Conservation genomics and evolutionary potential of the southern pygmy perch, *Nannoperca australis*, in the Murray-Darling Basin, Australia

Chris Brauer

School of Biological Sciences

Faculty of Science and Engineering

Flinders University

September 2016

Contents

List of figures

List of tables

List of appendices

Summary

Understanding how natural selection generates and maintains adaptive genetic diversity in heterogeneous environments is key to predicting the evolutionary responses of populations to rapid environmental change. The Murray-Darling Basin (MDB) is the largest source of water for agricultural irrigation in Australia and is one of the most fragile and threatened ecosystems in the country. Climate across the region is dominated by a steep gradient of aridity from east to west and hydroclimatic conditions are highly variable and unpredictable. Few areas remain unaffected by anthropogenic disturbance and current native fish populations represent just around 10% of pre-European settlement estimates. This work aims to investigate how environmental variation and human disturbance influence neutral and adaptive genetic variation, population connectivity and variation in gene expression of the southern pygmy perch (*Nannoperca australis*), a threatened freshwater fish undergoing rapid decline in the MDB. Here, I employed two next-generation sequencing methods, ddRAD and RNA-seq to assess adaptive resilience and evolutionary potential within an integrated riverscape genomics and comparative transcriptomics analytical framework. In the first riverscape genomics study of an Australian fish, high-resolution environmental data and 5,162 high-quality filtered SNPs were used to clarify spatial population structure and to assess footprints of selection associated with the hydroclimatic gradient and widespread human disturbance. Findings revealed strong neutral population structure consistent with spatial stream hierarchy, along with evidence that hydroclimatic variation and anthropogenic disturbance is

driving local adaptive divergence of populations. These results contribute to understanding of adaptive evolution in highly fragmented ecosystems and suggest polygenic selection may largely underpin adaptive divergence in the wild. The consequences of habitat fragmentation for population persistence of freshwater biodiversity were then examined with a combination of riverscape genomics and individual-based population genetic simulations. Populations most isolated by recently constructed in-stream barriers showed reduced genetic diversity and increased genetic differentiation, even after accounting for the effects of natural stream hierarchy and environmental variation. These results provide first evidence that the decline of freshwater biodiversity across a riverine ecosystem can be directly attributed to anthropogenic habitat fragmentation. Finally, a *de novo* transcriptome was assembled and used to explore the roles of genetic and environmental variation in the evolution of plastic and divergent gene expression profiles among wild populations. Gene expression plasticity appeared unconstrained by genetic diversity, and comparative transcriptomic analyses identified divergently expressed candidate genes involved in metabolic responses to variations in water quality. These findings suggest that phenotypic plasticity can contribute to evolutionary potential of small populations, and highlight that the compounding effects of climate change and pollution likely pose additional extinction risks for many threatened species. Overall, the results in this thesis challenge the assumption that genetic drift will usually overwhelm selection in small, poorly dispersive populations. Instead, environmental variability within fragmented and disturbed habitat patches may maintain

adaptive genetic variation, and drive the evolution of gene expression plasticity. The extent of recent demographic decline in response to severe habitat fragmentation however underscores the urgent need for well-considered proactive conservation measures to ensure persistence of the species in the MDB.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed.....

Date.....

Acknowledgements

I would like to thank my supervisors, Professor Luciano Beheregaray and Associate Professor Luciana Möller for their personal and academic support since I joined the lab. Your passion for science and all things aquatic is infectious and the enthusiasm with which you approach both the research, and social aspects of MELFU has been inspiring and rewarding.

Members of MELFU – past and present, especially Fabricius and Marina, Yuma, Minami (Mini, you owe me a bento box!), Cat, Steve and Julie, Peter Teske, Hilary, Leslie, Bree, Tess, Pedro and Juliana, Katie, Sean, Edu and Rafa. You guys have provided no end of fun and welcome distraction, and Astrid and I have made many life-long friends. Also thank you to Peter Unmack, Michael Hammer, Mark Adams, Nick Whiterod, Chris Bice, Scotte Wedderburn, Tom Barnes and Simon Westergaard who provided samples, advice and assistance in the field.

I am grateful for financial support provided by the Australian Research Council via a Future Fellowship project to Luciano Beheregaray (FT130101068), ARC Linkage program (LP100200409), and funding provided by the AJ & IM Naylon PhD Scholarship.

I am forever grateful to Karen Burke da Silva. You have been a friend and a mentor for me throughout my time at Flinders and your encouragement and support played no small part in my decision to continue with my studies.

Finally, my beautiful wife Astrid. Your unwavering belief in me, and endless patience has provided me with the opportunity to change my life. You have been there for me throughout, and this thesis is as much a product of your dedication as it is mine.

Its level lands are spread away
To meet the setting sun;
Fierce summers o'er them scorch and slay
The grass blades, one by one;
Long, wicked droughts have dried their breasts,
These virgin lands and clean;
But still a fertile promise rests
Upon the Riverine,
The fecund Riverine.
It rises, and its worth attests
The Phoenix Riverine.

E.J. Brady (1911)

Chapter 1: General introduction

Background

Evolutionary potential and rapid environmental change

As the human population continues to increase, the rapid escalation of development and exploitation of the Earth's natural resources is having a profound effect on every aspect of the environment (Foley *et al.* 2005; Davis *et al.* 2015). The threat these selective forces now pose to global biodiversity is increasing, and has been compounded by recent and rapid climate change (Vitousek *et al.* 1997; Walther *et al.* 2002; Thuiller 2007). As a result, species extinctions are now occurring at a rate far exceeding pre-anthropogenic estimates (Barnosky *et al.* 2011). Understanding the mechanisms by which species may persist in changed, and often sub-optimal conditions is therefore vital for identifying populations at risk and for improving conservation measures that are increasingly employed to mitigate biodiversity loss (Hoffmann & Sgro 2011; Sgrò *et al.* 2011).

Populations faced with environmental change can respond via three non-mutually exclusive mechanisms: genetic evolutionary adaptation, dispersal to more suitable habitat, or acclimation to the altered environment through phenotypic plasticity (Bellard *et al.* 2012; Pauls *et al.* 2013). Evolutionary adaptation is assumed to be essential for species persistence in the face of environmental change (Bradshaw & Holzapfel 2006). However, it remains unclear if natural selection can be effective in small populations with reduced genetic diversity and increased levels of inbreeding (Hoffmann & Willi 2008; Wood *et al.* 2016). Range shifts in response to climate change have already been observed for some species (Davis &

Shaw 2001; Cahill *et al.* 2012), but this may not provide a solution for many others where the opportunity for dispersal is naturally limited or is constrained by recent habitat fragmentation (Dawson *et al.* 2011). Phenotypic responses to environmental change have also been observed (Charmantier *et al.* 2008; Hendry *et al.* 2008), and while many suggest plasticity likely plays an important role in initial rapid responses to environmental change, there can be considerable fitness costs associated with plasticity (Gienapp *et al.* 2008). Additionally, some evidence suggests that the evolution of plasticity may be constrained by genetic effects, however few studies have examined this in wild populations (Papakostas *et al.* 2014; Wood & Fraser 2015). For many species, it is thus likely that a combination of these evolutionary processes will be required to withstand the rapid rate of global change (Visser 2008; Quintero & Wiens 2013), and understanding how evolutionary, demographic and environmental processes interact to shape species evolutionary potential is a key area of research (Harrisson *et al.* 2014).

Conservation genomics

Population genetics has traditionally relied on small numbers of selectively neutral markers (e.g. mitochondrial DNA (mtDNA), microsatellites) to examine demographic parameters such as effective population size, genetic drift, gene flow, and population structure (Avise 2010; Ouborg *et al.* 2010). Accordingly, the foundation of conservation genetics has also been built on investigating the demographic effects of genetic drift and gene flow on small populations (Frankham 1995). It is now well

established however, that these genetic markers often provide poor surrogates for quantitative genetic variation that provides the basis of evolutionary potential (Reed & Frankham 2001). As conservation genetics is evolving into conservation genomics, the large number of markers generated by next-generation sequencing (NGS) has improved the resolution of demographic inferences (Luikart *et al.* 2003; Allendorf *et al.* 2010). However, it is the ability to identify genomic regions putatively under selection that now allows us to begin to assess evolutionary potential for non-model and threatened species (Allendorf *et al.* 2010; Frankham 2010).

The data for this thesis were generated using two NGS methods, with all laboratory procedures and bioinformatic analyses performed in-house at the Molecular Ecology Lab at Flinders University (MELFU). Double-digest restriction site associated DNA (ddRAD) sequencing was used to simultaneously discover and genotype single nucleotide polymorphism (SNP) markers using a protocol based on Peterson *et al*. (2012). This method is one of the most popular of a number of reduced-representation sequencing approaches developed over the last few years. Such methods have transformed the field of molecular ecology by providing a fast and cost effective way of applying powerful NGS technology to non-model organisms (Andrews *et al.* 2016). Briefly, ddRAD uses two restriction enzymes to fragment genomic DNA followed by a size selection procedure to reduce the number of regions prior to sequencing on an Illumina

platform (Peterson *et al.* 2012). Loci identified using RAD methods are found throughout the genome and are useful for addressing a range of ecological and evolutionary questions as they represent the full history of evolutionary processes shaping a population or species (Narum *et al.* 2013). In addition to ddRAD data, RNA sequencing (RNA-seq) was also used in this thesis to quantify transcriptional variation among selected key populations. RNA-seq is a recently developed NGS method for simultaneously mapping and quantifying the transcriptome by sequencing cDNA reverse transcribed from RNA (Wang *et al.* 2009). While relatively expensive compared to RAD sequencing, there has recently been a rapid increase in RNA-seq studies of non-model species (Alvarez *et al.* 2015; DeBiasse & Kelly 2016). Gene expression levels are often inferred from RNA-seq data and when applied to non-model organisms this can contribute to our understanding of the molecular basis (both plastic and evolved) of physiological responses to environmental stressors (Whitehead *et al.* 2010; Romero *et al.* 2012).

Southern pygmy perch in the Murray-Darling Basin

The study system of this thesis is the southern pygmy perch, *Nannoperca australis* (Teleostei: Percichthyidae). This is a small-bodied freshwater fish (<85 mm) endemic to southeastern Australia, including the Murray-Darling Basin (MDB) (Unmack *et al.* 2013) (Figure 1.1A). Normally found in shallow streams and wetlands, this ecological specialist prefers sheltered micro-habitats and aquatic macrophyte cover, is relatively short lived (3–6 years) – reaching sexual maturity within one year, and possesses limited

dispersal ability (Lintermans 2007; Wedderburn *et al.* 2012). Climate in this region has been increasingly dry but highly variable throughout the late Holocene (the last ~3,000 years), with hydroclimatic conditions characterised by a steep gradient of aridity from east to west with more consistent rainfall and cooler temperatures in the southeast highlands and drier, semi-arid conditions in the western lowlands (Figure 1.1B) (Donders *et al.* 2007; Pittock & Finlayson 2011). Recent studies of wild populations of *N. australis* demonstrated that hydroclimatic-related factors, and in particular the variation in predictability of flow typical for many Australian rivers (Kennard *et al.* 2010), influence individual fitness and drive predictive patterns of local adaptation in key reproductive traits and life-history strategy (Morrongiello *et al.* 2010; Morrongiello *et al.* 2012; Morrongiello *et al.* 2013).

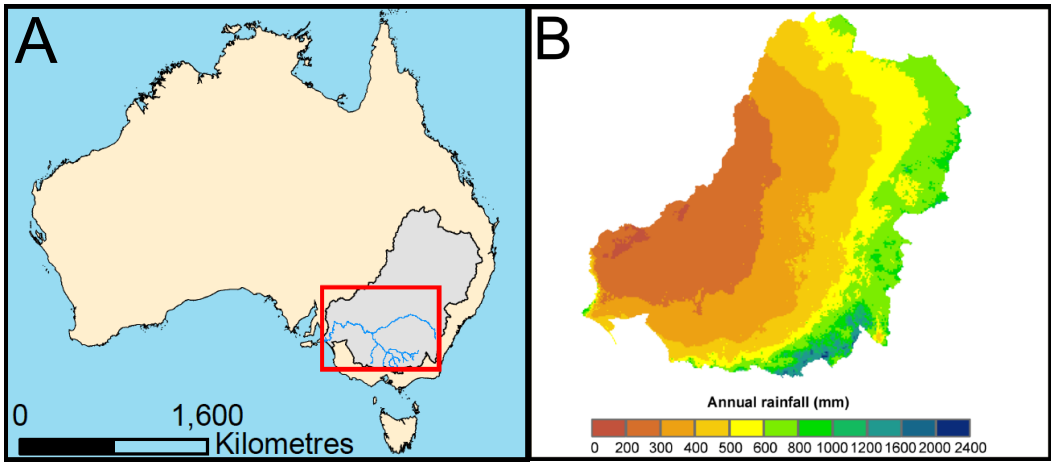


Figure 1.1: A) Location of study area (red box) in the Murray–Darling Basin (MDB; shaded area). B) The dominant climatic gradient represented by average annual rainfall across the MDB (reproduced from Chiew *et al.* 2008).

Since European settlement of this region began some 160 years ago, the MDB has suffered from extensive agricultural development (e.g. wetland reclamation, water abstraction), river regulation, construction of thousands of barriers to fish passage and the introduction of exotic species (Lintermans 2007; Balcombe *et al.* 2011; Laurance *et al.* 2011). These factors have likely synergistically contributed to the dramatic decline of native fishes in recent times with more than half of all MDB species, including *N. australis*, now listed as threatened or of conservation concern (Lintermans 2007). Compounding these issues, an unprecedented severe and prolonged drought between 1997 and 2010 caused catastrophic loss of habitat and local extinction for some *N. australis* populations, particularly in the Lower Murray (Wedderburn *et al.* 2012; Hammer *et al.* 2013). In response to the decline, several conservation-breeding and restoration programs were initiated (Hammer *et al.* 2013; Attard *et al.* 2016b) and additional translocations among populations of wild fish have recently been proposed.

The long-term metapopulation dynamics for *N. australis* has likely been strongly influenced by a combination of the hierarchical river network structure and the unpredictable and highly variable hydroclimatic conditions in the MDB. Recent studies within the MDB, based on allozymes, mtDNA and microsatellites, showed that *N. australis* has shallow basin-wide phylogeographic (i.e. historical) divergence (Figure 1.2) (Unmack *et al.* 2013; Cole *et al.* 2016). Similar phylogeographic

patterns have also been reported for many other widespread MDB fishes (e.g. Faulks *et al.* 2010a; Unmack *et al.* 2013) and are likely a result of the much greater flow of ancestral MDB rivers during the Pleistocene and also due to enhanced connectivity via the large inland lakes that inundated its lower reaches (Pels 1964). In contrast to low historical phylogeographic structure, strong contemporary population structure appears related to a recent history of isolation of populations (Cole *et al.* 2016). This is supported by coalescent analyses based on microsatellites that suggest that isolation and demographic declines observed for some *N. australis* populations are associated with modification and fragmentation of the MDB that post-dates the recent European settlement in the region (Attard *et al.* 2016b; Cole *et al.* 2016).

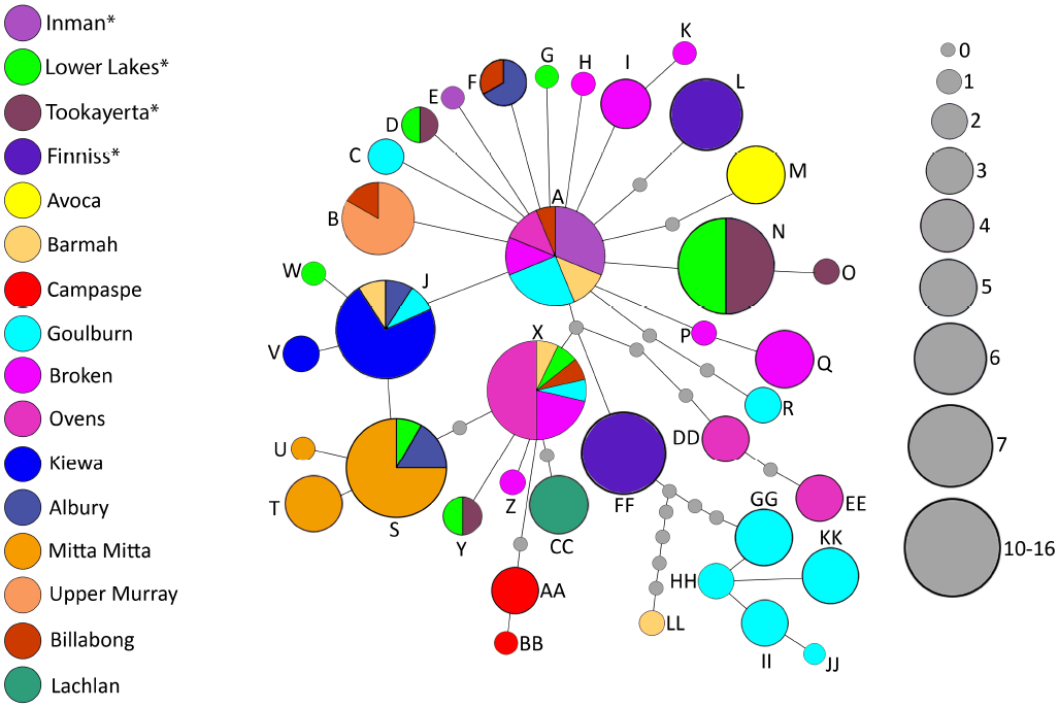


Figure 1.2: Network of cytochrome b mtDNA sequence data for *Nannoperca australis* demonstrating sharing of several haplotypes between multiple catchments and shallow phylogeographic divergence across the Murray–Darling Basin (MDB) (reproduced from Cole *et al.* 2016). Each mtDNA haplotype is colour-coded based on the catchment where it occurs and asterisks indicate that the river is in the Lower Murray region. Rivers are listed in their geographic order from downstream to upstream. Details for the specific geographic location for each haplotype are in Cole *et al*. 2016; Table 1. Circle size indicates haplotype abundance in the sample.

This overall biogeographic scenario indicates that historically, populations across the MDB were likely larger and more connected, despite highly variable natural environmental conditions. This can be considered a feature of this system (e.g. Faulks *et al.* 2010a; Unmack *et al.* 2013) that contrasts to scenarios in more commonly studied Northern Hemisphere fishes such as salmonids (e.g. Hecht *et al.* 2015) or sticklebacks (e.g. Raeymaekers *et al.* 2014; Ferchaud & Hansen 2016). In the latter

systems, genetic structure and patterns of diversity are often influenced by demographic expansion and secondary contact following extensive Pleistocene glaciations (Bernatchez & Wilson 1998).

Although the effects of drift have probably recently intensified in the MDB, highly variable hydroclimatic conditions, and evidence for local adaptation of populations to these conditions, suggest the genetic signal of local adaptation is unlikely to have been completely eroded. This makes *N. australis* an ideal and unique system to examine the genomic basis of evolutionary potential in a threatened, low dispersal species recently challenged by rapid and pervasive environmental change.

Significance and justification

Soulé (1985) laid out the biological and philosophical principles that he, and many others have since, regarded as central to the field of conservation biology. One of the most fundamental messages was that conservation measures should ensure the long-term viability of ecosystems, and that to do this, efforts should be directed at preserving the natural evolutionary processes that sustain biodiversity. That is not to say that conservation of single species, populations or even very small groups of individuals is not a worthwhile pursuit. Rather that these crisis-oriented objectives are a necessary response to symptoms of the decay of ecological and evolutionary systems operating on a much larger scale,

and that they must form only part of a more holistic approach.

Understanding the potential for populations to adapt to rapid environmental change and to persist in modified and often degraded habitats is now a global research priority. This is particularly the case for freshwater ecosystems, which are amongst the most highly impacted by human activities and where the decline of biodiversity has been most severe (Sala *et al.* 2000; Strayer & Dudgeon 2010). Soulé also appreciated that although science can provide an objective and rigorous framework with which to address conservation issues; science and society are inescapably entwined and, for conservation biology to remain relevant, it must incorporate and respect the sometimes-conflicting demands of political and socio-economic values of the time. Further complicating the dialogue between science and society is the common necessity that conservation management decisions are made quickly, often based on incomplete or imperfect data (Soulé 1985). The inherently complex nature of biological systems at every level, from molecules to biomes, ensures that our scientific understanding of the ecological implications of conservation management actions may never be absolute. From the outset however methods in conservation biology have constantly evolved in an effort to provide society with faster, cheaper, and more precise information with which to make more informed conservation decisions. It is in this context that conservation biology now enters the *'omics* era, and it was with these principles and objectives in mind that the following work presented in this thesis was undertaken.

This thesis represents the first conservation genomics study of an Australian fish and, as such, makes an original contribution to our understanding of the evolutionary, demographic and environmental determinants of species evolutionary potential across large freshwater ecosystems – a topic dominated by studies of Northern Hemisphere fishes. The recent transition from conservation genetics to conservation genomics has not only exposed new analytical challenges for the field, but also demands a renewed emphasis on strong communication between researchers and conservation managers (Shafer *et al.* 2015a). Here, several novel solutions to difficulties in inferring selection in complex spatial environments are demonstrated, and initial steps are made towards a framework for the practical implementation of modern genomic approaches to applied conservation. More specifically, recommendations based on several key findings here will inform ongoing, and recently proposed future conservation management actions for *N. australis*, and potentially for other threatened MDB species. Additionally, the work presented in this thesis will contribute to the current political and social dialogue in Australia concerning broader aspects of sustainable water resource management and freshwater biodiversity conservation.

This thesis forms part of a wider program of ecological genomics research at the MELFU aimed at understanding the evolutionary implications of rapid environmental changes for Australian freshwater fishes (Australian Research Council Future Fellowship project FT130101068). This broader

project capitalises on the diverse range of life history characteristics of five species of Percichthyidae, including the focal species of this thesis, *N. australis*, to increase our understanding of local adaptation to hydroclimatic variation, and population persistence in human modified ecosystems. Complementing the work presented here, another two low dispersal, small-bodied species are included in the project, the eastern and Yarra pygmy perches, *N. flindersi* and *N. obscura*, respectively. Together, the three *Nannoperca* species contrast with two larger-bodied and highly mobile species of golden perches (*M. ambigua* species complex) to provide a comparative framework with which to assess evolutionary potential of a family of endemic Australian freshwater fishes across a range of spatial scales and selective gradients.

During my candidature I have also contributed conceptual, analytical and written components of several companion publications (three published, one in review and one in preparation) related to this broader project, but not specifically part of my thesis. These are listed below and those already published are provided in Appendix 1.

Attard C, Möller L, Sasaki M, Hammer M, Bice C, Brauer C, Carvalho D, Harris J & Beheregaray L (2016) A novel holistic framework for genetic-based captive-breeding and reintroduction programs. *Conservation Biology*, 30, 1060–1069.

Cole T, Hammer M, Unmack P, Teske P, Brauer C, Adams M & Beheregaray L (2016) Range-wide fragmentation in a threatened fish associated with post-European settlement modification in the Murray-Darling Basin, Australia. *Conservation Genetics*, doi:10.1007/s10592-10016-10868-10598.

Attard C, Brauer C, Van Zoelen J, Sasaki M, Hammer M, Morrison L, Harris J, Möller L & Beheregaray L (2016) Multi-generational evaluation of genetic diversity and parentage in captive southern pygmy perch (*Nannoperca australis*). *Conservation Genetics*, doi:10.1007/s10592-10016-10873-y.

Sandoval-Castillo J, Attard C, Marri S, Brauer C, Möller L & Beheregaray L (in review) SWINGER: a user-friendly computer program to establish captive breeding groups that minimize relatedness without pedigree information. *Molecular Ecology Resources*.

Smith S, Brauer C, Sasaki M, Unmack P, Guillot G, Bernatchez L & Beheregaray L (in preparation) Latitudinal variation in adaptive resilience in an ecologically important aquatic species: a genomic test of the climatic variability hypothesis.

Thesis Outline

This thesis is comprised of a general introduction, three data chapters and a conclusion. It is intended that this first chapter is brief, providing general background information related to the study system and broader significance of the thesis. The following data chapters are written as stand-alone manuscripts for publication and more detailed background

information relevant to each chapter is provided in separate introductory sections. The final concluding chapter provides a synthesis of the findings from all three data chapters and outlines ongoing and future research efforts related to, and arising from the work presented here.

Chapter 2: *Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin*

A combination of population genomics and genotype–environment association analyses are used to assess the environmental factors influencing population structure and local adaptation for *N. australis* across the MDB. Strong neutral population structure consistent with spatial stream hierarchy was found, along with evidence that hydroclimatic variation and anthropogenic disturbance is driving local adaptive divergence of populations. This chapter has been published in *Molecular Ecology* (Brauer C, Hammer M, Beheregaray L (2016) Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, DOI:10.1111/mec.13830.).

Chapter 3: *Anthropogenic habitat fragmentation increases extinction risk for freshwater species*

In this chapter, population genomic analyses and individual-based population genetic simulations are employed to test several hypotheses related to how the recent and extensive construction of in-stream barriers across the MDB threatens the long-term persistence of *N. australis*. Here, for the first time, clear evidence is presented that recent anthropogenic habitat fragmentation has contributed to the decline of freshwater biodiversity across an entire riverine ecosystem.

Chapter 4: *Comparative ecological transcriptomics and the contribution of gene expression to evolutionary potential of a threatened freshwater fish*

RNA sequencing is used to construct a *de novo* transcriptome and to address several outstanding questions related to whether variation in gene expression might be constrained by genetic variation. The chapter also explores how plastic and divergent gene expression profiles may evolve in response to environmental and genetic variation in wild populations. Gene expression plasticity appeared unrelated to genetic diversity, however divergent patterns of gene expression were observed for genes potentially responding to variation in water quality.

Chapter 5: *Conservation genomics and evolutionary potential of a threatened Australian freshwater fish*

This chapter concisely summarises the major findings of the thesis in the context of evolutionary potential of *N. australis* and the three evolutionary responses to environmental change examined across the three data chapters: genetic adaptation, dispersal, and phenotypic plasticity. Concluding remarks outline the implications of this work for ongoing research and future conservation programs for *N. australis* and other freshwater fishes.

Chapter 2: Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin

Chris J. Brauer[1], Michael P. Hammer[2], Luciano B. Beheregaray[1]

[1] Molecular Ecology Laboratory, School of Biological Sciences, Flinders University, Adelaide, SA, 5042, Australia

[2] Natural Sciences, Museum and Art Gallery of the Northern Territory, Darwin, NT 0801, Australia

Molecular Ecology (2016) doi: 10.1111/mec.13830

This chapter has been published in Molecular Ecology and is reproduced with permission. I am the primary author with Dr Michael Hammer and Prof Luciano Beheregaray as co-authors. I was responsible for data collection, analyses and drafting the manuscript. Michael Hammer provided samples and provided comments on the manuscript. Luciano Beheregaray supervised the project and provided comments on the manuscript.

Abstract

Understanding how natural selection generates and maintains adaptive genetic diversity in heterogeneous environments is key to predicting the evolutionary response of populations to rapid environmental change. Detecting selection in complex spatial environments remains challenging, especially for threatened species where the effects of strong genetic drift may overwhelm signatures of selection. We carried out a basin-wide riverscape genomic analysis in the threatened southern pygmy perch (*Nannoperca australis*), an ecological specialist with low dispersal potential. High-resolution environmental data and 5,162 high-quality filtered SNPs were used to clarify spatial population structure and to assess footprints of selection associated to a steep hydroclimatic gradient and to human disturbance across the naturally and anthropogenically fragmented Murray-Darling Basin (Australia). Our approach included F_{ST} outlier tests to define neutral loci, and a combination of spatially explicit genotype-environment association analyses to identify candidate adaptive loci while controlling for the effects of landscape structure and shared population history. We found low levels of genetic diversity and strong neutral population structure consistent with expectations based on spatial stream hierarchy and life-history. In contrast, variables related to precipitation and temperature appeared as the most important environmental surrogates for putatively adaptive genetic variation at both regional and local scales. Human disturbance also influenced variation in candidate loci for adaptation, but only at a local scale. Our study contributes to understanding of adaptive evolution along naturally and

anthropogenically fragmented ecosystems. It also offers a tangible example of the potential contributions of landscape genomics for informing *in situ* and *ex situ* conservation management of biodiversity.

Introduction

The effects of human development and recent climate change on our natural environment are pervasive and the threat these selective forces pose to global biodiversity is increasing (Vitousek *et al.* 1997; Walther *et al.* 2002; Thuiller 2007). Populations faced with environmental change can respond through range shifts, acclimation through phenotypic plasticity or by genetic evolutionary adaptation to their new environment (Bellard *et al.* 2012; Pauls *et al.* 2013). If one, or some combination of these processes does not occur, populations and potentially entire species face an increased risk of extinction (Quintero & Wiens 2013). Range shifts in response to climate change have already been observed for some species (Davis & Shaw 2001; Cahill *et al.* 2012), but this can be problematic for those where the opportunity for dispersal is naturally limited or constrained by recent habitat fragmentation (Dawson *et al.* 2011). Phenotypic responses to environmental change have also been observed (Charmantier *et al.* 2008; Hendry *et al.* 2008), but there are costs and limits associated with plasticity (DeWitt *et al.* 1998) that make it unlikely to provide long-term solutions for many populations (Gienapp *et al.* 2008). This leads to several questions concerning the capacity of species to persist *in situ* and to adapt to altered environmental conditions. The currently rapid rate of climate and environmental change suggests that evolutionary adaptation will need to rely heavily on standing genetic variation (Barrett & Schluter 2008). However, for threatened species it is unclear if enough variation exists at adaptively important loci to facilitate an evolutionary response. Thus, it is important to ask how are threatened

populations locally adapted, what are the important environmental factors contributing to local adaptation, and how is adaptive genetic variation spatially distributed and maintained?

Landscape genomics (LG) provides an ideal framework for addressing questions in ecology and evolution, which have become particularly relevant in threatened biotas that are both naturally and anthropogenically fragmented. This rapidly growing research field combines information about environmental heterogeneity and genome-wide data of individuals sampled across the landscape to identify spatial patterns of neutral and adaptive variation (Manel *et al.* 2010; Sork & Waits 2010). Although there has been recent debate over the practical application of genomics to conservation (Garner *et al.* 2015; Shafer *et al.* 2015a; Shafer *et al.* 2015b), LG has increasingly been applied to threatened and non-model species (Cooke *et al.* 2012b; Cooke *et al.* 2012a; Limborg *et al.* 2012; Bourret *et al.* 2013; Cooke *et al.* 2014; Moore *et al.* 2014; Steane *et al.* 2014; Hand *et al.* 2015; Hecht *et al.* 2015; Laporte *et al.* 2015; Funk *et al.* 2016). In the context of conservation, the large number of markers generated by next-generation sequencing (NGS) can improve the resolution of demographic inferences (Luikart *et al.* 2003; Allendorf *et al.* 2010), but it is the identification of genomic regions under selection that has held great promise for increasing our understanding of the potential vulnerability or resilience of biodiversity to environmental change (Allendorf *et al.* 2010; Frankham 2010). There are however several characteristics common to

many threatened species, such as small effective population sizes, population isolation, repeated local extinction-recolonisation cycles and inbreeding, that may affect our ability to distinguish signals of selection from other confounding effects. These factors therefore need to be considered when selecting an analytical framework for LG studies of natural populations (Schoville *et al.* 2012).

The most common methods used to detect selection in LG studies are based on population genetics theory and the assumption that demographic processes such as migration and genetic drift should affect the genome uniformly while selection should act on specific regions or loci (Lewontin & Krakauer 1973). Known as F_{ST} outlier tests, they have become a standard feature of most LG studies. These tests however assume specific demographic models and may not be robust to violations imposed by non-equilibrium demographic scenarios (Lotterhos & Whitlock 2014; Whitlock & Lotterhos 2015). Nevertheless, aside from considering outliers as candidates for selection, these tests offer an effective solution for creating a large neutral dataset for improving inferences about population structure and demographic history (Luikart *et al.* 2003; Allendorf *et al.* 2010).

Genotype-environment association (GEA) approaches are an alternative strategy for detecting the signal of local adaptation by testing for direct associations between allele frequencies and environmental parameters.

These methods are generally free from the constraints of simple demographic models. They can be used to test specific hypotheses related to environmental heterogeneity, including the possibility that it shapes polygenic adaptation in natural populations (Lasky *et al.* 2012; Bourret *et al.* 2014; Hecht *et al.* 2015). Moreover, GEA approaches generally also incorporate means to account for the effects of shared population history and can separate geographic and environmental effects (Joost *et al.* 2007; Coop *et al.* 2010; Frichot *et al.* 2013; Guillot *et al.* 2014; Rellstab *et al.* 2015). This is particularly important for complex spatial environments such as dendritic river networks where physical landscape structure can greatly affect patterns of genetic variation (Hughes *et al.* 2009; Fourcade *et al.* 2013; Thomaz *et al.* 2016).

Here we use a framework that capitalizes on a high-resolution environmental dataset and on powerful LG approaches to assess footprints of selection in a threatened species found across a hydroclimatically heterogeneous and anthropogenically modified ecosystem. Our study system, the southern pygmy perch *Nannoperca australis*, is a small-bodied freshwater fish (<85 mm) endemic to southeastern Australia, including the Murray-Darling Basin (MDB) (Unmack *et al.* 2013). This ecological specialist is normally associated with streams and wetlands, sheltered micro-habitats and aquatic macrophyte cover, is relatively short lived (3–6 years; reaches maturity within one year), has large demersal eggs, and limited dispersal ability

(Lintermans 2007; Wedderburn *et al.* 2012). Climate in this region has been increasingly dry but highly variable throughout the late Holocene (~3,000 years), with conditions characterised by a steep gradient of aridity from east to west with higher, and more consistent rainfall and lower temperatures in the southeast highlands and drier, semi-arid conditions in the western lowlands (Donders *et al.* 2007; Pittock & Finlayson 2011).

