy in vertebrates

Heteroplasmy in chordates is not limited to hyper variable tandem repeats in the control region. Site heteroplasmy in multiple genes, and some length heteroplasmy outside of the major non-coding regions has been identified (Table 1). Numerous examples of nucleotide variability (site heteroplasmy) occur in the literature, often from whole mitogenome data, including in fishes, birds, mammals, amphibians, and turtles (Table 1). Interestingly, the

location of these variable sites is sometimes consistent within related taxa (e.g. Moreira et al. (2017)), but otherwise can occur throughout the mitogenome. Generally, the source of this nucleotide variation is not understood in these taxa.

In at least one example, no association between age and the frequency of heteroplasmic sites was found in the greater mouse-eared bats, *Myotis myotis* (Jebb et al. 2018), challenging the hypothesis that aging is the main mechanism producing somatic mtDNA mutations. SNPs were identified throughout the mitogenome of *M. myotis* (Jebb et al. 2018), with individuals between zero and six (or greater) years of age possessing similar numbers of heteroplasmic sites (<6 nucleotides). However, no other reports addressing the impacts of aging on heteroplasmic accumulation was recovered for wild animals. Conversely, in the case of Spicer et al. (2014), SNP variation was specifically associated with aging in domestic dogs (*Canis lupus familiaris*), supporting the notion of age-related accumulation of somatic mutations, as is observed in humans (Ye et al. 2014, Kang et al. 2016).

In most other examples of site heteroplasmy occurring in chordates, the proposed sources have either not been defined or are attributed to unknown mutagenetic processes leading to intra-individual variation (e.g. Shigenobu et al. (2005), Vollmer et al. (2011), Kesäniemi et al. (2018)). Whether heteroplasmy at these sites can be transmitted to offspring is not well explored, due to low sample sizes, difficulties in assessing the age of individuals, and a poor understanding of pedigree relationships among the individuals sampled. However, there are some examples were familial relationships within species are well assessed, and it has been demonstrated that heteroplasmic mitochondrial haplotypes are inherited, either maternally (e.g. Mayer and Kerth (2005), McLeod and White (2010), Lopes et al. (2011)) or paternally — although examples of paternal leakage are usually detect in hybridization events (e.g. Radojičić et al. (2015), Gandolfi et al. (2017), Päckert et al. (2019)). Evidence of *heritable* heteroplasmy (where heteroplasmic haplotypes are maintained at a population level) is not yet well established for wild chordates. However, examples such as *M. myotis* (Jebb et al. 2018) where heteroplasmic haplotypes were shown to be maternally transferred across generations, present strong evidence in support occurrences of some heritable systems.

Records recovered from squamates (snakes and lizards) are particularly unusual, because although length heteroplasmy is most common, it was not generally reported for the control region, but often in rRNA genes (e.g. Zevering et al. (1991), Kumazawa (2004)). Additionally, numerous studies failed to identify any heteroplasmy in squamates (e.g.

Kumazawa et al. (1996), Ujvari et al. (2007), Dubey et al. (2011); Table 1). In the parthenogenetic gecko, *Heteronotia binoei*, length heteroplasmy was identified (Moritz 1991, Zevering et al. 1991) occurring at low-medium frequencies within the sample populations (4.6% (n = 87) and 35.1% (n = 77), respectively; Table 1), eliminating paternal leakage as a possible mechanism. Unfortunately, contemporary studies of heteroplasmy in squamates were not recovered within our search parameters — with the exception of Miraldo et al. (2012) identifying low frequencies of site heteroplasmy (9.5% of n = 21) occurring in the cytb gene of *Timon lepidus*. Future research in this taxon group would be beneficial to explore why heteroplasmy might not occur commonly in the control region in squamates, compared to other chordates.

1.4.4 Site heteroplasmy is rampant in Arthropoda

In arthropods (and invertebrates generally), reports of assessments of mitochondrial heteroplasmy are less common than in vertebrates (Table 1). Compared to vertebrates where heteroplasmy is generally due to large tandem repeated units usually in the mitochondrial control region, heteroplasmy in arthropods is more commonly associated with synonymous substitutions at single nucleotides throughout the mitogenome (Figure 1A.), usually at third codon positions (Nunes et al. 2013, Robison et al. 2015). Additionally, the frequency of mitochondrial heteroplasmy within populations of arthropods was more often reported as *extensive* (100% of the sampled individuals; Figure 2B.), compared to chordates.

Heteroplasmy has been (comparatively) well reported in hymenopterans, predominantly in bees (Table 1). Site heteroplasmy has been reported in species from five of the world's seven bee families and at particularly high rates in the family Colletidae (Magnacca and Brown 2010a, Magnacca and Brown 2012). The frequency of heteroplasmic individuals in bees is variable (Table 1). But cases such as the Hawaiian *Hylaeus* are particularly unusual, where numerous rapidly and recently speciating taxa have extensive levels of heteroplasmy (generally 100%) and high mitochondrial haplotype diversity (Magnacca and Danforth 2006, Magnacca and Brown 2010a), suggesting possible ancestral drivers of heteroplasmy. NUMTs (nuclear inclusions of mitochondrial DNA) are also widely detected in bees, and a number of authors have highlighted the challenges in distinguishing between true mtDNA variation and nuclear pseudogenes (Magnacca and Brown 2010a, Ricardo et al. 2020a). Site heteroplasmy has also been recently detected in bumblebees (*Bombus*) (Williams et al. 2019, Ricardo et al. 2020a) which concurs with Magnacca and Brown (2010a) reporting that heteroplasmic

individuals can possess multiple mitochondrial haplotypes (e.g. an average of seven mtDNA haplotypes per heteroplasmic individual of *B. morio* (Ricardo et al. 2020a)). In honeybees (*Apis*), arguably the most extensively researched bee genus because of their agricultural use and invasive biology, reports of heteroplasmy are scarce. The Asian honeybee, *Apis cerana* is the exception, with length heteroplasmy reported in the ATPase gene region by Songram et al. (2006) in specific subpopulations, which the authors highlight had not been detected in other earlier investigations that included this heteroplasmic *A. cerana* population.

Interestingly, site heteroplasmy has been extensively detected in the well-known honeybee ectoparasitic mite, *Varroa destructor* (Gajić et al. 2016, Gajić et al. 2019, Muntaabski et al. 2020). Heteroplasmic levels are usually below ~50% of the sampled population (Table 1) and at least in these cases, heteroplasmy is likely maternally inherited (Gajić et al. 2019, Muntaabski et al. 2020). Numerous ticks have also been reported to have high rates of site heteroplasmy in studies where whole mitogenomes were sequenced, but in these cases, one individual was assessed for each species (Xiong et al. 2013, Williams-Newkirk et al. 2015). These heteroplasmic nucleotide sites occurred throughout the mitogenome, with as few as six, and up to 166 heteroplasmic sites recovered within any individual (Xiong et al. 2013). Conversely, tick studies targeting specific gene regions recover much lower frequencies (3.5% of individuals) with heteroplasmy (Nadolny et al. 2015, Mastrantonio et al. 2019), suggesting that methods targeting specific genes may miss intra-individual variation (either because barcoding methods are not sensitive enough, or variation occurs outside of the targeted gene region).

It is possible that heteroplasmy is an artefact of high rates of mtDNA variation facilitated by isolation, rapid adaptation, and speciation within some groups. For example, Van Leeuwen et al. (2008) attributed heteroplasmy in the two-spotted spider mite, *Tetranychus urticae* (Arachnida: Trombidiformes) to insecticide resistance i.e., rapidly evolved resistance to artificial culling attempts promoting high mtDNA variability and heteroplasmy. Numerous island-bound species (e.g. Magnacca and Brown (2010a)), pest/parasitic taxa (e.g. Xiong et al. (2013)), captive populations (e.g. Gawande et al. (2017)), and invertebrates with low dispersal ability (e.g. Doublet et al. (2012)) have been identified with high rates of heteroplasmy (Table 1). Given that heteroplasmy in arthropods is more frequently reported as extensive compared to other taxon groups, the role of heteroplasmy in preserving mtDNA variation in the context of heritable systems is an interesting and complex consideration. The

evolutionary roles of extensive heteroplasmy need to be further investigated in arthropod taxa and short generation times in most arthropods means that there is high potential for thorough investigations of these systems.

1.4.5 Other forms of heteroplasmy in arthropods

In arthropods, the mitochondrial control region is not commonly used for genetic studies and, therefore, heteroplasmy is not readily incidentally detected in this gene region. However, like that observed in chordates, some examples of length heteroplasmy in the control region have been identified in arthropods. This includes high rates (100%; Table 1) in three *Pissodes* species (Insecta: Coleoptera) (Boyce et al. 1989), however our search parameters did not recover any subsequent studies on this taxon group to investigate this further. Numerous 'early' studies (1985 ~ 2002) also detected length heteroplasmy in arthropod populations using restriction enzyme digests where the specific gene regions targeted were not identified in the publication (e.g. Harrison et al. (1985), Rand and Harrison (1989), Azeredo-Espin et al. (1991), Vargas and Espin (1995); Table 1). However, at least in the case of orthopterans, these early studies (Harrison et al. 1985, Rand and Harrison 1989) may have detected NUMTs, which are now known to occur rampantly across diverse lineages of Orthoptera (Bensasson et al. 2000, Song et al. 2014) — although Rand and Harrison (1989) did use a mitochondrial enrichment step in their protocol suggesting that mtDNA might truly be the source of the variation.

Heteroplasmy in the control region has been reasonably well explored in decapods (unlike other arthropod groups). Although length heteroplasmy (at the control region) has been detected (Shih et al. 2013), and one example of site heteroplasmy outside of the control region (Rodriguez-Pena et al. 2020), current evidence suggests that heteroplasmy may be the exception in decapods (Table 1). Most studies identify homoplasmy, including studies conducted on multiple species where some species are heteroplasmic (e.g. Katsares et al. (2003), Shih et al. (2013); Table 1). In these cases, the frequency of heteroplasmy was drastically different between species. For example, 10 individuals of *Homarus americanus* were identified as homoplasmic, but all 49 individuals of *H. gammarus* had length heteroplasmy (Katsares et al. 2003). Similarly, Shih et al. (2013) identified two homoplasmic *Paraleptuca* species (reported as *Uca*), and two heteroplasmic species, with 100% of *P. boninensis* and only ~7% of *P. crassipes* individuals having length variation in the control region. In all *Paraleptuca* species, the COI gene was also examined and was found to be

consistently homoplasmic (Shih et al. 2013). As is the case with most taxon groups, taxon sampling is too poor to suggest overall patterns nor shared molecular mechanisms.

D-loop length variation has also been identified in laboratory populations *Drosophila melanogaster* at low frequencies (16.7% (n = 144) and 1.5% (n = 194); Table 1) (Hale and Singh 1991, Townsend and Rand 2004). Townsend and Rand (2004) suggested that in these cases, that heteroplasmy was due to small laboratory populations accumulating mutations. Other studies have also identified heteroplasmy in *D. melanogaster* and other *Drosophila* species (Nunes et al. 2013, Wolff et al. 2013), but this was generally site heteroplasmy (Table 1) which occurred in wild-caught populations. Our search parameters excluded many records of laboratory/cultivated animal sample populations (see justification in the methods section). However, numerous examples of captive animal populations with high rates of heteroplasmy (both site and length) were retained in our search, as well as species vulnerable to extinction (i.e. have undergone substantial population reductions) (Table 1). It is possible that small population sizes and the subsequent loss of genetic diversity could contribute to occurrence of mitochondrial heteroplasmy, either because of disfunctions in normal mtDNA replication/repair/inheritance or as a mechanism to increase the rate of re-accumulation of mtDNA variation. However, this is purely speculative.

1.4.6 Doubly uniparental inheritance in Bivalvia

Records of heteroplasmy in molluscs are dominated by bivalves, with ~90% of mollusc papers recovered in our search being on bivalves. Bivalves are unique and have evolved a specific type of mitochondrial inheritance that results in sex-specific heteroplasmy — termed doubly uniparental inheritance (DUI) (Zouros et al. 1994). In DUI systems, females generally have only maternally derived mtDNA, but males have maternally derived mtDNA in somatic tissues and paternally derived mtDNA in gonadal tissues (Barr et al. 2005, Zouros 2013). Mitochondrial lineages in DUI systems can be highly divergent, separated into M-type and F-type mtDNA (male and female, respectively). For example, M- and F-type mtDNA in male freshwater mussels (Bivalvia: Unionoida) had an uncorrected nucleotide p-distance of >40% in most protein coding mitochondrial genes (Doucet-Beaupré et al. 2010).

Our data indicate that length heteroplasmy dominates mollusc records (Figure 1A), however realistically, the mtDNA types involved in DUI are so divergent, that it is likely a combination of site and length heteroplasmy which is not easily comparable with other animal taxa. Interestingly, this system appears susceptible to lineage replacement, with

"feminisation" or "masculinization" of the segregated lineages via reversals in the route of transmission (e.g. Hoeh et al. (1997), Breton et al. (2009)). The mechanisms behind this mitochondrial inheritance pattern are largely unknown, but it has been proposed the mechanisms are associated with patterns in germline formation and modification of mitochondrial transmission factors (Breton et al. 2014).

Possible DUI-like mechanisms have been identified in at least seven orders of bivalves (Table 1) however, it is not yet known how many times this trait has been gained (and lost) nor the mechanisms that facilitate DUI in this diverse group of molluscs (Doucet-Beaupré et al. 2010, Zouros 2013, Dégletagne et al. 2015). There has been substantial speculation about the evolutionary advantages and mechanisms that have led to the maintenance of DUI in many mollusc species. Bettinazzi et al. (2019) proposed that sustaining sex-linked mitochondria enable different gametes to meet their energetic requirements. The mechanisms underpinning DUI are specific in the context of animal mitochondrial heteroplasmy and require comprehensive reviews of their own (e.g. Zouros (2000), Breton et al. (2007), Passamonti and Ghiselli (2009), Breton et al. (2014)).

1.4.7 Heteroplasmy in other molluscs

DUI has been investigated in gastropods, where no evidence of any for type of heteroplasmy has yet been identified (Table 1; (Parakatselaki et al. 2016, Gusman et al. 2017)). Only one other publication of heteroplasmy within molluscs was recovered in our search, for the sparkling enope squid, *Watasenia scintillans* (Cephalopoda: Oegopsida). The entire mitogenome of one specimen revealed six heteroplasmic nucleotide sites occurring across the mitogenome (Hayashi et al. 2016). DUI is likely restricted to Bivalvia, and other molluscs with heteroplasmy may be susceptible to similar mutagenic mechanisms as other non-mollusc taxa. However, poor representation of non-bivalves in the heteroplasmic literature means it is too early to suggest what other mechanisms have evolved in molluscs to facilitate heteroplasmy.

1.4.8 Underrepresentation of other phyla in the literature

Taxa from seven additional phyla were recovered with our search parameters, but in all cases, there were fewer than two publications with low numbers of heteroplasmic taxa (many records identified only homoplasmic taxa). Length heteroplasmy in the COI gene was reported in a small proportion of the population (~2% of n = 117) of the basket star, *Astrobrachion constrictum* (Echinodermata: Ophiuroidea: Euryalida) (Steel et al. 2000)

(Table 1). Length heteroplasmy has also been reported in nematodes *Meloidogyne* spp. at reasonably high frequencies (between 60% and 80%; Table 1) (Dautova et al. 2002). However, in another study of the same species and its relatives, demonstrated only homoplasmy (Kiewnick et al. 2014) (Table 1) — although separate gene regions were targeted with different techniques. Additional Nematoda records were recovered in our search, but these were always of genetically monitored/modified lines in captive populations. The state of heteroplasmy in natural populations of nematodes remains unknown.

Site heteroplasmy has been reported in annelids, nemerteans and platyhelminths. In all cases, few taxa (<2 species) were determined to be heteroplasmic in any publication. In the case of annelids, one clitellate species, *Pontoscolex corethrurus* (Clitellata: Haplotaxida) was found to have low rates of site heteroplasmy in COI attributed to paternal leakage in a reasonably large sample of individuals (n = 61) (Taheri et al. 2018) (Table 1). However, Hurtado et al. (2004) identified only homoplasmy (assessing only COI) in five other annelid species (polychaetes) with similar sample sizes of individuals (Table 1). In Platyhelminthes, studies of two highly divergent species identified site heteroplasmy, both in the COI gene, at considerably different frequencies; 100% in *Echinococcus granulosus* (Bowles et al. 1994) and ~6% in *Dugesia japonic*a (Bessho et al. 1992) (Table 1). A single study of Nemertea was recovered and was represented by a single individual of *Carinina ochracea* (Palaeonemertea: Carinomiformes) found to have site heteroplasmy in the COI region (Fernandez-Alvarez et al. 2015) (Table 1). In all the above five phyla, it is difficult to discern patterns of mitochondrial heteroplasmy from so few examples in highly divergent species. The consistent identification of heteroplasmy in the COI gene is not surprising, given its prevalent use as a general barcoding region (Wilson 2012, Deagle et al. 2014, Andújar et al. 2018). It is notable that these unusual mitochondrial traits have not been widely captured in these taxon groups, some of which are hyper-diverse.

1.4.9 Heteroplasmy at the base of Animalia

Two basal animal clades, Cnidaria and Porifera, are poorly assessed for mitochondrial heteroplasmy. A single specimen of the cold water coral, *Desmophyllum pertusum* (Cnidaria: Anthozoa: Scleractinia) was found to have some site heteroplasmy where five heteroplasmic nucleotide sites were identified across the entire mitogenome (Emblem et al. 2012) (Table 1), the cause of which is unknown. Conversely, no evidence of mitochondrial heteroplasmy has been found in Porifera (demosponges). Fourteen species were assessed (each involving single

specimens), and most had whole/near-whole mitogenomes sequenced (Haen et al. 2014) (Table 1). Data is too deficient in these taxa to generalize about the extent and role of mitochondrial heteroplasmy in basal animal clades. The heteroplasmy literature reflects the general trends in most biological and ecological research; invertebrates, particularly marine, are heavily neglected compared to vertebrates (Donaldson et al. 2017, Titley et al. 2017, Beck and McCain 2020). However, we predict that as high-throughput DNA sequencing data becomes more widely used for general biodiversity monitoring, examples of deviant mitochondrial genetics (including heteroplasmy) will be passively collected during general biodiversity sampling, enlightening further patterns, and seeding future research directions.

1.4.10 Comprehensive mtDNA analyses also identify homoplasmy

Although it is not within the scope of this review to address taxa with an absence of heteroplasmy (i.e. homoplasmy), it is important to note that there are several examples of animals where only homoplasmy (or sometimes NUMTs) was identified despite analyses of whole mitogenomes. Most examples are from chordates, including the black lion tamarin (*Leontopithecus chrysopygus*) (de Freitas et al. 2018), Grey's beaked whale (*Mesoplodon grayi*) (Thompson et al. 2016), domestic cat (*Felis catus*) (Lopez et al. 1996), Atlantic bluefin tuna (*Thunnus thynnus*) (Manchado et al. 2004), Arctic cod (*Arctogadus glacialis*) (Breines et al. 2008), all of which have heteroplasmic relatives (to at least Order level; Table 1). One invertebrate, the isopod, *Cymothoa indica*, where the whole mitogenome was assessed and determined to be homoplasmic was recovered (Zou et al. 2018). However, in all these examples, only a single specimen was examined (Table 1) and that does not necessarily indicate the heteroplasmy is not present in any individuals within these species. In the case of *F. catus*, a later study identified low rates of heteroplasmy in the control region (Tarditi et al. 2011), in contrast to what Lopez et al. (1996) identified as a NUMT in the same gene region — although it is not clear if both studies were addressing the same location of nucleotide variation (Table 1). Nonetheless, it is important to highlight that interpretations of nucleotide variation can vary, and consideration of the impacts of laboratory-based protocols and bioinformatic pipelines which may undermine accurate interpretations of intra-individual nucleotide variation is crucial (Santibanez-Koref et al. 2019, González et al. 2020).

1.4.11 Reports on heteroplasmy are influenced by the methodology available

Before 2005, reports of length heteroplasmy in animals dominated the literature, and site heteroplasmy began to be increasingly reported after this period (Figure 3). This transition is likely explained by the continuous development of DNA visualisation and sequencing technologies (Figure 4), early taxon bias towards vertebrates and some cultivated invertebrates (which generally demonstrate length heteroplasmy (Table 1)), and increased confidence with interpreting unexpected outputs from developing techniques. The molecular techniques to accurately assess heterogeneity at single nucleotide positions, including numerous methods of polymerase chain reaction (PCR) and later, next-generation sequencing became widely available, affordable, and errors were better understood during the 2000's (Figure 4), limiting earlier identification of site heteroplasmy particularly outside of model organisms.

Most studies only assess a single tissue type or pool multiple tissues together, and rarely do studies investigate tissue specificity of heteroplasmy, with some notable exceptions (e.g. Jenuth et al. (1997), Magnacca and Brown (2010b)). Therefore, it is possible that studies that have identified homoplasmy may be analysing heteroplasmic individuals if their tissue sampling is too constrained. Additionally, molecular preparation, sequencing, and bioinformatic protocols influence the interpretation of intra-individual variation which may lead to misreporting of true mtDNA variation. This analysis identified that current knowledge of heteroplasmy may be limited by taxon bias, inconsistent and narrow methodologies (including tissue selection), and fidelities to consistent gene regions in some taxon groups. Therefore, determining broad, conclusive patterns of heteroplasmy within animals is difficult with the current published information. As whole-mitogenome sequencing becomes more widely used (particularly where multiple individuals of a species are assessed), more comprehensive interpretations of heteroplasmy will become possible.

1.4.12 Common misidentifications

Sequencing results can sometimes give a false appearance of heteroplasmy if primers amplify nuclear copies of mitochondrial genes (NUMTs/nuclear pseudogenes). Mitochondrial gene fragments are often incorporated into the nuclear genome where they are mostly non-functional, being not expressed, and can rapidly accumulate mutations (Brown et al. 1982, Sorenson and Fleischer 1996, Mundy et al. 2000). NUMTs are usually reported to be less than 1000 bp in length (Leite 2012) but larger NUMTs have occasionally been reported

(Lopez et al. 1994, Bensasson et al. 2001). Additionally, gene duplications within the mitogenome can be mistakenly identified as heteroplasmy, if DNA sequencing targets the gene region that has undergone duplication. When two copies of a gene that are highly convergent exists, these can be sequenced simultaneously and could appear to represent multiple mitogenomes.

Although heteroplasmy can be caused by a gene duplication where one lineage of mtDNA has multiple copies of the gene, and the other does not (this mtDNA without the duplication could be the original type), both must be present for this to be considered heteroplasmy. Like NUMTs, gene duplications can vary in length, but generally only occur for small fragments of the genome (Breton et al. 2014). Interpretation is complicated in cases where heteroplasmic sites occur in few or specific locations in the genome, making differentiation of localised heteroplasmy versus a gene duplication difficult. Therefore, analysing larger genome fragments or multiple genes can have limited utility when identifying between true heteroplasmy and a homoplasmic mitogenome with a gene duplication (e.g. Abbott et al. (2005)), and quantitative methods may be needed. Evidence of true mitochondrial heteroplasmy (i.e. not involving NUMTs or gene duplications) therefore needs to be examined very carefully to determine whether it involves actual variation in the mitogenome. There is the possibility that some studies attribute heteroplasmy incorrectly, as NUMTs and gene duplications are difficult to diagnose (Magnacca and Brown 2010a, Ricardo et al. 2020b), and there are examples of both NUMTs and heteroplasmy occurring between closely related taxa (Magnacca and Brown 2010a, Magnacca and Brown 2012) or even within single individuals (Ricardo et al. 2020a).

1.4.13 The persistence of homoplasmy

Although there are increasingly reports of intra-individual mtDNA variation occurring across animal taxa — particularly with increasingly affordable and efficient DNA sequencing technologies — homoplasmy remains the ‘default’ assumption for mtDNA. By and large, mitochondrial heteroplasmy appears to be detrimental in many systems (Sharpley et al. 2012, Christie et al. 2015, Christie and Beekman 2017) and mechanisms reducing intra-individual mtDNA variation remain widespread (Otten et al. 2016, De Fanti et al. 2017). Our search parameters did not target homoplasmic records, and we did not attempt to quantify the extent of homoplasmy across animal taxa.

There are three main hypothesis that have been proposed to explain conserved homoplasmy maintained through uniparental inheritance.

(i) The most widely accepted hypothesis suggests that when mtDNA mutations arise that increase mitochondrial replication rate, but concurrently decrease its metabolic contribution, the increased fitness to the mitochondria reduces the overall fitness of the cell or host organism. This genomic conflict (a variant of the selfish gene hypothesis) suggests that biparental inheritance would facilitate the spread of selfish mtDNA, thus the evolution of uniparental inheritance to combat selfish mitochondrial genes (Murlas Cosmides and Tooby 1981, Hastings 1992, Law and Hutson 1992).

(ii) A second hypothesis posits that selection against deleterious mtDNA mutations is facilitated by uniparental inheritance, as uniparental inheritance increases the variation of mtDNA between cells, but not within cells. Cells with high levels of mutations are removed during a 'bottleneck' in early germline development via purifying selection, and hence reduce mtDNA variation within a single individual (Bergstrom and Pritchard 1998, Roze et al. 2005, Christie et al. 2015).

(iii) A third hypothesis suggests that there is stringent coadaptation between the mitochondrial and nuclear genomes (Healy and Burton 2020), which is maintained via uniparental inheritance. The process of oxidative phosphorylation involves numerous interactions between proteins, DNA and RNA encoded by both genomes (Hadjivasiliou et al. 2012, Hadjivasiliou et al. 2013, Christie et al. 2015), and that mutations that compromise coevolved mitochondrial and nuclear genome combinations are purged via selection.

These hypotheses could help explain conserved homoplasmy and its maintenance through maternal transmission. However, pervasive heteroplasmic systems exist (such as DUI in bivalves) which challenge these concepts. There are increasing examples of heteroplasmic systems being maintained over multiple generations, both by biparental inheritance and maternal transmission (Table 1). The function of *advantageous* and/or *heritable* forms of mitochondrial heteroplasmy (if they exist, and evidence suggests they do), is not yet understood. How these alternative systems fit within the concepts of homoplasmic mtDNA inheritance is currently under investigation and is only just beginning to be addressed.

1.5 Conclusion and future directions

This review summarizes patterns of current heteroplasmic records in animals — where various breakdowns in standard molecular processes and specifically evolved strategies perpetuating unusual mtDNA traits have likely arisen multiple times. Because most cases of heteroplasmy are identified as a subsequent finding during standard genetic analyses, it is difficult to interpret the apparent discontinuity of the phylogenetic distribution of reported heteroplasmy-afflicted species, nor suggest the broad mechanisms that facilitate extensive levels of heteroplasmy in populations. The mechanisms that facilitate heteroplasmy may be generally different in invertebrates broadly, compared to vertebrates. Whereas arthropods may have less stringent repair mechanisms, higher mutation rates, and/or weaker effects of drift/selection (Meiklejohn et al. 2007, Thomas et al. 2010, Lynch et al. 2011) (Table 1), chordates may have increased instances of replication errors (Nabholz et al. 2009) (Table 1) that facilitate these differences in the characteristics of heteroplasmy in each group.

Although mitochondrial heteroplasmy is not a novel issue, its complexity has presented many challenges for our current knowledge of molecular processes and technology applied for such investigations. Our understanding of the mechanisms maintaining heteroplasmy and what function(s) heteroplasmy may serve in natural systems is generally underdeveloped. Studies addressing heteroplasmy are usually species specific, and these cases are often only explored because of accidental discoveries during other genetic analyses. Even organisms that have been extensively genotyped, such as *Drosophila* are only just being recognised for extensive heteroplasmy within their populations (Townsend and Rand 2004, Nunes et al. 2013), and model heteroplasmic systems such as bivalves remain highly active research fields. We propose some future directions of research in naturally occurring mitochondrial heteroplasmy in animal populations:

1. Expand sampling of species currently underrepresented in the literature to assess type and prevalence of heteroplasmy, particularly in basial animal lineages.
2. Increase the diversity of gene regions assessed (if not an assessment of entire mitogenomes) using technologies with the most reliable SNP identification capabilities, assessing variation within individuals most accurately.
3. Consider the impacts of tissue selection, DNA preparation, sequencing and bioinformatic protocols when designing experiments (even if detecting intra-individual variation is not necessarily the intended outcome of the research).

4. Using high-throughput DNA sequencing techniques, assess multiple individuals of a species to determine the extent of heteroplasmy.
5. Independent, reassessment of taxa identified as having unusual mitochondrial traits, particularly those where heteroplasmy was observed peripherally without an opportunity to explore further.
6. Select 'model' taxa across diverse lineages known to have extensive and (likely) heritable forms of heteroplasmy to explore the breadth of mechanisms facilitating *extensive* forms of mitochondrial heteroplasmy.

The implications for mitochondrial heteroplasmy for our understanding of genetic inheritance could be wide-reaching and has important implications for the use of mtDNA in a variety of research fields. The research directions for studies of naturally occurring mitochondrial heteroplasmy in animals in a broad evolutionary context have not been developed, and the implementation of more accurate and affordable sequencing technologies will make this a viable productive area of research. Mitochondrial heteroplasmy appears more prevalent that is currently appreciated and poses important considerations for our understanding of mitochondrial evolution and its link to the evolution of its animal (and other multicellular eukaryote) hosts.

1.6 Acknowledgements

We would like to thank Shannon Brown and the Flinders Library Service for their guidance in developing the search protocols for this review. Thank you also to Amy Butler for her comments on an earlier draft of this manuscript. This work was supported by the Australian Government Research Training Program Scholarship, as well as the Playford Memorial Trust PhD scholarship.

1.7 References

Abbott, C. L., M. C. Double, J. W. H. Trueman, A. Robinson, and A. Cockburn. 2005. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* albatrosses. Molecular Ecology **14**:3605–3613.

Andújar, C., P. Arribas, D. W. Yu, A. P. Vogler, and B. C. Emerson. 2018. Why the COI barcode should be the community DNA metabarcode for the metazoa. Molecular Ecology **27**:3968–3975.

Aryaman, J., I. G. Johnston, and N. S. Jones. 2019. Mitochondrial heterogeneity. Frontiers in Genetics **9**:718.

Azeredo-Espin, A. M. L., R. F. W. Schroder, M. D. Huettel, and W. S. Sheppard. 1991. Mitochondrial DNA variation in geographic populations of Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). Experientia **47**:483–485.

Balaban, R. S., S. Nemoto, and T. Finkel. 2005. Mitochondria, oxidants, and aging. Cell **120**:483–495.

Barr, C. M., M. Neiman, and D. R. Taylor. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist **168**:39–50.

Beck, J., and C. M. McCain. 2020. Just bird food? — On the value of invertebrate macroecology. Frontiers of Biogeography **12**:e47684.

Bensasson, D., D. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution **16**:314–321.

Bensasson, D., D. X. Zhang, and G. M. Hewitt. 2000. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Molecular Biology and Evolution **17**:406–415.

Berg, T., T. Moum, and S. Johansen. 1995. Variable numbers of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. Current Genetics **27**:257–262.

Bergstrom, C. T., and J. Pritchard. 1998. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics **149**:2135–2146.

Bessho, Y., T. Ohama, and S. Osawa. 1992. Planarian mitochondria I. Heterogeneity of cytochrome c oxidase subunit I gene sequences in the freshwater planarian, *Dugesia japonica*. Journal of Molecular Evolution **34**:324–330.

Bettinazzi, S., E. Rodriguez, L. Milani, P. U. Blier, and S. Breton. 2019. Metabolic remodelling associated with mtDNA: insights into the adaptive value of doubly uniparental inheritance of mitochondria. Proceedings of the Royal Society B: Biological Sciences **286**:20182708.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Research **27**:1767–1780.

Bowles, J., D. Blair, and D. P. McManus. 1994. Molecular-genetic characterization of the cervid strain (northern form) of *Echinococcus granulosus*. Parasitology **109**:215–221.

Boyce, T. M., M. E. Zwick, and C. F. Aquadro. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123**:825–836.

Breines, R., A. Ursvik, M. Nymark, S. D. Johansen, and D. H. Coucheron. 2008. Complete mitochondrial genome sequences of the Arctic Ocean codfishes *Arctogadus glacialis* and *Boreogadus saida* reveal oriL and tRNA gene duplications. Polar Biology **31**:1245–1252.

Breton, S., H. D. Beaupré, D. T. Stewart, W. R. Hoeh, and P. U. Blier. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? Trends in Genetics **23**:465–474.

Breton, S., L. Milani, F. Ghiselli, D. Guerra, D. T. Stewart, and M. Passamonti. 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends in Genetics **30**:555-564.

Breton, S., D. T. Stewart, and P. U. Blier. 2009. Role-reversal of gender-associated mitochondrial DNA affects mitochondrial function in *Mytilus edulis* (Bivalvia: Mytilidae). Journal of Experimental Zoology Part B: Molecular and Developmental Evolution **312**:108–117.

Brown, J. R., K. Beckenbach, A. T. Beckenbach, and M. J. Smith. 1996. Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). Genetics **142**:525–535.

Brown, W., E. Prager, A. Wang, and A. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. Journal of Molecular Evolution **18**:225–239.

Buroker, N. E., J. R. Brown, T. A. Gilbert, P. J. O'Hara, A. T. Beckenbach, W. K. Thomas, and M. J. Smith. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics **124**:157–163.

Chandel, N. S. 2014. Mitochondria as signaling organelles. BMC Biology **12**:34.

Chinnery, P., and D. Turnbull. 1999. Mitochondrial DNA and disease. The Lancet **354**:S17–S21.

Chinnery, P. F., D. C. Samuels, J. Elson, and D. M. Turnbull. 2002. Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? The Lancet **360**:1323–1325.

Christie, J. R., and M. Beekman. 2017. Selective sweeps of mitochondrial DNA can drive the evolution of uniparental inheritance. Evolution **71**:2090–2099.

Christie, J. R., T. M. Schaerf, and M. Beekman. 2015. Selection against heteroplasmy explains the evolution of uniparental inheritance of mitochondria. PLOS Genetics **11**:e1005112.

Çiftci, Y., O. Eroğlu, and Ş. Firidin. 2013. Heteroplasmy and length variation in the tRNApro- D-loop regions of three sturgeon species (*A. stellatus*, *A. gueldenstaedtii* and *H. huso*) from the Turkish coast of the Black Sea. Turkish Journal of Biochemistry **38**:250–257.

Dautova, M., H. Overmars, J. Bakker, G. Smant, and F. J. Gommers. 2002. Nuclear and mitochondrial DNA polymorphisms in three mitotic parthenogenetic *Meloidogyne* spp. European Journal of Plant Pathology **108**:209–220.

De Fanti, S., S. Vicario, M. Lang, D. Simone, C. Magli, D. Luiselli, L. Gianaroli, and G. Romeo. 2017. Intra-individual purifying selection on mitochondrial DNA variants during human oogenesis. Human Reproduction **32**:1100–1107.

de Freitas, P. D., F. L. Mendez, K. Chávez-Congrains, P. M. Galetti, Jr., L. L. Coutinho, A. Pissinatti, and C. D. Bustamante. 2018. Next-generation sequencing of the complete mitochondrial genome of the endangered species Black Lion Tamarin *Leontopithecus chrysopygus* (primates) and mitogenomic phylogeny focusing on the callitrichidae family. G3 Genes|Genomes|Genetics **8**:1985–1991.

Deagle, B. E., S. N. Jarman, E. Coissac, F. Pompanon, and P. Taberlet. 2014. DNA metabarcoding and the cytochrome *c* oxidase subunit I marker: not a perfect match. Biology Letters **10**:20140562.

Dégletagne, C., D. Abele, and C. Held. 2015. A distinct mitochondrial genome with DUI-like inheritance in the ocean quahog *Arctica islandica*. Molecular Biology and Evolution **33**:375–383.

Densmore, L. D., J. W. Wright, and W. M. Brown. 1985. Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics **110**:689–707.

DiMauro, S., and E. A. Schon. 2001. Mitochondrial DNA mutations in human disease. American Journal of Medical Genetics **106**:18–26.

Donaldson, M. R., N. J. Burnett, D. C. Braun, C. D. Suski, S. G. Hinch, S. J. Cooke, J. T. Kerr, and J. Hutchings. 2017. Taxonomic bias and international biodiversity conservation research. FACETS **1**:105–113.

Doublet, V., R. Raimond, F. Grandjean, A. Lafitte, C. Souty-Grosset, and I. Marcade. 2012. Widespread atypical mitochondrial DNA structure in isopods (Crustacea: Peracarida) related to a constitutive heteroplasmy in terrestrial species. Genome **55**:234–244.

Doublet, V., C. Souty-Grosset, D. Bouchon, R. Cordaux, and I. Marcadé. 2008. A thirty million year-old inherited heteroplasmy. PLOS ONE **3**.

Doucet-Beaupré, H., S. Breton, E. G. Chapman, P. U. Blier, A. E. Bogan, D. T. Stewart, and W. R. Hoeh. 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca):

searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evolutionary Biology **10**:50.
Dubey, B., P. R. Meganathan, and I. Haque. 2011. DNA mini-barcoding: An approach for forensic identification of some endangered Indian snake species. Forensic Science International: Genetics **5**:181–184.
Edwards, C. J., D. A. Magee, S. D. E. Park, P. A. McGettigan, A. J. Lohan, A. Murphy, E. K. Finlay, B. Shapiro, A. T. Chamberlain, M. B. Richards, D. G. Bradley, B. J. Loftus, and D. E. MacHugh. 2010. A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). PLOS ONE **5**.
Elson, J. L., D. C. Samuels, D. M. Turnbull, and P. F. Chinnery. 2001. Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. American Journal of Human Genetics **68**:802–806.
Emblem, Å., B. O. Karlsen, J. Evertsen, D. J. Miller, T. Moum, and S. D. Johansen. 2012. Mitogenome polymorphism in a single branch sample revealed by SOLiD deep sequencing of the *Lophelia pertusa* coral genome. Gene **506**:344–349.
Fernandez-Alvarez, F. A., R. Garcia-Jimenez, and A. Machordom. 2015. *Carinina ochracea* (Palaeonemertea: Tubulanidae) reaches its southernmost distribution: new morphological and molecular data. Zoological Science **32**:590–595.
Fontaine, K. M., J. R. Cooley, and C. Simon. 2007. Evidence for paternal leakage in hbrid periodical cicadas (Hemiptera: *Magicicada* spp.). PLOS ONE **2**:e892.
Gajić, B., I. Munoz, P. De la Rua, J. Stevanović, N. Lakic, Z. Kulišić, and Z. Stanimirović. 2019. Coexistence of genetically different *Varroa destructor* in *Apis mellifera* colonies. Experimental and Applied Acarology **78**:315–326.
Gajić, B., J. Stevanović, Ž. Radulović, Z. Kulišić, B. Vejnović, U. Glavinić, and Z. Stanimirović. 2016. Haplotype identification and detection of mitochondrial DNA heteroplasmy in *Varroa destructor* mites using ARMS and PCR–RFLP methods. Experimental and Applied Acarology **70**:287–297.
Gandolfi, A., B. Crestanello, A. Fagotti, F. Simoncelli, S. Chiesa, M. Girardi, E. Giovagnoli, C. Marangoni, I. Di Rosa, and L. Lucentini. 2017. New evidences of mitochondrial DNA heteroplasmy by putative paternal leakage between the rock partridge (*Alectoris graeca*) and the chukar partridge (*Alectoris chukar*). PLOS ONE **12**.
Gawande, S. J., S. Anandhan, A. A. Ingle, A. Jacobson, and R. Asokan. 2017. Heteroplasmy due to coexistence of mtCOI haplotypes from different lineages of the *Thrips tabaci* cryptic species group. Bulletin of Entomological Research **107**:534–542.
GBIF.org. 2020. GBIF Home Page. https://www.gbif.org accessed: 27 Nov 2020.
González, M. d. M., A. Ramos, M. P. Aluja, and C. Santos. 2020. Sensitivity of mitochondrial DNA heteroplasmy detection using Next Generation Sequencing. Mitochondrion **50**:88–93.
Green, D. R., and G. Kroemer. 2004. The pathophysiology of mitochondrial cell death. Science **305**:626–629.
Gusman, A., C. Azuelos, and S. Breton. 2017. No evidence of sex-linked heteroplasmy or doubly-uniparental inheritance of mtDNA in five gastropod species. Journal of Molluscan Studies **83**:119–122.
Hadjivasiliou, Z., N. Lane, R. M. Seymour, and A. Pomiankowski. 2013. Dynamics of mitochondrial inheritance in the evolution of binary mating types and two sexes. Proceedings of the Royal Society B: Biological Sciences **280**:1769.
Hadjivasiliou, Z., A. Pomiankowski, R. M. Seymour, and N. Lane. 2012. Selection for mitonuclear co-adaptation could favour the evolution of two sexes. Proceedings of the Royal Society B: Biological Sciences **279**:1865.

Haen, K. M., W. Pett, and D. V. Lavrov. 2014. Eight new mtDNA sequences of glass sponges reveal an extensive usage of +1 frameshifting in mitochondrial translation. Gene **535**:336–344.
Hale, L. R., and R. S. Singh. 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America **83**:8813–8817.
Hale, L. R., and R. S. Singh. 1991. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*: VI Mitochondrial DNA variation and the role of history vs selection in genetic structure of geographic populations. Genetics **129**:103–117.
Harrison, R. G., D. M. Rand, and W. C. Wheeler. 1985. Mitochondrial DNA size variation within individual crickets. Science **228**:1446–1448.
Hastings, I. M. 1992. Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genetical Research **59**:215–225.
Hatakeyama, H., and Y. i. Goto. 2016. Concise review: heteroplasmic mitochondrial DNA mutations and mitochondrial diseases: toward iPSC-based disease modeling, drug discovery, and regenerative therapeutics. STEM CELLS **34**:801–808.
Hayasaka, K., T. Ishida, and S. Horai. 1991. Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated sequences. Molecular Biology and Evolution **8**:399–415.
Hayashi, K., Y. L. Kawai, K. Yura, M. A. Yoshida, A. Ogura, K. Hata, K. Nakabayashi, and K. Okamura. 2016. Complete genome sequence of the mitochondrial DNA of the sparkling enope squid, *Watasenia scintillans*. Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis **27**:1842–1843.
He, X. L., C. Q. Ding, and J. L. Han. 2013. Lack of structural variation but extensive length polymorphisms and heteroplasmic length variations in the mitochondrial DNA control region of highly inbred Crested Ibis, *Nipponia nippon*. PLOS ONE **8**.
Healy, T. M., and R. S. Burton. 2020. Strong selective effects of mitochondrial DNA on the nuclear genome. Proceedings of the National Academy of Sciences of the United States of America **117**:6616–6621.
Hervé, M. 2020. RVAideMemoire: testing and plotting procedures for biostatistics. R package version 0.9-78. https://CRAN.R-project.org/package=RVAideMemoire accessed: 20 Nov 2020.
Hoarau, G., S. Holla, R. Lescasse, W. T. Stam, and J. L. Olsen. 2002. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. Molecular Biology and Evolution **19**:2261–2264.
Hoeh, W. R., D. T. Stewart, C. Saavedra, B. W. Sutherland, and E. Zouros. 1997. Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in *Mytilus* (Bivalvia: Mytilidae). Molecular Biology and Evolution **14**:959–967.
Horton, T., A. Kroh, S. Ahyong, N. Bailly, C. B. Boyko, S. N. Brandão, S. Gofas, J. N. A. Hooper, F. Hernandez, O. Holovachov, J. Mees, T. N. Molodtsova, G. Paulay, W. Decock, S. Dekeyzer, G. Poffyn, L. Vandepitte, B. Vanhoorne, and *e. al.* 2020. World Register of Marine Species (WoRMS). http://www.marinespecies.org accessed: 27 Nov 2020.
Hurtado, L. A., R. A. Lutz, and R. C. Vrijenhoek. 2004. Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. Molecular Ecology **13**:2603–2615.
IUCN. 2020. The IUCN Red List of Threatened Species. https://www.iucnredlist.org accessed: 27 Nov 2020.

Jebb, D., N. M. Foley, C. V. Whelan, F. Touzalin, S. J. Puechmaille, and E. C. Teeling. 2018. Population level mitogenomics of long-lived bats reveals dynamic heteroplasmy and challenges the Free Radical Theory of Ageing. Scientific Reports **8**.

Jenuth, J. P., A. C. Peterson, and E. A. Shoubridge. 1997. Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. Nature Genetics **16**:93–95.

Just, R. S., J. A. Irwin, and W. Parson. 2015. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. Forensic Science International-Genetics **18**:131–139.

Kang, A. R., M. J. Kim, I. A. Park, K. Y. Kim, and I. Kim. 2016. Extent and divergence of heteroplasmy of the DNA barcoding region in *Anapodisma miramae* (Orthoptera: Acrididae). Mitochondrial DNA Part A **27**:3405–3414.

Katsares, V., A. Apostolidis, A. Triantafyllidis, A. Kouvatsi, and C. Triantaphyllidis. 2003. Development of mitochondrial DNA primers for sse with homarid lobsters. Marine Biotechnology **5**:469–479.

Kesäniemi, J., Z. Boratynski, J. Danforth, P. Itam, T. Jernfors, A. Lavrinienko, T. Mappes, A. P. Moller, T. A. Mousseau, and P. C. Watts. 2018. Analysis of heteroplasmy in bank voles inhabiting the Chernobyl exclusion zone: a commentary on Baker et al. (2017) "Elevated mitochondrial genome variation after 50 generations of radiation exposure in a wild rodent.". Evolutionary Applications **11**:820–826.

Kiewnick, S., M. Holterman, S. van den Elsen, H. van Megen, J. E. Frey, and J. Helder. 2014. Comparison of two short DNA barcoding loci (COI and COII) and two longer ribosomal DNA genes (SSU & LSU rRNA) for specimen identification among quarantine root-knot nematodes (*Meloidogyne* spp.) and their close relatives. European Journal of Plant Pathology **140**:97–110.

Kmiec, B., M. Woloszynska, and H. Janska. 2006. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Current Genetics **50**:149–159.

Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata, and S. I. Chigusa. 1990. Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. Genetics **126**:657–663.

Kozhukhar, N., S. Mitta, and M. F. Alexeyev. 2020. Unusual mtDNA control Rregion length heteroplasmy in the COS-7 cell line. Genes **11**:607.

Kumazawa, Y. 2004. Mitochondrial DNA Sequences of Five Squamates: Phylogenetic Affiliation of Snakes. DNA Research **11**:137–144.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Molecular Biology and Evolution **13**:1242–1254.

Kvist, L., J. Martens, A. A. Nazarenko, and M. Orell. 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Molecular Biology and Evolution **20**:243-247.

Law, R., and V. Hutson. 1992. Intracellular symbionts and the evolution of uniparental cytoplasmic inheritance. Proceedings of the Royal Society of London. Series B: Biological Sciences **248**:69.

Leite, L. A. R. 2012. Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue. Biota Neotropica **12**:301–308.

Lopes, I. F., A. M. Tomasulo-Seccomandi, A. L. Bryan, I. L. Brisbin, T. C. Glenn, and S. N. Del Lama. 2011. Genetic status of the wood stork (*Mycteria americana*) from the southeastern United States and the Brazilian Pantanal as revealed by mitochondrial DNA analysis. Genetics and Molecular Research **10**:1910–1922.

Lopez, J., N. Yuhki, R. Masuda, W. Modi, and S. O'Brien. 1994. NUMT, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution **39**:174–190.

Lopez, J. V., S. Cevario, and S. J. Obrien. 1996. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. Genomics **33**:229–246.

Ludwig, A., B. May, L. Debus, and I. Jenneckens. 2000. Heteroplasmy in the mtDNA control region of sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics **156**:1933–1947.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates Sunderland.

Lynch, M., L.-M. Bobay, F. Catania, J.-F. Gout, and M. Rho. 2011. The repatterning of eukaryotic genomes by random genetic drift. Annual review of genomics and human genetics **12**:347–366.

Magnacca, K. N., and M. J. Brown. 2010a. Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). BMC Evolutionary Biology **10**:174.

Magnacca, K. N., and M. J. F. Brown. 2010b. Tissue segregation of mitochondrial haplotypes in heteroplasmic Hawaiian bees: implications for DNA barcoding. Molecular Ecology Resources **10**:60–68.

Magnacca, K. N., and M. J. F. Brown. 2012. DNA barcoding a regional fauna: Irish solitary bees. Molecular Ecology Resources **12**:990–998.

Magnacca, K. N., and B. N. Danforth. 2006. Evolution and biogeography of native Hawaiian Hylaeus bees (Hymenoptera: Colletidae). Cladistics **22**:393–411.

Manchado, M., G. Catanese, and C. Infante. 2004. Complete mitochondrial DNA sequence of the Atlantic bluefin tuna *Thunnus thynnus*. Fisheries Science **70**:68–73.

Martin, S., A. 2011. Mitochondrial DNA Repair. *in* DNA Repair: On the Pathways to Fixing DNA Damage and Errors F. Storici, editor. IntechOpen.

Mastrantonio, V., M. S. Latrofa, D. Porretta, R. P. Lia, A. Parisi, R. Iatta, F. Dantas-Torres, D. Otranto, and S. Urbanelli. 2019. Paternal leakage and mtDNA heteroplasmy in *Rhipicephalus* spp. ticks. Scientific Reports **9**:1460.

Maté, M. L., F. Di Rocco, A. Zambelli, and L. Vidal-Rioja. 2007. Mitochondrial heteroplasmy in Control Region DNA of South American camelids. Small Ruminant Research **71**:123–129.

Mayer, F., and G. Kerth. 2005. Microsatellite evolution in the mitochondrial genome of Bechstein's bat (*Myotis bechsteinii*). Journal of Molecular Evolution **61**:408–416.

McLeod, B. A., and B. N. White. 2010. Tracking mtDNA heteroplasmy through multiple generations in the north atlantic right whale (*Eubalaena glacialis*). Journal of Heredity **101**:235–239.

Meiklejohn, C. D., K. L. Montooth, and D. M. Rand. 2007. Positive and negative selection on the mitochondrial genome. Trends in Genetics **23**:259–263.

Melton, T. 2004. Mitochondrial DNA heteroplasmy. Forensic Sci Rev **16**:1–20.

Miraldo, A., G. M. Hewitt, P. H. Dear, O. S. Paulo, and B. C. Emerson. 2012. Numts help to reconstruct the demographic history of the ocellated lizard (*Lacerta lepida*) in a secondary contact zone. Molecular Ecology **21**:1005–1018.

Moreira, D. A., P. A. Buckup, C. Furtado, A. L. Val, R. Schama, and T. E. Parente. 2017. Reducing the information gap on Loricarioidei (Siluriformes) mitochondrial genomics. BMC Genomics **18**:345.

Moritz, C. 1991. Evoluationary dynamics of mitochondrial DNA duplications in parthenogenetic geckos, *Heteronotia binoei*. Genetics **129**:221–230.

Moritz, C., and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proceedings of the National Academy of Sciences of the United States of America **84**:7183–7187.

Mundy, N. I., A. Pissinatti, and D. S. Woodruff. 2000. Multiple nuclear insertions of mitochondrial cytochrome b sequences in callitrichine primates. Molecular Biology and Evolution **17**:1075–1080.

Muntaabski, I., R. M. Russo, M. C. Liendo, M. A. Palacio, J. L. Cladera, S. B. Lanzavecchia, and A. C. Scannapieco. 2020. Genetic variation and heteroplasmy of *Varroa destructor* inferred from ND4 mtDNA sequences. Parasitology Research **119**:411–421.

Murlas Cosmides, L., and J. Tooby. 1981. Cytoplasmic inheritance and intragenomic conflict. Journal of Theoretical Biology **89**:83–129.

Nabholz, B., S. Glémin, and N. Galtier. 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evolutionary Biology **9**:1–13.

Nadolny, R., H. Gaff, J. Carlsson, and D. Gauthier. 2015. Comparative population genetics of two invading ticks: Evidence of the ecological mechanisms underlying tick range expansions. Infection, Genetics and Evolution **35**:153–162.

Nesbø, C. L., M. O. Arab, and K. S. Jakobsen. 1998. Heteroplasmy, length and sequence variation in the mtDNA control regions of three percid fish species (*Perca fluviatilis*, *Acerina cernua*, *Stizostedion lucioperca*). Genetics **148**:1907–1919.

Nunes, M. D. S., M. Dolezal, and C. Schlotterer. 2013. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. Molecular Ecology **22**:2106–2117.

Otten, Auke B. C., Tom E. J. Theunissen, Josien G. Derhaag, Ellen H. Lambrichs, Iris B. W. Boesten, M. Winandy, Aafke P. A. van Montfoort, K. Tarbashevich, E. Raz, M. Gerards, Jo M. Vanoevelen, Bianca J. C. van den Bosch, M. Muller, and Hubert J. M. Smeets. 2016. Differences in strength and timing of the mtDNA bottleneck between zebrafish germline and non-germline cells. Cell Reports **16**:622–630.

Päckert, M., G. Giacalone, M. Lo Valvo, and C. Kehlmaier. 2019. Mitochondrial heteroplasmy in an avian hybrid form (*Passer italiae*: Aves: Passeriformes). Mitochondrial DNA Part B: Resources **4**:3809–3812.

Parakatselaki, M. E., C. Saavedra, and E. D. Ladoukakis. 2016. Searching for doubly uniparental inheritance of mtDNA in the apple snail *Pomacea diffusa*. Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis **27**:4000–4002.

Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M. R. Wilson, D. L. Berry, K. A. Holland, V. W. Weedn, P. Gill, and M. M. Holland. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nature Genetics **15**:363–368.

Passamonti, M., and F. Ghiselli. 2009. Doubly uniparental inheritance: Two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA and Cell Biology **28**:79–89.

Pierce, D., H. Werbin, and J. Shay. 1990. An improved procedure for quantitating mitochondrial DNA in cultured mammalian cells. Biotechniques **9**:724–729.

Piganeau, G. l., M. Gardner, and A. Eyre-Walker. 2004. A broad survey of recombination in animal mitochondria. Molecular Biology and Evolution **21**:2319–2325.

Pourkazemi, M., D. O. F. Skibinski, and J. A. Beardmore. 1999. Application of mtDNA D-loop region for the study of Russian sturgeon population structure from Iranian coastline of the Caspian Sea. Journal of Applied Ichthyology-Zeitschrift Fur Angewandte Ichthyologie **15**:23–28.

R Development Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Radojičić, J. M., I. Krizmanić, P. Kasapidis, and E. Zouros. 2015. Extensive mitochondrial heteroplasmy in hybrid water frog (*Pelophylax* spp.) populations from Southeast Europe. Ecology and Evolution **5**:4529–4541.
Ramsey, A. J., and J. R. Mandel. 2019. When one genome is not enough: organellar heteroplasmy in plants. Annual Plant Reviews Online **2**:619–658.
Rand, D. M., and R. G. Harrison. 1989. Molecular population genetics of mtDNA size variation in crickets. Genetics **121**:551.
Ricardo, C. P., E. Françoso, and M. C. Arias. 2020a. Mitochondrial DNA intra-individual variation in a bumblebee species: a challenge for evolutionary studies and molecular identification. Mitochondrion **53**:243–254.
Ricardo, P. C., E. Françoso, and M. C. Arias. 2020b. Fidelity of DNA polymerases in the detection of intraindividual variation of mitochondrial DNA. Mitochondrial DNA Part B: Resources **5**:108–112.
Robison, G. A., O. Balvin, C. Schal, E. L. Vargo, and W. Booth. 2015. Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the bed bug (Hemiptera: Cimicidae). Journal of Medical Entomology **52**:734–738.
Rodriguez-Pena, E., P. Verisimo, L. Fernandez, A. Gonzalez-Tizon, C. Barcena, and A. Martinez-Lage. 2020. High incidence of heteroplasmy in the mtDNA of a natural population of the spider crab *Maja brachydactyla*. PLOS ONE **15**.
Rokas, A., E. Ladoukakis, and E. Zouros. 2003. Animal mitochondrial DNA recombination revisited. Trends in Ecology & Evolution **18**:411–417.
Rollins, L. A., A. P. Woolnough, B. G. Fanson, M. L. Cummins, T. M. Crowley, A. N. Wiltonz, R. Sinclair, A. Butler, and W. B. Sherwin. 2016. Selection on mitochondrial variants occurs between and within individuals in an expanding invasion. Molecular Biology and Evolution **33**:995–1007.
Roze, D., F. Rousset, and Y. Michalakis. 2005. Germline bottlenecks, biparental inheritance and selection on mitochondrial variants. Genetics **170**:1385–1399.
RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.
Santibanez-Koref, M., H. Griffin, D. M. Turnbull, P. F. Chinnery, M. Herbert, and G. Hudson. 2019. Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution. Mitochondrion **46**:302–306.
Savolainen, P., L. Arvestad, and J. Lundeberg. 2000. mtDNA tandem repeats in domestic dogs and wolves: mutation mechanism studied by analysis of the sequence of imperfect repeats. Molecular Biology and Evolution **17**:474–488.
Sharpley, M. S., C. Marciniak, K. Eckel-Mahan, M. McManus, M. Crimi, K. Waymire, C. S. Lin, S. Masubuchi, N. Friend, M. Koike, D. Chalkia, G. MacGregor, P. Sassone-Corsi, and D. C. Wallace. 2012. Heteroplasmy of Mouse mtDNA Is Genetically Unstable and Results in Altered Behavior and Cognition. Cell **151**:333–343.
Shay, J. W., D. J. Pierce, and H. Werbin. 1990. Mitochondrial DNA copy number is proportional to total cell DNA under a variety of growth conditions. Journal of Biological Chemistry **265**:14802–14807.
Shigenobu, Y., K. Saitoh, K. I. Hayashizaki, and H. Ida. 2005. Nonsynonymous site heteroplasmy in fish mitochondrial DNA. Genes and Genetic Systems **80**:297–301.
Shih, H. T., T. Komai, and M. Y. Liu. 2013. A new species of fiddler crab from the Ogasawara (Bonin) Islands, Japan, separated from the widely-distributed sister species *Uca* (*Paraleptuca*) *crassipes* (White, 1847) (Crustacea: Decapoda: Brachyura: Ocypodidae). Zootaxa **3746**:175–193.

Shitara, H., J. I. Hayashi, S. Takahama, H. Kaneda, and H. Yonekawa. 1998. Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. Genetics **148**:851–857.

Signorell, A., K. Aho, A. Alfons, N. Anderegg, T. Argon, C. Archchige, A. Arppe, A. Baddeley, K. Barton, B. Bolker, H. W. Borchers, F. Caeiro, S. Champely, D. Chessel, L. Chhay, N. Cooper, C. Cummins, M. Dewey, H. C. Doran, S. Dray, C. Dupont, D. Eddelbuettel, C. Eskstrom, M. Elff, J. Enos, R. W. Farebrother, J. Fox, R. Francois, M. Friendly, T. Galili, M. Gamer, J. Gastwirth, and et al. 2021. DescTools: Tools for descriptive statistics. R package version 0.99.40. https://cran.r-project.org/package=DescTools accessed:

Sondheimer, N., C. E. Glatz, J. E. Tirone, M. A. Deardorff, A. M. Krieger, and H. Hakonarson. 2011. Neutral mitochondrial heteroplasmy and the influence of aging. Human Molecular Genetics **20**:1653–1659.

Song, H., J. E. Buhay, M. F. Whiting, and K. A. Crandall. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the National Academy of Sciences of the United States of America **105**:13486–13491.

Song, H., M. J. Moulton, and M. F. Whiting. 2014. Rampant nuclear insertion of mtDNA across diverse lineages within Orthoptera (Insecta). PLOS ONE **9**:e110508.

Songram, O., S. Sittipraneed, and S. Klinbunga. 2006. Mitochondrial DNA diversity and genetic differentiation of the honeybee (*Apis cerana*) in Thailand. Biochemical Genetics **44**:256–269.

Sorenson, M. D., and R. C. Fleischer. 1996. Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proceedings of the National Academy of Sciences of the United States of America **93**:15239–15243.

Spicer, A. M., T. J. Kun, B. N. Sacks, and E. J. Wictum. 2014. Mitochondrial DNA sequence heteroplasmy levels in domestic dog hair. Forensic Science International: Genetics **11**:7–12.

Steel, D. J., S. A. Trewick, and G. P. Wallis. 2000. Heteroplasmy of mitochondrial DNA in the ophiuroid *Astrobrachion constrictum*. Journal of Heredity **91**:146–149.

Stefano, G. B., and R. M. Kream. 2016. Mitochondrial DNA heteroplasmy in human health and disease. Biomedical Reports **4**:259–262.

Stewart, J. B., and P. F. Chinnery. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nature Reviews Genetics **16**:530–542.

Taheri, S., S. James, V. Roy, T. Decaëns, B. W. Williams, F. Anderson, R. Rougerie, C. H. Chang, G. Brown, L. Cunha, D. W. G. Stanton, E. Da Silva, J. H. Chen, A. R. Lemmon, E. Moriarty Lemmon, M. Bartz, D. Baretta, I. Barois, E. Lapied, M. Coulis, and L. Dupont. 2018. Complex taxonomy of the 'brush tail' peregrine earthworm *Pontoscolex corethrurus*. Molecular Phylogenetics and Evolution **124**:60–70.

Tarditi, C. R., R. A. Grahn, J. J. Evans, J. D. Kurushima, and L. A. Lyons. 2011. Mitochondrial DNA sequencing of cat hair: an informative forensic tool. Journal of Forensic Sciences **56**:36–46.

Thomas, J. A., J. J. Welch, R. Lanfear, and L. Bromham. 2010. A generation time effect on the rate of molecular evolution in invertebrates. Molecular Biology and Evolution **27**:1173–1180.

Thompson, K. F., S. Patel, L. Williams, P. Tsai, R. Constantine, C. S. Baker, and C. D. Millar. 2016. High coverage of the complete mitochondrial genome of the rare Grays beaked whale (*Mesoplodon grayi*) using Illumina next generation sequencing. Mitochondrial DNA **27**:128–129.

Titley, M. A., J. L. Snaddon, and E. C. Turner. 2017. Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. PLOS ONE **12**:e0189577.

Townsend, J. P., and D. M. Rand. 2004. Mitochondrial genome size variation in New World and Old World populations of *Drosophila melanogaster*. Heredity **93**:98–103.

Ujvari, B., M. Dowton, and T. Madsen. 2007. Mitochondrial DNA recombination in a free-ranging Australian lizard. Biology Letters **3**:189–192.

Van Blerkom, J. 2011. Mitochondrial function in the human oocyte and embryo and their role in developmental competence. Mitochondrion **11**:797–813.

Van Leeuwen, T., B. Vanholme, S. Van Pottelberge, P. Van Nieuwenhuyse, R. Nauen, L. Tirry, and I. Denholm. 2008. Mitochondrial heteroplasmy and the evolution of insecticide resistance: non-Mendelian inheritance in action. Proceedings of the National Academy of Sciences of the United States of America **105**:5980–5985.