Importantly, recent studies of wild populations of *N. australis* demonstrated that hydroclimatic-related factors, and in particular the variation in predictability of flow typical for many Australian rivers (Kennard *et al.* 2010), influence individual fitness and drive predictive patterns of local adaptation in key reproductive traits and life-history strategy (Morrongiello *et al.* 2010; Morrongiello *et al.* 2013). In addition, female reproductive investment in egg and clutch size in *N. australis* varies predictability among populations along gradients of stream flow (Morrongiello *et al.* 2012). Instead of merely reflecting spatial phenotypic plasticity, these findings also support bet-hedging as a co-evolved adaptive strategy in *N. australis*, a view consistent with increasing theoretical and empirical evidence about the consequences of female investment in the evolution of life-histories (Olofsson *et al.* 2009; Morrongiello *et al.* 2012).

Within the MDB, studies based on allozymes, mitochondrial DNA (mtDNA) and microsatellites showed that *N. australis* has very shallow basin-wide

phylogeographic divergence but strong contemporary population structure shaped by the hierarchical drainage network (Unmack *et al.* 2013; Cole *et al.* 2016). In fact, coalescent analyses based on microsatellites suggest that isolation and demographic decline observed for some *N. australis* populations is associated with modification and fragmentation of the MDB that post-dates the recent European settlement in Australia (Attard *et al.* 2016b; Cole *et al.* 2016). Genetic evidence thus indicates that the metapopulation structure of *N. australis* does not reflect deeply historic isolation across its range in the MDB. Similar phylogeographic patterns have also been reported for many other widespread MDB fishes (e.g. Faulks *et al.* 2010a; Unmack *et al.* 2013) and are likely a result of the much greater flow of ancestral MDB rivers during the Pleistocene and the large inland lakes that inundated its lower reaches (Pels 1964).

The overall biogeographic scenario indicates that historically, populations across the MDB were likely larger and more connected despite highly variable natural environmental conditions. This suggests that although the effects of drift have probably recently intensified, the signal of local adaptation is unlikely to have been completely eroded, and that appropriate LG frameworks have the potential to address questions about hydroclimatic adaptation in *N. australis*. Since European settlement the MDB has suffered from extensive development (e.g. wetland reclamation), river regulation, construction of thousands of barriers to fish passage and the introduction of exotic species (Lintermans 2007; Balcombe *et al.* 2011;

Laurance *et al.* 2011). These factors have likely synergistically contributed to the widespread decline of *N. australis* populations, and to its current listing as endangered in two Australian states (Hammer *et al.* 2013; Cole *et al.* 2016). An unprecedented severe and prolonged drought between 1997 and 2010 caused catastrophic loss of habitat and local extinction for some *N. australis* populations, particularly in the Lower Murray (Wedderburn *et al.* 2012; Hammer *et al.* 2013). In response to the decline several conservation-breeding and restoration programs were initiated (Hammer *et al.* 2013; Attard *et al.* 2016b) and additional translocations among populations of wild fish have been proposed. Given the ongoing conservation management of *N. australis*, and that climate change is expected to negatively impact its populations even further in the future (Perry & Bond 2009; Balcombe *et al.* 2011; Morrongiello *et al.* 2011a), it is important to understand how extant populations are adapted to local environmental conditions. More broadly, our study system also provides an opportunity for asking if recent human-driven selection has impacted the genome of extant populations.

In this study we test the core hypothesis that the steep hydroclimatic gradient across the MDB has contributed to adaptive genetic divergence of *N. australis* populations. This is based on the premise that the natural flow regime modulates many abiotic and biotic processes (Poff *et al.* 1997), such as habitat connectivity, physical disturbance, resource availability and ecological interactions, that have direct implications for

shaping the genetic architecture of widespread aquatic species. We also explore features of our study system to address three questions that have broad implications to conservation and ecological genomics. First, can LG be used to distinguish signals of selection from other confounding effects (e.g. strong drift) in a threatened, poor dispersive species? Second, can selection due to human disturbance be distinguished from selection due to natural environmental heterogeneity? Third, can GEA approaches detect genomic footprints of polygenic adaptation due to hydroclimatic heterogeneity? To answer these questions we employ a combination of recently developed spatially explicit GEA approaches within a riverscape genomics framework that integrates environmental and genome-wide datasets. These approaches are used to test for associations between population allele frequencies and a suite of environmental variables describing variation in climate, hydrology and human disturbance, while controlling for the effects of landscape structure and shared population history. We also discuss our results in the context of ongoing conservation efforts and the utility of genomics for guiding proactive conservation strategies such as translocations for genetic rescue, and for increasing the adaptive potential of populations in the face of ongoing climate change.

Methods

Sampling

Samples of *N. australis* were collected from the wild between 2000 and 2013 using netting, box trapping, or electrofishing. They were preserved either as frozen specimens or fin clips in 99% ethanol and curated at the South Australian Museum, Adelaide. Initially, 550 individuals were

sampled from 38 locations. A smaller sample was then selected to include all known populations in the MDB previously identified with allozyme, mtDNA and microsatellites (Unmack *et al.* 2013; Cole *et al.* 2016) while accommodating for unsatisfactory DNA quality for genomic analysis obtained from some individuals. This resulted in a final, high quality data set of 263 individuals sampled from 25 locations and encompassing 13 catchments across the entire current MDB distribution of *N. australis* (Figure 2.1; Table 2.1).

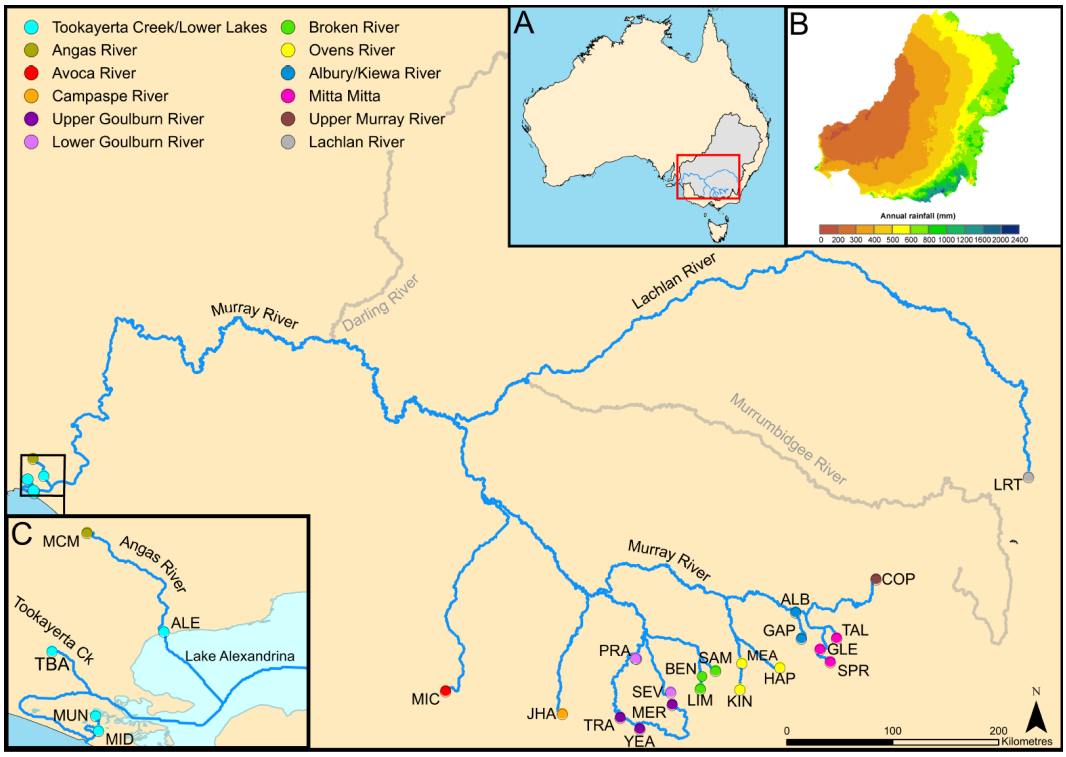


Figure 2.1: *Nannoperca australis* sampling locations covering the entire current distribution of the species in the Murray-Darling Basin (MDB). Sites are colour coded by catchment. Inset A shows the location of the MDB (shaded area) and inset B depicts the rainfall gradient across the basin. Inset C shows the Lower Murray sampling locations. The historical distribution of the species was essentially continuous within the MDB from the Lower Murray to the upper reaches of the Murray, Murrumbidgee and Lachlan Rivers (but excluding the Darling River system), although local abundance likely varied substantially across that range (Llewellyn 1974).

Table 2.1: Information about localities and sample sizes for *Nannoperca australis* from the Murray-Darling Basin (MDB). Lowland wetland sites referred to as Lower Murray in the text are indicated in bold.

Catchment	Site	Location	N	Latitude	Longitude
Tookayerta (TOO)	**TBA**	**Tookayerta Ck, Black Swamp**	**7**	**-35.428**	**138.834**
Lower Lakes (LMR)	**ALE**	**Turvey's Drain, L. Alexandrina**	**10**	**-35.395**	**139.008**
	MID	**Mundoo Is., L. Alexandrina**	**7**	**-35.549**	**138.915**
	MUN	**drain off Mundoo Channel**	**6**	**-35.520**	**138.904**
Angas (ANG)	MCM	Middle Ck	9	-35.250	138.887
Avoca (AVO)	MIC	trib to Middle Ck, Warrenmang	11	-37.028	143.338
Campaspe (CAM)	JHA	Jews Harp Ck, Sidonia	12	-37.139	144.578
Upper Goulburn (UGO)	MER	Merton Ck	17	-36.981	145.727
	TRA	Trawool Ck	10	-37.135	145.193
	YEA	Yea R., Yea	8	-37.213	145.414
Lower Goulburn (LGO)	PRA	Pranjip Ck	9	-36.623	145.309
	SEV	trib to Seven Creeks	11	-36.875	145.701
Broken (BRO)	BEN	Swanpool Ck, Swanpool	10	-36.723	146.022
	SAM	Sam Ck	10	-36.661	146.152
	LIM	Unnamed Ck, Lima South	18	-36.826	146.008
Ovens (OVE)	KIN	King R., Cheshunt	16	-36.795	146.424
	HAP	Happy Valley Ck	9	-36.579	146.824
	MEA	Meadow Ck, Moyhu	8	-36.573	146.423
Kiewa (KIE)	GAP	Gap Ck, Kergunyah	12	-36.317	147.022
Albury (ALB)	ALB	Murray R. lagoon, Albury	12	-36.098	146.928
Mitta Mitta (MIT)	SPR	Spring Ck	10	-36.499	147.349
	GLE	Glencoe Ck	10	-36.393	147.221
	TAL	Tallangatta Ck	7	-36.281	147.382
Upper Murray (COP)	COP	Coppabella Ck	16	-35.746	147.729
Lachlan (LAC)	LRT	Blakney Ck	8	-34.736	149.180

Molecular methods and bioinformatics

DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturers protocol. DNA integrity was assessed by

gel electrophoresis and purity measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

Double-digest Restriction-site Associated DNA (ddRAD) sequencing libraries were prepared following a protocol modified from Peterson et al. (2012). Libraries were multiplexed with 48 samples randomly assigned to each of six Illumina lanes and sequenced on a HiSeq2000 platform as paired-end, 100-bp reads. Raw sequences were demultiplexed using the *process_radtags* component of *Stacks* v.1.04 (Catchen *et al.* 2011) before *de novo* assembly of a reference catalogue and genotyping was performed with *dDocent.FB* v.1.2 (Puritz *et al.* 2014). Details about library preparation and bioinformatics are provided in Appendix 2. *dDocent* combines several existing software packages into a single pipeline designed specifically for paired-end RAD data; i.e. it takes advantage of both forward and reverse reads for SNP discovery. The resulting variant call file (VCF) was filtered to retain only variants present in at least 70% of individuals and in 70% of populations. Complex variants (multi-nucleotide polymorphisms and composite insertions and substitutions) were decomposed into SNP and indel representation following Puritz *et al*. (2014), retaining only one biallelic SNP per locus with a minimum minor allele frequency (MAF) of 0.05. A further six filtering steps were performed to remove SNPs likely to be the result of sequencing errors, paralogs, multi-copy loci or artefacts of library preparation (Appendix 2).

Detecting neutral and outlier loci

Loci not conforming to neutral expectations were detected using a Bayesian approach with *BayeScan* v.2.1 (Foll & Gaggiotti 2008), and the coalescent-based *FDIST* method (Beaumont & Nichols 1996) in *Arlequin* v.3.5 (Excoffier & Lischer 2010). *BayeScan* was run for 100,000 iterations using prior odds of 10,000 and loci significantly different from zero and with a *q*-value less than 0.1 (false discovery rate (FDR) of 10%) were considered outliers. *Arlequin* was run with 50,000 simulations of 13 groups, each with 100 demes and *P*-values were corrected for multiple testing using the *p.adjust* function in R (R Core Team 2015). The hierarchical island model was specified (Excoffier *et al.* 2009), as it allows for the assumption of lower migration among catchments than among sampling sites within catchments. Loci significantly outside the neutral distribution at a FDR of 10% were considered as outliers.

The remaining, putatively neutral SNPs were tested for departure from Hardy-Weinberg equilibrium (HWE) in *Genodive* v.2.0b27 (Meirmans & Van Tienderen 2004). Significance was tested using 10,000 random permutations and loci were subsequently removed if found to depart from HWE at a FDR of 10% in more than 50% of sampling locations.

Genetic diversity, N_e and population structure

Expected heterozygosity (H_E) and observed heterozygosity (H_O) were calculated in *Genodive* for both the neutral and the candidate loci.

Percentage of polymorphic loci was calculated in *GenAlEx* v.6.5 (Peakall & Smouse 2012). To evaluate whether populations have experienced recent genetic bottlenecks, we used *Bottleneck* 1.2.02 (Piry *et al.* 1999). *Bottleneck* was run using the infinite alleles model (IAM) and a Wilcoxon signed-rank test implemented using the *wilcox.test* function in R was used to test for significant heterozygosity excess compared to expectations under mutation-drift equilibrium. We estimated effective population size (N_e) using the linkage disequilibrium (LD) method in *NeEstimator* 2.01 (Do *et al.* 2014). This is based on the assumption that LD at independently segregating loci in a finite population is a function of drift, and performs particularly well with a large number of loci in situations where population sizes are expected to be small. *NeEstimator* was run assuming random mating and using a P_{crit} value of 0.075 following guidelines for small sample sizes (Waples & Do 2010). A Wilcoxon signed-rank test was used to test for differences in N_e estimates between lower and upper Murray regions.

Population-specific F_{ST} was estimated for each sampling site for both the neutral and candidate loci using the method of Weir and Hill (2002) calculated with the R package *hierfstat* (Goudet 2005). Population-specific F_{ST} estimates local population divergence from the whole metapopulation considering variation in the strength of genetic drift among demes due to differences in effective population size (Foll & Gaggiotti 2006).

Population genetic structure was assessed using the neutral loci with a combination of frequency- and genotype-based methods. Pairwise F_{ST} (Weir & Cockerham 1984) was estimated among sampling sites using *Genodive* with significance assessed using 10,000 permutations. *Genodive* was also used to perform a hierarchical AMOVA based on F_{ST} (Weir & Cockerham 1984) among catchments, among sites within catchments and among individuals within sites using 10,000 permutations. In all cases missing data were replaced with alleles drawn randomly from the overall allele frequency distribution. Bayesian clustering analysis of individual genotypes was then performed using *fastStructure* (Raj *et al.* 2014). This model-based method assumes there are *K* populations, and that population allele frequencies are in HWE. Individuals are assigned to one or more populations based on the probability of their genotypes belonging to each population. Ten independent runs for each value of *K* (1-25) were completed to ensure consistency and the most likely *K* was assessed by comparing the model complexity that maximised marginal likelihood across replicate runs. Isolation by distance (IBD) was assessed using multiple matrix regression with randomisation (MMRR) following Wang (2013). We examined the relationship between matrices of pairwise population F_{ST} calculated in *Genodive* and pairwise population distances along the river network calculated with *ArcMap* v.10.2 (ESRI 2012), and tested for significance using 10,000 random permutations.

Environmental data and interaction among variables

To characterise environmental conditions at each sampling site we used the comprehensive Australian hydrological geospatial fabric (Geoscience Australia 2011; Stein *et al.* 2014), which links spatial data depicting surface water features to a set of environmental attributes describing natural and anthropogenic characteristics of waterways. These include summary statistics on climate, land use, topography and hydrological characteristics organised according to stream hierarchy to allow assessment of environmental factors at multiple scales (i.e. stream vs. catchment level). Also included in the environmental attributes are series of river disturbance indices designed to evaluate the impact of human activities such as disturbance to the flow regime and the effect of land use on the health of freshwater ecosystems (Stein *et al.* 2002).

A subset of 40 candidate variables were selected based on those identified as important predictors of freshwater fish occurrence in south-eastern Australia (Bond *et al.* 2011) along with others that have previously been identified as influencing genetic diversity of freshwater fishes in the MDB (Appendix 3). These variables were divided into five categories concerning variation in temperature, precipitation, flow regime, human disturbance and topography. Within each category variance inflation factor (VIF) analysis was used to exclude highly correlated variables in a stepwise fashion until all remaining variables were below a VIF threshold of ten (Dyer *et al.* 2010). Principal components analyses (PCA) were then performed for the remaining variables in each category. This was carried

out using the *dudi.pca* function in the *ADE4* R package (Dray & Dufour 2007) and principal components (PCs) with eigenvalues greater than one were retained (Yeomans & Golder 1982) as synthetic environmental variables in GEA analyses. The *dimdesc* function in the *FactoMineR* R package (Lê *et al.* 2008) was used to identify individual variables significantly (P<0.05) associated with each PC.

Signatures of selection at local and regional scales

Evidence for local selection was assessed using both univariate and multivariate GEA methods to identify strong associations between allele frequencies and environmental variables. Firstly, we used a spatially explicit generalized linear mixed model approach implemented in *gINLAnd* (Guillot *et al.* 2014). This method generates two competing models for each locus; one in which the fixed effect of an environmental variable influences population allele frequencies, and one where the environmental variable has no effect. Both models account for spatial genetic structure by including a random spatial effect based on geographical coordinates. Due to the dendritic nature of the MDB river system, modelling the spatial arrangement of sites using xy coordinates provides a distorted measure of the true biological distance among sites. To overcome this, the *cmdscale* base function in R was used to perform multi-dimensional scaling (MDS) on the matrix of pairwise river distances. The MDS returned new coordinates that better represent among-site river distances. Using these coordinates, parameters describing the spatial covariance structure of the allele frequency data were estimated using a subset of 500 randomly

selected loci as recommended by Guillot *et al*. (2014). These spatial parameters were then used to control for spatial genetic structure in the final models. *gINLAnd* was run for each of the environmental PCs and log-Bayes factors were calculated for the two models for each locus and used to rank loci in terms of dependence of the allele frequencies on the environmental variable. Using a conservative interpretation of Bayes factors (Kass & Raftery 1995) loci with a log-Bayes factor >15 were considered strong candidates for selection.

Secondly, we used partial redundancy analysis (RDA) to assess the effect of environmental variation on patterns of genomic diversity while also controlling for the effects of spatial genetic structure using parameters describing the spatial distribution of sampling sites. The MDS spatial coordinates were expressed as third-degree polynomials, and subjected to a forward selection procedure based on the method of Meirmans (2015). To account for collinear explanatory variables in the RDA model, VIF analysis was again used, this time to identify environmental PCs strongly correlated with other PCs in the model. Initially RDA was performed with all retained environmental PCs before using a backwards-stepwise selection procedure implemented in *vegan* (Oksanen *et al.* 2015) to select the final model. The final RDA evaluated this reduced environmental model. Significance of the model, as well as marginal significance of each environmental PC were assessed by 1000 ANOVA permutations. The mean locus score across all loci was calculated for each of the first three

RDA axes and individual loci with a score greater than three standard deviations from the mean were considered candidates for selection (Forester *et al.* 2015).

Functional annotation and mode of selection

To examine gene ontology (GO) annotation terms associated with the SNP loci, *Blast2Go* (Conesa *et al.* 2005) was used to perform a BLAST search and annotation of the flanking sequences for all 5,162 SNPs against the NCBI non-redundant nucleotide database with the BLAST *e*-value threshold set to 1×10^{-3} and an annotation threshold *e*-value threshold of 1×10^{-6}. Fisher's exact test was then used with a FDR of 0.05 to assess the GEA candidate loci for enrichment of any biological processes, molecular functions or cellular components compared to the whole SNP data set.

Finally, to better understand the type of selection likely to have generated the genomic footprints detected by either F_{ST} outlier tests or GEA analyses we examined the distribution of F_{ST} values observed for each locus for i) the whole data set, ii) the outlier loci (identified by *Arlequin* and *Bayescan*), and iii) the GEA candidate loci (identified by *gINLAnd* and the RDA). Here we expect a relatively low average and broad distribution of F_{ST} values in the case of polygenic selection because in this scenario adaptation is expected to proceed without major changes in allele frequencies (Pritchard *et al.* 2010). On the other hand, much higher F_{ST}

values are expected for loci involved in 'hard' selective sweeps because alternate alleles in these loci should have approached or reached fixation (i.e. F_{ST} of 1) (Pritchard & Di Rienzo 2010; Messer & Petrov 2013). While this analysis cannot rule out alternative processes also capable of generating a broad F_{ST} distribution (e.g. balancing selection), it serves to highlight the increased power of GEA methods for detecting such a signal. F_{ST} for each locus was calculated in Arlequin and the density distribution kernel for each data set plotted in R.

Results

Genotyping, outlier detection and genome-wide variation

A total of 1,602,903,910 forward and reverse sequence reads were generated with the Illumina platform (detailed sequencing statistics are listed in Appendix 4). After filtering the data with stringent criteria, 5,162 SNPs were retained from 2,589,251 variant sites present in the VCF file produced by *dDocent* (Table 2.2). *BayeScan* identified 643 outlier loci while *Arlequin* identified 697, with 467 of these identified by both. Outliers from both methods were conservatively combined such that the 873 unique loci considered as outliers by either *BayeScan* or *Arlequin* were excluded to create a neutral dataset. After filtering the 4,289 remaining loci for HWE, 3,443 putatively neutral SNPs were retained for population structure and demographic analyses.

Table 2.2: Number of variant sites retained after each filtering step for *Nannoperca australis* from the Murray-Darling Basin (MDB). Detailed descriptions of each filtering step are included in Appendix 2.

Step	SNP count
Raw SNP catalogue	2 589 251
Genotyped in	
50% of individuals, base quality ≥30, minor allele count of 3	243 334
>70% of individuals and >70% of populations	112 557
Bi-allelic only	85 647
Single SNP per locus, MAF >0.05	24 315
Sequencing errors, paralogs, multi-copy loci and artefacts of library preparation	
1) Allele balance	20 828
2) Read orientation	12 878
3) Mapping quality	10 251
4) Paired reads	8876
5) Read quality	6905
6) Read depth	**5162**
Outlier detection	
BayeScan outliers	643
Arlequin outliers	697
Outliers identified in at least one method	873
Putatively neutral	4289
Putatively neutral in HWE	**3443**

Genetic variation based on 3,443 neutral SNPs was low with an average H_E of 0.161 (0.057-0.263), average H_O of 0.123 (0.043-0.206), and an average of 46.3% (19.0-71.7%) polymorphic loci (Table 3). There was however a striking contrast between regions with average H_E of 0.253 in the Lower Murray wetlands compared to 0.143 in the upper reaches (Table 3). Overall, genetic variation for the candidate loci was generally lower but followed a similar pattern to the neutral data (Table 3). Population specific F_{ST} estimated with both neutral and candidate SNPs was generally inversely proportional to genetic diversity, with the most highly differentiated populations also containing the least genetic variation

(Table 2.3). Results from the *Bottleneck* tests for excess heterozygosity confirmed recent reductions in population size at all sites ($P<1x10^{-10}$) except MER ($P=0.193$) and LRT ($P=0.748$) (Appendix 5). Estimates of N_e were generally low (Appendix 5), but varied between lower and upper Murray regions with average estimate of 194.75 (190.9-198.6) for Lower Murray sites significantly higher ($P=0.02$) than the upper reaches average estimate of 88.4 (13.7-305.4).

Table 2.3: Summary by sampling site of expected heterozygosity (H_E), observed heterozygosity (H_O), % polymorphic loci and population specific F_{ST} (Weir and Hill 2002) based on 3 443 neutral and 216 candidate adaptive SNPs for *Nannoperca* from the Murray-Darling Basin (MDB). Lowland wetland sites referred to as Lower Murray in the text are indicated in bold.

		H_E		H_O		% Polymorphic loci		F_{ST}	
Catchment	Site	Neutral	Adaptive	Neutral	Adaptive	Neutral	Adaptive	Neutral	Adaptive
TOO	**TBA**	**0.227**	**0.225**	**0.151**	**0.155**	**58.4**	**58.8**	**0.06**	**0.03**
LMR	**ALE**	**0.263**	**0.269**	**0.161**	**0.153**	**71.7**	**70.8**	**0.07**	**0.03**
	MID	**0.262**	**0.255**	**0.158**	**0.161**	**67.1**	**64.8**	**0.09**	**0.05**
	MUN	**0.260**	**0.271**	**0.167**	**0.178**	**61.6**	**63.0**	**0.03**	**0.02**
ANG	MCM	0.097	0.066	0.090	0.057	26.7	16.7	0.56	0.58
AVO	MIC	0.114	0.080	0.104	0.068	32.8	23.2	0.41	0.47
CAM	JHA	0.091	0.069	0.073	0.055	26.7	19.4	0.36	0.38
GOU	MER	0.075	0.041	0.062	0.035	30.4	19.9	0.47	0.52
	TRA	0.075	0.034	0.059	0.026	23.9	13.0	0.43	0.47
	YEA	0.087	0.049	0.066	0.027	24.4	14.8	0.36	0.43
	PRA	0.243	0.194	0.180	0.149	68.2	56.5	0.18	0.23
	SEV	0.218	0.173	0.170	0.125	59.7	48.6	0.12	0.17
BRO	BEN	0.236	0.203	0.191	0.153	68.5	57.4	0.16	0.20
	SAM	0.234	0.188	0.206	0.161	68.7	58.3	0.19	0.24
	LIM	0.118	0.105	0.094	0.075	39.7	38.0	0.34	0.36
OVE	KIN	0.104	0.091	0.077	0.068	36.2	34.3	0.30	0.29
	HAP	0.114	0.070	0.094	0.063	33.0	25.9	0.37	0.40
	MEA	0.158	0.182	0.129	0.137	43.6	44.4	0.25	0.14
KIE	GAP	0.168	0.102	0.145	0.094	51.8	35.2	0.30	0.39
ALB	ALB	0.226	0.140	0.182	0.106	66.9	43.5	0.30	0.47
MIT	SPR	0.152	0.087	0.119	0.066	48.1	32.9	0.26	0.35
	GLE	0.143	0.074	0.117	0.057	42.8	23.2	0.41	0.51
	TAL	0.164	0.092	0.135	0.068	46.7	27.3	0.48	0.58
COP	COP	0.133	0.100	0.111	0.079	39.9	32.4	0.30	0.35
LAC	LRT	0.057	0.040	0.043	0.031	19.0	16.7	0.67	0.71
Mean	Mean	0.161	0.128	0.123	0.094	46.3	37.6	0.30	0.34
Min	Min	0.057	0.034	0.043	0.026	19.0	13.0	0.03	0.02
Max	Max	0.263	0.271	0.206	0.178	71.7	70.8	0.67	0.71

Population genetic structure

High levels of population genetic structure were evident between most demes of *N. australis,* with population pairwise comparisons of F_{ST} ranging from 0-0.79 (global F_{ST} =0.48). All pairwise F_{ST} estimates were significant (P<0.003) except between Lower Lakes sites MID and MUN (F_{ST}= -0.002, P=0.66) (Appendix 6). Results of the MMRR indicated that river distance was not a good predictor of F_{ST} and that no significant pattern of IBD was apparent across the MDB (regression coefficient=0.108, P=0.342).

Based on F_{ST}, AMOVA calculated across all sites attributed 30.3% of the variation to differences among catchments (P<0.001), 10.7% to differences between sites within catchments (P<0.001), and 13.5% among individuals within sites (P<0.001) (Appendix 7). When calculated separately, the AMOVA for each of the catchments containing multiple sites suggest differences in levels of within-catchment connectivity across the MDB (Appendix 7). Sites within the Lower Murray appear to be highly connected, suggesting that TBA, ALE, MID and MUN constitute a single population. This is in contrast to less connected upper Murray catchments (Appendix 7).

Clustering analysis in *fastStructure* based on neutral SNPs identified 12 distinct populations that mostly correspond with the MDB catchment boundaries (Figure 2.2), except for the following. In the Lower Murray,

TBA (Tookayerta catchment) grouped together with ALE, MID and MUN (Lower Lakes). Goulburn River (five sites) and Broken River (three sites) catchments were split into three groups; an upper Goulburn cluster (MER/TRA/YEA), a distinct Broken River site (LIM) and an admixed group consisting of two lower Goulburn and two Broken River sites (PRA/SEV/BEN/SAM). The site at Albury is most similar to the Kiewa River site (GAP), however also shares some affinity with sites further upstream from the Mitta Mitta catchment.

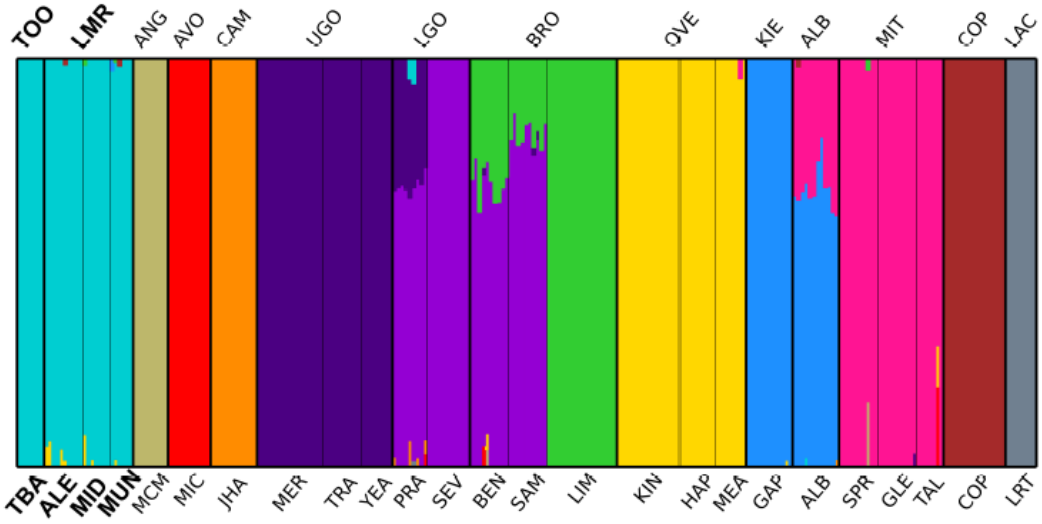


Figure 2.2: Admixture plot based on 3,443 'neutral' SNPs for *Nannoperca australis* from the MDB depicting K=13 clusters determined by maximum marginal likelihood using *fastStructure*. Codes above and below the plot refer to catchment and sampling site respectively (Table 2.1). Lowland wetland sites referred to as Lower Murray in the text are indicated in bold.

Environmental data and interactions among variables

Calculating pairwise distance among sampling sites with the revised MDS coordinates improved the correlation with along-river distances (R^2=0.97) compared to the original spatial coordinates (R^2=0.87) (Appendix 8). The

revised MDS coordinates were then substituted for downstream analyses requiring spatial coordinates.

Following VIF analyses, 19 variables representing all five environmental categories were retained for the environmental PCAs (see Appendix 3 for explanation of variable codes). These are four temperature variables (STRCOLDMTHMIN, STRDRYQTEMP, CATDRYQTEMP and STRWETQTEM), three precipitation variables (CATDRYQRAIN, STRWETQRAIN and CATCOLDQRAIN), two disturbance factors (CDI and FRDI), five flow variables (RUNCVMAXMTH, RUNPERENIALITY, RUNANNMEAN, SUBEROSIVITY and CATEROSIVITY), and five topographic variables (STRAHLER, SUBELEMAX, CATELEMEAN, VALLEYSLOPE and CATSLOPE).

The first two components of each PCA for the temperature, flow and topographic variables scored eigenvalues >1 and explained 75.3%, 83.1% and 83.3% of the total variance respectively. Just one component each for the precipitation and human disturbance PCAs scored an eigenvalue >1 and thus all individual variables rather than PCs for these categories were used for downstream analyses. The PCA plots for temperature and precipitation depict the climatic gradient across the MDB (Figure 2.3), with sites from the Lower Murray experiencing higher winter temperatures and lower rainfall than headwater sites such as those in the Ovens River and Mitta Mitta catchments.