Vargas, M. E. I., and A. Espin. 1995. Genetic variability in mitochondrial DNA of the screwworm, *Cochliomyia hominivorax* (Diptera: Calliphoridae), from Brazil. Biochemical Genetics **33**:237–256.

Vollmer, N. L., A. Viricel, L. Wilcox, M. K. Moore, and P. E. Rosel. 2011. The occurrence of mtDNA heteroplasmy in multiple cetacean species. Current Genetics **57**:115–131.

Wallace, D. C., M. D. Brown, and M. T. Lott. 1999. Mitochondrial DNA variation in human evolution and disease. Gene **238**:211–230.

Williams-Newkirk, A. J., M. Burroughs, S. S. Changayil, and G. A. Dasch. 2015. The mitochondrial genome of the lone star tick (*Amblyomma americanum*). Ticks and Tick-borne Diseases **6**:793–801.

Williams, P. H., M. V. Berezin, S. G. Cannings, B. Cederberg, F. Odegaard, C. Rasmussen, L. L. Richardson, J. Rykken, C. S. Sheffield, C. Thanoosing, and A. M. Byvaltsev. 2019. The arctic and alpine bumblebees of the subgenus *Alpinobombus* revised from integrative assessment of species' gene coalescents and morphology (Hymenoptera: Apidae: *Bombus*). Zootaxa **4625**:1–68.

Wilson, J. J. 2012. DNA barcodes for insects. Pages 17–46 *in* DNA Barcodes. Methods in Molecular Biology W. Kress and D. Erickson, editors. Springer, Totowa, NJ.

Wolff, J. N., M. Nafisinia, P. Sutovsky, and J. W. O. Ballard. 2013. Paternal transmission of mitochondrial DNA as an integral part of mitochondrial inheritance in metapopulations of *Drosophila simulans*. Heredity **110**:57–62.

Xiong, H., S. C. Barker, T. D. Burger, D. Raoult, and R. Shao. 2013. Heteroplasmy in the mitochondrial genomes of human lice and ticks revealed by high throughput sequencing. PLOS ONE **8**:e73329.

Ye, K., J. Lu, F. Ma, A. Keinan, and Z. Gu. 2014. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. Proceedings of the National Academy of Sciences of the United States of America **111**:10654–10659.

Zevering, C. E., C. Moritz, A. Heideman, and R. A. Sturm. 1991. Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*: Gekkonidae). Journal of Molecular Evolution **33**:431–441.

Zhang, F., and R. E. Broughton. 2013. Mitochondrial–nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? Genome Biology and Evolution **5**:1781–1791.

Zou, H., I. Jakovlic, D. Zhang, R. Chen, S. Mahboob, K. A. Al-Ghanim, F. Al-Misned, W. X. Li, and G. T. Wang. 2018. The complete mitochondrial genome of *Cymothoa indica* has a highly rearranged gene order and clusters at the very base of the Isopoda clade. PLOS ONE **13**.

Zouros, E. 2000. The exceptional mitochondrial DNA system of the mussel family Mytilidae. Genes and Genetic Systems **75**:313–318.
Zouros, E. 2013. Biparental inheritance through uniparental tansmission: The Doubly uniparental inheritance (DUI) of mitochondrial DNA. Evolutionary Biology **40**:1–31.
Zouros, E., A. Oberhauser Ball, C. Saavedra, and K. R. Freeman. 1994. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proceedings of the National Academy of Sciences of the United States of America **91**:7463–7467.

1.8 Tables

Table 1: Records of mitochondrial heteroplasmy across animal taxa recovered from published literature. The type of heteroplasmy (het.) for each taxon was reported as "site", "length", "S&L" (both types within the individuals sampled), or ∅ (no heteroplasmy i.e., homoplasmy only), with the mitochondrial fragment targeted included (e.g. "CR" = control region). Some records targeted multiple fragments or sequenced whole mitogenomes (denoted by "WGS") but detected heteroplasmy in only one or a few genes, which is summarized here. The percentage of heteroplasmy (% het.) within each 'population' sampled (with sample size, n) is given. The records are scaled from darkest green (100% heteroplasmic population) to white (0% heteroplasmy i.e. homoplasmy). Homoplasmic results are also given in grey. Note that here a "population" refers to all specimens of a species assessed within any publication and does not necessarily truly represent an ecological population.

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Annelida	Clitellata	Haplotaxida	*Pontoscolex corethrurus*	Site	COI	4.9	61	Paternal leakage	Taheri et al. (2018)
	Polycheata	Terebellida	*Alvinella pompejana*	∅	COI	0.0	93	N/A	Hurtado et al. (2004)
		Phyllodocida	*Branchipolynoe symmytilida*	∅	COI	0.0	61		
		Sabellida	*Oasisia alvinae*	∅	COI	0.0	43		
			Riftia pachyptila	∅	COI	0.0	135		
			Tevnia jerichonana	∅	COI	0.0	115		
Arthropoda (Chelicerata)	Arachnida	Ixodida	*Amblyomma americanum*	Site	Mitog wide (6 sites only)	100.0	1	Unknown	Williams-Newkirk et al. (2015)
			Amblyomma maculatum	Site	16S rRNA	3.5	370	Unknown	Nadolny et al. (2015)
			Amblyomma cajennense [+]	Site	Mitog (coding regions - 166 sites)	100.0	1	Unknown	Xiong et al. (2013)
			Argas sp.[+]	Site	Mitog (coding regions - 11 sites)	100.0	1		
			Haemaphysalis formosensis [+]	Site	Mitog (coding regions - 6 sites)	100.0	1		
			Haemaphysalis parva [+]	Site	Mitog (coding regions - 6 sites)	100.0	1		
			Otobius megnini [+]	Site	Mitog (coding regions - 24 sites)	100.0	1		
			Rhipicephalus geigy [+]	Site	Mitog (coding regions - 8 sites)	100.0	1		
			Rhipicephalus microplus [+]	Site	Mitog (coding regions - 40 sites)	100.0	1		
			Rhipicephalus spp.* (hybrid zone)	Length	12S rRNA	3.5	170	Paternal leakage via hybridization	Mastrantonio et al. (2019a)
		Parasitiformes	*Varroa destructor*	S&L	COI &/or cytb	31.3	64	Unknown	Gajić et al. (2016)
			Varroa destructor	Site	COI	51.8	245	Maternal inheritance	Gajić et al. (2019)
				Site	cytb	16.3			
			Varroa destructor	Site	COI & ND5, ND4-ND4L	18.7	182	Mutation and passage through generations	Muntaabski et al. (2020)
Arthropoda (Crustacea)	Branchiopoda	Anostraca	*Branchinecta paludosa*	Site	COI	26.5	49	Unknown	Lindholm et al. (2016)
		Decapoda	*Cancer irroratus*	∅	COI	0.0	95	NUMT	Gislason et al. (2013)
			Homarus americanus	∅	Multi mt frag	0.0	10	N/A	Katsares et al. (2003)
			Homarus gammarus	Length	Multi mt frag	100.0	49	Mitochondrial inheritance	
			Maja brachydactyla	Site	COI & 16S rRNA	26.5	83	Failure to eliminate male mtDNA	Rodriguez-Pena et al. (2020)
			Panulirus argus	∅	Multi mt frag	0.0	259	N/A	Silberman et al. (1994)
			Panulirus argus	∅	CR	0.0	28	N/A	Diniz et al. (2005)
			Paraleptuca boninensis (as *Uca boninensis*)	Length	CR (COI no heterop)	100.0	5	Unknown	Shih et al. (2013)
			Paraleptuca chlorophthalmus (as *Uca chlorophthalmas*)	∅	CR & COI	0.0	4		
			Paraleptuca crassipes (as *Uca crassipes*)	Length	CR (COI no heterop)	7.1	28		
			Paraleptuca splendida (as *Uca splendida*)	∅	CR & COI	0.0	16		
			Penaeus monodon	∅	CR & 16S rRNA	0.0	86	N/A	Alam et al. (2015)
			Penaeus monodon	∅	CR	0.0	115	NUMT	Walther et al. (2011)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Arthropoda (Crustacea)	Malacostraca	Isopoda	*Cymothoa indica*	Ø	Mitog wide	0.0	1	N/A	Zou et al. (2018)
			Armadillidium vulgare *	Site	tRNA anticodon	100.0	6	Unknown	Marcadé et al. (2007)
			Armadillidium assimile *	Site	tRNA anticodon	100.0	3	Balancing selection	Doublet et al. (2008, 2012)
			Armadillidium depressum *	Site	tRNA anticodon	100.0	3		
			Armadillidium nasatum *	Site	tRNA anticodon	100.0	2		
			Armadillidium vulgare *	Site	tRNA anticodon	100.0	26		
			Armadillo officinalis ⁺	Site	tRNA anticodon	100.0	2		
			Balloniscus sellowii ⁺	Site	tRNA anticodon	100.0	1		
			Cubaris murina *	Site	tRNA anticodon	100.0	2		
			Cylisticus convexus *	Site	tRNA anticodon	100.0	3		
			Helleria breviconis ⁺	Site	tRNA anticodon	100.0	2		
			Hemilepistus reaumuri ⁺	Site	tRNA anticodon	100.0	3		
			Platyarthrus caudatus ⁺	Site	tRNA anticodon	100.0	1		
			Platyarthrus hoffmannseggii ⁺	Site	tRNA anticodon	100.0	1		
			Porcellio gallicus *	Site	tRNA anticodon	100.0	3		
			Porcellio spinicornis *	Site	tRNA anticodon	100.0	2		
			Trachelipus rathkii ⁺	Site	tRNA anticodon	100.0	2		
			Trichoniscus pusillus pusillus ⁺	Site	tRNA anticodon	100.0	1		
			Trachelipus rathkeii	Site	tRNA anticodons (WGS)	100.0	2	Selection maintaining both variants	Chandler et al. (2015)
Arthropoda (Hexpoda)	Insecta	Coleoptera	*Gonioctena intermedia*	Site	COI	45.8	24	Balancing selection	Kastally & Mardulyn (2017)
			Inflata indica	Site	Mitog wide (11 sites only)	100.0	1	Unknown	Sriboonlert & Wonnapinij (2019)
			Leptinotarsa decemlineata	Length	Multi mt frag	1.5	535	Transient state between fixation of mutant haplotype	Azeredo-Espin et al. (1996)
			Monochamus galloprovincialis	Ø	COI & COII	0.0	148	NUMT	Koutroumpa et al. (2009)
			Monochamus sutor	Ø	COI & COII	0.0	49	N/A	
			Ochthebius quadricollis	Ø	COI	0.0	84	N/A	Mastrantonio et al. (2019b)
			Ochthebius urbanelliae	Site	COI	1.9	105	Paternal leakage due to hybridization	
			Pissodes strobi	Length	CR	100.0	142	Unknown	Boyce et al. (1989)
			Pissodes nemorensis *	Length	CR	100.0	54		
			Pissodes terminalis *	Length	CR	100.0	23		
		Diptera	*Aedes aegypti*	Site	ND4	2.4	125	Paternal leakage	Paduan & Ribolla (2008)
			Aedes aegypti	Site	ND4	1.7	234	Paternal leakage	Dalla Bona et al. (2012)
			Aedes aegypti	Ø	ND4	0.0	5840	NUMT	Black & Bernhardt (2009)
			Anopheles arabiensis	Ø	ND5	0.0	232	N/A	Aboud et al. (2014)
			Anopheles quadrimaculatus	Ø	Multi mt frag	0.0	288	N/A	Mitchell et al. (1992)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Arthropoda (Hexpoda)	Insecta	Diptera	*Anopheles stephensi* *	Ø	COI-II	0.0	40	N/A	Oshaghi (2005)
			Cochliomyia hominivorax	Length	Multi mt frag	4.2	405	Transient state after paternal leakage	Vargas & Espin (1995)
			Drosophila melanogaster *	Length	D-Loop	16.7	144	Unknown (small lab culture accumulating mutations (Townsend & Rand (2004))	Hale & Singh (1991)
			Drosophila melanogaster *	Length	D-Loop	1.5	194	Small laboratory populations accumulating mutations	Townsend & Rand (2004)
			Drosophila melanogaster	Site	COI	13.6	66	Paternal leakage	Nunes et al. (2013)
			Drosophila simulans and *D. mauritiana* (hybrid experiment)*	Length	Multi mt frag	0.5	191	Paternal leakage via hybridization	Mitrofanov et al. (2002)
			Drosophila mauritiana *	Length	Multi mt frag	50.0	60	Incomplete maternal inheritance	
			Drosophila simulans	Site	COI	0.7	4092	Paternal leakage	Wolff et al. (2013)
		Hemiptera	*Cimex lectularius*	Site	COI	17.2	29	Paternal leakage	Robison et al. (2015)
			Nezara viridula	Site	16S rRNA	3.3	61	Paternal leakage	Kavar et al. (2006)
			Philaenus spumarius	Site	cytb	16.7	108	Paternal leakage	Lis et al. (2014)
		Hymenoptera	*Amphylaeus morosus*	Site	COI	100.0	73	Endosymbiont facilitated, maternal inhertiance	Unpublished Davies
			Andrena tarsata	Site	COI	100.0	2	Unknown (some NUMTs present in homoplasmic species)	Magnacca & Brown (2012)
			Andrena spp. (16 species)	Ø	COI	0.0	1-8		
			Coelioxys elongata	Ø	COI	0.0	1		
			Colletes succinctus	Site	COI	100.0	11		
			Colletes spp. (2 species)	Ø	COI	0.0	4-5		
			Halictus rubicundus	Site	COI	16.7	6		
			Halictus tumulorum	Site	COI	20.0	5		
			Hylaeus spp. (2 species)	Ø	COI	0.0	2-4		
			Lasioglossum spp. (9 species)	Ø	COI	0.0	1-15		
			Megachile spp. (4 species)	Ø	COI	0.0	1-3		
			Nomada spp. (9 species)	Ø	COI	0.0	2-12		
			Osmia aurulenta	Site	COI	50.0	2		
			Osmia rufa	Ø	COI	0.0	3		
			Sphecodes geoffrellus	Site	COI	40.0	5		
			Sphecodes spp. (5 species)	Ø	COI	0.0	2-8		
			Apis cerana *	Length	ATPase6 - ATPase8	43.1	181	Unknown	Songram et al. (2006)
			Bombus morio	Site	COI	88.5	26	Unknown (NUMTs also present)	Ricardo et al. (2020a)
			Ectatomma ruidum complex	Site	Mitog wide (<7% sites)	33.3	21	Maternal Inheritance	Meza-Lázaro et al. (2018)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Arthropoda (Hexpoda)	Insecta	Hymenoptera	*Hylaeus akoko* [α1]	Site	COI	100.0	1	Unknown	Magnacca & Brown (2010a)
			Hylaeus andrenoides	Site	COI	100.0	3		
			Hylaeus angustulus	Site	COI	66.7	6		
			Hylaeus coniceps	Site	COI	100.0	9		
			Hylaeus connectens [α2]	Site	COI	100.0	12		
			Hylaeus dimidiatus [α2]	Site	COI	100.0	3		
			Hylaeus facilis	Site	COI	100.0	1		
			Hylaeus filicum	Site	COI	100.0	2		
			Hylaeus fuscipennis [α1]	Site	COI	100.0	9		
			Hylaeus haleakalae	Site	COI	44.4	9		
			Hylaeus hula	Site	COI	100.0	4		
			Hylaeus kauaiensis	Site	COI	100.0	4		
			Hylaeus kokeensis	Site	COI	100.0	4		
			Hylaeus kuakea	Site	COI	100.0	1		
			Hylaeus kukui	Site	COI	100.0	3		
			Hylaeus mimicus	Site	COI	100.0	3		
			Hylaeus muranus	Site	COI	100.0	3		
			Hylaeus mutatus	Site	COI	100.0	5		
			Hylaeus pubescens [α1]	Site	COI	100.0	5		
			Hylaeus setosifrons	Site	COI	100.0	4		
			Hylaeus unicus	Site	COI	84.6	13		
			Hylaeus spp. (28 species)	Ø	COI	0.0	1-16		
		Lepidoptera	*Jalmenus daemeli*	Ø	CR	0.0	9	N/A	Taylor et al. (1993)
			Jalmenus evagoras eubulus	Ø	CR	0.0	3		
			Jalmenus evagoras evagoras	Ø	CR	0.0	15		
			Jalmenus icilius	Ø	CR	0.0	6		
			Jalmenus ictinus	Ø	CR	0.0	6		
			Jalmenus lithochroa	Ø	CR	0.0	2		
			Jalmenus pseudictinus	Ø	CR	0.0	2		

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Arthropoda (Hexpoda)	Insecta	Orthoptera	*Arcyptera* sp.	Ø	ND5 (& COI)	0.0	1	Nuclear paralog	Bensasson et al. (2000)
			Calliptamus sp.	Ø	ND5 (& COI)	0.0	2		
			Cophopodisma pyrenea	Ø	ND5 (& COI)	0.0	1		
			Italopodisma sp.	Ø	ND5 (& COI)	0.0	13		
			Ognevia longipennis	Ø	ND5 (& COI)	0.0	1		
			Parapodisma mikado	Ø	ND5 (& COI)	0.0	5		
			Podisma pedestris	Ø	ND5 (& COI)	0.0	89		
			Primnoa fauriei (as *P. hayachinensis*)	Ø	ND5 (& COI)	0.0	1		
			Pseudochorthippus parallelus (as *Chorthippus parallelus*)	Ø	ND5 (& COI)	0.0	2		
			Gryllus firmus	Length	Multi mt frag	61.5	52	Unknown	Rand and Harrison (1986)
			Gryllus pennsylvanicus	Length	Multi mt frag	45.3	53		
			Gryllus pennsylvanicus and *G. firmus* (hybrid zone)	Length	Multi mt frag	12.0	100	Unknown	Harrison et al. (1985)
		Phthiraptera	*Geomydoecus aurei*	S&L	COI	100.0	2	Mitochondrial minichromosome	Pietan et al. (2016)
			Thomomydoecus minor	S&L	COI	100.0	1		
			Pediculus capitis +	Site	Mitog (coding regions <65 sites)	100.0	6	Unknown	Xiong et al. (2013)
			Pediculus humanus +	Site	Mitog (coding regions <76 sites)	100.0	6		
		Thysanoptera	*Thrips tabaci* *	Site	COI	100.0	500	Unknown	Gawande et al. (2017)
Chordata	Actinopterygii	Acipenseriformes	*Acipenser baerii* ^	Length	D-Loop	19.8	126	Intramolecular recombination	Ludwig & Jenneckens (2000), Ludwig et al. (2000a)
			Acipenser brevirostrum	Length	D-Loop	50.0	6		
			Acipenser gueldenstaedtii ^	Length	D-Loop	14.3	98		
			Acipenser fulvescens	Length	D-Loop	6.4	31		
			Acipenser medirostris	Length	D-Loop	54.1	37		
			Acipenser mikadoi ^	Length	D-Loop	50.0	6		
			Acipenser naccarii ^	Length	D-Loop	30.0	20		
			Acipenser nudiventris ^	Ø	D-loop	0.0	15		
			Acipenser oxyrinchus	Ø	D-loop	0.0	60		
			Acipenser persicus ^	Length	D-Loop	23.5	17		
			Acipenser ruthenus	Length	D-Loop	46.8	156		
			Acipenser stellatus ^	Length	D-Loop	25.8	93		
			Acipenser sturio ^	Ø	D-loop	0.0	44		
			Acipenser transmontanus	Length	D-Loop	16.6	12		
			Huso huso ^	Length	D-Loop	8.0	74		
			Acipenser brevirostrum	Length	CR	30.5	272	Unknown	Grunwald et al. (2002)
			Acipenser gueldenstaedtii ^	L&S	D-Loop	100.0	145	Paternal leakage and mutation	Pourkazemi et al. (1999)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Actinopterygii	Acipenseriformes	*Acipenser gueldenstaedtii* ^	Length	D-Loop	8.3	24	High mutation rate (via insertions and deletions)	Çiftci et al. (2013)
			Acipenser stellatus ^	Length	D-Loop	41.7	12		
			Huso huso ^	Length	D-Loop	2.2	45		
			Acipenser medirostris	Length	D-Loop	50.0	10	Unknown	Brown et al. (1996)
			Acipenser transmontanus	Length	D-Loop	42.0	174		
			Acipenser fulvescens	Ø	D-Loop	0.0	21		
			Acipenser oxyrinchus	Ø	D-Loop	0.0	19		
			Acipenser oxyrinchus oxyrinchus	Ø	D-loop	0.0	90	N/A	Ludwig et al. (2000b)
			Acipenser sturio	Ø	D-loop	0.0	12		
			Acipenser oxyrinchus desotoi	Length	D-Loop	18.5	168	Unknown	Miracle & Campton (1995)
			Acipenser oxyrinchus desotoi	Ø	CR	0.0	160	N/A	Stabile et al. (1996)
			Acipenser ruthenus	Site	cytb	10.0	30	Transitioinal heteroplasmy as the result of population hybridisation	Dudu et al. (2012)
			Acipenser sinensis	Length	D-Loop	57.4	47	Unknown	Zhang et al. (1999)
			Acipenser transmontanus	Length	D-loop	40.6	128	Dynamic, reversible equilibrium of D-loop strand and H-strand for pairing with L-strand	Buroker et al. (1990)
			Acipenser transmontanus	Length	D-Loop	42.0	174	Mutation mechanism that increases copy number	Brown et al. (1992)
		Amiiformes	*Amia calva*	Length	Multi mt frag	7.7	52	Unknown	Bermingham et al. (1986)
		Characiformes	*Brycon opalinus* *	Length	Multi mt frag	0.6	337	Intra-individual mutation	Hilsdorf & Krieger (2004)
			Megaleporinus piavussu (as *Leporinus piavussu*)	Length	16S rRNA (WGS - 1 site only)	100.0	1	Unknown	Yazbeck et al. (2016)
		Clupeiformes	*Alosa sapidissima*	Length	Multi mt frag	12.3	244	Incomplete segregation of variant mtDNA forms	Bentzen et al. (1988)
			Alosa sapidissima	Length	Multi mt frag	14.0	81	Unknown	Nolan et al. (1991)
			Engraulis encrasicolus	Length	Multi mt frag	0.7	435	Biparental inheritance	Magoulas & Zouros (1993)
		Cyprinodontiformes	*Gambusia holbrooki*	Site	CR	75.5	597	Chemical induced mutation	Rinner et al. (2011)
			Poecilia reticulata	Length	CR	37.0	46	Unknown	Taylor & Breden (2000)
			Poecilia reticulata	Length	CR	69.2	13	Inheritance (maternal or paternal)	Taylor & Breden (2002)
		Cypriniformes	*Acheilognathus typus*	Site	CR	0.6	541	Unknown	Saitoh et al. (2016)
			Formosania lacustre	Length	CR	35.5	68	Unknown	Wang et al. (2007)
			Cyprinella spiloptera	Ø	CR	0.0	38	Fragment duplication	Broughton & Dowling (1994)
			Cyprinella spiloptera	Ø	CR & tRNA Pro	0.0	50	N/A	Broughton & Dowling (1997)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
			Abramis brama	Ø	COI	0.0	9	Unknown	Behrens-Chapuis et al. (2018)
			Alburnus alburnus	Ø	COI	0.0	8		
			Alburnoides bipunctatus	Ø	COI	0.0	8		
			Blicca bjoerkna	Ø	COI	0.0	10		
			Carassius carassius	Ø	COI	0.0	10		
			Carassius gibelio	Ø	COI	0.0	10		
			Chondrostoma nasus	Ø	COI	0.0	10		
			Cyprinus carpio	Site	COI	11.1	9		
			Gobio gobio	Ø	COI	0.0	9		
			Leucaspius delineatus	Site	COI	10.0	10		
			Leuciscus aspius	Ø	COI	0.0	10		
			Leuciscus idus	Ø	COI	0.0	10		
			Leuciscus leuciscus	Site	COI	10.0	10		
		Cypriniformes	*Phoxinus phoxinus*	Ø	COI	0.0	13		
			Pseudorasbora parva	Ø	COI	0.0	8		
Chordata	Actinopterygii		*Rhodeus amarus*	Site	COI	10.0	10		
			Rutilus rutilus	Ø	COI	0.0	10		
			Scardinius erythrophthalmus	Site	COI	10.0	10		
			Squalius cephalus	Ø	COI	0.0	10		
			Tinca tinca	Ø	COI	0.0	10		
			Vimba vimba	Site	COI	20.0	10		
			Schizopyge niger (as *Schizothorax niger*)	Ø	16S rRNA, cytb, & D-loop	0.0	4	N/A	Ahmad et al. (2014)
			Schizothorax curvifrons	Ø	16S rRNA, cytb, & D-loop	0.0	4		
			Schizothorax esocinus	Ø	16S rRNA, cytb, & D-loop	0.0	4		
			Schizothorax labiatus	Ø	16S rRNA, cytb, & D-loop	0.0	4		
			Schizothorax plagiostomus	Ø	16S rRNA, cytb, & D-loop	0.0	5		
			Schizopyge niger (as *Schizothorax niger*)	Ø	D-loop	0.0	5	N/A	Syed et al. (2016)
			Schizothorax curvifrons	Ø	D-loop	0.0	5		
			Schizothorax esocinus	Ø	D-loop	0.0	5		
			Schizothorax labiatus	Ø	D-loop	0.0	5		
			Schizothorax plagiostomus	Ø	D-loop	0.0	5		
		Esociformes	*Esox lucius*	Length	CR	16.7	6	Unknown	Brzuzan et al. (1998)
			Arctogadus glacialis	Ø	Mitog wide	0.0	1	Unknown	Breines et al. (2008)
		Gadiformes	*Boreogadus saida*	Length	WANCY tRNA cluster (WGS)	50.0	2		
			Gadus morhua	Length	D-Loop	100.0	27	Mutation rate > genetic drift	Arnason & Rand (1992)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Actinopterygii	Gasterosteiformes	*Culaea inconstans*	Length	CR	43.8	121	Replication slippage	Gach & Brown (1997)
			Culaea inconstans	Length	D-loop	38.1	352	Unknown	Gach (1996)
			Gasterosteus aculeatus	Ø	D-loop	0.0	6		
			Pungitius pungitius	Ø	D-loop	0.0	4		
			Gasterosteus aculeatus	Length	CR	100.0	11	Illegitimate elongation resulting in misalignment	Starner et al. (2004)
		Gonorynchiformes	*Chanos chanos* *	Length	CR	36.4	96	Unknown	Ravago et al. (2002)
		Perciformes	*Centropomus undecimalis*	Length	CR	0.1	789	Unknown	Wilson et al. (1997)
			Dascyllus trimaculatus	Length	CR	58.3	36	Unknown	Chen et al. (2004)
			Dicentrarchus labrax *	Length	D-loop	52.2	209	Bias toward large mtDNA molecules	Cesaroni et al. (1997)
			Morone saxatilis	Length	Multi mt frag	14.1	821	Unknown	Wirgin et al. (1993)
			Morone saxatilis	Length	Multi mt frag	15.6	32	Unknown	Stellwag et al. (1994)
			Perca fluviatilis	Length	D-Loop (HVR)	73.8	149	Maternal Inheritance	Nesbø et al. (1998)
			Sciaenops ocellatus	Length	Multi mt frag	0.3	750	Unknown	Gold & Richardson (1990)
			Sander vitreus (as *Stizostedion vitreum*)	Ø	CR	0.0	199	N/A	Faber & Stepien (1998)
			Sander volgensis (as *Stizostedion volgensi*	Ø	CR	0.0	1		
			Stizostedion canadense	Ø	CR	0.0	4		
			Stizostedion lucioperca	Ø	CR	0.0	2		
			Thunnus thynnus ^	Ø	Mitog wide	0.0	1	N/A	Manchado et al. (2004)
		Pleuronectiformes	*Hippoglossus hippoglossus* *^	Length	CR	100.0	4	Slipped-strand mispairing and DNA recombination mechanisms	Mjelle et al. (2008)
			Hippoglossus stenolepis	Length	CR	100.0	4		
			Reinhardtius hippoglossoides	Length	CR	100.0	4		
			Paralichthys olivaceus	Site	ND4	2.6	39	Point mutation	Shigenobu et al. (2005)
			Platichthys flesus	Length	CR	98.2	168	Recombination of mtDNA inherited via paternal leakage	Hoarau et al. (2002)
		Salmoniformes	*Oncorhynchus keta*	Site	ND5	100.0	1	Point mutation	Shigenobu et al. (2005)
			Oncorhynchus mykiss	Ø	CR	0.0	15	N/A	Nielsen et al. (1998)
			Salmo trutta	Length	CR	11.7	60	Population specific selection for heteroplasmy	Wetjen et al. (2017)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
			Ancistrus sp. 1+	Site	Mitog wide (10 sites only)	100.0	1	Unknown	Moreira et al. (2015)
			Ancistrus sp. 2+	Site	Mitog wide (6-8 sites only)	100.0	2		
			Ancistrus multispinis	Site	Mitog wide (8 sites only)	100.0	1	Unknown	Moreira et al. (2017)
			Ancistrus sp. 1*	Site	Mitog wide (13 sites only)	100.0	1		
			Ancistrus sp. 2*	Site	Mitog wide (6 sites only)	100.0	1		
			Baryancistrus xanthellus *	Site	Mitog wide (10 sites only)	100.0	1		
			Corydoras nattereri	Site	Mitog wide (3 sites only)	100.0	1		
			Corydoras schwartzi *	Site	Mitog wide (7 sites only)	100.0	1		
			Dekeyseria amazonica	Site	Mitog wide (5 sites only)	100.0	1		
			Hemiancistrus snethlageae * (as *Ancistomus snethlageae*)	Site	Mitog wide (14 sites only)	100.0	1		
			Hemipsilichthys nimius	Site	Mitog wide (6 sites only)	100.0	1		
			Hisonotus thayeri	Site	Mitog wide (14 sites only)	100.0	1		
			Hypoptopoma incognitum	Site	Mitog wide (9 sites only)	100.0	1		
			Hypostomus affinis	Site	Mitog wide (6 sites only)	100.0	1		
Chordata	Actinopterygii	Siluriformes	*Hypostomus emarginatus* (as *Aphanotolurus emarginatus*)	Site	Mitog wide (11 sites only)	100.0	1		
			Hypostomus cf. *plecostomus*	Site	Mitog wide (8 sites only)	100.0	1		
			Hypostomus sp.*	Site	Mitog wide (5 sites only)	100.0	1		
			Kronichthys heylandi	Site	Mitog wide (13 sites only)	100.0	1		
			Loricaria cataphracta	Site	Mitog wide (8 sites only)	100.0	1		
			Loricariichthys castaneus	Site	Mitog wide (21 sites only)	100.0	1		
			Loricariichthys platymetopon	Site	Mitog wide (10 sites only)	100.0	1		
			Neoplecostomini gen. n.	Ø	Mitog wide	0.0	1		
			Neoplecostomus microps	Site	Mitog wide (6 sites only)	100.0	1		
			Otocinclus cf. *hoppei* *	Site	Mitog wide (11 sites only)	100.0	1		
			Panaqolus sp.*	Site	Mitog wide (15 sites only)	100.0	1		
			Pareiorhaphis garbei	Site	Mitog wide (12 sites only)	100.0	1		
			Parotocinclus maculicauda	Site	Mitog wide (6 sites only)	100.0	1		
			Peckoltia furcata *	Site	Mitog wide (6 sites only)	100.0	1		
			Pterygoplichthys pardalis	Site	Mitog wide (9 sites only)	100.0	1		
			Pterygoplichthys sp.*	Site	Mitog wide (7 sites only)	100.0	1		
			Rineloricaria cf. *lanceolata* *	Site	Mitog wide (6 sites only)	100.0	1		
			Rineloricaria sp.	Site	Mitog wide (5 sites only)	100.0	1		
			Schizolecis guntheri	Site	Mitog wide (15 sites only)	100.0	1		
			Hypancistrus zebra	Site	Mitog wide (21 sites only)	100.0	2	Unknown	Magalhães et al. (2017)
			Hypoptopoma incognitum	Site	Mitog wide (33 sites only)	100.0	1	Unknown	Moreira et al. (2016)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Actinopterygii	Scorpaeniformes	*Sebastes fasciatus* ^	Length	CR	38.9	36	Unknown replication error leading to secondary-structure stabilization	Bentzen et al. (1998)
			Sebastes mentella	Length	CR	42.3	52		
			Sebastes norvegicus (as *Sebastes marinus*	Length	CR	46.2	13		
		Teleostei	*Padogobius nigricans*	Ø	tRNA Pro, tRNA Phe & 12S rRNA	0.0	16	N/A	Cervelli et al. (2007)
	Amphibia	Anura	*Anaxyrus baxteri* ^ (as *Bufo baxteri*)	Site	12S & 16S rRNA	100.0	1	Unknown	Pauly et al. (2004)
			Incilius coniferus (as *Bufo coniferus*)	Site	12S & 16S rRNA	100.0	1		
			"*Bufo* " spp. (48 species)	Ø	12S & 16S rRNA	0.0	1-6		
			Bombina bombina	Length	CR	100.0	2	Unknown	Spolsky et al. (2006)
			Bombina variegata	Length	CR	100.0	3		
			Dryophytes cinereus (as *Hyla cinerea*)	Length	Multi mt frag	1.4	142	Unknown	Bermingham et al. (1986)
			Dryophytes gratiosus (as *Hyla gratiosa*)	Length	Multi mt frag	8.0	163		
			Pelophylax ridibundus (as *Rana ridibunda*)	Site	cytb	3.9	207	Chemical induced mutation	Matson et al. (2006)
			Pelophylax ridibundus and *P. lessonae* (hybrid zone)	Site	16S rRNA &/or Cytb	43.9	157	Paternal leakage via hybridization	Radojičić et al. (2015)
			Rana japonica	Ø	Multi mt frag	0.0	78	N/A	Sumida (1997)
		Urodela	*Proteus anguinus* ^	Length	CR & tRNA Pro	2.4	84	Unknown	Gorički &Trontelj (2006)
			Triturus cristatus	Length	Multi mt frag	1.1	185	Unknown	Wallis (1987)
	Aves	Bucerotiformes	*Aceros waldeni* *^	Length	CR	100.0	1	Unknown	Sammler et al. (2011)
			Penelopides panini *^	Length	CR	100.0	1		
		Ciconiiformes	*Mycteria americana*	Length	CR	35.2	88	Mutation & maternal inheritance	Lopes et al. (2011)
		Charadriiformes	*Alca torda*	Site	CR	4.1	123	Unknown	Moum & Bakke (2001)
			Aethia pusilla	Length	tRNA Phe	100.0	1	Replication slippage	Berg et al. (1995)
			Aethia psittacula (as *Cyclorrhyncus psittacula*)	Length	tRNA Phe	100.0	1		
			Alca torda	Length	tRNA Phe	100.0	3		
			Alle alle	Length	tRNA Phe	100.0	1		
			Brachyramphus brevirostris	Length	tRNA Phe	100.0	1		
			Brachyramphus marmoratus ^	Length	tRNA Phe	100.0	1		
			Calidris maritima	Length	tRNA Phe	100.0	1		
			Cepphus grylle	Length	tRNA Phe	100.0	1		
			Fratercula arctica	Length	tRNA Phe	100.0	1		
			Larus argentatus	Length	tRNA Phe	100.0	1		
			Larus canus	Length	tRNA Phe	100.0	1		
			Larus fuscus	Length	tRNA Phe	100.0	12		
			Rissa tridactyla	Length	tRNA Phe	100.0	1		
			Synthliboramphus antiquus	Length	tRNA Phe	100.0	1		
			Uria aalge	Length	tRNA Phe	100.0	1		

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Aves	Charadriiformes	*Cepphus carbo*	Length	CR	100.0	2	Unknown	Kidd & Friesen (1998)
			Cepphus columba	Length	CR	100.0	4		
			Cepphus grylle	Length	CR	100.0	10		
		Columbiformes	*Columba livia* *	Length	D-loop	41.2	131	Maternal inhertiance	Lee et al. (2010)
		Galliformes	*Alectoris chukar* and *A. graeca* (hybrid zone)	Site	COI	4.5	44	Paternal leakage via hybridization	Gandolfi et al. (2017)
			Colinus virginianus	Site	Multi mt genes	25.5	51	Unknown	Halley et al. (2015)
		Passeriformes	*Alauda arvensis*	Length	CR (WGS)	100.0	1	Unknown	Qian et al. (2013)
			Calliope pectoralis	Ø	cytb, ND6 & CR	0.0	80	NUMT	Spiridonova et al. (2019)
			Castanozoster thoracicus (as *Poospiza thoracica*)	Site	Mitog wide (2 sites only)	100.0	1	Unknown	do Amaral et al. (2015)
			Microspingus cabanisi (as *Poospiza cabanisi*)	Site	Mitog wide (1 site only)	100.0	1		
			Microspingus lateralis (as *Poospiza lateralis*)	Site	Mitog wide (2 sites only)	100.0	1		
			Lanius excubitor	Ø	CR	0.0	8	N/A	Mundy & Helbig (2004)
			Lanius ludovicianus gambeli	Length	CR	10.7	28	Unknown	Mundy et al. (1996)
			Lanius ludovicianus mearnsi ^	Ø	CR	0.0	16		
			Lanius meridionalis koenigi ^	Length	CR	12.1	174	Unknown	Hernandez et al. (2010)
			Luscinia calliope	Ø	cytb	0.0	21	NUMT	Spiridonova et al. (2016)
			Passer italiae	Site	ND2	3.8	52	Paternal leakage through a hybridisation event	Päckert et al. (2019)
			Pycnonotus sinensis hainanus	Length	CR (WGS)	100.0	1	Unknown	Ren et al. (2016c)
			Rubigula melanicterus (as *Pycnonotus melanicterus*)	Length	CR (WGS)	100.0	1	Unknown	Ren et al. (2016b)
			Spizixos semitorques	Length	CR (WGS)	100.0	1	Unknown	Ren et al. (2016a)
			Sturnus vulgaris	Site	CR	15.5	181	Unknown	Rollins et al. (2016)
			Sylviparus modestus	Ø	cytb, ND2, ATP6, COI, & CR	0.0	15	N/A	Wang et al. (2015)
		Pelecaniformes	*Nipponia nippon* *^	Length	CR	100.0	61	Replication slippage	He et al. (2013)
		Pleuronectiformes	*Solea senegalensis*	Ø	CR (WGS)	0.0	388	N/A	Manchado et al. (2007)
		Procellariiformes	*Puffinus lherminieri*	L&S	CR & 12S rRNA	100.0	1	Unknown	Torres et al. (2019)
			Thalassarche cauta	Ø	CR	0.0	27	Maternally inhertied gene duplication	Abbott et al. (2005)
	Cephalaspidomorphi	Petromyzontiformes	*Lampetra aepyptera*	Ø	ND3 & ND6	0.0	47	N/A	White & Martin (2009)
			Lethenteron appendix (as *Lampetra appendix*)	Ø	ND3 & ND6	0.0	3		
			Ichthyomyzon unicuspis	Ø	ND3 & ND6	0.0	1		

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Cephalaspidomorphi	Petromyzontiformes	*Lethenteron camtschaticum* (as *Lethenteron japonicum*)	Ø	CR	0.0	10	N/A	Okada et al. (2010)
			Lethenteron kessleri	Ø	CR	0.0	5		
			Lethenteron sp. N	Ø	CR	0.0	17		
			Lethenteron sp. S	Ø	CR	0.0	7		
			Lethenteron spp.	Site	COI	29.4	109	De novo mutagenesis with some vertical inheritance	Artamonova et al. (2015)
	Chondrichthyes	Orectolobiformes	*Rhincodon typus* ^	Ø	CR	0.0	70	N/A	Castro et al. (2007)
	Mammalia	Artiodactyla	*Bos primigenius* †	Site	CR (WGS)	100.0	1	Unknown	Edwards et al. (2010)
			Bos taurus *	L&S	D-loop & rRNA coding regions	30.5	36	Unknown	Wu et al. (2000)
			Bos taurus *	Length	D-Loop	4.6	219	DNA-strand slippage during replication	Seroussi & Yakobson (2010)
			Camelus bactrianus *	Length	CR	100.0	135	Unknown	He et al. (2017)
			Lama glama ⁺	Site	CR	20.0	5	Paternal leakage or de novo mutation	Maté et al. (2007)
			Lama guanicoe	Site	CR	60.0	5		
			Vicugna pacos ⁺	Site	CR	40.0	5		
			Vicugna vicugna ^	Site	CR	60.0	5		
			Sus scrofa domestica *	Length	CR (WGS)	100.0	1	Unknown	Ursing & Arnason (1998)
		Carnivora	*Canis lupus familiaris* *	Length	CR (WGS)	100.0	1	Unidirectional replication slippage	Kim et al. (1998)
			Canis lupus familiaris *	Length	CR	5.0	20	Mutation	Wetton et al. (2003)
			Canis lupus familiaris *	Site	CR [HV1]	33.3	6	Age related mutation	Spicer et al. (2014)
			Canis lupus familiaris⁺	Length	D-Loop	100.0	14	Replication slippage	Savolainen et al. (2000)
			Canis lupus ⁺	Length	D-Loop	100.0	5		
			Felis catus *	Length	CR	1.7	174	Unknown	Tarditi et al. (2011)
			Felis catus ⁺	Ø	Mitog wide	0.0	1	NUMT (CR)	Lopez et al. (1996)
			Lutra lutra *	Length	CR	68.8	32	Unknown	Ketmaier & Bernardini (2005)
			Lynx lynx	Length	CR	12.5	8	Replication slippage	Sindičić et al. (2012)
			Mirounga angustirostris	Length	CR	83.3	24	Replication slippage	Hoelzel et al. (1993)
			Mirounga leonina	Length	CR	100.0	24		
			Panthera tigris ^	Length	CR (WGS)	60.0	5	Unknown	Kitpipit & Linacre (2012)
			Ursus arctos	Length	CR	7.1	56	Replication slippage	Matsuhashi et al. (1999)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Mammalia	Cetacea	*Balaenoptera edeni*	Site	CR	7.1	14	Mutation and/or replication slippage	Vollmer et al. (2011)
			Delphinus delphis	S&L	CR	3.8	52		
			Globicephala macrorhynchus	Site	CR	1.2	335		
			Globicephala melas	Site	CR	0.5	173		
			Kogia breviceps	S&L	CR	3.5	204		
			Lagenorhynchus albirostris	Length	CR	50.0	2		
			Pseudorca crassidens	Site	CR	50.0	2		
			Stenella frontalis	S&L	CR	58.9	183		
			Tursiops truncatus	S&L	CR	2.7	4040		
			Eubalaena glacialis ^	Site	CR	42.9	14	Unknown	McLeod & White (2010)
			Mesoplodon grayi	Ø	Mitog wide	0.0	1	N/A	Thompson et al. (2016)
		Chiroptera	*Eptesicus fuscus*	Length	D-Loop	15.0	20	Replication rate bias for smaller repeat numbers	Wilkinson et al. (1997)
			Myotis bechsteini	Length	D-Loop	18.0	245		
			Myotis lucifugus	Length	D-Loop	31.6	19		
			Myotis myotis	Length	D-Loop	47.1	191		
			Nyctalus noctula	Length	D-Loop	42.0	112		
			Nycticeinops schlieffeni	Length	D-Loop	62.5	8		
			Nycticeius humeralis	Length	D-Loop	28.2	195		
			Pipistrellus pipistrellus	Length	D-Loop	62.5	8		
			Myotis bechsteinii	Length	CR	47.9	635	Maternal inheritance and high mutation rates of tandem repeats	Mayer & Kerth (2005)
			Myotis myotis	Site	Mitog wide (<6 sites)	100.0	195	Unknown	Jebb et al. (2018)
			Nycticeius humeralis	Length	D-Loop	28.2	195	Unidirectional replication slippage	Wilkinson & Chapman (1991)
			Rhinolophus sinicus complex	S&L	CR	66.7	18	Introgression via hybridization	Mao et al. (2014)
		Eulipotyphla	*Crocidura russula*	Length	CR	100.0	8	Replication slippage	Fumagalli et al. (1996)
			Sorex aruneus	Length	CR	12.5	8		
			Sorex cinereus	Length	D-Loop	6.7	15	Replication slippage	Stewart & Baker (1994)
			Talpa aquitania	Length	CR	100.0	1	Unknown	Aleix-Mata et al. (2020)
		Insectivora	*Suncus murinus* *	Length	CR	28.6	7	Unknown	Yamagata & Namikawa (1999)
		Lagomorpha	*Oryctolagus cuniculus* ^	Length	CR	100.0	1	Unknown	Gissi et al. (1998)
		Perissodactyla	*Ceratotherium simum*	Length	CR (WGS)	100.0	1	Unknown	Xu & Arnason (1997)
			Equus asinus *	Length	CR	100.0	1	Unknown	Xu et al. (1996)
			Equus ferus caballus *	S&L	Cytb	5.3	430	Unknown	Zhao et al. (2015)
		Primates	*Leontopithecus chrysopygus* *^	Ø	Mitog wide	0.0	1	N/A	de Freitas et al. (2018)
			Macaca fuscata	Length	MNR	41.0	100	L-strand displacement by H-strand	Hayasaka et al. (1991)
			Tarsius dentatus	Length	D-Loop	100.0	10	Unknown	Merker et al. (2014)
			Tarsius lariang	Length	D-Loop	100.0	10		
			Tarsius wallacei	Length	D-Loop	100.0	10		

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Mammalia	Rodentia	*Abrothrix jelskii* (as *Chroeomys jelskii*)	Ø	Multi mt frag	0.0	4	Nuclear paralog	Smith et al. (1992)
			Myodes glareolus	Site	Mitog wide (<26 sites)	63.2	115	Intra-individual mutation	Baker et al. (2017), Kesäniemi et al. (2018)
	Reptilia	Crocodilia	*Crocodylus moreletii*	Length	D-loop	10.3	29	Unknown	Ray & Densmore (2003)
			Tomistoma schlegelii *	Length	CR	3.4	29	Stable transfer of repetitive mt region	Kaur & Ong (2011)
		Squamata	*Chlamydosaurus kingii*	Ø	ND2 & ND4	0.0	47	N/A	Ujvari et al. (2007)
			Cnemidophorus tesselatus	Length	Multi mt genes	24.7	73	Unknown	Densmore et al. (1985)
			Eurolophosaurus divaricatus (as *Tropidurus divaricatus*)	Length	Multi mt frag	94.4	18	Unknown	Passoni et al. (2000)
			Eurolophosaurus amathites (as *Tropidurus amathites*)	Ø	Multi mt frag	0.0	16		
			Eurolophosaurus nanuzae (as *Tropidurus nanuzae*)	Ø	Multi mt frag	0.0	9		
			Gallotia galloti	Ø	Multi mt frag	0.0	10	N/A	Thorpe et al. (1993)
			Heteronotia binoei	Length	Multi mt frag	4.6	87	Unknown. Note: parthenogenetic	Moritz (1991)
			Heteronotia binoei	Length	16S rRNA	35.1	77	Unknown. Note: parthenogenetic	Zevering et al. (1991)
			Abronia graminea	Ø	Mitog wide	0.0	1	Replication slippage	Kumazawa (2004)
			Leptotyphlops dulcis	Length	12S rRNA (WGS)	100.0	1		
			Sceloporus occidentalis	Length	Noncoding region (WGS)	100.0	1		
			Shinisaurus crocodilurus ^	Length	12S rRNA (WGS)	100.0	1		
			Smaug warreni (as *Cordylus warreni*)	Length	16S rRNA (WGS)	100.0	1		
			Timon lepidus (as *Lacerta lepida*)	Site	cytb	9.5	21	Paternal leakage (NUMTs also present)	Miraldo et al. (2012)
			Urosaurus ornatus	Length	Multi mt frag	24.3	37	Unknown	Haenel (1997)
			Ahaetulla nasuta +	Ø	COI	0.0	1	N/A	Dubey et al. (2011)
			Bungarus fasciatus +	Ø	COI	0.0	1		
			Daboia russelii +	Ø	COI	0.0	1		
			Eryx johnii +	Ø	COI	0.0	1		
			Indotyphlops braminus + (as *Ramphotyphlops braminus*)	Ø	COI	0.0	1		
			Naja kaouthia +	Ø	COI	0.0	2		
			Naja naja +	Ø	COI	0.0	1		
			Ptyas mucosa + (as *P. mucosus*)	Ø	COI	0.0	1		
			Python molurus +	Ø	COI	0.0	1		
			Xenochrophis piscator +	Ø	COI	0.0	1		
			Xenochrophis schnurrenbergeri +	Ø	COI	0.0	1		

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Chordata	Reptilia	Squamata	*Boa constrictor* +	Ø	tRNA clusters - IQM, WANCY, & cytb-12S rRNA	0.0	1	Duplication of gene region	Kumazawa et al. (1996)
			Crotalus viridis +	Ø	tRNA clusters - IQM, WANCY, & cytb-12S rRNA	0.0	1		
			Lycodon semicarinatus + (as *Dinodon semicarinatus*)	Ø	tRNA clusters - IQM, WANCY, & cytb-12S rRNA	0.0	1		
			Ovophis okinavensis +	Ø	tRNA clusters - IQM, WANCY, & cytb-12S rRNA	0.0	3		
			Pituophis melanoleucus +	Ø	tRNA clusters - IQM, WANCY, & cytb-12S rRNA	0.0	1		
			Python regius +	Ø	tRNA clusters - IQM, WANCY, & cytb-12S rRNA	0.0	1		
		Testudines	*Chelonia mydas* *^	Length	CR	100.0	82	Maintanance of high genetic diversity in low population sizes	Tikochinski et al. (2020)
			Caretta caretta ^	Length	CR	20.6	34	Unknown	Drosopoulou et al. (2012)
			Caretta caretta ^	S&L	Mitog wide	100.0	3	Unknown	Delgado-Cano et al. (2020)
			Eretmochelys imbricata ^	S&L	Mitog wide	100.0	1		
			Chelonia mydas *^	Length	CR	12.1	289	Unknown	Tikochinski et al. (2012)
			Malaclemys terrapin	Length	Multi mt frag	32.1	53	Unknown	Lamb & Avise (1992)
Cnidaria	Anthozoa	Scleractinia	*Desmophyllum pertusum* (as *Lophelia pertusa*)	Site	Mitog wide (5 sites only)	100.0	1	Unknown	Emblem et al. (2012)
Echinodermata	Ophiuroidea	Euryalida	*Astrobrachion constrictum*	Length	COI	1.8	117	Paternal leakage	Steel et al. (2000)
Mollusca	Bivalvia	Adapedonta	*Solen marginatus*	Length	COI & 16S rRNA	83.3	6 ♂	Doubly Uniparental inhertiance	Lucentini et al. (2020)
				Ø	COI & 16S rRNA	0.0	3 ♀		
		Arcida	*Anadara broughtonii* (as *Scapharca broughtonii*)	Length	Heterop. tandem repeat unit (WGS)	100.0	12	mtDNA recombination	Liu et al. (2013)
		Cardiida	*Cerastoderma glaucum*	Ø	COI & 16S rRNA	0.0	7 ♂	N/A	Lucentini et al. (2020)
				Ø	COI & 16S rRNA	0.0	5 ♀		
		Mytilida	*Arcuatula senhousia* (as *Musculista senhousia*)	Site	Cytb	100.0	8 ♂	Doubly Uniparental inhertiance	Passamonti (2007)
				Ø	Cytb	0.0	4 ♀		
			Bathymodiolus thermophilus	Length	Multi mt frag	1.7	58	Paternal transmission	Craddock et al. (1995)
			Mytilus californianus	Length	COI & ND5	100.0	5 ♂	Doubly Uniparental inhertiance	Beagley et al. (1997)
				Ø	COI & ND5	0.0	5 ♀		
			Mytilus californianus	Ø	16S, cytb, & D-loop	0.0	6	(DUI species - somatic homoplasmy)	Rawson (2005)
			Mytilus edulis	Ø	16S, cytb, & D-loop	0.0	18	(DUI species - somatic homoplasmy)	
			Mytilus trossulus	Length	16S, cytb, & D-loop	58.2	98	(DUI species) Somatic heteroplasmy cause unknown	

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Mollusca	Bivalvia	Mytilida	*Mytilus galloprovincialis*	Length	ND2 & COIII	100.0	26 ♂	Doubly uniparental inheritance	Quesada et al. (1996)
				Length	ND2 & COIII	15.4	26 ♀	Paternal leakage	
			Mytilus galloprovincialis	Length	16S, COIII-ND2, COI	100.0	15 ♂	Doubly uniparental inheritance	Obata et al. (2006)
				Length	16S, COIII-ND2, COI	100.0	15 ♀	No elimination of male mtDNA	
			Mytilus galloprovincialis	Length	CR	64.7	17 ♂	Doubly Uniparental inhertiance with sperm derived mtDNA leakage	Kyriakou et al. (2010)
				Length	CR	10.5	19 ♀	Unknown	
			Mytilus edulis	Length	COIII-ND2 & multi mt frag	100.0	28 ♂	Doubly uniparental inheritance	Skibinski et al. (1994)
				Ø	COIII-ND2 & multi mt frag	0.0	24 ♀		
			Mytilus edulis	Length	COI and COIII	100.0	7 ♂	Doubly uniparental inheritance	Batista et al. (2011)
				Length	COI and COIII	25.0	12 ♀	Low levels of paternal mtDNA through leakage	
			Mytilus edulis	Length	COIII & multi mt frag	100.0	31 ♂	Doubly uniparental inheritance	Stewart et al. (1995)
				Length	COIII & multi mt frag	2.1	47 ♀	Unknown	
			Mytilus trossulus	Length	COIII & multi mt frag	76.2	42 ♂	Doubly uniparental inheritance	
				Ø	COIII & multi mt frag	0.0	30 ♀		
			Mytilus edulis	Length	COIII & multi mt frag	100.0	34 ♂	Doubly uniparental inheritance	Saavedra et al. (1996)
				Length	COIII & multi mt frag	2.1	47 ♀	Unknown	
			Mytilus trossulus	Length	COIII & multi mt frag	100.0	32 ♂	Doubly uniparental inheritance	
				Ø	COIII & multi mt frag	0.0	30 ♀		
			Mytilus edulis	S&L	Multi mt frag & lrRNA	100.0	6 ♂	Doubly uniparental inheritance	Quesada et al. (2003)
				Ø	Multi mt frag & lrRNA	0.0	8 ♀		
			Mytilus trossulus	S&L	Multi mt frag & lrRNA	63.2	19 ♂	Doubly uniparental inheritance w/ paternal co-transmission (transmitting multiple mitogenomes)	
				S&L	Multi mt frag & lrRNA	5.3	19 ♀	Unknown	
			Mytilus trossulus	Length	Multi mt frag	32.7	52 ♂	Doubly uniparental inheritance	Wenne & Skibinski (1995)
				Length	Multi mt frag	3.1	32 ♀	Unknown	
			Mytilus trossulus	Length	ND2-COIII	46.9	241 ♂	Doubly uniparental inheritance	Zbawicka et al. (2003)
				Length	ND2-COIII	6.9	102 ♀	Feminization of male mitogenome	
			Mytilus trossulus	Length	CR	100.0	68 ♂	Doubly uniparental inheritance	Burzynski et al. (2006)
				Ø	CR	0.0	52 ♀		
			Mytilus galloprovincialis	Length	ND2-COIII	100.0	38 ♂	Doubly uniparental inheritance	Quesada et al. (1998)
				Ø	ND2-COIII	0.0	39 ♀		
			Mytilus galloprovincialis	Length	COIII & 16S rRNA	100.0	137 ♂	Doubly uniparental inheritance	Ladoukakis et al. (2002)
				Length	COIII & 16S rRNA	1.6	124 ♀	Paternal leakage	

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Mollusca	Bivalvia	Mytilida	*Mytilus* spp. (mutliple species, hybrid zones)	Length	Multi mt frag	33.3	108 ♂	*Pre-DUI discovery*	Fisher & Skibinski (1990)
				Length	Multi mt frag	4.8	104 ♀	*Pre-DUI discovery*	
			Mytilus spp. (multiple species, unidentified)	Length	Multi mt frag	56.7	150	Biparental inheritance, enabled by hybrid zones	Hoeh et al. (1991)
			Mytilus spp. (mutliple species, hybrid zones)	Site	16S rRNA	89.0	821	DUI with some breakdown in strict inheritance patterns	Brannock et al. (2013)
		Nuculanida	*Ledella ultima*	Length	16S rRNA & cytb	55.8	52	Possible DUI	Boyle & Etter (2013)
		Pectinida	*Aequipecten opercularis*	Length	Multi mt frag	9.7	31	Weak selection for small molecule sizes	Gjetvaj et al. (1992)
			Argopecten irradians	Ø	Multi mt frag	0.0	24		
			Crassadoma gigantea	Length	Multi mt frag	16.7	12		
			Chlamys hastata	Length	Multi mt frag	9.5	21		
			Chlamys islandica	Length	Multi mt frag	15.1	119		
			Pecten maximus	Ø	Multi mt frag	0.0	26		
			Pecten maximus	Length	Multi mt frag	3.7	27	Unknown	Rigaa et al. (1997)
			Placopecten magellanicus	Length	Multi mt frag	6.4	280	Unknown	Fuller & Zouros (1993)
		Unionida	*Pyganodon grandis* (as *Anodonta grandis grandis*)	Length	Multi mt frag	100.0	6 ♂	Doubly uniparental inheritance	Liu & Mitton (1996)
				Ø	Multi mt frag	0.0	2 ♀		
			Unio crassus ^	S&L	COI	95.0	20 ♂	Doubly uniparental inheritance	Mioduchowska et al. (2016)
				Ø	COI	0.0	21 ♀		
			Utterbackia peninsularis	Length	COI	100.0	8 ♂	Doubly uniparental inheritance	Breton et al. (2017)
				Length	COI	100.0	5 ♀	Failure in exclusion mechansisms of male mtDNA	
			Venustaconcha ellipsiformis	Length	COI	100.0	6 ♂	Doubly uniparental inheritance	
				Length	COI	66.7	6 ♀	Failure in exclusion mechansisms of male mtDNA	
		Venerida	*Meretrix lusoria*	Length	CR (WGS)	3.5	8	Unknown	Wang et al. (2010)
			Polititapes rhomboides	Site	COI	100.0	114 ♂	Doubly uniparental inheritance	Chacón et al. (2020)
				Ø	COI	0.0	60 ♀		
			Ruditapes decussatus	Length	"Largest unassigned region" adj. CR (WGS)	15.4	13 ♀	Replication slippage	Ghiselli et al. (2017)
			Spisula sachalinensis (as *Pseudocardium sachalinense*)	Length	COI, rrnL, & rrnS	100.0	2 ♂	Doubly uniparental inheritance	Plazzi (2015)
				Ø	COI, rrnL, & rrnS	0.0	2 ♀		
			Venerupis philippinarum (as *Tapes philippinarum*)	L&S	mt-l-rRNA	100.0	10 ♂	Doubly uniparental inheritance	Passamonti & Scali (2001)
				Ø	mt-l-rRNA	0.0	5 ♀		
	Cephalopoda	Oegopsida	*Watasenia scintillans*	Site	Mitog wide (6 sites only)	100.0	1	Unknown	Hayashi et al. (2016)
	Gastropoda	Architaenioglossa	*Pomacea diffusa*	Ø	16S rRNA & COI	0.0	4 ♂	N/A	Parakatselaki et al. (2016)
				Ø	16S rRNA & COI	0.0	3 ♀		
			Viviparus ater	Ø	16S rRNA	0.0	4	N/A	Gusman et al. (2017)

Phyla	Class	Order	Taxa	het.	mt fragment	% het.	n	Proposed source	Reference
Mollusca	Gastropoda	Littorinimorpha	*Euspira heros* (as *Lunatia heros*)	Ø	COI & 16S rRNA	0.0	5	N/A	Gusman et al. (2017)
			Littorina littorea	Ø	COI & 16S rRNA	0.0	8		
		Neogastropoda	*Nucella lapillus*	Ø	COI & 16S rRNA	0.0	7		
		Patellogastropoda	*Testudinalia testudinalis* (as *Tectura testudinalis*)	Ø	COI & 16S rRNA	0.0	8		
		Stylommatophora	*Cepaea nemoralis*	Ø	16S rRNA	0.0	459	N/A	Davison (2000)
Nematoda	Secernentea	Tylenchida	***Meloidogyne arenaria*** *	Length	Multi mt frag	70.5	61	Intra-individual mutation Note: parthenogenetic	Dautova et al. (2002)
			Meloidogyne incognita *	Length	Multi mt frag	79.1	134		
			Meloidogyne javanica *	Length	Multi mt frag	60.0	60		
			Meloidogyne arenaria *	Ø	COI, COII, SSU & LSU rRNA	0.0	16	N/A	Kiewnick et al. (2014)
			Meloidogyne chitwoodi *	Ø	COI, COII, SSU & LSU rRNA	0.0	39		
			Meloidogyne enterolobii *	Ø	COI, COII, SSU & LSU rRNA	0.0	19		
			Meloidogyne ethiopica *	Ø	COI, COII, SSU & LSU rRNA	0.0	4		
			Meloidogyne fallax *	Ø	COI, COII, SSU & LSU rRNA	0.0	23		
			Meloidogyne hapla *	Ø	COI, COII, SSU & LSU rRNA	0.0	25		
			Meloidogyne incognita *	Ø	COI, COII, SSU & LSU rRNA	0.0	25		
			Meloidogyne javanica *	Ø	COI, COII, SSU & LSU rRNA	0.0	16		
			Meloidogyne maritima *	Ø	COI, COII, SSU & LSU rRNA	0.0	5		
			Meloidogyne minor *	Ø	COI, COII, SSU & LSU rRNA	0.0	11		
			Meloidogyne naasi *	Ø	COI, COII, SSU & LSU rRNA	0.0	6		
Nemertea	Palaeonemertea	Carinomiformes	*Carinina ochracea*	Site	COI	100.0	1	Unknown	Fernandez-Alvarez et al. (2015)
Platyhelminthes	Cestoda	Cyclophyllidea	*Echinococcus granulosus*	Site	COI (& ND1)	100.0	4	Unknown	Bowles et al. (1994)
	Rhabditophora	Tricladida	*Dugesia japonica*	Site	COI	5.6	18	High substiution rate of mtDNA Note: asexual	Bessho et al. (1992)
Porifera	Hexactinellida	Amphidiscosida	f. Hyalonematidae sp.	Ø	Mitog wide	0.0	1	N/A	Haen et al. (2014)
			Tabachnickia sp.	Ø	Mitog (coding regions)	0.0	1		
		Hexactinosida	*Aphrocallistes beatrix*	Ø	Mitog wide	0.0	1		
			Aphrocallistes vastus	Ø	Mitog wide	0.0	1		
			f. Euretidae n. gen. n. sp.	Ø	Mitog wide	0.0	1		
			Farrea sp.	Ø	COI	0.0	1		
			Iphiteon panicea	Ø	Mitog wide	0.0	1		
			Psilocalyx wilsoni	Ø	COI	0.0	1		
		Lyssacinosida	*Hertwigia falcifera*	Ø	Mitog wide	0.0	1		
			Oopsacas minuta	Ø	Mitog wide	0.0	1		
			f. *Regadrella* sp.	Ø	Mitog (coding regions)	0.0	1		
			f. Rossellidae sp.	Ø	Mitog wide	0.0	1		
			Sympagella nux	Ø	Mitog wide	0.0	1		
			Vazella pourtalesii	Ø	Mitog wide	0.0	1		

1.9 Figures

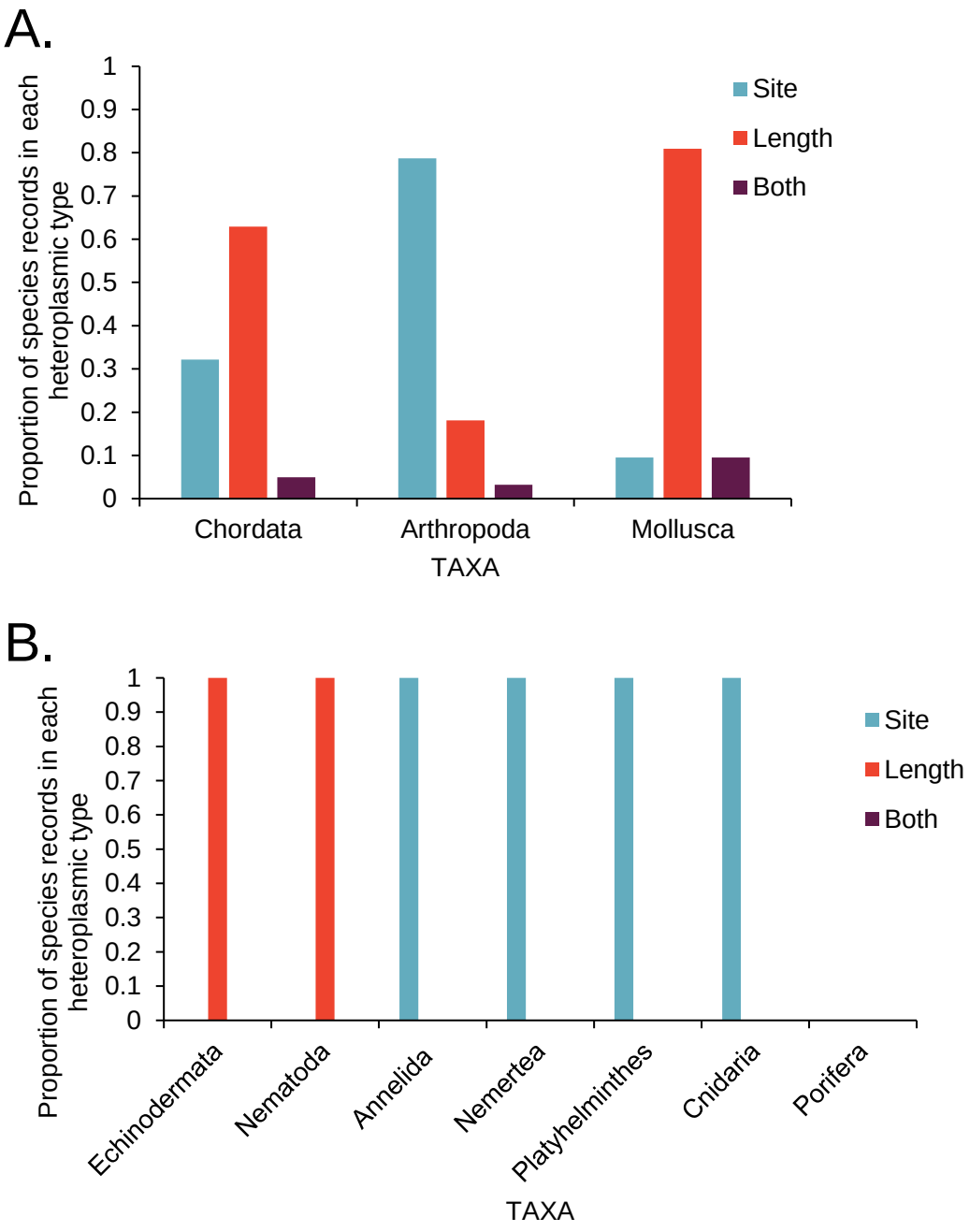


Figure 1: Summary of the proportion of species records in each heteroplasmic mtDNA type (site, length, or both site and length heteroplasmy within one species/individual) recovered in the literature search. Records were dominated by A. Chordata with 244 heteroplasmic taxon records, Arthropoda with 94, and Mollusca with 44. Very few records were recovered from the remaining phyla; B. with only one heteroplasmic taxon recorded each from Echinodermata, Annelida, Nemertea, and Cnidaria. Two and three taxa records were reported from Platyhelminthes and Nematoda, respectively. Only homoplasmic (heteroplasmy absent) taxa were reported in Porifera. The totals including records recovered that were homoplasmic are reported in Supp. Table 1.