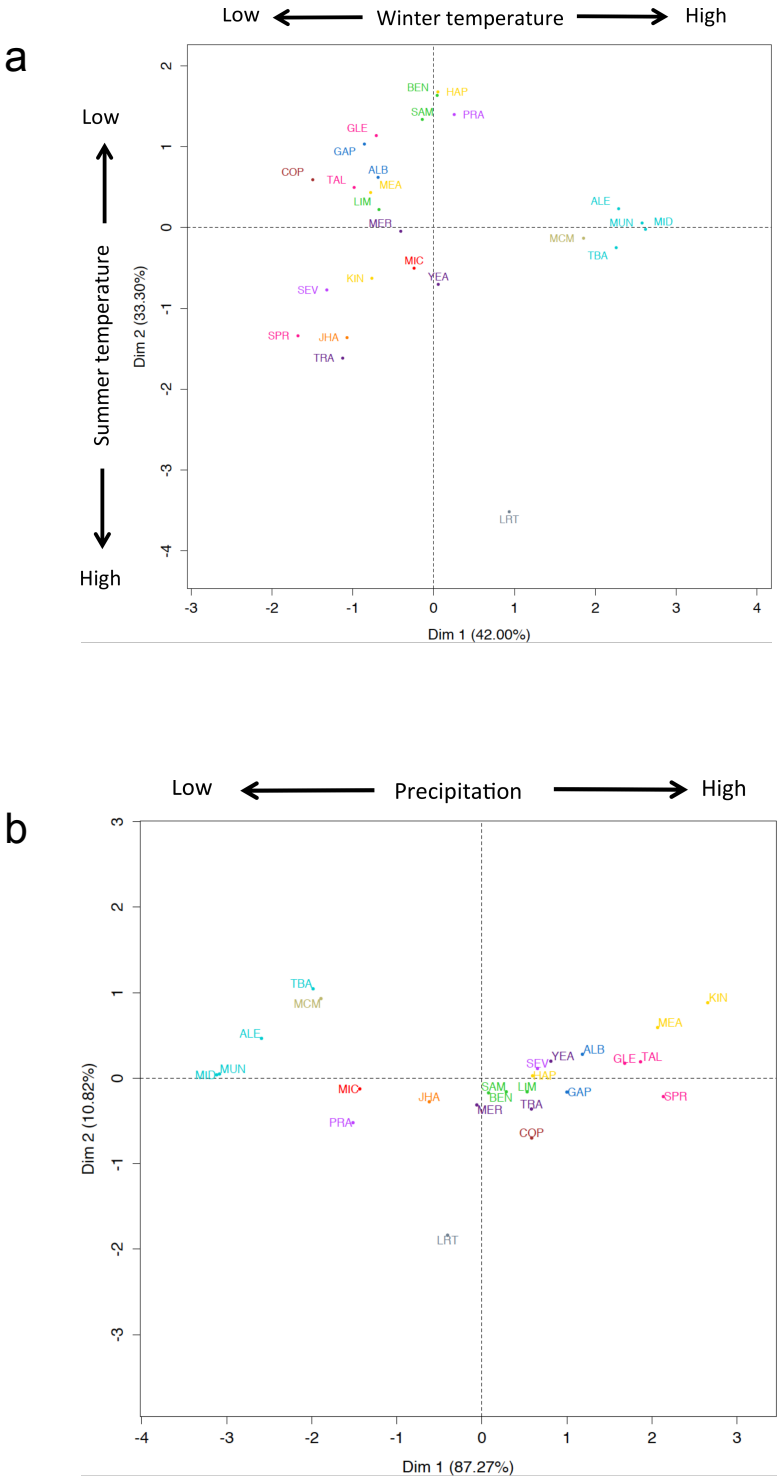


Figure 2.3: Environmental PCAs of the Murray-Darling Basin (MDB) describing the relationship between each *Nannoperca australis* sampling location based on variables related to a) temperature and b) precipitation. Site names are colour coded based on the colours used in Figure 2.1. Annotations above and to the left of plots describe which variables contribute the most to each axis. Environmental PCAs based on variables related to flow, human disturbance and topography are in Appendix 9.

Measures of flow variation (erosivity and pereniality) dominate the first PC of the flow PCA and demonstrate that Lower Murray sites experience lower variation in flow than headwater locations (Appendix 9). The second flow PC explains differences in average annual flow, which tend to reflect accumulated increases in total flow from headwaters to the main channel rather than variation in climate (Appendix 9). The topographic PCA was most influenced by elevation and Strahler stream order, which describes intrinsic physical factors related to each site's position in the river network (Appendix 9). In contrast to the other categories there was no evidence of any regional spatial pattern for the human disturbance PCA confirming that human disturbance mostly affects conditions at a local scale (Appendix 9). This resulted in a final list of 11 environmental variables (six PCs and five individual precipitation and disturbance variables) that describe variation in the environment across the MDB, and the correlations between individual variables and the PCs are described in Appendix 10.

Signatures of selection at local and regional scales

gINLAnd provided strong evidence (log-Bayes factor >15) for associations between allele frequencies of 178 unique loci and the 11 environmental variables identified based on PCAs, as above (Appendix 11). Candidate loci were identified by *gINLAnd* for all environmental variables, with precipitation related variables associated with 85 loci, temperature (53 loci), flow (39 loci) and topography (26 loci). Variables associated with high numbers of loci included CATCOLDQRAIN (74), Temp2 (41), Flow1

(35) and CDI (35) (Figure 2.4). As expected, there was also a high degree of overlap with 95 loci identified as candidates associated with more than one variable. Human disturbance variables describe mostly local scale disturbance and were associated with 41 loci, of which 22 were not associated with any other variables.

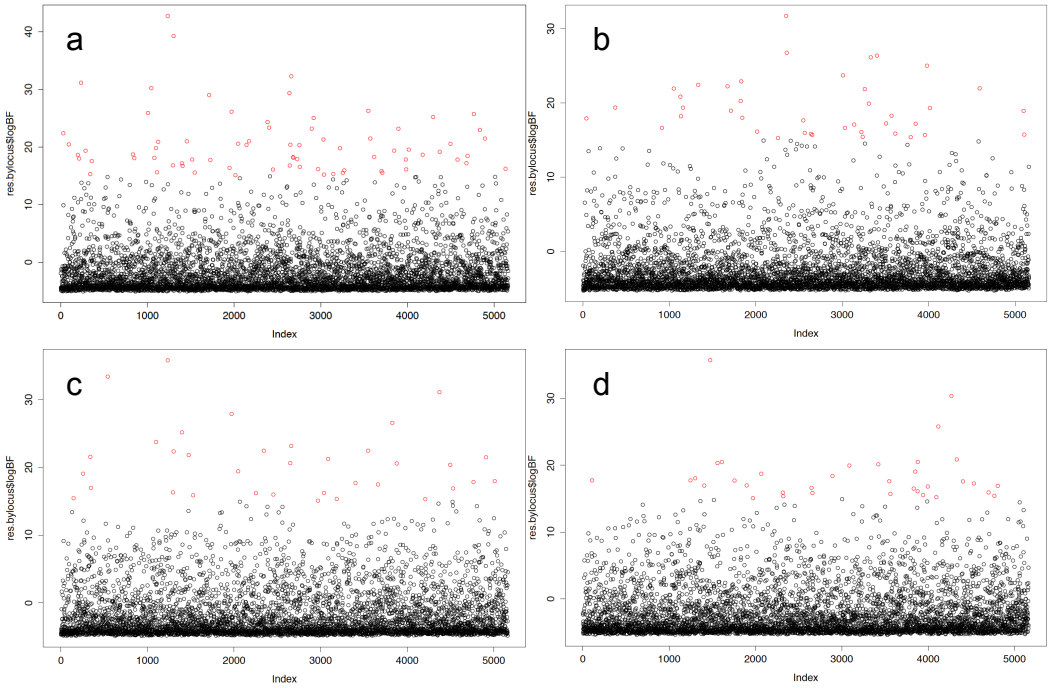


Figure 2.4: log Bayes factor scores for each of 5,162 SNPs for their association with environmental variables. a) Catchment average coldest quarter rainfall (CATCOLDQRAIN), b) average summer temperature (Temp2 PC), c) stream flow variation (Flow1 PC) and d) catchment disturbance index (CDI). Loci with a log Bayes factor >15 are highlighted in red and were considered as candidates for selection.

The RDA triplot summarises the first two axes of the final model and indicates temperature, rainfall and topography are the major environmental factors influencing genetic variation of 42 candidate adaptive loci (Figure 2.5). Winter rainfall (CATCOLDQRAIN) and summer temperature (Temp2) were the most influential predictive variables in the

model suggesting that the climate is the major factor driving selection across the region. Following VIF analyses Temp1, Temp2, CATCOLDQRAIN, Topo1 and Topo2 were retained as predictive variables in the final model, along with two spatial conditional variables. The RDA was globally significant (P=0.007) with environmental variation explaining 23.83% of the total genetic variation after accounting for spatial structure (30.07% of total genetic variation). Assessment of the marginal significance of each explanatory variable revealed that Temp2, CATCOLDQRAIN, Topo1, and Topo2 were each significant predictors of allele frequencies (P<0.05). The first three RDA axes explained 85.41% of the variation (33.89%, 28.88% and 22.65% respectively) and individual locus scores for 42 loci (9, 17 and 16 for each of the three RDA axes) were further than three standard deviations from the mean (Appendix 12) and were considered as candidate loci potentially under selection. Triplots including RDA3 are presented in Appendix 13.

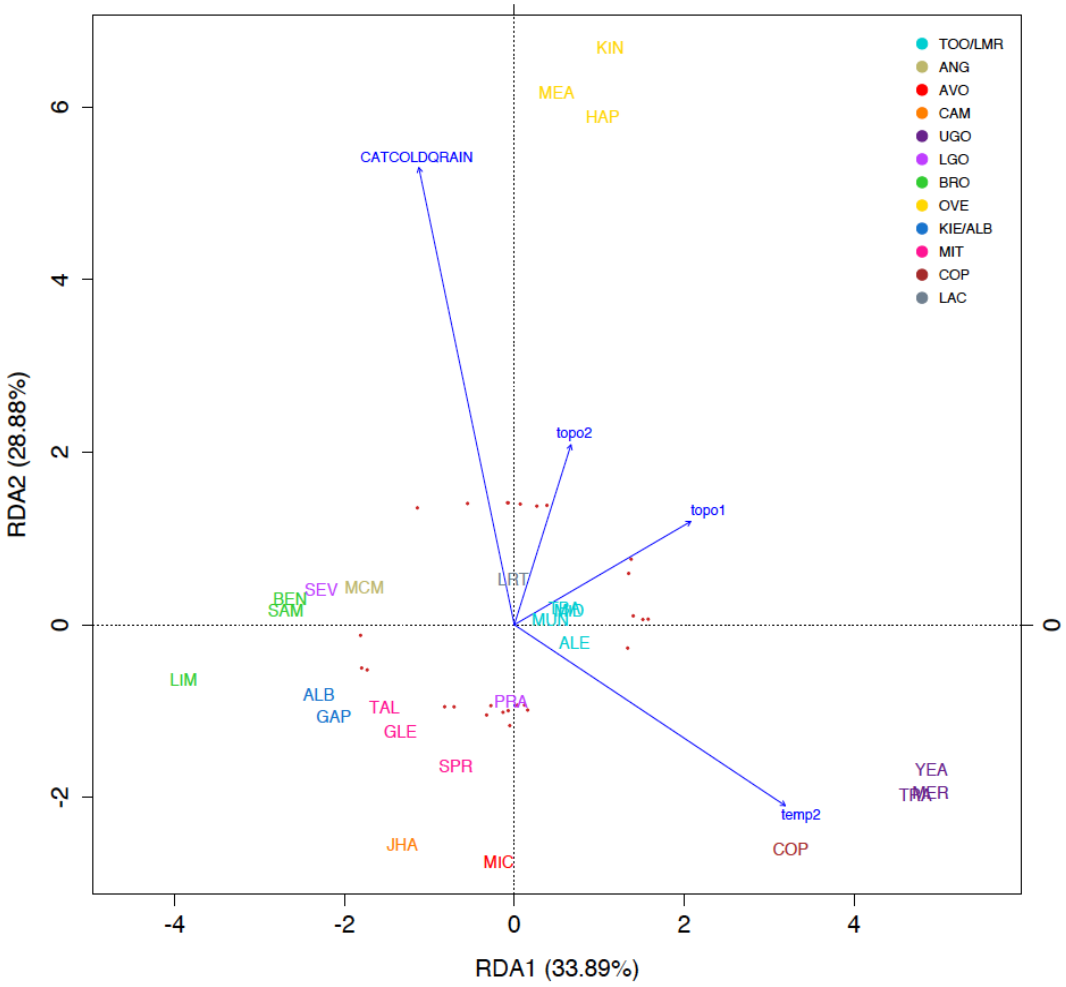


Figure 2.5: Triplot summarising the first two axes of the partial redundancy analysis (RDA). Sampling sites are colour coded according to Figure 2.1 and depict each sites position in the environmental model. Significant environmental factors (P<0.05) are represented as blue vectors where the direction of the arrowhead indicates high values (e.g. site KIN receives the high rainfall while site MIC receives low rainfall). The length of each vector represents the magnitude of their contribution to the model and the angle between each vector represents the correlation among variables. Allele frequency vectors for individual SNPs significantly associated with the model have been re-scaled to the same ordination space and are indicated by red markers. Their position depicts the direction of allele frequency variation in relation to the environmental model. Plots including RDA3 are in Appendix 13.

Functional annotation and mode of selection

The GEA analyses together identified 216 unique candidate adaptive loci (178 *gINLAnd*, 42 RDA with four loci identified by both methods). *Blast2Go* recorded blast hits for 1289 (*e*-value<1 x 10^{-3}) of the 5,162 loci (Table S8) of which 638 could be annotated and were assigned 885 GO terms (*e*-value<1 x 10^{-6}) (Table S9), and 49 blast hits for the 216 candidate loci (Table S10) of which 24 were annotated and were assigned 138 GO terms (Table S11). Enrichment analysis did not find any GO terms significantly (FDR of 0.05) under- or over-represented in the candidate adaptive data set compared to all loci. Tables S8–S11 are available on Dryad: DRYAD entry doi:10.5061/dryad.3dp50.

The distribution of F_{ST} values for the entire SNP dataset is broad and extends across the entire theoretical range (i.e. from zero to one). This contrasts with the narrow distribution for the F_{ST} outlier loci, which include several values close to one. On the other hand, the distribution of F_{ST} values for the GEA candidate loci is not only broader than the F_{ST} outliers, but also peaks at much lower F_{ST} values (Figure 2.6). The average single locus F_{ST} for all 5,162 loci was 0.461 (0-1), compared to 0.826 (0.695-1) for the 177 F_{ST} outliers and to 0.634 (0.356-1) for the 216 GEA candidate loci.

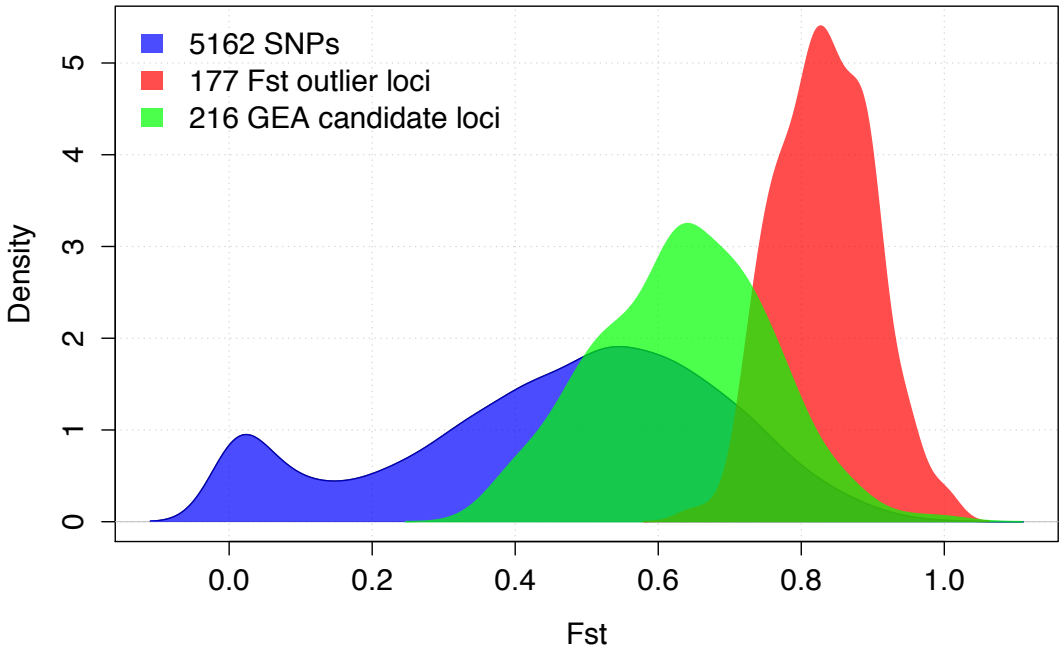


Figure 2.6: Density distribution of F_{ST} values for all 5,162 SNPs (blue), 177 F_{ST} outlier loci (red) and 216 candidate loci identified using genotype-environment association analyses (green).

Discussion

The rapid rate of climate change and other anthropogenic threats suggests that evolutionary adaptation will be required for many species to persist into the future (Stockwell *et al.* 2003; Mergeay & Santamaria 2012; Losos *et al.* 2013). However, in order to gauge the potential of species to adapt to environmental change we need to first understand how the environment shapes intraspecific variation across the genome. Here, replicate populations of a threatened and poor dispersive Australian freshwater fish sampled across a steep hydroclimatic gradient were examined using 5,162 high quality SNPs and compared with high-resolution environmental data in a riverscape genomics framework. Overall, strong population structure associated with the hierarchical river

network and low genetic variation was identified with putatively neutral SNPs. This is consistent with findings from studies based on other classes of selectively neutral markers (Attard *et al.* 2016b; Cole *et al.* 2016), confirming that drift is a major evolutionary process shaping genetic diversity in this system.

On the other hand, evidence is also provided for a marked pattern of hydroclimatic driven-selection, with temperature and precipitation emerging as the most important of several environmental parameters influencing the allele frequencies of 216 candidate adaptive loci, both at regional (basin-wide) and local (catchment) scales. Human disturbance also influenced putatively adaptive variation, but for a smaller number of candidate loci and only at a local scale. In addition, despite strong drift and geographic isolation, adaptive divergence among populations appears to be explained by a pattern of non-allelic fixation consistent with a genomic footprint of polygenic adaptation. This represents the first riverscape genomics study of an Australian fish and, as such, makes an original contribution to our understanding of adaptation across large freshwater ecosystems – a topic dominated by studies of Northern Hemisphere fishes, in particular salmonids (e.g. Hecht *et al.* 2015) and sticklebacks (e.g. Raeymaekers *et al.* 2014; Ferchaud & Hansen 2016). More broadly, our results highlight the utility of spatially explicit GEA methods for elucidating the signal of selection in spatially complex and

anthropogenically-modified ecosystems and for informing conservation management of endangered biodiversity.

Boom-bust cycles and dendritic landscapes influence genome-wide variation

Understanding the relationship between landscape heterogeneity, environmental variability and population genetic diversity in river basins is an important topic in ecology and evolution because these are among the most diverse and yet most threatened ecosystems (Palmer *et al.* 2008; Strayer & Dudgeon 2010). Levels of genome-wide variation in *N. australis* (Table 2.3) are lower than those reported in other population genomic studies of freshwater fishes (Matala *et al.* 2014; Skovrind *et al.* 2016). This is unsurprising and consistent with low levels of microsatellite DNA variation reported for *N. australis* (Cook *et al.* 2007; Attard *et al.* 2016b; Cole *et al.* 2016), and more broadly for other MDB fishes (Faulks *et al.* 2010b; Faulks *et al.* 2011; Brauer *et al.* 2013; Coleman *et al.* 2013; Sasaki *et al.* 2016), which exhibit generally very low genetic variation compared to non-Australian freshwater fishes (DeWoody & Avise 2000). This emerging paradigm is likely due to the naturally variable hydroclimatic environment of the MDB and of several other Australian river systems (Kennard *et al.* 2010) that result in frequent cycles of population booms and busts. These cycles cause fluctuations in population size that produce bottlenecks and affect spatial patterns of gene flow in Australian freshwater fishes (Huey *et al.* 2008; Faulks *et al.* 2010b). Natural boom-bust cycles, likely in combination with recent and localized human-disturbance (e.g. Attard *et*

al. 2016b), have no doubt contributed to patterns observed here, as evident in the signal of genetic bottlenecks and small N_e inferred for populations across the basin.

Despite this general pattern, genome-wide variation in *N. australis* was markedly reduced in the upper compared to the lower reaches of the MDB. The Lower Murray is composed of a large system of linked wetlands and lakes, whereas the upper reaches of the MDB consist of small, often disconnected rivers and creeks (e.g. Hammer *et al.* 2013). As expected based on landscape configuration, N_e estimates obtained for *N. australis* from the lower MDB were significantly larger (P=0.02) compared to estimates for the smaller upper MDB waterways. In addition, historical demographic analyses indicate that Lower Murray *N. australis* maintained relatively stable N_e until before European settlement (Attard *et al.* 2016b), followed by very recent bottlenecks and near local extirpation (Hammer *et al.* 2013; Attard *et al.* 2016b). Theoretical (Morrissey & de Kerckhove 2009; Paz-Vinas & Blanchet 2015; Thomaz *et al.* 2016) and empirical studies (Crispo *et al.* 2006; Barson *et al.* 2009; Osborne *et al.* 2014) of the effects of landscape structure on genetic variation of fishes suggest that not only intrinsic physical landscape properties but also asymmetrical downstream migration generate higher variation downstream compared to headwater populations. For *N. australis*, we detected strong and hierarchical population structure (i.e. differentiation was much greater among than within catchments) and no migration between most

catchments (Figure 2), consistent with nil contemporary microsatellite-based gene flow observed in a larger sample (*n*=578) (Cole *et al.* 2016). Nonetheless, a more contiguous meta-population occupying and dispersing along the Murray River corridor prior to European Settlement, enhanced in wetter periods over evolutionary time scales (Unmack *et al.* 2013; Cole *et al.* 2016), has also probably contributed to higher genome-wide variation and lower population specific F_{ST} observed in the Lower Murray (i.e. the latter is, on average, the most similar to all other populations in the basin; Table 2.3).

Detecting the signal of selection across a large and heterogeneous river basin

Environmental variability and instability likely exacerbate the effects of drift for *N. australis*, yet average hydroclimatic conditions vary substantially among catchments across the MDB. In this case, natural selection is also expected to contribute to population divergence, especially when gene flow among populations is restricted (Willi *et al.* 2007; Blanquart *et al.* 2012; Harrisson *et al.* 2014). Detecting the signal of selection in complex river networks however is particularly challenging and inferences can be misleading if based on approaches using inappropriate null models (Fourcade *et al.* 2013; Thomaz *et al.* 2016). Through the use of spatially explicit GEA methods, we aimed to disentangle the signal of adaptive variation responding to the environment from the strong spatial pattern of neutral genetic variation. The RDA confirmed that spatial population structure was responsible for patterns of genome-wide diversity (30.07%

of total genetic variation), however temperature, precipitation and topography were also important factors accounting for a large amount of the residual variation (23.83% of the total). This is reinforced by the *gINLAnd* results, where precipitation and temperature variables were associated with the majority of candidate adaptive loci (106 and 58, respectively, of 178). Our suggestion of adaptive population divergence is further strengthened by the fact that the loci identified by the GEA methods are responding in parallel to the environment across a number of demographically independent populations. This also builds on studies showing that local adaptation of traits related to reproductive fitness in *N. australis* vary predictably along gradients of variability in water flow (Morrongiello *et al.* 2010; Morrongiello *et al.* 2012; Morrongiello *et al.* 2013) and supports the hypothesis that hydroclimatic selection has driven adaptive genetic differentiation of populations. These results also add to a growing body of evidence that climate is a major factor contributing to adaptive divergence among freshwater fish populations. For example, Bourret *et al*. (2013) found that climate and geology were associated with adaptive divergence of Atlantic salmon (*Salmo salar*) populations, and Hecht *et al*. (2015) identified precipitation and temperature as significant factors shaping adaptive variation of Chinook salmon (*Oncorhynchus tshawytscha*).

Powerful GEA methods have also recently shown promise in detecting polygenic adaptation in natural populations (Lasky *et al.* 2012; Bourret *et*

al. 2014; Hecht *et al.* 2015). Empirical and modelling studies suggest that local adaptation to environmental change may predominantly arise through a process of polygenic selection (Hermisson & Pennings 2005; Pritchard & Di Rienzo 2010). This mechanism involves relatively small changes in allele frequencies at a large number of loci underlying the trait under selection. On the other hand, genome scans based on F_{ST} outlier tests are primarily designed to identify 'hard' selective sweeps that lead to fixation or near fixation of alternate alleles (Messer & Petrov 2013). For our dataset, the distribution of F_{ST} values for the vast majority of GEA loci is inconsistent with alternate fixation of candidate adaptive alleles (Figure 2.6; mean F_{ST} of 0.634;) and instead supports recent views that adaptation of complex quantitative traits probably takes place by simultaneous selection acting on variants at many loci of small effects (Pritchard *et al.* 2010). Rapid adaptation to environmental change due to polygenic selection is possible if sufficient standing genetic variation exists in the population (Pritchard *et al.* 2010; Crisci *et al.* 2016), underscoring the potential benefits of incorporating GEA methods into conservation studies of adaptation (Le Corre & Kremer 2012).

Are small fragmented populations subjected to more divergent selection?

As the field of landscape genomics is evolving, so to is the idea that fragmentation not only reduces habitat size and quality, but also increases environmental variation within, and among habitat fragments (Wood *et al.* 2014). An emerging paradigm challenges the classical view of population

genetic theory and predicts that natural selection can promote adaptive diversity, even in small populations where strong drift is expected to constrain adaptive evolution (Koskinen *et al.* 2002; Fraser *et al.* 2014; Wood *et al.* 2016). Our results support this idea and the hypothesis that, despite reduced genetic diversity due to drift in the small and fragmented populations typically found in catchments from the upper MDB, heterogeneous selection pressure is also driving local adaptive divergence in response to the increased environmental variation resulting from decreasing local habitat size. The environmental analysis shows that many upper catchments appear to harbour unique and divergent habitats, especially in regard to precipitation (Figure 2.3) and to flow and human disturbance (Appendix 9). This is in part supported by the RDA results, where several upper MDB catchment populations showed the most divergent GEA profiles across the basin (e.g. Ovens and upper Goulburn rivers). Accordingly, fragmentation and habitat quality are known to impose divergent selection that alters microgeographic adaptation in isolated populations (Willi & Hoffmann 2012). An alternative, but not mutually exclusive, view is that habitat complexity, rather than size only, might have impacted on evolutionary persistence of these small fragmented populations. Here, fragmentation would lead to random subsampling of habitat types, with population persistence at small N_e dependent on the quality of the habitat sampled in the resulting occupied fragment (Fraser *et al.* 2014). For instance, adaptive differentiation in fragmented brook trout (*Salvelinus fontinalis*) populations was greater among small than among large populations, with very small populations

still very much affected by natural selection (Fraser *et al.* 2014). Whether this could lead to more variable evolutionary responses (and even perhaps enhance persistence) in fragmented *N. australis* is a hypothesis that remains to be assessed.

Regardless of the mechanisms, the small and fragmented upper MDB populations have comparatively little standing variation at hydroclimatically selected loci compared to the significantly larger Lower Murray populations (Table 2.3). This highlights that these populations may have reduced adaptive potential to respond to rapid climate change. Conversely, in the Lower Murray, the combination of higher diversity and low population specific F_{ST} at candidate loci (i.e. low adaptive divergence (Funk *et al.* 2012)) indicates these populations may be a reservoir of adaptive variation for the species and reinforces the critical nature of ongoing conservation efforts in this region (e.g. Bice *et al.* 2012; Hammer *et al.* 2013; Attard *et al.* 2016b). The adaptive sink hypothesis is also consistent with recent findings of unexpectedly high levels of variation and strong positive selection at the MHC IIB gene of Lower Murray *N. australis* (Bracamonte *et al.* 2015).

Implications and recommendations for conservation

Despite ongoing conservation efforts including habitat restoration, environmental water allocation, captive breeding and reintroductions (Bice *et al.* 2012; Hammer *et al.* 2013; Pearce 2015; Attard *et al.* 2016b), *N.*

australis remains endangered or threatened across the MDB. In fact, during sampling for this study we observed at least 10 populations that are now extirpated due to loss of habitat associated with river regulation and drought. To promote long-term persistence of remnant populations, conservation efforts need to be proactive and should focus on maintaining natural habitat and restoring evolutionary processes to avoid further loss of genetic diversity and to increase resilience to environmental change (Crook *et al.* 2010; Morrongiello *et al.* 2011a; Hammer *et al.* 2013). In this sense, carefully considered translocations provide an attractive option for conservation management of small and fragmented populations (Sgrò *et al.* 2011; Weeks *et al.* 2011; Frankham *et al.* 2014; Frankham 2015). Genetic rescue (Tallmon *et al.* 2004) (here we also include the closely related concept of genetic restoration proposed by Hedrick (2005)), can occur when translocations are used to restore gene flow between recently isolated populations. This can thereby reduce the genetic consequences associated with small population sizes such as inbreeding depression, reduced genetic variation and genetic load (Weeks *et al.* 2011; Whiteley *et al.* 2015). For *N. australis*, translocations among populations within catchments would replicate natural evolutionary and demographic processes by restoring connectivity among recently isolated but historically connected demes. This could be achieved by translocating several individuals to provide ~20% gene flow initially, followed by a small number of migrants per generation thereafter (Hedrick 1995; Lopez *et al.* 2009). It has been suggested that outbreeding depression could lead to reduced fitness in target populations (Edmands 2007), however Frankham

(2015) argues that the risk of outbreeding depression has likely been overemphasized in the literature (see also Weeks *et al.* 2016). Populations of *N. australis* have clearly been historically connected at the catchment level and given observed historic and ongoing declines we argue that their risk of extirpation due to inbreeding depression, loss of genetic diversity and stochastic demographic events outweighs risks posed by outbreeding depression.

Where species inhabit a wide range of environments, the potential also exists to select source populations based on information from LG and predictions of future environmental conditions to build evolutionary resilience to future environmental change (Aitken & Whitlock 2013). In addition to genetic rescue, our findings provide the opportunity to also consider a second strategy, and use translocations to introduce new alleles that may increase the potential for populations to adapt to environmental change (Sgrò *et al.* 2011; Aitken & Whitlock 2013). Populations of *N. australis* in the Lower Murray experience hotter and drier conditions than elsewhere in the MDB (Figure 2.3) and our results suggest that these populations are locally adapted. Increasing aridity and climate variability are predicted for the whole MDB in the future (Kershaw *et al.* 2003; Morrongiello *et al.* 2011a; Davis *et al.* 2015) and we propose that translocations within catchments could be additionally supplemented with a small number of individuals harbouring adaptive genetic variation from other populations. In this case, the Lower Murray would provide an ideal

source population due to their higher genetic diversity and low average neutral, and potentially adaptive divergence.

Conclusions

Understanding the evolutionary potential of populations to respond to rapid climate change demands knowledge of how environmental factors contribute to local adaptation of populations. The recent transition from landscape genetics to landscape genomics has already yielded strong evidence for the role of climate in shaping patterns of intraspecific genetic variation. Inferring selection in complex spatial environments however remains challenging. Our riverscape genomic approach used spatially explicit GEA methods to control for the effects of landscape structure and shared population history. It showed that hydroclimatic conditions influence the population genetic architecture of *N. australis* in the MDB. We revealed precipitation and temperature as the most important of several environmental parameters influencing adaptive genetic variation, both at local and regional scales. Human disturbance also influenced putatively adaptive variation, but only at a local scale. The 216 candidate loci we identified provide a basis for further work exploring the functional significance of genomic regions involved in local adaptation to hydroclimatic heterogeneity. Recently, there has been a call for genomic approaches currently used to address questions in ecology and evolution to move beyond the realm of academic research and contribute more to solving the practical issues of conservation biology (Shafer *et al.* 2015a).

Our work is an initial step towards that goal and will hopefully inspire further debate and research into how knowledge of adaptive genetic variation may best be incorporated into species conservation.

Chapter 3: Anthropogenic habitat fragmentation increases extinction risk for freshwater species

Abstract

Habitat fragmentation is a key threatening process implicated in the current global extinction crisis. Freshwater ecosystems are amongst the most highly impacted by fragmentation, and the decline of freshwater biodiversity has been far greater than for either terrestrial or marine biomes. The genetic signals of natural historical, and contemporary anthropogenic processes can however be difficult to separate, especially in complex spatial environments that have only recently been impacted by fragmentation. Here, we use high-resolution genomic data to examine the demographic and genetic effects of very recent habitat fragmentation on a threatened freshwater fish across an entire river basin. We employ a combination of population and landscape genetic methods to show that the construction of thousands of in-stream barriers over the last ~160 years has contributed to the isolation of populations of southern pygmy perch, *Nannoperca australis* in the Murray–Darling Basin. Furthermore, populations most isolated by fragmentation have reduced genetic diversity. Individual-based simulations and evidence of bottlenecks in most populations provide additional support that recent anthropogenic habitat fragmentation is increasing the extinction risk of freshwater biodiversity.