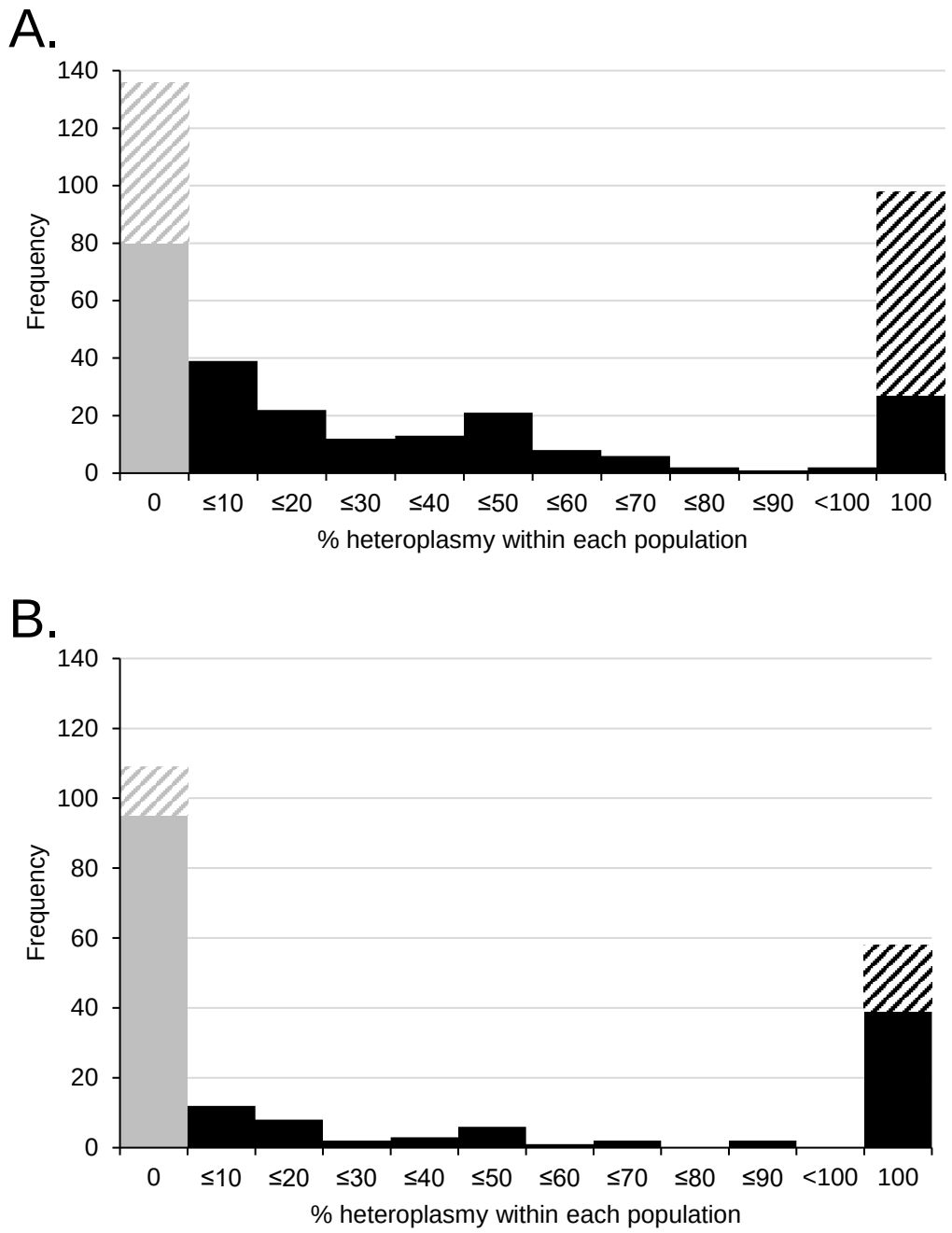


Figure 2: The frequency of taxon records collected from the literature in A. Chordata (with 360 records in total) and B. Arthropoda (203 records) assessed for the percentage of reported heteroplasmic individuals within each assessed "population" ranging from zero (a homoplasmic population – grey) to 100% of individuals identified as heteroplasmic. Crosshatch pattern indicates singleton records – where only one individual was assessed in any report (this individual would be found to be either homoplasmic or heteroplasmic). Homoplasmic records (grey) are underrepresented as our search parameters were not intended to assess homoplasmy. Note that here a "population" refers to all specimens of a species assessed within any publication and does not necessarily truly represent an ecological population.

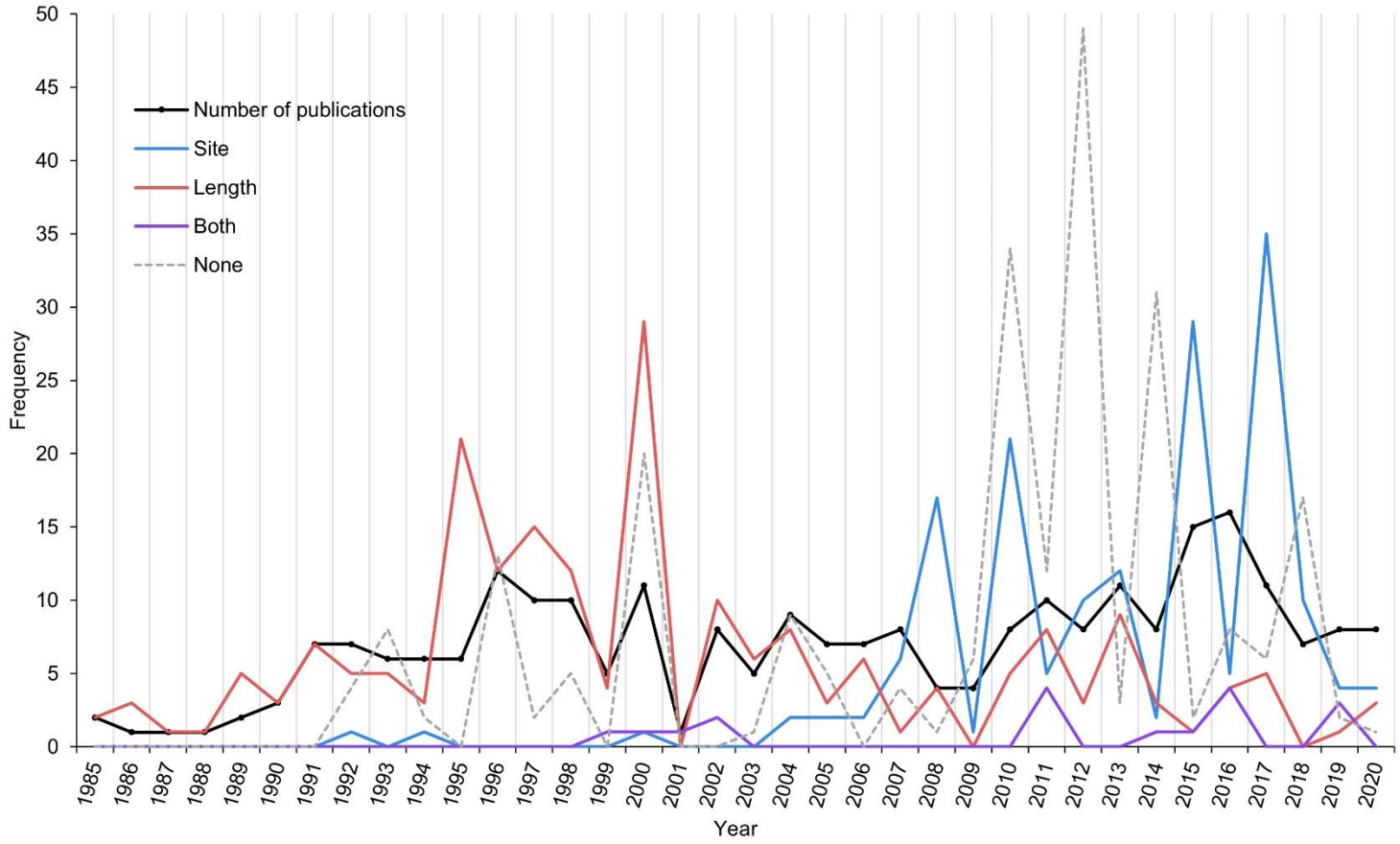


Figure 3: Frequency of each type of heteroplasmic record from all animal taxa (species) recovered in the literature search for each year they were published. This shows the progress in the number of records taxa with of site, length, and both (site and length heteroplasmic within a single species/individual) types of heteroplasmy being reported in the literature as recovered in our search, as well as the homoplasmic results ("none"). The number of publications that for each year — those that met our search parameters — is also shown (indicating that some publications include reports of mitochondrial heteroplasmy in multiple species).

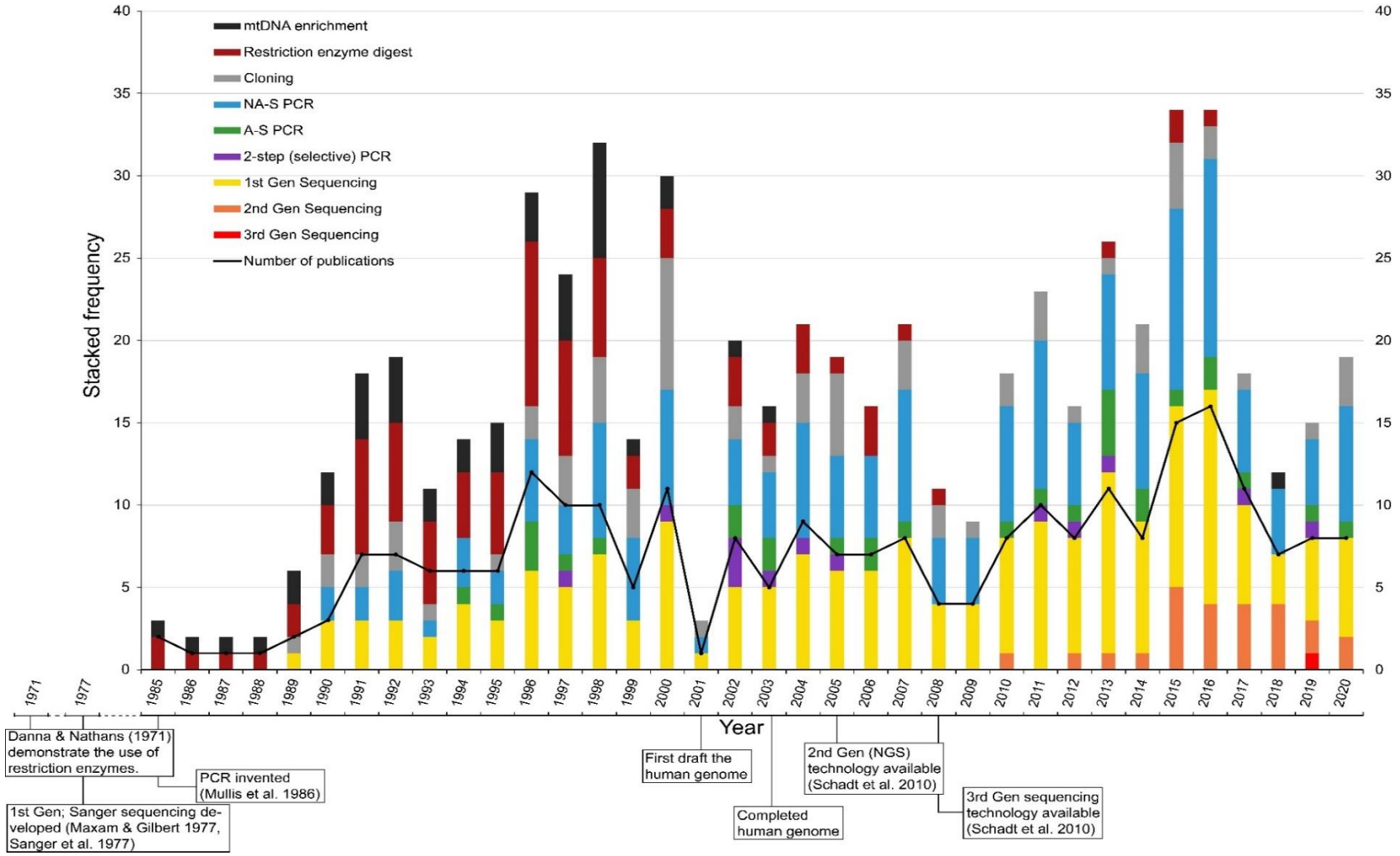


Figure 4: Stacked frequency of the number of times (publications) a particular molecular technique was applied to the records of mitochondrial heteroplasmy recovered in our search for each year. This demonstrates the transition of molecular techniques through time as DNA sequencing knowledge develops — a timeline of major genetic and genomic technological innovation is included. Multiple techniques could be used in any publication (as observed as the number of publications published in any year is generally lower than the techniques). These techniques have been generalized and how specific techniques were classified can be found in Supp. Table 3.

1.10 Supplementary Material

Supp. Table 1: Number of taxa in each type of mitochondrial heteroplasmy (or homoplasmy) recovered across ten phyla in the systematic review of literature on heteroplasmy in animals, including the total number of taxa determined to have heteroplasmy (Het) and the number of taxa recovered in the search that were homoplasmic (None). Multiple taxa could be sourced from a single publication.

Taxa	Site	Length	Both	Total (Het)	None	Total (all records)
Annelida	1	0	0	1	5	6
Arthropoda	74	17	3	94	109	203
Chordata	72	141	11	224	136	360
Cnidaria	1	0	0	1	0	1
Echinodermata	0	1	0	1	0	1
Mollusca	4	34	4	44	12	54
Nemertea	1	0	0	1	0	1
Nematoda	0	3	0	3	11	14
Platyhelminthes	2	0	0	2	0	2
Porifera	0	0	0	0	14	14

Supp. Table 2: P-values from Fisher's *post hoc* tests using a Bonferroni correction, performed using R package RVAideMemoire for three types of heteroplasmy: site, length, or both, in the taxon groups with the highest number of records, Chordata, Arthropoda, and Mollusca. Significant values (at significance level 0.05) are bolded, with an asterisk.

Comparisons	Site : Length	Site : Both	Length : Both
Chordata : Arthropoda	**1.01×10^{-13}***	0.44	1
Chordata : Mollusca	**2.04×10^{-2}***	0.21	1
Arthropoda : Mollusca	**5.54×10^{14}***	**0.01***	1

Supp. Table 3: Classification of molecular and laboratory techniques applied to investigate mitochondrial heteroplasmy in animal taxa, collated from the literature assessed in this review. The changes in laboratory techniques applied to investigations of mitochondrial heteroplasmy through time (from 1985 to 2020) are summarized in Figure 4.

Technique classification (Fig. 4)	**Full description**	**Specific techniques included under the classification**
mtDNA enrichment	Mitochondrial DNA enrichment	Multiple techniques isolating mtDNA from a DNA extraction
Restriction enzyme digest	Restriction enzyme digest	• Multiple techniques of restriction enzyme digest • RLFP analysis
Cloning	DNA cloning	Multiple techniques using plasmids to make copies of DNA fragments
NA-S PCR	Non-allele specific PCR	• General PCR (non-allele specific/untargeted primers) • Long range PCR without targeted primers • Long and accurate PCR without targeted primers
A-S PCR	Allele specific PCR	• Qualitative PCR (using allele specific/targeted primers/probes) • General PCR with specific primers developed for each allele • ARMS (amplification refractory mutation system)
2-step (selective) PCR	PCR procedures with two steps to select for specific products	• Single molecule PCR • Nested PCR • ALFP (amplified fragment length polymorphism) PCR
1st Gen Sequencing	Sequencing methods classified as “first generation”	• Sanger sequencing including: • BigDye Terminator sequencing • ABI Prism sequencing • Capillary sequencing, etc
2nd Gen Sequencing	Sequencing methods classified as “second generation”	• Illumina Miseq, Hiseq, & Novaseq • Massively parallel sequencing • ABI SOLiD sequencing NOTE: for shotgun, multiplexed, or UCE sequencing, the sequencing platform was recorded
3rd Gen Sequencing	Sequencing methods classified as “third generation”	Nanopore

Chapter 2: Extensive and widespread mitochondrial heteroplasmy in a colletid bee, *Amphylaeus morosus*

Olivia K. Davies[a*], Michael G. Gardner[a,b], Mark I. Stevens[b,c], Tessa M. Bradford[b] and Michael P. Schwarz[a]

[a]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

[b]South Australian Museum, GPO Box 234, Adelaide SA 5001, Australia

[c]School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia

*Corresponding author. Email: davies.oliviak@gmail.com

Keywords: Hylaeinae, site polymorphism, DNA barcoding, mitochondrial heterogeneity, Hymenoptera

2.1 Abstract

Mitochondrial heteroplasmy is the occurrence of more than one type of mitochondrial DNA within a single individual. Although generally reported to occur in a small subset of individuals within a species, there are some instances of widespread heteroplasmy across entire populations. *Amphylaeus morosus* (Smith, 1879) is an Australian native bee species in the diverse and cosmopolitan bee family Colletidae. This species has an extensive geographical range along the eastern Australian coast, from southern Queensland to western Victoria, covering approximately 2,000 km. Seventy-three individuals were collected from five localities across this geographical range and sequenced using Sanger sequencing for the mitochondrial cytochrome *c* oxidase subunit I (COI) gene, initially to detect population variation for historical demography studies. These data indicate that every individual had the same consistent heteroplasmic sites but no other nucleotide variation, suggesting two conserved and widespread heteroplasmic mitogenomes. Ion Torrent shotgun sequencing revealed that heteroplasmy occurred across multiple mitochondrial protein-coding genes and is unlikely explained by transposition of mitochondrial genes into the nuclear genome (NUMTs). Our data are consistent with the presence of two mitogenomes within all individuals examined in this species and suggest a major divergence from standard patterns of mitochondrial inheritance.

2.2 Introduction

Mitochondrial heteroplasmy is the presence of more than one type of mitochondrial DNA (mtDNA) within a single cell or individual. It is sporadically reported in the literature, and early reports were usually attributed to a spontaneous accumulation of tandem repeat mutations in highly variable gene regions (Avise et al. 1987, Boyce et al. 1989). Further research has also attributed heteroplasmy to biparental inheritance (Gyllensten et al. 1991, Kvist et al. 2003, Boyle and Etter 2013), replication errors (Hoelzel et al. 1993, Irwin et al. 2009), mutagenetic processes (Muntaabski et al. 2020), or recombination (Zsurka et al. 2005). It has generally been interpreted as an intermittent, transient condition in natural populations (Hoarau et al. 2002, Song et al. 2008), and which might be facilitated by hybridization (e.g. Gandolfi et al. (2017), Mastrantonio et al. (2019)). Occasionally, mitochondrial heteroplasmy is reported to be widespread in some taxa (e.g. Doublet et al. (2012), Xiong et al. (2013), Pietan et al. (2016)) and this poses problems for standard mitochondrial inheritance models. Increasingly, examples of non-standard mitochondrial patterns in natural systems are being reported (e.g. Bandelt et al. (2005), Hoolahan et al. (2012), Sańko and Burzyński (2014)). This has important implications for the application of mtDNA in analyses such as species delineations, population genetics, and phylogenetics (e.g. Hassanin (2006), Leaché and McGuire (2006), Hlaing et al. (2009), Pedraza-Marrón et al. (2019)). However, our understanding of how these 'abnormal' mtDNA systems have arisen and the biological roles of mitochondrial heteroplasmy and other mitochondrial deviancy is currently in its infancy.

Mitochondrial heteroplasmy can be broadly grouped into two types; length heteroplasmy and site heteroplasmy (Barr et al. 2005). The former can be observed as small tandem repeated units, duplications, or deletions and has been more readily detected than site heteroplasmy (Lunt et al. 1998, Barr et al. 2005, Parson 2013). Length heteroplasmy is also more frequently reported in vertebrates than invertebrates (Chapter 1) but these biases could be due to taxon selection, targeted methodology with expectant outcomes, and the limitations of historic genetic techniques. Site heteroplasmy is most commonly observed as substitutions at single nucleotides, usually occurring at the third codon position (Nunes et al. 2013, Robison et al. 2015). More often reported in invertebrates (Chapter 1), site heteroplasmy can be maintained within a single heteroplasmic codon (e.g. Doublet et al. (2008), Doublet et al. (2012)) or occur across the mitogenome (e.g. Xiong et al. (2013), Jebb et al. (2018), Meza-Lázaro et al.

(2018)). Although high rates of heteroplasmy within populations are reported across numerous animal taxa, populations where all individuals appear heteroplasmic are mostly reported from arthropods (Chapter 1).

Site mitochondrial heteroplasmy is being increasingly reported, probably due more accessible and reliable second (next generation sequencing – NGS) and third generation sequencing technologies. In humans, it has been demonstrated that somatic heteroplasmic mutations become more prevalent with age, often without expressing any clinical symptoms of mitochondrial disease (Ye et al. 2014, Kang et al. 2016). It is likely that other organisms also develop similarly and have varying rates of susceptibility to mitochondrial mutations during their development, through exposure to oxidative damage (Santos et al. 2013, Shokolenko et al. 2014), insufficient repair mechanisms (Kmiec et al. 2006), or errors caused by mitochondrial polymerases during mtDNA replication (Trifunovic et al. 2004, Kennedy et al. 2013). However, these mutations will not necessarily be heritable (Ju et al. 2014, Pinto and Moraes 2015). Extensive heteroplasmy — where mitochondrial heteroplasmy is maintained in most individuals throughout a population, and the resultant heteroplasmic lineages potentially maintained through inheritance — is difficult to explain, particularly when much of the variation exists as synonymous mutations as often observed in invertebrates (Robison et al. 2015).

Here we explore distribution-wide mitochondrial heteroplasmy in *Amphylaeus morosus* (Smith, 1879) from the cosmopolitan bee family Colletidae (Hylaeinae). *Amphylaeus morosus* is widely distributed along the eastern coast of Australia, from subcoastal subtropical heathlands in southern Queensland to the north, through to wet-montane and subcoastal forests in Victoria in the south (Houston 1975; Figure 1). This univoltine species is unusual, as it is the only known colletid species to nest socially, with colony sizes consisting of a single female, up to six females together nesting in a lateral cavity (Spessa et al. 2000). Barcoding of the mitochondrial cytochrome *c* oxidase subunit I (mt-COI) gene was initially intended to detect population variation for historical demography studies. However, a lack of mitochondrial variation and consistent double peaks throughout the generated chromatograms indicated non-standard mitochondrial inheritance patterns, which we describe here.

2.3 Methods

2.3.1 Sampling

Intact colonies of *Amphylaeus morosus* were sampled from five regions along the eastern coast of Australia from southern Queensland (QLD) to southwestern Victoria (VIC) (Figure 1). Colonies collected from Tin Can Bay, QLD on 10th December 2013 and Cobboboonee State Forest, VIC (Figure 1) on 22nd–24th February 2017 were collected from dead flower scapes of *Xanthorrhoea* spp. in coastal heath habitat. In the central localities in New South Wales (NSW) including Enfield State Forest on 22nd December 2018 and Blue Mountains on 24th July 2017, as well as multiple collections from the Dandenong Ranges, VIC (Figure 1), specimens were collected from fallen dead fronds of the tree fern *Cyathea australis* (R.Br.) Domin in wet montane forest. Intact bee colonies were kept at 4°C until opening and adults were transferred directly to 99% ethanol for preservation. Four collections from the Dandenong Ranges were conducted on 14th–18th August 2014, 6th–10th November 2014, 26th February – 2nd March 2015, and 21st–24th November 2016. Immature individuals from the latter collection were reared in controlled conditions at Flinders University campus, South Australia to adulthood, to obtain adult males. All individuals were identified as *A. morosus* using Houston (1975).

2.3.2 DNA extraction

Total DNA was extracted from the tissue of a single hind leg from adult *A. morosus* specimens. Sixty-nine females from across all regions and four males from the Dandenong Ranges were included (total n = 73). Extractions used an adapted Gentra Puregene Cell Kit procedure (Qiagen) at the South Australian Regional Facility of Molecular Ecology and Evolution (SARFMEE) following manufactures recommendations. Extracts were stored at 4°C prior to Polymerase Chain Reaction (PCR) amplification.

2.3.3 PCR and Sanger sequencing

PCR amplifications of mt-COI were carried out in a total volume of 25 µL, as follows: 1x MRT Buffer ($MgCl_2$), primers (0.4 µM each), 1 U Immolase DNA Polymerase (Bioline 5 U/µL), and template DNA. The universal degenerate primer set COIF-PR115 5'-TCWACNAAYCAYAARGAYATTGG-3' and COIR-PR114 5'-ACYTCNGGRTGNCCRAARARYCA-3' (Folmer et al. 1994) was used. PCR cycling conditions for specimens from the Tin Can Bay and Dandenong Ranges were one cycle (10 min at 95 ºC), 38 cycles (45 sec at 94 ºC, 45 sec at 48 ºC, 60 sec at 75 ºC) and one cycle (6

min at 72 ºC, 2 min at 25 ºC). For the remaining specimens, PCR conditions were one cycle (10 min at 94 ºC), five cycles (60 sec at 94°C, 90 sec at 45°C, 90 sec at 72°C), 35 cycles (60 sec at 94°C, 90 sec at 51°C, 60 sec at 72°C) and one cycle (10 min at 72 ºC, 2 min at 20 ºC). PCR amplified reaction products were visualised using 1.5% agarose gel. Successful PCR reaction products were purified using Multiscreen PRC384 Filter Plate (Millipore) and re-suspended in 20–25 µL of 10 mM TRIS. Amplicons were sent to the Australian Genome Research Facility (AGRF) in Adelaide, South Australia and sequenced with Applied Bio-Systems 3730 and 3730 xl capillary sequencers. Forward mt-COI sequences were edited and aligned using Geneious version 10.2.2 (https://www.geneious.com). Reverse sequences were also obtained for a subset of samples (12 individuals from the Dandenong Ranges) to confirm base identity.

2.3.4 Sequence analysis

To confirm the identity of the source species for each sequence, edited sequences were screened against the BLAST database (http://www.ncbi.nlm.nih.gov/blast) to eliminate the possibility of contamination or amplification of internal parasites such as *Wolbachia*. To explore whether sequences comprised mitochondrial genes or potential NUMTS, nucleotide sequences were translated into amino acids using both invertebrate mitochondrial and nuclear coding schemes to explore amino acid changes and identify potential stop codons using Geneious. Nucleotide sequence divergences between haplotypes were calculated using uncorrected pairwise distance using Geneious.

2.3.5 Ion Torrent shotgun sequencing and mitogenome alignment

Total DNA was extracted from four legs and thoracic tissue of a single female bee (collected from the Dandenong Ranges in March 2015), using the DNeasy Blood and Tissue extraction kit (Qiagen), following the manufacturer's protocols. The sample (2.6 µg) was sequenced on an Ion Torrent PGM platform (Life Technologies) using a 318 chip with 400 bp chemistry using standard protocols at the AGRF facility. This sample was one of four runs in unison and shearing of DNA was conducted with a Covaris S2 model (Covaris) with shearing parameters slightly modified (Duty Cycle 10%, Intensity 4,100 cycles per burst, 80 sec time) to provide more DNA fragments in the 350–400 bp range. A Pippin Prep (Sage Science) was used on these fragments to size select to ensure most fragments were greater than 300 bp and did not exceed 400 bp. Fragments were tagged by ligating with standard Ion Torrent barcode

adaptors. Equal molar ratios of fragments were mixed prior to sequencing. Post sequencing fragments were demultiplexed, base called and aligned using Torrent Suite Software.

The resultant shotgun sequence fragments were mapped onto a *Hylaeus dilatatus* (Kirby, 1802) (Colletidae: Hylaeinae) reference mitogenome (Tan et al. 2015) using the reference mapping function in Geneious. Trimmed contigs were mapped using Geneious' low sensitivity mapping function with minimum mapping quality of ten, including flexibility for changes in fragment lengths compared to the reference genome, and minimum support for structural variation set to two reads. The alignment was then manually assessed and edited. Potential single nucleotide polymorphisms (SNPs) i.e. heteroplasmic sites were identified using the Geneious SNP/Variation function (Supp. Table 1). Geneious settings included a minimum read coverage set to six reads due to low read coverage, and a minimum frequency variant of 0.15. The maximum variant p-value and minimum strand-bias p-value were set to 10^{-5} with a 65% bias. Sequence variation between the *A. morosus* shotgun reads and the *H. dilatatus* reference genome was ignored.

2.3.6 Mitochondrial gene isolation of shotgun sequences

Four mitochondrial genes were isolated from the aligned shotgun sequences produced from the female *A. morosus* specimen: cytochrome *c* oxidase subunit I (mt-COI), cytochrome *c* oxidase subunit II (mt-COII), cytochrome b (mt-cytb), and NADH-ubiquinone oxidoreductase chain 4 (mt-ND4). These genes had the highest read depths (>15 contigs) and were individually isolated from the alignment to be examined thoroughly for evidence of recombination, NUMTs, and ambiguities.

Conducted in two steps, the aligned contigs within these four genes were firstly reduced to show only suspected heteroplasmic sites (i.e. all non-variable sites within the contigs were hidden) (Supp. Table 2, Supp. Figures 1–4). The (aligned) heteroplasmic sites retained for assessment included both those detected by the Geneious SNP/Variation function and those that were manually observable but were not detected. This is because the Geneious SNP/Variation function demonstrated a bias against detecting potential heteroplasmic sites that included a T nucleotide, even if they occurred at the same frequency as other nucleotide combinations across the aligned contigs. A lower quality score assigned to T nucleotides by Ion Torrent (which is a factor the Geneious SNP/Variation detection function includes when assessing potential variable sites) could be responsible for this apparent bias. This reduced the length of the gene fragments by ~90–97%, so that contigs could be sorted into analogues

based on *variable sites only* (Supp. Table 2; Supp. Figures 1–4). An assessment of the number of 'lineages' (i.e. the number of types of mitogenomes or possible NUMTs) that were present within the sample for each of the four genes was then evaluated. Secondly, these aligned, analogous contigs were then assessed in their entirety (i.e. non-variable sites were included) and the mitochondrial protein translation scheme was applied to investigate whether there were any mitochondrial stop codons which could indicated NUMTs, rather than multiple mitogenomes.

2.4 Results

2.4.1 Chromatogram evidence of heteroplasmy

We obtained a total of 73 forward mt-COI fragments (four of these specimens were also re-sequenced to check nucleotide base pair assignments when initial sequencing chromatograms were ambiguous) and 12 reverse mt-COI Sanger sequences from the *Amphylaeus morosus* specimens, with between 613 bp and 658 bp from an alignment including forward and reverse sequences. Every specimen (females and males) exhibited double peaks at specific sites throughout the mt-COI fragment. Because these heteroplasmic sites could be identified in individuals in every population implies that these mitochondrial haplotypes are shared across the species range. For this reason, we grouped all 'shared' double peaks (observed in most individuals from *all* geographic regions) as being stable heteroplasmic sites across the species. In total, 24 heteroplasmic sites were consistently recovered across the 658 bp mt-COI fragment (Figure 2). Although every chromatogram contained double peaks and these double peaks were always consistently placed throughout the mt-COI fragment involving only two nucleotides, not all double peaks were clearly recovered in every produced sequence. These double peaks were also consistently recovered in reverse sequences and after re-sequencing tissue extracts. Heteroplasmic sites that were not detected in the original sequence were often detected in re-sequencing attempts or reverse sequences. These inconsistencies could be due to tissue, extraction, PCR, or sequencing biases of each mtDNA type. An additional double peak was detected in a single specimen from the Blue Mountains, NSW (position 310 bp; Figure 2), resulting in a novel synonymous-coding heteroplasmic site.

Mitochondrial protein translations of heteroplasmic sites indicted four amino acid changes, using the invertebrate mitochondrial coding scheme, detected via Sanger sequencing (Figure 2) — none of which involved mitochondrial stop codons (Supp. Table 3). An uncorrected 'p' pairwise distance (i.e. percentage difference in nucleotide sites) for the two mitochondrial

strains was calculated using Geneious and gave a sequence divergence of 3.8% for the 658 bp mt-COI region that was sequenced.

2.4.2 Nucleotide variation throughout the mitogenome

The Ion Torrent shotgun run produced a total number of 782,607 reads. In order to examine the extent of the potential heteroplasmic variation beyond the mt-COI gene, shotgun sequences were aligned to the *Hylaeus dilatatus* mitogenome (Tan et al. 2015). A total of 5,151 contigs aligned to the *H. dilatatus* mitogenome and after conservative editing, 466 high quality contigs were retained. Most of contigs aligned to protein-coding regions, but with some poorly matched alignments to the two rRNA regions, 16S and 12S rRNA (Figure 3). The mitochondrial protein-coding genes are more conserved in placement and composition than the rRNA and tRNA genes (Zhang and Broughton 2013). Therefore, the position and composition of *A. morosus* protein-coding genes in the mitogenome is probably similar to that of *H. dilatatus.*

Ion Torrent shotgun data indicated that heteroplasmic sites are distributed throughout the mitogenome of *A. morosus*, with *all* mitochondrial genes with aligned contigs containing variable (heteroplasmic) sites. A total of 214 heteroplasmic sites were identified using Geneious (Supp. Table 1), consisting entirely of single nucleotide substitutions (i.e. site heteroplasmy only). All of the heteroplasmic sites detected by the Geneious SNP/Variant function involved only two base pairs (Figure 3). Ion Torrent data were consistent with the Sanger sequencing data (Figure 2) but further suggested that mitochondrial heteroplasmy occurs across multiple mitochondrial genes.

2.4.3 Identification of heteroplasmic sites using two techniques

Results from the Ion Torrent shotgun alignment of the mt-COI gene were compared to the corresponding Sanger sequenced data (Supp. Table 3). Heteroplasmic sites between the two methods were generally consistent, however Ion Torrent data suggested overall more variable sites (28 compared with 24 detected using Sanger) (Figure 2; Supp. Table 3). These sites could represent either novel heteroplasmic sites in this individual or sequencing error. The Ion Torrent platform has been shown to have a higher SNP call of true positives, but also higher call of false positives, compared to the Illumina platform (Quail et al. 2012). However, given that all but one of the 24 (Sanger) heteroplasmic sites in mt-COI were otherwise recovered using both techniques, we interpreted our NGS data as being generally reliable, but possibly slightly overestimates SNP variation.

2.4.4 Ion Torrent gene isolates

For mitochondrial fragments isolated from the shotgun alignment (genes with the highest read depth, viz. mt- COI, COII, cytb and ND4), analogue contigs were sorted into two common, separate 'lineages' where adjacent heteroplasmic sites were generally not shared between lineages. Within each lineage, contigs could be further separated by within-lineage variation into 'groups' (Supp. Figures 1–4), however, these within-lineage differences presented as minor variation (dissimilar in 1–3 consecutive variable nucleotides, see arrows in Supp. Figures 1–4). The source of this within-lineage variation is unknown, but it could be due to recombination (see Supp. Figure 3 for an example), naturally occurring variation due to mutagenetic processes, or sequencing error. Further next-generation sequencing efforts to produce datasets across multiple individuals at greater read depth is recommended to address these variable sites more thoroughly. Some additional contigs (usually present as singletons; Supp. Figures 1–4) were recovered that could not be assign to either of the two main lineages. These could be the result of NUMTs (nuclear inclusion of mtDNA), additional minor mtDNA variants, or sequencing error, however these data are not extensive enough to determine their true origins.

Mitochondrial protein translations (using the invertebrate mitochondrial coding scheme) of each of the four mitochondrial gene fragments did not reveal mitochondrial stop codons at heteroplasmic sites within either of the lineages, that would suggest one lineage is due to a NUMT. However, stop codons were detected earlier in the reading frame of the gene isolates (both lineages) than would be expected compared to the *H. dilatatus* reference mitogenome (Tan et al. 2015). When these stop codons occurred in a gene, they were positioned in homologous locations in both mitochondrial lineages and succeeded AT rich regions with homopolymers, where Ion Torrent sequencing is known to be prone to errors (Quail et al. 2012, Bragg et al. 2013). Hymenopteran mitogenomes are known have a strong AT biased composition (Dowton and Austin 1997, Ma et al. 2019). Adjusting these homopolymer regions by changing the length by one base always shifted the stop codon so that it occurred at the approximately expected position in the gene.

2.5 Discussion

Our analyses of *Amphylaeus morosus* mitochondrial data indicate two very unusual phenomena; (i) a near-complete lack of mtDNA variation throughout a very large geographical range (~2,000 km), and (ii) the widespread prevalence of mt-COI heteroplasmy

restricted to consistent nucleotide sites, which shotgun sequencing data suggests extends to multiple genes in the mitogenome. The finding of uniform heteroplasmic sites maintained across both individuals and localities are suggestive of widespread heritable mitochondrial heteroplasmy retained across generations.

The lack of almost any mitochondrial haplotype variation (excepting heteroplasmy) in *A. morosus* suggests a mechanism that has removed mtDNA variation across a huge geographical range (~2,000 km) in this species. It seems unlikely that a recent population bottleneck has caused the lack of mtDNA variation, given the very extensive geographical distribution of this species (Houston 1975, Spessa 1999) (Figure 1). Mitochondrial haplotype variation within species is not well reported for bees, particularly species outside of Europe and North America. Within Hawaiian *Hylaeus* species (Colletidae: Hylaeinae) (a closely related genus to *Amphylaeus*), intra-island populations were up to 3.1% divergent, and inter-island populations up to 9.5% (Magnacca and Brown 2010a) for the mt-COI gene. It is interesting to note that these *Hylaeus* species have very recent origins of less than 700,000 mya (Magnacca and Danforth 2006) and have high rates heteroplasmy (Magnacca and Brown 2010a). A study by Dew et al. (2016) demonstrated that the Australian ceratinine bee, *Ceratina australensis* (Perkins, 1912) (Apidae: Xylocopinae) had total of 24 haplotypes among 91 specimens for the same mt-COI barcode region and included specimens from southern Queensland through to southern South Australia, which is similar to the geographical range inhabited by *A. morosus*. Therefore, the maintenance of a widespread and consistent heteroplasmic mitochondrial 'haplotype' is very unusual and the mechanisms that enable its persistence are not yet understood.

In insect studies, reports of mitochondrial heteroplasmy generally occur for the widely sequenced mt-COI 'barcode' region because of its prevalent use in systematic studies (Hendrich et al. 2011, De Mandal et al. 2014, Shokralla et al. 2014). Whole mitogenome exploration of heteroplasmy is not yet widely performed in insects. Within Hymenoptera, mitochondrial heteroplasmy has almost exclusively been explored in bees (with a single report in ants (Meza-Lázaro et al. 2018)) and exploration is generally limited to the mt-COI gene. At least 29 bee species from five of the seven bee families have been identified with unusually high rates of mitochondrial heteroplasmy (Songram et al. 2006, Magnacca and Brown 2010a, Magnacca and Brown 2012, Françoso et al. 2016, Ricardo et al. 2020), but the most extensively studied group are the *Hylaeus* of Hawai'i (Magnacca and Brown 2010a).

Magnacca and Brown (2010a) assayed Hawaiian *Hylaeus* species to detect heteroplasmy, with 21 of 49 species examined containing heteroplasmic individuals. Although only a few specimens (n = 1–13) of each species were analysed, in ~85% of tested species, all specimens were heteroplasmic in the mt-COI gene, suggesting extensive heteroplasmy in this bee group. Magnacca and Brown (2010a) also found evidence of more than two mitochondrial haplotypes within heteroplasmic *Hylaeus* individuals (as well as NUMTs in some species) and suggested that this could be explained by the accumulation of mutations in hypervariable regions of the mitochondrial genome producing variation among haplotypes. This has also been widely reported in humans and other studies (Li et al. 2015, Wachsmuth et al. 2016, Morris et al. 2017) where heteroplasmy features nucleotide variability, producing more than two haplotypes within individuals. Within *A. morosus*, the nucleotide variation appears to be highly conserved throughout the mitogenome. However, gene isolates from Ion Torrent sequencing data suggest there may be haplotype variants within mitochondrial lineages (Supp. Figures 1–4), but the source(s) of this variation have not been confidently identified and could be due to sequencing error rather than genuine intra-individual variation.

Although alternative explanations remain, it seems most likely the patterns observed in *A. morosus* are due to widespread and consistent mitochondrial heteroplasmy. Nuclear inclusions of mitochondrial genes (NUMTs) are common sources of intra-individual variation (Lopez et al. 1994, Berthier et al. 2011, Cihlar et al. 2020). Although the presence of a large NUMT is possible, NUMTs are reportedly rarely larger than 1,000 bp (Leite 2012), although some near-entire mitogenome inclusions have been documented (Kim et al. 2006, Verscheure et al. 2015). Within animals, NUMTs are not known to retain functionality (Herrnstadt et al. 1999, Erpenbeck et al. 2010) and are hence free to accumulate mutations in the absence of purifying selection (Brown et al. 1982, Sorenson and Fleischer 1996, Mundy et al. 2000, Strugnell and Lindgren 2007). A functional total mitogenome nuclear inclusion with a subsequent loss of the mitochondrial genome has been documented in a marine dinoflagellate (John et al. 2019), but seems unlikely to be applicable to animals with higher respiratory needs. “Mega-NUMTs” which are concatenated multicopy mtDNA integrated into the nuclear genome have also been postulated in humans (Balciuniene and Balciunas 2019). However, these options are very unlikely to explain our results, because:

(i) a similar number of contigs were sorted into each mitochondrial 'lineage' suggesting they occurred at a similar frequency within the *A. morosus* sample, inconsistent with one of the sources being from single nuclear DNA insert,
(ii) whole-genome (particularly multicopy) transferrals into the nucleus are very rare,
(iii) no mitochondrial stop codons were associated within any heteroplasmic sites within Ion Torrent gene isolates of either lineage,
(iv) mitochondrial heteroplasmy has been documented in other hylaeine (and bee) species, and
(v) it would require the 'NUMT' spreading over an enormous geographical range without acquiring additional mutations that would have been detected as sequence variation.

Additional Ion Torrent contigs within gene isolates were also identified which did not match either proposed mitochondrial lineage, usually possessing inconsistent combinations of SNPs with either mitochondrial lineage (Supp. Figures 1–4). In all cases, between one and two 'outlier' contigs were produced. Unfortunately, these data are not comprehensive enough to investigate the sources of this additional 'outlier' variation. Although Ion Torrent sequencing has now been surpassed by numerous improved NGS technologies, its ability to detect mitochondrial heteroplasmy is more sensitive and accurate than Sanger sequencing (Magalhães et al. 2015). Our data are limited by the known constraints of Ion Torrent and Sanger sequencing, low sample sizes within some localities, and our tissue selection (tissue segregation has been shown for heteroplasmy (Magnacca and Brown 2010b) but was not explored this study). However, there are worthy patterns within this study system which present interesting challenges for our understanding of the inheritance of mtDNA.

The effective population size of mtDNA during embryogenesis is very small, with estimates being a few hundred copies per cell in the early stages of primordial germ cell production (Jansen 2000, Cao et al. 2007, Cree et al. 2008). Such a small effective population size should lead to very strong genetic drift, such that if two different mitogenomes are present at any one time, either of them could be rapidly lost across multiple cell-division cycles. Consequently, if there are indeed two divergent mitogenomes present in *A. morosus*, there must be some mechanism(s) that operate to maintain their dual presence in a way that counteracts genetic drift. The heteroplasmic sites identified in *A. morosus* were also detected consistently in the four males. This species, like other hymenopterans, is sexually determined by haplodiploidy

(where males are produced from unfertilized eggs which means they are produced with no paternal contribution). Therefore, because males are also heteroplasmic, mitochondrial heteroplasmy must be maternally inherited in this species. This is unusual because in these kinds of systems, heteroplasmy has sometimes been suggested to be due to paternal leakage (e.g. Kvist et al. (2003), Hoolahan et al. (2011), Mastrantonio et al. (2019)), although experimental confirmation of this is rare (Nunes et al. 2013).

Maternal inheritance of mitochondrial heteroplasmy has also been proposed for the ant species *Ectatomma ruidum* (Meza-Lázaro et al. 2018), however for other Hymenoptera and haplodiploid groups, the mechanisms that might enable heteroplasmy to persist with populations are still unknown (e.g. Magnacca and Brown (2010a), Magnacca and Brown (2012), Gawande et al. (2017)). Although these mechanisms may explain the occurrence of heteroplasmy, those that *maintain* mitochondrial heteroplasmy are unclear — mechanisms which preserve widespread, consistent, and heritable heteroplasmy. It could be that there are some ancestral or shared mechanisms within bees/hymenopterans/haplodiploid species which enable high rates of heteroplasmy. Unfortunately, there are not enough definitive studies of heteroplasmy to determine what such factors might be.

2.6 Conclusions and suggestions for future research

A complete lack of mt-COI variation other than widespread and consistent mitochondrial heteroplasmy was identified in every individual (n = 73) of *Amphylaeus morosus* in five localities across a ~2000 km range (Figure 1). Additionally, Ion Torrent shotgun sequencing data obtained from one female, further suggested extensive heteroplasmic sites (>200 SNPs) occurring throughout the mitogenome, far more than has been reported in other heteroplasmic systems (Xiong et al. 2013, Moreira et al. 2017, Meza-Lázaro et al. 2018, Sriboonlert and Wonnapinij 2019). The maintenance of divergent heteroplasmic mitogenomes, excluding all other mtDNA variation, suggests complex, non-standard mitochondrial inheritance mechanisms taking place in this species. A number of mechanisms have been proposed that could maintain certain (heritable) mitochondrial haplotypes (and even heteroplasmic haplotypes) in arthropods including:

(i) ‘Standard’ positive and/or balancing selection: coexistence of multiple mitogenomes, each providing unique selective advantages whereby heteroplasmic individuals are favoured in the population (e.g. Doublet et al. (2012), Ma et al. (2014), Kastally and Mardulyn (2017)).

(ii) Breakdown of purifying or positive selective processes (Burr et al. 2018, Palozzi et al. 2018). For example, this could include selection biases in the germline genetic bottleneck (Ashley et al. 1989, Freyer et al. 2012), an inability to purge deleterious mitochondrial mutations (Chen et al. 2014, Lin et al. 2016, Samstag et al. 2018), or the spread of selfish mitochondrial genes/genomes at the expense of those with high respiratory function (Gitschlag et al. 2016, Ma and O'Farrell 2016). This hypothesis favours a temporary heteroplasmic phase, unlike what we observe in *A. morosus*.

(iii) Genetic hitchhiking (induced-positive selection) of mitogenomes facilitated by reproductive parasites; a selective sweep on co-inherited mitogenomes as they are 'dragged' through by heritable endosymbionts (Marcadé et al. 2007, Schuler et al. 2016).

We have identified that *A. morosus* is also infected with the intracellular bacteria *Wolbachia* (Rickettsiales: Anapasmataceae) (Chapter 3), which is known to cause selective sweeps on host mtDNA removing genetic variation from the population (Jiggins 2003, Schuler et al. 2016). *Wolbachia* seems a likely candidate to explain the lack of mtDNA variation within this species but does not easily explain the prevalent mitochondrial heteroplasmy. This will be discussed further in Chapters 3 and 4.

2.7 Acknowledgements

We would like to thank Rebecca Dew, Ben Parslow, and James Dorey for their assistance in fieldwork. Thank you to Celina Rebola for providing comments on a draft of this manuscript. This work was completed under awards from the Australian Federal Government's Research Training Program Scholarship, as well as the Playford Memorial Trust PhD scholarship. Funding for this project was provided by the Australian Pacific Science Foundation and Holsworth Wildlife Trust. Insect specimens were collected under the sampling permits SL101857 (NSW) and SL102137 (NSW).

2.8 References

Ashley, M. V., P. J. Laipis, and W. W. Hauswirth. 1989. Rapid segregation of heteroplasmic bovine mitochondria. Nucleic Acids Research **17**:7325–7331.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics **18**:489–522.

Balciuniene, J., and D. Balciunas. 2019. A nuclear mtDNA concatemer (mega-NUMT) could mimic paternal inheritance of mitochondrial genome. Frontiers in Genetics **10**:1–3.
Bandelt, H. J., Q. P. Kong, W. Parson, and A. Salas. 2005. More evidence for non-maternal inheritance of mitochondrial DNA? Journal of Medical Genetics **42**:957–960.
Barr, C. M., M. Neiman, and D. R. Taylor. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist **168**:39–50.
Berthier, K., M. Chapuis, S. M. Moosavi, D. Tohidi-Esfahani, and G. A. Sword. 2011. Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. Systematic Entomology **36**:285–299.
Boyce, T. M., M. E. Zwick, and C. F. Aquadro. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123**:825–836.
Boyle, E. E., and R. J. Etter. 2013. Heteroplasmy in a deep-sea protobranch bivalve suggests an ancient origin of doubly uniparental inheritance of mitochondria in Bivalvia. Marine Biology **160**:413–422.
Bragg, L. M., G. Stone, M. K. Butler, P. Hugenholtz, and G. W. Tyson. 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLOS Computational Biology **9**:e1003031.
Brown, W., E. Prager, A. Wang, and A. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. Journal of Molecular Evolution **18**:225–239.
Burr, S. P., M. Pezet, and P. F. Chinnery. 2018. Mitochondrial DNA heteroplasmy and purifying selection in the mammalian female germ line. Development, Growth & Differentiation **60**:21–32.
Cao, L., H. Shitara, T. Horii, Y. Nagao, H. Imai, K. Abe, T. Hara, J. Hayashi, and H. Yonekawa. 2007. The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. Nature Genetics **39**:386–390.
Chen, Z., Y. Qi, S. French, G. Zhang, R. C. Garcia, R. Balaban, and H. Xu. 2014. Genetic mosaic analysis of a deleterious mitochondrial DNA mutation in *Drosophila* reveals novel aspects of mitochondrial regulation and function. Molecular Biology of the Cell **26**:674–684.
Cihlar, J. C., C. Strobl, R. Lagacé, M. Muenzler, W. Parson, and B. Budowle. 2020. Distinguishing mitochondrial DNA and NUMT sequences amplified with the precision ID mtDNA whole genome panel. Mitochondrion **55**:122–133.
Cree, L. M., D. C. Samuels, S. C. de Sousa Lopes, H. K. Rajasimha, P. Wonnapinij, J. R. Mann, H. H. Dahl, and P. F. Chinnery. 2008. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nature Genetics **40**:249–254.
De Mandal, S., L. Chhakchhuak, G. Gurusubramanian, and N. S. Kumar. 2014. Mitochondrial markers for identification and phylogenetic studies in insects – A Review. DNA Barcodes **2**:1–9.
Dew, R. M., S. M. Rehan, and M. P. Schwarz. 2016. Biogeography and demography of an Australian native bee *Ceratina australensis* (Hymenoptera: Apidae) since the last glacial maximum. Journal of Hymenoptera Research **49**:25–41.
Domin, K. 1929. Pteridophyta (in Czech). Nákladem České Akademie Vêd a Umêni. , Praha, Czech Republic
Doublet, V., R. Raimond, F. Grandjean, A. Lafitte, C. Souty-Grosset, and I. Marcade. 2012. Widespread atypical mitochondrial DNA structure in isopods (Crustacea: Peracarida) related to a constitutive heteroplasmy in terrestrial species. Genome **55**:234–244.
Doublet, V., C. Souty-Grosset, D. Bouchon, R. Cordaux, and I. Marcadé. 2008. A thirty million year-old inherited heteroplasmy. PLOS ONE **3**:e2938.

Dowton, M., and A. D. Austin. 1997. The evolution of strand-specific compositional bias. A case study in the Hymenopteran mitochondrial 16S rRNA gene. Molecular Biology and Evolution **14**:109–112.

Erpenbeck, D., O. Voigt, M. Adamski, B. J. Woodcroft, J. N. A. Hooper, G. Wörheide, and B. M. Degnan. 2010. NUMTs in the sponge genome reveal conserved transposition mechanisms in metazoans. Molecular Biology and Evolution **28**:1–5.

Folmer, O., W. Hoeh, M. Black, and R. Vrijenhoek. 1994. Conserved primers for PCR amplification of mitochondrial DNA from different invertebrate phyla. Molecular Marine Biology and Biotechnology **3**:294–299.

Françoso, E., A. R. Zuntini, A. C. Carnaval, and M. C. Arias. 2016. Comparative phylogeography in the Atlantic forest and Brazilian savannas: pleistocene fluctuations and dispersal shape spatial patterns in two bumblebees. BMC Evolutionary Biology **16**:267.

Freyer, C., L. M. Cree, A. Mourier, J. B. Stewart, C. Koolmeister, D. Milenkovic, T. Wai, V. I. Floros, E. Hagström, E. E. Chatzidaki, R. J. Wiesner, D. C. Samuels, N. G. Larsson, and P. F. Chinnery. 2012. Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. Nature Genetics **44**:1282–1285.

Gandolfi, A., B. Crestanello, A. Fagotti, F. Simoncelli, S. Chiesa, M. Girardi, E. Giovagnoli, C. Marangoni, I. Di Rosa, and L. Lucentini. 2017. New evidences of mitochondrial DNA heteroplasmy by putative paternal leakage between the rock partridge (*Alectoris graeca*) and the chukar partridge (*Alectoris chukar*). PLOS ONE **12**:e0170507.

Gawande, S. J., S. Anandhan, A. A. Ingle, A. Jacobson, and R. Asokan. 2017. Heteroplasmy due to coexistence of mtCOI haplotypes from different lineages of the *Thrips tabaci* cryptic species group. Bulletin of Entomological Research **107**:534–542.

Gitschlag, Bryan L., Cait S. Kirby, David C. Samuels, Rama D. Gangula, Simon A. Mallal, and Maulik R. Patel. 2016. Homeostatic responses regulate selfish mitochondrial genome dynamics in *C. elegans*. Cell Metabolism **24**:91–103.

Gyllensten, U., D. Wharton, A. Josefsson, and A. C. Wilson. 1991. Paternal inheritance of mitochondrial DNA in mice. Nature **352**:255–257.

Hassanin, A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: Strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Molecular Phylogenetics and Evolution **38**:100–116.

Hendrich, L., J. Pons, I. Ribera, and M. Balke. 2011. Mitochondrial *COX1* sequence data reliably uncover patterns of insect dversity but suffer from high lineage-idiosyncratic error rates. PLOS ONE **5**:e14448.

Herrnstadt, C., W. Clevenger, S. S. Ghosh, C. Anderson, E. Fahy, S. Miller, N. Howell, and R. E. Davis. 1999. A Novel mitochondrial DNA-like sequence in the human nuclear genome. Genomics **60**:67–77.

Hlaing, T., W. Tun-Lin, P. Somboon, D. Socheat, T. Setha, S. Min, M. S. Chang, and C. Walton. 2009. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. BMC Genetics **10**:11.

Hoarau, G., S. Holla, R. Lescasse, W. T. Stam, and J. L. Olsen. 2002. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. Molecular Biology and Evolution **19**:2261–2264.

Hoelzel, A. R., J. M. Hancock, and G. A. Dover. 1993. Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. Journal of Molecular Evolution **37**:190–197.

Hoolahan, A. H., V. C. Blok, T. Gibson, and M. Dowton. 2011. Paternal leakage of mitochondrial DNA in experimental crosses of populations of the potato cyst nematode *Globodera pallida*. Genetica **139**:1509–1519.

Hoolahan, A. H., V. C. Blok, T. Gibson, and M. Dowton. 2012. Evidence of animal mtDNA recombination between divergent populations of the potato cyst nematode *Globodera pallida*. Genetica **140**:19–29.

Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.

Irwin, J. A., J. L. Saunier, H. Niederstätter, K. M. Strouss, K. A. Sturk, T. M. Diegoli, A. Brandstätter, W. Parson, and T. J. Parsons. 2009. Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. Journal of Molecular Evolution **68**:516–527.

Jansen, R. P. S. 2000. Germline passage of mitochondria: quantitative considerations and possible embryological sequelae. Human Reproduction **15**:112–128.

Jebb, D., N. M. Foley, C. V. Whelan, F. Touzalin, S. J. Puechmaille, and E. C. Teeling. 2018. Population level mitogenomics of long-lived bats reveals dynamic heteroplasmy and challenges the Free Radical Theory of Ageing. Scientific Reports **8**:13634.

Jiggins, F. M. 2003. Male-killing *Wolbachia* and mitochondrial DNA: sselective sweeps, hybrid introgression and parasite population dynamics. Genetics **164**:5–12.

John, U., Y. Lu, S. Wohlrab, M. Groth, J. Janouškovec, G. S. Kohli, F. C. Mark, U. Bickmeyer, S. Farhat, M. Felder, S. Frickenhaus, L. Guillou, P. J. Keeling, A. Moustafa, B. M. Porcel, K. Valentin, and G. Glöckner. 2019. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. Science Advances **5**:eaav1110.

Ju, Y. S., L. B. Alexandrov, M. Gerstung, I. Martincorena, S. Nik-Zainal, M. Ramakrishna, H. R. Davies, E. Papaemmanuil, G. Gundem, A. Shlien, N. Bolli, S. Behjati, P. S. Tarpey, J. Nangalia, C. E. Massie, A. P. Butler, J. W. Teague, G. S. Vassiliou, A. R. Green, M. Du, A. Unnikrishnan, J. E. Pimanda, B. T. Teh, N. Munshi, M. Greaves, P. Vyas, A. K. El-Naggar, T. Santarius, V. P. Collins, R. Grundy, J. A. Taylor, D. N. Hayes, D. Malkin, I. B. C. Group, I. C. M. D. Group, I. P. C. Group, C. S. Foster, A. Y. Warren, H. C. Whitaker, D. Brewer, R. Eeles, C. Cooper, D. Neal, T. Visakorpi, W. B. Isaacs, G. S. Bova, A. M. Flanagan, P. A. Futreal, A. G. Lynch, P. F. Chinnery, U. McDermott, M. R. Stratton, and P. J. Campbell. 2014. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. eLife **3**:e02935.

Kang, A. R., M. J. Kim, I. A. Park, K. Y. Kim, and I. Kim. 2016. Extent and divergence of heteroplasmy of the DNA barcoding region in *Anapodisma miramae* (Orthoptera: Acrididae). Mitochondrial DNA Part A **27**:3405–3414.

Kastally, C., and P. Mardulyn. 2017. Widespread co-occurrence of two distantly related mitochondrial genomes in individuals of the leaf beetle *Gonioctena intermedia*. Biology Letters **13**:20170570.

Kennedy, S. R., J. J. Salk, M. W. Schmitt, and L. A. Loeb. 2013. Ultra-sensitive sequencing reveals an age-related increase in Somatic mitochondrial mutations that are inconsistent with oxidative damage. PLOS Genetics **9**:e1003794.

Kim, J., A. Antunes, S. Luo, J. Menninger, W. G. Nash, S. J. O'Brien, and W. E. Johnson. 2006. Evolutionary analysis of a large mtDNA translocation (NUMT) into the nuclear genome of the *Panthera* genus species. Gene **366**:292–302.

Kirby, W. 1802. *Monographia Apum Angliae*. J. Raw Printers, Ipswich (UK).
Kmiec, B., M. Woloszynska, and H. Janska. 2006. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Current Genetics **50**:149–159.
Kvist, L., J. Martens, A. A. Nazarenko, and M. Orell. 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Molecular Biology and Evolution **20**:243–247.
Leaché, A. D., and J. A. McGuire. 2006. Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. Molecular Phylogenetics and Evolution **39**:628–644.
Li, M., R. Schröder, S. Ni, B. Madea, and M. Stoneking. 2015. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proceedings of the National Academy of Sciences of the United States of America **112**:2491–2496.
Lin, Y. F., A. M. Schulz, M. W. Pellegrino, Y. Lu, S. Shaham, and C. M. Haynes. 2016. Maintenance and propagation of a deleterious mitochondrial genome by the mitochondrial unfolded protein response. Nature **533**:416–419.
Lopez, J. V., N. Yuhki, R. Masuda, W. Modi, and S. J. O'Brien. 1994. NUMT, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution **39**:174–190.
Lunt, D. H., L. E. Whipple, and B. C. Hyman. 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Molecular Ecology **7**:1441–1455.
Ma, H., H. Xu, and P. H. O'Farrell. 2014. Transmission of mitochondrial mutations and action of purifying selection in *Drosophila melanogaster*. Nature Genetics **46**:393.
Ma, H. S., and P. H. O'Farrell. 2016. Selfish drive can trump function when animal mitochondrial genomes compete. Nature Genetics **48**:798–802.
Ma, Y., B.-y. Zheng, J.-c. Zhu, C. van Achterberg, P. Tang, and X.-x. Chen. 2019. The first two mitochondrial genomes of wood wasps (Hymenoptera: Symphyta): novel gene rearrangements and higher-level phylogeny of the basal hymenopterans. International Journal of Biological Macromolecules **123**:1189–--1196.
Magalhães, S., S. L. Marques, C. Alves, A. Amorim, L. Alvarez, and A. Goios. 2015. Evaluation of heteroplasmy detection in the Ion Torrent PGM. Forensic Science International: Genetics Supplement Series **5**:e13–e15.
Magnacca, K. N., and M. J. Brown. 2010a. Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). BMC Evolutionary Biology **10**:1–16.
Magnacca, K. N., and M. J. F. Brown. 2010b. Tissue segregation of mitochondrial haplotypes in heteroplasmic Hawaiian bees: implications for DNA barcoding. Molecular Ecology Resources **10**:60–68.
Magnacca, K. N., and M. J. F. Brown. 2012. DNA barcoding a regional fauna: Irish solitary bees. Molecular Ecology Resources **12**:990–998.
Magnacca, K. N., and B. N. Danforth. 2006. Evolution and biogeography of native Hawaiian *Hylaeus* bees (Hymenoptera: Colletidae). Cladistics **22**:393–411.
Marcadé, I., R. Cordaux, V. Doublet, C. Debenest, D. Bouchon, and R. Raimond. 2007. Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda: Crustacea). Journal of Molecular Evolution **65**:651–659.
Mastrantonio, V., M. S. Latrofa, D. Porretta, R. P. Lia, A. Parisi, R. Iatta, F. Dantas-Torres, D. Otranto, and S. Urbanelli. 2019. Paternal leakage and mtDNA heteroplasmy in *Rhipicephalus* spp. ticks. Scientific Reports **9**:1460–1460.

Meza-Lázaro, R. N., C. Poteaux, N. J. Bayona-Vásquez, M. G. Branstetter, and A. Zaldívar-Riverón. 2018. Extensive mitochondrial heteroplasmy in the neotropical ants of the *Ectatomma ruidum* complex (Formicidae: Ectatomminae). Mitochondrial DNA Part A **29**:1203–1214.

Moreira, D. A., P. A. Buckup, C. Furtado, A. L. Val, R. Schama, and T. E. Parente. 2017. Reducing the information gap on Loricarioidei (Siluriformes) mitochondrial genomics. BMC Genomics **18**:345.

Morris, J., Y. Na, H. Zhu, J. Lee, H. Giang, A. V. Ulyanova, G. H. Baltuch, S. Brem, H. I. Chen, D. K. Kung, T. H. Lucas, D. M. O'Rourke, J. A. Wolf, M. S. Grady, J. Sul, J. Kim, and J. Eberwine. 2017. Pervasive within-mitochondrion single-nucleotide variant heteroplasmy as revealed by single-mitochondrion sequencing. Cell Reports **21**:2706–2713.

Mundy, N. I., A. Pissinatti, and D. S. Woodruff. 2000. Multiple nuclear insertions of mitochondrial cytochrome b sequences in callitrichine primates. Molecular Biology and Evolution **17**:1075–1080.

Muntaabski, I., R. M. Russo, M. C. Liendo, M. A. Palacio, J. L. Cladera, S. B. Lanzavecchia, and A. C. Scannapieco. 2020. Genetic variation and heteroplasmy of *Varroa destructor* inferred from ND4 mtDNA sequences. Parasitology Research **119**:411–421.

Nunes, M. D., M. Dolezal, and C. Schlotterer. 2013. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. Molecular Ecology **22**:2106–2117.

Palozzi, J. M., S. P. Jeedigunta, and T. R. Hurd. 2018. Mitochondrial DNA purifying selection in mammals and invertebrates. Journal of Molecular Biology **430**:4834–4848.

Parson, W. 2013. Mitochondrial DNA. Pages 264–271 *in* Encyclopedia of Forensic Sciences (Second Edition) J. A. Siegel, P. J. Saukko, and M. M. Houck, editors. Academic Press, Waltham.

Pedraza-Marrón, C. D. R., R. Silva, J. Deeds, S. M. Van Belleghem, A. Mastretta-Yanes, O. Domínguez-Domínguez, R. A. Rivero-Vega, L. Lutackas, D. Murie, D. Parkyn, L. H. Bullock, K. Foss, H. Ortiz-Zuazaga, J. Narváez-Barandica, A. Acero, G. Gomes, and R. R. Betancur. 2019. Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. Proceedings of the Royal Society B: Biological Sciences **286**:20182924.

Perkins, R. C. L. 1912. XII.—Notes, with descriptions of new species, on Aculeate Hymenoptera of the Australian region. Annals and Magazine of Natural History **9**:96–121.

Pietan, L. L., T. A. Spradling, and J. W. Demastes. 2016. The mitochondrial cytochrome oxidase subunit I gene occurs on a minichromosome with extensive heteroplasmy in two species of chewing lice, *Geomydoecus aurei* and *Thomomydoecus minor*. PLOS ONE **11**:e0162248.

Pinto, M., and C. T. Moraes. 2015. Mechanisms linking mtDNA damage and aging. Free Radical Biology and Medicine **85**:250–258.

Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics **13**:341.

Ricardo, P. C., E. Françoso, and M. C. Arias. 2020. Fidelity of DNA polymerases in the detection of intraindividual variation of mitochondrial DNA. Mitochondrial DNA Part B: Resources **5**:108–112.

Robison, G. A., O. Balvin, C. Schal, E. L. Vargo, and W. Booth. 2015. Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the bed bug (Hemiptera: Cimicidae). Journal of Medical Entomology **52**:734–738.

Samstag, C. L., J. G. Hoekstra, C.-H. Huang, M. J. Chaisson, R. J. Youle, S. R. Kennedy, and L. J. Pallanck. 2018. Deleterious mitochondrial DNA point mutations are overrepresented in *Drosophila* expressing a proofreading-defective DNA polymerase γ. PLOS Genetics **14**:e1007805.

Sańko, T. J., and A. Burzyński. 2014. Co-expressed mitochondrial genomes: recently masculinized, recombinant mitochondrial genome is co-expressed with the female – transmitted mtDNA genome in a male *Mytilus trossulus* mussel from the Baltic Sea. BMC Genetics **15**:1–40.

Santos, R. X., S. C. Correia, X. Zhu, M. A. Smith, P. I. Moreira, R. J. Castellani, A. Nunomura, and G. Perry. 2013. Mitochondrial DNA oxidative damage and repair in aging and Alzheimer's disease. Antioxid Redox Signal **18**:2444–2457.

Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S. Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffer, W. Arthofer, and M. Riegler. 2016. The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Molecular Ecology **25**:1595–1609.

Shokolenko, I. N., G. L. Wilson, and M. F. Alexeyev. 2014. Aging: a mitochondrial DNA perspective, critical analysis and an update. World Journal of Experimental Medicine **4**:46–57.

Shokralla, S., J. F. Gibson, H. Nikbakht, D. H. Janzen, W. Hallwachs, and M. Hajibabaei. 2014. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular Ecology Resources **14**:892–901.

Smith, F. 1879. Descriptions of New Species of Hymenoptera in the Collection of the British Museum, London: British Museum.

Song, H., J. E. Buhay, M. F. Whiting, and K. A. Crandall. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the National Academy of Sciences of the United States of America **105**:13486–13491.

Songram, O., S. Sittipraneed, and S. Klinbunga. 2006. Mitochondrial DNA diversity and genetic differentiation of the honeybee (*Apis cerana*) in Thailand. Biochemical Genetics **44**:256–269.

Sorenson, M. D., and R. C. Fleischer. 1996. Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proceedings of the National Academy of Sciences of the United States of America **93**:15239–15243.

Spessa, A. 1999. The behavioural and population ecology of an Australian native bee, *Amphylaeus morosus* Smith (Colletidae: Hylaeinae). Thesis. Australian National University, Australia.

Spessa, A., M. P. Schwarz, and M. Adams. 2000. Sociality in *Amphylaeus morosus* (Hymenoptera: Colletidae: Hylaeinae). Annals of the Entomological Society of America **93**:684–692.

Sriboonlert, A., and P. Wonnapinij. 2019. Comparative mitochondrial genome analysis of the firefly, *Inflata indica* (Coleoptera: Lampyridae) and the first evidence of heteroplasmy in fireflies. International Journal of Biological Macromolecules **121**:671–676.

Strugnell, J., and A. Lindgren. 2007. A barcode of life database for the Cephalopoda? Considerations and concerns. Reviews in Fish Biology and Fisheries **17**:337–344.

Tan, M., R. Zhang, C. Hardman, and X. Zhou. 2015. Mitochondrial genome of *Hylaeus dilatatus* (Hymenoptera: Colletidae). Mitochondrial DNA:1–2.

Trifunovic, A., A. Wredenberg, M. Falkenberg, J. N. Spelbrink, A. T. Rovio, C. E. Bruder, M. Bohlooly-Y, S. Gidlöf, A. Oldfors, R. Wibom, J. Törnell, H. T. Jacobs, and N.-G. Larsson. 2004. Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature **429**:417–423.

Verscheure, S., T. Backeljau, and S. Desmyter. 2015. In silico discovery of a nearly complete mitochondrial genome NUMT in the dog (*Canis lupus familiaris*) nuclear genome. Genetica **143**:453–458.

Wachsmuth, M., A. Hübner, M. Li, B. Madea, and M. Stoneking. 2016. Age-related and heteroplasmy-related variation in human mtDNA copy number. PLOS Genetics **12**:e1005939.

Xiong, H., S. C. Barker, T. D. Burger, D. Raoult, and R. Shao. 2013. Heteroplasmy in the mitochondrial genomes of human lice and ticks revealed by high throughput sequencing. PLOS ONE **8**:e73329.

Ye, K., J. Lu, F. Ma, A. Keinan, and Z. Gu. 2014. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. Proceedings of the National Academy of Sciences of the United States of America **111**:10654–10659.

Zhang, F., and R. E. Broughton. 2013. Mitochondrial–nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? Genome Biology and Evolution **5**:1781–1791.

Zsurka, G., Y. Kraytsberg, T. Kudina, C. Kornblum, C. E. Elger, K. Khrapko, and W. S. Kunz. 2005. Recombination of mitochondrial DNA in skeletal muscle of individuals with multiple mitochondrial DNA heteroplasmy. Nature Genetics **37**:873–877.

2.9 Figures

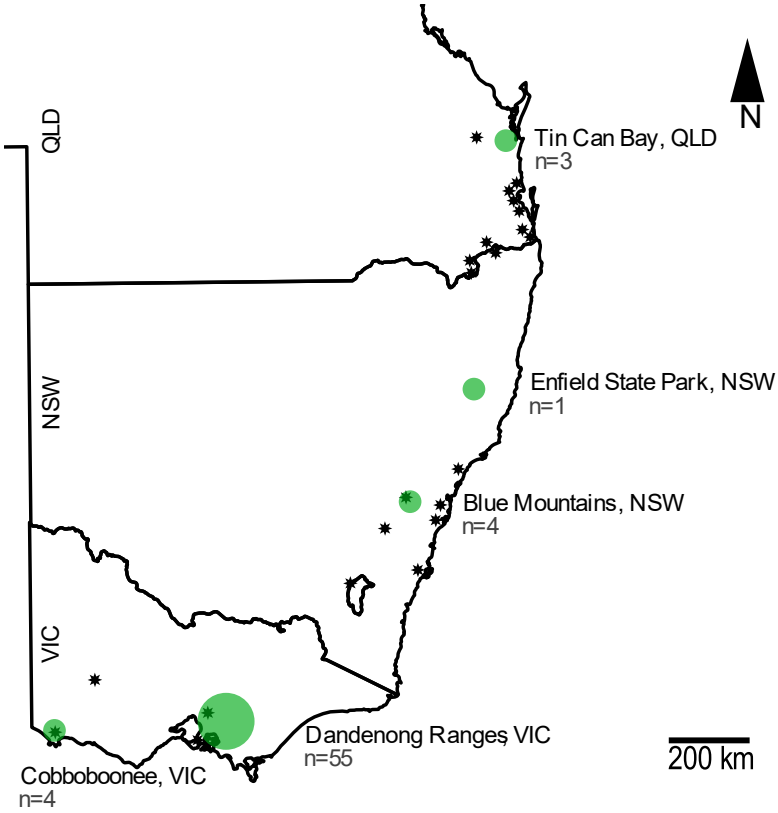


Figure 1: Recorded distribution (star points) of *Amphylaeus morosus* as reported by Houston (1975). Collection sites (green circles) for this study occurred across five locations of eastern Australia. Sample sizes of individuals barcoded from each location are included.

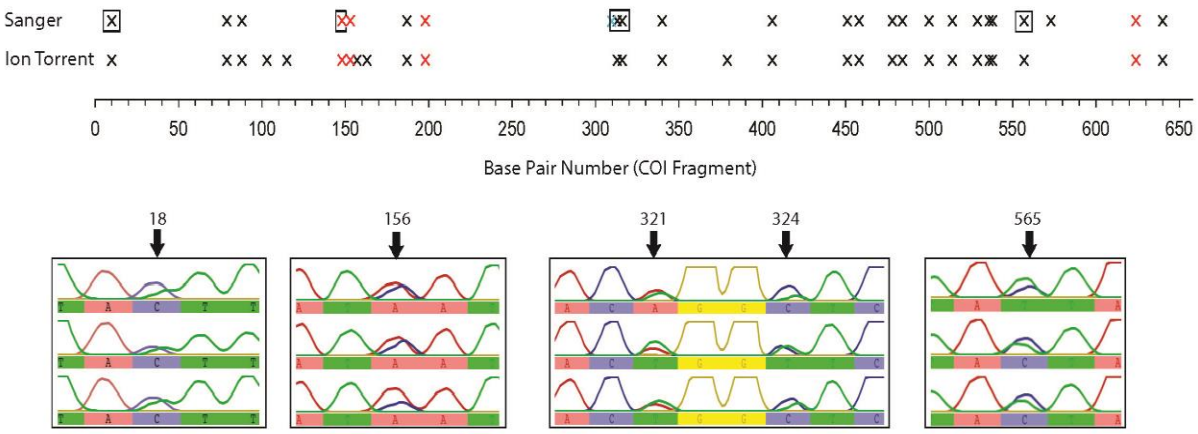


Figure 2: Summary of mt-COI heteroplasmic sites found using two techniques; Sanger sequencing and Ion Torrent shotgun sequencing. Black crosses represent nucleotide substitutions that were synonymous, and red crosses represent nucleotide substitutions that were non-synonymous (amino acid changing). An additional synonymous nucleotide was detected at 310 bp in an individual from the Blue Mountains, New South Wales (in blue). The chromatograms below are examples of the double peaks observed in the chromatogram of Sanger sequences. Their nucleotide position in the COI fragment is indicated by the sites encapsulated in boxes in the technique summary, which correspond to the numbers above the chromatograms.

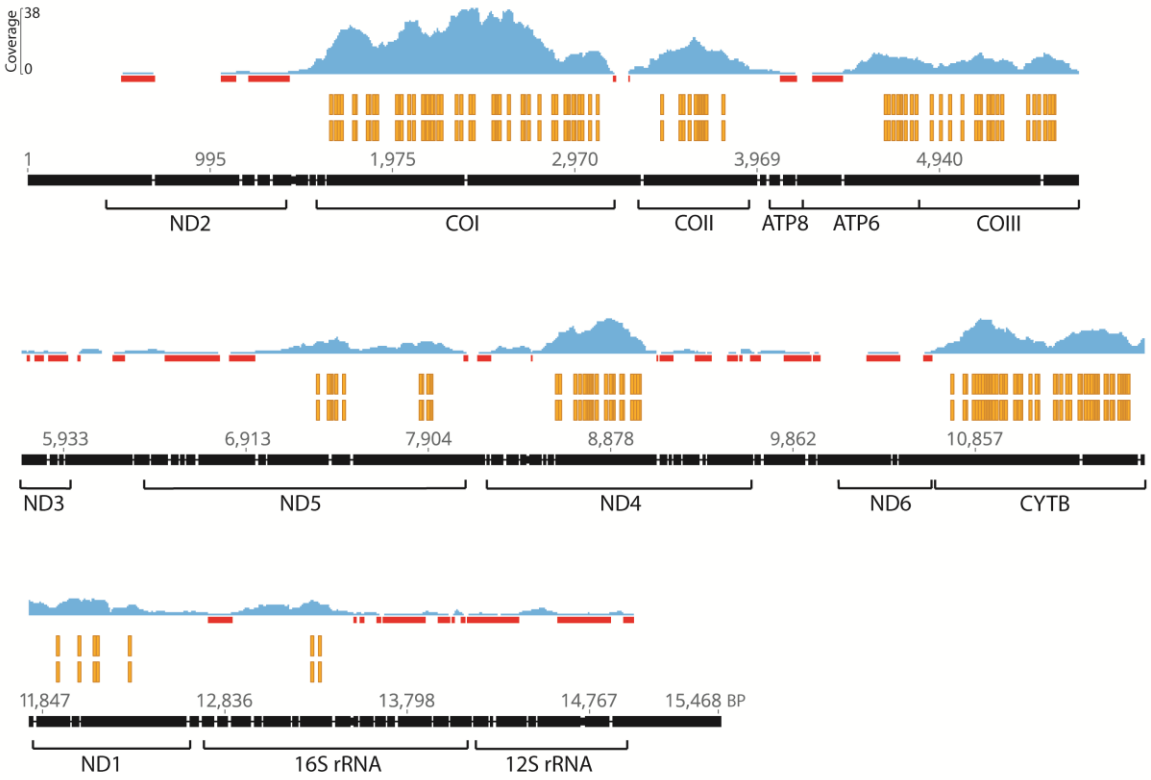


Figure 3: *Amphylaeus morosus* Ion Torrent shotgun sequencing alignment to the *Hylaeus dilatatus* reference mitogenome (Tan *et al.* 2015). The locations of the major protein-coding and rRNA genes of *H. dilatatus* mitochondrial genome are shown. Read coverage (blue) represents the comparative amount of the *A. morosus* shotgun contigs that aligned to the reference genome and red bars indicate where coverage was only a single contig. SNP/variation was determined using Geneious version 10.2.2 (https://www.geneious.com) which is indicated by the yellow markers and suggests widespread heteroplasmic sites throughout the mitogenome, only ever involving two base pairs.

2.10 Supplementary Material

Supp. Table 4: SNP/Variation data determined using Geneious version 10.2.2 for the *Amphylaeus morosus* Ion Torrent shotgun mitogenome alignment. This table provides information for each nucleotide involved in a variable 'heteroplasmic' site (all determined to be SNPs (single nucleotide polymorphisms)), with every two rows reporting the two nucleotides associated with a heteroplasmic site. Sequence position (Min., Max.) provides the location of these variable sites in relation to the alignment consensus. The frequency for each nucleotide in the variable site and given as well as the approximate P-value of the nucleotide identified as a SNP as determined by the analysis.

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
A	1635	1635	1	15	SNP	53.30%	1.00E-17
T	1635	1635	1	15	SNP	33.30%	9.40E-12
C	1666	1666	1	20	SNP	55.00%	1.70E-28
T	1666	1666	1	20	SNP	45.00%	2.50E-15
C	1678	1678	1	23	SNP	69.60%	2.40E-35
T	1678	1678	1	23	SNP	30.40%	1.50E-13
C	1699	1699	1	25	SNP	56.00%	1.70E-30
T	1699	1699	1	25	SNP	44.00%	4.80E-15
A	1768	1768	1	26	SNP	46.20%	1.50E-28
T	1768	1768	1	26	SNP	53.80%	3.50E-23
C	1777	1777	1	25	SNP	56.00%	1.10E-27
T	1777	1777	1	25	SNP	44.00%	5.50E-26
C	1842	1842	1	22	SNP	54.50%	9.90E-24
T	1842	1842	1	22	SNP	45.50%	6.30E-21
C	1846	1846	1	20	SNP	30.00%	8.80E-09
T	1846	1846	1	20	SNP	70.00%	1.80E-14
C	1852	1852	1	20	SNP	15.00%	0.0000022
T	1852	1852	1	20	SNP	85.00%	4.50E-35
C	1876	1876	1	17	SNP	58.80%	1.90E-24
T	1876	1876	1	17	SNP	41.20%	7.30E-12
G	1887	1887	1	15	SNP	46.70%	1.60E-16
T	1887	1887	1	15	SNP	53.30%	2.30E-11
A	2002	2002	1	22	SNP	27.30%	1.90E-14
T	2002	2002	1	22	SNP	72.70%	7.40E-44
C	2005	2005	1	23	SNP	26.10%	4.00E-16
T	2005	2005	1	23	SNP	69.60%	1.50E-30
C	2029	2029	1	24	SNP	29.20%	6.10E-10
T	2029	2029	1	24	SNP	70.80%	2.10E-39
A	2068	2068	1	31	SNP	64.50%	8.20E-45
G	2068	2068	1	31	SNP	32.30%	4.10E-17
C	2095	2095	1	29	SNP	31.00%	3.00E-16
T	2095	2095	1	29	SNP	69.00%	9.70E-44
A	2140	2140	1	24	SNP	62.50%	1.20E-24
T	2140	2140	1	24	SNP	37.50%	8.80E-12
C	2147	2147	1	22	SNP	59.10%	7.70E-29
T	2147	2147	1	22	SNP	40.90%	6.20E-23

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	2167	2167	1	22	SNP	59.10%	3.00E-26
T	2167	2167	1	22	SNP	40.90%	1.10E-08
C	2173	2173	1	21	SNP	66.70%	2.80E-29
T	2173	2173	1	21	SNP	33.30%	4.30E-11
C	2189	2189	1	21	SNP	42.90%	1.40E-19
T	2189	2189	1	21	SNP	57.10%	1.70E-20
A	2203	2203	1	24	SNP	62.50%	4.10E-38
T	2203	2203	1	24	SNP	37.50%	1.60E-22
C	2225	2225	1	26	SNP	65.40%	1.90E-38
T	2225	2225	1	26	SNP	34.60%	2.70E-12
A	2227	2227	1	26	SNP	30.80%	2.20E-11
T	2227	2227	1	26	SNP	69.20%	6.10E-45
C	2246	2246	1	23	SNP	60.90%	1.30E-39
T	2246	2246	1	23	SNP	39.10%	3.20E-18
C	2329	2329	1	33	SNP	66.70%	7.50E-52
T	2329	2329	1	33	SNP	33.30%	2.40E-24
C	2353	2353	1	36	SNP	66.70%	1.90E-56
T	2353	2353	1	36	SNP	33.30%	7.50E-23
C	2399	2399	1	36	SNP	69.40%	5.90E-57
T	2399	2399	1	36	SNP	27.80%	2.20E-14
C	2419	2419	1	36	SNP	72.20%	1.50E-49
T	2419	2419	1	36	SNP	27.80%	1.40E-08
C	2531	2531	1	33	SNP	15.20%	7.20E-09
T	2531	2531	1	33	SNP	84.80%	5.80E-57
C	2537	2537	1	31	SNP	22.60%	6.40E-14
G	2537	2537	1	31	SNP	77.40%	1.60E-61
C	2542	2542	1	33	SNP	18.20%	1.10E-12
T	2542	2542	1	33	SNP	81.80%	2.80E-70
C	2560	2560	1	30	SNP	16.70%	0.0000038
T	2560	2560	1	30	SNP	83.30%	4.40E-53
C	2609	2609	1	36	SNP	44.40%	1.10E-35
T	2609	2609	1	36	SNP	55.60%	7.10E-45
C	2686	2686	1	28	SNP	60.70%	2.60E-39
T	2686	2686	1	28	SNP	39.30%	2.20E-14
A	2698	2698	1	26	SNP	69.20%	3.90E-43
G	2698	2698	1	26	SNP	30.80%	5.70E-13
A	2719	2719	1	27	SNP	66.70%	6.90E-32
T	2719	2719	1	27	SNP	33.30%	2.30E-18
C	2722	2722	1	27	SNP	33.30%	4.60E-21
T	2722	2722	1	27	SNP	66.70%	1.10E-33
A	2782	2782	1	20	SNP	50.00%	1.80E-23
C	2782	2782	1	20	SNP	50.00%	1.80E-21
C	2854	2854	1	14	SNP	57.10%	2.80E-13
T	2854	2854	1	14	SNP	35.70%	0.0000018
C	2875	2875	1	8	SNP	50.00%	6.80E-07
T	2875	2875	1	8	SNP	50.00%	0.0000017

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	2923	2923	1	10	SNP	40.00%	3.30E-09
T	2923	2923	1	10	SNP	60.00%	2.10E-13
C	2932	2932	1	9	SNP	55.60%	1.30E-13
T	2932	2932	1	9	SNP	44.40%	0.000003
C	2933	2933	1	9	SNP	44.40%	1.30E-10
T	2933	2933	1	9	SNP	55.60%	1.20E-10
C	2938	2938	1	10	SNP	40.00%	5.30E-12
T	2938	2938	1	10	SNP	60.00%	8.20E-13
C	2962	2962	1	11	SNP	54.50%	1.80E-15
T	2962	2962	1	11	SNP	45.50%	4.60E-12
C	2971	2971	1	11	SNP	54.50%	2.80E-11
T	2971	2971	1	11	SNP	45.50%	1.40E-10
C	3001	3001	1	12	SNP	50.00%	1.50E-14
T	3001	3001	1	12	SNP	41.70%	7.80E-12
A	3016	3016	1	10	SNP	40.00%	8.30E-09
G	3016	3016	1	10	SNP	60.00%	3.30E-15
C	3055	3055	1	13	SNP	61.50%	8.10E-21
T	3055	3055	1	13	SNP	38.50%	1.20E-07
C	3058	3058	1	13	SNP	38.50%	4.00E-11
T	3058	3058	1	13	SNP	61.50%	2.00E-22
A	3102	3102	1	13	SNP	76.90%	2.80E-25
G	3102	3102	1	13	SNP	23.10%	5.70E-07
C	3450	3450	1	11	SNP	27.30%	0.0000026
T	3450	3450	1	11	SNP	72.70%	1.50E-10
C	3552	3552	1	15	SNP	66.70%	3.00E-25
T	3552	3552	1	15	SNP	33.30%	9.40E-12
A	3567	3567	1	15	SNP	73.30%	6.70E-23
T	3567	3567	1	15	SNP	26.70%	2.10E-08
C	3600	3600	1	17	SNP	41.20%	8.70E-10
T	3600	3600	1	17	SNP	58.80%	1.90E-24
G	3635	3635	1	21	SNP	19.00%	0.0000014
T	3635	3635	1	21	SNP	81.00%	1.90E-39
A	3642	3642	1	19	SNP	73.70%	1.00E-17
G	3642	3642	1	19	SNP	21.10%	3.70E-07
C	3644	3644	1	19	SNP	78.90%	3.80E-36
G	3644	3644	1	19	SNP	15.80%	0.0000019
A	3649	3649	1	18	SNP	16.70%	8.10E-07
G	3649	3649	1	18	SNP	83.30%	8.10E-37
G	3651	3651	1	18	SNP	16.70%	0.0000063
T	3651	3651	1	18	SNP	83.30%	2.60E-35
A	3654	3654	1	18	SNP	77.80%	7.30E-24
T	3654	3654	1	18	SNP	16.70%	0.0000016
G	3655	3655	1	18	SNP	16.70%	4.10E-07
T	3655	3655	1	18	SNP	77.80%	7.30E-24
C	3656	3656	1	18	SNP	16.70%	4.10E-07
T	3656	3656	1	18	SNP	83.30%	2.50E-26

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
A	3657	3657	1	18	SNP	83.30%	2.40E-20
G	3657	3657	1	18	SNP	16.70%	8.10E-07
A	3658	3658	1	18	SNP	77.80%	7.60E-31
T	3658	3658	1	18	SNP	16.70%	8.10E-07
A	3659	3659	1	17	SNP	82.40%	1.00E-21
G	3659	3659	1	17	SNP	17.60%	0.0000027
G	3664	3664	1	17	SNP	17.60%	0.0000053
T	3664	3664	1	17	SNP	82.40%	4.20E-30
A	3666	3666	1	17	SNP	76.50%	1.20E-24
G	3666	3666	1	17	SNP	17.60%	3.40E-07
C	3667	3667	1	17	SNP	82.40%	1.70E-31
T	3667	3667	1	17	SNP	17.60%	2.10E-08
A	3669	3669	1	15	SNP	20.00%	2.30E-07
T	3669	3669	1	15	SNP	80.00%	1.10E-19
C	3670	3670	1	15	SNP	33.30%	2.90E-09
T	3670	3670	1	15	SNP	66.70%	3.00E-22
A	3671	3671	1	16	SNP	18.80%	0.0000044
T	3671	3671	1	16	SNP	81.30%	1.10E-22
A	3672	3672	1	16	SNP	75.00%	2.80E-26
G	3672	3672	1	16	SNP	18.80%	0.0000044
A	3678	3678	1	16	SNP	81.30%	4.40E-33
G	3678	3678	1	16	SNP	18.80%	0.0000011
A	3695	3695	1	13	SNP	23.10%	0.0000044
T	3695	3695	1	13	SNP	76.90%	2.80E-25
C	3786	3786	1	13	SNP	76.90%	2.90E-27
T	3786	3786	1	13	SNP	23.10%	0.0000011
C	4651	4651	1	11	SNP	54.50%	1.10E-13
T	4651	4651	1	11	SNP	45.50%	1.40E-08
C	4662	4662	1	11	SNP	45.50%	1.40E-09
T	4662	4662	1	11	SNP	45.50%	1.40E-07
C	4669	4669	1	10	SNP	60.00%	2.10E-16
T	4669	4669	1	10	SNP	40.00%	0.000002
C	4689	4689	1	10	SNP	40.00%	8.00E-07
T	4689	4689	1	10	SNP	60.00%	8.00E-10
C	4720	4720	1	9	SNP	55.60%	4.00E-14
T	4720	4720	1	9	SNP	44.40%	0.0000075
C	4725	4725	1	10	SNP	50.00%	7.60E-08
T	4725	4725	1	10	SNP	50.00%	2.50E-15
C	4737	4737	1	10	SNP	50.00%	2.50E-09
T	4737	4737	1	10	SNP	50.00%	0.0000023
C	4755	4755	1	10	SNP	40.00%	5.20E-08
T	4755	4755	1	10	SNP	60.00%	3.30E-15
C	4792	4792	1	9	SNP	44.40%	3.10E-10
T	4792	4792	1	9	SNP	55.60%	3.90E-11
C	4818	4818	1	9	SNP	44.40%	0.0000075
T	4818	4818	1	9	SNP	55.60%	3.90E-09

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	4902	4902	1	6	SNP	50.00%	0.000005
T	4902	4902	1	6	SNP	50.00%	3.20E-07
A	4951	4951	1	6	SNP	33.30%	9.50E-07
C	4951	4951	1	6	SNP	66.70%	9.40E-09
C	4952	4952	1	6	SNP	33.30%	0.0000094
T	4952	4952	1	6	SNP	66.70%	1.50E-11
A	5006	5006	1	6	SNP	50.00%	4.00E-08
T	5006	5006	1	6	SNP	50.00%	1.00E-08
C	5067	5067	1	8	SNP	62.50%	1.80E-12
T	5067	5067	1	8	SNP	37.50%	0.0000069
C	5143	5143	1	10	SNP	50.00%	2.50E-09
T	5143	5143	1	10	SNP	50.00%	7.90E-11
A	5158	5158	1	11	SNP	54.50%	2.90E-14
T	5158	5158	1	11	SNP	45.50%	1.50E-12
C	5167	5167	1	12	SNP	58.30%	7.90E-19
T	5167	5167	1	12	SNP	41.70%	2.50E-12
A	5215	5215	1	9	SNP	44.40%	7.80E-08
T	5215	5215	1	9	SNP	55.60%	1.20E-07
A	5217	5217	1	9	SNP	55.60%	3.90E-09
G	5217	5217	1	9	SNP	44.40%	0.0000012
C	5224	5224	1	9	SNP	55.60%	1.20E-11
T	5224	5224	1	9	SNP	44.40%	3.10E-08
C	5233	5233	1	9	SNP	55.60%	1.30E-12
T	5233	5233	1	9	SNP	44.40%	2.00E-09
A	5260	5260	1	7	SNP	57.10%	2.20E-08
T	5260	5260	1	7	SNP	42.90%	1.40E-07
A	5266	5266	1	7	SNP	57.10%	1.40E-07
G	5266	5266	1	7	SNP	42.90%	0.0000043
C	5287	5287	1	9	SNP	44.40%	2.00E-11
T	5287	5287	1	9	SNP	55.60%	3.90E-10
C	5434	5434	1	6	SNP	50.00%	1.60E-07
T	5434	5434	1	6	SNP	50.00%	0.0000013
C	5467	5467	1	9	SNP	33.30%	3.30E-07
T	5467	5467	1	9	SNP	55.60%	4.00E-13
C	5491	5491	1	10	SNP	50.00%	2.50E-11
T	5491	5491	1	10	SNP	30.00%	4.70E-07
C	5518	5518	1	11	SNP	54.50%	2.70E-08
T	5518	5518	1	11	SNP	45.50%	1.40E-10
A	5540	5540	1	11	SNP	54.50%	7.20E-12
T	5540	5540	1	11	SNP	45.50%	1.40E-10
C	5557	5557	1	10	SNP	40.00%	5.20E-08
T	5557	5557	1	10	SNP	60.00%	5.20E-14
C	5563	5563	1	9	SNP	55.60%	1.30E-12
T	5563	5563	1	9	SNP	44.40%	7.90E-10
C	5569	5569	1	9	SNP	55.60%	4.00E-12
T	5569	5569	1	9	SNP	44.40%	1.20E-08

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	7305	7305	1	7	SNP	42.90%	0.0000043
T	7305	7305	1	7	SNP	57.10%	2.20E-10
A	7368	7368	1	7	SNP	71.40%	6.60E-11
T	7368	7368	1	7	SNP	28.60%	0.0000033
C	7373	7373	1	8	SNP	37.50%	2.20E-07
T	7373	7373	1	8	SNP	37.50%	0.0000017
A	7380	7380	1	8	SNP	62.50%	1.80E-12
T	7380	7380	1	8	SNP	37.50%	7.00E-09
C	7407	7407	1	10	SNP	50.00%	7.90E-13
T	7407	7407	1	10	SNP	50.00%	2.40E-07
C	7443	7443	1	7	SNP	57.10%	8.80E-11
T	7443	7443	1	7	SNP	42.90%	5.50E-07
C	7862	7862	1	7	SNP	42.90%	0.0000043
T	7862	7862	1	7	SNP	57.10%	2.20E-08
C	7874	7874	1	7	SNP	42.90%	1.40E-07
T	7874	7874	1	7	SNP	57.10%	5.50E-12
A	7904	7904	1	7	SNP	57.10%	3.50E-09
T	7904	7904	1	7	SNP	42.90%	2.80E-07
A	7910	7910	1	7	SNP	42.90%	2.80E-07
T	7910	7910	1	7	SNP	57.10%	1.40E-09
A	7925	7925	1	7	SNP	57.10%	0.0000053
T	7925	7925	1	7	SNP	42.90%	7.00E-08
C	8584	8584	1	9	SNP	66.70%	8.30E-14
T	8584	8584	1	9	SNP	33.30%	0.0000052
A	8599	8599	1	12	SNP	41.70%	0.0000022
T	8599	8599	1	12	SNP	50.00%	3.70E-15
A	8683	8683	1	14	SNP	50.00%	1.90E-08
T	8683	8683	1	14	SNP	50.00%	4.30E-16
C	8716	8716	1	11	SNP	45.50%	4.50E-09
T	8716	8716	1	11	SNP	54.50%	1.00E-07
C	8740	8740	1	13	SNP	46.20%	4.20E-13
T	8740	8740	1	13	SNP	53.80%	1.90E-09
C	8744	8744	1	13	SNP	46.20%	1.10E-13
T	8744	8744	1	13	SNP	53.80%	1.30E-13
C	8746	8746	1	13	SNP	46.20%	4.20E-13
T	8746	8746	1	13	SNP	53.80%	1.30E-13
A	8758	8758	1	13	SNP	46.20%	1.70E-12
C	8758	8758	1	13	SNP	53.80%	9.50E-09
A	8761	8761	1	13	SNP	53.80%	1.10E-15
T	8761	8761	1	13	SNP	38.50%	1.30E-09
C	8770	8770	1	14	SNP	50.00%	1.30E-12
T	8770	8770	1	14	SNP	50.00%	2.60E-13
C	8776	8776	1	14	SNP	42.90%	6.30E-07
T	8776	8776	1	14	SNP	57.10%	1.20E-15
C	8803	8803	1	18	SNP	50.00%	7.30E-16
T	8803	8803	1	18	SNP	44.40%	1.00E-13

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	8857	8857	1	19	SNP	47.40%	2.90E-10
T	8857	8857	1	19	SNP	47.40%	7.30E-22
A	8878	8878	1	20	SNP	40.00%	4.20E-10
C	8878	8878	1	20	SNP	50.00%	1.80E-19
A	8890	8890	1	19	SNP	47.40%	3.60E-19
T	8890	8890	1	19	SNP	52.60%	8.80E-18
A	8891	8891	1	19	SNP	52.60%	9.00E-22
G	8891	8891	1	19	SNP	47.40%	3.60E-19
A	8892	8892	1	19	SNP	47.40%	3.60E-19
T	8892	8892	1	19	SNP	52.60%	9.00E-22
C	8893	8893	1	19	SNP	52.60%	8.90E-20
T	8893	8893	1	19	SNP	47.40%	2.20E-17
A	8941	8941	1	17	SNP	47.10%	9.60E-19
T	8941	8941	1	17	SNP	47.10%	5.20E-10
C	8950	8950	1	16	SNP	50.00%	4.90E-15
T	8950	8950	1	16	SNP	50.00%	7.40E-12
C	9001	9001	1	10	SNP	40.00%	2.10E-10
T	9001	9001	1	10	SNP	60.00%	3.30E-15
C	9013	9013	1	10	SNP	40.00%	0.000005
T	9013	9013	1	10	SNP	60.00%	3.30E-15
C	9033	9033	1	8	SNP	25.00%	0.0000028
T	9033	9033	1	8	SNP	75.00%	6.70E-09
A	9043	9043	1	8	SNP	62.50%	1.70E-08
G	9043	9043	1	8	SNP	37.50%	0.0000035
C	10729	10729	1	6	SNP	33.30%	0.0000038
T	10729	10729	1	6	SNP	66.70%	5.90E-08
C	10798	10798	1	13	SNP	38.50%	4.00E-12
T	10798	10798	1	13	SNP	61.50%	1.80E-10
C	10801	10801	1	13	SNP	38.50%	1.30E-11
T	10801	10801	1	13	SNP	61.50%	8.00E-17
A	10805	10805	1	13	SNP	46.20%	4.10E-10
G	10805	10805	1	13	SNP	53.80%	3.30E-12
C	10846	10846	1	15	SNP	53.30%	1.60E-14
T	10846	10846	1	15	SNP	46.70%	6.00E-11
C	10850	10850	1	15	SNP	66.70%	3.00E-22
T	10850	10850	1	15	SNP	33.30%	8.90E-08
C	10864	10864	1	17	SNP	41.20%	1.20E-14
T	10864	10864	1	17	SNP	58.80%	1.90E-22
A	10880	10880	1	19	SNP	15.80%	0.0000019
T	10880	10880	1	19	SNP	84.20%	6.10E-41
A	10897	10897	1	19	SNP	47.40%	3.60E-19
T	10897	10897	1	19	SNP	52.60%	8.10E-14
C	10903	10903	1	19	SNP	52.60%	8.90E-20
T	10903	10903	1	19	SNP	47.40%	6.50E-13
A	10909	10909	1	19	SNP	47.40%	7.30E-22
T	10909	10909	1	19	SNP	52.60%	8.90E-19

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
A	10913	10913	1	19	SNP	47.40%	1.80E-16
T	10913	10913	1	19	SNP	52.60%	8.70E-17
A	10927	10927	1	19	SNP	47.40%	1.40E-15
C	10927	10927	1	19	SNP	52.60%	8.30E-15
A	10933	10933	1	19	SNP	52.60%	7.80E-13
G	10933	10933	1	19	SNP	42.10%	2.90E-14
C	10935	10935	1	19	SNP	47.40%	1.40E-15
T	10935	10935	1	19	SNP	52.60%	9.10E-23
A	10941	10941	1	19	SNP	47.40%	1.10E-14
G	10941	10941	1	19	SNP	52.60%	9.10E-25
C	10962	10962	1	15	SNP	60.00%	9.70E-18
T	10962	10962	1	15	SNP	40.00%	4.90E-12
A	10973	10973	1	14	SNP	42.90%	1.20E-11
T	10973	10973	1	14	SNP	57.10%	7.30E-15
A	10978	10978	1	14	SNP	42.90%	2.80E-09
T	10978	10978	1	14	SNP	57.10%	1.90E-16
A	11005	11005	1	14	SNP	50.00%	4.30E-16
T	11005	11005	1	14	SNP	42.90%	1.90E-16
C	11006	11006	1	14	SNP	50.00%	8.50E-17
T	11006	11006	1	14	SNP	50.00%	6.80E-19
C	11029	11029	1	14	SNP	64.30%	6.20E-20
T	11029	11029	1	14	SNP	35.70%	2.00E-10
C	11074	11074	1	11	SNP	36.40%	0.0000077
T	11074	11074	1	11	SNP	63.60%	3.20E-12
C	11075	11075	1	11	SNP	36.40%	0.0000077
T	11075	11075	1	11	SNP	54.50%	7.30E-15
C	11087	11087	1	9	SNP	33.30%	0.0000052
T	11087	11087	1	9	SNP	66.70%	8.20E-11
C	11101	11101	1	8	SNP	62.50%	1.80E-13
T	11101	11101	1	8	SNP	37.50%	2.80E-08
C	11113	11113	1	8	SNP	62.50%	5.60E-13
T	11113	11113	1	8	SNP	37.50%	5.60E-08
A	11159	11159	1	9	SNP	55.60%	1.20E-11
C	11159	11159	1	9	SNP	44.40%	5.00E-09
C	11198	11198	1	7	SNP	57.10%	3.50E-09
T	11198	11198	1	7	SNP	42.90%	0.0000086
C	11203	11203	1	7	SNP	42.90%	0.0000022
T	11203	11203	1	7	SNP	57.10%	1.40E-07
A	11296	11296	1	6	SNP	33.30%	0.0000094
T	11296	11296	1	6	SNP	66.70%	2.40E-08
A	11299	11299	1	6	SNP	66.70%	1.50E-09
C	11299	11299	1	6	SNP	33.30%	0.000006
A	11304	11304	1	6	SNP	33.30%	0.0000038
T	11304	11304	1	6	SNP	66.70%	2.40E-10
C	11305	11305	1	7	SNP	42.90%	2.20E-09
T	11305	11305	1	7	SNP	57.10%	1.40E-07

Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	11326	11326	1	8	SNP	62.50%	1.80E-13
T	11326	11326	1	8	SNP	37.50%	0.0000017
A	11327	11327	1	8	SNP	37.50%	8.80E-07
T	11327	11327	1	8	SNP	62.50%	1.70E-10
A	11359	11359	1	10	SNP	50.00%	7.70E-09
T	11359	11359	1	10	SNP	40.00%	3.20E-07
A	11360	11360	1	10	SNP	50.00%	7.70E-09
G	11360	11360	1	10	SNP	50.00%	7.90E-11
C	11365	11365	1	10	SNP	50.00%	2.40E-08
T	11365	11365	1	10	SNP	50.00%	2.50E-14
C	11383	11383	1	12	SNP	50.00%	3.70E-15
T	11383	11383	1	12	SNP	50.00%	2.30E-13
T	11432	11431	0	12	Indel	33.30%	4.90E-08
	11432	11431	0	12	Indel	66.70%	9.90E-08
C	11437	11437	1	13	SNP	46.20%	4.20E-13
T	11437	11437	1	13	SNP	53.80%	4.00E-10
C	11461	11461	1	13	SNP	46.20%	9.60E-08
T	11461	11461	1	13	SNP	53.80%	1.10E-15
A	11476	11476	1	13	SNP	46.20%	1.70E-12
C	11476	11476	1	13	SNP	53.80%	5.30E-15
C	11492	11492	1	14	SNP	57.10%	1.20E-19
T	11492	11492	1	14	SNP	42.90%	2.90E-12
C	11496	11496	1	14	SNP	42.90%	4.70E-14
T	11496	11496	1	14	SNP	57.10%	3.00E-21
A	11507	11507	1	13	SNP	53.80%	8.00E-11
G	11507	11507	1	13	SNP	46.20%	4.10E-10
C	11509	11509	1	13	SNP	46.20%	2.60E-11
T	11509	11509	1	13	SNP	53.80%	2.70E-14
C	11517	11517	1	13	SNP	46.20%	1.10E-13
T	11517	11517	1	13	SNP	53.80%	8.50E-18
C	11518	11518	1	13	SNP	53.80%	1.70E-18
T	11518	11518	1	13	SNP	46.20%	6.30E-09
A	11533	11533	1	13	SNP	46.20%	9.60E-08
C	11533	11533	1	13	SNP	53.80%	1.60E-11
C	11572	11572	1	13	SNP	46.20%	2.60E-11
T	11572	11572	1	13	SNP	53.80%	5.30E-15
C	11582	11582	1	13	SNP	46.20%	2.50E-08
T	11582	11582	1	13	SNP	53.80%	2.10E-16
C	11608	11608	1	11	SNP	45.50%	4.60E-15
T	11608	11608	1	11	SNP	54.50%	4.60E-16
A	11609	11609	1	11	SNP	45.50%	4.60E-13
T	11609	11609	1	11	SNP	45.50%	1.50E-13
C	11647	11647	1	12	SNP	41.70%	7.50E-08
T	11647	11647	1	12	SNP	50.00%	9.20E-16
A	11656	11656	1	12	SNP	50.00%	2.20E-10
C	11656	11656	1	12	SNP	50.00%	5.60E-11

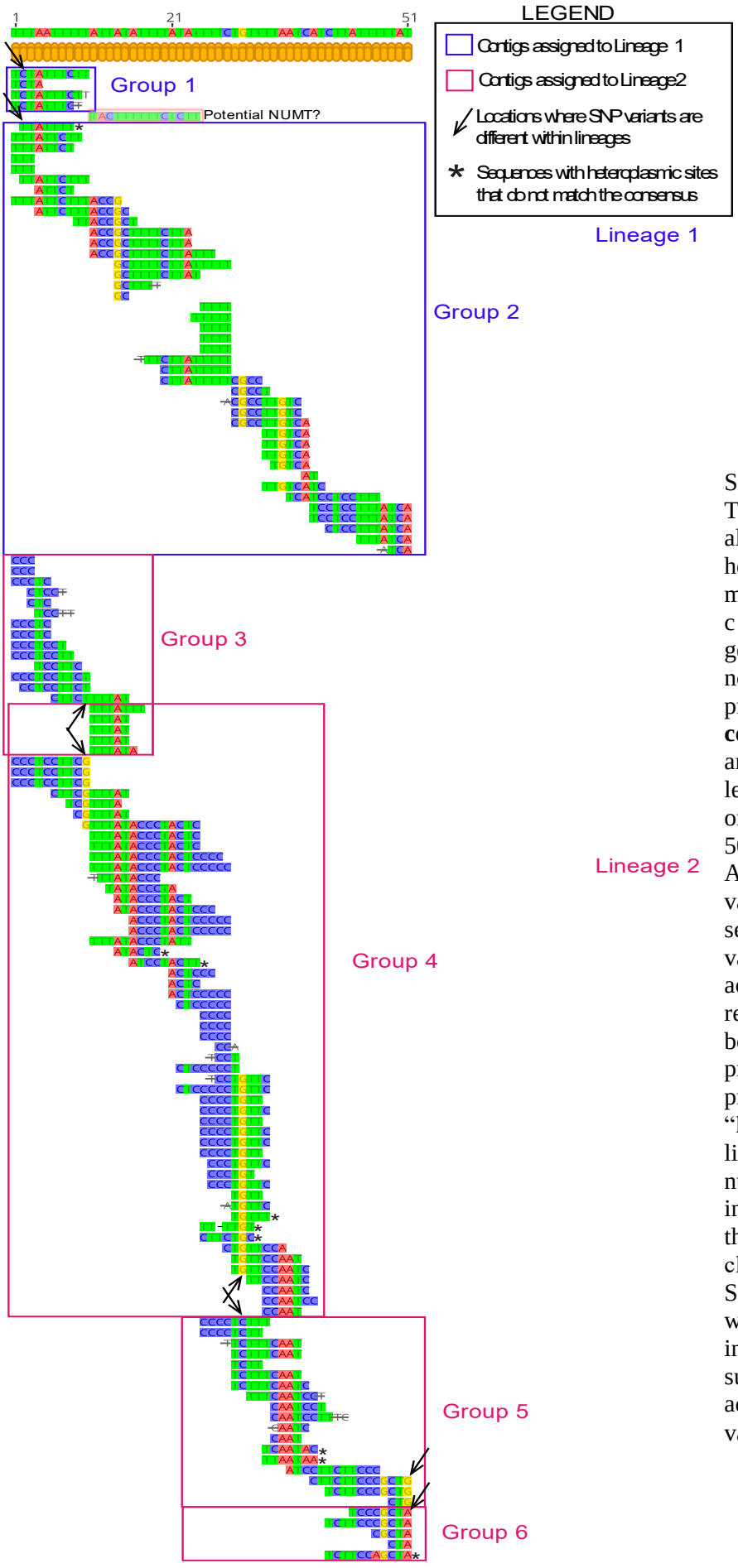
Name	Min.	Max.	Length	Coverage	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)
C	11660	11660	1	12	SNP	50.00%	2.30E-13
T	11660	11660	1	12	SNP	50.00%	5.60E-11
C	11674	11674	1	10	SNP	60.00%	3.30E-12
T	11674	11674	1	10	SNP	40.00%	0.000002
A	11683	11683	1	10	SNP	60.00%	1.30E-14
T	11683	11683	1	10	SNP	40.00%	8.30E-09
C	11699	11699	1	8	SNP	75.00%	1.70E-12
T	11699	11699	1	8	SNP	25.00%	0.0000011
A	11932	11932	1	6	SNP	66.70%	5.90E-08
G	11932	11932	1	6	SNP	33.30%	0.0000038
C	12049	12049	1	9	SNP	33.30%	0.0000013
T	12049	12049	1	9	SNP	66.70%	2.00E-08
A	12127	12127	1	9	SNP	44.40%	5.00E-11
C	12127	12127	1	9	SNP	55.60%	3.90E-09
C	12130	12130	1	9	SNP	66.70%	2.00E-08
T	12130	12130	1	9	SNP	33.30%	0.0000013
A	12151	12151	1	9	SNP	44.40%	2.00E-09
T	12151	12151	1	9	SNP	33.30%	0.0000013
A	12325	12325	1	6	SNP	50.00%	5.00E-09
G	12325	12325	1	6	SNP	50.00%	1.30E-09
C	13292	13292	1	9	SNP	44.40%	7.90E-10
T	13292	13292	1	9	SNP	55.60%	1.20E-07
A	13335	13335	1	7	SNP	42.90%	0.0000011
T	13335	13335	1	7	SNP	57.10%	3.50E-09

Supp. Table 5: The number of contigs recovered for each of the Ion Torrent mitochondrial gene isolates examined. Total length of aligned contigs were reduced to only show suspected SNPs (variable nucleotide sites indicating heteroplasmy). The number of SNPs detected by the Geneious "SNP/Variation" function are also reported. By isolating only the SNP's to sort contigs into lineages, the gene fragments were reduced by 90–97% of their original length. Note: once the lineages were established, the full length of the gene isolate were restored to assess for coding changings associated with variable sites.

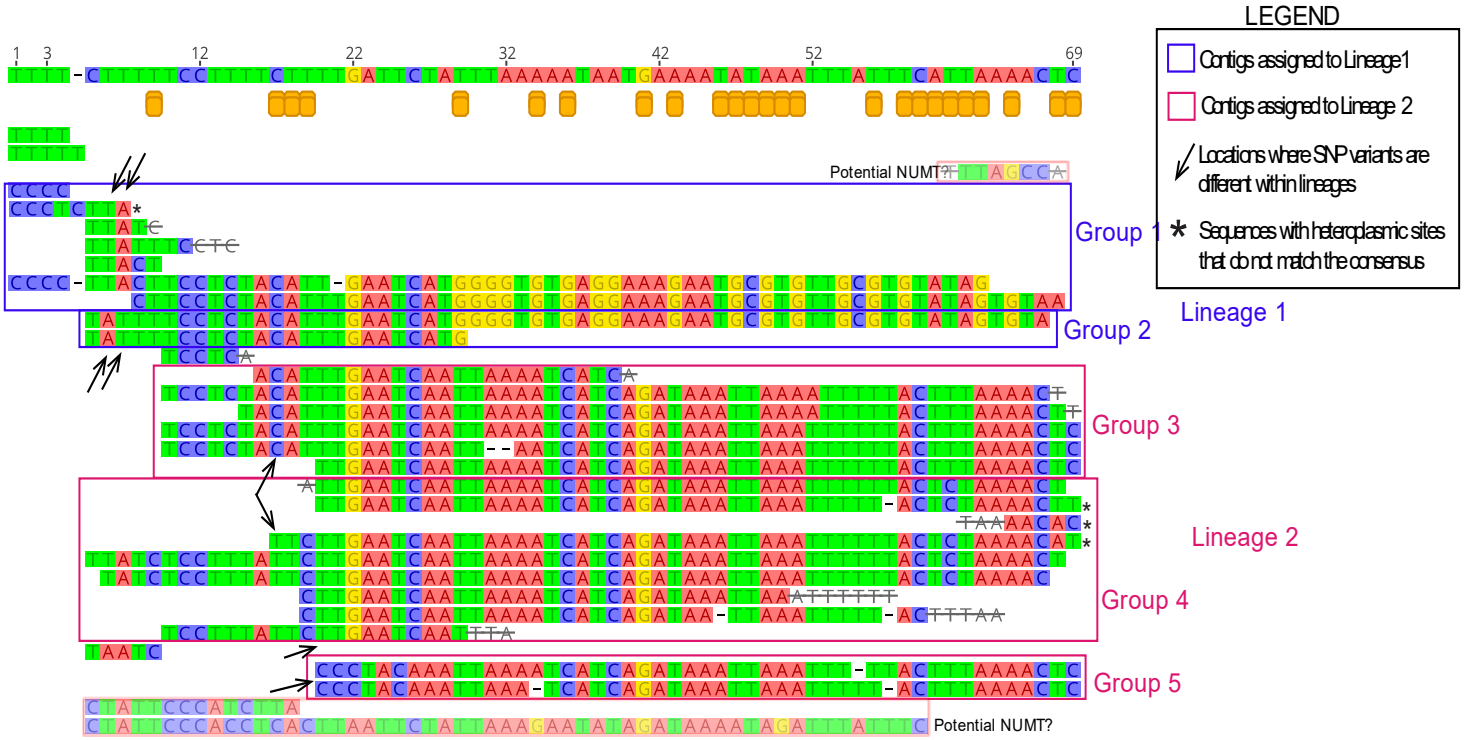
Gene	**No. contigs**	**Max. contig depth**	**Total fragment length**	**All suspected SNPs**	**Geneious SNPs only**	**Fragment length reduction**
COI	150	38	1,610	51	51	97%
COII	34	21	668	69	25	90%
cytb	52	19	1,224	81	58	93%
ND4	32	20	572	27	24	95%

Supp. Table 6: Summary of heteroplasmic sites detected in mitochondrial cytochrome *c* oxidase subunit 1 (COI) gene fragment using both techniques: Sanger sequencing and Ion Torrent shotgun sequencing. Base pair (bp) positions are in reference to the combined Sangers sequences (forward + reverse sequence) which was 658 bp. Heteroplasmic site detected at 310 bp was identified in a single individual (AM6) from the Blue Mountains, NSW. The heteroplasmic site at 624 bp was observed in the Sanger sequence chromatograms, but was observed as within lineage variation (minor variation) within the Ion Torrent shotgun data.

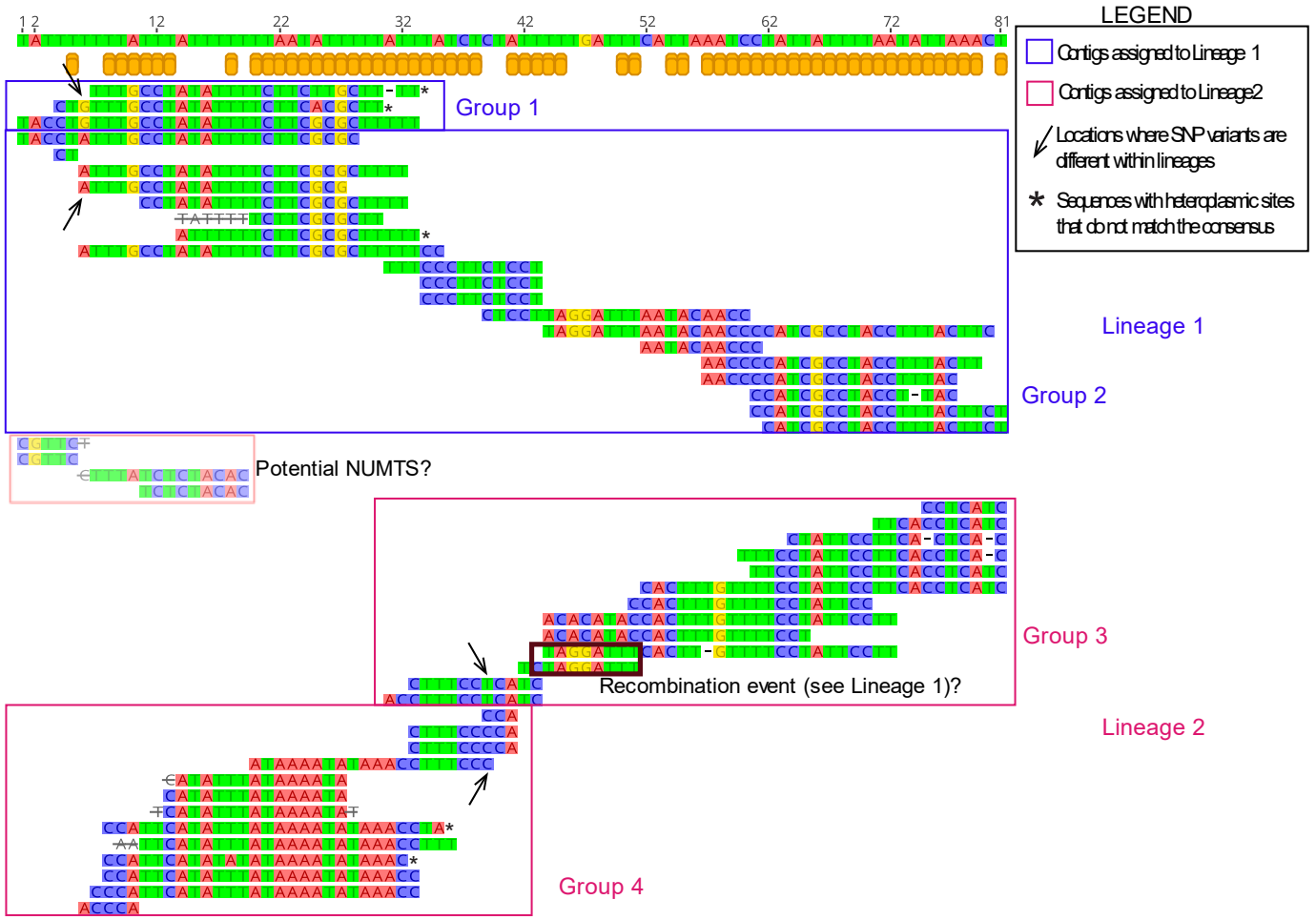
Base Pair (bp)	**Sanger**	**Ion Torrent**	**Allele 1**	**Allele 2**	**Non-synonymous**	**Amino Acid**
10	✓	✓	C	T	×	Y
79	✓	✓	T	A	×	S
88	✓	✓	C	T	×	N
103	×	✓	T	C	×	N
115	×	✓	T	C	×	Y
148	✓	✓	A	C	✓	**M-I**
153	✓	✓	C	T	✓	**S-F**
157	×	✓	C	T	×	F
163	×	✓	T	C	×	V
187	✓	✓	C	T	×	F
198	✓	✓	G	T	✓	**W-L**
310	*AM6 (NSW) only*	×	C	T	×	Y
313	✓	✓	T	A	×	T
316	✓	✓	T	C	×	G
340	✓	✓	T	C	×	Y
379	×	✓	G	A	×	S
406	✓	✓	T	C	×	H
451	✓	✓	A	T	×	V
458	✓	✓	C	T	×	L
478	✓	✓	C	T	×	N
484	✓	✓	C	T	×	N
500	✓	✓	T	C	×	L
514	✓	✓	A	T	×	A
529	✓	✓	C	T	×	A
536	✓	✓	C	T	×	L
538	✓	✓	T	A	×	L
557	✓	✓	C	T	×	L
573	✓	×	C	G	×	T
624	✓	~ (minor)	G	A	✓	**G-D**
640	✓	✓	C	T	×	I
TOTAL	24	28			4	



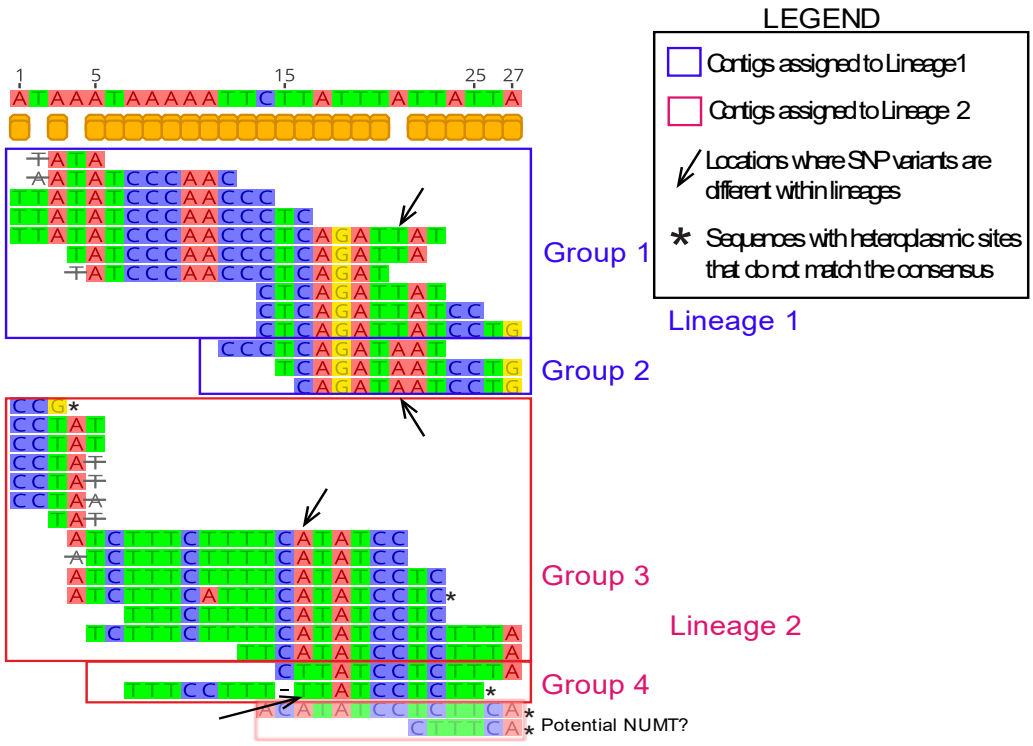
Supp. Figure 1: Ion Torrent shotgun sequence alignment of suspected heteroplasmic sites in the mitochondrial cytochrome c oxidase subunit I (COI) gene. It is important to note that the contigs presented here **only consist of variable sites** and have been reduced in length by ~90% of their original length (~150 – 500 bp per contig). Almost all of these variable sites are separated by long, non-variable regions shared across the dataset. These reduced-contigs have been sorted into analogues producing the two proposed heteroplasmic "lineages". Within lineages, variable nucleotide sites have been indicated by an arrow and these contigs were classified into "groups". Some reduced-contigs were not able to be sorted into any lineage and are suspected to be NUMTs, additional mtDNA variants, or errors.