Introduction

There can be little doubt that we are now on the brink of the sixth global mass extinction with the current rate of species extinctions far exceeding pre-anthropogenic estimates (Barnosky *et al.* 2011). As the human population increases in size, our veracious appetite for the Earth's natural resources is intensifying pressure on the natural environment (Foley *et al.* 2005; Davis *et al.* 2015). Habitat fragmentation is a frequent side effect of this proliferation of human development and is a key factor leading to the genetic and demographic decline of populations that together increase the risk of local population extirpation, and eventually species extinctions (Hanski 1998; Lande 1998). Roads, fences, dams and other infrastructure associated with the conversion of natural habitat for human use fragment previously continuous habitat into small, and increasingly isolated patches (Fischer & Lindenmayer 2007). This can lead to increased inbreeding and genetic drift within populations and reduced gene flow among populations which, in turn, can result in reduced fitness due to inbreeding depression and lowered evolutionary potential due to loss of quantitative genetic diversity (Frankham 2005; Keyghobadi 2007). Additionally, small populations become vulnerable to extirpation due to stochastic demographic events (Lande 1993) and if this occurs on a regional scale then species extinctions are an inevitable outcome (Hanski 1998). It is thus crucial, if we are to curtail the current rate of biodiversity loss, that we improve our understanding of the relationship between anthropogenic habitat fragmentation, population isolation and genetic diversity in threatened populations.

Landscape genetics has contributed much to our understanding of how features of the landscape influence patterns of population structure and connectivity. Isolation by barrier (IBB) is one of the most fundamental landscape genetics models and assumes that physical landscape structures prevent or restrict dispersal and gene flow among populations (Manel *et al.* 2003; Storfer *et al.* 2007). In the context of conservation we are most interested in examining the effects of habitat fragmentation driven by human-mediated landscape change, with several studies providing evidence that anthropogenic barriers pose a serious threat to population persistence (Meldgaard *et al.* 2003; Keller *et al.* 2004; Epps *et al.* 2005; McCraney *et al.* 2010; Gouskov *et al.* 2016). The time lag between the construction of a barrier and any resulting detectable genetic signal can however make it difficult to disentangle historical and contemporary processes, and the choice of molecular marker may exacerbate this issue (Landguth *et al.* 2010; Epps & Keyghobadi 2015). Earlier studies investigating the consequences of recent habitat fragmentation relied on low-resolution markers, such as mtDNA or allozymes, and often failed to detect any genetic signal (Cunningham & Moritz 1998; Driscoll & Hardy 2005). Similarly, life history traits, such as high dispersal or long generation times, may also affect our ability to detect the effects of recent fragmentation (Whiterod *et al.* 2016). However, the ongoing transition from landscape genetics (i.e. based on tens of DNA markers) to landscape genomics (i.e. based on hundreds to thousands of DNA markers) has increased both the spatial and temporal resolutions at which demographic processes can be examined, providing a more

powerful framework with which to quantify the effects of very recent landscape change on genetic diversity and population structure (Allendorf *et al.* 2010; Wang 2010).

Freshwater ecosystems have been highly impacted by human development (Dudgeon *et al.* 2006) and the aquatic habitat sustained by 65% of global freshwater runoff is considered either moderately or highly threatened (Vorosmarty *et al.* 2010). The effects of habitat fragmentation can be more severe in freshwater than in terrestrial or marine environments as the physical structure of dendritic river systems naturally fragments populations and reduces the likelihood of re-colonisation following local extirpations (Fagan 2002). Of particular concern are freshwater fishes, a group recognized as highly susceptible to decline following disturbance and that shows the highest extinction rate among vertebrates (Burkhead 2012). Furthermore, a disproportionately high level of biodiversity inhabits freshwater ecosystems with a high level of endemism, and as a consequence the decline of freshwater biodiversity has been far greater than for either terrestrial or marine ecosystems (Sala *et al.* 2000; Strayer & Dudgeon 2010). Over the last century close to one million dams have been constructed globally and more than 50% of freshwater runoff is now diverted for human use (Jackson *et al.* 2001; Nilsson *et al.* 2005). These barriers not only prevent longitudinal connectivity along river networks important for migratory species, but also restrict lateral movements between stream and ecologically important

floodplain habitats (Liermann *et al.* 2012). The homogenization of natural seasonal flow variability can also have devastating consequences for spawning and recruitment of many species and results in improved conditions for ecological generalist, and often invasive species at the expense of specialist endemic species (Poff *et al.* 2007). Previous work demonstrated an effect of in-stream barriers isolating fish populations, however these studies have mainly assessed large migratory species of high economic value and concerned only major barriers situated on main river channels (Faulks *et al.* 2011; Horreo *et al.* 2011; Torterotot *et al.* 2014; Gouskov *et al.* 2016). Research to date has largely ignored the cumulative effects of multiple in-stream barriers on the many other ecologically important aquatic species likely to be affected by fragmentation. As a result these species often receive little attention from conservation managers (Olden *et al.* 2007; Saddlier *et al.* 2013).

The Murray-Darling Basin (MDB) is the largest river network in Australia and one of the country's most impacted ecosystems (Laurance *et al.* 2011). The MDB has been heavily modified with ~4000 dams, weirs and levies constructed since 1857 when European settlement of this region began (Arthington & Pusey 2003). This has contributed to the dramatic decline of native fishes in recent times with more than half of all MDB species now listed as threatened or of conservation concern (Lintermans 2007). Considering that it can take decades following fragmentation for the demographic decline of populations to progress to the point of extinction

(Matthews & Marsh-Matthews 2007), it is likely that many remnant MDB populations are already well down that path, as recently observed in the lower MDB (Hammer *et al.* 2013; Attard *et al.* 2016a) This highlights the urgency with which we need to identify and conserve populations or species most at risk. In this chapter I present a case study of the effects of habitat fragmentation for the southern pygmy perch (*Nannoperca australis*), a threatened native fish inhabiting small streams and wetlands and typical of many small-bodied fishes in the MDB. Landscape and environmental factors, including human disturbance are already known to have impacted genetic diversity and connectivity for this species in the MDB (Brauer *et al.* 2016). However little is known about the population-level effects of recent and widespread habitat fragmentation. The low-lying topography of the MDB presents few natural barriers in this system (McLaren *et al.* 2011) so we know that fragmentation has occurred only very recently. The latter is consistent with microsatellite-based coalescent estimates that suggest a recent isolation (~125 years) between *N. australis* populations from the upper and lower reaches of MDB (Cole *et al.* 2016).

A key question that should be addressed in studies assessing the consequences of anthropogenic habitat fragmentation is whether or not patterns of population differentiation and isolation evident in the contemporary genetic architecture of the species predate the onset of human impacts in the ecosystem. Here, the MDB provides a unique

opportunity to examine the consequences of recent habitat fragmentation without the confounding influence of more prolonged human disturbance common to many northern hemisphere river basins (e.g. Hansen *et al.* 2014). In addition, small poorly dispersive species such as *N. australis* offer a conservative benchmark for guiding broader conservation strategies in aquatic ecosystems, as they provide insights into the impacts of habitat fragmentation that are less evident in more dispersive species. It is however important to use methods that can account for the strong signal of historical population structure related to natural stream hierarchy expected in these systems. Using a large and highly resolving neutral SNP data set we test several hypotheses related to how the recent and extensive construction of in-stream barriers across the MDB has contributed to the loss of genetic diversity and isolation of populations, and how it threatens the long-term persistence of *N. australis*. Firstly we test the hypothesis that genetic differentiation among demes increases with the number of in-stream barriers separating them. Secondly, that populations most isolated by fragmentation will exhibit reduced levels of genetic diversity. Thirdly, that the effective population size (N_e) of demes affected by fragmentation has recently declined. Additionally, we use individual-based population genetic simulations to investigate whether the level of differentiation observed for *N. australis* could have arisen in the short time since the construction of in-stream barriers began in the MDB.

Methods

Samples and SNP data

The same 263 individuals examined in chapter two are again used in this chapter and include samples from 25 locations (n=7-18), encompassing 13 catchments across the entire current MDB distribution of *N. australis* (Table 2.1; Figure 3.1). The 3443 neutral SNP markers genotyped for the previous chapter were also used in the following analyses.

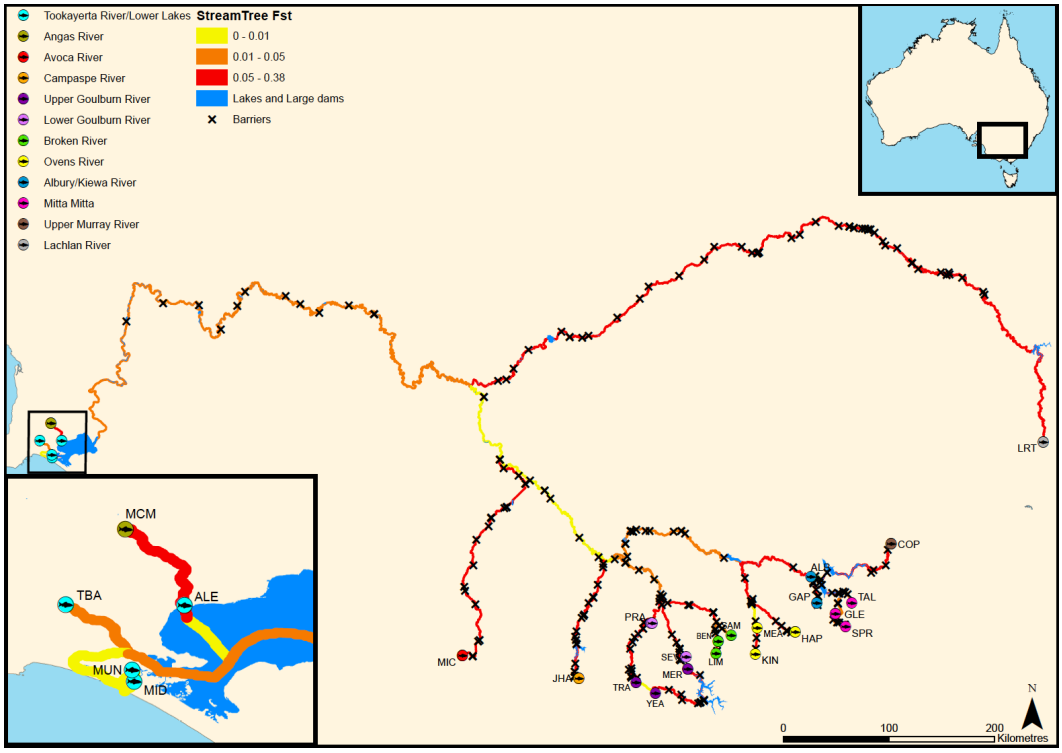


Figure 3.1: *Nannoperca australis* sampling locations in the Murray-Darling Basin (MDB). Stream sections are colour coded according to F_{ST} estimated using the *StreamTree* model. Cross markers represent the location of artificial in-stream barriers.

Anthropogenic isolation of populations

If anthropogenic habitat fragmentation has affected population connectivity and dispersal, we should expect genetic distance to increase in response to the number of in-stream barriers separating populations. Before this expectation can be tested, genetic differences between populations need to be initially mapped to the sections of the streams that connect them so other variables (e.g. spatial distance) can be accounted for. To determine if local characteristics of the stream network better explain F_{ST} than isolation by distance (IBD), we first estimated F_{ST} for each stream section following the *StreamTree* model of Kalinowski *et al*. (2008). The *StreamTree* model assigns genetic distance values to individual stream sections and can identify those parts of the network that contribute more to F_{ST} (e.g. restricted dispersal due to barriers or other local environmental conditions). Model fit was assessed by plotting the *StreamTree* fitted pairwise F_{ST} against observed F_{ST} and calculating the regression coefficient of determination (R^2). This model was then compared with a model of IBD calculated using multiple matrix regression with randomisation (MMRR) following the method of Wang (2013). This method is similar to correlation-based Mantel tests but instead implements multiple regression to quantify how genetic distance responds to multiple independent variables such as geographic and environmental distance matrices. To test for IBD we calculated pairwise population F_{ST} in *Genodive* (Meirmans & Van Tienderen 2004) and pairwise population distances along the river network with *ArcMap* v.10.2 (ESRI 2012). Both the dependent (F_{ST}) and independent (distance) variables were

standardized to mean=0 and standard deviation=1 and model significance was assessed using 10,000 random permutations.

To test the hypothesis that genetic differentiation increases with the number of in-stream barriers separating populations and to evaluate the relative contributions of anthropogenic habitat fragmentation, natural stream hierarchy and environmental variation we again used MMRR. In addition to IBD, we tested for the dependence of pairwise population F_{ST} on distance matrices calculated for the number of in-stream barriers, catchment membership, and a range of environmental variables. The number of in-stream barriers separating each pairwise combination of sites was determined using *ArcMap* and spatial data from the Murray-Darling Basin Weir Information System (Murray–Darling Basin Authority 2013). To account for the effect of natural stream hierarchy a binary model matrix describing catchment membership was constructed such that pairwise comparisons of sampling sites from within the same catchment were assigned a value of zero, and those among catchments were scored as one (Appendix 14). Finally, to test for any additional effect of isolation by environment, pairwise Euclidean distance matrices were constructed for each of the environmental variables and PCs described in chapter two of this thesis. These include temperature, rainfall, river flow and topography related variables (Appendix 10). Continuous variables were standardized to mean=0 and standard deviation=1. Each variable was first tested independently before significant factors were combined in a

multivariate model with 10,000 random permutations used to test significance.

Habitat fragmentation, genetic diversity and population size

Populations affected by habitat fragmentation are expected to lose genetic diversity due to inbreeding and drift associated with reduced population size (Frankham 2005; Fischer & Lindenmayer 2007). To test the hypothesis that the most isolated populations exhibit reduced genetic diversity we examined the relationship between population-specific F_{ST} and expected heterozygosity (H_E). Population-specific F_{ST} is a measure of genetic differentiation that estimates local population divergence from the whole metapopulation and reflects variation in the strength of genetic drift among demes due to differences in N_e (Foll & Gaggiotti 2006). This was estimated for each sampling site using the method of Weir and Hill (2002) implemented in the R package *hierfstat* (Goudet 2005) and H_E was calculated using *Genodive*.

Effective population size was estimated using the linkage disequilibrium (LD) estimator implemented in *NeEstimator* 2.01 (Do *et al.* 2014). This method is based on the assumption that LD at independently segregating loci in a finite population is a function of genetic drift, and performs particularly well with a large number of loci in situations where population sizes are expected to be small (Waples & Do 2010). In the absence of significant F_{ST} (Appendix 6), the lower Murray sites MID and MUN were

considered as one population and these samples were combined for the N_e estimates. *NeEstimator* was run assuming random mating and using a P_{crit} value of 0.075 following guidelines for small sample sizes suggested by Waples and Do (2010).

Individual-based simulations of anthropogenic population isolation

Simulation studies are becoming an increasingly important part of landscape genetics as a wide range of parameters can be explored for key evolutionary processes such as gene flow, genetic drift, mutation and selection (Epperson *et al.* 2010; Balkenhol & Landguth 2011; Hoban *et al.* 2012). When used for assessing the consequences of anthropogenic habitat fragmentation, simulations provide a means to evaluate the likelihood that the genetic signal of population isolation predates or not the onset of human impacts in the ecosystem. To address this question we simulated upland metapopulations at the within-catchment scale and assessed the impact of increasing levels of fragmentation and concurrent decreasing local population sizes on population differentiation. By simulating populations within catchments we controlled for the influence of hierarchical catchment structure typical of dendritic networks. We assumed that historical connectivity among populations within catchments was higher than today. The latter is markedly consistent with results from *fastStructure* and the AMOVAs obtained in chapter two that support within catchment connectivity (Figure 2.2; Appendix 7) and is also consistent with findings of previous studies of *N. australis* in the MDB (Attard *et al.* 2016b; Cole *et al.* 2016). Three metapopulations were modeled with i)

N_e=1000 representing approximate historical population sizes in upper Murray catchments (Appendix 15), before also simulating ii) N_e=500 (Appendix 16) and iii) N_e=100 (Appendix 17) representing more contemporary and declining N_e estimates. The historical *Ne* estimates are based on coalescent simulations using SNP sequences for *N. australis* from the MDB's Ovens River catchment (Lodge 2015; unpublished thesis). Simulations were run using *QuantiNemo* 1.04 (Neuenschwander *et al.* 2008) implemented in *Marlin* 2.0 (Meirmans 2011) and were based on a stepping stone model assuming equal N_e for each sub-population (while maintaining a constant metapopulation size to simulate the concurrent reduction in habitat patch size with increasing fragmentation). During simulations we sampled 3443 biallelic loci with a mutation rate of 10^{-8} (Martínez-Arias *et al.* 2001; Brumfield *et al.* 2003) for each individual every ten generations. Nine runs with an increasing number of demes (2-10) were completed for each metapopulation to examine the effect of increasing the number of barriers. Each simulation was allowed to run for 900 generations with 90% migration to allow F_{ST} to reach an equilibrium value of zero, before simulating the construction of barriers by reducing migration rate to zero for another 300 generations. Average pairwise F_{ST} among sites within upper Murray catchments is 0.196 (Appendix 6), therefore the number of generations (assuming a generation time of one year (Humphries 1995)) needed to achieve F_{ST}=0.2 was plotted against the number of barriers for each simulation for the three metapopulation models.

Results

Anthropogenic isolation of populations

High levels of population genetic structure were evident between most demes of *N. australis,* with pairwise comparisons of F_{ST} among sampling sites ranging from 0 to 0.79 (global F_{ST}=0.48). All pairwise F_{ST} estimates were significant (P<0.003), except between lower MDB sites MID and MUN (F_{ST}= -0.002, P=0.66) (Appendix 6). The *StreamTree* model indicated that local characteristics of the stream network better explain F_{ST} than the null hypothesis of IBD. Figure 3.1 provides a visual representation of the relationship between *StreamTree* fitted F_{ST} and the density of artificial in-stream barriers. Stream sections were colour coded in this figure according to F_{ST} as estimated by the model (yellow represents a local F_{ST} range of 0-0.01, orange: 0.01-0.05 and red: 0.05-0.38) and the location of barriers marked with X. The *StreamTree* model was significantly related to observed F_{ST} (R^2=0.947, P<1X10^{-15}) (Figure 3.2a), whereas basin-wide IBD was not significant (R^2=0.0139, P=0.334) (Figure 3.2b). Although there was significant IBD within catchment groups (i.e. the first cluster in Figure 3.2b, P=6.54X10^{-8}), IBD is not significant in the MMRR models across the whole basin whereas stream hierarchy and barriers are (see below). In addition, if only comparisons within catchments are considered, the barriers still represent a better model than IBD (R^2=0.8 compared to R^2=0.72, respectively). These findings, which indicate that populations within catchments were more connected at this regional scale until recently, are highly anticipated for freshwater fishes (Thomaz *et al.* 2016).

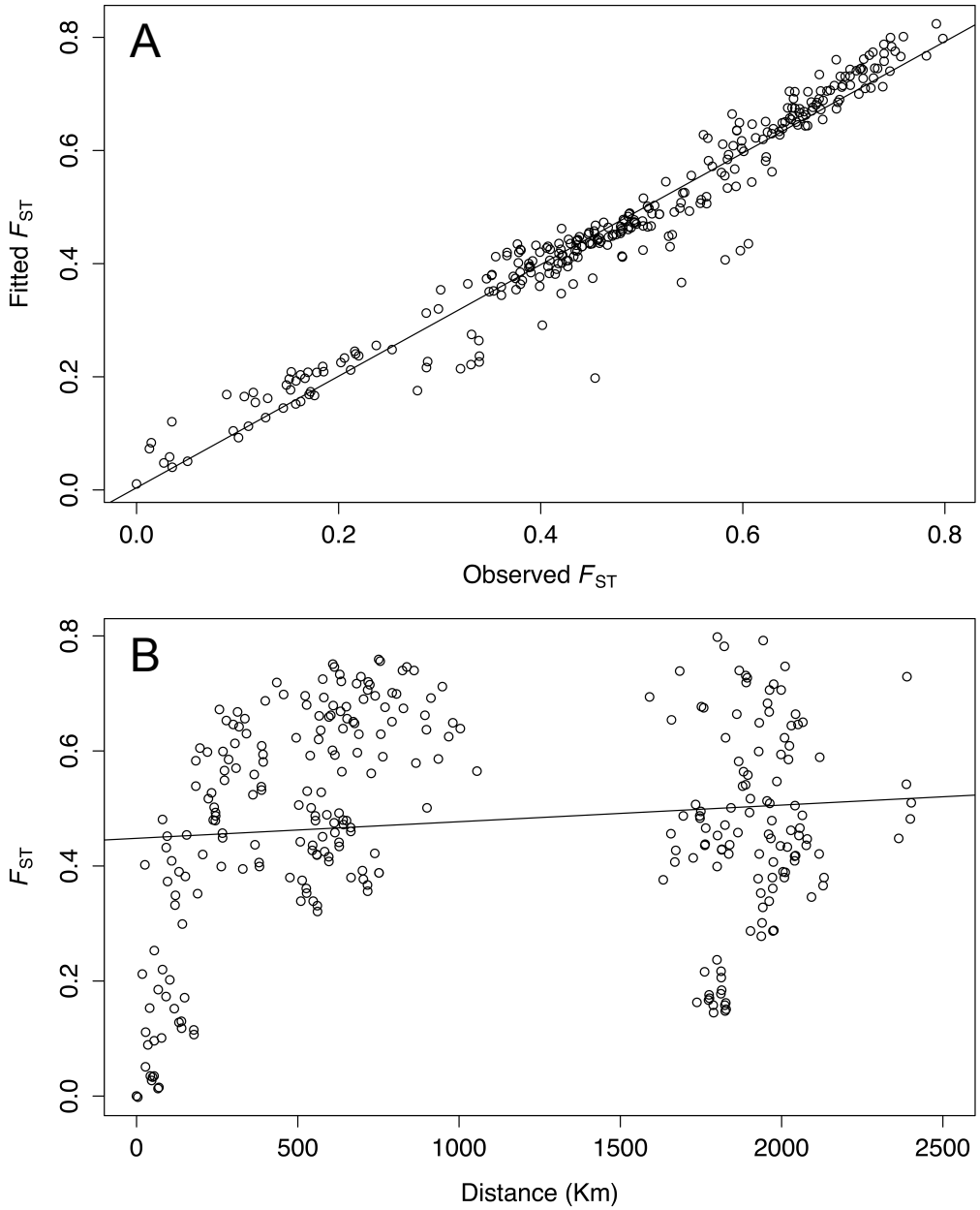


Figure 3.2: Plots of A) *StreamTree* analyses and B) isolation by distance (IBD) for *Nannoperca australis* in the MDB. The *StreamTree* plot compares fitted F_{ST} based on the *StreamTree* model with observed pairwise F_{ST} values (R^2=0.947, P<1X10^{-15}). The IBD plot depicts the relationship between pairwise F_{ST} and riverine distance between sampling sites (R^2=0.0139, P=0.334).

The MMRR results indicate that the contemporary genetic architecture of the species in the MDB is influenced by a combination of natural stream hierarchy and the recent construction of in-stream barriers. For the individual tests, catchment membership (R_2=0.170, P<0.0001) and number of barriers (R_2=0.322, P<0.0001) were both good predictors of population differentiation, while there was no evidence for isolation by environment for any of the 11 environmental variables tested (Table 3.1).

Table 3.1: Results of multiple matrix regression with randomisation (MMRR) tests for the relationship between pairwise genetic distance (F_{ST}) and geographic distance, catchment membership, number of in-stream barriers and environmental distances. *P*-values <0.0001 are indicated in bold.

Model	Variable	Coefficient	*P*-value	R^2	Model *P*-value
	Distance	0.108	0.3340	0.014	
	Catchment	1.431	**0.0001**	0.170	
	Barriers	0.548	**0.0001**	0.322	
	TempPC1	-0.119	0.2776	0.017	
	TempPC2	0.180	0.1372	0.039	
	CATCOLDQRAIN	0.098	0.3730	0.011	
	CATDRYQRAIN	-0.061	0.5539	0.004	
	STRWETQRAIN	-0.058	0.5441	0.004	
	FlowPC1	-0.053	0.6583	0.003	
	FlowPC2	-0.125	0.3601	0.019	
	CDI	0.037	0.6622	0.002	
	FRDI	-0.087	0.4627	0.009	
	TopoPC1	-0.121	0.2414	0.017	
	TopoPC2	0.021	0.8622	0.001	
Catchment+Barriers				0.358	**0.0001**
	Catchment	0.725	**0.0045**		
	Barriers	0.462	**0.0001**		

Including both catchment membership and number of barriers in a single model improved model fit (R_2=0.358, P<0.0001) (Figure 3.3). The regression coefficients (β) indicated that although historical population structure related to natural stream hierarchy strongly influences

contemporary population structure (β=0.725, P=0.0045), the number of barriers between populations contributes substantially to population isolation (β=0.462, P=0.0001) (Table 3.1).

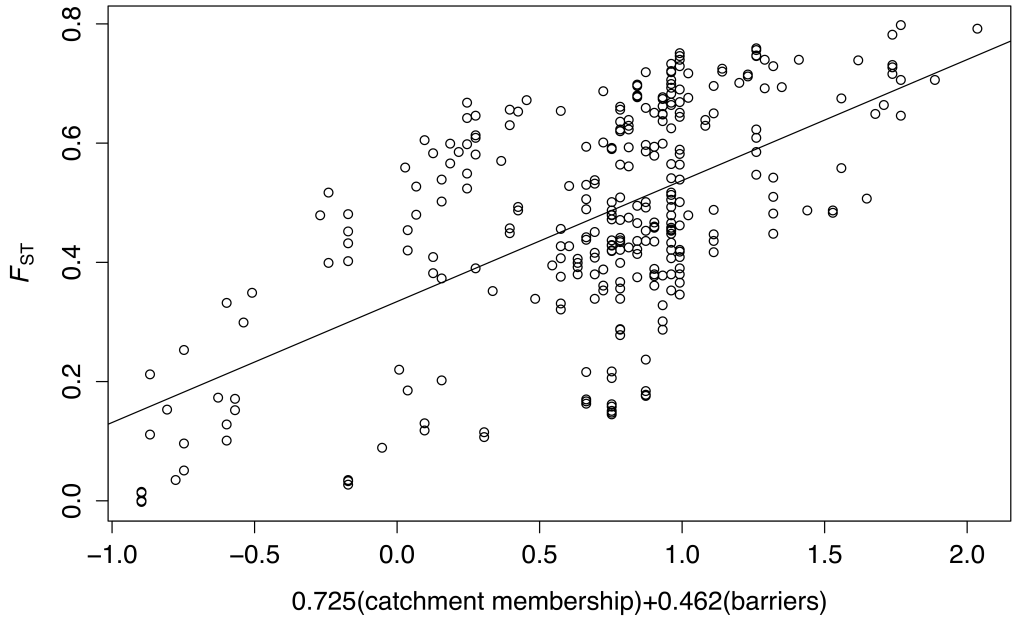


Figure 3.3: Multiple matrix regression with randomisation (MMRR) plot for the combined effects of natural stream hierarchy (model matrix of catchment membership) and number of barriers on F_{ST} (y =0.725(catchment membership)+0.462(number of barriers), R_2=0.358, P<0.0001).

Habitat fragmentation, genetic diversity and population size

Genetic variation based on 3,443 neutral SNPs was generally low but varied across the MDB with an average expected heterozygosity (H_E) of 0.161 (0.057-0.263). There was a striking contrast between regions with average H_E of 0.253 (0.227-0.263) for sites in the more connected Lower Murray wetlands compared to an average H_E 0.143 (0.057-0.243) for sites in the highly fragmented upper reaches (Table 3.2).

Table 3.2: Sample size (N), expected heterozygosity (H_E), population-specific F_{ST}(Weir & Hill 2002) and effective population size estimates (N_e). Lowland wetland sites referred to as Lower Murray in the text are indicated in bold. *MID and MUN samples combined for N_e estimation.

Catchment	Site	N	H_E	F_{ST}	N_e (95% CI)
Tookayerta (TOO)	**TBA**	**7**	**0.227**	**0.059**	∞
Lower Lakes (LMR)	**ALE**	**10**	**0.263**	**0.066**	**198.6 (158.6–264.9)**
	MID	**7**	**0.262**	**0.092**	**190.9 (163.3–229.4)***
	MUN	**6**	**0.260**	**0.034**	
Angas (ANG)	MCM	9	0.097	0.555	76.3 (61.0–101.3)
Avoca (AVO)	MIC	11	0.114	0.409	13.7 (13.2–14.4)
Campaspe (CAM)	JHA	12	0.091	0.364	393.8 (184.0–∞)
Upper Goulburn (UGO)	MER	17	0.075	0.467	70.4 (61.4–82.2)
	TRA	10	0.075	0.433	50.7 (41.2–65.3)
	YEA	8	0.087	0.364	260.4 (111.1–∞)
Lower Goulburn (LGO)	PRA	9	0.243	0.179	114.9 (98.4–137.9)
	SEV	11	0.218	0.119	54.8 (50.8–59.4)
Broken (BRO)	BEN	10	0.236	0.159	117.2 (101.7–138.2)
	SAM	10	0.234	0.188	124.7 (108.0–147.2)
	LIM	18	0.118	0.337	99.1 (88.5–112.5)
Ovens (OVE)	KIN	16	0.104	0.297	69.9 (62.1–79.8)
	HAP	9	0.114	0.369	∞
	MEA	8	0.158	0.245	53.4 (45.7–64)
Kiewa (KIE)	GAP	12	0.168	0.305	122.5 (105.3–146.2)
Albury (ALB)	ALB	12	0.226	0.299	305.4 (241.8–413.4)
Mitta Mitta (MIT)	SPR	10	0.152	0.262	98.1 (80.5–125)
	GLE	10	0.143	0.408	51.1 (46.1–57.2)
	TAL	7	0.164	0.479	31.9 (29.1–35.2)
Upper Murray (COP)	COP	16	0.133	0.297	118.7 (102.2–141.1)
Lachlan (LAC)	LRT	8	0.057	0.672	18.1 (15.3–21.8)

A strong and significant negative relationship between population specific F_{ST} and genetic diversity was evident (R^2=0.737, P<1x10^{-7}) with the most isolated populations also containing the least genetic variation (Figure 3.4; Table 3.2). Estimates of N_e were generally low with an average of 194.75 for Lower Murray sites and 112.26 for sites in the upper reaches with many of the latter <100 (Table 3.2).

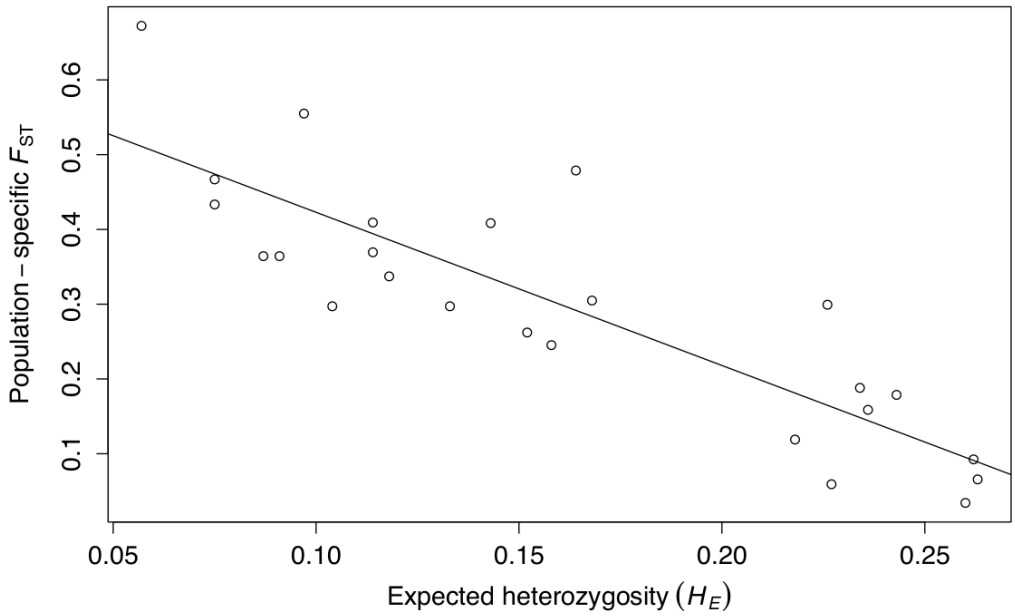


Figure 3.4: Regression plot of population-specific F_{ST} vs. expected heterozygosity (H_E) (R^2=0.737, P<0.0000001) for populations of *Nannoperca australis*.

Individual-based simulations of anthropogenic population isolation

Results of the simulations demonstrate that average observed population differentiation among sites within upper Murray catchments (F_{ST} =0.196) could have easily developed within the time since the construction of in-

stream barriers began ~120 generations ago (Figure 3.5). The metapopulation model assuming an historical N_e of 1000 indicated that F_{ST} increased from zero to 0.2 in less than 120 generations with only three barriers fragmenting the population (Figure 3.5; Appendix 18). Models assuming more contemporary estimates of population size (N_e=500 and N_e=100) indicated that substantially fewer generations following fragmentation were required for F_{ST} to reach 0.2. At N_e=500, F_{ST} =0.2 occurred after 101 generations with one barrier and after just 21 generations with nine barriers (Figure 3.5; Appendix 19). For the simulation assuming N_e=100 the results were even more striking with F_{ST} =0.2 achieved within 21 generations with just one barrier and only four generations if fragmented by nine barriers (Figure 3.5; Appendix 20).