Supp. Figure 2: Ion Torrent shotgun sequence alignment of suspected heteroplasmic sites in the mitochondrial cytochrome *c* oxidase subunit II (COII) gene. It is important to note that the contigs presented here **only consist of variable sites** and have been reduced in length by ~90% of their original length (~150 – 500 bp per contig). Almost all of these variable sites are separated by long, non-variable regions shared across the dataset. These reduced-contigs have been sorted into analogues producing the two proposed heteroplasmic "lineages". Within lineages, variable nucleotide sites have been indicated by an arrow and these contigs were classified into "groups". Some reduced-contigs were not able to be sorted into any lineage and are suspected to be NUMTs, additional mtDNA variants, or errors.



Supp. Figure 3: Ion Torrent shotgun sequence alignment of suspected heteroplasmic sites in the mitochondrial cytochrome b (cytb) gene. It is important to note that the contigs presented here **only consist of variable sites** and have been reduced in length by ~90% of their original length (~150 – 500 bp per contig). Almost all of these variable sites are separated by long, non-variable regions shared across the dataset. These reduced-contigs have been sorted into analogues producing the two proposed heteroplasmic "lineages". Within lineages, variable nucleotide sites have been indicated by an arrow and these contigs were classified into "groups". Some reduced-contigs were not able to be sorted into any lineage and are suspected to be NUMTs, additional mtDNA variants, or errors.



Supp. Figure 4: Ion Torrent shotgun sequence alignment of suspected heteroplasmic sites in the mitochondrial NADH-ubiquinone oxidoreductase chain 4 (ND4) gene. It is important to note that the contigs presented here **only consist of variable sites** and have been reduced in length by ~90% of their original length (~150 – 500 bp per contig). Almost all of these variable sites are separated by long, non-variable regions shared across the dataset. These reduced-contigs have been sorted into analogues producing the two proposed heteroplasmic "lineages". Within lineages, variable nucleotide sites have been indicated by an arrow and these contigs were classified into "groups". Some reduced-contigs were not able to be sorted into any lineage and are suspected to be NUMTs, additional mtDNA variants, or errors.

Chapter 3: Co-infection of *Wolbachia* in the Australian bee *Amphylaeus morosus* (Colletidae: Hylaeinae)

Olivia K. Davies[a*], James B. Dorey[a,b], Mark I. Stevens[b,c], and Michael P. Schwarz[a]

[a]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

[b]Biological and Earth Sciences, South Australian Museum, SA 5000, Australia

[c]School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia

*Corresponding author. Email: davies.oliviak@gmail.com

Keywords: Hymenoptera, population-wide, endosymbiont, mitochondrial heteroplasmy, selective sweep

3.1 Abstract

The α-proteobacteria genus *Wolbachia* (Rickettsiales: Anaplasmataceae) is known to exclusively infect invertebrates where it can present itself as either a reproductive parasite or a mutualist. It is the most common maternally inherited intracellular endosymbiont in the world, infecting as many as 60% of insect species. At present, little is known about the infection pathways of this bacterium, nor the full breadth of impacts it can have on its hosts. *Amphylaeus morosus* (Smith, 1879) is a hylaeine bee (Hymenoptera: Colletidae) with an extensive geographical range along the eastern coast of Australia. This bee species is known to have extremely high rates (100%) of conserved mitochondrial heteroplasmy (two or more types of mitogenome within a single individual) across its distribution, involving only two consistent mitochondrial haplotypes. Using DNA sequencing and phylogenetic techniques, we assess (i) the distribution a *Wolbachia* co-infection across this host's distribution and (ii) the phylogenetic relationship between the two *Wolbachia* strains infecting this bee host. DNA sequence data demonstrated a consistent co-infection of *Wolbachia* across the *A. morosus* ~2,000 km distribution with every individual infected with both bacterial strains. Furthermore, phylogenetic analyses indicate that these two *Wolbachia* strains did not diverge from each other within this host. Date estimates from phylogenetic data provide some evidence of two independent introductions (divergence times of each *Wolbachia* within this host are approximately 2 mya compared to 0.5 mya. As the host's mitogenome and the *Wolbachia* genome are genetically linked through maternal inheritance, we propose three possible hypotheses that could explain maintenance of the widespread and conserved co-occurring bacterial and mitochondrial genomes in this species.

3.2 Introduction

Obligate intracellular bacteria are widespread within invertebrates (Goebel and Gross 2001) and can function as pathogens or as mutualists. Although they are phylogenetically divergent, these endosymbionts often share common genomic similarities, such as AT-biased compositions, small genomes, and high rates of evolution (Moran and Baumann 2000, Moran and Wernegreen 2000). These traits generally result in specialized lifestyles within any particular bacterial genus (Moran and Wernegreen 2000), where endosymbionts become highly obligate and are strictly pathogenic or mutualistic within their hosts (Gil et al. 2004). The α-proteobacteria genus *Wolbachia* (Rickettsiales: Anaplasmataceae) is an unusual exception (Gil et al. 2004), as it seems to have repeatedly horizontally transferred into new host lineages (Heath et al. 1999, Werren et al. 2008) and presents itself as either parasitic or an obligate mutualist across numerous host taxa (e.g. O'Neill et al. (1997), Bandi et al. (1999)).

Within most insect lineages, *Wolbachia* are primarily maternally inherited reproductive parasites. Notably, this infection can affect the sexual differentiation of its host. The transmission of the endosymbiont through the population is promoted by biasing for infected female offspring relative to uninfected individuals (Werren 1997, Werren et al. 2008). This is achieved via four known phenotypic alterations — cytoplasmic incompatibility, feminization of genetic males, thelytokous parthenogenesis induction, and male killing — although the mechanisms behind these effects are poorly understood (Werren 1997, Stouthamer et al. 1999). These reproductive manipulation strategies enable *Wolbachia* to maximize its inheritance rate and largely depend on (i) the host's sex determination system, (ii) the *Wolbachia* strain, and (iii) the interactions between the host and bacterium (Charlat et al. 2003).

Wolbachia has been documented to reach infection fixation within laboratory populations (e.g. Xi et al. (2005), Bian et al. (2013)). However, the proportion of infection within host populations can become stable without reaching complete fixation (e.g. Shoemaker et al. (2003), Unckless et al. (2009)). Unfortunately, infection prevalence has not been widely assessed across the distribution of wild hosts. Additionally, it is unknown how reproductive phenotypic manipulations are adjusted once *Wolbachia* completely infects a population (e.g. whether phenotypic manipulation ceases once *Wolbachia* can no longer infect *Wolbachia*-free individuals). Phenotypic changes expressed by hosts can be quite variable, even within

closely-related taxa (e.g. Fujii et al. (2001), Sasaki et al. (2002), Sasaki et al. (2005), Veneti et al. (2012)) and some *Wolbachia* strains might have more modification potential than others (Werren et al. 2008, Hamm et al. 2014). Hence, despite its prevalence in insects, the interactions of *Wolbachia* and its hosts are poorly understood (Werren et al. 2008).

The transmission of *Wolbachia* within hosts can lead to rapid selection on host mitogenomes. These 'selective mitochondrial sweeps' occur when *Wolbachia*-infected host lineages replace non-infected linages and cause bacterially-facilitated dominance of the co-inherited mitogenome in the host population (Jiggins 2003). The mitogenomes can 'hitchhike' their way to fixation resulting in a host population with low mitochondrial haplotype variation (e.g. Ballad et al. (1996), Graham and Wilson (2012), Schuler et al. (2016)). Because mitochondria and *Wolbachia* are genetically linked through maternal transmission (Cariou et al. 2017), mitochondrial selective sweeps can occur when there are mechanisms that selectively favour *Wolbachia* (e.g. reproductive phenotypic manipulation) (Jiggins 2003, Hurst and Jiggins 2005). Consequently, these mitochondrial selective sweeps are a major hurdle for studies of host demography that use only mitochondrial markers (Jiggins 2003, Charlat et al. 2009, Toews and Brelsford 2012). Unlike mitochondrial DNA (mtDNA), nuclear loci are expected to largely follow Mendelian inheritance patterns despite any *Wolbachia* infections (unless *Wolbachia* induces parthenogenesis, deviating from sexual reproduction (Rokas et al. 2001)).

Wolbachia can provide diverse benefits and costs to its hosts, but these effects depend on the taxonomic lineage, sex, and life stage of the host, as well as environmental stressors (e.g. Gavotte et al. (2010), White et al. (2011), Joshi et al. (2014)). *Wolbachia*-induced effects are variable, and infections can persist in host populations with or without apparent reproductive alterations to the host (e.g. Gomi et al. (1997), Hoffmann et al. (1998), Harcombe and Hoffmann (2004), Hedges et al. (2008)). These effects include either a reduction (Fleury et al. 2001) or increase in the fecundity of infected females (Dobson et al. 2002), providing nutritional benefits (Brownlie et al. 2009, Nikoh et al. 2014), antiviral protection (Hedges et al. 2008, Hoffmann et al. 2015, Dutra et al. 2016), and can reduce defences against other parasites or parasitoids (Fytrou et al. 2006). However, there is limited understanding of how these alterations (independent of reproductive modifications) affect the transmission rate of the endosymbiont or its persistence in host lineages (Werren et al. 2008, Zug and Hammerstein 2015). Assessing the effects of *Wolbachia* is further complicated in situations

where hosts are infected with multiple endosymbionts. Co-infections are reasonably common (Duron et al. 2008), but our understanding of the impacts and origins of multiple infections is limited (White et al. 2011).

In this study, we aimed to assess (i) the extent of the co-infection of two *Wolbachia* strains in the hylaeine bee, *Amphylaeus morosus* (Smith, 1879) (Hymenoptera: Colletidae) throughout its range and (ii) the divergence between these two *Wolbachia* strains from their geographical extremities (north and south). *Amphylaeus morosus* occupies warm-temperate/subtropical subcoastal heathland and cool temperate montane forests along the eastern coast of Australia (Houston 1975). Our previous research identified that this bee species has widespread and conserved mitochondrial heteroplasmy (Chapter 2) where every individual across the distribution appears to have the same two mitochondrial haplotypes (with minor nucleotide variation detected in some individuals). Here we explore the extensive and consistent co-infection of two *Wolbachia* strains infecting *A. morosus* and discuss the possible effects of this pervasive endosymbiont on this host.

3.3 Methods

3.3.1 Specimen sampling

Adult female specimens of *Amphylaeus morosus* were collected in nesting substrate across five regions of eastern Australia, from southern Queensland to southwestern Victoria (Table 1). Specimen collections (Table 1) from the northern-most (Tin Can Bay, QLD) and southwestern-most (Cobboboonee, VIC) locations occurred in subcoastal heathland habitat where colonies nest in dead flower scapes of grass tree *Xanthorrhoea* spp. (Asparagales: Asphodelaceae). Collections from all central localities (Enfield State Forest and Blue Mountains, NSW, and Dandenong Ranges, VIC; Table 1) were collected in wet montane forests from fallen excised fronds of the tree fern *Cyathea australis* (R.Br.) Domin (Cyatheales: Cyatheaceae). To obtain adult male specimens for these analyses, immature individuals collected from the Dandenong Ranges in 2016 were reared to adulthood in controlled conditions at the Flinders University Bedford Park campus, South Australia.

3.3.2 *Wolbachia* screening across the host's distribution

Wolbachia screening was performed on genomic DNA extracts (leg tissue) previously used to assess mitochondrial heteroplasmy (Chapter 2; i.e. all specimens analysed for *Wolbachia* infection had also been assessed for mitochondrial heteroplasmy). Twenty-seven individuals

(24 females and 3 males) were selected from five regions across the *A. morosus* distribution for assessment of the co-infection of *Wolbachia* using the COI-like *Wolbachia* gene region. This DNA region is highly conserved i.e. fragment sizes were consistent between different bacterial strains. This gene region was reliably recovered using common arthropod mitochondrial (mt-) COI primer sets (a common problem with arthropod mt primers (Smith et al. 2012)), originally intended to recover *A. morosus* mtDNA. Polymerase chain reaction (PCR) amplification protocols for the COI-like region of *Wolbachia* were performed using the arthropod (Lepidoptera) primer set Lep-F1 5'-ATTCAACCAATCATAAAGATAT-3' and Lep-R1 5'-TAAACTTCTGGATGTCCAAAAA-3' (Hebert et al. 2004) for most specimens. The COI-like *Wolbachia* region was also recovered using the universal primers LCO1490 5'-GGTCAACAAATCATAAAGATATTGG-3' and HCO2198 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' (Folmer et al. 1994) for five of the specimens, although it produced lower quality sequences.

PCR amplifications of the *Wolbachia* COI-like region were carried out in a total volume of 25 µL, as follows: 1x MRT Buffer ($MgCl_2$), primers (0.4 µM each), 1 U Immolase DNA Polymerase (Bioline 5 U/µL), and template DNA. PCR cycling conditions were one cycle (10 min at 95°C), 35 cycles (45 sec at 94°C, 45 sec at 48°C, 60 sec at 75°C) and one cycle (6 min at 72°C, 2 min at 25°C). PCR products were visualised using 1.5% agarose gel. Successful PCR products were purified using Multiscreen PRC384 Filter Plate (Millipore) and re-suspended in 20–25 µL of 10 mM TRIS. Purified amplicons were sent to the Australian Genome Research Facility (AGRF) in Adelaide, South Australia for Sanger sequencing with Applied Bio-Systems 3730 and 3730 xl capillary sequencers.

3.3.3 Cloning to separate *Wolbachia* strains

One individual each from the northern- and southern- most localities (Tin Can Bay and Dandenong Ranges) were selected for cloning. These specimens had previously been Sanger sequenced and their chromatograms had confirmed co-infections of two *Wolbachia* strains. For comparison with our other *Wolbachia* data, we produced clones for the *Wolbachia* COI-like region, as well as *Wolbachia* surface protein (*wsp*) gene region (see below). The COI-like region was recovered using the Lep primer set (Hebert et al. 2004) and PCR amplification conditions were repeated from the Sanger sequencing above. Four additional *Wolbachia* COI-like sequences were recovered from cloning products using the universal degenerate primer set COIF-PR115 5'-TCWACNAAYCAYAARGAYATTGG-3' and COIR-

PR114 5'-ACYTCNGGRTGNCCRAARARYCA-3' (Folmer et al. 1994) which was originally intended to recover bee mtCOI (cloning was not able to recover host mtDNA).

The *wsp* gene was also analysed as it is one of the fastest evolving *Wolbachia* genes, and can provide a good indication of phylogenetic relationships among closely related *Wolbachia* strains (Zhou et al. 1998). The primer set wsp-81F 5'-TGGTCCAATAAGTGATGAAGA-3' and wsp-691R 5'-AAAAATTAAACGCTACTCCA-3' (Braig et al. 1998, Zhou et al. 1998) was used and PCR amplification conditions were one cycle (3 min at 95°C), 35 cycles (60 sec at 95°C, 60 sec at 55°C, 60 sec at 72°C) and one cycle (30 min at 72°C) (Zhou et al. 1998). All reactions were carried out in PCR volumes as described above.

All amplified DNA was purified using Multiscreen PRC384 Filter Plate. PCR products 1 μL neat from the COI-like gene with Hebert et al. (2004) primers and 2 μL neat from the *wsp* gene with Zhou et al. (1998) primers were cloned into pGEM-T Easy (Promega) and blue/white screening following manufactures instructions. Transformants were harvested from Petri dishes and suspended in 25 mL of 10 mM TRIS and then heat treated to lyse cells. All colonies were amplified using the primer set T7 5'-TAATACGACTCACTATAGGG-3' and SP6 5'-TTCTATAGTGTCACCTAAAT-3' (Promega) and the following PCR conditions applied: one cycle (10 min at 95°C), 34 cycles (30 sec at 94°C, 30 sec at 60°C, 90 sec at 72°C) and one cycle (20 min at 72°C). Purified cloned amplicons were sequenced (Sanger) at AGRF.

3.3.4 Sequence analysis

To assess the extent of the two *Wolbachia* strains in this bee species, forward *Wolbachia* COI-like sequences and cloning products (COI-like and *wsp* genes) were edited and aligned using Geneious Prime version 2020.1 (https://www.geneious.com). To confirm the identity of the source species for each sequence, edited sequences were BLAST screened against the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/blast).

Both of our cloned bee specimens had two haplotypes for both non-contiguous *Wolbachia* gene regions (COI-like and *wsp*). To identify which of the two *Wolbachia* strains each of our non-contiguous gene fragments belonged to, BLAST results to published *Wolbachia* genomes were used. BLAST results of the *wsp* gene were also used to attempt to identify the

Wolbachia strain's possible reproductive manipulation phenotypes by looking for similar strains identified in other hosts.

For the *Wolbachia* COI-like and *wsp* genes recovered from *A. morosus* specimens, one of the two haplotypes from both gene regions was highly similar to two NCBI published *Wolbachia* genomes (>99% identify match; *w*CauA; accession PRJNA550963 and *w*Ha; accession PRJNA176303 (Ellegaard et al. 2013)). There was no match (BLAST) to any published whole *Wolbachia* genomes for the remaining haplotype of each gene fragment. Therefore, for our phylogenetic analyses, we grouped the first two haplotypes (one sequence from both COI-like and *wsp* genes) as one of the *Wolbachia* strains — which we labelled *w*AmHa — and grouped the remaining two haplotypes as the second *Wolbachia* strain — which we labelled *w*Amor. These *Wolbachia* gene fragments will be submitted to the NCBI database. A concatenated 1,271 bp nucleotide alignment was then created in Geneious Prime with best alignment inferred from amino acids. This alignment included 15 supergroup A *Wolbachia* strains comprised of the <665 bp COI-like and <606 bp *wsp* genes; four samples from our data — two *Wolbachia* strains from both the northern- and southern-most *A. morosus* populations — and 11 trimmed *Wolbachia* genomes used in Gerth and Bleidorn (2016) (Supp. Table 1).

3.3.5 Phylogenetic analysis of *Wolbachia* strains

To find the best partition schemes and DNA substitution models for these sequence data (nucleotide alignment of two *Wolbachia* genes; COI-like and *wsp*), PartitionFinder version 2.1.1 (Lanfear et al. 2017) was employed with an AICc (corrected Akaike information criterion) and a greedy algorithm from the available BEAST models (Guindon et al. 2010, Lanfear et al. 2012, Lanfear et al. 2017). All codon positions for both gene fragments were assigned individual partitions. The first and third COI-like codons as well as the second *wsp* codon positions were each assigned an HKY+I+X model. The second COI-like, first and third *wsp* codon positions were assigned HKY+X, HKY+Γ+X and TRN+I+X substitution models, respectively. BEAST run files, parameters and optimisations for phylogenetic analyses were set in BEAUTi version 2.6.2 (Bouckaert et al. 2019). All partitions were assigned a relaxed log normal clock model. Known node ages were assigned based on the dated phylogeny in Gerth and Bleidorn (2016) with either a normal or log normal distribution to best fit the 95% highest posterior densities (HPDs) of each node. An additional uniform distribution was set at the boundaries of these 95% HPDs.

Phylogenetic analyses were implemented in BEAST version 2.6.2 (Bouckaert et al. 2019) with 100 million iterations, sampled every 20,000th iteration. The resulting log files were analysed in Tracer version 1.7 (Rambaut et al. 2018) and a 10% burnin was employed, which was always after stationarity had been achieved. Maximum clade credibility trees were produced using TreeAnnotator version 2.6.2 (Bouckaert et al. 2019). This analysis was performed four times to check for consistent results and stationarity. All four post-burnin log and tree files were combined using LogCombiner version 2.6.2 (Bouckaert et al. 2019).

FigTree version 1.4.4 (Drummond 2016) was used to produce trees. Our recovered tree height was about half of that recovered in Gerth and Bleidorn (2016) (Supp. Figure 1). Gerth and Bleidorn (2016) used entire *Wolbachia* genomes and included *Wolbachia* supergroup B in their analyses, and likely recovered a more reliable tree height. To correct for this, we used FigTree to set the root height at 76 million years ago (mya) to be concordant with Gerth and Bleidorn (2016).

3.4 Results

3.4.1 Widespread co-occurring *Wolbachia* infection

Sequence evidence of widespread co-infection. Sequences were recovered for 27 *Amphylaeus morosus* specimens for the *Wolbachia* COI-like gene (amplicon lengths between 567 bp and 613 bp) which was clearly distinguished from host mitochondrial DNA based on BLAST results. All 27 specimens were observed to have a co-infection of *Wolbachia* as they showed clear double peaks at consistent nucleotide sites in their chromatograms (Figure 1) indicating a persistent co-infection of two *Wolbachia* strains across the host's distribution.

Cloning of the bacterial co-infection. Cloning data further supported that there are dual *Wolbachia* infections occurring in *A. morosus*, with both the Tin Can Bay (northern) and Dandenong Ranges (southern) specimens infected with the same two *Wolbachia* strains. When aligned to the COI-like Sanger sequences, double peaks in the chromatograms occurred at variable sites detected through cloning (e.g. Figure 1). Clones were successfully recovered for both the COI-like (with two primer sets) and *wsp* gene regions for both *A. morosus* specimens.

Within and between strain variation. **Between** the distinct *Wolbachia* strains, there were 13 variable sites in the conserved COI-like gene fragment. Of these variable nucleotides, three

coded for amino acid changes (using the bacterial protein translation scheme). No variation **within either** strain between the two geographical extremities was detected for the COI-like gene. One exception was detected: one COI-like cloned amplicon from the southern specimen did possess a single synonymous 'mutation' in the *Wolbachia* strain *w*AmHa. The other *w*AmHa amplicon for this specimen in this gene region was consistent with the equivalent amplicons of the northern specimen. Although possibly suggesting intra-individual *Wolbachia* variation within this southern specimen, it is likely that this represents a cloning, PCR or sequencing error (Eckert and Kunkel 1991, Clarke et al. 2001). Low sample sizes of cloned amplicons (Supp. Table 2) make it very difficult to distinguish between error and genuine intra-individual nucleotide variation. Therefore, the most conservative option was implemented for all analyses.

High base pair variation was seen in the more-variable *wsp* gene **between** the *Wolbachia* strains, with over 100 variable sites including multiple insertions and deletions. This indicated that although the *Wolbachia* strains are related, both being supergroup A *Wolbachia* strains (as designated by Werren et al. (1995b)), they are clearly divergent. Like the COI-like gene, the *wsp* gene indicated variation **within** each *Wolbachia* strain to be highly conserved across the two geographical locations. One synonymous and two non-synonymous substitutions were identified between the southern and northern localities in one *Wolbachia* strain (*w*Amor). In the other *Wolbachia* strain (*w*AmHa), a single non-synonymous substitution was identified between the specimens. As with the COI-like gene, 'intra-individual' variation was detected, but in both individuals. An additional non-synonymous base pair change was present in the *w*Amor strain from the northern specimen and another in the *w*AmHa strain from the southern specimen. However, as with the COI-like gene this is likely the results of error and the most conservative options were implemented for all analyses.

3.4.2 Identifying *Wolbachia* strains

Using the phylogenetically-informative *wsp* gene (Zhou et al. 1998), BLAST screening identified that one of the *Wolbachia* strains infecting *A. morosus* (*w*AmHa) is most similar to strains *w*CauA (completely identical for *wsp*) described from the moth *Cadra cautella* (Walker, 1863) (Lepidoptera: Pyralidae) (Sasaki and Ishikawa 1999), and *w*Ha (with only a single non-synonymous base change) described from *Drosophila simulans* Sturevant, 1919 (Diptera: Drosophilidae) (O'Neill and Karr 1990, Ellegaard et al. 2013). Interestingly, in both

the host moth *C. cautella* and in *D. simulans*, these strains often co-occur with other *Wolbachia* strains (Poinsot and Merçot 1997, Sasaki and Ishikawa 1999).

The second strain of *Wolbachia* infecting *A. morosus* (*w*Amor) could not be matched to a published bacterial genome on the NCBI BLAST database and has not yet been phylogenetically classified. BLAST results of the *wsp* gene closely matched (~99%) strains infecting several arthropods including Hymenoptera (parasitic wasps (Yang et al. 2012, Mohammed et al. 2017)) and Lepidoptera (Russell et al. 2012, Gutzwiller et al. 2015) species.

3.4.3 Phylogenetic analysis of *Wolbachia* co-infection

The maximum-credibility tree shows high posterior probability (PP) values at most nodes (PP ≥ 0.95; Supp. Figure 2). The uncorrected tree height was 32.7 mya (95% HPD = 26, 44.1 mya; Supp. Figure 1), but this was scaled to match the tree produced by Gerth and Bleidorn (2016) at 76 mya (95% HPD = 60.4, 102.5 mya; see methods). The Gerth and Bleidorn (2016) phylogenetic tree was dated using the host divergence times and demonstrated that *Wolbachia* strains diverged within host sister species in the bee genus *Nomada* (Hymenoptera: Apidae). Our tree (Figure 2) demonstrates that the co-infecting *Wolbachia* strains in *A. morosus* did not diverge within this host and their most recent common ancestor (MRCA) was approx. 61.5 mya (95% HPD = 39.6, 85.8). Additionally, there is some evidence that these strains might have entered this host species in two separate horizontal transfer events (Figure 2). The MRCA within *w*Amor strain is approximately 2 mya (95% HPD = 0.3, 4.1 mya) indicating a longer divergence time within this host, whereas and *w*AmHa strain is approximately 0.52 mya (95% HPD = 0, 1.3 mya) (Figure 2). However, the 95% HPDs overlap (Figure 2) and these results should be interpreted cautiously.

3.5 Discussion

3.5.1 Widespread co-infection of *Wolbachia* in Australian hylaeine bee

Amphylaeus morosus was observed to be consistently infected with two supergroup A *Wolbachia* strains in samples taken from across its entire ~2,000 km range (Figure 1). Given that both the COI-like and *wsp* gene regions indicated that there were two widespread copies of each gene within all 27 individuals tested (as evident from both the Sanger sequences and cloning data), it is unlikely that this represents *Wolbachia*-derived pseudogenes within the parasite or host genomes. Assuming mostly vertical transmission of the bacteria within this

host (Tolley et al. 2019), for this co-infection to have spread so completely through the *A. morosus* population, these *Wolbachia* strains would have had to overcome major biological and geographical limitations of this host species.

Our phylogeny (Figure 2) assessing the divergence patterns of the co-infection of *Wolbachia* in *A. morosus* (two gene fragments; COI-like and *wsp*) was produced by integrating our data with the dated *Wolbachia* phylogeny produced by Gerth and Bleidorn (2016) based on whole genome data. It indicated that the two *Wolbachia* strains did not diverge from each other within *A. morosus* (Figure 2). These results suggest that these *Wolbachia* strains have entered this host independently (rather than a co-infection of unrelated *Wolbachia* strains transferring into a novel host — *Amphylaeus morosus* — simultaneously). For novel strains to independently infect a host lineage, *Wolbachia* needs to be transmitted via either horizontal or paternal transmission (Keeling et al. 2003).

The divergence time estimates within the *Wolbachia* strains indicate that *w*Amor has been diverging within *A. morosus* for approximately 2 million years (my), whereas *w*AmHa has been diverging for only approximately 0.5 my. This provides further support of two independent transmission events, rather than one horizontal transfer event from another co-infected host. However, this should be interpreted cautiously, as the inferred dates are based on limited data and 95% HPDs overlap between the divergence estimates for the two strains. Although these results concur with scenarios raised by Bailly-Bechet et al. (2017) where most *Wolbachia* infection events have occurred within the last few million years, these dates were developed using only ~1,200 bp, two host specimens, and less than three variable nucleotide sites *within* either strain. Full bacterial genome sequencing of more host specimens throughout the *A. morosus* distribution would be beneficial to (i) establish the phylogenetic relationships of these *Wolbachia* strains (particularly *w*Amor) compared to other supergroup A *Wolbachia*; (ii) better assess the divergence times within each *Wolbachia* strain, and (iii) attempt to examine the possible epicentre of each *Wolbachia* strain entering this host (i.e. if there were separate introductions of each *Wolbachia* strain, and if their initial infection occur in the same geographical location?).

3.5.2 Potential reproductive phenotypic manipulation of host

BLAST results indicated that similar *Wolbachia* strains to those in *A. morosus* have reproductive manipulation potential in their respective hosts. Relatives of the *w*AmHa strain, *w*CauA and *w*Ha, cause cytoplasmic incompatibility in their respective hosts, phycitinid moth

Cadra cautella and *Drosophila simulans* (Sasaki and Ishikawa 1999, Charlat et al. 2002). In another phycitinid moth host, *Ephestia kuehniella* (Zeller, 1879) (Lepidoptera: Pyralidae), *w*CauA causes male killing (Sasaki et al. 2002, Sasaki et al. 2005) when artificially introduced. BLAST results indicate that this *Wolbachia* group (labelled '*Haw*' by Zhou et al. (1998)) has a wide geographical and taxonomic range and BLAST results indicated infections in other Hymenoptera (parasitic wasps and ants (Wang et al. 2016)), Coleoptera (Floate et al. 2011) and Arachnida (scorpions (Ashtian et al. 2017)), although the phenotypic effects on these hosts are not known.

Interestingly, although divergent from the *Haw* group, relatives of the second strain infecting *A. morosus* (*w*Amor) are known to naturally infect the phycitinid moth *E. kuehniella*. In this host it causes cytoplasmic incompatibility (Sasaki and Ishikawa 1999), however it is not known what effects occur in other hosts (numerous Hymenoptera and Lepidoptera species). Although both *Wolbachia* strains in *A. morosus* have relatives within the phycitinid moths *C. cautella* and *E. kuehniella*, they do not appear to naturally co-occur within these moth species (Sasaki et al. 2002, Sasaki et al. 2005). Nonetheless, both *Wolbachia* strains are known to cause reproductive phenotypic manipulations of other hosts. It is feasible that in *A. morosus*, either/both strains could have induced a reproductive alteration in order to reach infection fixation across this bee's distribution.

3.5.3 Lack of knowledge of the distribution and effects of endosymbionts in bees

Wolbachia infections appear to occur frequently within bees, with supergroup A strains being the most common (Gerth et al. 2011, Gerth et al. 2013). Co-infections of different *Wolbachia* strains are also reported within bees (Jeyaprakash et al. 2009, Gerth et al. 2011, Gerth et al. 2013). But like other insect taxa, attempts to identify endosymbiont infections within bees has been generally limited to broad and general sampling within narrow geographical areas of interest (with low intra-species sampling), or within specific taxa (often those with close ties to human activities). Therefore, the biological roles of these endosymbionts within bee hosts have not been explored. Widespread taxonomic patterns of *Wolbachia* and their hosts have been difficult to interpret (Werren et al. 1995a, Werren et al. 1995b, Werren and Windsor 2000). This is further hampered by our considerable lack of knowledge about *Wolbachia* diversity and global host association patterns (Detcharoen et al. 2019). Additionally, other divergent reproductive manipulating endosymbionts occur in arthropods and can commonly co-occur with *Wolbachia* (Duron et al. 2008, Russell et al. 2012). We did not assess for the

presence of other reproductive parasites within *A. morosus*. The phenotypic impacts of *Wolbachia* (and other reproductive parasites) on their hosts is difficult to determine and extensive bacterial exclusion experiments are often needed to assess the effects of individual strains within a particular host (e.g. Dedeine et al. (2001), Koukou et al. (2006), Li et al. (2014), Li et al. (2019)).

3.5.4 *Wolbachia*'s relationship with heteroplasmic mitogenomes

It is likely that the reproductive strategy of *A. morosus* has been modified by one or both *Wolbachia* strains for the following reasons. Firstly, the extremely low mtDNA diversity across its entire ~2,000 km range (Chapter 2) is concordant with a bacteria-induced selective sweep (Jiggins 2003, Schuler et al. 2016). Secondly, relatives of both *Wolbachia* strains infecting *A. morosus* are known to have reproductive modification potential (Charlat et al. 2002, Sasaki et al. 2002, Sasaki et al. 2005). Thirdly, it is doubtful that a recent population bottleneck is a sufficient explanation for the lack of mtDNA variation (Chapter 2), given the large geographical distribution of *A. morosus* (Houston 1975). However, a *Wolbachia*-induced selective sweep alone does not explain the persistence of mitochondrial heteroplasmy.

We propose several likely hypotheses that might explain the widespread fixation AND maintenance of conserved heteroplasmic mtDNA in *A. morosus* (Figure 3).

(i) **The heteroplasmic founder hypothesis (H1)**: An *A. morosus* individual already possessed two mitogenomes (was heteroplasmic) and a single *Wolbachia* infection (which could have already swept through the population). A second *Wolbachia* strain entered this lineage (via horizonal or paternal transfer) and spread to fixation within the population, dragging along both mitogenomes and the initial *Wolbachia* strain. Mitochondrial heteroplasmy is maintained because of a mutualism with one of the *Wolbachia* strains (probably the ancestral infection).

(ii) **The co-inheritance hypothesis (H2)**: Two *A. morosus* individuals with different *Wolbachia* strains mated (each with a single divergent mitogenome), and both mitogenomes and *Wolbachia* strains were maintained in the progeny via paternal transmission of both a mitochondrial and bacterial genome. These co-occurring *Wolbachia* strains and their co-inherited mitogenomes then swept (via maternal inheritance) through the population to fixation. Each mitogenome is maintained by a mutualism with its corresponding *Wolbachia* strain.

(iii) **The advantageous heteroplasmy hypothesis (H3)**: The maintenance of mitochondrial heteroplasmy in *A. morosus* is unrelated to either *Wolbachia* strain. A selective sweep via either *Wolbachia* and/or an advantageous mutation (heteroplasmy) could be responsible for the overall loss of mitochondrial diversity. However, one of the two mitogenomes was never lost (post-fixation) because heteroplasmy has a fitness advantage over homoplasmy.

Our hypotheses include many assumptions about the biology of the host and parasite, and we cannot confirm which hypothesis applies to our system, or to similar systems. Each hypothesis will likely require extensive manipulative or modelling experiments to test. However, *A. morosus* presents a unique opportunity to investigate these complex interactions.

3.6 Concluding remarks

The Australian hylaeine bee, *Amphylaeus morosus* is consistently infected with two strains of the common insect intracellular parasite *Wolbachia* across its ~2,000 km range. Phylogenetic data indicates that these strains likely entered this host independently (Figure 2), with both strains having close relatives known to cause reproductive phenotypic alterations in other insect species. *Amphylaeus morosus* also lacks any mtDNA variation except for consistent mitochondrial heteroplasmy (found in every specimen). *Wolbachia* are known to cause extensive selective sweeps on host mtDNA, due to the linkage between the two genomes through maternal transmission. However, what mechanism(s) are maintaining widespread mitochondrial heteroplasmy despite the elimination of all other mtDNA variation remains unknown. Given the co-occurrence of two widespread bacterial and mitogenomes, it is possible that the mechanisms maintaining them are intertwined. We present three possible hypotheses (Figure 3) to explain the widespread loss of mtDNA haplotype diversity and consistent maintenance of mitochondrial heteroplasmy in this bee.

3.7 Acknowledgements

This work was supported by the Australian Government Research Training Program Scholarship, as well as the Playford Memorial Trust PhD scholarship. Funding for this research was provided by the Holsworth Wildlife Trust, Mark Mitchell Foundation, and Royal Society of South Australia. We extent our appreciation to Rebecca Dew and Ben Parslow for their assistance in fieldwork, and Tessa Bradford, Peter Anderson and James

Herbert for their guidance in laboratory protocols. Insect specimens were collected under the sampling permits SL101857 (NSW) and SL102137 (NSW)

3.8 References

Ashtian, N., S. Rajaie, A. T. Moghadam, and M. A. Bayatzadeh. 2017. Isolation and identification of *Wolbachia* from *Hemiscorpius lepturus* scorpion in Khuzestan province with PCR method. Journal of RNA and Genomics **13**:573–577.

Bailly-Bechet, M., P. Martins-Simoes, G. J. Szollosi, G. Mialdea, M. F. Sagot, and S. Charlat. 2017. How long does *Wolbachia* remain on board? Molecular Biology and Evolution **34**:1183–1193.

Ballad, J. W. O., J. Hatzidakis, T. L. Karr, and M. Kreitman. 1996. Reduced variation in *Drosophila simulans* mitochondrial DNA. Genetics **144**:1519–1528.

Bandi, C., J. W. McCall, C. Genchi, S. Corona, L. Venco, and L. Sacchi. 1999. Effects of tetracycline on the filarial worms *Brugia pahangi* and *Dirofilaria immitis* and their bacterial endosymbionts *Wolbachia* fn1. International Journal for Parasitology **29**:357–364.

Bian, G., D. Joshi, Y. Dong, P. Lu, G. Zhou, X. Pan, Y. Xu, G. Dimopoulos, and Z. Xi. 2013. *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. Science **340**:748–751.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C. Wu, D. Xie, C. Zhang, T. Stadler, and A. J. Drummond. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology **15**.

Braig, H. R., W. Zhou, S. L. Dobson, and S. L. O'Neill. 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. Journal of Bacteriology **180**:2373–2378.

Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe, E. A. McGraw, and S. L. O'Neill. 2009. Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. PLOS Pathogens **5**:e1000368.

Cariou, M., L. Duret, and S. Charlat. 2017. The global impact of *Wolbachia* on mitochondrial diversity and evolution. Journal of Evolutionary Biology **30**:2204–2210.

Charlat, S., A. Duplouy, E. A. Hornett, E. A. Dyson, N. Davies, G. K. Roderick, N. Wedell, and G. D. D. Hurst. 2009. The joint evolutionary histories of *Wolbachia* and mitochondria in *Hypolimnas bolina*. BMC Evolutionary Biology **9**:64.

Charlat, S., G. D. D. Hurst, and H. Merçot. 2003. Evolutionary consequences of *Wolbachia* infections. Trends in Genetics **19**:217–223.

Charlat, S., A. Nirgianaki, K. Bourtzis, and H. Mercot. 2002. Evolution of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila simulans* and *D. schellia*. Evolution **56**:1735–1742.

Clarke, L. A., C. S. Rebelo, J. Gonçalves, M. G. Boavida, and P. Jordan. 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Molecular Pathology **54**:351–353.

Dedeine, F., F. Vavre, F. Fleury, B. Loppin, M. E. Hochberg, and M. Boulétreau. 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic

wasp. Proceedings of the National Academy of Sciences of the United States of America **98**:6247–6252.
Detcharoen, M., W. Arthofer, B. C. Schlick-Steiner, and F. M. Steiner. 2019. *Wolbachia* megadiversity: 99% of these microorganismic manipulators unknown. FEMS Microbiology Ecology **95**.
Dobson, S. L., E. J. Marsland, and W. Rattanadechakul. 2002. Mutualistic *Wolbachia* infection in *Aedes albopictus*: accelerating cytoplasmic drive. Genetics **160**:1087–1094.
Domin, K. 1929. Pteridophyta (in Czech). Nákladem České Akademie Vêd a Umêni. , Praha, Czech Republic
Drummond, A. J. 2016. FigTree version 1.4.3. http://tree.bio.ed.ac.uk/software/figtree/ accessed: 25 May 2019.
Duron, O., D. Bouchon, S. Boutin, L. Bellamy, L. Zhou, J. Engelstädter, and G. D. Hurst. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biology **6**:27.
Dutra, Heverton Leandro C., Marcele N. Rocha, Fernando Braga S. Dias, Simone B. Mansur, Eric P. Caragata, and Luciano A. Moreira. 2016. *Wolbachia* blocks currently circulating Zika virus isolates in Brazilian *Aedes aegypti* mosquitoes. Cell Host & Microbe **19**:771–774.
Eckert, K. A., and T. A. Kunkel. 1991. DNA polymerase fidelity and the polymerase chain reaction. PCR Methods and Applications **1**:17–24.
Ellegaard, K. M., L. Klasson, K. Näslund, K. Bourtzis, and S. G. E. Andersson. 2013. Comparative genomics of *Wolbachia* and the bacterial species concept. PLOS Genetics **9**:e1003381.
Fleury, F., F. Vavre, N. Ris, P. Fouillet, and M. BoulÉTreau. 2001. Physiological cost induced by the maternally-transmitted endosymbiont *Wolbachia* in the *Drosophila* parasitoid *Leptopilina heterotoma*. Parasitology **121**:493–500.
Floate, K. D., P. C. Coghlin, and L. Dosdall. 2011. A test using *Wolbachia* bacteria to identify Eurasian source populations of cabbage seedpod weevil, *Ceutorhynchus obstrictus* (Marsham), in North America. Environmental Entomology **40**:818–823.
Folmer, O., W. Hoeh, M. Black, and R. Vrijenhoek. 1994. Conserved primers for PCR amplification of mitochondrial DNA from different invertebrate phyla. Molecular Marine Biology and Biotechnology **3**:294–299.
Fujii, Y., D. Kageyama, S. Hoshizaki, H. Ishikawa, and T. Sasaki. 2001. Transfection of *Wolbachia* in Lepidoptera: the feminizer of the adzuki bean borer *Ostrinia scapulalis* causes male killing in the Mediterranean flour moth *Ephestia kuehniella*. Proceedings of the Royal Society B: Biological Sciences **268**:855–859.
Fytrou, A., P. G. Schofield, A. R. Kraaijeveld, and S. F. Hubbard. 2006. *Wolbachia* infection suppresses both host defence and parasitoid counter-defence. Proceedings of the Royal Society B: Biological Sciences **273**:791–796.
Gavotte, L., D. R. Mercer, J. J. Stoeckle, and S. L. Dobson. 2010. Costs and benefits of *Wolbachia* infection in immature *Aedes albopictus* depend upon sex and competition level. Journal of Invertebrate Pathology **105**:341–346.
Gerth, M., and C. Bleidorn. 2016. Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. Nature Microbiology **2**:16241.
Gerth, M., A. GeißLer, and C. Bleidorn. 2011. *Wolbachia* infections in bees (Anthophila) and possible implications for DNA barcoding. Systematics and Biodiversity **9**:319–327.

Gerth, M., J. Röthe, and C. Bleidorn. 2013. Tracing horizontal *Wolbachia* movements among bees (Anthophila): a combined approach using multilocus sequence typing data and host phylogeny. Molecular Ecology **22**:6149–6162.
Gil, R., A. Latorre, and A. Moya. 2004. Bacterial endosymbionts of insects: insights from comparative genomics. Environmental Microbiology **6**:1109–1122.
Goebel, W., and R. Gross. 2001. Intracellular survival strategies of mutualistic and parasitic prokaryotes. Trends in Microbiology **9**:267–273.
Gomi, K., T. Gotoh, and H. Noda. 1997. *Wolbachia* having no effect on reproductive incompatibility in *Tetranychus kanzawai* Kishida (Acari: Tetranychidae). Applied Entomology and Zoology **32**:485–490.
Graham, R. I., and K. Wilson. 2012. Male-killing *Wolbachia* and mitochondrial selective sweep in a migratory African insect. BMC Evolutionary Biology **12**:204.
Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology **59**:307–321.
Gutzwiller, F., F. Dedeine, W. Kaiser, D. Giron, and C. Lopez-Vaamonde. 2015. Correlation between the green-island phenotype and *Wolbachia* infections during the evolutionary diversification of Gracillariidae leaf-mining moths. Ecology and Evolution **5**:4049–4062.
Hamm, C. A., D. J. Begun, A. Vo, C. C. R. Smith, P. Saelao, A. O. Shaver, J. Jaenike, and M. Turelli. 2014. *Wolbachia* do not live by reproductive manipulation alone: infection polymorphism in *Drosophila suzukii* and *D. subpulchrella*. Molecular Ecology **23**:4871–4885.
Harcombe, W., and A. A. Hoffmann. 2004. *Wolbachia* effects in *Drosophila melanogaster*: in search of fitness benefits. Journal of Invertebrate Pathology **87**:45–50.
Heath, B. D., R. D. J. Butcher, W. G. F. Whitfield, and S. F. Hubbard. 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Current Biology **9**:313–316.
Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the United States of America **101**:14812–14817.
Hedges, L. M., J. C. Brownlie, S. L. Neill, and K. N. Johnson. 2008. *Wolbachia* and virus protection in insects. Science **322**:702.
Hoffmann, A. A., M. Hercus, and H. Dagher. 1998. Population dynamics of the *Wolbachia* infection causing cytoplasmic incompatibility in *Drosophila melanogaster*. Genetics **148**:221–231.
Hoffmann, A. A., P. A. Ross, and G. Rašić. 2015. *Wolbachia* strains for disease control: ecological and evolutionary considerations. Evolutionary Applications **8**:751–768.
Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.
Hurst, G. D. D., and F. M. Jiggins. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences **272**:1525–1534.
Jeyaprakash, A., M. A. Hoy, and M. H. Allsopp. 2009. Multiple *Wolbachia* strains in *Apis mellifera capensis* from South Africa. Apidologie **40**:178–183.

Jiggins, F. M. 2003. Male-killing *Wolbachia* and mitochondrial DNA: sselective sweeps, hybrid introgression and parasite population dynamics. Genetics **164**:5–12.
Joshi, D., M. J. McFadden, D. Bevins, F. Zhang, and Z. Xi. 2014. *Wolbachia* strain *w*AlbB confers both fitness costs and benefit on *Anopheles stephensi*. Parasites & Vectors **7**:336.
Keeling, M. J., F. M. Jiggins, and J. M. Read. 2003. The invasion and coexistence of competing *Wolbachia* strains. Heredity **91**:382–388.
Koukou, K., H. Pavlikaki, G. Kilias, J. H. Werren, K. Bourtzis, and S. N. Alahiotis. 2006. Influence of antibiotic treatment and *Wolbachia* curing on sexual isolation among *Drosophila melanogaster* cage populations. Evolution **60**:87–96.
Lanfear, R., B. Calcott, S. Y. Ho, and S. Guindon. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution **29**:1695–1701.
Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution **34**:772–773.
Li, Y., X. Liu, and H. Guo. 2019. Population dynamics of *Wolbachia* in *Laodelphax striatellus* (Fallén) under successive stress of antibiotics. Current Microbiology **76**:1306–1312.
Li, Y. Y., K. D. Floate, P. G. Fields, and B. P. Pang. 2014. Review of treatment methods to remove *Wolbachia* bacteria from arthropods. Symbiosis **62**:1–15.
Mohammed, M. A., A. Aman-Zuki, S. Yusof, B. M. Md-Zain, and S. Yaakop. 2017. Prevalence and evolutionary history of endosymbiont *Wolbachia* (Rickettsiales: Anaplasmataceae) in parasitoids (Hymenoptera: Braconidae) associated with *Bactrocera* fruit flies (Diptera: Tephritidae) infesting carambola. Entomological Science **20**:382–395.
Moran, N. A., and P. Baumann. 2000. Bacterial endosymbionts in animals. Current Opinion in Microbiology **3**:270–275.
Moran, N. A., and J. J. Wernegreen. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends in Ecology & Evolution **15**:321–326.
National Center for Biotechnology Information (NCBI). 1988. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/blast accessed: 20/09/2020.
Nikoh, N., T. Hosokawa, M. Moriyama, K. Oshima, M. Hattori, and T. Fukatsu. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proceedings of the National Academy of Sciences of the United States of America **111**:10257–10262.
O'Neill, S. L., A. Hoffmann, and J. Werren. 1997. Influential passengers: inherited microorganisms and arthropod reproduction. Oxford University Press.
O'Neill, S. L., and T. L. Karr. 1990. Bidirectional incompatibility between conspecific populations of *Drosophila simulans*. Nature **348**:178–180.
Poinsot, D., and H. Merçot. 1997. *Wolbachia* infection in *Drosophila simulans*: does the female host bear a physiological cost? Evolution **51**:180–186.
Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology **67**:901–904.
Rokas, A., R. J. Atkinson, G. S. Brown, S. A. West, and G. N. Stone. 2001. Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? Heredity **87**:294–304.
Russell, J. A., C. F. Funaro, Y. M. Giraldo, B. Goldman-Huertas, D. Suh, D. J. C. Kronauer, C. S. Moreau, and N. E. Pierce. 2012. A veritable menagerie of heritable bacteria

from ants, butterflies, and beyond: Broad molecular surveys and a systematic review. PLOS ONE **7**:e51027.
Sasaki, T., and H. Ishikawa. 1999. *Wolbachia* infections and cytoplasmic incompatibility in the almond moth and the Mediterranean flour moth. Zoological Science **16**:739–744.
Sasaki, T., T. Kubo, and H. Ishikawa. 2002. Interspecific transfer of *Wolbachia* between two lepidopteran insects expressing cytoplasmic incompatibility: a *Wolbachia* variant naturally infecting *Cadra cautella*; causes male killing in *Ephestia kuehniella*. Genetics **162**:1313–1319.
Sasaki, T., N. Massaki, and T. Kubo. 2005. *Wolbachia* variant that induces two distinct reproductive phenotypes in different hosts. Heredity **95**:389–393.
Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S. Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffer, W. Arthofer, and M. Riegler. 2016. The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Molecular Ecology **25**:1595–1609.
Shoemaker, D. D., M. Ahrens, L. Sheill, M. Mescher, L. Keller, and K. G. Ross. 2003. Distribution and prevalence of *Wolbachia* infections in native populations of the fire ant *Solenopsis invicta* (Hymenoptera: Formicidae). Environmental Entomology **32**:1329–1336.
Smith, F. 1879. Descriptions of New Species of Hymenoptera in the Collection of the British Museum, London: British Museum.
Smith, M. A., C. Bertrand, K. Crosby, E. S. Eveleigh, J. Fernandez-Triana, B. L. Fisher, J. Gibbs, M. Hajibabaei, W. Hallwachs, K. Hind, J. Hrcek, D. Huang, M. Janda, D. H. Janzen, Y. Li, S. E. Miller, L. Packer, D. Quicke, S. Ratnasingham, J. Rodriguez, R. Rougerie, M. R. Shaw, C. Sheffield, J. K. Stahlhut, D. Steinke, J. Whitfield, M. Wood, and X. Zhou. 2012. *Wolbachia* and DNA barcoding insects: patterns, potential, and problems. PLOS ONE **7**:e36514.
Stouthamer, R., J. A. Breeuwer, and G. D. Hurst. 1999. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Annual Review of Microbiology **53**:71–102.
Sturtevant, A. H. 1919. A new species closely resembling *Drosophila melanogaster*. Psyche (Cambridge) **26**:153–155.
Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology **21**:3907–3930.
Tolley, S. J. A., P. Nonacs, and P. Sapountzis. 2019. *Wolbachia* horizontal hransmission events in ants: what do we know and what can we learn? Frontiers in Microbiology **10**:296.
Unckless, R. L., L. M. Boelio, J. K. Herren, and J. Jaenike. 2009. *Wolbachia* as populations within individual insects: causes and consequences of density variation in natural populations. Proceedings of the Royal Society B: Biological Sciences **276**:2805–2811.
Veneti, Z., S. Zabalou, G. Papafotiou, C. Paraskevopoulos, S. Pattas, I. Livadaras, G. Markakis, J. K. Herren, J. Jaenike, and K. Bourtzis. 2012. Loss of reproductive parasitism following transfer of male-killing *Wolbachia* to *Drosophila melanogaster* and *Drosophila simulans*. Heredity **109**:306–312.
Walker, F. 1863. Crambites & Tortricites: List of the Specimens of Lepidopterous Insects in the Collection of the British Museum., London: British Museum.
Wang, L., J. Jiang, Y. Xu, L. Zeng, and Y. Lu. 2016. Occurrence of three intracellular symbionts (*Wolbachia*, *Arsenophonus*, *Cardinium*) among ants in southern China. Journal of Asia-Pacific Entomology **19**:981–988.
Werren, J. H. 1997. Biology of *Wolbachia*. Annual Review of Entomology **42**:587–609.

Werren, J. H., L. Baldo, and M. E. Clark. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature Reviews Microbiology **6**:741–751.
Werren, J. H., D. Windsor, and L. Guo. 1995a. Distribution of *Wolbachia* among neotropical arthropods. Proceedings of the Royal Society B: Biological Sciences **262**:197–204.
Werren, J. H., and D. M. Windsor. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? Proceedings of the Royal Society B: Biological Sciences **267**:1277–1285.
Werren, J. H., W. Zhang, and L. R. Guo. 1995b. Evolution and Phylogeny of *Wolbachia*: Reproductive Parasites of Arthropods. Proceedings of the Royal Society B: Biological Sciences **261**:55–63.
White, J. A., S. E. Kelly, S. N. Cockburn, S. J. Perlman, and M. S. Hunter. 2011. Endosymbiont costs and benefits in a parasitoid infected with both *Wolbachia* and *Cardinium*. Heredity **106**:585–591.
Xi, Z., C. C. H. Khoo, and S. L. Dobson. 2005. *Wolbachia* establishment and invasion in an *Aedes aegypti* laboratory population. Science **310**:326.
Yang, X.-H., D.-H. Zhu, Z. Liu, and L. Zhao. 2012. Sequencing and phylogenetic analysis of the *wsp* gene of *Wolbachia* in three geographic populations of an oak gall wasp, *Andricus mairei* (Hymenoptera: Cynipidae), from Hunan, South China. Acta Entomologica Sinica **55**:247–254.
Zeller, P. C. 1879. Lepidopterologische Bemerkungen. Stettin Entomologische Zeitung.
Zhou, W., F. Rousset, and S. O'Neil. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proceedings of the Royal Society B: Biological Sciences **265**:509–515.
Zug, R., and P. Hammerstein. 2015. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. Biological Reviews **90**:89–111.

.

3.9 Tables

Table 1: Northern- to southern-most collection localities of the Australian native bee *Amphylaeus morosus* specimens assessed for endosymbiont infection using the *Wolbachia* COI-like gene

Location	**Lat, Long**	**Collection dates**	**Specimens sequenced**
Tin Can Bay, QLD	-25.92, 153.00	10 Dec 2013	3♀
Enfield State Forest, NSW	-31.31, 151.95	22 Dec 2018	1♀
Blue Mountains region, NSW	-33.64, 150.30	24 Jul 2017	4♀
Dandenong Ranges region, VIC	-37.94, 145.50	14–18 Aug 2014, 21–24 Nov 2016	12♀, 3♂
Cobboboonee State Forest and Narrawong, VIC	-38.21, 141.58	22–24 Feb 2017	4♀

3.10 Figures

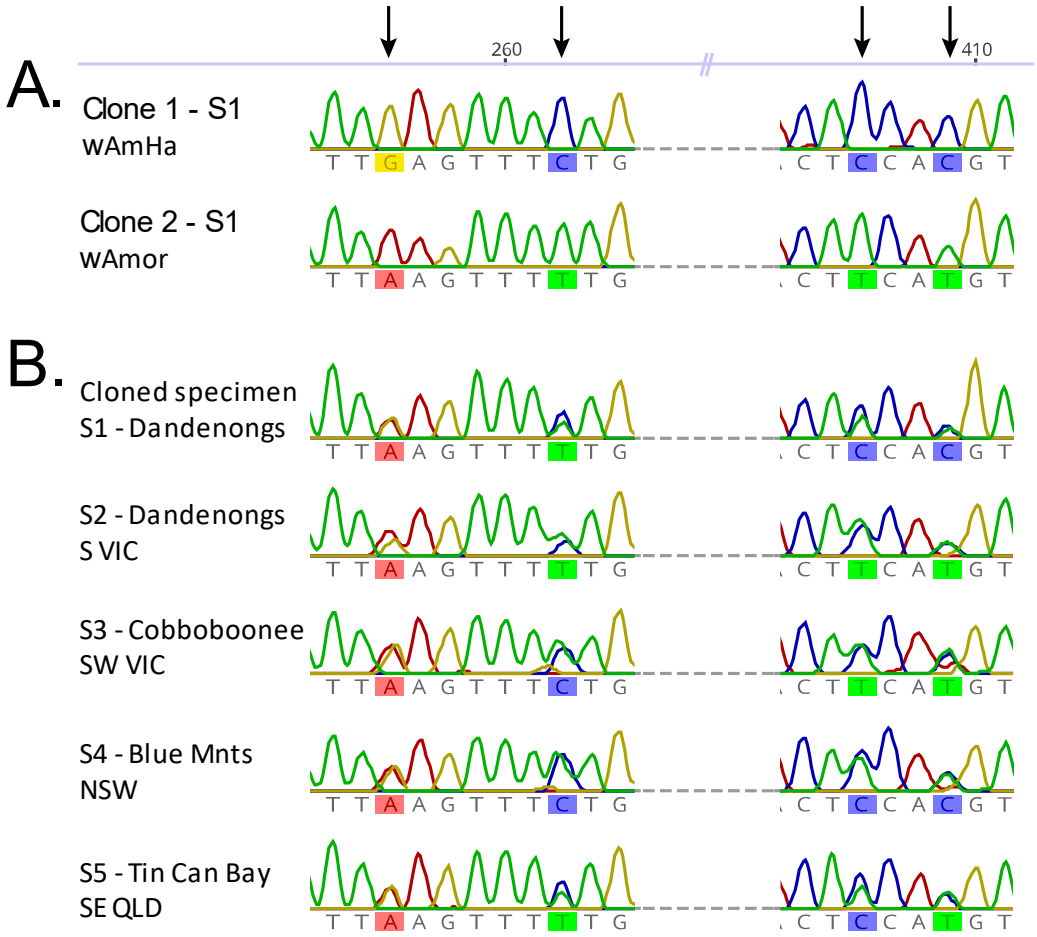


Figure 1: A. Examples of some variable nucleotide sites (indicated by arrows) in sequences recovered from a single cloned individual (S1) for the *Wolbachia* COI-like gene region. B. Aligned examples of the concurrent double peaks in chromatograms from Sanger sequencing detected in a cloned specimen (S1) and four example specimens (S2–5) from four *Amphylaeus morosus* populations along the east coast of Australia. All 27 specimens sequenced had consistently occurring double peaks.

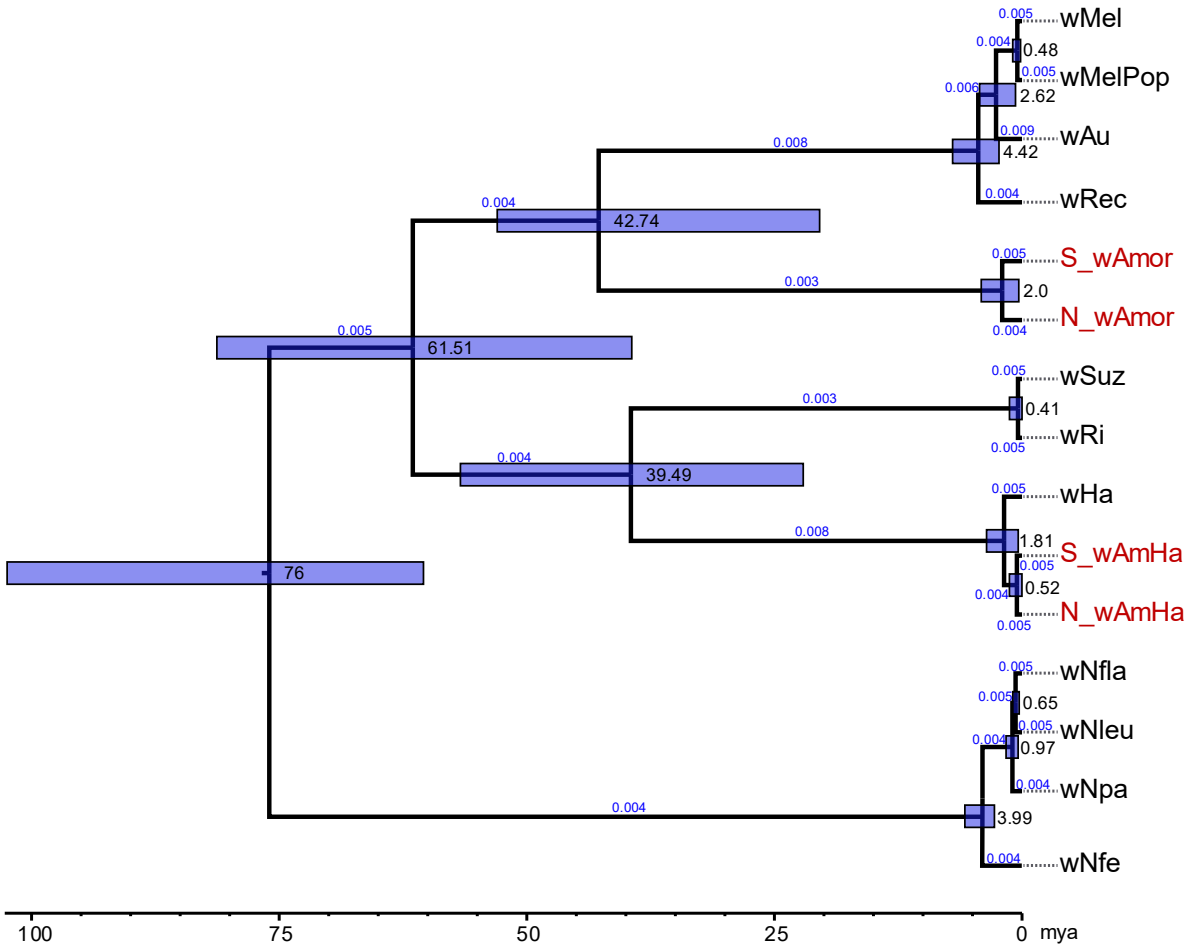


Figure 2: Adjusted dated phylogeny of *Wolbachia* supergroup A based on Gerth and Bleidorn (2016), with the addition of two bacterial strains (red; *w*Amor and *w*AmHa) infecting Australian colletid bee *Amphylaeus morosus* recovered in a specimen from each the southern- (S_) and northern-most (N_) localities of its distribution. Purple bars indicate 95% highest posterior density (HPD) values at each node of the phylogeny with values in black indicating node ages (mya). Blue text along branches indicates branch rates. Sequence data composed of two non-contiguous bacterial genes — *wsp* and COI-like (concatenated 1,271 bp).

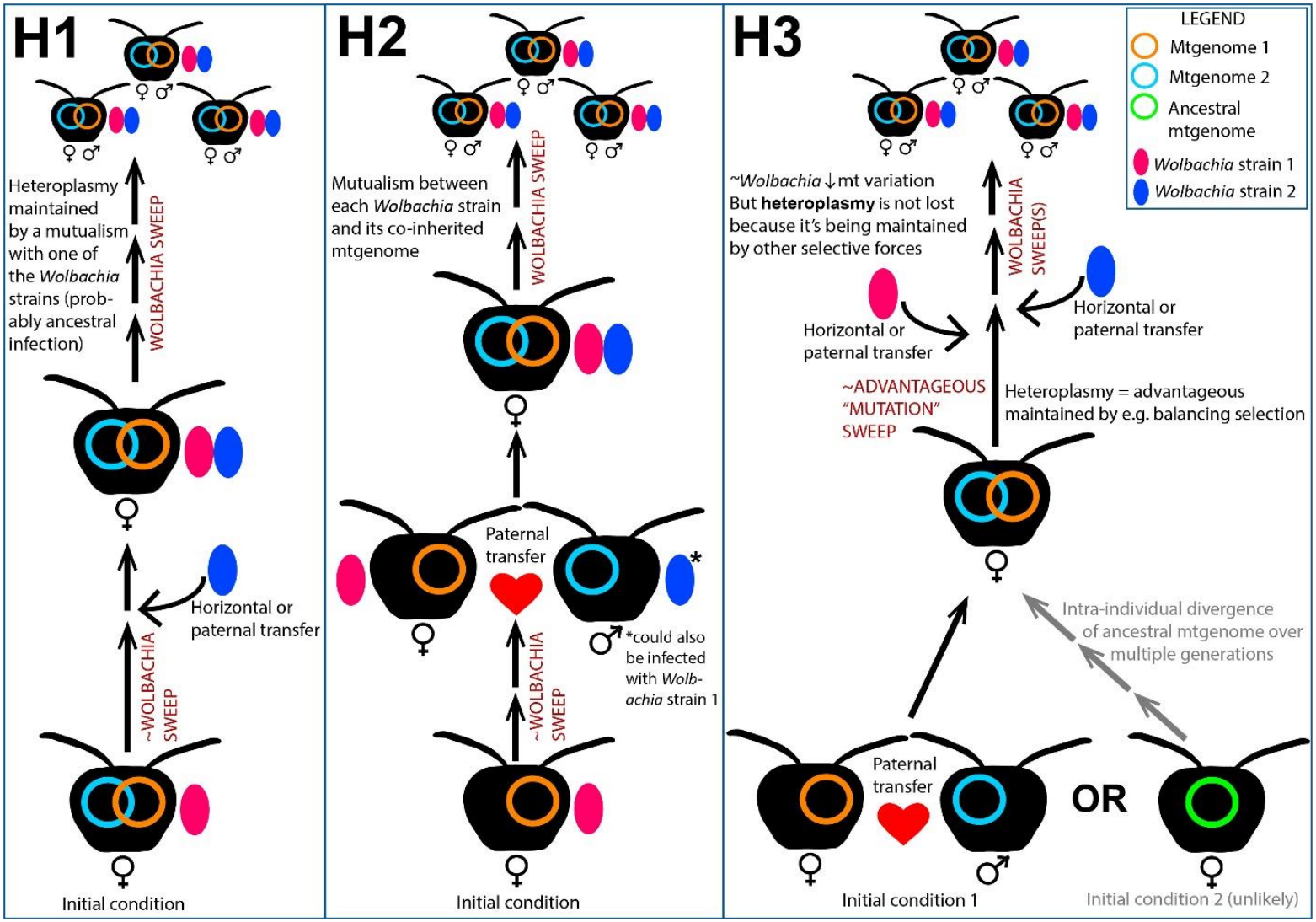


Figure 3: Three proposed hypotheses that could explain the widespread and consistent co-occurring heteroplasmic mitogenomes and *Wolbachia* strains in *Amphylaeus morosus*: H1: The heteroplasmic founder hypothesis; H2: The co-inheritance hypothesis; and H3: The null-*Wolbachia* hypothesis

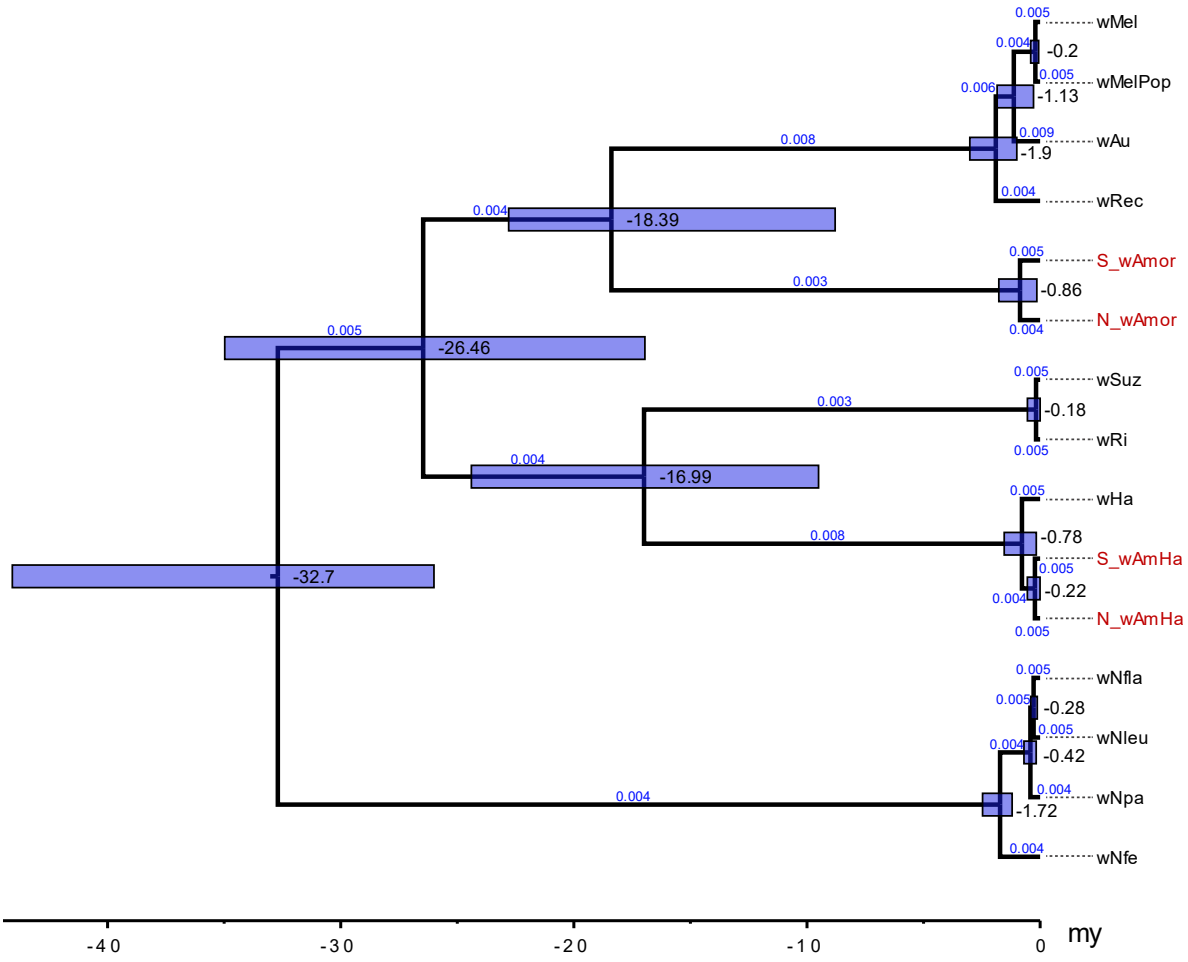
3.11 Supplementary Material

Supp. Table 1: Full names and NCBI Bioproject accession numbers of all supergroup A *Wolbachia* genomes used in this study (where relevant) for phylogenetic analysis based on Gerth and Bleidorn (2016). Full genomes were trimmed into two non-contiguous gene regions (COI-like and *wsp*) and fragment lengths are given for each. Total lengths of each gene were 665 bp total and 606 bp total, respectively.

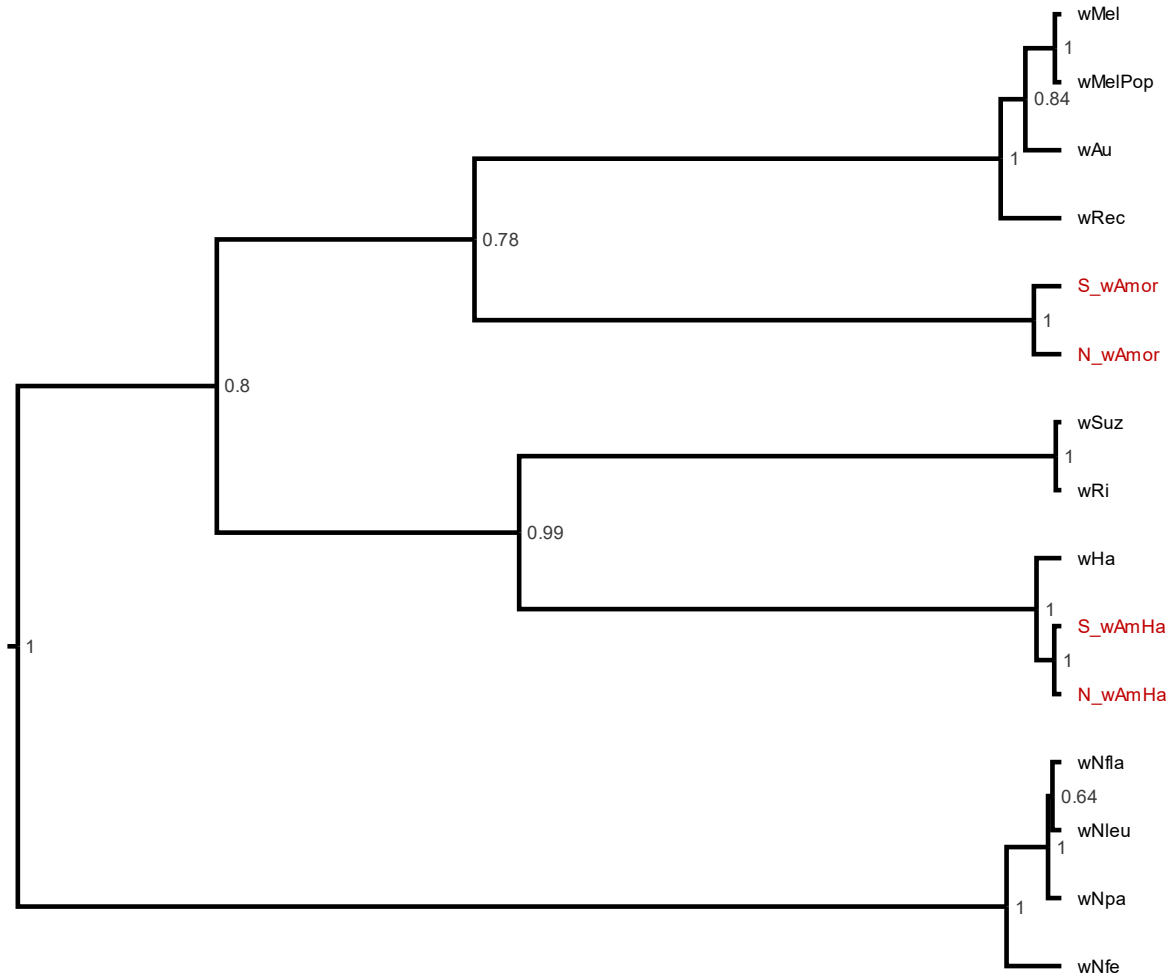
Name	**Host**	**Accession**	**Reference**	**COI-like**	***wsp***
*w*AmHa	*Amphylaeus morosus*	N/A	N/A	665 bp	579 bp
*w*Amor	*Amphylaeus morosus*	N/A	N/A	665 bp	564 bp
*w*Au	*Drosophila simulans*	PRJEB6321	Sutton et al. (2014)	665 bp	591 bp
*w*Ha	*Drosophila simulans*	PRJNA176303	Ellegaard et al. (2013)	665 bp	579 bp
*w*Mel	*Drosophila melanogaster*	PRJNA272	Wu et al. (2004)	665 bp	591 bp
*w*Melpop	*Drosophila melanogaster*	PRJNA196671	Woolfit et al. (2013)	665 bp	591 bp
*w*Nfe	*Nomada ferruginata*	PRJNA322628	Gerth and Bleidorn (2016)	665 bp	576 bp
*w*Nfla	*Nomada flava*	PRJNA322628	Gerth and Bleidorn (2016)	665 bp	576 bp
*w*Nleu	*Nomada leucophthalma*	PRJNA322628	Gerth and Bleidorn (2016)	665 bp	576 bp
*w*Npa	*Nomada panzeri*	PRJNA322628	Gerth and Bleidorn (2016)	665 bp	576 bp
*w*Rec	*Drosophila recens*	PRJNA254527	Metcalf et al. (2014)	665 bp	585 bp
*w*Ri	*Drosophila simulans*	PRJNA33273	Klasson et al. (2009)	665 bp	570 bp
*w*Suz	*Drosophila suzukii*	PRJEB596	Siozios et al. (2013)	665 bp	570 bp

Supp. Table 2: The number of amplicons of clones recovered for each strain of *Wolbachia* from each *Amphylaeus morosus* specimen from the northern-most (Tin Can Bay, QLD) and southern- (Dandenong Ranges, VIC) localities for both the COI-like and *wsp* gene regions.

		Number of amplicons recovered for each gene	
***Wolbachia* strain**	**Specimen location**	**COI-like**	***wsp***
*w*Amor	QLD	1	2
	VIC	5	1
*w*AmHa	QLD	1	1
	VIC	2	2



Supp. Figure 1: Dated phylogeny of *Wolbachia* supergroup A based on Gerth and Bleidorn (2016), with the addition of two bacterial strains (red; *w*Amor and *w*AmHa) infecting Australian colletid bee *Amphylaeus morosus* recovered in a specimen from each the southern- (S_) and northern-most (N_) localities of its distribution. Dates are those originally generated in our analyses, before adjustment to match those in Gerth and Bleidorn (2016). Purple bars indicate 95% highest posterior density (HPD) values at each node of the phylogeny with black values within indicating our generated node ages (mya). Blue text along branches indicates branch rates. Sequence data composed of two non-contiguous bacterial genes — *wsp* and COI-like (concatenated 1,271 bp).



Supp. Figure 2: Phylogeny of *Wolbachia* supergroup A based on Gerth and Bleidorn (2016), with the addition of two bacterial strains (red; *w*Amor and *w*AmHa) infecting Australian colletid bee *Amphylaeus morosus* recovered in a specimen from each the southern- (S_) and northern-most (N_) localities of its distribution. Sequence data composed of two non-contiguous bacterial genes — *wsp* and COI-like (concatenated 1,271 bp). Posterior probabilities for each node are included in black.

Supplementary References

Ellegaard, K. M., L. Klasson, K. Näslund, K. Bourtzis, and S. G. E. Andersson. 2013. Comparative genomics of *Wolbachia* and the bacterial species concept. PLOS Genetics **9**:e1003381.

Gerth, M., and C. Bleidorn. 2016. Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. Nature Microbiology **2**:16241.

Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes, A. C. Darby, Z. Veneti, L. Chen, H. R. Braig, R. Garrett, K. Bourtzis, and S. G. E. Andersson. 2009. The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. Proceedings of the National Academy of Sciences of the United States of America **106**:5725–5730.

Metcalf, J. A., M. Jo, S. R. Bordenstein, J. Jaenike, and S. R. Bordenstein. 2014. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. PeerJ **2**:e529.

Siozios, S., A. Cestaro, R. Kaur, I. Pertot, O. Rota-Stabelli, and G. Anfora. 2013. Draft genome sequence of the *Wolbachia* endosymbiont of *Drosophila suzukii*. Genome Announcements **1**:e00032.

Sutton, E. R., S. R. Harris, J. Parkhill, and S. P. Sinkins. 2014. Comparative genome analysis of *Wolbachia* strain *w*Au. BMC Genomics **15**:928.

Woolfit, M., I. Iturbe-Ormaetxe, J. C. Brownlie, T. Walker, M. Riegler, A. Seleznev, J. Popovici, E. Rancès, B. A. Wee, J. Pavlides, M. J. Sullivan, S. A. Beatson, A. Lane, M. Sidhu, C. J. McMeniman, E. A. McGraw, and S. L. O'Neill. 2013. Genomic evolution of the pathogenic *Wolbachia* strain, *w*MelPop. Genome Biology and Evolution **5**:2189–2204.

Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill, and J. A. Eisen. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLOS Biology **2**:e69.

Chapter 4: Incomplete removal of the reproductive parasite *Wolbachia* from a non-model organism following antibiotic treatment

Olivia K. Davies[a*], Michael G. Gardner[a,b], Mark I. Stevens[b,c], and Michael P. Schwarz[a]

[a]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

[b]South Australian Museum, GPO Box 234, Adelaide SA 5001, Australia

[c]School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia

*Corresponding author. Email: davies.oliviak@gmail.com

Keywords: Hymenoptera, bee, endosymbiont, tetracycline, mitochondrial heteroplasmy

4.1 Abstract

Wolbachia is a common α-proteobacteria genus known to exclusively infect invertebrates. Although *Wolbachia* has a wide range of effects on its hosts, it is best known for its ability to induce reproductive phenotypic manipulations to increase its own inheritance rate. *Wolbachia* is also known to cause extensive mitochondrial selective sweeps in host species, as genetic linkage through maternal inheritance causes co-inherited mitogenomes to 'hitchhike' with *Wolbachia* across reproductive events. However, assessing the impacts of *Wolbachia* on specific hosts is challenging and often requires diligent antibacterial regimens and/or multiple generations of observation. *Amphylaeus morosus* is a widespread hylaeine bee (Hymenoptera: Colletidae) that occurs along the eastern Australia coast. We have previously identified that that every individual appears to have two conserved and co-occurring *Wolbachia* strains and mitochondrial haplotypes (in the form of mitochondrial heteroplasmy) across its entire geographic distribution. To explore whether conserved mitochondrial heteroplasmy is being maintained by dual-strain *Wolbachia* co-inheritance, we proposed conducting an antibacterial treatment to eradicate *Wolbachia* from host individuals and their subsequent offspring and assess the effects on mitochondrial inheritance, post-infection. However, *A. morosus* is not able to be reared in laboratory conditions, and we endeavoured to assess the utility of traditional antibiotic treatment to remove *Wolbachia* infections in a non-model organism outside of a strictly controlled environment. Within one generation, *Wolbachia* were not removed from this host, however one *Wolbachia* strain appears consistently more susceptible to antibiotic treatment than the other. We propose multiple improvements for this treatment regimen however, the inability to successfully rear this bee species for multiple generations in captivity remains a major limitation for future attempts at antibacterial treatment.

4.2 Introduction

Wolbachia is a common intracellular bacterium that infects arthropods and nematodes. These α-proteobacteria belong to the diverse order Rickettsiales, which includes species demonstrating a range of relationships with their hosts, including parasitic, mutualistic, and commensal relationships (Yu and Walker 2006). *Wolbachia* is known to exclusively infect invertebrates (Werren et al. 2008) unlike many of its bacterial relatives (Yu and Walker 2006). As many as 16 supergroups (A – Q, excluding G) of *Wolbachia* have been proposed (Baldo and Werren 2007, Ros et al. 2009, Glowska et al. 2015), exhibiting various host-endosymbiont relationships which are generally characteristic within supergroups (Werren et al. 2008). *Wolbachia* in supergroups A and B are the most common infection within arthropods, particularly insects (Werren and Windsor 2000). The effects of *Wolbachia* infections on hosts are variable, but supergroup A and B bacteria often display parasitic traits associated with reproductive manipulation (Werren et al. 1995).

When functioning as reproductive parasites, *Wolbachia* can induce four known phenotypic manipulations; cytoplasmic incompatibility, feminization of genetic males, thelytokous parthenogenesis induction, and male killing (Werren 1997, Stouthamer et al. 1999). *Wolbachia* can also have other (and often multiple, simultaneous) effects on their hosts, including providing resistance to pathogens (e.g. Chrostek et al. (2013), Ye et al. (2013), Johnson (2015)), influencing host mating behaviour (e.g. Vala et al. (2004), Rohrscheib et al. (2015)) and causing various — positive, negative, and intermediate — fitness effects on their hosts (e.g. Tagami et al. (2001), Fry et al. (2004), Xie et al. (2011)). Assessing the influence of *Wolbachia* in host populations is challenging, and often requires antibacterial treatments and/or multiple generations of observation (Li et al. 2014). The presence of endosymbiont co-infections is reasonably common (Duron et al. 2008) and can further complicate attempts to unravel the roles of these endosymbionts within their hosts.

Amphylaeus morosus (Smith, 1879) belongs to the cosmopolitan bee family Colletidae and is distributed widely along the eastern coast of Australia (Houston 1975). This species has widespread and conserved mitochondrial heteroplasmy where every individual appears to have the same two mitochondrial haplotypes across the entire geographic distribution (Chapter 2). This bee is also consistently infected with two distinct strains of supergroup A *Wolbachia* across this distribution (Chapter 3). In summary, these unusual traits in *A. morosus* prompts four questions: (i) How did mitochondrial heteroplasmy arise? (ii) What

are the mechanisms that continue to maintain widespread mitochondrial heteroplasmy? (iii) What processes contributed to the loss of all other mitochondrial variation? (iv) How (if at all) has the *Wolbachia* co-infection contributed to the unusual mitochondrial traits in *A. morosus*?

Wolbachia is known to cause extensive selective sweeps on host mitochondrial DNA (mtDNA) (e.g. Jiggins (2003); Raychoudhury et al. (2010); Graham and Wilson (2012); Schuler et al. (2016)). The lack of mtDNA variation (except mitochondrial heteroplasmy) could be explained by a *Wolbachia*-induced selective sweep. However, the widespread presence of conserved mitochondrial heteroplasmy remains puzzling.

We have previously presented multiple hypotheses (Table 1) that might explain the widespread occurrence *and* maintenance of consistent mitochondrial heteroplasmy in *A. morosus*. This included hypotheses where *Wolbachia* is the agent facilitating widespread mitochondrial heteroplasmy (Table 1: H1 or H2) — alternative to the possibility that mitochondrial heteroplasmy being maintained by the selection of advantageous genotypes (Table 1: H3). In these cases (H1 or H2), *Wolbachia* are the selective forces maintaining heteroplasmic mitochondrial inheritance, whereby one or both *Wolbachia* strain(s) 'drag' their co-inherited mitochondrial haplotype across reproductive events. To explore whether conserved mitochondrial heteroplasmy is being maintained by dual-strain *Wolbachia* co-inheritance, we attempted an antibacterial treatment to eradicate *Wolbachia* from host individuals and their subsequent brood and assess the effects on mitochondrial inheritance within the host, post-infection. However, *A. morosus* is not able to be reared in laboratory conditions, and we endeavoured to assess the utility of traditional antibiotic treatment to remove *Wolbachia* infections in this non-model organism.

4.3 Methods

4.3.1 Nest sampling and treatment preparation

An experiment was conducted on adult colonies of wild *Amphylaeus morosus* collected from the Dandenong Ranges in Victoria, Australia between 1st to 5th Aug 2016 (late winter) before brood rearing commences (this bee species overwinters as adults). Schwarz and Overholt (1993) found that colony transfer of overwintering allodapine bees (Hymenoptera: Apidae) were most successful, so we replicated this with *A. morosus* colonies which were transferred into artificial Perspex®-topped pine nests (Schwarz and Overholt 1993) at 10°C in a

temperature-controlled room at Flinders University, South Australia. Artificial nests (Supp. Figure 1) were made from untreated square pine dowel with the dimensions (H×W×D) 18×17×400 mm. A rounded nesting cavity was routed out of the upper surface of the dowel with the dimensions 6×7×380 mm, and then smoothed using coarse sandpaper. To seal the nest cavity, clear Perspex® was cut to the dimensions of the dorsal surface of the dowel and taped securely into place, which allowed nest construction and occupation to be observed. Weather-flashing (consolidated alloy 0.3 mm thick) followed by forest green corflute was then folded around the lateral and dorsal sides of the dowel to protect against rain, ambient temperature fluctuations, and solar radiation. The nest opening was left unobstructed (with no overhang from the Perspex®, weather-flashing, corflute layers) to replicate natural nest entrances. A unique simple symbol (e.g. triangle, star, square etc) was drawn on the anterior surface near each opening (Supp. Figure 1) to provide a possible nest identification tool for the bees (nests may be slightly repositioned during the experiment, making potential 'natural' landmarks less reliable for the bees).

A total of 112 nests consisting of solitary (one female) and social nests (with up to five females per nest) were evenly, in no particular order, allocated into three antibiotic treatment groups. Each treatment group was fed over two days on of the following tetracycline doses: 10 mg/mL (high dose), 2 mg/mL (low dose), or 0 mg/mL ("control") antibiotic tetracycline. A range of doses was applied as we were considerate of the potential toxic effects of antibiotics (Dedeine et al. 2001, Stouthamer and Mak 2002) in this species. To expose the adult female bees (F0) to the antibiotic, tetracycline was dissolved in diluted commercial honey (1:1) solution at the designated antibiotic concentration and presented in a 1 mL transfer pipette positioned at the entrance of each nest following a week of fasting. Although the volume of the solution was marked on the pipette tip to monitor the consumption by the bees, the solution would often leak, and we could not reliably assess the volume consumed. Every bee was marked on the scutum with enamel Testors® paint corresponding to their dose; red paint was 10 mg/mL, blue paint was 2 mg/mL, and yellow paint was 0 mg/mL.

Nests were distributed at an outdoor compound at Flinders University Bedford campus and in the gardens of three residential properties throughout the Adelaide Hills region. These residential locations included Eden Hills, Belair, and Macclesfield, South Australia. Each location contained the two treatment groups and control. All residential properties were <200 m from native bushland reserves and all locations had diverse native flora on the

property. The distribution of nests across multiple properties was to ensure that there were no confounding local effects that would impact the entire experiment, such as unfavourable food sources, local pathogens or predators, or weather events. Nests were positioned between 0.5 to 1.8 m off the ground under partial cover provided by surrounding vegetation and structures at between 45^{o} to 90^{o} to prevent becoming waterlogged (Supp. Figure 1C.).

These nests were collected 10 days post-release and refed the antibiotic diluted honey solution. However, we found that nest switching was common at all sites, resulting in cases of nest abandonment and females of different doses taking up residences in other nests. As such, the 0 mg/mL ("control") nests were treated with 5 mg/mL tetracycline to ensure all bees (F0) previously treated with antibiotics were exposed in the second treatment. Colony members were fed for two days and were presented with antibiotics based on the original dose prescribed for that nest, rather than who the current residents were — for example a 'blue' colony was treated with 2 mg/mL tetracycline in diluted honey solution regardless of whether blue-painted individuals currently resided there. Completely abandoned nests were removed from the experiment as they were usually superseded by other insects, such as ants or Dermaptera.

Once nests were repositioned in their original locations in their outdoor plots, to limit stress on the bees, risks of nest abandonment and nest swapping, bee activity was only monitored visually upon their second and final treatment. These adult females (F0) were then undisturbed during egg laying and provisioning of offspring from local nectar and pollen sources in artificial nests with limited human interference for three months. The F1 progeny were collected in early November 2016 and once food provisions were totally consumed, juveniles were transitioned to Petri dishes where they were reared to adulthood under controlled conditions. F1 adults (antibiotic treated) were then stored in 99% ethanol for subsequent analyses. Due to the nest switching behaviour, and the loss of the experimental "control" group (0 mg/mL tetracycline), a subsequent fieldtrip was conducted from 21st to 24th Nov 2016 to collect provisioned nests from the same sampling location in the Dandenong Ranges, to serve as a replacement control group (labelled "W" for "wild type"). These individuals (W1) were also reared to adulthood as above and then stored in 99% ethanol.

4.3.2 DNA Extraction

Total DNA was extracted from the tissue of a single hind leg of 29 antibiotic treated F1 (3 complete nests) and 17 untreated W1 (4 complete nests) individuals. Extractions were completed using an adapted Gentra Puregene Cell Kit procedure (Qiagen) at the South Australian Regional Facility of Molecular Ecology and Evolution (SARFMEE), following the same protocols that were originally used to detect *Wolbachia* and mitochondrial heteroplasmy (see Chapter 2 and 3). Only F1 offspring from the 10 mg/mL (red) tetracycline treated nests were selected, and nests that yielded high numbers of offspring, contained male and female progeny, and consistently had residing mothers marked with red paint were used for this analysis. The adult females (F0) were not used for DNA analysis due to mass die off following the end of their reproductive cycle and nest swapping behaviour that makes assigning reliable parentage, with few surviving adult females, impossible with the applied genetic techniques. Extracts were stored at 4^{o}C prior to Polymerase Chain Reaction (PCR) amplification.

4.3.3 PCR amplification and Sanger sequencing

PCR amplification of mitochondrial cytochrome *c* oxidase subunit 1 (mtCOI), *Wolbachia* COI-like gene, and *Wolbachia* surface protein (*wsp*) was carried out on the 46 F1/W1 individuals using the PCR conditions previously optimized for detecting heteroplasmy and *Wolbachia* infection in *A. morosus* (Chapter 2 and 3). PCR amplification of all DNA fragments was carried out in a total volume of 25 µL, as follows: 1x MRT Buffer ($MgCl_2$), primers (0.4 µM each), 1 U Immolase DNA Polymerase (Bioline 5 U/µL), and 2 µL template DNA. Primer specifications summarised in Table 2.

PCR cycling conditions for primer sets COIF-PR115/COIR-PR114 and Lep-F1/Lep-R1were one cycle (10 min at 95^{o}C), 35 cycles (45 sec at 94^{o}C, 45 sec at 48^{o}C, 60 sec at 75^{o}C) and one cycle (6 min at 72^{o}C, 2 min at 25^{o}C). Conditions for primer set wsp-81F/wsp-691R were one cycle (3 min at 95^{o}C), 35 cycles (60 sec at 95^{o}C, 60 sec at 55^{o}C, 60 sec at 72^{o}C) and one cycle (30 min at 72^{o}C) (Zhou et al. 1998). Visualisation of PCR product were shown on 1.5% agarose gel and successful PCR reaction products then were purified using Multiscreen PRC384 Filter Plate (Millipore) and re-suspended in 25 µL of 10 mM TRIS.

All visualised PCR products for *Wolbachia* genes indicated bacterial presence. To confirm potential retention of *Wolbachia* co-infections across the experiment, amplicons from 17 individuals (both control, W1 (n = 7) and experimental, F1 (n = 10) treatments) were selected

and successfully Sanger sequenced for bacterial and mitochondrial gene regions by the Australian Genome Research Facility (AGRF) in Adelaide, South Australia. Reaction mixtures were in volumes of 12 μL (11 μL PCR product, 1 μL primers) and were sequenced with Applied Bio-Systems 3730 and 3730 xl capillary sequencers. Only the forward sequences were obtained and were edited and aligned using Geneious Prime 2020.1 (https://www.geneious.com) to confirm mitochondrial heteroplasmy and co-infecting *Wolbachia* strains in each individual by looking for double peaks in the chromatograms indicating heterogeneous bases.

4.3.4 Statistical analysis of nest activity

All statistical tests were performed in SPSS version 25.0 and as our data did not meet the requirements of parametric tests, non-parametric tests were used (Figure 1). To assess the composition of adult females during the experiment, Kruskal Wallis tests were performed. These tests assessed whether the experimental antibiotic dose or location impacted (i) nest occupation — the number of females per nest at each stage of the experiment and (ii) nest switching — changes in the number of females per nest during the early stages of the experiment (Figure 1A).

The impact of antibiotic dose and location of nests on brood production (number of brood and sex ratios) was also assessed. Brood sex ratios were arcsine square-root transformed prior to statistical tests. Kruskal Wallis tests were used to assess whether location or antibiotic dose (control and each experimental antibiotic concentration) impacted (i) the number of adult females remaining at the termination of the experiments, (ii) the number of brood produced in occupied nests, or (iii) the (transformed) sex ratio of the brood produced (Figure 1B). For any tests that indicated significant differences, Dunn-Bonferroni *post hoc* tests were conducted. Mann-Whitney U tests were also used to assess if there was an overall difference in number of adult females remaining, brood produced, in the experimental versus control (untreated) nests and (transformed) brood sex ratio (Figure 1B). Additionally, a one-sample Wilcoxon signed-rank test was used to estimate whether the sex ratio (untransformed) of the brood significantly deviated from the expected median of 0.5 (Spessa 1999) (Figure 1B).

4.4 Results

4.4.1 Evidence of partial removal of *Wolbachia*

Twelve individuals were successfully sequenced for both *Wolbachia* genes (COI-like and *wsp* genes) and, of those, ten were reared F1 offspring from antibiotic treated F0 mothers, and two were control (untreated), W1 individuals. Five additional control W1 individuals were also sequenced for the COI-like gene. Both gene sequences were assessed for double peaks or ambiguous nucleotide sections at known regions of variation (Figure 2) identified in each *Wolbachia* strain infecting *Amphylaeus morosus*, previously separated through cloning (Chapter 3). These data suggested that of the ten F1 individuals that were assayed, both the COI-like and *wsp* genes showed evidence of the loss or suppression of one of the *Wolbachia* strains in six of these individuals. The sequences of the remaining four F1 individuals were consistent with co-infected control samples (although one of these F1 specimens was somewhat ambiguous, appearing suppressed in the *wsp* sequence, but showing some double peaks consistent with a double *Wolbachia* infection in the COI-like gene). In all cases, the same *Wolbachia* strain was "lost" (*w*Amor), suggesting partial removal or suppression of *Wolbachia* where one strain is more suspectable to antibiotic treatment.

4.4.2 No effect on mitochondrial heteroplasmy

A total of 17 sequences from the same W1/ F1 individuals (control W1, n = 7 and experimental F1, n = 10) were obtained for the mt-COI (forward) fragment Sanger sequences, with lengths of 609 bp. Every specimen exhibited double peaks throughout the mt-COI gene which were comparable to those found in previous *A. morosus* samples. In total, 24 heteroplasmic sites were recovered for the mt-COI fragment, consistent with previous results (see Chapter 2). There is no evidence that antibiotic treated F1 offspring that exhibited partial removal of *Wolbachia* had any changes in the presence of heteroplasmic mitogenomes in one generation. An additional novel heteroplasmic site (synonymous) in two untreated (control group) specimens reared from the same nest (female and male) from the Dandenong Ranges, VIC was identified. This is only the second novel nucleotide variant to be identified in all *A. morosus* specimens sequenced for the 658 bp mt-COI fragment across Australia — the other being from an individual from a nest collected in the Blue Mountains, NSW (Chapter 2).

4.4.3 Statistical analysis of female activity in artificial nests

Experimental nests only

Location of artificial nests (four locations in Adelaide) nor antibiotic doses had any impact on nest activity within the experimental cohort (Table 3). The initial number of adult females (F0) per nest ranged between one and five individuals, and two weeks post-release, between zero and six. Of the nests recovered at this stage, 46 of 112 no longer contained adult females and multiple nests had a different number of females than the original nest arrangement. Statistical tests demonstrated that these changes in the number of females per nest did not vary significantly between locations nor antibiotic dose applied at the two-week re-feeding period (Table 3). Upon the termination of the experiment, only 35 of the initial 207 adult females (F0) remained. Thirty of the original 112 nests contained provisioned brood cells (F1; 222 brood cells in total). Many nests had no resident F0 females remaining, but 23 brood-containing nests still contained at least one resident female (an additional nest with no brood contained a deceased female). At the termination of the experiment, there was no significant different in the remaining number of adult females (F0) nor the number of brood (F1) produced at each location or from any particular antibiotic treatment (Table 3).

Replacement control nests

A total of 31 nests were collected from the Dandenong Ranges, VIC to replace the control nests lost in the experimental cohort due to nest switching behaviour. Of these, one nest contained cellophane-like cell lining but was empty, three nests had a residing adult female but no attempt at brood production and in the remaining 27 nests, brood production had commenced to some degree. Of these 27 successful nests, the number of residing adult females (W0) ranged between zero and three at the time of collection. Brood cell numbers ranged between one cell still being provisioned (not yet sealed) to 11 provisioned cells (however, only eight W1 offspring survived to adulthood in this nest — the most productive control nest). No parasitoids known to impact *A. morosus* nests in this region were reared from these nests e.g. *Gasteruption* sp. (Hymenoptera: Gasteruptiidae), *Ephutomorpha* spp. (Hymenoptera: Mutillidae) or *Anthrax* sp. (Diptera: Bombyliidae) (Spessa 1999, Hearn et al. 2019).

Comparison of control and experimental nests

No effect of tetracycline dose (specific regimen or broadly grouped) was detected on the number of adult females (W0/F0) remaining in nests at termination (Table 4). Location was found to have a significant effect (Table 4) however, exploration of these differences (with

Dunn-Bonferroni *post hoc* tests, Supp. Table 1) demonstrated that these significant differences were driven the small ranges in the remaining adult females (between 0–3 individuals) across these treatments and small sample sizes within locations.

Location was found to have no significant impact on F1 brood production however, tetracycline dose (specific regimen and broadly grouped) was found to have an effect (Table 4). These results were also explored further (with Dunn-Bonferroni *post hoc* tests, Supp. Table 2 and within-group distribution of data, Supp. Figure 2). Overall, the experimental nests produced significantly more brood (n = 31, $\bar{x}$ = 7.2, SD = 4.5) than the control (untreated) nests (n = 29, $\bar{x}$ = 4.4, SD = 2.7) (Table 4, Supp. Figure 2). Within treatment groups, only antibiotic doses with the highest and lowest average number of brood (W1/F1) produced were significantly different (i.e. 5 mg/mL versus control (untreated) respectively) (Table 4). Again, these differences should be interpreted cautiously due to small sample sizes within groupings. Location nor antibiotic dose (specific regimen or broadly grouped) had any effect on the sex ratio of the brood (W1/F1) (Table 4). No deviation for an expected sex ratio of 1:1 was detected in these brood (W1/F1) however, small sample sizes limit the reliability of interpreting these sex ratio data.

4.5 Discussion

Our study attempted to assess the viability of antibiotic treatment to remove *Wolbachia* in a non-model species. Within one generation, the *Wolbachia* infection was not completely removed from this host. However, there is evidence that one strain may have been more susceptible to antibiotic therapy than the other. Six of the tetracycline-treated F1 individuals (n = 10) assayed lacked (i) heterogeneous base calls in their *Wolbachia* COI-like sequences and (ii) the ambiguous nucleotide sections at known regions of variation in *wsp* sequences which are observed in the control (W1, untreated) samples (Figure 2). Upon comparison with known isolates of each *Wolbachia* strain for each gene, these data suggest that the *Wolbachia* strain *w*Amor is more susceptible to tetracycline treatment. The other strain, *w*AmHa was present in the somatic (leg) tissues of all individuals. Other studies also provide evidence that some *Wolbachia* strains may be antibiotic resistant or hosts may be unable to sequester the antibiotic (Dyson et al. 2002). However, to assess the extent of antibiotic effectiveness in *Amphylaeus morosus*, quantitative analyses (e.g. quantitative PCR) would need to be conducted with a specific marker/probe developed for each *Wolbachia* strain.

Additionally, no loss of mitochondrial heteroplasmy was detected in F1 progeny (within one generation). The combination of methodology of antibiotic application and only one generation of treated individuals limited our results. *Amphylaeus morosus* females (F0) were treated with tetracycline with the intention of also removing bacteria from the oocytes in their reproductive system (Ghedin et al. 2009). As heteroplasmic mitochondria (which is likely maternally inherited in this species — Chapter 2) are probably present in the oocytes, it is unlikely that heteroplasmy will be lost within such a short generation turn-over. Other studies addressing the loss of mitochondrial heteroplasmy in model systems have demonstrated it takes multiple generations to achieve shifts in mitochondrial inheritance patterns (e.g. Clark et al. (2012), Ma et al. (2014)). Evidence of the facilitation of atypical mitochondrial inheritance by *Wolbachia* in other taxa is not extensive. In the study systems where patterns were initially supportive, these traits were subsequently found to be due to other mechanisms mostly unrelated to *Wolbachia* (e.g. Marcadé et al. (2007), Doublet et al. (2012)).

Our antibiotic treatment methodology was purposely conservative, as we were cautious of the health and reproductive effects that tetracycline may have on this bee species. Statistical analyses of our data suggested that the dose of tetracycline had no significant effect on the nesting behaviour in *A. morosus*. No difference in changes in composition of adult females (F0) in nests nor brood production (F1) was detected between doses or nesting location in the experimental nests (Table 3). Comparing control and experimental groups, some differences in the number of remaining adult females (F0) and brood (F1) produced in nests at termination were detected (Table 4). However, these differences were associated with the control (untreated) versus specific cohorts within the experimental group and were overinflated by small sample sizes.

Nests collected from the Dandenong Ranges as our replacement control group were less developed (juveniles in earlier stages of development and some cells still being provisioned) than those in South Australia, and this is likely due to climate differences and/or reduced time spent on nest construction in South Australia as nests were artificial. Fewer brood and more residing adult females still occupying nests in these less developed control (untreated) nests were observed to drive the differences in the statistical analyses. Offspring sex ratios of brood did not deviate from the expected ratio of 1:1 (Spessa 1999), indicating that treatment with antibiotics has not severely disrupted sex allocation of offspring in this species (*A. morosus* can control the sex of its offspring (Spessa 1999)), although low sample sizes limit the

reliability of our sex ratio data. A brief discussion about our recommended amendments and limitations of the methodology are included as Appendix 1. However, the major limitation of antibacterial treatment on this bee species is the difficulty rearing multiple generations in a controlled environment and we currently have no comprehensive solution to address this.

The addition of tetracycline to diets is the most common type of *Wolbachia* treatment, followed by using rifampicin as an alternative antibiotic when tetracycline is not successful (Li et al. 2014). However, exposure to antibiotics is known to cause disruptions to some key biological functions in insects. In honeybees (*Apis mellifera* Linnaeus, 1758), routine antibiotic treatment caused the depletion of numerous symbiotic bacteria resulting in the dysfunction of the immune and digestive systems (Daisley et al. 2020). Antibiotic treatment of host larvae was shown to reduce survivorship of reared endoparasitoid *Pimpla turionellae* (Linnaeus, 1758) (Hymenoptera: Ichneumonidae), particularly those exposed to the highest neomycin dose (0.5 g/100 g of diet) (Harmancı et al. 2019). Similarly to *A. morosus*, after one generation of antibiotic treatment in the thelytokous wasp *Encarsia formosa* Gahan, 1924 (Hymenoptera: Aphelinidae), its *Wolbachia* infection was reduced but not totally removed (Wang et al. 2017). Furthermore, tetracycline (at concentrations of 10 and 20 mg/mL) was found to reduce longevity and oocyte load of F1 wasps compared to the untreated control (Wang et al. 2017). The effects of a full course of antibiotics in *A. morosus* on the microbial, molecular, and reproductive systems is unknown.

4.6 Conclusion

Although our methods of antibiotic therapy were unable to remove the *Wolbachia* co-infection within this bee host within one generation, one *Wolbachia* strain (*w*Amor) showed evidence of being reduced in some offspring of antibiotic-treated mothers. Outside of a few agriculturally important bee species (e.g. Evans (2003), Palmer-Young et al. (2019)), antibacterial treatment in bees has not been applied. A major challenge is that many wild bee species are difficult to cultivate in controlled laboratory settings. Select insect taxa have been extensively investigated for the impacts of *Wolbachia* infections however, in all cases, these species are able to be kept as laboratory colonies for multiple generations (e.g. *Drosophila*, numerous parasitic wasps, and pest Coleoptera and Lepidoptera etc). We have recommended adjustments to our experimental procedures to increase the success of antibiotic therapy on this novel host (Appendix 1). However, a major limitation remains difficult to overcome — rearing multiple generations of *A. morosus* in captivity (especially considering this species in

univoltine). However, to determine whether *Wolbachia* facilitates the occurrence and maintenance of widespread mitochondrial heteroplasmy in this species, antibacterial treatment is the most direct and convincing tool to start answering these complex questions.

4.7 Acknowledgements

This work was supported by the Australian Government Research Training Program Scholarship, well as the Playford Memorial Trust PhD Scholarship. Research funding was provided by the Holsworth Wildlife Trust and the Joyce W. Vickery Fund from the Linnean Society of New South Wales. We would like to thank Diana Fusco for allowing us access to her residence for this study. Also, thank you to Ary Hoffmann for his advice on *Wolbachia* treatment protocols and Martin Davies, Rebecca Dew, Ben Parslow, Bella Davies, Steven Preston, and Jessica Wong for assisting with fieldwork and/or maintenance for the experiment.

4.8 References

Baldo, L., and J. H. Werren. 2007. Revisiting *Wolbachia* supergroup typing based on *WSP*: spurious lineages and discordance with MLST. Current Microbiology **55**:81–87.

Braig, H. R., W. Zhou, S. L. Dobson, and S. L. O'Neill. 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. Journal of Bacteriology **180**:2373–2378.

Chrostek, E., M. S. P. Marialva, S. S. Esteves, L. A. Weinert, J. Martinez, F. M. Jiggins, and L. Teixeira. 2013. *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. PLOS Genetics **9**:e1003896.

Clark, K. A., D. K. Howe, K. Gafner, D. Kusuma, S. Ping, S. Estes, and D. R. Denver. 2012. Selfish little circles: transmission bias and evolution of large deletion-bearing mitochondrial DNA in *Caenorhabditis briggsae* nematodes. PLOS ONE **7**:e41433.

Daisley, B. A., A. P. Pitek, J. A. Chmiel, S. Gibbons, A. M. Chernyshova, K. F. Al, K. M. Faragalla, J. P. Burton, G. J. Thompson, and G. Reid. 2020. *Lactobacillus* spp. attenuate antibiotic-induced immune and microbiota dysregulation in honey bees. Communications Biology **3**:534.

Dedeine, F., F. Vavre, F. Fleury, B. Loppin, M. E. Hochberg, and M. Boulétreau. 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proceedings of the National Academy of Sciences of the United States of America **98**:6247–6252.

Doublet, V., R. Raimond, F. Grandjean, A. Lafitte, C. Souty-Grosset, and I. Marcade. 2012. Widespread atypical mitochondrial DNA structure in isopods (Crustacea: Peracarida) related to a constitutive heteroplasmy in terrestrial species. Genome **55**:234–244.

Duron, O., D. Bouchon, S. Boutin, L. Bellamy, L. Zhou, J. Engelstädter, and G. D. Hurst. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biology **6**:27.

Dyson, E. A., M. K. Kamath, and G. D. D. Hurst. 2002. *Wolbachia* infection associated with all-female broods in *Hypolimnas bolina* (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. Heredity **88**:166–171.
Evans, J. D. 2003. Diverse origins of tetracycline resistance in the honey bee bacterial pathogen *Paenibacillus larvae*. Journal of Invertebrate Pathology **83**:46–50.
Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology **3**:294–299.
Fry, A. J., M. R. Palmer, and D. M. Rand. 2004. Variable fitness effects of *Wolbachia* infection in *Drosophila melanogaster*. Heredity **93**:379–389.
Gahan, A. B. 1924. Some parasitic Hymenoptera with notes on several described forms. Proceedings of the United States National Museum **65**:1–23.
Ghedin, E., T. Hailemariam, J. V. DePasse, X. Zhang, Y. Oksov, T. R. Unnasch, and S. Lustigman. 2009. *Brugia malayi* gene expression in response to the targeting of the *Wolbachia* endosymbiont by tetracycline treatment. PLOS Neglected Tropical Diseases **3**:e525.
Glowska, E., A. Dragun-Damian, M. Dabert, and M. Gerth. 2015. New *Wolbachia* supergroups detected in quill mites (Acari: Syringophilidae). Infection, Genetics and Evolution **30**:140–146.
Graham, R. I., and K. Wilson. 2012. Male-killing *Wolbachia* and mitochondrial selective sweep in a migratory African insect. BMC Evolutionary Biology **12**:204.
Harmancı, C., K. Büyükgüzel, and E. Büyükgüzel. 2019. The effect of neomycin on survival and development of *Pimpla turionellae* L. (Hymenoptera: Ichneumonidae) reared on a natural host. Journal of Economic Entomology **112**:1081–1088.
Hearn, L. R., K. A. Williams, M. I. Stevens, M. P. Schwarz, O. K. Davies, and B. A. Parslow. 2019. Description and novel host records for a new species of Australian mutillid wasp (Hymenoptera: Mutillidae) from hylaeine bee nests (Hymenoptera: Colletidae). Austral Entomology **58**:524–532.
Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences of the United States of America **101**:14812–14817.
Hertig, M., and S. B. Wolbach. 1924. Studies on *Rickettsia*-like micro-organisms in insects. The Journal of Medical Research **44**:329–374.
Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.
Jiggins, F. M. 2003. Male-killing *Wolbachia* and mitochondrial DNA: sselective sweeps, hybrid introgression and parasite population dynamics. Genetics **164**:5–12.
Johnson, K. N. 2015. The impact of *Wolbachia* on virus infection in mosquitoes. Viruses **7**:5705–5717.
Li, Y. Y., K. D. Floate, P. G. Fields, and B. P. Pang. 2014. Review of treatment methods to remove *Wolbachia* bacteria from arthropods. Symbiosis **62**:1–15.
Linnaeus, C. 1758. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. 10th edition, Laurentius Salvius: Stockholm.
Ma, H., H. Xu, and P. H. O'Farrell. 2014. Transmission of mitochondrial mutations and action of purifying selection in *Drosophila melanogaster*. Nature Genetics **46**:393.

Marcadé, I., R. Cordaux, V. Doublet, C. Debenest, D. Bouchon, and R. Raimond. 2007. Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda: Crustacea). Journal of Molecular Evolution **65**:651–659.

Palmer-Young, E. C., L. Ngor, R. Burciaga Nevarez, J. A. Rothman, T. R. Raffel, and Q. S. McFrederick. 2019. Temperature dependence of parasitic infection and gut bacterial communities in bumble bees. Environmental Microbiology **21**:4706–4723.

Raychoudhury, R., B. K. Grillenberger, J. Gadau, R. Bijlsma, L. van de Zande, J. H. Werren, and L. W. Beukeboom. 2010. Phylogeography of *Nasonia vitripennis* (Hymenoptera) indicates a mitochondrial-*Wolbachia* sweep in North America. Heredity **104**:318–326.

Rohrscheib, C. E., E. Bondy, P. Josh, M. Riegler, D. Eyles, B. van Swinderen, M. W. Weible, and J. C. Brownlie. 2015. *Wolbachia* influences the production of octopamine and affects *Drosophila* male aggression. Applied and Environmental Microbiology **81**:4573.

Ros, V. I., V. M. Fleming, E. J. Feil, and J. A. Breeuwer. 2009. How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae). Applied and Environmental Microbiology **75**:1036–1043.

Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S. Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffer, W. Arthofer, and M. Riegler. 2016. The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Molecular Ecology **25**:1595–1609.

Schwarz, M. P., and L. A. Overholt. 1993. Methods for rearing allodapine bees in artificial nests (Hymenoptera: Anthophoridae). Australian Journal of Entomology **32**:357–363.

Smith, F. 1879. Descriptions of New Species of Hymenoptera in the Collection of the British Museum, London: British Museum.

Spessa, A. 1999. The behavioural and population ecology of an Australian native bee, *Amphylaeus morosus* Smith (Colletidae: Hylaeinae). Thesis. Australian National University, Australia.

Stouthamer, R., J. A. Breeuwer, and G. D. Hurst. 1999. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Annual Review of Microbiology **53**:71–102.

Stouthamer, R., and F. Mak. 2002. Influence of antibiotics on the offspring production of the *Wolbachia*-infected parthenogenetic parasitoid *Encarsia formosa*. Journal of Invertebrate Pathology **80**:41–45.

Tagami, Y., K. Miura, and R. Stouthamer. 2001. How does infection with parthenogenesis-inducing *Wolbachia* reduce the fitness of *Trichogramma*? Journal of Invertebrate Pathology **78**:267–271.

Vala, F., M. Egas, J. A. J. Breeuwer, and M. W. Sabelis. 2004. *Wolbachia* affects oviposition and mating behaviour of its spider mite host. Journal of Evolutionary Biology **17**:692–700.

Wang, X., L. Qi, R. Jiang, Y. Du, and Y. Li. 2017. Incomplete removal of *Wolbachia* with tetracycline has two-edged reproductive effects in the thelytokous wasp *Encarsia formosa* (Hymenoptera: Aphelinidae). Scientific Reports **7**:44014.

Werren, J. H. 1997. Biology of *Wolbachia*. Annual Review of Entomology **42**:587–609.

Werren, J. H., L. Baldo, and M. E. Clark. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nature Reviews Microbiology **6**:741–751.

Werren, J. H., D. Windsor, and L. Guo. 1995. Distribution of *Wolbachia* among neotropical arthropods. Proceedings of the Royal Society B: Biological Sciences **262**:197–204.

Werren, J. H., and D. M. Windsor. 2000. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? Proceedings of the Royal Society B: Biological Sciences **267**:1277–1285.

Xie, R.-R., X.-L. Chen, and X.-Y. Hong. 2011. Variable fitness and reproductive effects of *Wolbachia* infection in populations of the two-spotted spider mite *Tetranychus urticae* Koch in China. Applied Entomology and Zoology **46**:95–102.

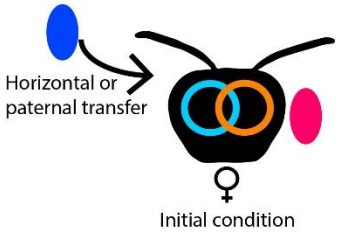
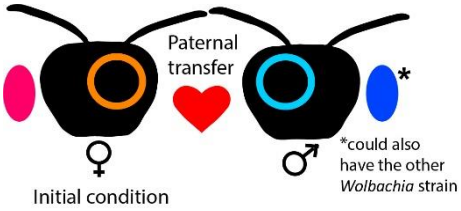
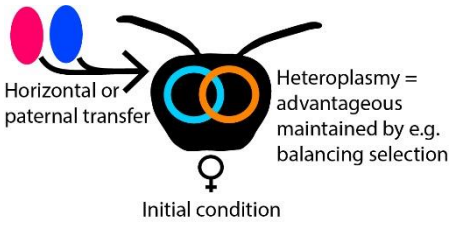
Ye, Y. H., M. Woolfit, E. Rancès, S. L. O'Neill, and E. A. McGraw. 2013. *Wolbachia*-associated bacterial protection in the mosquito *Aedes aegypti*. PLOS Neglected Tropical Diseases **7**:e2362.

Yu, X., and D. H. Walker. 2006. The Order Rickettsiales. Pages 493–528 *in* The Prokaryotes M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt, editors. Springer, New York.

Zhou, W., F. Rousset, and S. O'Neil. 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proceedings of the Royal Society B: Biological Sciences **265**:509–515.

4.9 Tables

Table 1: Three proposed hypotheses that could explain the widespread occurrence of two consistent co-occurring mitochondrial haplotypes (as heteroplasmy) and *Wolbachia* strains in *Amphylaeus morosus*. Each hypothesis is summarized in three steps: (1.) initial condition, (2.) characteristics of the transitional phase in this lineage, and (3.) the outcome — mechanism(s) that continue to maintain present heteroplasmic patterns.

H1: The heteroplasmic founder hypothesis  	(1.) The ancestor of this lineage possessed two mitogenomes (was heteroplasmic — see H3 (1.) for mechanisms) **and** a single *Wolbachia* infection (which may have already swept through the population). (2.) A second *Wolbachia* strain entered this lineage (via horizonal or paternal transfer) and spread to fixation within the population, dragging along both mitogenomes and the initial *Wolbachia* strain. (3.) Heteroplasmy is maintained because of **a mutualism with one of the *Wolbachia* strains**.
H2: The co-inheritance hypothesis  	(1.) Two individuals from separate lineages mated: each with a different *Wolbachia* strain **and** a single divergent mitogenome. (2.) Paternal transmission results in both mitogenomes and *Wolbachia* strains being maintained in the progeny. These co-occurring *Wolbachia* strains and their co-inherited mitogenomes then swept through the population to fixation. (3.) Each mitogenome is maintained by **a mutualism with its corresponding *Wolbachia* strain**.
H3: The advantageous heteroplasmy hypothesis  	(1.) Mitochondrial heteroplasmy had arisen in an ancestral lineage (via either paternal leakage or intraindividual variation maintained for multiple generations). (2.) *Wolbachia* entered this lineage. A selective sweep via either *Wolbachia* and/or an advantageous mutation (heteroplasmy) could be responsible for the overall loss of mitochondrial diversity. (3.) The maintenance of mitochondrial heteroplasmy in *A. morosus* is unrelated to either *Wolbachia* strain. One of the two mitogenomes is never lost (post-fixation) because **heteroplasmy is advantageous over homoplasmy**.

NOTE: The two different *Wolbachia* strains are represented by ovals (pink and blue), and heteroplasmic mitogenomes as rings (orange and light blue) within the bee-head outline representing the founder(s) of a host lineage. The bacterium next to a bee indicates a hypothesised initial infection condition and arrows indicate the transition of novel *Wolbachia* into this host lineage.

Table 2: Primer sets used got generate both host mitochondrial and *Wolbachia* loci used for DNA analyses.

Locus	**Primer Name**	**Sequence**	**Reference**
mt-COI	COIF-PR115 COIR-PR114	5'-TCWACNAAYCAYAARGAYATTGG-3' 5'-ACYTCNGGRTGNCCRAARARYCA-3'	Folmer et al. (1994)
Wolbachia COI-like	Lep-F1 Lep-R1	5'-ATTCAACCAATCATAAAGATAT-3' 5'-TAAACTTCTGGATGTCCAAAAA-3'	Herbert *et al* (2004)
Wolbachia *wsp*	wsp-81F wsp-691R	5'-TGGTCCAATAAGTGATGAAGA-3' 5'-AAAAATTAAACGCTACTCCA-3'	Braig *et al.* (1998); Zhou *et al.* (1998)

Table 3: Kruskal Wallis statistical test results for **experimental** nests only. The impact of nest location (four Adelaide, SA properties) and antibiotic regimen (three doses — 2, 5, 10 mg/mL) on bee activity within nests, specifically changes in the number of adult females (F0) and number of brood (F1) produced during the experiment. No significant differences were observed in any of the parameters compared.

Experimental nest activity	**Location (4 properties in Adelaide, SA)**	**Experimental antibiotic treatment (2, 5, 10 mg/mL)**
Initial number of females (F0) per nest	Kruskal Wallis H3 = 5.277, P = 0.153	Kruskal Wallis H3 = 0.565, P = 0.754
Re-checked (two weeks later) number of females (F0) per nest	Kruskal Wallis H3 = 2.137, P = 0.545	Kruskal Wallis H3 = 0.307, P = 0.858
Difference between initial and re-checked number of adult females (F0) per nest	Kruskal Wallis H3 = 1.300, P = 0.729	Kruskal Wallis H3 = 2.087, P = 0.352
Final number of adult females (F0) per nest	Kruskal Wallis H3 = 3.720, P = 0.293	Kruskal Wallis H3 = 3.276, P = 0.194
Number of brood (F1) produced per nest	Kruskal Wallis H3 = 0.475, P = 0.924	Kruskal Wallis H3 = 1.492, P = 0.474

Table 4: Test statistics and p-values for statistical comparisons of **experimental and control** *Amphylaeus morosus* nests. The effects of nest location (control — Dandenong Ranges, VIC, and experimental — 4 Adelaide, SA properties) and antibiotic treatment regimen (specific regimen — untreated (control), 2, 5, 10 mg/mL, and broadly grouped — control versus all experimental combined) were assessed for number of remaining adult females (W0/F0), brood (W1/F1) produced per nest and sex ratio of W1/F1 offspring in occupied nests at the termination of the experiment. Statistical tests that were significant are bolded with an asterisk.