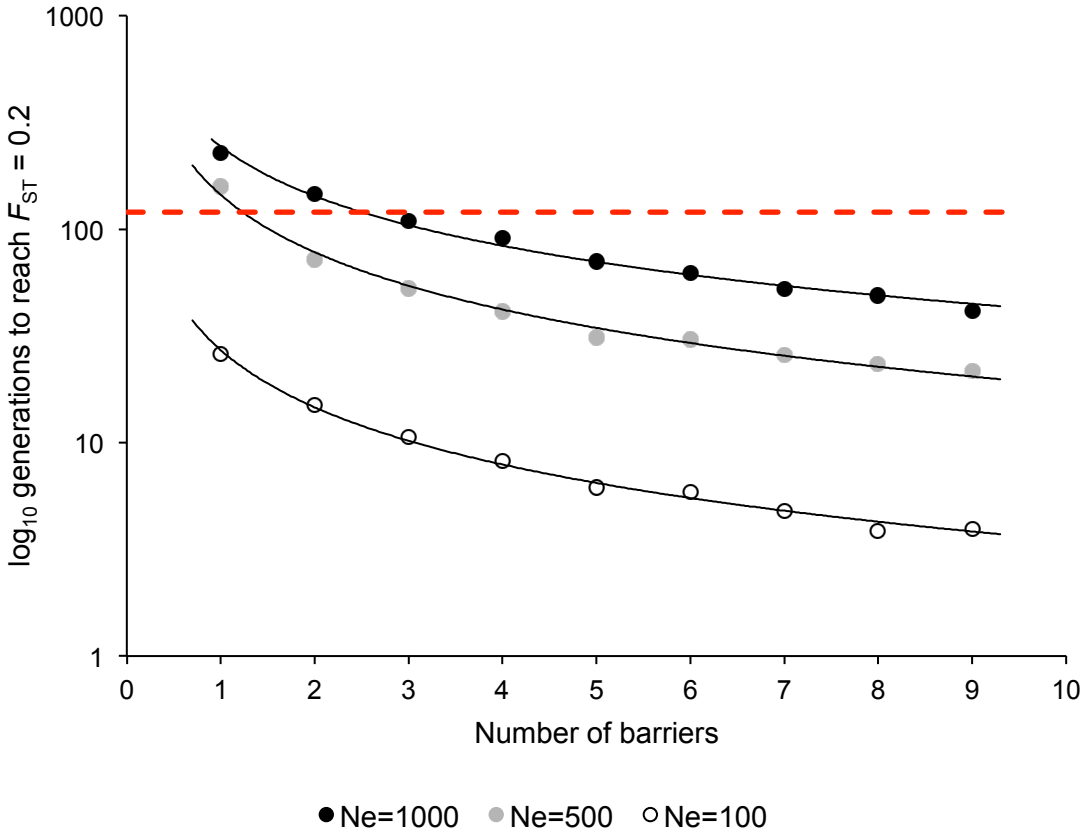


Figure 3.5: Number of generations (log_{10}) for global F_{ST} to reach 0.2 with increasing levels of habitat fragmentation for simulated *N. australis* metapopulations of N_e=1000, N_e =500 and N_e =100. Simulations were based on a stepping stone model assuming equal N_e for each sub-population and were allowed to run for 900 generations with 90% migration before 300 generations with no migration. Red dashed line indicates the approximate number of generations since construction of in-stream barriers began in the MDB (120 generations).

Discussion

Habitat fragmentation is a key process implicated in the current and unprecedented worldwide decline of freshwater biodiversity (Sala *et al.* 2000; Fischer & Lindenmayer 2007). Linking recent human activities to the genetic decline of wild populations is however challenging, as natural historic spatial population structure and demographic stochasticity can confound the effects of more recent disturbance, particularly for species

with limited dispersal abilities (Landguth *et al.* 2010). Here we demonstrated that recent anthropogenic habitat fragmentation has likely contributed to the decline of freshwater biodiversity across one of Australia's most important and threatened riverine ecosystems. Populations of *N. australis* most isolated by habitat fragmentation showed reduced genetic diversity and increased population differentiation, and this signal remained strong after accounting for the historical effects of natural stream hierarchy. Additionally, simulations confirm that current levels of population differentiation could have evolved within the time since European settlement of the region. Many populations sampled for this study have subsequently suffered local extirpation during prolonged drought, and the small N_e estimates obtained for most remnant populations indicate they are at high risk of extinction. These findings not only highlight the critical nature of conservation efforts for this species in the MDB, but also suggest that many aquatic species in the region likely face an increased risk of extinction as a direct result of anthropogenic habitat fragmentation reducing connectivity and genetic diversity of populations.

A fundamental objective in conservation biology is to understand how humans are impacting the distribution of biodiversity and threatening population, and ultimately species persistence (Soulé 1985). To achieve that goal it is necessary to consider both historic and contemporary evolutionary processes likely to have contributed to the spatial distribution

of genetic diversity (Waples *et al.* 2008; Storfer *et al.* 2010). Natural hydrological structure influences population connectivity for many freshwater species (Hughes *et al.* 2009) and the resulting genetic signal of spatial population structure can persist for many generations following disturbance (Landguth *et al.* 2010). For *N. australis*, this is reflected in the strong association between genetic population structure and natural stream hierarchy (Table 3.1). Ideally, a combination of historical and contemporary samples could be used to examine the relative contribution of natural hydrological structure and recent habitat fragmentation to contemporary population genetic structure. For instance, Fountain *et al*. (2016) compared historical samples with more recent data to identify changes in genetic diversity related to recent habitat fragmentation. It should be noted however that for *N. australis* (and indeed for most, if not all Australian flora and fauna), historical data with which to make temporal comparisons before, and after human disturbance do not exist. In this case, the use of high resolution SNP data enables clarification of the relative influence of historic and contemporary processes and several key results from this study indicate anthropogenic habitat fragmentation is contributing to population genetic divergence and demographic decline of this species. The number of in-stream barriers between populations was a significant predictor of F_{ST} in the MMRR (Table 3.1) and contributed substantially to the model that also included natural stream hierarchy (Figure 3.3). This finding is further supported by the *StreamTree* model that indicates local stream characteristics better explain F_{ST} than IBD (Figure 3.2). Previous studies based on microsatellites suggest that

historical population sizes in the Lower MDB were much larger before European settlement (Attard *et al.* 2016b), and that populations across the MDB were also more connected until that time (Cole *et al.* 2016). These studies support the notion that the low contemporary N_e estimates obtained for many populations in this study are most likely the result of recent processes rather than due to natural demographic variability over longer evolutionary time scales. Finally, pairwise F_{ST} estimates among sites within catchments are well within the range obtained by simulating population isolation due to the construction of in-stream barriers over the time since European settlement (Figure 3.5). Together these results provide strong evidence that anthropogenic habitat fragmentation has contributed to the recent isolation of populations, resulting in reduced genetic diversity and increased the risk of extinction for *N. australis* in the MDB.

Habitat fragmentation is considered one of the greatest threats to freshwater biodiversity worldwide (Vorosmarty *et al.* 2010) and as climate change intensifies, proactive conservation management interventions will be increasingly required (Palmer *et al.* 2008). Managing regulated river systems to provide environmental flows and other measures to restore connectivity among habitat patches such as fishways can potentially address some impacts. However these long-term, landscape-scale measures are often constrained by political and socio-economic issues (Dudgeon *et al.* 2006; Davis *et al.* 2015) and many species may be

already depleted to the point where improved environmental conditions alone will not be enough to facilitate recovery. Genetic rescue via translocations offers a potential solution for a broad range of threatened taxa, however despite strong evidence supporting the benefits of genetic rescue for fragmented populations, conservation managers have been reluctant to adopt these measures (Frankham 2015). Instead, management of threatened species has focused on defining conservation units on the basis that historically isolated evolutionary lineages (i.e. evolutionary significant units, ESUs), and demographically independent populations (i.e. management units, MUs) should be managed separately (Funk *et al.* 2012). The widespread use of conservation units has recently been criticized due to the risk of preserving genetic uniqueness at the expense of species level genetic diversity (Weeks *et al.* 2016) and our findings also suggest caution here. We argue that the influence of recent habitat fragmentation on patterns of spatial population structure has likely been underestimated for many species. In this case, estimates of population structure erroneously attributed to natural evolutionary processes may have potentially led to management initiatives that actually promote fragmentation of historically connected populations.

A related discussion emerging from current advances in genomic methods is how to best utilize information concerning adaptive genetic variation to potentially build evolutionary resilience of species and even ecological communities (Webster *et al.* 2017). Contrary to prevailing theory, it is

becoming increasingly apparent that selection can still influence the evolutionary trajectory of small and fragmented populations (Fraser *et al.* 2014; Wood *et al.* 2015; Brauer *et al.* 2016; Wood *et al.* 2016; Sætre *et al.* 2017). The molecular mechanisms underpinning adaptive evolution of small populations are not yet clear. However it seems possible that altered selection regimes resulting from increased environmental variability within fragmented habitat patches may favour the maintenance of adaptive genetic variation, while environmental heterogeneity among isolated populations may be simultaneously driving adaptive divergence. Critically for conservation, this indicates that adaptive divergence of small populations can occur quickly following fragmentation, and that even very recently isolated populations may harbor unique combinations of alleles. Rather than manage these populations separately however, it may be more beneficial to facilitate genetic exchange among demes, thereby restoring natural evolutionary processes and maintaining species-level genetic variation potentially valuable under a range of future selection regimes. In this case conservation strategies need to evolve and become more proactive to build evolutionary resilience of managed populations (Aitken & Whitlock 2013; Harrisson *et al.* 2014; Webster *et al.* 2017).

Chapter 4: Comparative ecological transcriptomics and the contribution of gene expression to the evolutionary potential of a threatened freshwater fish

Abstract

The ability for species to adapt and persist is being challenged by a combination of rapid environmental change, widespread habitat loss and fragmentation. Populations trapped in isolated and often poor quality habitat fragments are more susceptible to stochastic demographic and genetic declines. Understanding the capacity for small populations with low genetic diversity to respond to rapid environmental change through phenotypic plasticity is a key research question in conservation biology. Information on adaptive traits for threatened species is often limited or non-existing. However, RNA sequencing (RNA-seq) has recently provided the opportunity to examine variation in gene expression, a surrogate for phenotypic variation, in non-model species. Gene expression data can provide insight into the most basic link between genotypes and complex phenotypic traits shaped by ecological and evolutionary processes. Here we used RNA-seq to assemble the first transcriptome for an Australian percichthyid, a dominant and ecologically important group of freshwater fishes in Australia. Expression variation within- and among five select populations of *Nannoperca australis* was assessed, and several hypotheses related to how plastic and divergent gene expression profiles may evolve in response to environmental and genetic variation in wild populations were tested in a comparative transcriptomics framework. The results show that variation in gene expression was not constrained by genetic diversity. After accounting for phylogenetic structure, we found evidence for high expression plasticity in 24 genes, and identified 165 candidates for divergent gene expression. Functional annotation of the

candidate genes suggests populations are responding to variations in water quality through both plastic and evolved variation in gene expression. These findings indicate that despite strong drift, plastic and evolved phenotypic responses do contribute to evolutionary potential of small and isolated populations of *N. australis*.

Introduction

Understanding the mechanisms by which species may persist in changed, and often sub-optimal conditions is vital for identifying populations at risk of extinction and for improving conservation measures increasingly employed to mitigate the current loss of biodiversity (Hoffmann & Sgro 2011). As more land is converted from a natural state to agricultural, industrial and urban environments, habitats have become fragmented, limiting species opportunity for dispersal and migration (Fischer & Lindenmayer 2007). Adaptation from standing genetic variation is one way species can respond to environmental change, and it is increasingly suggested that even small and threatened populations may retain variation at potentially adaptive loci and be able to respond to rapid change (Koskinen *et al.* 2002; Fraser *et al.* 2014; Brauer *et al.* 2016; Wood *et al.* 2016). Phenotypic plasticity – the ability for multiple phenotypes to arise from a single genotype – is another mechanism that may facilitate population persistence in altered environments and potentially lead to evolutionary adaptation (Chevin *et al.* 2010; Dayan *et al.* 2015; Ghalambor *et al.* 2015). These two mechanisms are not mutually exclusive and empirical examples featuring wild populations suggest that rapid phenotypic changes often involve a combination of genetic adaptation and phenotypic plasticity (Réale *et al.* 2003; Charmantier *et al.* 2008; Gienapp *et al.* 2008; van de Pol *et al.* 2012). While the former has received substantial recent attention, relatively few studies have examined the role or extent of phenotypic plasticity in wild populations (Wood & Fraser 2015).

Our ability to predict the capacity of species to use phenotypic plasticity when responding to environmental change depends on the identification of traits affecting fitness in the new environment. For cryptic or threatened species especially, information linking phenotypic variation and fitness may often be scarce. In such cases, any investigation will always represent a balance between exploring and explaining trait variation (Houle *et al.* 2010). An alternative strategy for study systems for which the knowledge of important traits is lacking or the ability to measure them is limited is to consider gene expression measurements as phenotypic traits. The advent of high-throughput genomic methods such as microarrays and, more recently, RNA sequencing (RNA-seq) has seen an increase in gene expression studies of non-model species (Alvarez *et al.* 2015). RNA-seq measures global levels of mRNA transcription that are often used as a surrogate for gene expression. These data can provide insight into the most basic link between genotypes and complex phenotypic traits shaped by ecological and evolutionary processes (Whitehead 2012). Additionally, the simultaneous measurement of vast numbers of traits facilitated by RNA-seq may reveal cryptic evolutionary patterns not discernible when fewer phenotypic traits are considered (Houle *et al.* 2010).

Comparative gene expression or transcriptomic analyses of wild populations can contribute to our understanding of the molecular basis (both plastic and evolved) of physiological responses to environmental

stressors (Whitehead *et al.* 2010; Romero *et al.* 2012). However, several challenges exist in analyzing and interpreting comparative gene expression data, especially in the case of wild populations of non-model species. For instance, phylogenetic distance needs to be accounted for when comparing gene expression among groups (Dunn *et al.* 2013); cryptic or transient environmental factors, or developmental effects may bias results due to sampling just one time point (DeBiasse & Kelly 2016); and the biological interpretation of functional annotations derived from distantly related taxa may be misleading (Pavey *et al.* 2012). Nevertheless, comparative studies of wild populations can provide information concerning the effects of multiple and dynamic environmental conditions on gene expression not otherwise obtainable in more controlled experimental conditions (Alvarez *et al.* 2015). However, when populations sampled over an environmental gradient are also isolated by geographic and phylogenetic distance, it becomes difficult to determine if differences in expression represent plastic or adaptive responses to variation in the environment, or are simply due to neutral drift (Khaitovich *et al.* 2004).

While significant research effort in genetics has been directed towards understanding the effects of population size and drift on evolutionary potential, surprisingly few studies have considered the impact of drift on phenotypic plasticity (Chevin *et al.* 2013; Wood & Fraser 2015). On the one hand, small fragmented populations are likely to exhibit reduced genetic diversity. Given the mounting evidence for an underlying heritable

component of variation in gene expression (Gibson & Weir 2005; Leder *et al.* 2015; McCairns *et al.* 2016), it might therefore be reasonable to expect that where genetic diversity has been eroded by drift, plasticity in gene expression may also be impaired (Bijlsma & Loeschcke 2012). On the other hand, if fragmentation causes an overall decrease in habitat quality and also increases variation in environmental conditions, natural selection may maintain phenotypic plasticity in small populations for traits important in responding to environmental stressors (Paschke *et al.* 2003; Chevin & Lande 2011).

Studies of differential expression among populations inhabiting different environments can potentially identify important genes because of absolute differences in expression between populations or lineages. It is however necessary to account for the phylogenetic component of trait variation when assessing environmental influences on gene expression variation among wild populations (Martins *et al.* 2002). Analyses based on tests of ANOVA have often been used in comparative transcriptomics studies that contrast variation in gene expression within species to variation among species. In many cases these analyses have incorporated a correction for phylogenetic effects (Oleksiak *et al.* 2002; Oleksiak *et al.* 2005; Dayan *et al.* 2015; Uebbing *et al.* 2016). A recent extension of ANOVA-based methods is the Expression Variance and Evolution model (*EVE*) (Rohlfs & Nielsen 2015). *EVE* models gene expression as a quantitative trait across a phylogeny; considering the ratio (β) of among lineage expression

divergence to within lineage expression diversity in a similar manner to the Hudson Kreitman Aguadé (HKA) test used to detect molecular evolution in DNA sequences (Hudson *et al.* 1987). The expectation is that *β* should be consistent among the majority of genes that have undergone similar evolutionary and demographic processes, but higher for genes with more variance within- than among-lineages, and lower for genes with more variance among- compared to within-lineages. While models such as *EVE* are primarily intended for interspecific studies, Leder *et al*. (2015) recently suggested that the effects of demography and natural selection may exert the greatest influence on expression divergence during, or immediately following, lineage divergence. Therefore, comparative gene expression studies among recently isolated but demographically independent populations may offer insights into key genetic and environmental elements of expression plasticity during rapid environmental, ecological and evolutionary change. This might be particularly the case for populations spanning environmentally heterogeneous landscapes.

The freshwater fish *Nannoperca australis* represents a good system for examining variation in gene expression in the wild and in the context of conservation. Despite indications that populations across the Murray-Darling basin (MDB) were more connected in the past (Attard *et al.* 2016b; Cole *et al.* 2016), the impact of drift due to recent demographic decline and isolation of populations has resulted in remarkable population differentiation (Brauer *et al.* 2016; Cole *et al.* 2016). Additionally, there is

evidence for adaptive divergence of isolated populations occupying a range of naturally variable hydroclimatic environments subjected to varying degrees of human impacts (Morrongiello *et al.* 2012; Brauer *et al.* 2016). The long-term natural climatic variability and unpredictability found across the MDB also suggest that selection may not only lead to local adaptation, but may also influence plasticity of traits related to maintaining fitness in a variable environment.

Determining the role of phenotypic plasticity in facilitating population persistence and adaptive evolution in changing environments is a key research question in ecology and evolution (Alvarez *et al.* 2015; DeBiasse & Kelly 2016). An initial step towards this goal, and the aim of this study, is to understand how patterns of gene expression vary within- and among-populations, and to examine the relative contribution of plastic (environmental) and evolutionary (genetic) components in shaping these patterns in the wild. In this chapter an RNA-seq approach is used to construct and functionally annotate a *de novo* transcriptome for *Nannoperca*. We then use this resource to examine patterns of global (i.e. transcriptome-wide) gene expression variation among select wild populations of *N. australis*.

The question of how population size affects the evolutionary potential of populations has received recent attention, with some evidence now suggesting that adaptive potential can be maintained even in very small

populations (Fraser *et al.* 2014; Wood *et al.* 2015). This subsequently raises the question of whether phenotypic responses to environmental change are also affected by population size? Here we test the hypotheses that variance in global gene expression, in this case a surrogate for phenotypic plasticity, varies among isolated populations, and that variance in gene expression is correlated with levels of genetic diversity. This study also addresses several outstanding questions related to how plastic and divergent gene expression profiles may evolve in response to environmental and genetic variation in wild populations. First, multivariate models of gene expression profiles were used to examine the relationship between variance in global gene expression and genetic diversity within, and among five populations of wild-born *N. australis*. Second, using a recently developed comparative phylogenetic ANOVA-based expression variance and evolution model, we identify candidate genes potentially under divergent selection for expression level, and other genes with a signal of high expression level plasticity. This work represents an initial examination of the role and extent of gene expression variation in wild populations of Australian percichthyid fish, and results here lay the foundation for further research into the conservation, ecology and evolution of this group.

Methods

Sampling

Nannoperca australis were collected from five populations selected

specifically to capture maximum variation in the hydroclimatic gradient across the MDB and to include populations with high (SPBR, SEV), intermediate (LIM, KIN) and low (MER) levels of genetic diversity (Table 4.1). The lower Murray environment is considered semi-arid with warmer winter temperatures and far less rainfall than elsewhere in the MDB (Figure 2.3). The relatively well-connected lakes and wetlands of this region contrast with the small, isolated rivers and creeks typical of headwater habitats (Hammer *et al.* 2013). Hydroclimatic conditions are wetter in the headwater sites (especially in KIN), with generally warmer summer temperatures but cooler winter temperatures (Figure 2.3).

Table 4.1: Information about localities and sample sizes (N) and mean individual heterozygosity (IH) for *Nannoperca australis* from the Murray-Darling Basin (MDB). Captive fish from lower Murray sites referred to as SPBR in the text are indicated in bold.

Population	Location	N	IH (SD)	Latitude	Longitude
SPBR	**Turvey's Drain, L. Alexandrina**	**10**	**0.16 (0.04)**	**-35.395**	**139.008**
	Mundoo Is., L. Alexandrina	**7**		**-35.549**	**138.915**
MER	Merton Ck	17	0.06 (0.01)	-36.981	145.727
SEV	trib to Seven Creeks	11	0.17 (0.04)	-36.875	145.701
LIM	Unnamed Ck, Lima South	18	0.09 (0.03)	-36.826	146.008
KIN	King R., Cheshunt	16	0.08 (0.02)	-36.795	146.424

Fish were collected directly from the wild at four upper MDB sites, while wild-born but captive held individuals were sampled from the lower Murray (Figure 4.1). Although the use of captive fish may potentially influence some results, there was no other alternative as this population is critically endangered after being locally extirpated from the wild during the recent drought. The fish sampled here were rescued just prior to the complete loss of habitat in the lower Murray, and were part of the founding captive

breeding population (Hammer *et al.* 2013; Attard *et al.* 2016b). It was considered important to include this population, as these are the only representatives of this geographically isolated, environmentally divergent and ecologically important region.

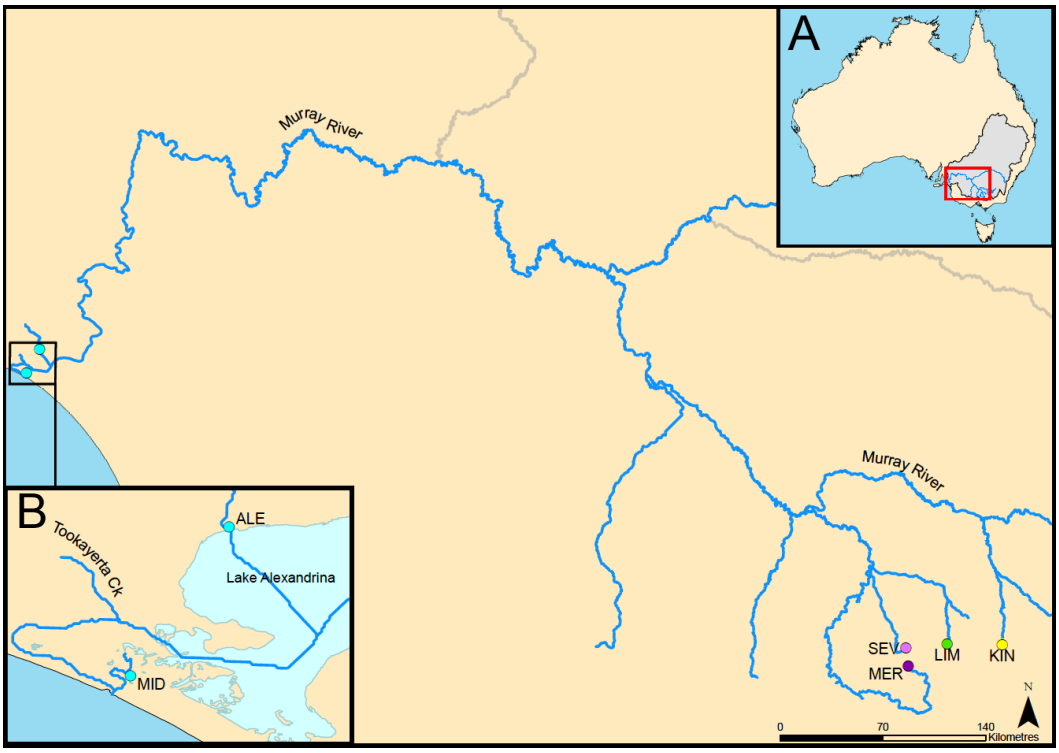


Figure 4.1 *Nannoperca australis* RNA-seq sampling locations. Inset A shows the location of the Murray-Darling Basin (shaded area). Captive fish from lower Murray sites referred to as SPBR (southern pygmy perch breeders) in the text originate from sites ALE and MID in the lower Murray (inset B) and were part of the founding captive population contributing to conservation breeding efforts described in Chapter 2. Sites are colour coded by catchment based on the colours in Figure 2.1.

A combination of dip netting and electrofishing was used to collect approximately ten adult males from each site. Only males were used to account for putative differences in expression between males and females (Smith *et al.* 2013). Fish were euthanized in an overdose of AQUI-S®

solution (50% isoeugenol) and immediately dissected to extract the liver and confirm the sex. Liver tissue was incubated at 4°C for 12 hours in RNA*later* (Ambion) following the manufacturer's protocol before freezing in liquid nitrogen for transport and subsequent laboratory storage at -80°C. Liver tissue was selected because gene expression in liver is known to respond to environmental stimuli in fish, such as variation in temperature (Rabergh *et al.* 2000; Smith *et al.* 2013; McCairns *et al.* 2016). In addition to the five wild populations, samples were also obtained from F_1 generation captive born *N. australis*, and wild and F_1 Yarra pygmy perch, *N. obscura*. These extra samples were included to increase the quality of the *de novo* transcriptome assembly, but were not used for further downstream analyses in this study.

RNA extraction and library preparation

Total RNA was extracted from approximately 5mg of tissue for each sample using a MagMax 96 Total RNA Extraction Kit (Life Sciences) following the manufacturer's protocol. Integrity and concentration of RNA was evaluated with RNA Nano assay kit on an Agilent Bioanalyzer 2100 (Agilent Technologies) and RNA purity was assessed using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

Samples were normalized to a starting quantity of 0.9 µg total RNA and individual Illumina sequencing libraries were prepared for 48 samples using a TruSeq RNA sample preparation kit (Illumina) following the low

sample protocol. Briefly, poly-A containing mRNA was first purified with magnetic beads before fragmenting the RNA by incubating at 94°C. SuperScript II reverse transcriptase was used to synthesize the first strand of cDNA after which the RNA template was removed and replaced with a second cDNA strand to produce double strand cDNA. Illumina adapter indices 2, 4–7, 12–16, 18 and 19 were ligated to the cDNA with the 12 barcodes assigned to samples for pooled sequencing across four Illumina lanes. Fragments with adapters at both ends were amplified using PCR and the resulting libraries validated using an Agilent Bioanalyzer 2100 before normalizing and pooling 12 individual libraries for each sequencing lane. Paired-end, 100 base-pair sequencing was performed on an Illumina Hiseq2000 at the Genome Quebec/McGill University Innovation Center (Montreal, Canada).

Read trimming, de novo transcriptome assembly and quality assessment

Raw sequence reads were demultiplexed according to individual indices at the sequencing facility. *Trimmomatic* v.0.33 (Bolger *et al.* 2014) was used to remove adapter sequences and trim low quality bases. Bases with a PHRED score <23 were trimmed from the ends of each read before further trimming to also remove any reads with a total length shorter than 45 bases, and any bases where the average PHRED score within a five base sliding window fell below 23. Both R1 and R2 reads from all 48 libraries were combined and assembled using *Trinity* v2.0.6 (Grabherr *et al.* 2011). The *Trinity de novo* assembly pipeline (Haas *et al.* 2013)

combines three separate software modules; *Inchworm*, *Chrysalis*, and *Butterfly*. *Inchworm* assembles the raw reads into longer transcripts, which get passed to *Chrysalis* where similar contigs are sorted into clusters. *Chrysalis* then constructs de Bruijn graphs for each cluster, representing the full transcriptional complexity of a 'gene', where multiple paths through each graph track alternative isoforms. Finally, *Butterfly* processes the individual graphs to report dominant, and alternative isoforms of each gene and to identify transcripts belonging to paralogous genes. Prior to assembly, *in silico* normalization was implemented to reduce redundancy in the data by limiting read coverage to a maximum of 50x, in order to reduce memory requirements and improve computational time.

In addition to the main assembly software, *Trinity* also includes a number of utilities to facilitate integrated transcriptome quality assessment and other downstream analyses such as individual sample transcript quantification, differential expression analyses and functional annotation. Using these utilities, statistics describing read representation, N50 values, the number of full-length protein transcripts and the number of BUSCO (Benchmarking Universal Single-Copy Orthologs) conserved orthologs (Simão *et al.* 2015), were generated to assess quality of the transcriptome assembly. Sequence reads retained after quality filtering were mapped back to the assembled transcripts to examine the overall number of reads mapping to the assembly and also the proportion of those mapped reads occurring as proper forward and reverse pairs. Read mapping was

performed using *Bowtie* v1.1.2 (Langmead *et al.* 2009) with default settings specified in the *Trinity* bowtie_PE_separate_then_join.pl script. The length distribution of assembled transcript contigs was assessed by calculating the N10 through to N50 values (N50 is the minimum length of transcript contigs in which 50% of all assembled bases are found) using the TrinityStats.pl script. A BLAST (Tao 2014) search of transcripts was performed against the SwissProt protein database (UniProt Consortium 2015) to examine the length distribution and number of unique, top matching proteins covered by the transcriptome. The *Trinity* script analyze_blastPlus_topHit_coverage.pl was used to call *blastx* with an *e*-value threshold of $1x10^{-20}$. Finally, to quantify completeness of the assembly in terms of gene content, the transcriptome was assessed against the vertebrate BUSCO database (http://busco.ezlab.org/). This database consists of 3023 evolutionarily conserved genes expected to be found as single-copy orthologs in >90% of vertebrate species (Simão *et al.* 2015).

Functional annotation and gene ontology

Homology searches of several sequence and protein databases were performed using *Trinotate* v2.0.2 to assign functional annotations to the transcriptome. *Transdecoder* v2.0.1 was used to extract open read frames >100 bases from the *Trinity* transcripts and identify candidate coding regions. Both the *Trinity* transcripts, and the *Transdecoder* predicted coding regions were then blasted against the SwissProt and smaller Uniref90 sequence databases (UniProt Consortium 2015) to provide gene

annotation and assign gene ontology terms. Additionally, the *Transdecoder*-predicted coding regions were searched for homologies with the Pfam protein family domain (Bateman *et al.* 2004), protein signal peptide (Petersen *et al.* 2011) and transmembrane protein domain (Krogh *et al.* 2001) databases (*e*-value thresholds of $1x10^{-5}$). The resulting BLAST homologies were loaded into a SQLite database along with the transcriptome to generate an annotation report and to provide gene ontology (GO) information (Botstein *et al.* 2000) for downstream functional enrichment analyses.

Transcript quantification and differential expression analysis

To quantify the level of transcription for individual samples, reads for each sample were first mapped back to the transcriptome using *Bowtie* v1.1.2 (Langmead *et al.* 2009), before abundance estimations at both transcript- and gene-level were performed with *RSEM* v1.2.19 (Li & Dewey 2011). To enable comparison of expression level among samples, the resulting read count estimations were also cross sample normalized using the trimmed mean of M-values (TMM) method implemented in *edgeR* v3.12.0 (Robinson *et al.* 2010).

Pairwise comparisons of differential expression (DE) among populations were estimated in both *edgeR* (Robinson *et al.* 2010), and *DESeq2* v1.10.1 (Love *et al.* 2014). Genes with a minimum $\log_2$ fold change of two between any two populations were considered differentially expressed at a

false discovery threshold of $1x10^{-3}$. Heatmaps describing the correlation among samples, and gene expression per sample were generated using the *Trinity* analyze_diff_expr.pl utility to allow visual analysis of patterns of expression.

Functional GO enrichment analysis for DE genes was performed using the Bioconductor R package *GOseq* v1.22.0 (Young *et al.* 2010). *Goseq* is a popular method for GO analyses and can account for the bias in DE gene detection for long and highly expressed transcripts common to RNA-seq data (Young *et al.* 2010). Gene ontology terms for the DE genes were retrieved from the earlier BLAST annotation results and tested for enrichment compared to all GO term assignments for the transcriptome assembly.

Is variance in gene expression constrained by drift?

A multivariate analogue of Levene's test for homogeneity of variances was used to test for differences in inter-individual gene expression variance among populations. A Euclidean distance matrix was first constructed for all samples based on the TMM normalized expression matrix. Using the *betadisper* function in the *vegan* R package the distance matrix was then reduced to principal coordinates and the distance of each individual to the population centroid (average population multivariate expression profile) was calculated and subjected to ANOVA. A total of 9,999 permutations was used to test for significant departure from the null hypothesis of no

difference in variation among populations. Tukey's test for significant differences between groups was also applied using the *TukeyHSD* R function (also part of *vegan*) to test for pairwise population differences in within-population mean expression variance.

To test the hypothesis that gene expression variance, here a surrogate for phenotypic plasticity, is constrained by genetic diversity (Bijlsma & Loeschcke 2012), individual heterozygosity was regressed against a Euclidean distance matrix based on TMM normalized expression values at all *Trinity* 'genes' using the *adonis* function in *vegan*. This function performs an analysis of variance using distance matrices and allows linear models to be fitted to multiple matrices. Individual heterozygosity was calculated as the proportion of heterozygous loci per individual at 3443 neutral, and at 216 GEA candidate SNP loci (SNP data described in chapter two). The test was performed separately for each data set to assess the possibility that expression variance may respond differently to putatively neutral and candidate loci. One individual from site MER was dropped from these analyses, as reliable estimates of genetic diversity were unable to be obtained due to a high proportion of missing data in its SNP dataset. A stratified permutation test with pseudo-*F* ratios was performed to test significance of the portion of gene expression attributed to variation in genetic diversity using 9,999 permutations within each population.

Gene expression plasticity and adaptation

The advent of RNA-seq provides a powerful platform for examining the mechanisms behind non-model species' capacity to persist in variable environments (Harrisson *et al.* 2014). Here the recently developed *EVE* model (Rohlfs & Nielsen 2015) was used to parameterize the ratio *β* across the five populations of *N. australis*, to identify genes potentially under divergent selection for expression level (low *β*) or genes with high expression plasticity (high *β*). The model requires gene expression data and a phylogeny as input. In order to improve computational time whilst maintaining maximum biological information, a reduced expression matrix was used based on genes for which gene ontology terms had been assigned. Following TMM normalization, some lowly expressed genes can end up with an effective expression level of zero across all samples and these genes were also filtered from the data. To construct the phylogenetic tree, *pyRAD* v3.0.6 (Eaton 2014) was first used to align ddRAD sequences from the same set of individuals and an additional five Yarra pygmy perch (*N. obscura*), included as an out-group (Unmack *et al.* 2011) (*pyRAD* parameters are specified in Appendix 21). Maximum likelihood phylogenetic analyses were then run with *RAxML* v8.0.26 (Stamatakis 2014), specifying a GTRGAMMA model and 1000 bootstrap replicates. The majority-rule consensus tree was used as the input phylogeny. For each gene (*i*), maximum likelihood values were calculated and a likelihood ratio test (LRT) was performed to test the null hypothesis that β_i is equal to *β* for all genes. Under the null model, the LRT statistic follows a χ_1^2 distribution (Rohlfs & Nielsen 2015) and a custom R script

(Appendix 22) was used to identify genes where the LRT statistic deviated from this distribution at a FDR of 10%. Functional GO enrichment analysis was performed as previously described for the DE analysis. Interpreting GO information can be difficult as many terms can be assigned to one gene, and many genes may share the same GO term. To aid interpretation, enriched GO terms were submitted to the REVIGO web server (Supek *et al.* 2011) where treemaps were generated to enable visual summaries of the main GO categories based on semantic similarity of terms. Finally, a BLAST search was performed of flanking sequences from the 216 GEA candidate SNP loci against the transcriptome assembly (*e*-value threshold 1 x 10^{-3}) to identify any sequence homologies between the candidate SNP and candidate gene data sets.

Results

Sequencing and assembly

The four Illumina lanes produced over 770 million paired end reads (2x100 bp) and, after trimming and quality filtering, 707 million read pairs (91.8%) were retained (Table 4.2). This resulted in an average of 14,744,482 read pairs per individual (min=5,789,189, max=35,515,038) retained for downstream analyses (Appendix 23). Close to 1.2 billion reads aligned to the *de novo* assembly (64.6% mapping as proper pairs), consisting of 445,324 unique transcripts that clustered into 273,111 Trinity 'genes' (Table 4.2; Appendix 24). Based on all transcript contigs, a N50 of 1810 bases (mean=933, total assembled bases=415,611,353) was achieved

(Table 4.2; Appendix 25). A BLAST search of the SwissProt database returned 22,590 top matching hits (*e*-value threshold of $1x10^{-20}$) with 6,106 of these covering >80% of the full protein length (Table 4.2; Appendix 26). A search of the vertebrate BUSCO database revealed at least partial hits for 2377 (79%) orthologs, including 1889 (62%) complete orthologs (Table 4.2; Appendix 27).

Functional annotation, gene ontology and differential expression analysis

Transdecoder predicted 187,767 coding regions of at least 100 amino acids (Table 4.2). BLAST searches resulted in 434,450 annotations of 306,989 unique transcripts (68.9%) and 167,648 annotations of 109,348 unique transcripts (24.6%) to the SwissProt, and the smaller Uniref90 sequence databases, respectively. A total of 1,882,664 functional GO terms associated with the BLAST homologies could be assigned to 36,679 Trinity genes (Table 4.2).

Table 4.2: Sequencing, *de novo* assembly and annotation statistics for the *Nannoperca* liver transcriptome.

Sequencing	Total read pairs (2x100 bp)	771,455,524
	Quality trimmed reads	760,191,485
	Retained read pairs	707,735,147
Assembly	Total aligned reads	1,191,758,692
	Trinity transcripts	445,324
	Trinity 'genes'	273,111
	Percent GC	44.15
	Swiss-Prot blast hits >80% of protein length	6,106
	Complete BUSCO conserved orthologs	1889 (62%)
All transcript contigs	N50	1,810
	Median length	445
	Mean length	933.28
	Total assembled bases	415,611,353
Longest isoform	N50	926
	Median length	320
	Mean length	605.17
	Total assembled bases	165,277,448
Annotation	Predicted coding regions >100 amino acids	187,767
	Genes with functional gene ontology terms	36,679

DESeq and *edgeR* results were remarkably similar, with *DESeq* identifying 6900 genes differentially expressed in at least one pairwise population comparison (FDR 5%) compared to 7243 for *edgeR*, with 6358 common to both methods (Appendix 28–30). The slightly more conservative *DESeq* results were retained for downstream analyses. Within populations, expression profiles of DE genes among samples were very similar with all individuals clustering within their population of origin, and clear distinctions among populations (Figure 4.2).

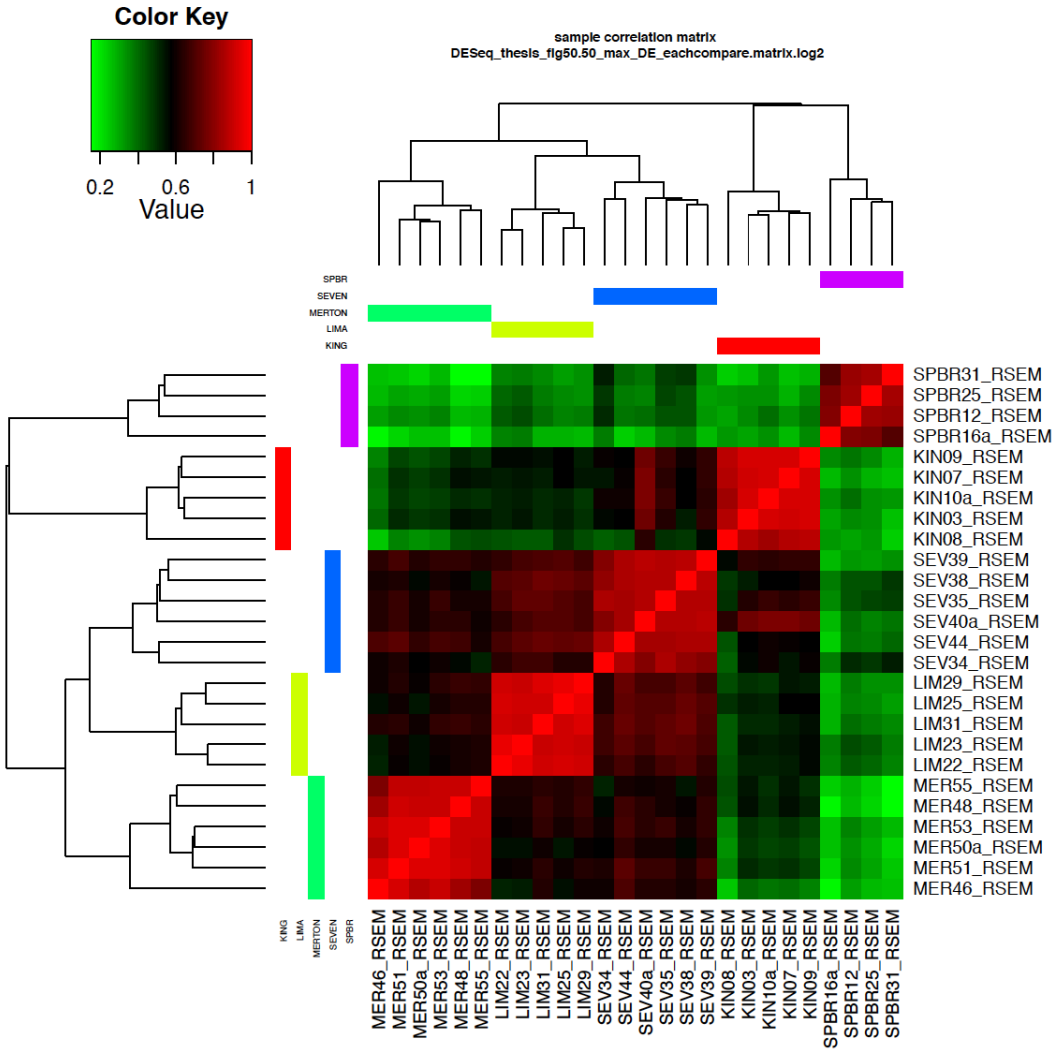


Figure 4.2: Heatmap summarising the correlation in gene expression profiles for *Nannoperca australis* samples based on the top 50 differentially expressed genes identified by *DESeq*.

Expression levels for the top 50 DE genes are contrasted in Figure 4.3 where clear patterns emerge among populations for several clusters of genes. Plots depicting the $\log_2$ fold change in expression versus the $\log_2$ mean expression counts for each pairwise comparison are shown in Appendix 31.

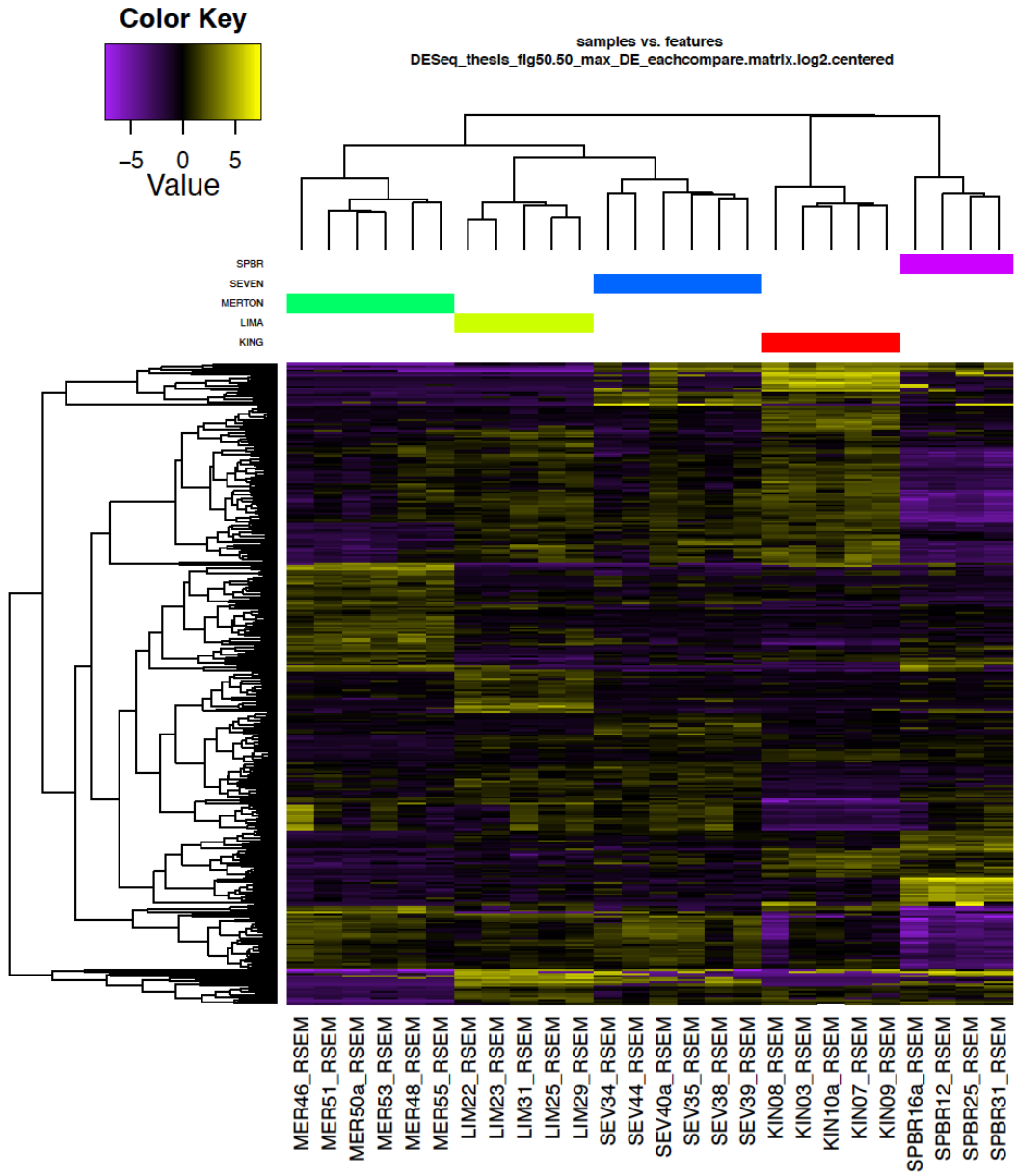


Figure 4.3: Clustered heatmap contrasting $\log_2$ gene expression levels for *Nannoperca australis* samples (columns) based on the top 50 differentially expressed genes (rows) identified by *DESeq*.

Gene expression variance and genetic diversity

The multivariate homogeneity of variances test found no significant differences (P=0.263) in gene expression variance among populations (Table 4.3). One individual from SEV exhibited an extremely different expression profile from individuals within that population (Figure 4.4).

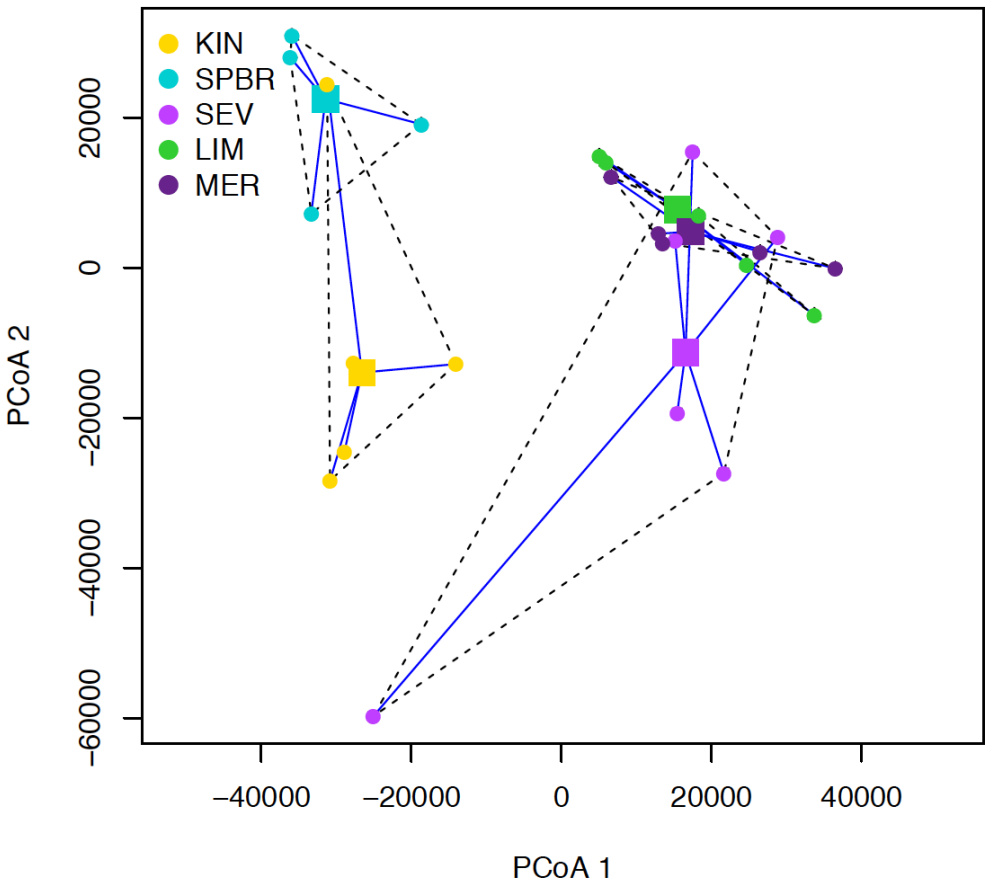


Figure 4.4: Multivariate homogeneity of variance analysis of gene expression profiles for *Nannoperca australis* based on the first two principal coordinate axes summarizing 273,111 *Trinity* 'genes'. Population centroids (squares) depict the mean expression profiles for each population relative to all others. The length of the blue vectors connecting individuals (circles) to the population centroids depicts the relative divergence in expression of each sample from the population mean, and the area within each hull (dashed lines) is proportional to the total within-population variance in gene expression. Colours are based on those used in Figure 4.1.

The cause of this variation is unclear however as there appears no significant difference in either sequencing quality or number of reads, this individual (SEV34; Appendix 23) was retained in the analyses. For each pairwise Tukey's 95% confidence intervals included zero, confirming the

null hypothesis of no significant difference in gene expression variance among any populations (Figure 4.5).

Table 4.3: Multivariate analysis of homogeneity of variance of gene expression based on 273,111 *Trinity* 'genes' for populations of *Nannoperca australis*.

	Df	SumsOfSqs	MeanSqs	F value	Pr(>F)
Groups	4	1194906297	298726574.1	1.421	0.263
Residuals	20	4201884862	210094243.1		

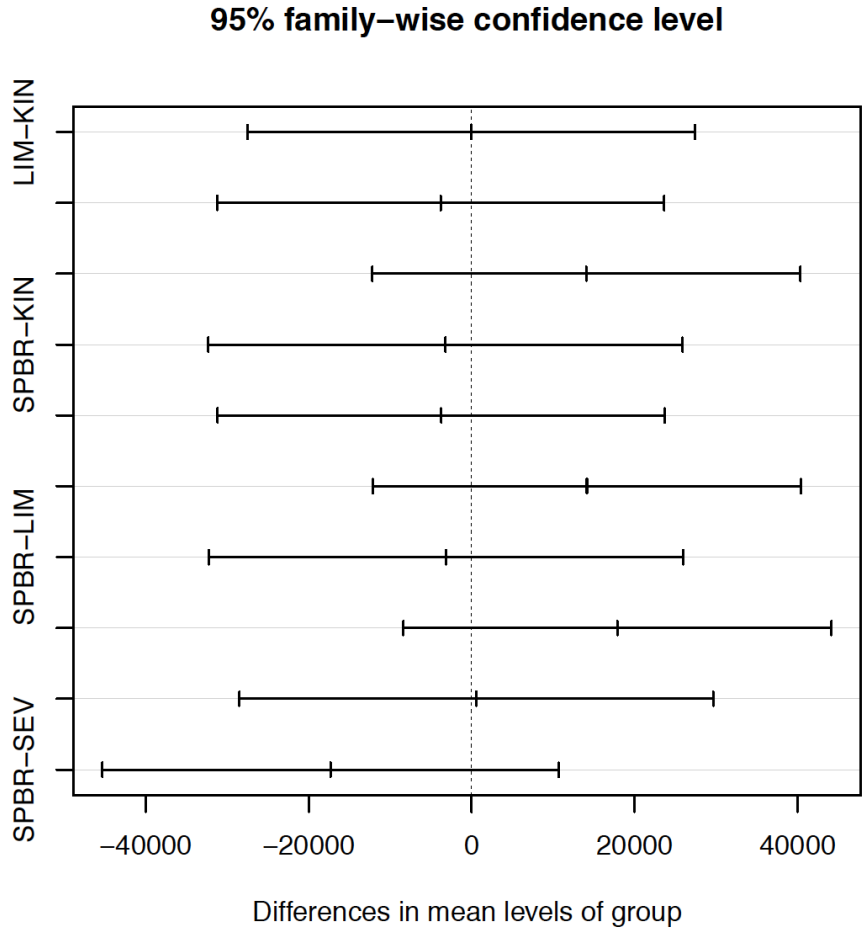


Figure 4.5: Results of Tukey's honest significant differences test for pairwise population differences in mean expression variance for *Nannoperca australis* populations based on the first two principal coordinate axes summarizing 273,111 *Trinity* 'genes'. In all pairwise comparisons, 95% confidence intervals included zero, supporting the null hypothesis of no significant difference in gene expression variance among populations.

The analysis of variance using distance matrices permutation test found no significant relationship between genetic diversity and gene expression. Variation in individual heterozygosity (i.e. the proportion of heterozygous loci per individual) based on 3443 neutral and 216 candidate SNPs is summarised in Figure 4.6A and 4.6B. Variance in population multivariate expression profiles is summarised in Figure 4.6C. Individual heterozygosity was a poor predictor of gene expression variance for both neutral (R^2=0.05, *P*=0.771) (Table 4.4) and candidate loci (R^2=0.09, *P*=0.601) (Table 4.5).

Table 4.4: Multivariate analysis of variance test for association of gene expression variance and genetic diversity (proportion of heterozygous loci at 3443 putitatively neutral SNPs) for *Nannoperca australis*.

	Df	SumsOfSqs	MeanSqs	F.Model	R^2	Pr(>F)
H_E	1	1952875488	1952875488	1.284	0.0529	0.771
Residuals	23	34986420287	1521148708		0.947	
Total	24	36939295775			1	

Table 4.5: Multivariate analysis of variance test for association of gene expression variance and genetic diversity (proportion of heterozygous loci at 216 GEA candidate SNPs) for *Nannoperca australis*.

	Df	SumsOfSqs	MeanSqs	F.Model	R^2	Pr(>F)
H_E	1	3.36E+09	3360355429	2.3017	0.09097	0.601
Residuals	23	3.36E+10	1459953928	0.90903		
Total	24	3.69E+10	1			

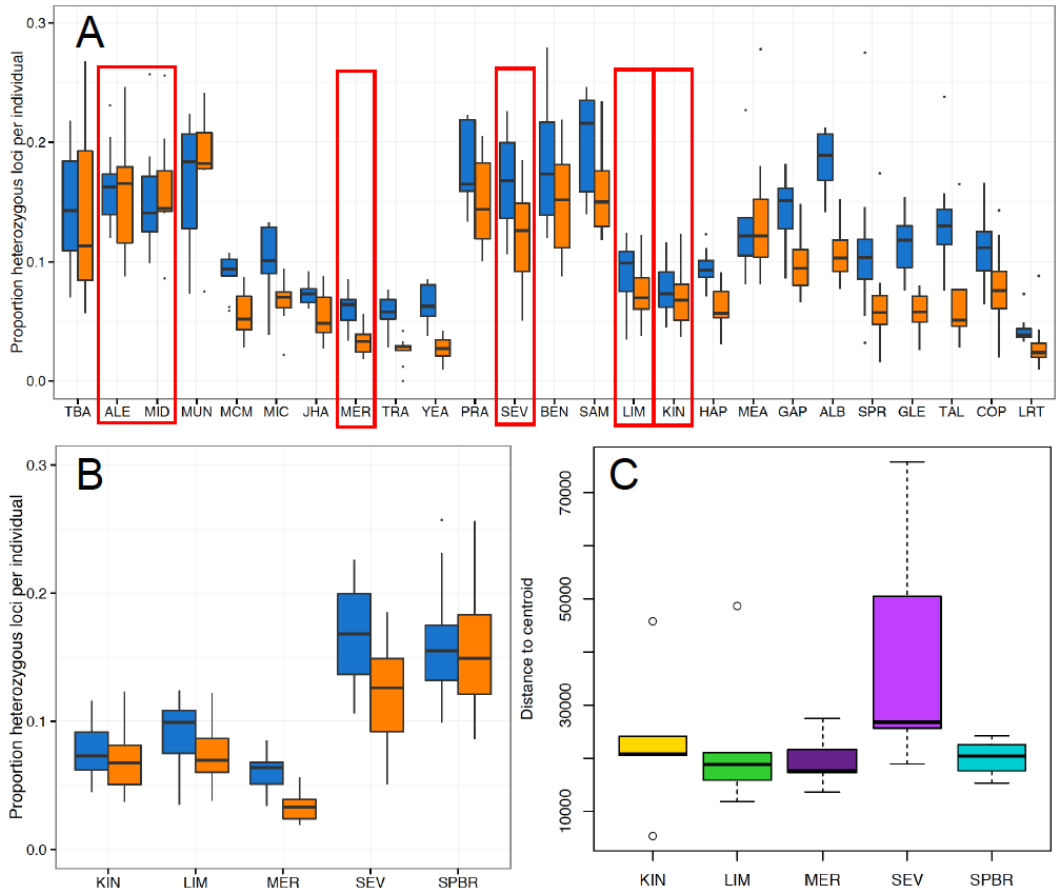


Figure 4.6: Boxplots summarizing population variance for *Nannoperca australis* in individual heterozygosity at 3443 neutral (blue) and 216 GEA candidate (orange) SNP loci for A) all MDB populations, B) populations with RNA data, and C) population variance in gene expression based on the first two principal coordinate axes summarizing 273,111 *Trinity* 'genes'. Populations present in B and C are highlighted in red (SPBR in B and C contains individuals from sites ALE and MID). Colours in C are based on those used in Figure 4.1. There was no significant relationship between gene expression variance and genetic diversity at neutral (R^2=0.05, *P*=0.771), or candidate loci (R^2=0.09, *P*=0.6012).

Gene expression plasticity and adaptation

Bioinformatic processing of the ddRAD sequences produced 30,870 distinct alignments with a total of 384,998 sites, of which 14,997 were variable and 12,244 were parsimony informative. Phylogenetic analysis with *RAxML* supported reciprocal monophyly for all populations, which are

hereafter referred to as lineages (Appendix 32). The *RAxML* majority-rule consensus tree was used as the input phylogeny for the *EVE* analysis (Appendix 33).

Of the 32,914 genes assessed with the *EVE* phylogenetic ANOVA, 189 showed a significant departure from the null hypothesis of a constant expression divergence to diversity ratio. Of these, 24 were identified as candidates for high expression level plasticity, demonstrating significantly (FDR 10%) greater expression variance within- than among-lineages. The hierarchical sample dendogram that clusters individuals based on these genes showed no consistent spatial phylogenetic patterns. This suggests that the expression of these genes is plastic and can vary in response to local variations in environmental conditions (Figure 4.7).

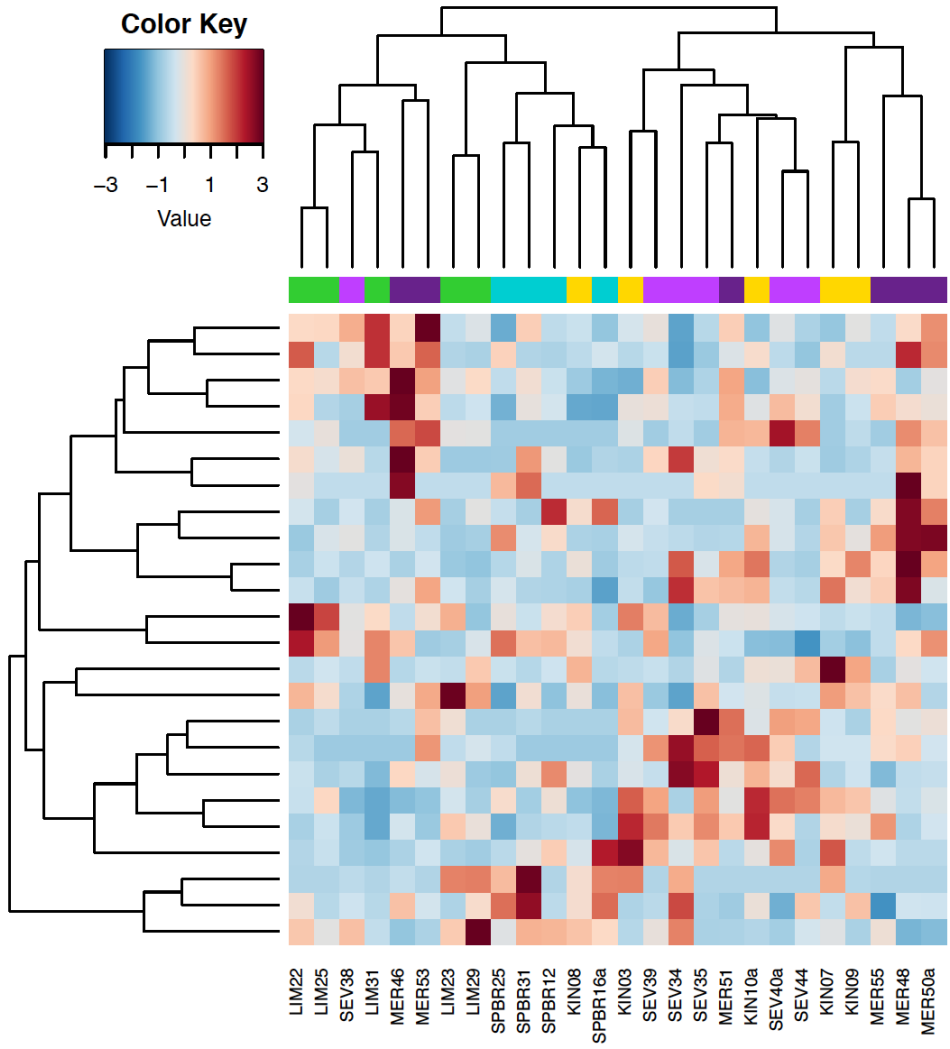


Figure 4.7: Hierarchical cluster of 24 genes identified as candidates (FDR 10%) for expression level plasticity (greater variance within lineages that among lineages after controlling for phylogenetic structure) by the EVE model. Individual samples with similar patterns of expression among genes are grouped together (columns), and genes with similar expression among individuals are grouped together (rows). Coloured bars under the sample dendogram represent the five lineages and are based on those used in Figure 4.1.

Functional annotation enrichment analysis of GO terms for these genes identified 107 significantly enriched terms ($P<0.05$), however none remained significant at a FDR of 10%. Functional categories consisted mainly of terms related to general metabolic activities and cell cycle

regulation, but several involving response to oxidative stress, hormone metabolism and regulation of reproductive processes stand out as key functional terms associated with these genes (Figure 4.8; Appendix 34).

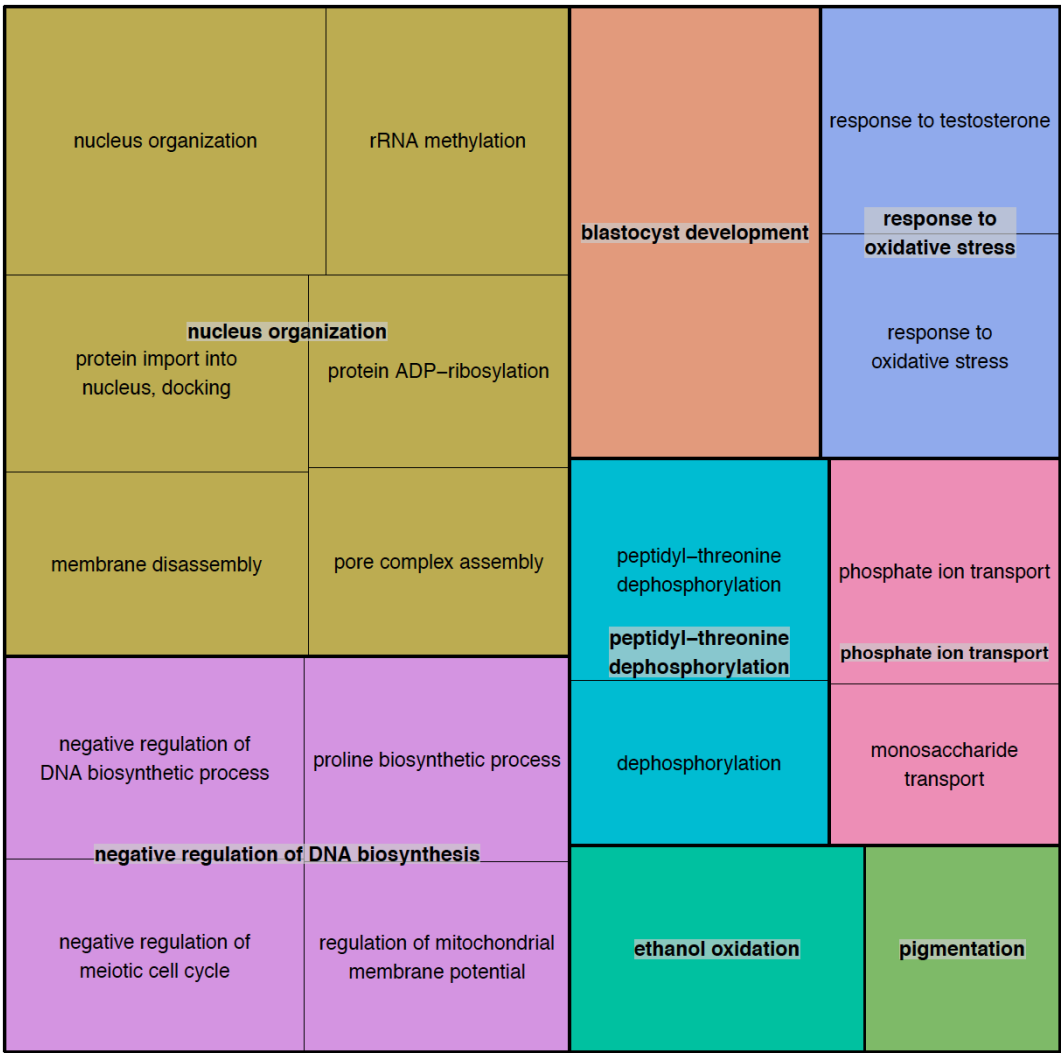


Figure 4.8: Treemap summarising categories of enriched GO terms (*P*<0.05) for 24 genes identified as candidates (FDR 10%) for divergent selection for expression level (greater variance among lineages that within lineages after controlling for phylogenetic structure) by the EVE model. Similar functional GO terms are grouped by colour and size of the boxes is proportional to the level of statistical support for enrichment of the terms relative to GO terms for all genes.

The remaining 165 genes identified with *EVE* showed significantly (FDR 10%) greater expression variance among- than within-lineages. This suggests adaptive evolution of expression level of these genes in response to environmental differences among catchments (Figure 4.9).