Complete nest activity at termination	**Location ('wild' (control) and 4 Adelaide properties)**	**Specific antibiotic treatment (0 (control), 2, 5 or 10 mg/mL)**	**Broad antibiotic treatment (control vs experimental)**
Final number of females (W0/F0) per **occupied** nest *(excluding all nests with no adult females AND no brood cells at termination)*	Kruskal Wallis H3 = 16.333, **P = 0.003***	Kruskal Wallis H3 = 2.766, P = 0.429	Mann-Whitney U = 372, P = 0.208
Number of brood (W1/F1) produced per **occupied** nest *(excluding all nests with no adult females AND no brood cells at termination)*	Kruskal Wallis H3 = 8.410, P = 0.078	Kruskal Wallis H3 = 13.242, **P = 0.004***	Mann-Whitney U = 281, **P = 0.012***
Sex ratio of offspring (W1/F1)	Kruskal Wallis H3 = 3.511, P = 0.476	Kruskal Wallis H3 = 3.339, P = 0.342	Mann-Whitney U = 228.5, P = 0.152

4.10 Figures

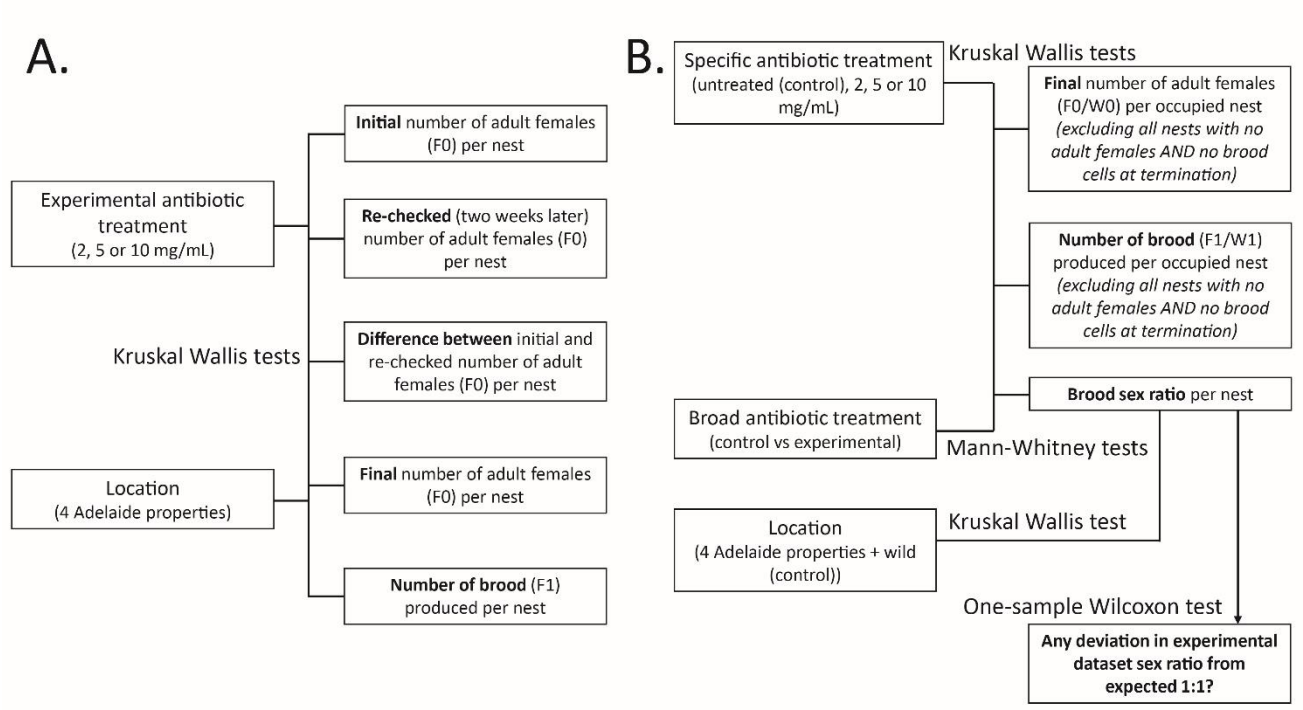


Figure 1: Summary of the statistical tests performed on all parameters in SPSS version 25.0. The data was subset into A. location and antibiotic treatment within the **experimental** nests only, and B. location and antibiotic treatments (specific regimen and broadly grouped) for **experimental and control (untreated)** nests. The impact of these parameters was assessed for (i) the number of females per nest during specific stages of the experiment, (ii) the changes in the number of adult females (W/F0) per nest in the early stages of the experiment, (iii) the number of brood (W1/F1) produced per nest, and (iv) the sex ratio of the brood (W1/F1) produced at the termination of the experiment.

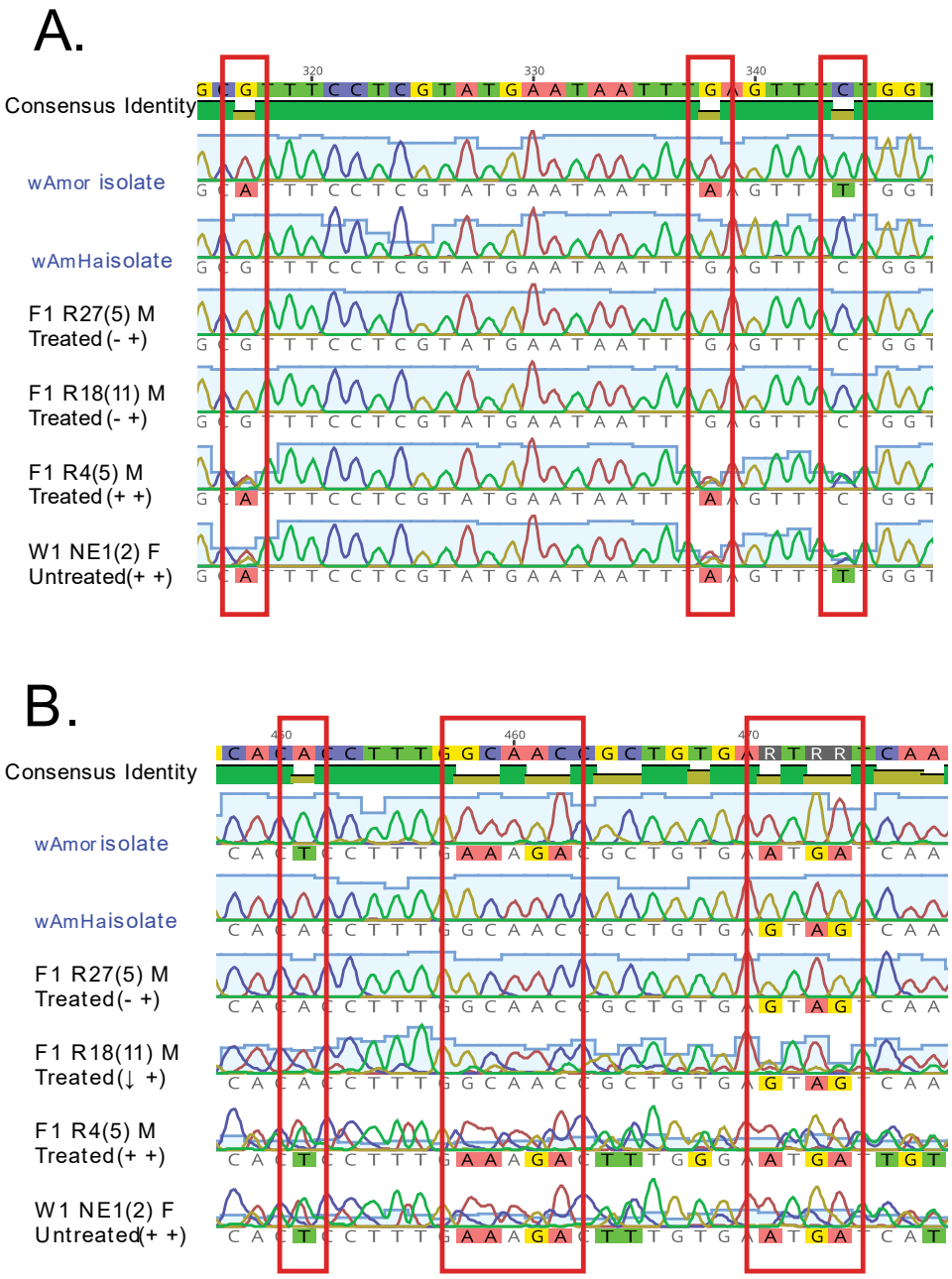


Figure 2: Chromatogram evidence of a reduction of *Wolbachia* strain *w*Amor but persistence of *w*AmHa in *Amphylaeus morosus* F1 offspring (male, "M" and female, "F") from tetracycline-treated (10 mg/mL) F0 mothers. Known isolates of each *Wolbachia* strain were used for comparison of the range of outcomes, post-antibiotic treatment. An example of an antibiotic-treated F1 individual that showed strong evidence of *w*Amor reduction (− +), an equivocal reduction (↓ +), no reduction (+ +), and an *untreated* (control) W1 individual (+ +). Red boxes demonstrate examples of sites of variation in **A.** COI-like *Wolbachia* gene where heterogeneous bases can be seen in co-infected individuals but not in those "successfully" treated, and **B.** *wsp* gene where ambiguous sequence is prominent in co-infected individuals, but not in "successfully" treated individuals (note that the *wsp* gene is difference lengths in these *Wolbachia* strains, hence sequences overlap incongruently when both are present in the sample).

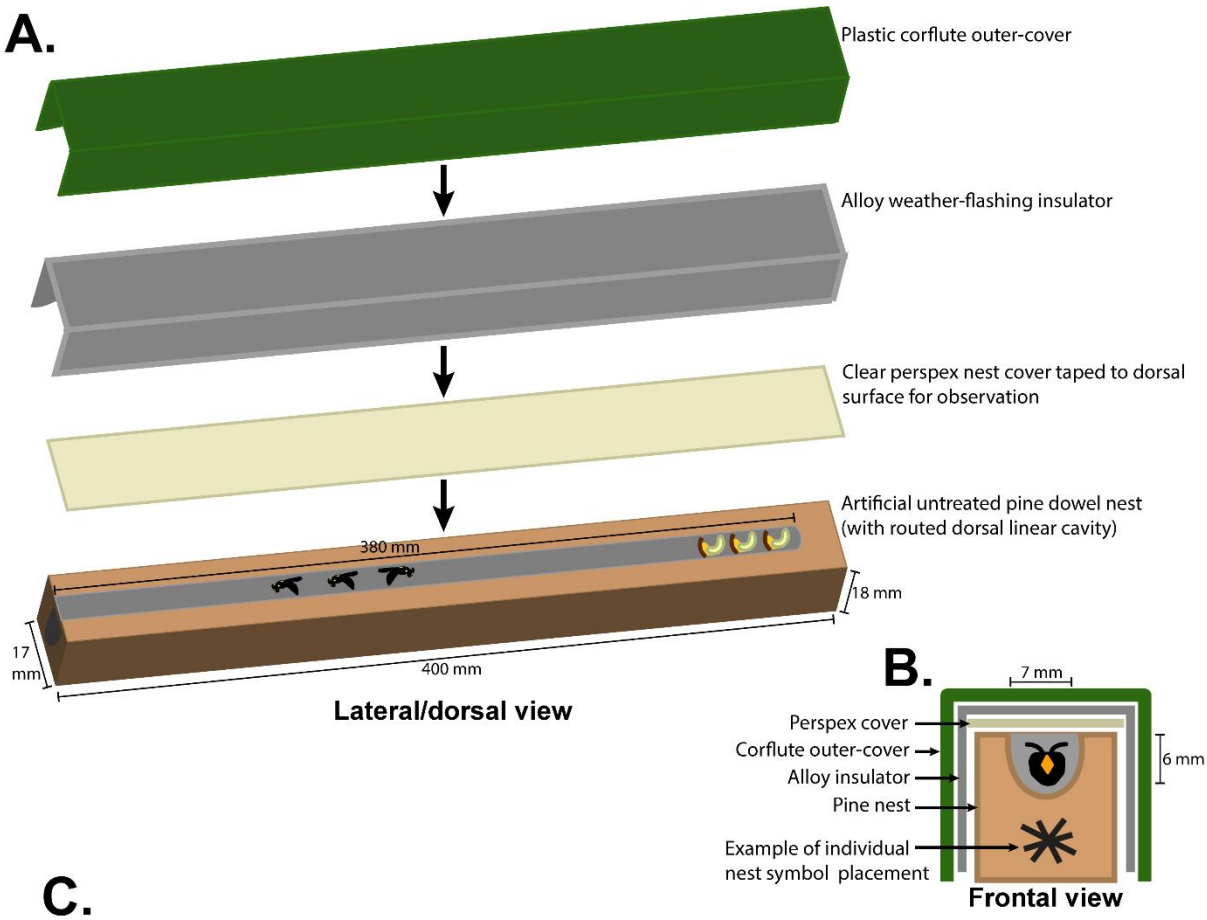
4.11 Supplementary Material

Supp. Table 1: Dunn-Bonferroni post hoc tests for the number of brood (W1/F1) produced in occupied nests across each antibiotic treatment applied (untreated control and doses of 2, 5 and 10 mg/mL with n = 29, 8, 9, and 14 nests, respectively). Asymptotic significance (2-sided tests) values adjusted by the Bonferroni correction for multiple tests are displayed where significant differences are bolded, with an asterisk. Significant level is 0.05.

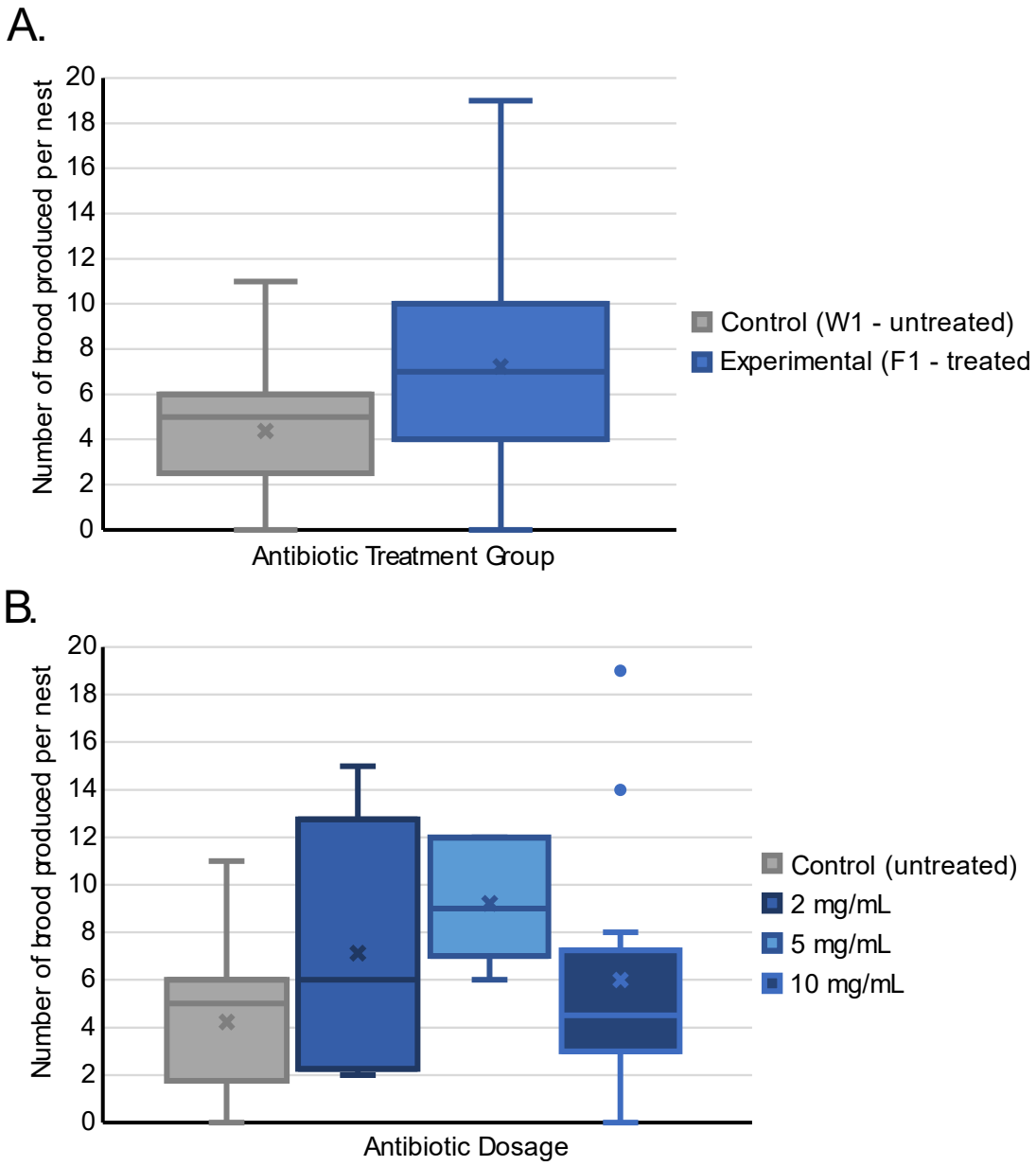
Sample comparisons	**Test statistic**	**Standard error**	**Adjusted significance (Bonferroni corrected)**
0 vs 2 mg/mL	-9.25	6.94	1.0
0 vs 5 mg/mL	23.76	6.63	**0.002***
0 vs 10 mg/mL	-4.35	5.66	1.0
2 vs 5 mg/mL	14.51	8.45	0.515
2 vs 10 mg/mL	4.90	7.71	1.0
5 vs 10 mg/mL	19.41	7.43	0.054

Supp. Table 2: Dunn-Bonferroni post hoc tests for the number adult females (F0) remaining in occupied nests at each location at the termination of the experiment. These sites include the control (wild/untreated from Dandenong Ranges, VIC), and four properties in Adelaide, SA including Eden Hills, Belair, Macclesfield, and Flinders University in Bedford Park. "(*Integer*)" indicates the label given to each property in the analysis and for easier comparison between properties. Sample sizes of nests (n) at each location was 29, 7, 8, 8, and 8, respectively. Asymptotic significance (2-sided tests) values adjusted by the Bonferroni correction for multiple tests are displayed where significant differences are bolded, with an asterisk. Significant level is 0.05.

Property comparisons	**Test statistic**	**Standard error**	**Adjusted significance (Bonferroni corrected)**
Control vs Eden Hills (1)	-20.32	6.69	**0.024***
Control vs Belair (2)	-7.80	6.35	1.0
Control vs Macclesfield (3)	10.89	6.35	0.862
Control vs Flinders Uni (4)	-5.36	6.35	1.0
Eden Hills (1) vs Belair (2)	-12.52	8.23	1.0
Eden Hills (1) vs Macclesfield (3)	-31.21	8.23	**0.001***
Eden Hills (1) vs Flinders Uni (4)	-14.96	8.23	0.690
Belair (2) vs Macclesfield (3)	-18.69	7.95	0.187
Belair (2) vs Flinders Uni (4)	-2.44	7.95	1.0
Macclesfield (3) vs Flinders Uni (4)	16.25	7.95	0.409



Supp. Figure 1: Artificial nests developed for *Amphylaeus morosus* for controlled antibiotic treatment. A. Layers of material for observation and protection of artificial nests with *A. morosus* colonies. B. Frontal view of artificial nest appearance. C. Examples of antibiotic-treated artificial nests with active *A. morosus* colonies distributed through vegetation at Adelaide properties for observation during brood production.



Supp. Figure 2: Boxplots of the number of brood (W1/F1) produced in occupied nests across each antibiotic treatment applied. A. Broad antibiotic treatment — control/untreated (n = 29 nests) and experimental/treated (n = 31 nests), and B. specific antibiotic treatments — control (untreated) and doses of 2, 5 and 10 mg/mL with n = 29, 8, 9, and 14 nests, respectively.

4.12 Appendix 1: Future amendments to the antibiotic treatment methodology

I will discuss the limitations of the antibiotic treatment methodology used in *Amphylaeus morosus* for *Wolbachia* removal. This will include both justifications for the implemented protocols or proposed improved/alternative methodology. This content will be discussed in the order that the methodology is presented. The extra information provided here is specific but may provide insights for those interested in attempting a similar experiment.

4.12.1 Selecting nesting material and encouraging natural nesting behaviour

For successful reproduction in *A. morosus* (and other stem-nesting bees), using nesting materials that are both structurally and compositionally similar to natural nests is recommended. Originally, I used untreated pine dowels for nesting material due to its durability, accessibility, and affordability. However, I would recommend using balsa wood or another soft wood as an alternative. Females of this species naturally construct their own nests (Spessa 1999) by using their mandibles to remove the softer pithy internal contents of their natural nesting substrates (*Cyathea australis* and *Xanthorrhoea* spp.). I would recommend pre-routing a nesting chamber as described, but females would now be able to modify the length and the width of the nesting chamber. Additionally, the excess balsa shaving removed by the females can be used as nest building material with which they create 'plugs' between some brood cells (Hearn et al. 2019).

4.12.2 Location of the experiment

Most *Wolbachia* researchers would expect antibacterial treatments for *Wolbachia* depletion to be conducted under strict laboratory conditions, and I am not aware of any studies in which antibiotic treatment for a *Wolbachia* infection was attempted under 'open' conditions. However, unlike some other bee species (e.g. Maeta et al. (1985), Fauser-Misslin et al. (2014)), *A. morosus* appears to be unable to be reared in confined conditions. When attempted, *A. morosus* was consistently observed to agitatedly fly at parameter material until exhausted and death followed, despite successes with other bee species within the same type of confinement (Schwarz pers. comm.). It is possible that individuals who eclose within confined conditions may behave differently, but this would need to be trialled. Additionally, little is known about the food (floral) preferences of this species (Houston 1975), which complicates assigning a containment protocol.

4.12.3 Grouping the treatments and using individual identifiers

For this species, I would not recommend mixing treatment groups because females were observed to switch nests. Although my intension was that by mixing treatment groups over four treatment sites any local confounding effects on the experiment would be mitigated, I would not recommend this procedure for this species in future. Furthermore, I would recommend marking each specimen with an individual identifier to monitor their movement between nests. I used enamel paints (three colours corresponding to each antibiotic dose) to create a small mark on the dorsal side of mesosoma. However, this protocol became problematic as some markings were lost and females were likely switching into nests with both the same and different treatment colours. Numbered tags glued onto the mesosoma is a possible option (like those used in honeybees (Switzer and Combes 2016)), however the tags I trailed were too large for this species. Our group has since used combinations of two colours of enamel paint placed side by side on the scutum of each individual which has been successful.

4.12.4 Timing of antibiotic treatment

In laboratory settings, antibiotic therapy to remove *Wolbachia* with a feeding solution is commonly performed on either larval or adults with variable success (Li et al. 2014). Application of antibiotics at the larval stage in *A. morosus* could increase the successful of completely removing *Wolbachia*, particularly in developing gametes. However, the lifecycle of this hylaeine bee likely inhibits applying treatment in this development stage.

Firstly, hylaeine bees produce a water-resistant cell lining, in which liquid food provisions are deposited from the crop of the mother and an egg is placed before the cell is sealed (Almeida 2008). Once larvae have consumed all provisions and have reached their final instar before pupating, these cells can be unsealed, and larvae can be reared to adulthood externally with very low mortality rates. However, I have never successfully reared any hylaeine larvae taken out of a sealed cell when wet provisions are still available, and these larvae appear very sensitive to disruptions of their cell during their growth. Transferring antibiotics into the food provisions of developing larvae via e.g. a needle injection, could be trialled, however natural nests occur in hard, woody material and cell partitions would need to somehow be mapped before application. Therefore, accessing the provisions to treat with antibiotics without affecting larval development would be very difficult and could risk high rates of mortality.

Secondly, if antibiotic therapy were applied during the larval stage, mating would need to take place in captivity and females would need to then be encouraged to disperse within the confinements of the provided artificial nesting sites where they will overwinter. *Amphylaeus morosus* is very unusual for hylaeine bees, in that females overwinter as adults (mating occurs in late summer — early autumn and then the males perish), whereas I have observed most other hylaeine bees to overwinter as late-stage larvae. The potential female dispersal events pre- and post-winter would leave the experiment vulnerable to mass loss of individuals. Therefore, I suggest that the antibiotic treatment in *A. morosus* (and other hylaeines) should take place soon after mating has occurred as adult females are establishing their nests. In conclusion, careful investigation is needed into the most effective introduction of antibiotics into this system as well as regimen of the treatment throughout the relevant stages.

4.12.5 Mechanics of antibacterial treatment

Less commonly, extreme (high or low) temperatures are used to removed *Wolbachia* (Li et al. 2014). However, these temperature-based methods often take more generations to remove the infection and it can be difficult to find an optimum temperature that is effective but does not damage the host (Pijls et al. 1996, Kyei-poku et al. 2003, Li et al. 2014). Therefore, antibiotic treatment is a preferable option. To treat adult females, antibiotics were presented in a diluted honey solution, which I would not recommend. Australian honeybees (and therefore Australian honey) are known to harbour pathogens (Giersch et al. 2009, Roberts et al. 2017) but are not known to possess considerable *Wolbachia* infections (Wenseleers and Billen 2000). Exposing *A. morosus* or any other bee species to honey could present an opportunity introduce honeybee pathogens (e.g. Purkiss and Lach (2019)), which could affect their performance in these trials. I could recommend using a diluted sucrose solution (1:1 ratio) as a delivery medium.

Individuals were observed to interact with the pipettes and although I attempted to monitor the consumption of the treatment solution, I could not assess (i) whether each bee in the nest consumed solution, (ii) how much solution each individual was consuming, and (iii) whether the volume was reduced by evaporation or leaking. I recommend that a pilot study should be conducted where the antibiotic solution is coloured with dye and these specimens are destructively sampled to assess if they are imbibing the solution. Ideally, each individual should be exposed separately to the antibiotic solution to better monitor their consumption.

Our highest tetracycline concentration (10 mg/mL) did not appear to affect nesting behaviour in this species; however, the sample sizes were low and the effects of the antibiotic on other biological factors were not assessed. The above pilot study could be paired with a more thorough assessment of the impact of different antibiotic doses on adult females (before brood production). A quantitative assessment of *Wolbachia* depletion in somatic and gonadal tissues, as well as differences in ovary size and survivorship between different treatment regimens would establish which antibiotic concentration is the most suitable for this species.

4.12.6 Placement of nests and managing nest occupation

When placed in the experimental plots, I placed nest at 45° to 90° to prevent waterlogging. In a natural setting, these nests would be vertical (180°), with entrances protected by canopy vegetation. I recommend that artificial nests replicate this, rather than being placed horizontally. The specimens used in this experiment were originally from the Dandenong Ranges, VIC where they naturally nest in fallen fern fronds of *Cyathea australis* (R.Br.) Domin. These large fronds cluster around the base of the fern and I usually observe multiple *A. morosus* nests under the same tree. Therefore, future attempts should replicate these natural conditions (consider that nest switching is likely). My previous research has demonstrated that *A. morosus* are unable to recognition nestmates (Davies et al. 2021) and individuals do not necessarily nest with close genetic relatives (Spessa et al. 2000). Therefore, the consideration of which colonies to include in each cluster is not necessary, however control and treatment groups cannot be mixed.

Additionally, nest supersedure by other invertebrates was a major problem for these artificial nests (particularly by ants). Monitoring and intervention of these events is important to discourage nest abandonment. Unfortunately, I did not record which nest placements were the most successful however, generally nests closer to the ground (<1 m) were observed to be invaded by ants. Including additional empty artificial nests in the area might alleviate the impact of nest invasion or abandonment due to poor nest placement.

4.12.7 Maintaining multiple generations

Amphylaeus morosus, like many other wild bee species, is difficult to keep in captivity. Additionally, this species is univoltine and only produces one set of brood per year under natural climate conditions (Spessa 1999). Maintaining colonies for multiple generations, particularly under controlled conditions would be an arduous task. It is likely that I did not

observe any effect on the presence of mitochondrial heteroplasmy because changes to patterns of mitochondrial inheritance would need to occur over multiple generations.

Keeping colonies of *A. morosus* even in an 'open' setting (outside of their home range to prevent integration) would also be incredibility difficult but would likely be more successful in rearing multiple generations compared to controlled, laboratory containment. It is worth considering the risk that *A. morosus* may naturalize outside of an allocated 'open' experimental plot. However, this species is likely dependant on appropriate nesting substrate (habitat) to persist (Houston 1975, Spessa 1999), which has been considerably reduced across eastern Australia (Chapter 5). I do not believe this species can be successfully maintained in over multiple generations in either a contained or 'open' captive setting (and isolated from naturally occurring populations) without extensive development of appropriate animal husbandry protocols.

4.12.8 References

Almeida, E. A. B. 2008. Colletidae nesting biology (Hymenoptera: Apoidea). Apidologie **39**:16–29.

Davies, O. K., M. G. Gardner, and M. P. Schwarz. 2021. No evidence for nestmate recognition in the only social bee in the family Colletidae, with implications for very early stages in social evolution. *In preparation*

Domin, K. 1929. Pteridophyta (in Czech). Nákladem České Akademie Vêd a Umêni. , Praha, Czech Republic

Fauser-Misslin, A., B. M. Sadd, P. Neumann, and C. Sandrock. 2014. Influence of combined pesticide and parasite exposure on bumblebee colony traits in the laboratory. Journal of Applied Ecology **51**:450–459.

Giersch, T., T. Berg, F. Galea, and M. Hornitzky. 2009. *Nosema ceranae* infects honey bees (*Apis mellifera*) and contaminates honey in Australia. Apidologie **40**:117–123.

Hearn, L. R., K. A. Williams, M. I. Stevens, M. P. Schwarz, O. K. Davies, and B. A. Parslow. 2019. Description and novel host records for a new species of Australian mutillid wasp (Hymenoptera: Mutillidae) from hylaeine bee nests (Hymenoptera: Colletidae). Austral Entomology **58**:524–532.

Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.

Kyei-poku, G. K., K. D. Floate, B. Benkel, and M. S. Goettel. 2003. Elimination of *Wolbachia* from *Urolepis rufipes* (Hymenoptera: Pteromalidae) with heat and antibiotic treatments: Implications for host reproduction. Biocontrol Science and Technology **13**:341–354.

Li, Y. Y., K. D. Floate, P. G. Fields, and B. P. Pang. 2014. Review of treatment methods to remove *Wolbachia* bacteria from arthropods. Symbiosis **62**:1–15.

Maeta, Y., S. F. Sakagami, and C. D. Michener. 1985. Laboratory studies on the life cycle and nesting biology of *Braunsapis sauteriella*, a social xylocopine bee (Hymenoptera: Apidae). Sociobiology **10**:17–41.

Pijls, J. W. A. M., H. J. van Steenbergen, and J. J. M. van Alphen. 1996. Asexuality cured: the relations and differences between sexual and asexual *Apoanagyrus diversicornis*. Heredity **76**:506–513.

Purkiss, T., and L. Lach. 2019. Pathogen spillover from *Apis mellifera* to a stingless bee. Proceedings of the Royal Society B: Biological Sciences **286**:20191071.

Roberts, J. M. K., D. L. Anderson, and P. A. Durr. 2017. Absence of deformed wing virus and *Varroa destructor* in Australia provides unique perspectives on honeybee viral landscapes and colony losses. Scientific Reports **7**:6925.

Spessa, A. 1999. The behavioural and population ecology of an Australian native bee, *Amphylaeus morosus* Smith (Colletidae: Hylaeinae). Thesis. Australian National University, Australia.

Spessa, A., M. P. Schwarz, and M. Adams. 2000. Sociality in *Amphylaeus morosus* (Hymenoptera: Colletidae: Hylaeinae). Annals of the Entomological Society of America **93**:684–692.

Switzer, C. M., and S. Combes. 2016. *Bombus impatiens* (Hymenoptera: Apidae) display reduced pollen foraging behavior when marked with bee tags vs. paint. Journal of Melittology **62**:1–13.

Wenseleers, T., and J. Billen. 2000. No evidence for *Wolbachia*-induced parthenogenesis in the social Hymenoptera. Journal of Evolutionary Biology **13**:277–280.

Chapter 5: Population structure in a native bee species with no mitochondrial variation?

Olivia K. Davies[a*], James B. Dorey[a,b], Lucas R. Hearn[a], and Michael P. Schwarz[a]

[a]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide SA 5001, Australia

[b]Biological and Earth Sciences, South Australian Museum, SA 5000, Australia

*Corresponding author. Email: davies.oliviak@gmail.com

Keywords: Hymenoptera, isolation by distance, *Wolbachia*, mitochondrial sweep, endosymbiont, GENELAND, TESS3

5.1 Abstract

The native hylaeine bee *Amphylaeus morosus* (Hymenoptera: Colletidae) occurs across the eastern coast of Australia, spanning a north-south ~2,000 km geographical range. The characteristics of this bees' mitochondrial DNA (mtDNA) are unusual, possessing little haplotype diversity across its entire distribution except for consistent, widespread mitochondrial heteroplasmy (presenting as two mitogenomes in every single individual). *Amphylaeus morosus* is also consistently infected with two strains of the insect endosymbiont *Wolbachia* (Rickettsiales: Anaplasmataceae). The combination of ultra-conserved mitochondrial heteroplasmy and a widespread *Wolbachia* co-infection is perplexing. These traits could suggest either the loss of mtDNA variation is due to (i) a population bottleneck followed by a massive range expansion or (ii) a historic species-wide mitochondrial selective sweep facilitated by the *Wolbachia* infection(s). We used genome-wide single nucleotide polymorphism data (SNPs) generated using the DArTseq™ platform for individuals collected aross its geographical range to assess for evidence of population differentiation along its distribution. We sampled 52 individuals from six locations, and found pronounced isolation by distance, as well as strong genetic clustering indicating population structure from nuclear (but not mitochondrial) DNA data. Furthermore, we used habitat modelling to show that pre-European settlement in Australia, *A. morosus* habitats were better connect than at present. These two pieces of evidence suggest that a lack of mtDNA diversity observed in this species is more likely the result of a *Wolbachia*-induced selective sweep across the whole species' distribution in the pre-European past, rather than a recent population bottleneck. This provides further, but indirect, evidence that *Wolbachia* has played a key role in the evolution the unusual *A. morosus* mtDNA traits.

5.2 Introduction

Population genetic analyses are powerful tools for understanding patterns of gene flow and the divergence of genetic lineages within species. However, several studies have shown that traditional genetic methods are susceptible to several biases and weaknesses which can make the generation of reliable results difficult (Bossart and Pashley Prowell 1998, Zwickl and Hillis 2002, Teske et al. 2018). In animals, mitochondrial DNA (mtDNA) data has been widely used to assess evolutionary relationships because, generally, mtDNA can be easily amplified, establishing homology of gene sequences is often straightforward, inheritance patterns are generally simple, and it usually exhibits higher evolutionary rates compared to most nuclear markers (Moritz et al. 1987, Harrison 1989). However, mtDNA has limitations to its utility including a small genome size which is susceptible to biased base composition, occurrences of paralogous nuclear loci (Lin and Danforth 2004), and a lack of recombination that regulates selfish genomes (those that have replication and transmission advantage over other genomes) (Barr et al. 2005). Additionally, mtDNA can be influenced by independent selective forces (e.g. co-inheritance with parasitic endosymbionts) causing a break-down of linkage disequilibrium creating discrepancies between mitochondrial and nuclear DNA lineages (Hurst and Jiggins 2005). In an attempt to overcome these limitations, high-throughput, large-scale sequencing arrays have been used to generate multi-locus data which densely span whole genomes (Melville et al. 2017) and can be widely applied to non-model organisms (Ekblom and Galindo 2011).

The use of SNP markers is being increasingly applied to assessments of gene flow, population structure, and delimiting boundaries of conservation relevance in wild species — particularly those with large distributional ranges and where individuals occur at low densities (e.g. Park et al. (2015), Viengkone et al. (2016), Torati et al. (2019), Di Santo and Hamilton (2021)). A complicating factor is that genetic variation is expected to be influenced by geographic distances between individuals or populations (Guillot et al. 2009). This isolation by distance (IBD) is where geographic distance limits dispersal and thus reduces mating probabilities between individuals, particularly in species with linear distributions where heterogeneous gene flow cannot occur (Wright 1943, van Strien et al. 2015, Perez et al. 2018). Isolation by distance is known to over-inflate the interpretation of genetic clusters (i.e. population structure), termed the “cline versus cluster dilemma” (Guillot et al. 2009), particularly in cases of linear geographical distributions. In datasets where IBD has been

identified, usually by Mandel tests (Guillot et al. 2009), special consideration of appropriate clustering models (models detecting population structure) needs to be made which consider the impact of geographic distance on population assemblages (Frantz et al. 2009, Guillot et al. 2009, Perez et al. 2018).

The Australian bee species *Amphylaeus morosus* (Hymenoptera: Colletidae) occurs in highland wet-montane and lowland, sub-coastal heathland habitats across a ~2,000 km cline along the eastern Australian coast (Houston 1975). Our previous data have indicated that there is little mtDNA diversity (excluding mitochondrial heteroplasmy) in *A. morosus* across its entire geographical range (Chapter 2), which is unexpected given its large population range from southern Queensland to western Victoria. *Amphylaeus morosus* individuals also possess a consistent co-infection of the common insect parasite *Wolbachia*. We have hypothesised that this unusual lack of geographical variation in mtDNA is best explained by (i) a mitochondrial selective sweep occurring throughout the whole species, possibly caused by a widespread *Wolbachia* infection, rather than (ii) a population bottleneck in recent history followed by a massive range expansion.

For *Wolbachia* (and the consistent heteroplasmic mtDNA 'haplotype') to occur in every individual across the entire *A. morosus* distribution would require a well-connected, widespread, single population or regular integration from adjacent populations with minimal barriers to geneflow. Given that *A. morosus* appears to exhibit habitat specificity and individuals are not readily encountered/collected throughout most of its fragmented distribution (Houston 1975, Spessa et al. 2000, ALA 2020), high connectivity seems unlikely. However, the ecology of *A. morosus* is largely unknown, it is difficult to infer its current extent of occurrence, let alone its historic population characteristics. Here we use genome-wide single nucleotide polymorphisms (SNPs) to explore geographical population genetic structure in *A. morosus* and examine whether the pattern of mitochondrial uniformity across this species' range is concordant with nuclear DNA (nuDNA) variation. Using these data, we infer the most likely history of the unusual uniformity observed in mtDNA across the *A. morosus* distribution.

5.3 Methods

5.3.1 Specimen collection

Amphylaeus morosus specimens were collected from six locations across the eastern coast of Australia between 2013 and 2018 (Supp. Table 1). Specimens collected in the northern and south-western limits (Tin Can Bay and Cobboboonee region, respectively) were collected in dead *Xanthorrhoea* spp. flower scapes in coastal heathland. Specimens collected from the central locations (north to south; Enfield State Forest, Blue Mountains, and Snowy Mountains, and the Dandenong Ranges) were collected from nests of dead and excised fronds of the tree fern *Cyathea australis* in wet montane forests. Any immature brood were reared in Petri dishes at ambient room temperature to adulthood at Flinders University Bedford Campus, South Australia and all resulting adults were stored in 100% ethanol at ~2˚C prior to DNA sequencing.

5.3.2 Specimen selection and preparation

DNA was extracted from head or thorax tissue of 96 *A. morosus* individuals, completed by Diversity Arrays Technology Pty. Ltd. (Canberra, Australia). Approximately half of these specimens (51) were adult females collected from six locations across the geographic distribution (see below). Head tissue was used from adult females, each was selected from an individual colony (i.e. no two individuals selected were collected in the same nest). The remaining 45 *A. morosus* specimens were from the Dandenong Ranges (from complete nests, including all resident female and male adults and pupae with DNA extracted from thoracic tissue) and were included for a separate project addressing questions on colony pedigrees. Of these additional 45 individuals, three adult females, each from separate colonies, were included in this analysis of population structure.

A preliminary analysis using the DArTseq™ (Diversity Array Technology sequencing) platform had been conducted by Diversity Arrays on five adult females from the Dandenong Ranges comparing the similarity of SNP calls of the head versus thoracic tissues, and to determine the risk of contamination of non-genomic DNA. Head/thorax tissues were used to avoid potentially complicating inclusion of DNA from pollen in the gut. No significant difference was identified between the two tissue types, therefore SNPs from individuals generated from different tissues could be combined. Four of these preliminary individuals (retained after filtering protocols) were included in the final dataset for this analysis.

5.3.3 Genomic analysis

High-throughput microarray sequencing was used to call SNPs applying the protocols described by Jaccoud et al. (2001) using the DArTseq™ platform, completed by Diversity Arrays Technology Pty. Ltd. (Canberra, Australia). DNA was fragmented using the restriction enzymes *PstI* and *SphI* which target low-copy DNA regions throughout the nuclear genome to reduce genome complexity and identify DNA polymorphisms without the need for prior DNA sequencing information (Kilian et al. 2012, Melville et al. 2017). This generated sequences of approximately 69 bp containing an informative single nucleotide polymorphism (SNP) in two alleles from loci across the genome, comparable across individuals (Shams et al. 2019). These reads were processed using an in-house pipeline (Cruz et al. 2013) prior to quality filtering and population analyses. Although to date, DArTseq™ has mostly been applied to plant systems, many studies have also applied this method successfully to numerous animal species (Grewe et al. 2015, Lambert et al. 2016, Melville et al. 2017, Mulvena et al. 2020, van Deventer et al. 2020).

5.3.4 Sequence quality control and loci filtering

A total of 12,587 SNP loci were called across the combined samples and all collection localities. Populations were not sorted by collection locality, but rather SNP quality filtering was conducted on all 52 individuals selected for population analyses — those that met the initial data quality requirements. Final sample sizes for most of locations were small (Tin Can Bay, n = 4, Enfield State Forest, n = 4, Blue Mountains, n = 8, Snowy Mountains, n = 1, Dandenong Ranges, n = 30, and Cobboboonee, n = 5), but reflect the scarcity of sampled *A. morosus* in most localities.

SNP loci were assigned the following DArT scores: “0” = reference allele homozygote, “1” = SNP allele homozygote, and “2” = heterozygote based on their allelic variation. All genetic and population analyses were conducted using packages implemented in R version 3.6.3 (R Development Core Team 2020) via RStudio version 1.3.1056 (RStudio Team 2020). Using the R package dartR version 1.3.5 (Gruber et al. 2018), DArTseq raw data was filtered for repeatability (100%), call-rate (>99%), minor allele frequency (>2%), all secondaries and monomorphic loci removed, and Hamming distance >20%, with 1,364 loci retained.

5.3.5 Analyses of population genetics

Due to low sample sizes in most collection localities, linkage disequilibrium was assessed in two data groupings — combined dataset (n = 52) and Dandenong Ranges population only (n = 30) with $r^2 = 0.4$ and 0.7, respectively, to reflect the estimated degrees of relatedness within the samples. Linked loci that were shared in both the combined and Dandenong Ranges samples were removed from the whole dataset. On these filtered data, we calculated mean observed heterozygosity (H_O) and expected heterozygosity (H_S) using dartR, as well as the inbreeding coefficient (F_{IS}), calculated using the formula $F_{IS} = (H_S - H_O)/H_S$ (Wright 1931) (Supp. Table 2). To assess for differences between mean observed and expected heterozygosity, paired two sample t-tests with a Bonferroni correction were applied to each collection locality dataset (Supp. Table 2). Because of the sample size limitations, Hardy-Weinberg equilibrium was assessed using dartR for only individuals from the Dandenong Ranges (n = 30) where no significant deviation was found within these data. Pairwise F_{ST}'s were calculated using dartR, on the filtered dataset as a measure for genetic distances between our six collection localities.

5.3.6 Non-spatial analyses of population structure

To visualise the genetic similarity between *A. morosus* individuals, we conducted a principal coordinate analysis (PCoA) using a Euclidean distance model (Supp. Figure 1) in dartR (Gruber et al. 2018) on the filtered dataset. We also conducted a discriminate analysis of principal components (DAPC) (Jombart et al. 2010) in the package adegenet version 2.1.3 (Jombart 2008). This multivariant analysis attempts to find clusters (inferable without prior group information by K-means) (Supp. Figure 2) that maximize the genetic differentiation between groups, whist minimizing the effect of within group variation (Jombart et al. 2010). Both analyses allow an estimation of the number of genetic clusters within the *A. morosus* data, but do not consider the effect of spatial parameters driving genetic differentiation.

5.3.7 Population genetic structure with spatial parameters

To assess the degree of IBD (isolation by distance i.e. whether genetic distance is correlated with geographic distance) of the *A. morosus* sampling localities, two Mantel tests were performed. Both were run with 9,999 permutations; firstly using the R package vegan (Oksanen et al. 2019) wrapped in dartR (Gruber et al. 2018) and secondly, in adegenet (Jombart 2008). Because the Mantel tests determined that IBD is a significant factor in our

dataset, to assess for population structure, we used methods that consider both spatial and genetic data in their clustering analyses (Perez et al. 2018).

Markov chain Monte Carlo (MCMC) simulations were conducted using GENELAND version 4.9.2 (Guillot et al. 2005a, Guillot et al. 2005b) in R to estimate genetic boundaries across our *A. morosus* samples. The number of population clusters were estimated by running ten MCMC iterations, modifying the following parameters: 100,000 MCMC iterations per run, with thinning set to 100, and allowing *K* to vary between minimum *K* = 1, maximum *K* = 10. Of these ten MCMC runs, the run with the highest posterior probability was used to infer the number of *A. morosus* populations.

We also used TESS3 version 1.1.0 (Caye et al. 2016) in R. Using the *tess3* function which uses graph-based non-negative matrix factorization to estimate the number of spatial population clusters, we ran 10 replicates for each *K* (allowing *K* to vary between minimum *K* = 1, maximum *K* = 8), with tolerance set to 1×10^{-7} and maximum iterations set to 1,000, using a projected least squares algorithm. Cross-entropy criterion were generated with 5% of genotypes masked, and our optimal *K*-values for exploration were chosen based on cross-validation scores (Caye et al. 2017).

5.3.8 Pre- and post-European habitat modelling

To model the potential habitat of *A. morosus* we combined our own records and those from the Atlas of Living Australia (ALA 2020). We filtered Atlas of Living Australia records following the methods of Dorey et al. (2020) by only including reliable (preserved specimens, machine observations, and those from publications) and present records. These records were then overlaid with both pre-1788 and more-recent vegetation maps of Australia (NMD 2003b, a). All major vegetation subgroups (MVSs) inhabited by *A. morosus* were identified (NMD 2003b, a). The geographic extent of these occurrences was then buffered by 20% for GIS analyses. Major vegetation subgroups were then mapped for both time periods to visualise how habitat might have changed since European clearing. The GIS analyses and visualisations were completed using raster package version 3.3-13 (Hijmans and Etten 2012) and sp package version 1.4-4 (Pebesma and Bivand 2005, Bivand et al. 2013) as implemented in R (R Development Core Team 2020). Fragmentation statistics of inhabited MVSs were calculated using SDMTools package version 1.1-221 (VanDerWal et al. 2014) in R.

5.4 Results

5.4.1 Genetic diversity analyses

The filtered dataset consisted of 1,243 loci retained for 52 adult female *Amphylaeus morosus*, sampled across six locations along the eastern coast of Australia. Paired two sample t-tests demonstrated that the observed heterozygosity was significantly different from the expected heterozygosity in all locations south of the Blue Mountains (Supp. Table 2). Inbreeding coefficients (F_{IS}) ranged from -0.140 for specimens from Cobboboonee (indicating a slight excess in heterozygotes in this sample), to 0.000 in Enfield State Forest, and 0.053 in the Blue Mountains (Supp. Table 2). Pairwise values of F_{ST} ranged from 0.035 (Enfield State Forest versus Blue Mountains) to 0.500 (at the geographic margins; Tin Can Bay versus Cobboboonee) (Table 1).

5.4.2 Non-spatial analyses of genetic clustering

Our PCoA results of the filtered SNP data formed distinct clusters (Figure 1A), with the first principal coordinate explaining 15.6% of variation detected and the first three principal coordinates combined explained 24.4% of the total variation. Data clusters along the first principal coordinate reflected the latitudinal gradient of the sampling localities, and likely represent geographic distance between the genetic clusters. All localities sampled formed distinct clusters along the first principal coordinate, with the exception of the Blue Mountains and Enfield State Forest (i.e. five clusters were recovered from the six sampling localities) (Figure 1A).

The DAPC was also able to recover distinct clusters within the *A. morosus* dataset. Although the BIC plot displayed the lowest value a $K = 2$, BIC scores were not informatively different for $K = 2$ to 4 (Supp. Figure 2). Clustering models were visualised for all K, with $K = 4$ being the most biologically informative (Figure 1B). The four genetic clusters identified were distinct, with the northern-most locality (Tin Can Bay) and south-western-most (Cobboboonee) distinct from the central localities that that clustered into two additional groups; Enfield State Forest + Blue Mountains, and Snowy Mountains + Dandenong Ranges, respectively (Figure 1B).

5.4.3 Incorporation of spatial parameters maintains population structure

Mantel tests to assess IBD were performed in dartR (Gruber et al. 2018) and adegenet (Jombart 2008) found significant evidence of genetic divergence associate with geographic

distance (r = 0.72, P-value = 0.0069, and r = 0.66, P-value = 0.027, respectively). IBD is not compatible with programs that are not spatially explicit, such as STRUCTURE, because it can cause overinflation of genetic clustering (Perez et al. 2018). Therefore, we used alternative programs which included spatial parameters to estimate the number of genetic clusters.

With the incorporation of spatial information, our analyses retained evidence of distinct genetic clusters (population structure). Highest log posterior probabilities generated across the ten simulations in GENELAND indicted five genetic clusters across our *A. morosus* dataset (Supp. Figure 3). These five clusters reflected the same groupings which were recovered in our PCoA, with all localities forming distinct groupings, except for Enfield State Forest and Blue Mountains (Figure 2). Alternatively, cross-validation scores generated in TESS3 indicated $K = 4$ as the most informative number of clusters (Supp. Figure 4) and reflected the groupings recovered by the DAPC analysis; northern- to southern-most clusters being Tin Can Bay, Enfield State Forest + Blue Mountains grouped, Snowy Mountains + Dandenong Ranges grouped, and Cobboboonee (Figure 3). Ancestry proportions of individuals from each locality to their designated cluster were generally distinct, except for the Snowy Mountains individual which also shares ancestry with the Enfield State Forest + Blue Mountains cluster (Figure 3), indicating potential connectivity between these central regions.

5.4.4 Modelling of *Amphylaeus morosus* habitat

The MVSs in which *A. morosus* were found have reduced in total area by ~39% since European settlement (from 11.4×10^9 km^2 to 7.0×10^9 km^2; Figure 4). This habitat has also become more fragmented with the number of patches increasing by a factor of ~4 (from 1.1×10^5 patches to 4.3×10^5 patches) and mean patch size decreasing by ~85% (from 10.9×10^4 km^2 to 1.6×10^4 km^2; Figure 4). The landscape shape index also increased by a factor of ~1.8 (from 377 to 675; Figure 4).

5.5 Discussion

Our study highlights that patterns of population structure can appear different when conflicting selective forces shape the evolution of mitochondrial and nuclear DNA. All analyses conducted on these *Amphylaeus morosus* SNP data (non-spatial and spatial) indicated that there is strong nuclear genetic clustering. We have interpreted this as evidence

of population structure but are aware that our dataset is strongly influence by IBD (Frantz et al. 2009, van Strien et al. 2015, Perez et al. 2018). We selected clustering models that consider spatial parameters (Guillot et al. 2005a, Guillot et al. 2005b, Caye et al. 2016, Caye et al. 2017), however sampling of specimens from areas between these locations (if such populations still exist) may dissolve some of these 'distinct' population boundaries. Nonetheless, the patterns of population assemblages as demonstrated by mtDNA and nuDNA are discordant. Genomic data reflects distributional patterns that would be expected in a species with a vast geographical range, such as *A. morosus*.

Wolbachia has been shown to obscure or mislead patterns of population structure when using mtDNA markers (Sucháčková Bartoňová et al. 2021). The lack of mtDNA variation across the geographical range of *A. morosus*, as well as a widespread *Wolbachia* infection, might first suggest that this species forms a single, large population with few or no barriers to mitochondrial gene flow. However, genomic SNP markers demonstrate clear genetic clustering across the *A. morosus* distribution. In contrast to population bottlenecks, nuclear genes are not expected to be affected by spreading *Wolbachia* infections, because continuous gene flow should be maintained as mating remains successful between uninfected males and infected females (Sinkins and Godfray 2004) — unless *Wolbachia* induces parthenogenesis (Rokas et al. 2001). Our study further highlights that the use of mtDNA can be unreliable for understanding population differentiation (e.g. Bensch et al. (2006), Adams et al. (2013), Teske et al. (2018)), and potentially misinform conservation management of species.

SNP analyses are yet to be widely applied to bee species for population genetic analyses, particularly outside of honeybees. Population structure assessed using microsatellites in the wide-ranging bumble bee, *Bombus ephippiatus*, (Hymenoptera: Apidae) was found to correspond with habitat specificity and body colour variation across the Mesoamerican region (Duennes et al. 2012), but these patterns were not consistently reflected in the mtDNA data. Many studies have identified limited to no population structure in numerous wide-ranging bee species using microsatellite data, including *Amegilla dawsoni* (Beveridge and Simmons 2006), *Andrena vaga* (Exeler et al. 2008, Černá et al. 2013), *Euglossa cordata* (Boff et al. 2014), *Osmia rufa* (Neumann and Seidelmann 2006), and numerous *Bombus* species (Estoup et al. 1996, Lozier and Cameron 2009, Cameron et al. 2011, Françoso et al. 2016). Except in examples where populations were geographically isolated by mountain ranges or oceans separating landmasses (e.g. Widmer et al. (1998), Widmer and Schmid-Hempel (1999), Shao

et al. (2004)), dispersal by long-ranging males (particularly in relatively large species) has been proposed to explain the lack of broad population structure (Françoso et al. 2016). However, the geographical extent of *A. morosus* exceeds that of most species in the above studies. Little is known about the biology and ecology of *A. morosus*; however, reliance on particular nesting substrates and floral resources are likely major barriers to dispersal.

Amphylaeus morosus is unusual, being the only known social species in the family Colletidae — with colonies ranging from solitary, to up to six females together in a single tubular nest cavity (Spessa et al. 2000). Spessa et al. (2000) proposed that limitations in available nesting material — *Cyathea australis* and *Xanthorrhoea* spp. (Spessa 1999) — is the main driver behind the unusual social behaviour of *A. morosus*. Our PCoA results suggest that the main driver of genetic divergence in *A. morosus* might be geographic distance, because clusters along PCo1 (15.6%) coincide with the latitudinal gradient. PCo2 (5%) further separated our marginal sampling locations — Tin Can Bay and Cobboboonee — from the central samples. These populations nest in *Xanthorrhea* spp. and occupy sub-tropical and warm temperate coastal habitats, respectively. In these locations, other similarly sized hylaeine species were collected more frequently. This might indicate niche competition and marginal habitat for *A. morosus*. Environmental differences and competition could be driving selection and divergence from the central localities. However, we did not assess the influence of isolation by environment (IBE) — isolation due to restricted niche tolerances, e.g. climate or habitat, independent of geographic distance (Wang and Bradburd 2014) — as we do not have sufficient environmental data.

Across the four central localities (where *A. morosus* occurs in montane forests, nesting in *C. australis*), our analyses of population structure demonstrated some inconsistencies when classing clusters. All analyses were unable to distinguish between individuals sampled from the Enfield State Forest and Blue Mountains (which are ~300 km apart, the same approximate distance as all localities in our study) (Figures 1–3). Pairwise F_{ST} values between all central localities were <0.132 (Table 1) and might indicate that these montane habitats have some historic connectivity. Although, we did not address fine-scale population structure in our dataset, further research would be beneficial to better-understand the historic population conditions of this species and its role in facilitating the spread of *Wolbachia*, particularly in these montane habitats. To further understand the selective process(es) that

have homogenised the heteroplasmic mtDNA in *A. morosus*, historical demography studies should be conducted to investigate the role of population declines or expansions.

Understanding population connectivity is important for informing conservation management and disease (or parasite) susceptibility (Kool et al. 2013, Lilley et al. 2020). We present three avenues of evidence that support high levels of habitat connectivity for *A. morosus* across its ~2,000 km range in the recent (post-Pliocene) past. Firstly, the lack of mitochondrial diversity (Chapter 2) in *A. morosus* implies that the event that led to the homogenisation of mitochondrial heteroplasmy was able to occur throughout the entire species. Secondly, that a widespread *Wolbachia* co-infection could reach fixation in *A. morosus* (Chapter 3) also supports historically, high levels of habitat connectivity. Finally, our habitat modelling supports high pre-European habitat connectivity (Figure 4). However, our fragmentation analyses indicated a major reduction in habitat area (39%) and connectivity (four-fold increase in number of patches with mean patch size decreasing by ~85%) since European settlement in Australia (Figure 4). The landscape shape index (a measurement of patch aggregation) demonstrates *A. morosus* patches have become more disjunctive and could be an indication that it will be more difficult for individuals to move between patches — inhibiting gene flow and replenishment of 'sink' populations (Howe et al. 1991). Given the possible habitat dependency of this species and habitat loss across its range, it seems unlikely that population connectivity is currently maintained. The value of *A. morosus* habitats should be considered when making development and conservation decisions.

The Hylaeinae are highly ecologically successful, occupying all habitat types within Australia, and are one of the most species-rich bee subfamilies in the world (Michener 2007). However, many Australian hylaeine species are likely threatened by anthropogenic activity due to their specific nesting requirements and floral-food preferences (Houston 1975, 1981, Almeida 2008). *Amphylaeus morosus* is only one of four species in its genus, making it a phylogenetically-interesting species. Our results indicate that *A. morusus* might have undergone major declines (due to reduced habitat availability and quality) since European arrival and that the remaining habitat has become increasingly fragmented and dispersed. Following the revision by Houston (1975), no further assessment of the current extent of occurrence has been conducted for this species.

5.6 Concluding remarks

Amphylaeus morosus provides an opportunity to assess population structure in a species lacking mtDNA variation across a wide geographical range, and also raises questions about parasite and disease susceptibility. We have hypothesized that the lack of mtDNA in *A. morosus* is the result of a species-wide, mitochondrial selective sweep, induced by *Wolbachia* (Chapter 3), rather than a massive population bottleneck event. Examination of past demography in each population could be applied to further understand the selective process(es) that have homogenised the heteroplasmic mtDNA in *A. morosus*. These analyses could also investigate the degree to which past clearing has impacted *A. morosus* and provide insight into how other wild bees cope with massively anthropogenically-damaged landscapes. Nonetheless, these results indicate a historic, species-wide spread of *Wolbachia* is a feasible explanation for the usual mtDNA patterns observed in *A. morosus*.

5.7 Acknowledgements

This work was supported by the Australian Federal Governments Research Training Program Scholarship, as well as the Playford Memorial Trust PhD scholarship. Thank you to the Holsworth Wildlife Trust, Mark Mitchell Foundation, and Ecological Society of Australia for providing research funding for this project. We extend our appreciation to Rebecca Dew and Ben Parslow for their assistance in fieldwork, and to Robert O'Reilly for his assistance with SNP analyses. Insect specimens were collected under the sampling permits SL101857 (NSW) and SL102137 (NSW).

5.8 References

Adams, M., T. J. Page, D. A. Hurwood, and J. M. Hughes. 2013. A molecular assessment of species boundaries and phylogenetic affinities in *Mogurnda* (Eleotridae): a case study of cryptic biodiversity in the Australian freshwater fishes. Marine and Freshwater Research **64**:920–931.

ALA. 2020. *Amphylaeus morosus* — Atlas of Living Australia occurrence download at https://biocache.ala.org.au/occurrences/search?&q=Amphylaeus+morosus&disableAllQualityFilters=true accessed on 29 December 2020. .

Almeida, E. A. B. 2008. Colletidae nesting biology (Hymenoptera: Apoidea). Apidologie **39**:16–29.

Barr, C. M., M. Neiman, and D. R. Taylor. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist **168**:39–50.

Bensch, S., D. E. Irwin, J. H. Irwin, L. Kvist, and S. Åkesson. 2006. Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus* warblers. Molecular Ecology **15**:161–171.

Beveridge, M., and L. W. Simmons. 2006. Panmixia: an example from Dawson's burrowing bee (*Amegilla dawsoni*) (Hymenoptera: Anthophorini). Molecular Ecology **15**:951–957.
Bivand, R. S., E. J. Pebesma, V. Gomez-Rubio, and E. J. Pebesma. 2013. Applied spatial data analysis with R. Springer.
Boff, S., A. Soro, R. J. Paxton, and I. Alves-dos-Santos. 2014. Island isolation reduces genetic diversity and connectivity but does not significantly elevate diploid male production in a neotropical orchid bee. Conservation Genetics **15**:1123–1135.
Bossart, J. L., and D. Pashley Prowell. 1998. Genetic estimates of population structure and gene flow: Limitations, lessons and new directions. Trends in Ecology & Evolution **13**:202–206.
Cameron, S. A., J. D. Lozier, J. P. Strange, J. B. Koch, N. Cordes, L. F. Solter, and T. L. Griswold. 2011. Patterns of widespread decline in North American bumble bees. Proceedings of the National Academy of Sciences of the United States of America **108**:662–667.
Caye, K., T. M. Deist, H. Martins, O. Michel, and O. François. 2016. TESS3: fast inference of spatial population structure and genome scans for selection. Molecular Ecology Resources **16**:540–548.
Caye, K., F. Jay, and O. François. 2017. Analyzing population genomic data with tess3r. https://bcm-uga.github.io/TESS3_encho_sen/articles/main-vignette.html accessed: 24 Mar 2021.
Černá, K., J. Straka, and P. Munclinger. 2013. Population structure of pioneer specialist solitary bee *Andrena vaga* (Hymenoptera: Andrenidae) in central Europe: the effect of habitat fragmentation or evolutionary history? Conservation Genetics **14**:875–883.
Cruz, V. M. V., A. Kilian, and D. A. Dierig. 2013. Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop *Lesquerella* and related species. PLOS ONE **8**:e64062.
Di Santo, L. N., and J. A. Hamilton. 2021. Using environmental and geographic data to optimize *ex situ* collections and preserve evolutionary potential. Conservation Biology **35**:733–744.
Dorey, J. B., E. P. Fagan-Jeffries, M. I. Stevens, and M. P. Schwarz. 2020. Morphometric comparisons and novel observations of diurnal and low-light-foraging bees. Journal of Hymenoptera Research **79**:117–144.
Duennes, M. A., J. D. Lozier, H. M. Hines, and S. A. Cameron. 2012. Geographical patterns of genetic divergence in the widespread Mesoamerican bumble bee *Bombus ephippiatus* (Hymenoptera: Apidae). Molecular Phylogenetics and Evolution **64**:219–231.
Ekblom, R., and J. Galindo. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity **107**:1–15.
Estoup, A., M. Solignac, J. Cornuet, J. Goudet, and A. Scholl. 1996. Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. Molecular Ecology **5**:19–31.
Exeler, N., A. Kratochwil, and A. Hochkirch. 2008. Strong genetic exchange among populations of a specialist bee, *Andrena vaga* (Hymenoptera: Andrenidae). Conservation Genetics **9**:1233–1241.
Françoso, E., A. R. Zuntini, A. C. Carnaval, and M. C. Arias. 2016. Comparative phylogeography in the Atlantic forest and Brazilian savannas: pleistocene fluctuations and dispersal shape spatial patterns in two bumblebees. BMC Evolutionary Biology **16**:267.