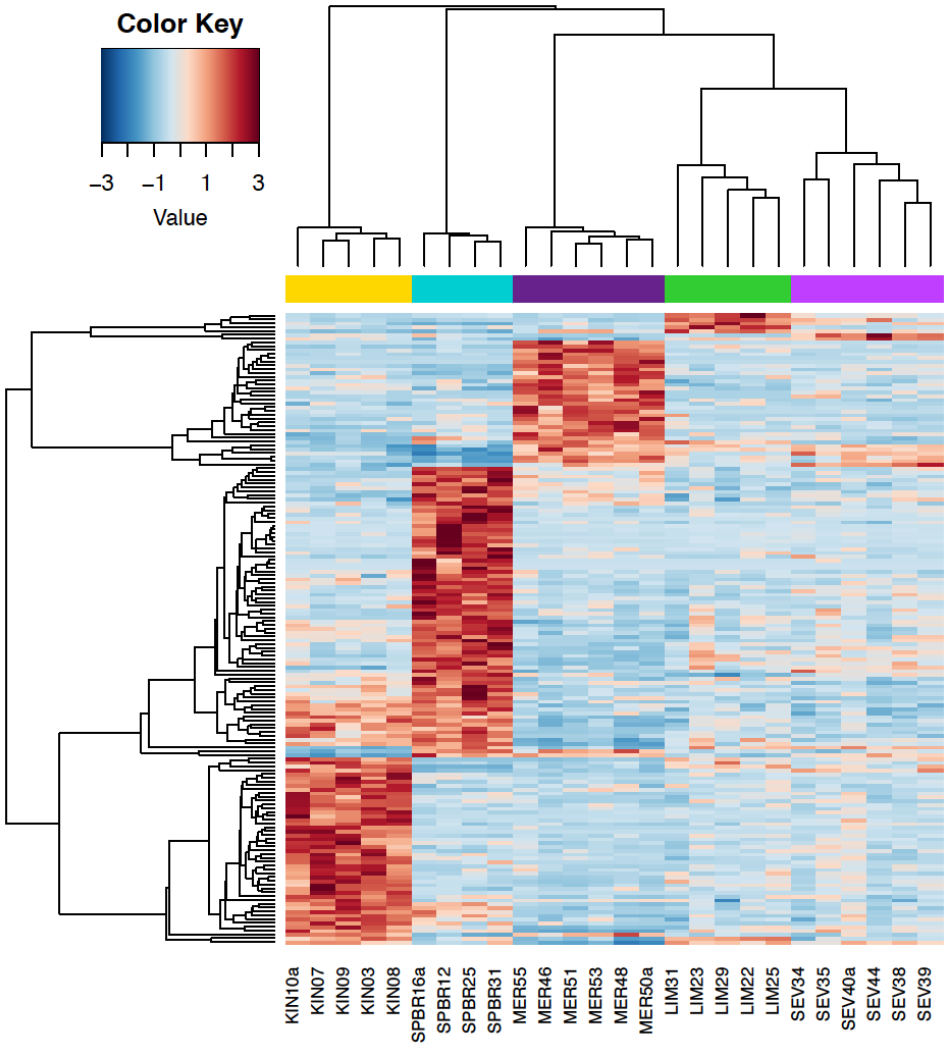


Figure 4.9: Hierarchical cluster of 165 genes identified as candidates (FDR 10%) for divergent selection for expression level (greater variance among lineages that within lineages after controlling for phylogenetic structure) by the EVE model. Individual samples with similar patterns of expression among genes cluster together (columns), and genes with similar expression among individuals are also clustered (rows). Coloured bars under the sample dendogram represent the five lineages and are based on those used in Figure 4.1.

Enrichment analysis of GO terms assigned to this group of genes recovered 347 significantly enriched terms (P<0.05), with 10 remaining significant at a FDR of 10%. These 10 terms fall into two broad categories, response to oxidative stress, and DNA integration (Figure 4.10; Appendix 35).

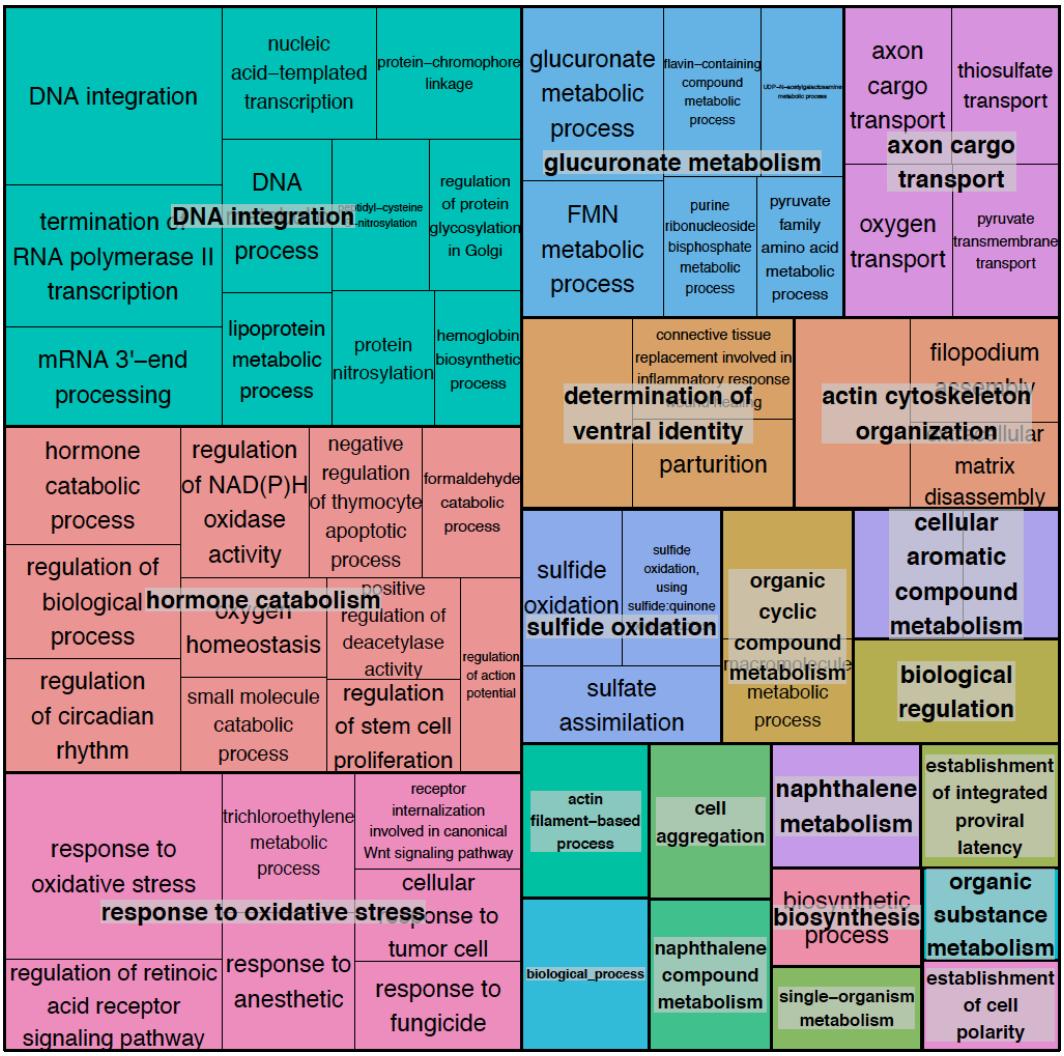


Figure 4.10: Treemap summarising categories of enriched GO terms (P<0.05) for 165 genes identified as candidates (FDR 10%) for divergent selection for expression level (greater variance among lineages that within lineages after controlling for phylogenetic structure) by the EVE model. Similar functional GO terms are grouped by colour and size of the boxes is proportional to the level of statistical support for enrichment of the terms relative to GO terms for all genes.

DNA integration covers a broad group of DNA metabolic processes related to regulation of functions such as recombination and transcription, but also molecular functions initiated by, and in response to viral infections, and others related to protein metabolism (Mommsen 2004). Response to oxidative stress is another broad category including functional terms involved in the processing of reactive oxygen species, derived either endogenously, mainly as a by-product of mitochondrial processes, or exogenously from a range of natural and xenobiotic compounds (Finkel & Holbrook 2000).

The BLAST search of GEA candidate SNP flanking sequences against the transcriptome assembly returned hits for two candidate SNPs to one of the divergently expressed genes (a homologue for human ALPK1), and single hits for another two GEA SNPs to genes exhibiting significant expression level plasticity (IR3IP and BNIP2). The SNPs mapping to ALPK1 were both associated with rainfall variables in the *gINLAnd* analysis and one was also associated with flow and present in the RDA candidates. The SNP mapping to IR3IP was identified in the RDA while the one mapping to BNIP2 was associated with rainfall in the *gINLAnd* analysis.

Discussion

The long-term persistence of populations trapped by habitat fragmentation and threatened by the combination of rapid climate change and widespread habitat degradation likely depends on both adaptive genetic, and phenotypic responses (Chevin *et al.* 2010). The extent to which phenotypic plasticity contributes to evolutionary potential, and the relationship between plastic and evolved responses to rapid environmental change however remains unresolved and is a key research priority (Merilä & Hendry 2014; Alvarez *et al.* 2015). Comparative transcriptomics provides a powerful platform with which to address these issues, as gene expression measurements can be considered as phenotypic traits resulting from a combination of genotype, environment, and genotype–environment interactions (DeBiasse & Kelly 2016). Here we present a *de novo* liver transcriptome for *Nannoperca*, the first transcriptome assembled for any member of Percichthyidae, one of the dominant freshwater fish families in Australia. This represents a valuable resource for ongoing work examining ecology and evolution in both wild and captive populations of the threatened *N. australis* and *N. obscura* (see Chapter 5). No significant relationship between global gene expression variance and genetic diversity was evident for *N. australis*, suggesting that despite strong drift, small and isolated populations retain some capacity for phenotypic plasticity. We also provide insights into patterns of gene expression variation within- and among-populations of *N. australis* sampled across a gradient of hydroclimatic variability. We identify 24 genes as candidates for high expression plasticity and 165 candidates for

divergent selection on expression level. Functional GO analyses identified that many of these 189 candidates are involved in regulatory pathways related to oxidative stress responses and metabolism of a range of natural and xenobiotic compounds. In addition, three of these genes appear homologous with previously identified GEA candidate loci (Brauer *et al.* 2016) thought to be under selection due to hydroclimatic variation. This is suggestive of heritable genotype–environment interactions and provides strong candidates for adaptive phenotypic plasticity in gene expression.

Is variance in gene expression constrained by genetic diversity?

A growing body of evidence indicates that gene expression has a large heritable component (Gibson & Weir 2005; Leder *et al.* 2015; McCairns *et al.* 2016), suggesting that if genetic diversity is lost due to drift, plasticity in gene expression may also be reduced (Bijlsma & Loeschcke 2012). Few studies have examined this issue using wild populations and the relationship between genetic diversity and phenotypic plasticity remains unclear (Chevin *et al.* 2013). For instance, Wood and Fraser (2015) recently addressed this question using a common-garden experiment with populations of brook trout (*Salvelinus fontinalis*) and found little evidence for differences in plasticity for several life history traits in relation to population size. They suggested that increased habitat variability in smaller habitat fragments likely favours higher plasticity.

The populations used in this study span the range of genetic diversity found in the MDB (Figure 4.6) and also include sites at the extreme ends of the hydroclimatic gradient (Figure 2.5). When considered in the context of the naturally variable environment that *N. australis* have evolved in, along with more recent impacts of fragmentation and population size reductions, the finding of this study that gene expression variance is not constrained by genetic diversity supports the hypothesis that selection may maintain plasticity in small populations exposed to variable conditions. This conclusion is supported by empirical evidence revealing spatial patterns of adaptive plasticity in egg size and fecundity for *N. australis* in response to harsh and unpredictable environmental conditions (Morrongiello *et al.* 2012). It is also consistent with theoretical predictions that higher levels of phenotypic plasticity should evolve in marginal and highly variable environments despite reduced population sizes (Chevin & Lande 2011). Furthermore, high levels of plasticity have also been reported in several empirical studies of range-margin populations where genetic diversity is often reduced, and decreased habitat quality and increased habitat variability are common (Nilsson-Örtman *et al.* 2012; Valladares *et al.* 2014; Lázaro-Nogal *et al.* 2015). Phenotypic responses to rapidly changing and poor quality environments can facilitate population persistence (Whitehead *et al.* 2010), and our results add to an emerging trend suggesting that small populations may maintain some capacity for phenotypic plasticity despite reduced levels of genetic variation.

Comparative transcriptomics, evolution and gene expression

Transcriptomic responses to environmental stressors are well documented in fishes (Oleksiak 2008; Whitehead *et al.* 2010; Bozinovic & Oleksiak 2011; Pujolar *et al.* 2012; Smith *et al.* 2013; Baillon *et al.* 2015; Leder *et al.* 2015). For species evolving in variable and naturally harsh environments, the ability to respond rapidly to often-abrupt changes in water quality should provide a distinct evolutionary advantage. Accordingly, several recent studies have provided evidence suggesting that natural selection influences patterns of gene expression variation. For example, killifish (*Fundulus heteroclitus*) inhabit highly variable tidal marshes and are well known for their ability to tolerate extreme conditions and rapid changes in water quality such as variation in pH, temperature, salinity and dissolved oxygen (Burnett *et al.* 2007). Experimental work revealed that complex patterns of gene expression and genetic variation in killifish are underpinned by locally adapted transcriptomic responses to osmotic shock (Whitehead *et al.* 2010). Similarly, Leder *et al*. (2015) found substantial genetic variance in gene expression among populations of threespine stickleback (*Gasterosteus aculeatus*) for genes associated with temperature stress. Heritable patterns of gene expression have also been documented for an Australian rainbowfish (*Melanotaenia duboulayi*) at candidate genes for thermal adaptation (McCairns *et al.* 2016). In that study additive genetic variance and transcriptional plasticity explained variation in gene expression associated with long-term exposure to a future climate, providing pedigree-based support that transcriptional variation has an underlying heritable basis. These studies suggest that

gene expression can evolve in response to natural selection, and that both genomic and transcriptomic variation contribute to species' evolutionary potential.

Results in the present study raise the possibility that the divergent expression profiles of 165 candidate genes represent adaptive shifts in expression levels among populations of *N. australis*. An alternative interpretation, and one that cannot be presently ruled out, is that these genes are responding plastically to local and transient environmental challenges present at the time of sampling. This may be particularly plausible considering a large number of these genes are putatively involved in responses to xenobiotic toxins and, more broadly, oxidative stress; both of which may be induced as a response to variations in water quality over relatively short time scales (Whitehead *et al.* 2011). While historical water quality data exist for some locations in the MDB, the records are inconsistent in both spatial and temporal scale, and lack of specific information on chemical pollutants limits the potential for these data to provide insight.

Another possibility is that genetic drift is responsible for the patterns of gene expression divergence. Although the *EVE* model used here does account for phylogenetic divergence, it assumes one consistent phylogeny for all genes (Rohlfs & Nielsen 2015). While this could be a particular issue for intraspecific studies, for *N. australis* the phylogeny is well

resolved and is unlikely to bias the results (Appendix 32). Additionally, *EVE* does not control for expression covariance among genes. This assumption of independence would likely be violated in most cases and improving the method to account for complex correlation structures in the data would prove beneficial. Regardless of these limitations the divergently expressed candidate genes identified here support the hypothesis that evolutionary adaptation is driving patterns of gene expression for *N. australis*. This provides a basis for further work and testing this hypothesis within a common-garden experimental framework (e.g. Whitehead *et al.* 2011; Smith *et al.* 2013; McCairns *et al.* 2016) represents a next step in assessing the potential role of natural selection in shaping gene expression in this system.

Functional analysis and environmental stressors

Functional annotations based on distantly related species should be interpreted with caution as the extent to which gene functions are conserved among divergent taxa remains largely unknown (Primmer *et al.* 2013). A general assessment of putative functional categories characterising candidate genes in an ecological context can however still provide useful information and hypotheses regarding important environmental factors influencing patterns of gene expression (Pavey *et al.* 2012). Several of the most significantly enriched GO terms for the divergently expressed candidate genes belong to a group of aspartic-type endopeptidase and peptidase enzymes involved in protein digestion (Appendix 35). This class of enzyme is known to be important in other

fishes for muscle proteolysis associated with physiological challenges such as starvation, migration or reproductive activity (Mommsen 2004; Wang *et al.* 2007), and is likely to play an important role in survival in variable environments. Many of the other divergent expression candidates are associated with metabolism of organic, and synthetic compounds and with response to oxidative stress. Challenging environmental conditions such as thermal stress or exposure to pollution can induce oxidative stress (Hermes-Lima & Zenteno-Savín 2002), and heritable variation in expression of genes associated with oxidative stress was identified in *M. duboulayi* (McCairns *et al.* 2016). Glucuronate metabolism is related to oxidative stress response and was enriched in the divergent expression candidates, with these genes playing a role in the de-activation and excretion of a broad range of toxins (Meech & Mackenzie 1997). These can include endogenous cellular compounds as well as toxic xenobiotics such as plant-based flavonoids and tannins (Buckley & Klaassen 2007). Tannins and polyphenols leaching from *Eucalyptus* leaves are naturally present in waters inhabited by *N. australis* and are known to impact reproductive success, affect juvenile growth and survival and drive variation in male nuptial colouration in this species (Morrongiello *et al.* 2010; Morrongiello *et al.* 2011b, 2013). The magnitude of response to *Eucalyptus* leachate exposure also varies for *N. australis* populations across a natural gradient of water quality (Morrongiello *et al.* 2013). The latter suggests populations are adapted to local variations in water quality, which is consistent with our findings that genes apparently involved in metabolising flavonoids and tannins are divergently expressed among

populations.

Many of the candidate genes related to oxidative stress responses are known to be important in metabolism of industrial chemicals such as pesticides and petroleum products. Organochlorine pesticides were used heavily throughout the MDB during the mid-to-late 1900s, and residues remaining in sediments today are known to increase concentrations in waterways after heavy rainfall events (McKenzie-Smith *et al.* 1994). These chemicals have been linked to invertebrate larval deformities across the MDB (Pettigrove 1989), and are known to cause oxidative stress in fish (Slaninova *et al.* 2009). Naphthalene is a water-soluble by-product of oil and gas production and is also a constituent of some pesticides (Gavin *et al.* 1996). This compound is toxic to fish and has been shown to induce developmental abnormalities and affect reproduction in another MDB fish *Melanotaenia fluviatilis* (Pollino & Holdway 2002; Pollino *et al.* 2009). Estrogens and other endocrine disrupting chemicals are recognised as a global issue for freshwater fishes and are known to adversely affect native fish reproduction in the MDB (Vajda *et al.* 2015). Low concentrations of these substances were also implicated in the feminization of males in a population of fathead minnows (*Pimephales promelas*) in Canada, leading to the eventual collapse of the population (Kidd *et al.* 2007). Several genes identified here are involved in cellular responses and metabolism of steroidal estrogens, suggesting populations may be exposed to these pollutants. While it is possible these genes are up regulated due to natural

reproductive cycles, several GO terms related to parturition and blastocyst development were also enriched. These processes are associated with female reproduction and, given the samples in this study consist only of males it is likely that environmental estrogens are impacting reproductive health of *N. australis*, and probably other MDB fishes. Sulfide compound transport and metabolism is another group of important functional terms enriched in the divergently expressed genes. Sulfide compounds in freshwaters can originate from natural decomposition of organic matter, and also industrial and urban pollution such as sewage wastewater. Sulfides are toxic to fish and can reduce juvenile survival (Smith & Oseid 1972). Smith and Oseid (1972) also found the effects of sulfides were exacerbated by increased temperatures and reduced dissolved oxygen levels, suggesting these toxins may become more virulent under a changed climate regime. This also points to the broader implications of widespread habitat degradation concomitant with rapid climatic changes, and supports recent evidence that the compounding effects of climate change and pollution pose additional extinction risks for many threatened species (Brown *et al.* 2015).

Conclusions

Proactive conservation management strategies need to incorporate assessments and predictions of species evolutionary potential. Plastic and evolutionary components of gene expression can now be explored for non-model species using RNA-seq and, as new analytical approaches

evolve, we stand to gain insight into how gene expression variation in the wild contributes to evolutionary potential of populations. Although the work presented here represents an initial step towards that understanding, it is particularly important given that it pioneers this type of research in poorly studied but highly diverse southern hemisphere fishes. The first transcriptome assembly for any percichthyid, generated as part of this thesis, provides a valuable resource for ongoing work with this study system, both in the context of conservation and for addressing broader questions related to the ecology and evolution of freshwater biodiversity. The comparative transcriptomic analyses identified 165 genes as candidates for divergent selection on expression level, 24 candidates for high expression plasticity, and support mounting evidence that phenotypic plasticity may not be constrained by population size. In addition, functional GO analyses of the candidate genes suggests water quality is driving patterns of divergent and plastic gene expression for genes related to metabolic and reproductive traits for populations of *N. australis*, and likely for other MDB fishes.

Chapter 5: Conservation genomics and evolutionary potential of a threatened Australian freshwater fish

General discussion

As natural habitat is increasingly degraded, fragmented and converted for human use, rapid climate change is also challenging many species phenological and physiological limits. Our ability to predict species capacity to respond to environmental change and assess the likelihood of long-term persistence has become a key research focus. Populations can typically respond to environmental challenges via genetic adaptation, dispersal to more suitable habitat, or phenotypic plasticity (Pauls *et al.* 2013). The central goal of this thesis was to examine how these three mechanisms influence evolutionary potential for an Australian temperate freshwater fish threatened by a cocktail of rapid environmental changes.

Nannoperca australis has suffered recent and severe demographic decline in the Murray–Darling Basin and many populations included in this study are now thought to be locally extirpated. In response to this widespread decline and to further losses resulting from catastrophic loss of habitat during recent prolonged drought, several captive breeding programs were initiated and additional translocations among populations of wild fish have been proposed. In this context, the work presented here provides an empirical example of applied ecological genomics and comparative transcriptomics that advances our understanding of the genetic, demographic and environmental aspects of evolutionary resilience of freshwater biodiversity. Critically, it also offers a timely and in-depth first genomic assessment of evolutionary potential for an

ecologically important endemic freshwater fish on the brink of extinction. The findings can directly contribute to ongoing conservation efforts for this species and provide a tangible example of the integration of genomic approaches with applied conservation. Here, I reiterate the major discoveries from each data chapter, highlighting their contribution to our understanding of the genetic, demographic and environmental determinants of species evolutionary potential. I also discuss implications for conservation of *N. australis* in the MDB, and then outline future research opportunities arising from this thesis.

Chapter 2: Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin

For genetic evolutionary adaptation to keep pace with the rapid rate of climate change and other anthropogenic threats, it is likely that selection will need to act mainly on standing genetic variation (Barrett & Schluter 2008). However, it is still unclear if threatened species may retain enough adaptive genetic variation to facilitate an evolutionary response. Chapter two of this thesis examined this question using replicate populations of *N. australis* sampled across a steep hydroclimatic gradient. Genome-wide SNP genotype data was integrated with high-resolution environmental layers in the first riverscape genomics study of an Australian fish, and likely, of a southern hemisphere fish. Neutral population structure across the MDB was consistent with expectations based on spatial stream hierarchy and life history. In contrast, variables related to precipitation and temperature were identified as the most important environmental

surrogates for putatively adaptive genetic variation at both regional and local scales. Human disturbance also influenced variation in candidate loci, but the lack of any regional spatial pattern suggested allele frequencies at these loci might be responding to local scale variation in habitat disturbance. This suggests that although drift is clearly a dominant evolutionary process in this system, it's effects have probably only recently intensified and the signal of local adaptation has not been completely eroded. Across the MDB, *N. australis* spans a range of ecologically diverse, and hydroclimatically variable environments, and results here suggest this heterogeneous selection pressure is also driving local adaptive divergence of populations. We proposed a combination of within-catchment genetic rescue, and translocations among adaptively divergent populations to increase evolutionary potential and adaptive resilience of the species. This represents a paradigm shift away from the long-held view in conservation biology that local–is–best. However, where pervasive threats such as climate change and wide-scale habitat loss affect recently fragmented and isolated populations, adaptive differences between populations should be seen as an opportunity to increase adaptive resilience, rather than as a barrier to proactive management options (sensu Weeks *et al.* 2016). For instance, captive insurance populations of lower Murray (Hammer *et al.* 2013), and also several headwater (Coppabella and Lachlan catchments; Figure 2.1) (Pearce 2015) *N. australis* are currently maintained in semi-wild conditions in dams. Evidence presented in this thesis suggests that lower Murray populations may be locally adapted to hotter and drier conditions, such as are

predicted for the rest of the MDB in the near future. This represents an ideal opportunity to introduce new, and potentially adaptively important alleles to the headwater populations, while being able to closely monitor them for signs of outbreeding depression.

The biogeographic scenario of recent and rapid range-reduction, severe habitat fragmentation and demographic decline of *N. australis* mirrors predictions for many species now threatened by climate change and human alteration of the environment. This provided a unique and useful model to test and extend the utility of landscape genomics methods for systems with complex spatial structure and other confounding demographic effects common to threatened and poor dispersive species. Through the use of spatially explicit GEA methods, we were able to disentangle the signal of adaptive variation responding to hydroclimatic variation from the effects of landscape structure and shared population history. The distribution of F_{ST} values for the GEA candidate loci support recent views that adaptation of complex quantitative traits predominantly arises through polygenic selection, highlighting benefits of incorporating GEA methods into conservation studies of adaptation.

Chapter 3: Anthropogenic habitat fragmentation increases extinction risk for freshwater species

Habitat fragmentation is a key process implicated in the current and unprecedented worldwide decline of freshwater biodiversity (Sala *et al.* 2000; Fischer & Lindenmayer 2007). Linking recent human activities to the decline of wild populations is however challenging, as natural historic spatial population structure and demographic stochasticity can confound the effects of more recent disturbance, particularly for species with limited dispersal abilities (Landguth *et al.* 2010). In chapter three, population genomic analyses and individual-based population genetic simulations were employed to examine how recent and extensive construction of in-stream barriers across the MDB threaten the long-term persistence of *N. australis*. The findings provide for the first time clear evidence that recent anthropogenic habitat fragmentation has contributed to the decline of freshwater biodiversity across a riverine ecosystem. The populations most isolated by habitat fragmentation showed reduced genetic diversity and increased population differentiation, and this signal remained strong after accounting for the effects of natural stream hierarchy and environmental variation. Earlier studies investigating the consequences of recent habitat fragmentation have relied on low-resolution markers, such as mtDNA or allozymes, and have often failed to detect any genetic signal (Cunningham & Moritz 1998; Driscoll & Hardy 2005). The high-resolution SNP data used in this study provided a powerful framework with which to quantify the effects of very recent fragmentation on genetic diversity and population structure. Evidence for recent bottlenecks, and results of the simulations provided crucial support for our assertion that recent anthropogenic

habitat fragmentation is driving population genetic divergence and demographic decline of this species. The very low N_e estimates for most remnant populations add additional support when considered along with coalescent-based estimates of much larger, and more connected populations in the past (Attard *et al.* 2016b; Cole *et al.* 2016). Together, results from chapter two and evidence presented in chapter three suggest that not only drift, but also selection is influencing the evolution of populations, and that adaptive divergence of small populations can occur quickly following fragmentation. This has broad implications for the conservation of freshwater biodiversity, as the long-term persistence of many species (perhaps even the majority?) is likely also similarly threatened by anthropogenic habitat fragmentation.

Chapter 4: Comparative ecological transcriptomics and the contribution of gene expression to the evolutionary potential of a threatened freshwater fish

Phenotypic plasticity can provide a rapid response to environmental change and is another mechanism that may facilitate population persistence in altered environments and potentially lead to evolutionary adaptation (Chevin *et al.* 2010; Dayan *et al.* 2015; Ghalambor *et al.* 2015). Chapter four of this thesis used RNA-seq to quantify patterns of gene expression within, and among selected populations spanning a gradient of environmental variability. The results provide insight into how plastic and evolutionary components of gene expression interact and combine to produce an adaptive evolutionary response in wild populations. In spite of strong drift, the results that small and isolated populations retain some

phenotypic variation suggests that phenotypic plasticity can, to some extent, buffer *N. australis* populations against rapid environmental change. This conclusion is supported by empirical evidence for plasticity in reproductive traits in response to naturally harsh and unpredictable environmental conditions for this species (Morrongiello *et al.* 2012). This is also consistent with theoretical predictions that even for small populations, phenotypic plasticity should evolve in variable environments (Chevin & Lande 2011). Although the functional annotation performed here should be interpreted cautiously, the general pattern emerging from the comparative transcriptomics analyses indicate that both natural and anthropogenic variation in water quality is driving patterns of gene expression for populations of *N. australis*. These findings highlight the vulnerability of aquatic biodiversity to the combined effects of anthropogenic disturbance and rapid climate change, and suggest that compounding effects of climate change and pollution likely pose additional extinction risks for many threatened species.

Future directions and concluding remarks

Predicting species evolutionary potential and understanding the mechanisms that promote population persistence in the face of rapid and widespread environmental change is necessary to identify priorities for conservation management. Although the work presented here represents only an initial step towards that understanding, it is particularly significant given that it pioneers this type of research in poorly studied but highly

diverse southern hemisphere fishes. The lack of genomic resources previously available for this species necessarily limited the scope of some aspects of this study, and as such several findings represent hypotheses in need of functional verification through common-garden type experiments. The catalogue of SNP loci, including 216 GEA candidates, the *de novo* transcriptome assembly and candidate genes from the comparative transcriptomics thus all provide valuable resources and represent a major contribution of this thesis. These resources will be vital for ongoing work with this study system, both in the context of conservation and for addressing broader questions related to the ecology and evolution of freshwater biodiversity.

One promising avenue of research related to the work presented here concerns examining the genomic consequences of captive breeding. A conservation-breeding program for *N. australis* (and two other species) was initiated at Flinders University following catastrophic loss of habitat in the Lower Murray during recent drought (Hammer *et al.* 2013). Microsatellite data were used to estimate relatedness among the founding individuals and 11 family groups were selected on the basis of minimising inbreeding (Attard *et al.* 2016b). Offspring from this program (i.e. from the 11 family groups) were subsequently released to restored habitat in the Lower Murray. Monitoring of the reintroduced population over several years has resulted in the capture of 71 *N. australis*. Subsequent microsatellite analysis confirmed 19 were released fish, originating from

the breeding program, and 52 were wild-born offspring from the reintroduced population (Attard *et al.* 2016b). Preliminary analyses using 210 of the 216 GEA candidate SNPs (six loci were monomorphic for these samples and were omitted) has been completed by the PhD candidate, but only a subset of three of the original family groups has been sequenced to date. Discriminant analysis of principal components (DAPC) results provide evidence for rapid changes in allele frequencies at candidate adaptive loci but, importantly, suggest that once reintroduced to the wild, the population has responded to selection (Figure 5.1).

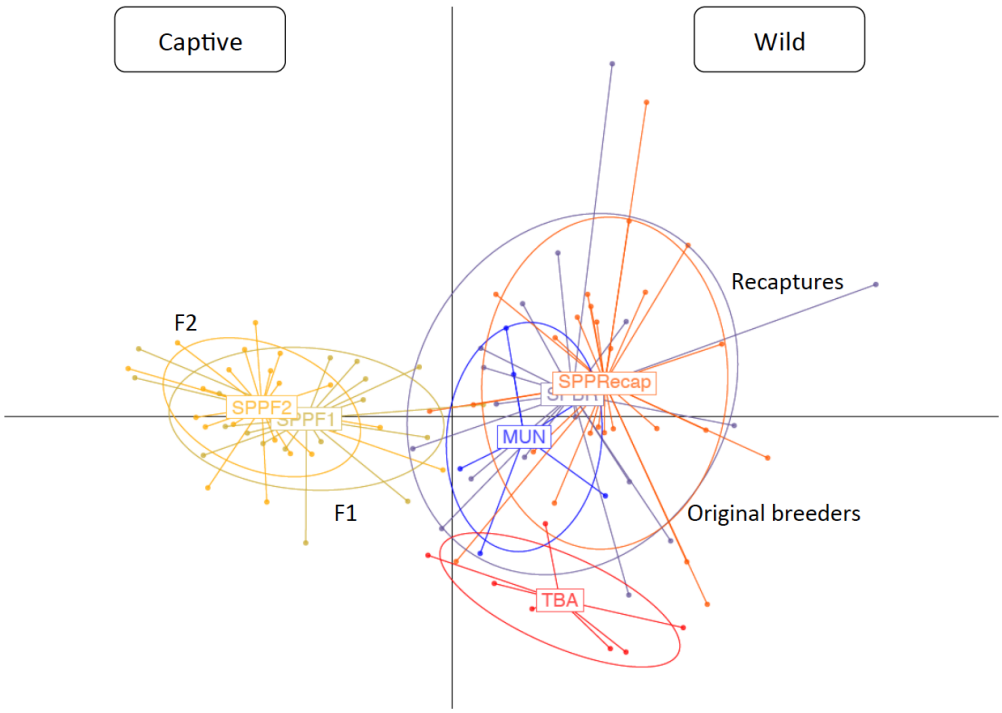


Figure 5.1: Discriminant analysis of principal components plot based on 210 of 216 previously identified GEA candidate SNPs. The wild caught founding population (i.e. the 'original breeders'; 18 samples included here) and individuals sampled during monitoring of the re-introduced population (i.e. the 'recaptures'; 26 samples) cluster with other wild populations sampled in the Lower Murray before the drought. Captive F_1 and F_2 generations (20 and 24 samples) based on three of the original brood groups form a second and clearly distinguishable group.