Frantz, A. C., S. Cellina, A. Krier, L. Schley, and T. Burke. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? Journal of Applied Ecology **46**:493–505.

Grewe, P. M., P. Feutry, P. L. Hill, R. M. Gunasekera, K. M. Schaefer, D. G. Itano, D. W. Fuller, S. D. Foster, and C. R. Davies. 2015. Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. Scientific Reports **5**:16916.

Gruber, B., P. J. Unmack, O. F. Berry, and A. Georges. 2018. dartrR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources **18**:691–699.

Guillot, G., A. Estoup, F. Mortier, and J. Cosson. 2005a. A spatial statistical model for landscape genetics. Genetics **170**:1261–1280.

Guillot, G., R. Leblois, A. Coulon, and A. C. Frantz. 2009. Statistical methods in spatial genetics. Molecular Ecology **18**:4734–4756.

Guillot, G., F. Mortier, and A. Estoup. 2005b. Geneland: a computer package for landscape genetics. Molecular Ecology Notes **5**:712–715.

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology & Evolution **4**:6–11.

Hijmans, R. J., and J. v. Etten. 2012. raster: Geographic analysis and modeling with raster data.

Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.

Houston, T. 1981. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). II. Australian Journal of Zoology Supplementary Series **29**:1–128.

Howe, R. W., G. J. Davis, and V. Mosca. 1991. The demographic significance of 'sink' populations. Biological Conservation **57**:239–255.

Hurst, G. D. D., and F. M. Jiggins. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proceedings of the Royal Society B: Biological Sciences **272**:1525–1534.

Jaccoud, D., K. Peng, D. Feinstein, and A. Kilian. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research **29**:e25.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11**:94.

Kilian, A., P. Wenzl, E. Huttner, J. Carling, L. Xia, H. Blois, V. Caig, K. Heller-Uszynska, D. Jaccoud, C. Hopper, M. Aschenbrenner-Kilian, M. Evers, K. Peng, C. Cayla, P. Hok, and G. Uszynski. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. Methods in Molecular Biology **888**:67–89.

Kool, J. T., A. Moilanen, and E. A. Treml. 2013. Population connectivity: recent advances and new perspectives. Landscape Ecology **28**:165–185.

Lambert, M. R., D. K. Skelly, and T. Ezaz. 2016. Sex-linked markers in the North American green frog (*Rana clamitans*) developed using DArTseq provide early insight into sex chromosome evolution. BMC Genomics **17**:844.

Lilley, T. M., T. Sävilammi, G. Ossa, A. S. Blomberg, A. Vasemägi, V. Yung, D. L. J. Vendrami, and J. S. Johnson. 2020. Population connectivity predicts vulnerability to white-nose syndrome in the Chilean Myotis (*Myotis chiloensis*) — A genomics approach. G3 Genes|Genomes|Genetics **10**:2117–2126.

Lin, C., and B. N. Danforth. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Molecular Phylogenetics and Evolution **30**:686–702.

Lozier, J. D., and S. A. Cameron. 2009. Comparative genetic analyses of historical and contemporary collections highlight contrasting demographic histories for the bumble bees *Bombus pensylvanicus* and *B. impatiens* in Illinois. Molecular Ecology **18**:1875–1886.

Melville, J., M. L. Haines, K. Boysen, L. Hodkinson, A. Kilian, K. L. Smith Date, D. A. Potvin, and K. M. Parris. 2017. Identifying hybridization and admixture using SNPs: application of the DArTseq platform in phylogeographic research on vertebrates. Royal Society Open Science **4**:161061.

Michener, C. D. 2007. The Bees of the World. 2 edition. The Johns Hopkins University Press, Baltimore, USA.

Moritz, C., T. Dowling, and W. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annual Review of Ecology and Systematics **18**:269–292.

Mulvena, S. R., J. C. Pierson, K. A. Farquharson, E. A. McLennan, C. J. Hogg, and C. E. Grueber. 2020. Investigating inbreeding in a free-ranging, captive population of an Australian marsupial. Conservation Genetics **21**:665–675.

Neumann, K., and K. Seidelmann. 2006. Microsatellites for the inference of population structures in the Red Mason bee *Osmia rufa* (Hymenoptera, Megachilidae). Apidologie **37**:75–83.

NMD. 2003a. Vegetation - Post-European Settlement (1988).*in* Geoscience Australia: National Mapping Division, editor.

NMD. 2003b. Vegetation - Pre-European Settlement (1788).*in* Geoscience Australia: National Mapping Division, editor.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2019. vegan: Community Ecology Package. R package version 2.5-6.

Park, S. D. E., D. A. Magee, P. A. McGettigan, M. D. Teasdale, C. J. Edwards, A. J. Lohan, A. Murphy, M. Braud, M. T. Donoghue, Y. Liu, A. T. Chamberlain, K. Rue-Albrecht, S. Schroeder, C. Spillane, S. Tai, D. G. Bradley, T. S. Sonstegard, B. J. Loftus, and D. E. MacHugh. 2015. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. Genome Biology **16**:234.

Pebesma, E., and R. S. Bivand. 2005. S classes and methods for spatial data: the sp package. R news **5**:9-13.

Perez, M. F., F. F. Franco, J. R. Bombonato, I. A. S. Bonatelli, G. Khan, M. Romeiro-Brito, A. C. Fegies, P. M. Ribeiro, G. A. R. Silva, and E. M. Moraes. 2018. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? Diversity and Distributions **24**:1883–1889.

R Development Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rokas, A., R. J. Atkinson, G. S. Brown, S. A. West, and G. N. Stone. 2001. Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? Heredity **87**:294–304.

RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.
Shams, F., F. Dyer, R. Thompson, R. P. Duncan, J. D. Thiem, A. Kilian, and T. Ezaz. 2019. Application of DArT seq derived SNP tags for comparative genome analysis in fishes; An alternative pipeline using sequence data from a non-traditional model species, *Macquaria ambigua*. PLOS ONE **14**:e0226365.
Shao, Z., H. Mao, W. Fu, M. Ono, D. Wang, M. Bonizzoni, and Y. Zhang. 2004. Genetic structure of Asian populations of *Bombus ignitus* (Hymenoptera: Apidae). Journal of Heredity **95**:46–52.
Sinkins, S. P., and H. C. J. Godfray. 2004. Use of *Wolbachia* to drive nuclear transgenes through insect populations. Proceedings of the Royal Society of London. Series B: Biological Sciences **271**:1421–1426.
Spessa, A. 1999. The behavioural and population ecology of an Australian native bee, *Amphylaeus morosus* Smith (Colletidae: Hylaeinae). Thesis. Australian National University, Australia.
Spessa, A., M. P. Schwarz, and M. Adams. 2000. Sociality in *Amphylaeus morosus* (Hymenoptera: Colletidae: Hylaeinae). Annals of the Entomological Society of America **93**:684–692.
Sucháčková Bartoňová, A., M. Konvička, J. Marešová, M. Wiemers, N. Ignatev, N. Wahlberg, T. Schmitt, and Z. Faltýnek Fric. 2021. *Wolbachia* affects mitochondrial population structure in two systems of closely related Palaearctic blue butterflies. Scientific Reports **11**:3019.
Teske, P. R., T. R. Golla, J. Sandoval-Castillo, A. Emami-Khoyi, C. D. van der Lingen, S. von der Heyden, B. Chiazzari, B. Jansen van Vuuren, and L. B. Beheregaray. 2018. Mitochondrial DNA is unsuitable to test for isolation by distance. Scientific Reports **8**:8448.
Torati, L. S., J. B. Taggart, E. S. Varela, J. Araripe, S. Wehner, and H. Migaud. 2019. Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. BMC Genetics **20**:13.
van Deventer, R., C. Rhode, M. Marx, and R. Roodt-Wilding. 2020. The development of genome-wide single nucleotide polymorphisms in blue wildebeest using the DArTseq platform. Genomics **112**:3455–3464.
van Strien, M. J., R. Holderegger, and H. J. Van Heck. 2015. Isolation-by-distance in landscapes: considerations for landscape genetics. Heredity **114**:27–37.
VanDerWal, J., L. Falconi, S. Januchowski, L. Shoo, and C. Storlie. 2014. SDMTools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises. R package version **1**:1–221.
Viengkone, M., A. E. Derocher, E. S. Richardson, R. M. Malenfant, J. M. Miller, M. E. Obbard, M. G. Dyck, N. J. Lunn, V. Sahanatien, and C. S. Davis. 2016. Assessing polar bear (*Ursus maritimus*) population structure in the Hudson Bay region using SNPs. Ecology and Evolution **6**:8474–8484.
Wang, I. J., and G. S. Bradburd. 2014. Isolation by environment. Molecular Ecology **23**:5649–5662.
Widmer, A., and P. Schmid-Hempel. 1999. The population genetic structure of a large temperate pollinator species, *Bombus pascuorum* (Scopoli) (Hymenoptera: Apidae). Molecular Ecology **8**:387–398.
Widmer, A., P. Schmid-Hempel, A. Estoup, and A. Scholl. 1998. Population genetic structure and colonization history of *Bombus terrestris* s.l. (Hymenoptera: Apidae) from the Canary Islands and Madeira. Heredity **81**:563–572.
Wright, S. 1931. Evolution in Mendelian populations. Genetics **16**:97–159.
Wright, S. 1943. Isolation by Distance. Genetics **28**:114–138.

Zwickl, D. J., and D. M. Hillis. 2002. Increased taxon sampling greatly reduces phylogenetic error. Systematic Biology **51**:588–598.

5.9 Tables

Table 1: Comparison of pairwise F_{ST} values of filtered SNP data of *Amphylaeus morosus* collected at the six localities across its distribution along the eastern coast of Australia. All F_{ST} statistics were significant (P-values < 0.0001).

LOCATION	Tin Can Bay	Enfield State Forest	Blue Mountains	Snowy Mountains	Dandenong Ranges	Cobboboonee
Tin Can Bay	-					
Enfield State Forest	0.206	-				
Blue Mountains	0.177	0.035	-			
Snowy Mountains	0.277	0.108	0.075	-		
Dandenong Ranges	0.257	0.125	0.132	0.065	-	
Cobboboonee	0.500	0.391	0.341	0.434	0.117	-

5.10 Figures

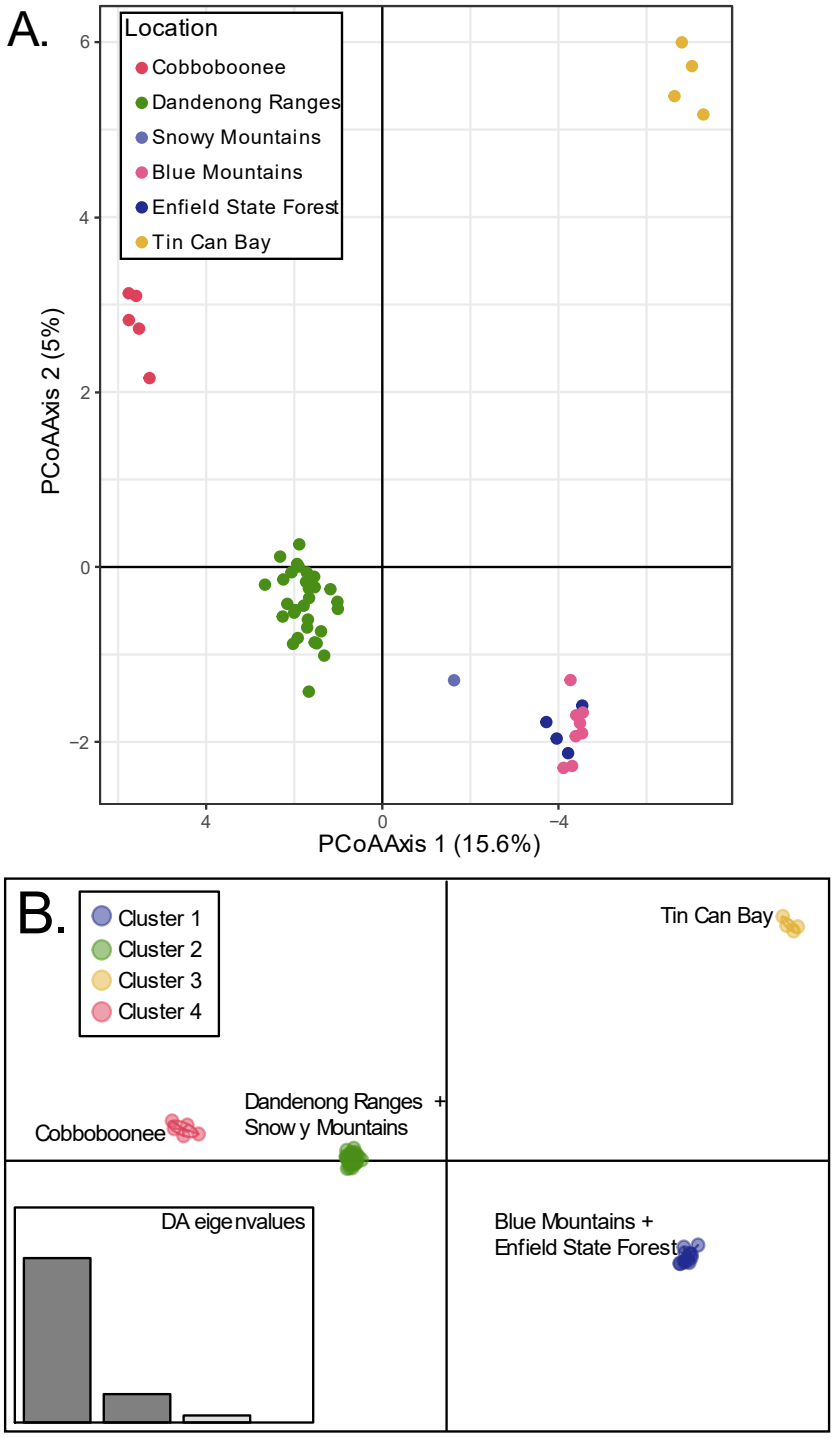


Figure 1: Non-spatial analyses of filtered SNP markers of *Amphylaeus morosus* individuals (n = 52) collected across the eastern coast of Australia. A. Scatterplot of principal coordinate analysis (PCoA) with individuals coloured by collection location, where the first two principal coordinate explain ~20% of the total variation. B. Scatterplot from discriminant analysis of principle components (DAPC) indicating four genetic clusters (captured with inertia ellipses). Individuals (dots) were not predefined by location, and the collection locality of individuals in each cluster is given.

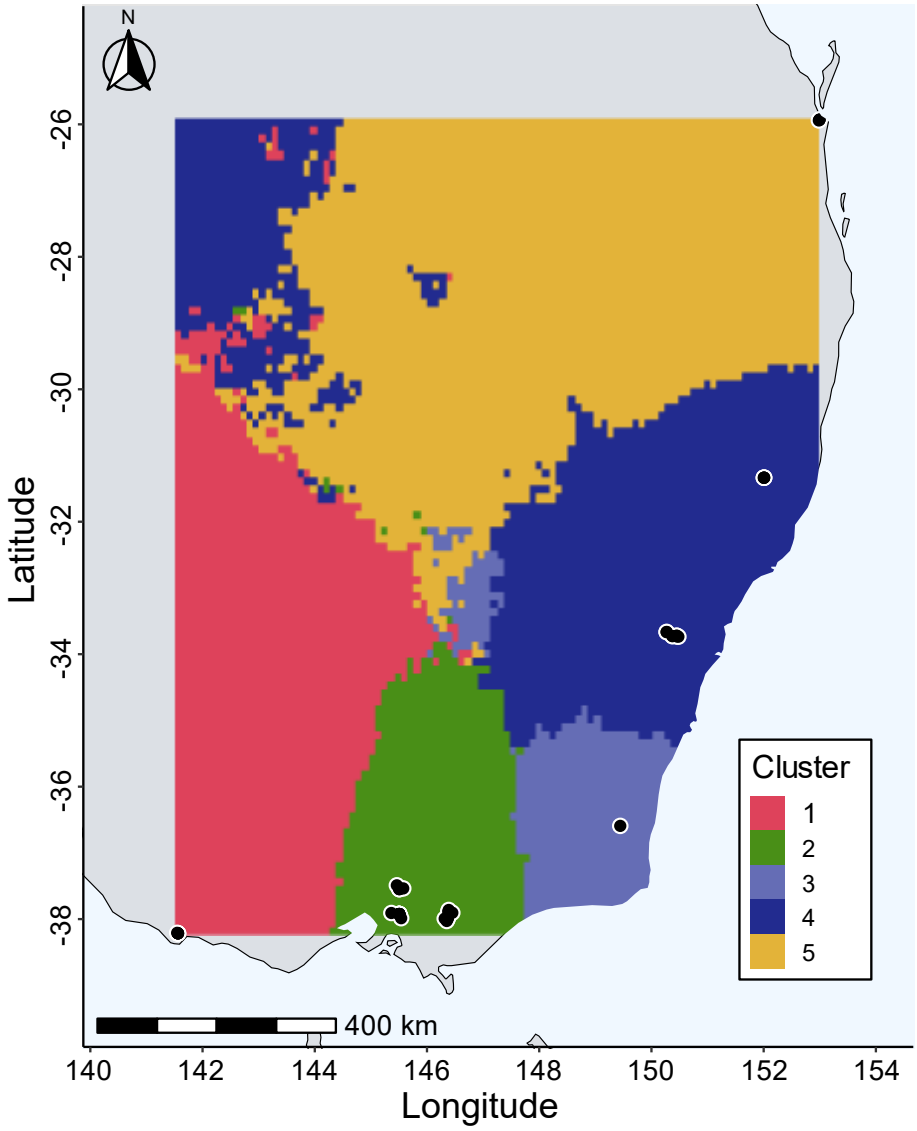


Figure 2: Baysian analysis of population structure in GENELAND (Guillot et al. 2005a, Guillot et al. 2005b) incorportating spatial data as a parameter. Ten simulations were performed with the clustering estimate from the simulation with the highest prosterior probability taken; this simulation suggested five genetic clusters across the *Amphylaeus morosus* distribution. Note: this model has incorrectly extrapolated equivilant populations towards central Australia where *A. morosus* does not occur.

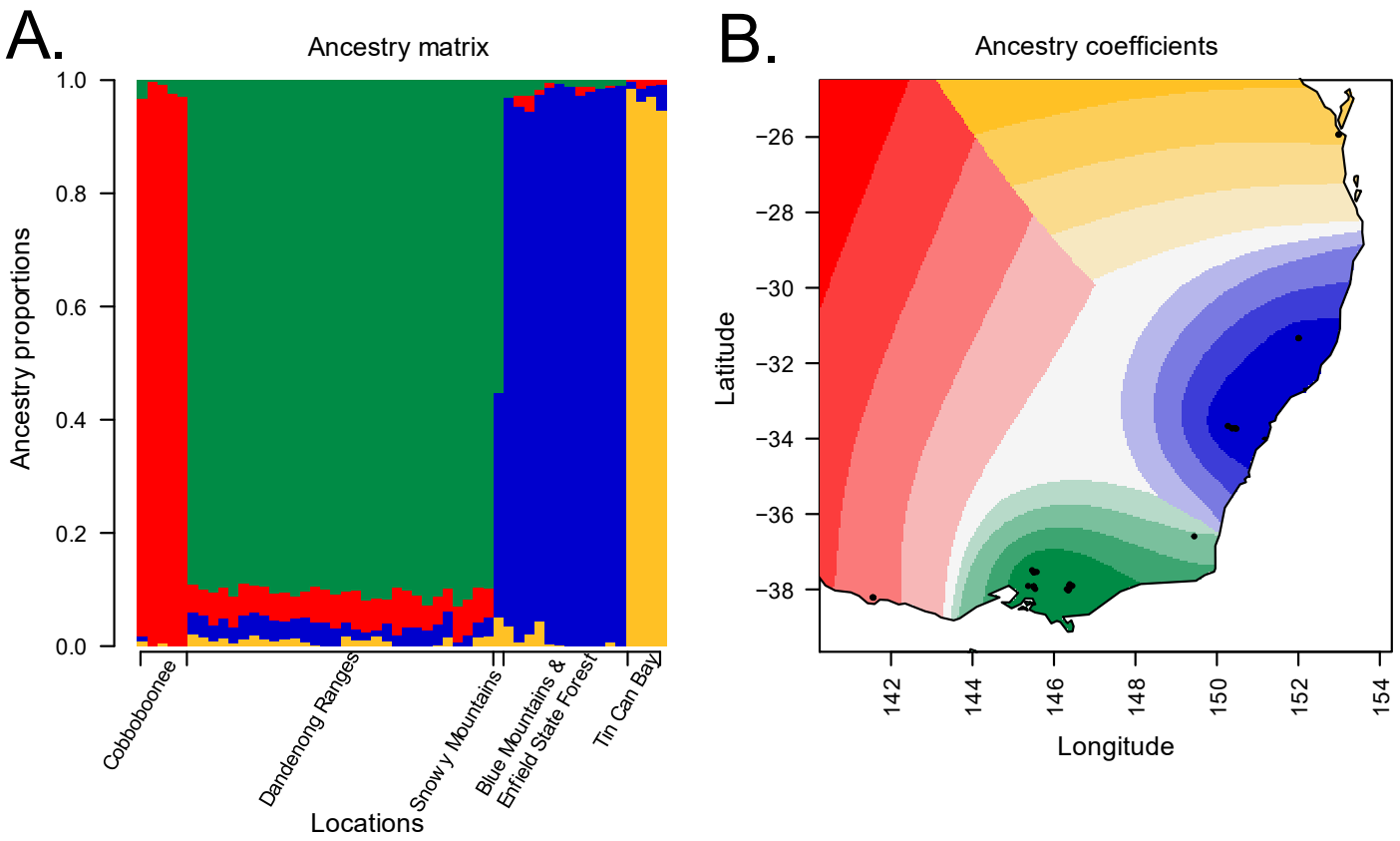


Figure 3: TESS3 analysis (Caye et al. 2016) of population structure, with A. a barplot of proportions of genetic ancestry for four genetic clusters assigned to n = 52 individuals from across the *Amphylaeus morosus* geographical range, and B. the geographical boundaries of the four genetic clustered determined for these individuals, sampled along the eastern Australia coast (black dots). Note: as with Figure 2 this model has incorrectly extrapolated equivilant populations towards central Australia where *A. morosus* does not occur.

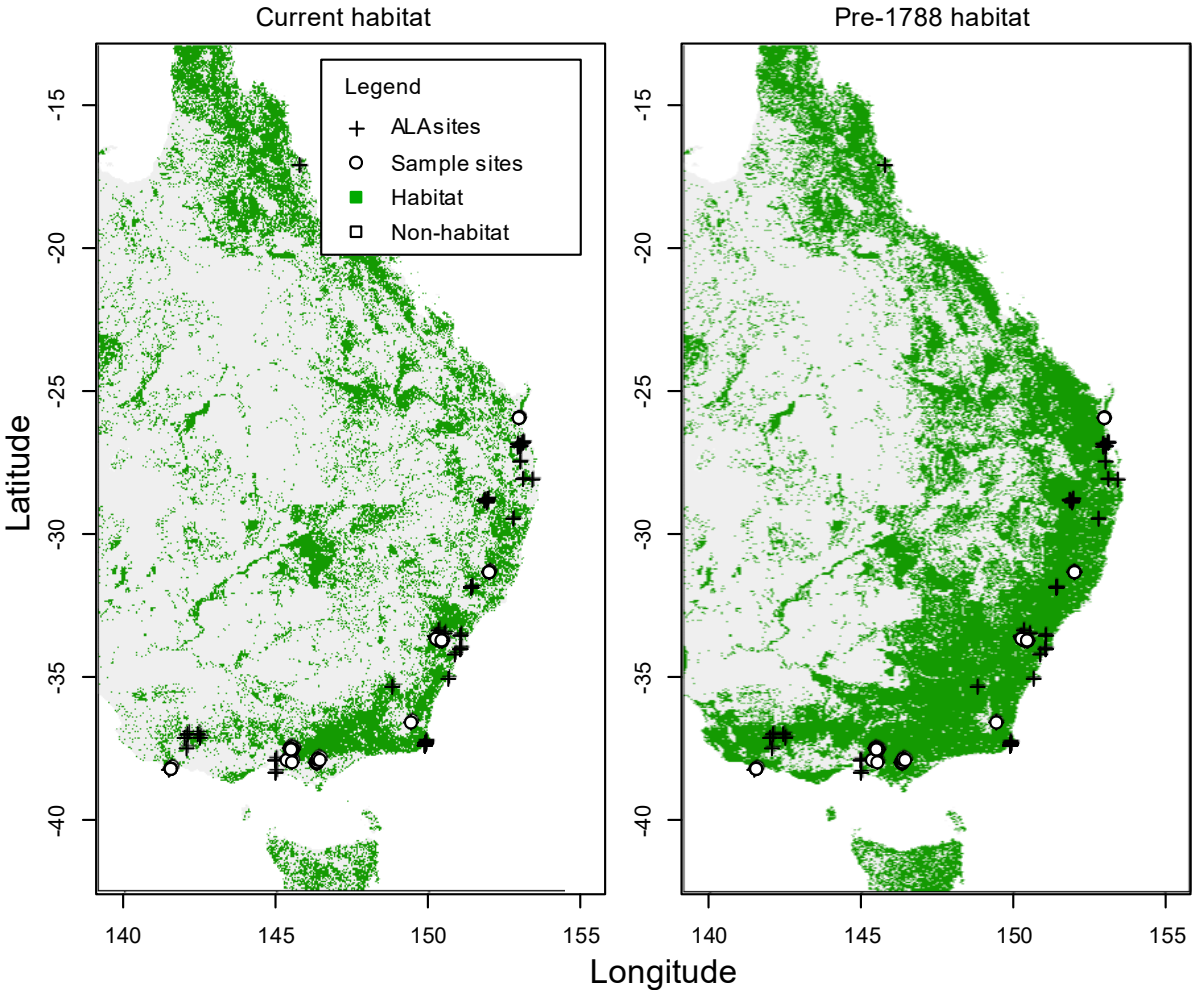


Figure 4: Current versus pre-European habitat of *Amphylaeus morosus* using major vegetation sub (MVS) groups (NMD 2003b, a).

5.12 Supplementary Material

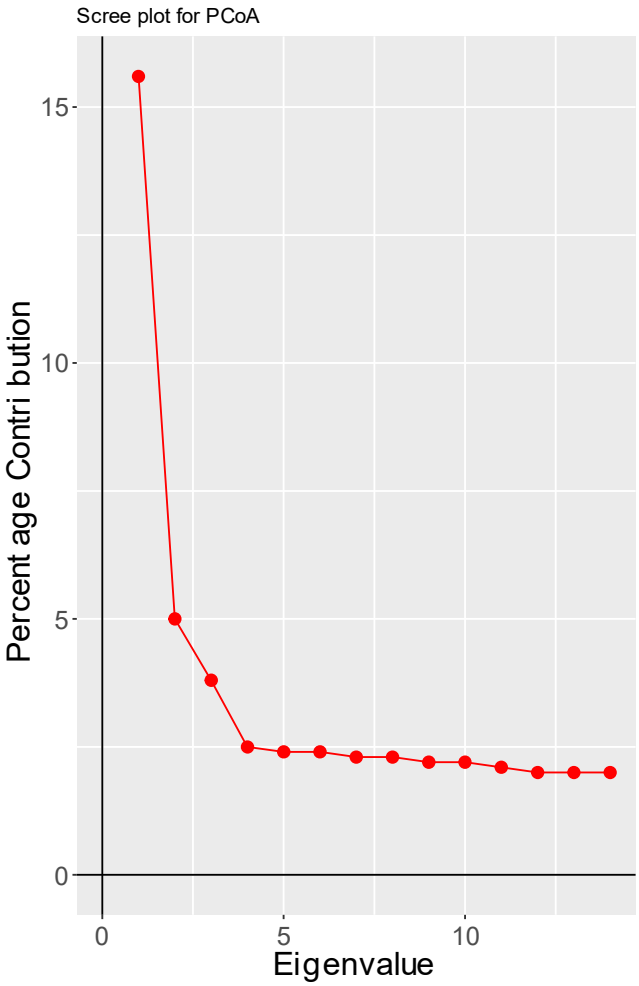
Supp. Table 1: Location and dates of collection for each of the 52 *Amphylaeus morosus* female specimens collected from six localities from between southern Queensland to south-western Victoria, Australia used in the population structure analyses.

Location	Specimen	Latitude	Longitude	Collection Date
Tin Can Bay	1	-25.9301	152.9942	10-Dec-13
	2	-25.9301	152.9942	10-Dec-13
	3	-25.9301	152.9942	10-Dec-13
	4	-25.9412	152.9837	29-Jun-16
Enfield State Forest	5	-31.335	152.0076	22-Dec-18
	6	-31.335	152.0076	22-Dec-18
	7	-31.335	152.0076	22-Dec-18
	8	-31.336	152.0015	22-Dec-18
Blue Mountains	9	-33.6642	150.2712	25-Jul-17
	10	-33.666	150.2707	25-Jul-17
	11	-33.7205	150.4502	24-Jul-17
	12	-33.7234	150.4495	24-Jul-17
	13	-33.7261	150.3734	25-Jul-17
	14	-33.7264	150.3734	25-Jul-17
	15	-33.7307	150.4502	24-Jul-17
	16	-33.7358	150.4796	24-Jul-17
Snowy Mountains	17	-36.5922	149.4444	8-Dec-17
Dandenong Ranges	18	-37.4897	145.4628	8-Nov-14
	19	-37.5328	145.5093	8-Nov-14
	20	-37.5362	145.5098	5-Aug-18
	21	-37.5365	145.5762	8-Nov-14
	22	-37.5456	145.505	8-Nov-14
	23	-37.5457	145.505	5-Aug-18
	24	-37.5476	145.5078	5-Aug-18
	25	-37.8619	146.3877	9-Nov-14
	26	-37.8835	146.3816	9-Nov-14
	27	-37.8942	146.3838	9-Nov-14
	28	-37.9041	146.4511	9-Nov-14
	29	-37.9078	145.3649	4-Aug-18
	30	-37.924	145.5042	26-Dec-16
	31	-37.924	145.5042	26-Dec-16
	32	-37.924	145.5042	3-Jan-17
	33	-37.924	145.5042	31-Dec-16
	34	-37.924	145.5042	3-Jan-17
	35	-37.924	145.5042	6-Dec-17
	36	-37.924	145.5042	6-Dec-17
	37	-37.924	145.5042	6-Dec-17
	38	-37.924	145.5042	6-Dec-17
	39	-37.924	145.5042	6-Dec-17
	40	-37.924	145.5042	6-Dec-17
	41	-37.924	145.5042	6-Dec-17

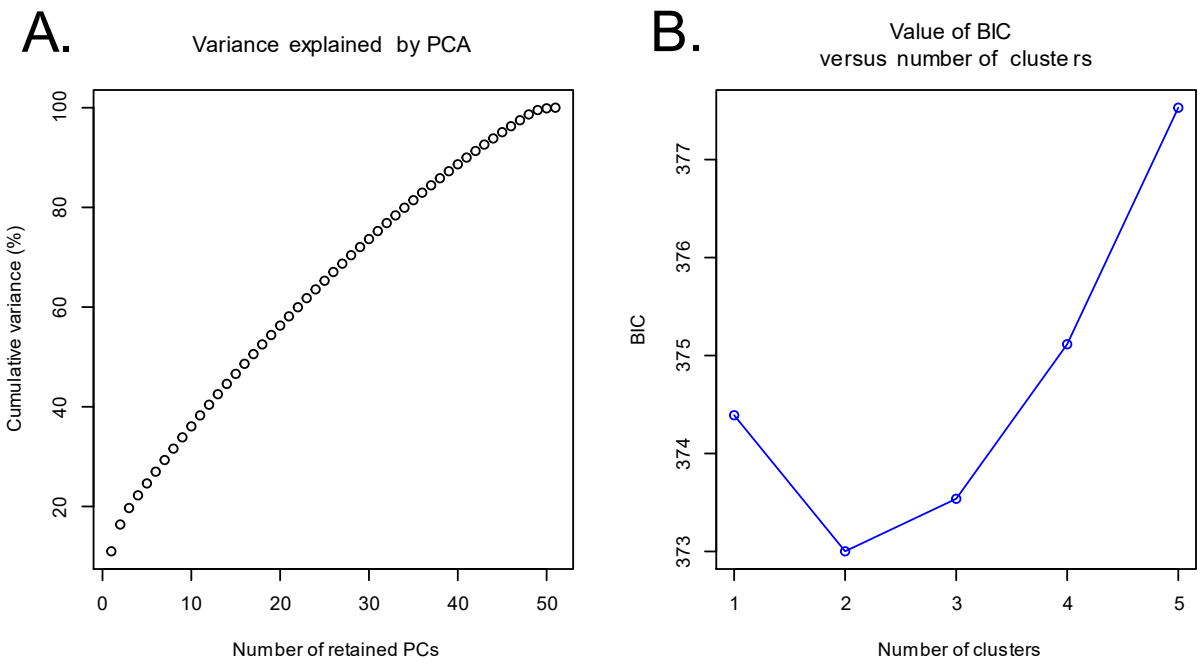
	42	-37.9815	146.3713	9-Nov-14
	43	-37.9849	146.3683	9-Nov-14
	44	-37.9859	145.5394	2-Aug-18
	45	-37.9927	146.3156	7-Nov-14
	46	-37.9977	146.3686	7-Nov-14
	47	-38.0281	146.3517	7-Nov-14
Cobboboonee	48	-38.2072	141.5477	2-Feb-17
	49	-38.2151	141.5624	2-Feb-17
	50	-38.2151	141.5624	2-Feb-17
	51	-38.2151	141.5624	2-Feb-17
	52	-38.2151	141.5624	2-Feb-17

Supp. Table 2: Mean observed heterozygosity (H_O), mean expected heterozygosity (H_S), results from two sample t-tests comparing for significant differences between H_O and H_S in each collection locality, and inbreeding coefficient (F_{IS}) of *Amphylaeus morosus* individuals collected across six locations along the eastern coast of Australia. A significant difference between mean H_O and H_S (significance level 0.05) for all locations south of the Blue Mountains, NSW was detected. Note (**) only one specimen was collected in the Snowy Mountains, NSW and although reported, results for this specimen are not interpretable.

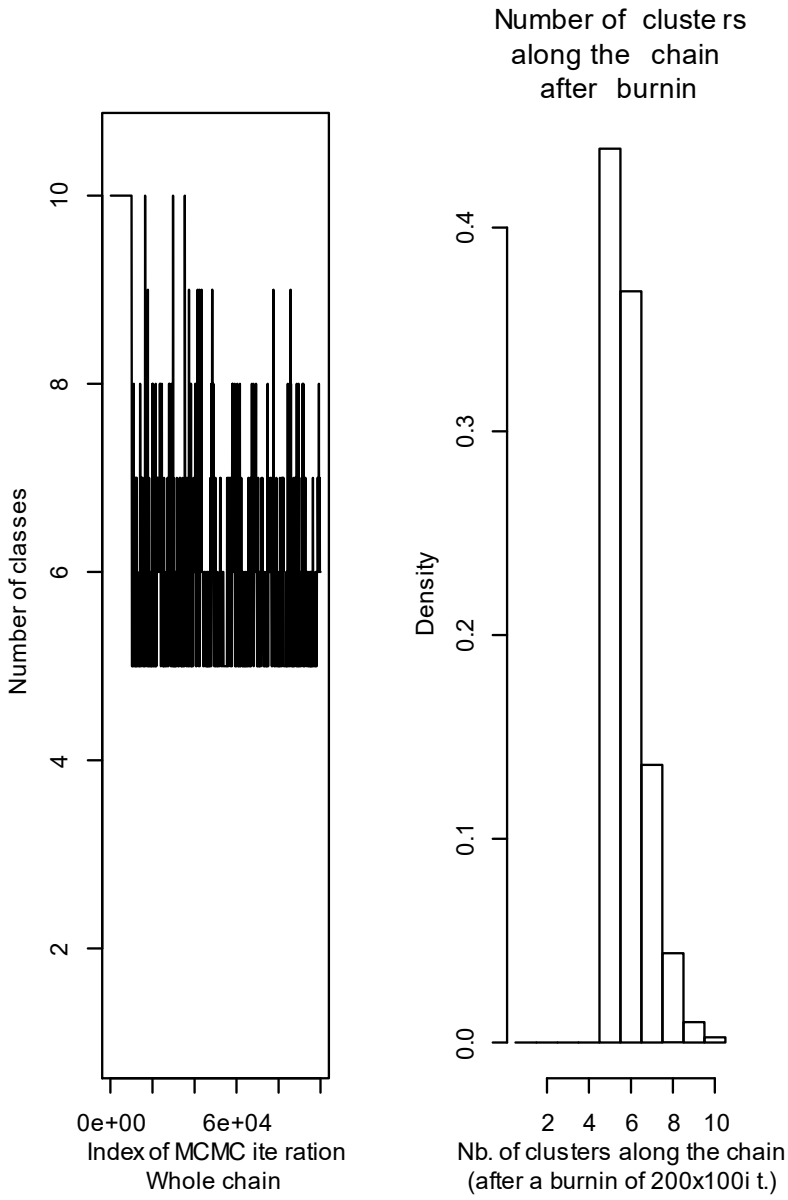
Location	**No. individuals**	**No. loci**	**Observed heterozygosity (Ho)**	**Expected heterozygosity (Hs)**	**Test statistic**	**P-value**	**Inbreeding coefficient (F_{IS})**
Tin Can Bay	4	1243	0.136	0.140	$t_{1242} = -1.214$	0.225	0.029
Enfield State Forest	4	1243	0.156	0.156	$t_{1242} = -0.0812$	0.935	0.000
Blue Mountains	8	1243	0.178	0.188	$t_{1242} = -3.787$	**1.63×10^{-4} ***	0.053
Snowy Mountains	1	1243	*0.191***	*0.095***	$t_{1242} = 17.11$	*$<2.2\times10^{-16}$***	*-1.011***
Dandenong Ranges	30	1243	0.208	0.216	$t_{1242} = -5.802$	**8.33×10^{-9} ***	0.037
Cobboboonee	5	1243	0.122	0.107	$t_{1242} = 4.674$	**3.23×10^{-6} ***	-0.140



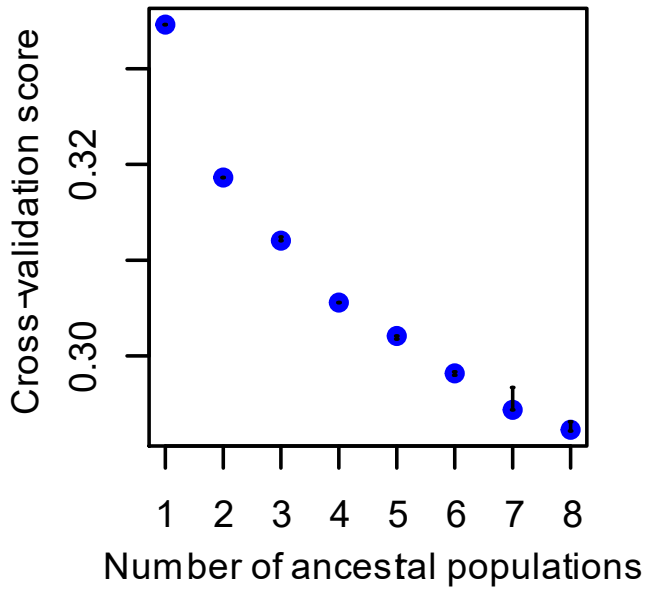
Supp. Figure 1: Percentage of total variance in *Amphylaeus morosus* SNP data explained by each of the principal components in the PCoA.



Supp. Figure 2: A. Cumulative variance (%) explained by the principal components (PCs) and B. Bayesian informative criterion (BIC) for each value of *K* (number of clusters) for DAPC analysis of *Amphylaeus morosus* SNP data.



Supp. Figure 3: Bayesian analysis in GENELAND to estimate the number of clusters (K = 1–10) from 10 runs of 100,000 MCMC per run. Barplots of the cluster results for the run with the highest posterior probability, indicating five genetic clusters as the most probable in these *Amphylaeus morosus* SNP data.



Supp. Figure 4: Cross-validation scores estimated for *K* = 1–8 (number of ancestral populations) averaged over 10 runs using TESS3 analysis for *Amphylaeus morosus* SNP data.

General Discussion

This thesis explores the complex connection between unusual mitochondrial traits and heritable endosymbionts in their insect hosts from an evolutionary and ecological context. These factors have broad implications for invertebrate research, particularly in studies of parasite-host interactions and the use of mitochondrial markers for population differentiation, species delineation, and to infer phylogenetic relationships.

In this thesis, I described widespread and conserved mitochondrial heteroplasmy in the hylaeine bee, *Amphylaeus morosus* (Smith, 1879) (Hymenoptera: Colletidae). In *A. morosus*, mitochondrial heteroplasmy is present as two *conserved* mtDNA haplotypes over the species' entire range which has not been documented for any other animal species. Furthermore, the ubiquitous presence of two co-occurring supergroup A strains of *Wolbachia* (Rickettsiales: Anaplasmataceae) in *A. morosus* raises the possibility that they may play some role in the unusual heteroplasmy occurring in this host. Using molecular based methods, I have explored the incidences of these two sets of co-occurring genomes across the *A. morosus* distribution. I have also investigated the utility of different experimental approaches to demonstrate the relationship between *Wolbachia* and the mitogenomes of this non-model host. Below, I discuss the specific research findings and the broader implications of this complex system.

Key research findings

Mitochondrial heteroplasmy is not as uncommon as once thought

My comprehensive literature review (Chapter 1) captured the extent of mitochondrial heteroplasmy reports in animal taxa. Once thought to be uncommon, heteroplasmy occurs widely across multiple phyla. Different taxon groups generally preferentially exhibit specific forms of heteroplasmy; with chordates predominantly described as having length heteroplasmy and invertebrates in general with site heteroplasmy. Furthermore, extensive forms of heteroplasmy, where all individuals in a population are heteroplasmic, are being increasingly reported across multiple phyla. This challenges the idea that heteroplasmy is a transient condition that may arise sporadically only in some individuals.

Additionally, heteroplasmic systems may have evolved in some specific taxon lineages, such as bivalves, bees, fishes, and mammals — however in the three latter cases, taxon sampling

bias and molecular techniques may have over-inflated the significance of these groups in relation to other taxa. Although bivalves are the most thoroughly interrogated heteroplasmic system (exhibiting doubly uniparental inheritance, DUI) little is understood about the mechanisms nor evolutionary origins this system (Doucet-Beaupré et al. 2010, Breton et al. 2014). In mammals and fishes, broad taxonomic sampling of heteroplasmy has been conducted, but an organised approach to better understand these currently loose patterns is yet to be conducted. Bees present a similar conundrum, except with even fewer taxa assessed and the experimental protocols applied have been more limited.

In this thesis I present an extreme example of naturally occurring, extensive heteroplasmy in the Australian native bee species, *Amphylaeus morosus* (Chapter 2). Two conserved mt-COI haplotypes occur in every individual across the ~2,000 km range, and shotgun sequencing data further suggests that these heteroplasmic sites occur throughout the mitogenome. The combination of widespread heteroplasmy and conserved mtDNA haplotypes is unusual, and the mechanisms maintaining these traits are currently unknown.

Wolbachia co-infection is widespread but its impact is challenging to investigate

Amphylaeus morosus is consistently infected with two distantly related strains of supergroup A *Wolbachia* across its distribution (Chapter 3). Because *Wolbachia* are known to be able to influence mitochondrial inheritance in their host, we hypothesized that *Wolbachia* could be enabling the maintenance of conserved mitochondrial heteroplasmy (Chapter 3). However, producing experimental evidence of changes to mitochondrial inheritance in response to *Wolbachia* presence/absence is difficult. Antibacterial treatment to remove *Wolbachia* and assess changes to mitochondrial inheritance, post-infection, is one of the most common, and convincing tools to demonstrate this relationship. I assessed the utility of antibiotic treatment methods to remove *Wolbachia* from this host and the practicality of maintaining this system over multiple generations (Chapter 4). However, *A. morosus*, like other wild bee species, is difficult to maintain in closed, captive settings. This is made even more difficult because *A. morosus* is univoltine, such that modifications to experimental treatments can only be made once a year. Furthermore, I found that females of *A. morosus* frequently switch nests, so keeping differently dosed or control populations separate is logistically very challenging. Without extensive efforts trialling animal husbandry for this species, maintaining captive populations of *A. morosus* over multiple generations is not possible. However, antibacterial

treatment remains the most convincing tool to conclusively support the impact of *Wolbachia* on the inheritance of heteroplasmic mitogenomes in *A. morosus.*

Evidence of *Wolbachia's* influence on this hosts mitogenome(s)

Because antibiotic treatment was not a feasible experimental tool to investigate the relationship between heteroplasmic mtDNA and maternal inheritance of endosymbionts in *A. morosus*, I explored other avenues of evidence to address the relationship between these genomes. The process(es) that led to the lack of mtDNA variation is *A. morosus* are an important component to investigate. The lack of variation could be the result of either a severe population bottleneck or a mitochondrial selective sweep induced by *Wolbachia.* Investigating patterns of diversity in nuclear DNA can be used to distinguish these hypotheses, because genomic variation is not affected by *Wolbachia* inheritance, as long as sexual reproduction is retained (Rokas et al. 2001) — but would be affected by a population bottleneck. I produced genome-wide single nucleotide polymorphisms (SNPs) using the DArTseq™ platform for individuals of *A. morosus* collected across its geographic range (Chapter 5). These results showed strong patterns of geographic divergence, indicating that a single, large population (as possibly suggested by the mtDNA) that has expanded recently from a severe bottleneck is unlikely. Furthermore, habitat modelling showed that pre-European settlement in Australia, *A. morosus* habitats were better connected, making a *Wolbachia* sweep across the whole species more plausible.

Discussion of the research findings

The combination of the lack of mtDNA variation and *Wolbachia* infection is compelling, because *Wolbachia* has been shown to induce mitochondrial selective sweeps in some host taxa (e.g. Jiggins (2003), Raychoudhury et al. (2010), Schuler et al. (2016)), removing mtDNA variation. A *Wolbachia*-induced selective sweep might explain the lack the mtDNA variation within *A. morosus*, however the persistent of widespread heteroplasmy is very puzzling. It would be expected that even if the initial infection occurred in a heteroplasmic host lineage, without selective processes favouring heteroplasmy, homoplasmy would be restored in some subsequent lineages as one mitochondrial haplotype is lost to genetic drift among the various lineages.

In this thesis, I presented three hypotheses (Chapter 3) that could explain the widespread and consistent mitochondrial heteroplasmy occurring in *A. morosus*.

(i) **The heteroplasmic founder hypothesis (H1)**: An *A. morosus* individual already possessed two mitogenomes (was heteroplasmic) and a single *Wolbachia* infection (which could have already swept through the population). A second *Wolbachia* strain entered this lineage (via horizonal or paternal transfer) and spread to fixation within the population, dragging along both mitogenomes and the initial *Wolbachia* strain. Mitochondrial heteroplasmy is maintained because of a mutualism with one of the *Wolbachia* strains (probably the ancestral infection).

(ii) **The co-inheritance hypothesis (H2)**: Two *A. morosus* individuals with different *Wolbachia* strains mated (each with a single divergent mitogenome), and both mitogenomes and *Wolbachia* strains were maintained in the progeny via paternal transmission of both a mitochondrial and bacterial genome. These co-occurring *Wolbachia* strains and their co-inherited mitogenomes then swept (via maternal inheritance) through the population to fixation. Each mitogenome is maintained by a mutualism with its corresponding *Wolbachia* strain.

(iii) **The advantageous heteroplasmy hypothesis (H3)**: The maintenance of mitochondrial heteroplasmy in *A. morosus* is unrelated to either *Wolbachia* strain. A selective sweep via either *Wolbachia* and/or an advantageous mutation (heteroplasmy) could be responsible for the overall loss of mitochondrial diversity. However, one of the two mitogenomes was never lost (post-fixation) because heteroplasmy has a fitness advantage over homoplasmy.

There is increasing evidence that *Wolbachia* can interact closely with mitochondria; for example influencing oxidative phosphorylation activity (Uribe-Alvarez et al. 2019), mtDNA density (Ballard and Melvin 2007), and the production of mitochondria-derived ROS (reactive oxygen species) (Kremer et al. 2009). However, the molecular processes that might enable *Wolbachia* to maintain the two (heteroplasmic) mtDNA haplotypes in *A. morosus* are currently unknown. The experimental protocol to assess these potential interactions in this host species would be challenging given the current limitations of maintaining captive host populations. Nonetheless, the impact of *Wolbachia* on *A. morosus* appears highly complex and presents a unique system to explore these potential molecular interactions.

It is important to note that although *Wolbachia* is a strong candidate for the mechanism maintaining the unusual mtDNA traits observed in *A. morosus*, other explanations remain. Maintenance of widespread heteroplasmy could be caused by other selective processes, such as mechanisms similar to heterozygote advantage as seen in nuclear genomes, positive or balancing selection. For example, in *Drosophila* it has been demonstrated that some mtDNA haplotypes have selective advantages when host individuals are exposed to different thermal conditions (Camus et al. 2017, Lajbner et al. 2018). These patterns of temperature-dependant selection have also been demonstrated in other taxa, such as the seed beetle *Callosobruchus maculatus* (Immonen et al. 2020) and yeasts in the genus *Saccharomyces* (Baker et al. 2019). *Amphylaeus morosus* occupies areas susceptible to extreme temperature ranges; for example, temperatures in the Dandenong Ranges, VIC, can range between from -2.7°C to 46.1°C (Bureau of Meteorology 2020). Therefore, maintenance of two mtDNA haplotypes could potentially occur through balancing selection under these extreme thermal conditions. However, in regions that have lower seasonal temperature variation, such as sub-tropical heathlands, the maintenance of heteroplasmy via temperature-dependent selection seems unlikely. Given these two mtDNA haplotypes are present across the entire distribution of *A. morosus*, the selective driver maintaining heteroplasmy would need to be in operation in all habitat and climate conditions. Identifying such a driver (beyond the hypothesized *Wolbachia*) is challenging.

Future research directions

This thesis describes the correlation of a widespread co-infection of *Wolbachia* and conserved mitochondrial heteroplasmy across the *A. morosus* distribution. Furthermore, I presented hypotheses that could explain the maintenance of this unusual mitochondrial system. This research is the first to analyse the co-occurrence of heteroplasmic mtDNA and *Wolbachia* across a species' distribution and these results present challenging considerations for our understanding of mitochondrial-endosymbiont-host relationships. Although I have endeavoured to provide strong evidence of the characteristics of these traits, there are numerous future research directions that can be applied to better understand this complex system.

Although numerous other controlled experimental protocols could provide convincing evidence towards characterizing the impact (if any) *Wolbachia* has on this hosts heteroplasmic mtDNA, protocols previously used in similar (but model) systems (e.g.

Reynolds et al. (2003), Dean (2006), Ilinsky (2013)) have limited application in this bee species because of the difficulty maintaining captive populations. Therefore, genetic and genomic sequencing as well as microscopy and micro-imagery techniques of wild-caught individuals are the most realistic experimental approaches to further explore these traits in *A. morosus* and similar non-model systems.

Ion Torrent shotgun sequencing data provided insights into the heteroplasmic variation beyond the mt-COI region, and whole mitogenome sequencing using newer next generation approaches would be beneficial to expand this research. Furthermore, obtaining both whole mitogenomes and *Wolbachia* genomes from *A. morosus* individuals across the geographic range will allow more in-depth exploration of (i) geographic patterns of minor variation within both types of genomes, (ii) potential geographic origins of each bacterial infection, and (iii) potential functional differences in these genomes. Additionally, mitochondrial heteroplasmy was detected in all genes recovered in our Ion Torrent data, suggesting that functional differences in *A. morosus* are operating at a whole genome level. In this research, I have only sampled for heteroplasmy (and *Wolbachia*) in muscle tissue in the leg and I did not explore tissue specificity of heteroplasmy. If heteroplasmy is being maintained because of selective retention of functional differences, these mtDNA haplotypes could be serving tissue-specific roles, or whole-body responses. Exploring tissue specificity of this heteroplasmy and the *Wolbachia* infection could provide more insights into the roles of these genomes in this host.

This research explored the distributional pattern of nuclear DNA across the range of *A. morosus* and shows population structure contrary to a 'panmictic' population indicated by mtDNA. This retention of genomic DNA structure as well as habitat modelling demonstrating pre-European habitat connectivity provide evidence that *Wolbachia* could have induced a selective-sweep across this hosts distribution — rather than the loss of mtDNA diversity being the result of a population bottleneck. However, this conclusion is broadly inferred and more specific analyses to better describe the mechanisms responsible for the traits observed in all genomes analysed should be implemented. Historical demography studies of nuclear DNA markers for *A. morosus* should be applied to provide more conclusive evidence that a population bottleneck followed by a range expansion is not responsible for the loss of mtDNA.

Intriguingly, widespread heteroplasmy has also been reported in a remarkable radiation of hylaeines in Hawaii (>80 species within ~700,000 years; Magnacca and Danforth (2006)), but unfortunately that study did not attempt to characterize within-species patterns or explore if they were linked to *Wolbachia* infections or other reproductive parasites. Furthermore, heteroplasmy has been reported in numerous other bee species from five of the world's seven bee families (Songram et al. 2006, Magnacca and Brown 2012, Françoso et al. 2016, Ricardo et al. 2020). Given that hylaeine bees originated in Australia and Australia contains all their deepest divergences (Kayaalp et al. 2013), explorations of the extent and origin of heteroplasmy in this bee group should focus on Australian taxa. Australia has an estimated 500 species of hylaeine bees (Michener 2007) and the taxonomy of this group was last addressed by Houston (1975, 1981) and now needs a comprehensive phylogenetic treatment. However, with increased affordability of DNA sequencing technologies means that documenting (using e.g. next or third generation sequencing DNA barcoding (Shokralla et al. 2014, Laforest-Lapointe et al. 2016, Hebert et al. 2018)) mitochondrial heteroplasmy and endosymbiont infections across this diverse taxon group is feasible. Furthermore, these techniques can be broadly applied to most animal taxa to investigate these traits.

Conclusion

This thesis is the first step in describing the association between widespread mitochondrial heteroplasmy and the co-infecting *Wolbachia* strains in the bee host *A. morosus*. The results are concordant with the notion that *Wolbachia* has played a role in producing the unusual mitochondrial traits observed in *A. morosus* — but verifying or eliminating that possibility will be very challenging. Nevertheless, this bee-*Wolbachia* system presents a remarkable opportunity to explore the complex interaction between endosymbionts and host mitogenomes, but future studies will need to develop methodologies that have not been previously applied to other model systems. An understanding of the roles of these unusual mitochondrial systems, as well as the interactions of hosts with endosymbionts are becoming increasingly relevant, and DNA sequencing technologies enable broad and rapid identification of these traits in natural systems. These complex systems challenge the long stranding traditional views of the roles of mtDNA and intracellular bacteria in their hosts and will enable a greater understanding of the generally poorly described biology of invertebrates.

References

Baker, E. C. P., D. Peris, R. V. Moriarty, X. C. Li, J. C. Fay, and C. T. Hittinger. 2019. Mitochondrial DNA and temperature tolerance in lager yeasts. Science Advances **5**:eaav1869.

Ballard, J. W. O., and R. G. Melvin. 2007. Tetracycline treatment influences mitochondrial metabolism and mtDNA density two generations after treatment in *Drosophila*. Insect Molecular Biology **16**:799–802.

Breton, S., L. Milani, F. Ghiselli, D. Guerra, D. T. Stewart, and M. Passamonti. 2014. A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends in Genetics **30**:555–564.

Bureau of Meteorology. 2020. Climate statistics for Australian locations – Scoresby Research Institution. http://www.bom.gov.au/climate/averages/tables/cw_086104.shtml accessed: 25 May 2020.

Camus, M. F., J. N. Wolff, C. M. Sgrò, and D. K. Dowling. 2017. Experimental support that natural selection has shaped the latitudinal distribution of mitochondrial haplotypes in Australian *Drosophila melanogaster*. Molecular Biology and Evolution **34**:2600–2612.

Dean, M. D. 2006. A *Wolbachia*-associated fitness benefit depends on genetic background in *Drosophila simulans*. Proceedings of the Royal Society B: Biological Sciences **273**:1415–1420.

Doucet-Beaupré, H., S. Breton, E. G. Chapman, P. U. Blier, A. E. Bogan, D. T. Stewart, and W. R. Hoeh. 2010. Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evolutionary Biology **10**:50.

Françoso, E., A. R. Zuntini, A. C. Carnaval, and M. C. Arias. 2016. Comparative phylogeography in the Atlantic forest and Brazilian savannas: pleistocene fluctuations and dispersal shape spatial patterns in two bumblebees. BMC Evolutionary Biology **16**:267.

Hebert, P. D. N., T. W. A. Braukmann, S. W. J. Prosser, S. Ratnasingham, J. R. deWaard, N. V. Ivanova, D. H. Janzen, W. Hallwachs, S. Naik, J. E. Sones, and E. V. Zakharov. 2018. A Sequel to Sanger: amplicon sequencing that scales. BMC Genomics **19**:219.

Houston, T. 1975. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). I. Introductory material and the genera *Heterapoides* Sandhouse, *Gephyrohylaeus* Michener, *Hyleoides* Smith, *Pharohylaeus* Michener, *Hemirhiza* Michener, *Amphylaeus* Michener and *Meroglossa* Smith. Australian Journal of Zoology Supplementary Series **23**:1–135.

Houston, T. 1981. A revision of the Australian hylaeine bees (Hymenoptera: Colletidae). II. Australian Journal of Zoology Supplementary Series **29**:1–128.

Ilinsky, Y. 2013. Coevolution of *Drosophila melanogaster* mtDNA and *Wolbachia* Genotypes. PLOS ONE **8**:e54373.

Immonen, E., D. Berger, A. Sayadi, J. Liljestrand-Rönn, and G. Arnqvist. 2020. An experimental test of temperature-dependent selection on mitochondrial haplotypes in *Callosobruchus maculatus* seed beetles. Ecology and Evolution **10**:11387–11398.

Jiggins, F. M. 2003. Male-killing *Wolbachia* and mitochondrial DNA: sselective sweeps, hybrid introgression and parasite population dynamics. Genetics **164**:5–12.

Kayaalp, P., M. P. Schwarz, and M. I. Stevens. 2013. Rapid diversification in Australia and two dispersals out of Australia in the globally distributed bee genus, *Hylaeus* (Colletidae: Hylaeinae). Molecular Phylogenetics and Evolution **66**:668–678.
Kremer, N., D. Voronin, D. Charif, P. Mavingui, B. Mollereau, and F. Vavre. 2009. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. PLOS Pathogens **5**:e1000630.
Laforest-Lapointe, I., C. Messier, and S. W. Kembel. 2016. Tree phyllosphere bacterial communities: exploring the magnitude of intra-and inter-individual variation among host species. PeerJ **4**:e2367.
Lajbner, Z., R. Pnini, M. F. Camus, J. Miller, and D. K. Dowling. 2018. Experimental evidence that thermal selection shapes mitochondrial genome evolution. Scientific Reports **8**:9500.
Magnacca, K. N., and M. J. F. Brown. 2012. DNA barcoding a regional fauna: Irish solitary bees. Molecular Ecology Resources **12**:990–998.
Magnacca, K. N., and B. N. Danforth. 2006. Evolution and biogeography of native Hawaiian Hylaeus bees (Hymenoptera: Colletidae). Cladistics **22**:393–411.
Michener, C. D. 2007. The Bees of the World. 2 edition. The Johns Hopkins University Press, Baltimore, USA.
Raychoudhury, R., B. K. Grillenberger, J. Gadau, R. Bijlsma, L. van de Zande, J. H. Werren, and L. W. Beukeboom. 2010. Phylogeography of *Nasonia vitripennis* (Hymenoptera) indicates a mitochondrial-*Wolbachia* sweep in North America. Heredity **104**:318–326.
Reynolds, K. T., L. J. Thomson, and A. A. Hoffmann. 2003. The effects of host age, host nuclear background and temperature on phenotypic effects of the virulent *Wolbachia* strain *popcorn* in *Drosophila melanogaster*. Genetics **164**:1027–1034.
Ricardo, C. P., E. Françoso, and M. C. Arias. 2020. Mitochondrial DNA intra-individual variation in a bumblebee species: a challenge for evolutionary studies and molecular identification. Mitochondrion **53**:243–254.
Rokas, A., R. J. Atkinson, G. S. Brown, S. A. West, and G. N. Stone. 2001. Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? Heredity **87**:294–304.
Schuler, H., K. Köppler, S. Daxböck-Horvath, B. Rasool, S. Krumböck, D. Schwarz, T. S. Hoffmeister, B. C. Schlick-Steiner, F. M. Steiner, A. Telschow, C. Stauffer, W. Arthofer, and M. Riegler. 2016. The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. Molecular Ecology **25**:1595–1609.
Shokralla, S., J. F. Gibson, H. Nikbakht, D. H. Janzen, W. Hallwachs, and M. Hajibabaei. 2014. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. Molecular Ecology Resources **14**:892–901.
Smith, F. 1879. Descriptions of New Species of Hymenoptera in the Collection of the British Museum, London: British Museum.
Songram, O., S. Sittipraneed, and S. Klinbunga. 2006. Mitochondrial DNA diversity and genetic differentiation of the honeybee (*Apis cerana*) in Thailand. Biochemical Genetics **44**:256–269.
Uribe-Alvarez, C., N. Chiquete-Félix, L. Morales-García, A. Bohórqueš Hernández, N. Delgado-Buenrostro, L. Vaca, A. Peña, and S. Uribe-Carvajal. 2019. *Wolbachia pipientis* grows in *Saccharomyces cerevisiae* evoking early death of the host and deregulation of mitochondrial metabolism. MicrobiologyOpen **8**:e00675.