Although these results are very preliminary, it seems that by limiting the number of generations in captivity, not only the loss of diversity but also the adaptation to captivity can be minimised. Intriguingly, these findings suggest that natural selection can rapidly promote persistence of reintroduced populations in the wild. Incorporating transcriptomics into this study promises to also add to our understanding of the combined genomic and transcriptomic aspects of adaptation to captivity and threatened species evolutionary potential.

Another beneficial extension of this work, and possibly of other ecological genomics studies, would be to jointly analyse genomic and transcriptomic data with physical phenotypic traits. There is clear evidence that body shape in fishes is a key phenotypic trait subject to selection related to hydrodynamic variation (Webb 1982; McGuigan *et al.* 2003; McGuigan *et al.* 2005; McGuigan *et al.* 2011). Morphometric data have been integrated with landscape genomic analyses for another Australian fish within our lab, with results identifying several candidate loci strongly associated with both hydroclimatic variation and body shape (Smith *et al. in preparation*). Museum specimens are available for most populations included in this thesis, presenting an exciting opportunity to incorporate morphometric analyses with the genomic and transcriptomic data following a similar framework.

The integration of results across the three data chapters of this thesis provides a novel framework for understanding how environmental variation and human disturbance have influenced genetic diversity, population connectivity and adaptive potential of wild populations of a threatened aquatic species. The overwhelmingly consistent pattern emerging from the work presented here is that evolutionary potential can be maintained in small populations. The recent disruption of historic meta-population dynamics and demographic decline of populations has resulted in the widespread loss of genetic variation. Despite this, altered selection regimes resulting from increased environmental variability within fragmented habitat patches may have favoured the maintenance of adaptive genetic variation, and the capacity for gene expression plasticity. Concomitantly, environmental heterogeneity among isolated populations may be driving local adaptive divergence. For *N. australis* continued survival in the MDB however, bold and proactive conservation measures are urgently required to mitigate the increased extinction risk posed by recent widespread and severe habitat fragmentation.

Appendices

Appendix 1: Companion publications contributed to during candidature

Conservation Biology

Contributed Paper

A novel holistic framework for genetic-based captive-breeding and reintroduction programs

C.R.M. Attard,* L.M. Möller,* M. Sasaki,* M.P. Hammer,† C.M. Bice,‡ C.J. Brauer,* D.C. Carvalho,*§ J.O. Harris,* and L.B. Beheregaray* ¶

*School of Biological Sciences, Flinders University, Adelaide, SA 5042, Australia
†Museum and Art Gallery of the Northern Territory, P.O. Box 4646, Darwin, NT 0801, Australia
‡Inland Waters and Catchment Ecology Program, SARDI Aquatic Sciences, P.O. Box 120, Henley Beach, SA 5022, Australia
§Pontificia Universidade Catolica de Minas Gerais, Belo Horizonte, MG 30535-610, Brazil

Abstract: *Research in reintroduction biology has provided a greater understanding of the often limited success of species reintroductions and highlighted the need for scientifically rigorous approaches in reintroduction programs. We examined the recent genetic-based captive-breeding and reintroduction literature to showcase the underuse of the genetic data gathered. We devised a framework that takes full advantage of the genetic data through assessment of the genetic makeup of populations before (past component of the framework), during (present component), and after (future component) captive-breeding and reintroduction events to understand their conservation potential and maximize their success. We empirically applied our framework to two small fishes: Yarra pygmy perch* (Nannoperca obscura) *and southern pygmy perch* (Nannoperca australis)*. Each of these species has a locally adapted and geographically isolated lineage that is endemic to the highly threatened lower Murray–Darling Basin in Australia. These two populations were rescued during Australia's recent decade-long Millennium Drought, when their persistence became entirely dependent on captive-breeding and subsequent reintroduction efforts. Using historical demographic analyses, we found differences and similarities between the species in the genetic impacts of past natural and anthropogenic events that occurred in situ, such as European settlement (past component). Subsequently, successful maintenance of genetic diversity in captivity—despite skewed brooder contribution to offspring—was achieved through carefully managed genetic-based breeding (present component). Finally, genetic monitoring revealed the survival and recruitment of released captive-bred offspring in the wild (future component). Our holistic framework often requires no additional data collection to that typically gathered in genetic-based breeding programs, is applicable to a wide range of species, advances the genetic considerations of reintroduction programs, and is expected to improve with the use of next-generation sequencing technology.*

Keywords: conservation genetics, extinction, fish, genetic diversity, management, restoration genomics

Un Marco de Referencia Holístico Novedoso para Programas de Reproducción en Cautiverio Basada en Genética y de Reintroducción

Resumen: *Investigaciones sobre biología de la reintroducción han proporcionado un mejor entendimiento del, a menudo, éxito limitado de las reintroducciones de especies y han resaltado la necesidad de aproximaciones rigorosas científicamente en los programas de reintroducción. Examinamos la literatura reciente sobre reproducción en cautiverio basada en genética y reintroducción para exhibir la subutilización de los datos genéticos. Diseñamos un marco de referencia que obtiene la mayor ventaja de los datos genéticos mediante la evaluación de la composición genética de las poblaciones antes (componente pasado del marco de referencia), durante (componente presente), y después (componente futuro) de eventos de reproducción en cautiverio y de reintroducción para entender su potencial de conservación y maximizar su éxito. Aplicamos nuestro marco de referencia empíricamente con dos especies de peces pequeños:* Nannoperca obscura

¶*email luciano.beheregaray@flinders.edu.au*
Paper submitted January 19, 2015; revised manuscript accepted January 6, 2016.

Conservation Biology, Volume 30, No. 5, 1060–1069

DOI: 10.1111/cobi.12699

Conserv Genet
DOI 10.1007/s10592-016-0868-8

RESEARCH ARTICLE

Range-wide fragmentation in a threatened fish associated with post-European settlement modification in the Murray–Darling Basin, Australia

Theresa L. Cole[1,7] · Michael P. Hammer[2,3] · Peter J. Unmack[4] · Peter R. Teske[1,5] · Chris J. Brauer[1] · Mark Adams[2,6] · Luciano B. Beheregaray[1]

Received: 22 January 2016 / Accepted: 12 July 2016

Abstract Distinguishing the relative influence of historic (i.e. natural) versus anthropogenic factors in metapopulation structure is an important but often overlooked step in management programs of threatened species. Biotas in freshwater wetlands and floodplains, such as those in the Murray–Darling Basin (MDB)—one of Australia's most impacted ecosystems, are particularly susceptible to anthropogenic fragmentation. Here we present a comprehensive multilocus assessment of genetic variation in the threatened southern pygmy perch *Nannoperca australis* (578 individuals; 45 localities; microsatellite, allozyme and mitochondrial DNA datasets), an ecological specialist with low dispersal potential. We assess patterns of spatial structure and genetic diversity in populations spanning the highly fragmented MDB and test whether recent anthropogenic modification has disrupted range-wide connectivity. We detected strong and hierarchical population structure, very low genetic diversity and lack of contemporary gene flow across the MDB. In contrast, the apparent absence of pronounced or long-term phylogeographic structure suggests that observed population divergences generally do not reflect deeply historic natural fragmentation. Coalescent-based analyses supported this inference, revealing that divergence times between populations from the upper and lower MDB fall into the period of European settlement. It appears that the observed contemporary isolation of populations is partly explained by the severe modification of the MDB post-dating the onset of European settlement. Our integrated approach substantially improves the interpretation of how fragmentation impacts present-day biodiversity. It also provides novel contributions for risk-assessing management actions in the context of captive breeding and translocations of small freshwater fishes, a group of increasing global conservation concern.

Keywords Conservation genetics · Connectivity · Translocations · Ecological genetics · Climate change

Electronic supplementary material The online version of this article (doi:10.1007/s10592-016-0868-8) contains supplementary material, which is available to authorized users.

✉ Luciano B. Beheregaray
Luciano.Beheregaray@flinders.edu.au

[1] Molecular Ecology Laboratory, School of Biological Sciences, Flinders University, Bedford Park, SA 5042, Australia

[2] Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, SA 5000, Australia

[3] Museum and Art Gallery of the Northern Territory, PO Box 4646, Darwin, NT 0801, Australia

[4] Institute for Applied Ecology and Collaborative Research Network for Murray–Darling Basin Futures, University of Canberra, Bruce, ACT 2601, Australia

[5] Department of Zoology, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa

[6] Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

[7] Department of Zoology, Otago University, PO Box 56, Dunedin, New Zealand

Introduction

Natural landscape fragmentation plays an important role in shaping metapopulation and ecosystem boundaries through its influence on many micro-evolutionary processes, including migration, genetic drift and inbreeding (Lande 1988). The dispersal of individuals between habitat patches increases the probability of species persistence by reducing

Published online: 20 July 2016

Conserv Genet
DOI 10.1007/s10592-016-0873-y

SHORT COMMUNICATION

Multi-generational evaluation of genetic diversity and parentage in captive southern pygmy perch (*Nannoperca australis*)

Catherine R. M. Attard[1] · Chris J. Brauer[1] · Jacob D. Van Zoelen[1] · Minami Sasaki[1] · Michael P. Hammer[2] · Leslie Morrison[1] · James O. Harris[1] · Luciana M. Möller[1] · Luciano B. Beheregaray[1]

Received: 25 April 2016 / Accepted: 26 July 2016

Abstract Maintaining genetic diversity within captive breeding populations is a key challenge for conservation managers. We applied a multi-generational genetic approach to the captive breeding program of an endangered Australian freshwater fish, the southern pygmy perch (*Nannoperca australis*). During previous work, fish from the lower Murray-Darling Basin were rescued before drought exacerbated by irrigation resulted in local extinction. This endemic lineage of the species was captive-bred in genetically designed groups, and equal numbers of F1 individuals were reintroduced to the wild with the return of favourable habitat. Here, we implemented a contingency plan by continuing the genetic-based captive breeding in the event that a self-sustaining wild population was not established. F1 individuals were available as putative breeders from the subset of groups that produced an excess of fish in the original restoration program. We used microsatellite-based parentage analyses of these F1 fish to form breeding groups that minimized inbreeding. We assessed their subsequent parental contribution to F2 individuals and the maintenance of genetic diversity. We found skewed parental contribution to F2 individuals, yet minimal loss of genetic diversity from their parents. However, the diversity was substantially less than that of the original rescued population. We attribute this to the unavoidable use of F1 individuals from a limited number of the original breeding groups. Alternative genetic sources for supplementation or reintroduction should be assessed to determine their suitability. The genetic fate of the captive-bred population highlights the strong need to integrate DNA-based tools for monitoring and adaptive management of captive breeding programs.

Keywords Restoration genetics · Pedigree · Kinship · Relatedness · Fish · Biodiversity extinction

Introduction

Captive breeding programs aim to ensure the persistence of a population when it is endangered in the wild (Frankham et al. 2009). Captive-bred individuals may be used to supplement, reintroduce, or establish new wild populations. However, maintaining genetic diversity in captivity and successful supplementation or reintroduction is fraught with difficulties and compromises (Fraser 2008; Williams and Hoffman 2009), such as competing ecological and anthropogenic interests (Hobbs et al. 2009). A decline in genetic diversity and therefore evolutionary potential, or the detrimental consequences of inbreeding depression in captive, supplemented or reintroduced populations, can lead to the extinction of a population or species (Frankham 2005).

The Murray-Darling Basin (MDB) in Australia is a fundamentally altered and highly threatened ecosystem that is of great economic importance for agriculture. Human-made structures such as dams have been constructed since European settlement to regulate the naturally variable water flow for irrigation purposes (Kingsford et al. 2011).

Electronic supplementary material The online version of this article (doi:10.1007/s10592-016-0873-y) contains supplementary material, which is available to authorized users.

✉ Luciano B. Beheregaray
luciano.beheregaray@flinders.edu.au

[1] School of Biological Sciences, Flinders University, Adelaide, SA 5042, Australia

[2] Museum and Art Gallery of the Northern Territory, Darwin, NT 0801, Australia

Published online: 29 July 2016

Appendix 2: ddRAD library preparation and bioinformatics

Double digest RAD sequencing libraries were prepared following a protocol modified from Peterson et al. (2012). Using custom, in-house designed six-nucleotide barcodes with a minimum distance of three nucleotides 48 samples were multiplexed per Illumina lane. Samples were allocated to lanes randomly so that variation among sampling locations would not be confounded by artefacts arising due to differences among sequencing runs. For each lane of 48 individuals, 300ng total DNA from each sample was digested individually with two restriction enzymes *SbfI* and *MseI* (New England Biolabs) before ligation of the individual barcodes and RAD adapter sequences. Samples were then pooled into multiplex libraries of 12 uniquely barcoded individuals before purification with AMPure XP beads (Beckman Coulter Genomics) to remove any unligated adapters and adapter-to-adapter ligation products. Libraries were then size selected for an average of 500 base pair fragments (300-700bp) with a 1.5% Pippin prep electrophoresis gel (Sage Science) and quantified with a Qubit 2.0 fluorometer using the double-strand DNA broad range assay (Life Technologies). Polymerase chain reactions were then performed using two 25uL reactions per library (to reduce PCR bias associated with larger reaction volumes) before another bead purification step and fragment size evaluation with a Bioanalyzer 2100 (Agilent Technologies) using a DNA 7500 assay kit. Lastly, DNA quantity was assessed using real-time PCR to accurately equalize library concentrations before pooling four libraries of 12 samples together to create multiplex libraries of 48

uniquely barcoded samples for sequencing per one Illumina Hiseq2000 lane.

Libraries were sequenced as paired-end, 100-bp reads at the Genome Quebec/McGill University Innovation Center (Montreal, Canada). The raw data files were demultiplexed using the *process_radtags* component of *Stacks* v.1.04 (Catchen *et al.* 2011). Here, *process_radtags* was run twice utilizing the 'rescue barcodes' flag (-r), initially to recover reads with up to two errors in the individual barcode sequences before subsequently using the same flag to recover reads with up to three errors in the RAD-tag. The demultiplexed sequences were then processed using *dDocent.FB* v.1.2 (Puritz *et al.* 2014). *dDocent* combines several existing bioinformatics software packages into a single pipeline and was designed specifically for paired-end RAD data and is thus able to take advantage of both forward and reverse reads for SNP discovery. The pipeline consists of four basic steps; quality trimming, de novo assembly of a catalogue of reference contigs, mapping of the trimmed reads to the reference catalogue, and variant calling. After demultiplexing, the raw reads were quality trimmed using *Trimmomatic* v.0.33 (Bolger *et al.* 2014). *Trimmomatic* was configured to remove adapter sequences and then remove low quality bases (PHRED <20) from the beginning and end of the reads, before using a sliding window approach (five nucleotide window) to trim the total read length if the average PHRED score drops below ten. Next *Rainbow* v.2.0.2 (Chong *et al.* 2012) was used to cluster reads based on similarity and then assemble the clusters into longer reference contigs with the

maximum number of mismatches (-m) set to six. *CD-HIT* v.4.6 (Li *et al.* 2001) was then used to cluster the reference contigs based on 90% sequence similarity, retaining only the longest contig from each cluster in the final assembled reference contig data set. Using the MEM algorithm (Li 2013) implemented in *BWA* v.0.7.12-r1044 (Li & Durbin 2010), the quality trimmed reads were mapped to the reference contigs for each individual using default settings (match score -A=1, mismatch score -B=4, and gap-opening penalty -O=6). The resulting alignments were passed as indexed BAM files to the Bayesian variant caller *FreeBayes* v.0.9.20-8-gfef284a (Garrison & Marth 2012) to simultaneously detect SNPs, Indels, and more complex multi-nucleotide polymorphisms (MNPs) with default settings (minimum mapping quality -m=5, minimum base quality -q=5, and maximum complex gap -E=3). The resulting variant call file (VCF) containing information on sequence variation across all project samples (TotalRawSNPs.vcf) was subsequently filtered using custom BASH scripts utilizing *VCFtools* (Danecek *et al.* 2011) and *vcflib* (included in the *FreeBayes* package) separately for the landscape and conservation genomics chapters, and for the restoration genomics chapter. Details of these downstream filtering steps are reported in the methods sections of each data chapter.

The following steps were implemented in order to filter SNPs likely to be the result of sequencing errors, paralogs, multi-copy loci and artefacts of library preparation and are based on scripts from the *dDocent* GitHub

page (https://github.com/jpuritz/dDocent/). 1) Allele balance: for each locus, you should expect an approximately equal number of reads for the reference and alternate alleles for individuals called as heterozygotes. Loci were therefore removed if the proportion of alternate to reference allele was <0.25 or >0.75 across all heterozygote individuals. 2) Read orientation: each SNP should only occur in either forward or reverse reads. Those occurring in both are potentially paralogs and were accordingly filtered to retain only loci with at least 100 times more forward than reverse reads (or alternately, the opposite of 100 times more reverse than forward reads). 3) Mapping quality: as both alleles of a locus should start from the same RAD cut site, mapping quality scores (probability that the *BWA* alignment is correct) for the two alleles should be similar. Loci with a mapping quality score ratio (alternate allele mapping score/ reference allele mapping score) <90% or >110% were discarded. 4) Paired reads: loci where properly paired reads map to the reference allele but only unpaired reads map to the alternate allele are also indicative of potential paralogs and were removed. 5) Read quality: loci with overall low read quality scores (less than 25% of read depth) were discarded. Additionally, Li (2014), found a predictable relationship between Illumina read quality scores and read depth, such that where loci are covered by a high number of reads, quality scores are likely to be inflated. In this case, a higher quality score threshold is required to distinguish true variants from errors. Consequently, for loci with unusually high read depths (greater than the mean depth plus three times the square root of the mean), those with quality scores less than two times their read depth were

also removed. 6) Read depth: finally, the read depth of each locus was recalculated and the frequency distribution of mean depth per locus, averaged over all individuals was plotted to identify and remove loci with abnormally high coverage. Details of the number of SNPs retained after each filtering step are summarized in Table 2.2).

Appendix 3: Details of environmental variables considered for landscape genomics analysis of *Nannoperca australis* from the Murray-Darling Basin (MDB).

Category	Variable	Description
Climate		
Temperature	STRANNTEMP	Stream average annual mean temperature
	CATANNTEMP	Catchment average annual mean temperature
	STRCOLDMTHMIN	Stream average coldest month minimum temperature
	CATCOLDMTHMIN	Catchment average coldest month minimum temperature
	STRHOTMTHMAX	Stream average hottest month maximum temperature
	CATHOTMTHMAX	Catchment average hottest month maximum temperature
	STRCOLDQTEMP	Stream average coldest quarter mean temperature
	CATCOLDQTEMP	Catchment average coldest quarter mean temperature
	STRDRYQTEMP	Stream average driest quarter mean temperature
	CATDRYQTEMP	Catchment average driest quarter mean temperature
	STRWETQTEMP	Stream average wettest quarter mean temperature
	CATWETQTEMP	Catchment average wettest quarter mean temperature
Precipitation	STRANNRAIN	Stream average annual mean rainfall
	CATANNRAIN	Catchment average annual mean rainfall
	STRDRYQRAIN	Stream average driest quarter rainfall
	CATDRYQRAIN	Catchment average driest quarter rainfall
	STRWETQRAIN	Stream average wettest quarter rainfall
	CATWETQRAIN	Catchment average wettest quarter rainfall
	STRWARMQRAIN	Stream average warmest quarter rainfall
	CATWARMQRAIN	Catchment average warmest quarter rainfall
	STRCOLDQRAIN	Stream average coldest quarter rainfall
	CATCOLDQRAIN	Catchment average coldest quarter rainfall
Flow	RUNANNCOFV	Coefficient of variation of annual totals of accumulated soil water surplus
	RUNCVMAXMTH	Coefficient of variation of annual max monthly soil water surplus
	RUNMTHCOFV	Coefficient of variation of monthly totals of accumulated soil water surplus
	RUNPERENIALITY	% Contribution to mean annual flow by the six driest months of the year
	RUNANNMEAN	Annual mean accumulated soil water surplus
	SUBEROSIVITY	Sub-catchment average rainfall erosivity R-factor (rainfall intensity)
	CATEROSIVITY	Catchment average rainfall erosivity R-factor (rainfall intensity)
Disturbance	CDI	Catchment disturbance index
	FRDI	Flow regime disturbance index
	RDI	River disturbance index
Topography	STRAHLER	Strahler stream order
	SUBELEMAX	Maximum sub-catchment elevation
	SUBELEMEAN	Mean sub-catchment elevation
	CATELEMEAN	Mean catchment elevation
	VALLEYSLOPE	Stream segment slope
	CATSLOPE	Catchment average slope
	SUBAREA	Sub-catchment area
	CATAREA	Catchment area

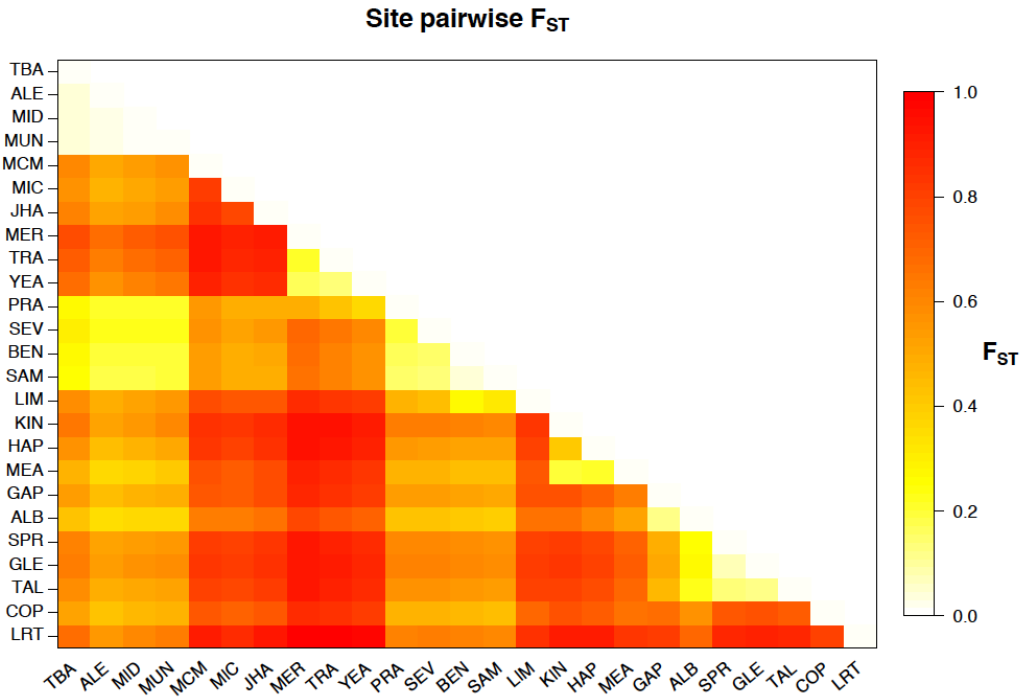
Appendix 4: Illumina HiSeq2000 sequencing statistics for 263 *Nannoperca australis* ddRAD libraries. Number of retained reads refers to properly paired reads remaining after demultiplexing described in the supplementary methods.

	Raw	Retained
Total reads (forward and reverse)	1,602,903,910	805,070,034
Average reads per sample	6,094,692	3,061,103
Minimum reads per sample	1,570,456	630,296
Maximum reads per sample	18,339,432	11,497,184

Appendix 5: Effective population size estimates (N_E) and *P*-values from bottleneck test for excess heterozygosity. *MID and MUN samples combined for N_e estimation. ** indicates $P<1x10^{-10}$. Lowland wetland sites referred to as Lower Murray in the text are indicated in bold.

Site	N_E (95% CI)		*P*-value
TBA	**∞**		**
ALE	**198.6**	**(158.6–264.9)**	**
MID	**190.9**	**(163.3–229.4)***	**
MUN			**
MCM	76.3	(61.0–101.3)	**
MIC	13.7	(13.2–14.4)	**
JHA	393.8	(184.0–∞)	**
MER	70.4	(61.4–82.2)	0.193
TRA	50.7	(41.2–65.3)	**
YEA	260.4	(111.1–∞)	**
PRA	114.9	(98.4–137.9)	**
SEV	54.8	(50.8–59.4)	**
BEN	117.2	(101.7–138.2)	**
SAM	124.7	(108.0–147.2)	**
LIM	99.1	(88.5–112.5)	**
KIN	69.9	(62.1–79.8)	**
HAP	∞		**
MEA	53.4	(45.7–64)	**
GAP	122.5	(105.3–146.2)	**
ALB	305.4	(241.8–413.4)	**
SPR	98.1	(80.5–125)	**
GLE	51.1	(46.1–57.2)	**
TAL	31.9	(29.1–35.2)	**
COP	118.7	(102.2–141.1)	**
LRT	18.1	(15.3–21.8)	0.748

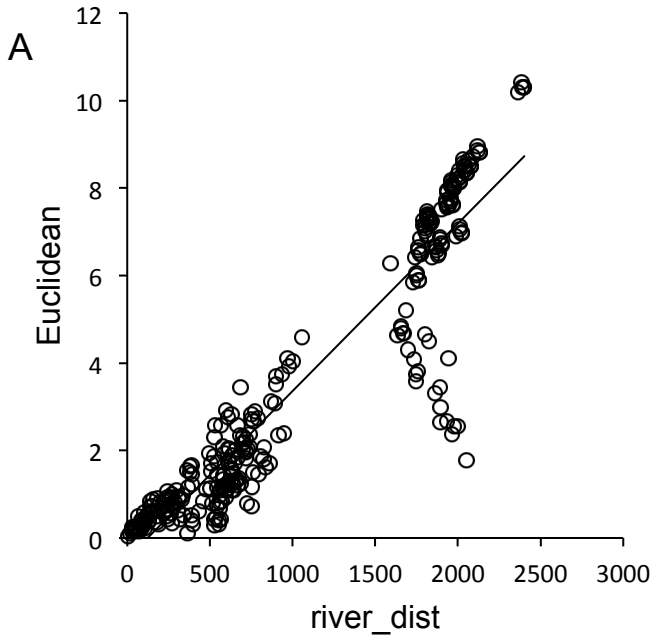
Appendix 6: Pairwise population F_{ST} among *Nannoperca australis* sampling sites.

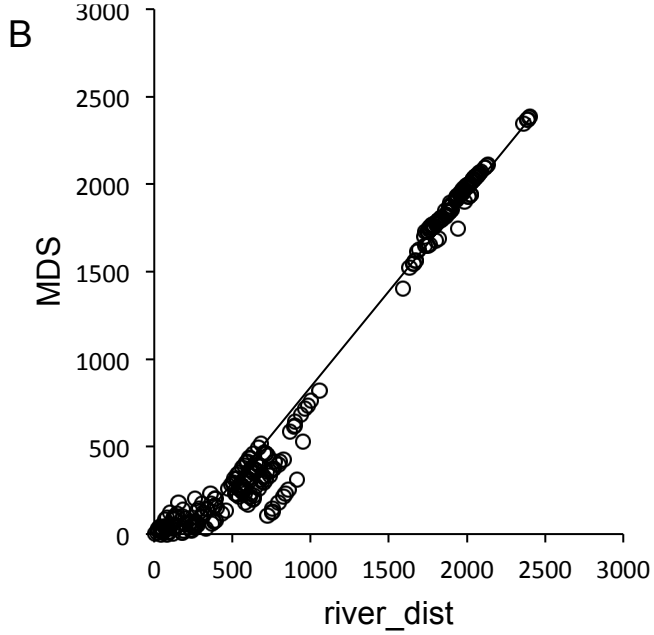


Appendix 7: Hierarchical analysis of molecular variance (AMOVA) based on neutral (3443 SNPs) F_{ST} for all samples and those catchments containing multiple sampling sites for *Nannoperca australis* from the Murray-Darling Basin (MDB).

Group	Source of variation	% variance	*P* value
All sites	Among catchments	30.3%	0.001
	Among sites within catchments	10.7%	0.001
	Among individuals within sites	13.5%	0.001
LMR	Among sites within catchment	1.3%	0.001
	Among individuals within sites	29.7%	0.001
GOU	Among sites within catchment	39.7%	0.001
	Among individuals within sites	13.6%	0.001
BRO	Among sites within catchment	14.3%	0.001
	Among individuals within sites	12.8%	0.001
OVE	Among sites within catchment	17.1%	0.001
	Among individuals within sites	15.5%	0.001
MIT	Among sites within catchment	6.6%	0.001
	Among individuals within sites	16.3%	0.001

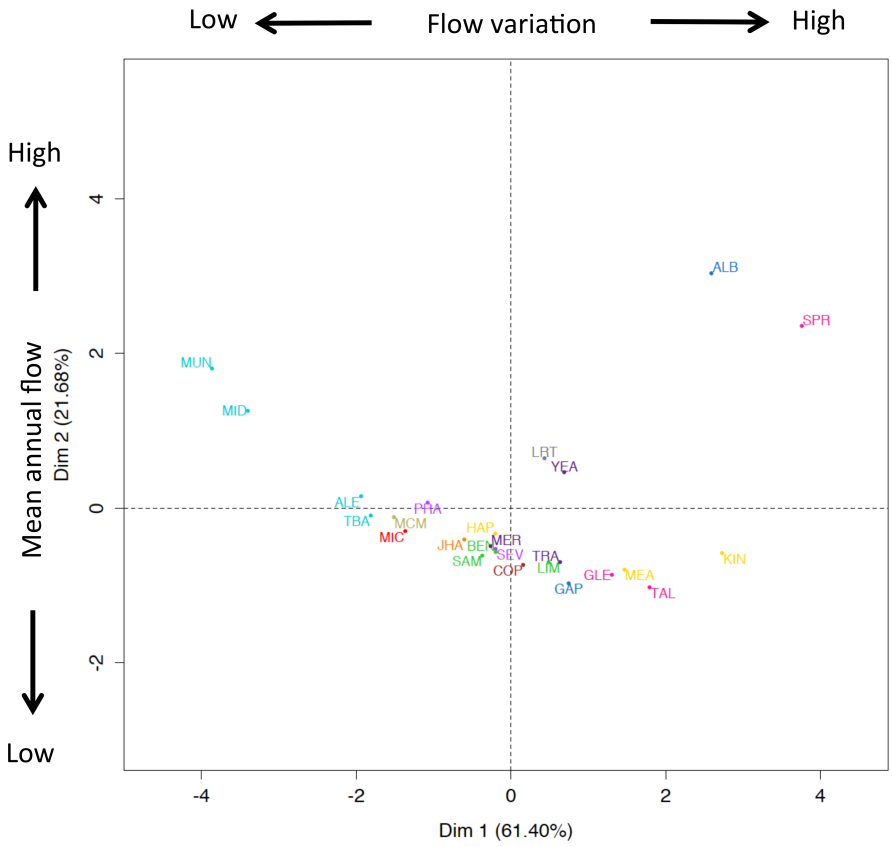
Appendix 8: Correlation between river distance between sampling sites and A) Euclidean distance (R^2=0.87), and B) distance calculated using multi-dimensional scaling (MDS) (R^2=0.97).



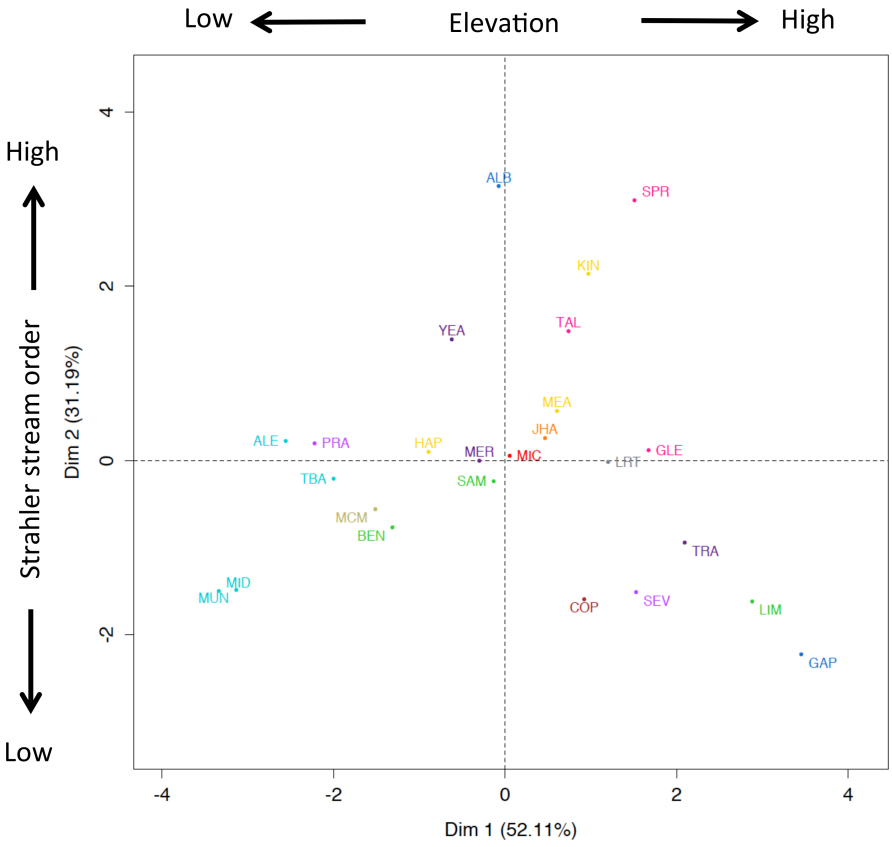


Appendix 9: Environmental PCAs describing the relationship between each *Nannoperca australis* sampling location based on variables related to a) flow, b) human disturbance and c) topography. Site names are colour coded based on the colours used in Figure 1. Annotations above and to the left of each plot describe which variables contribute the most to each axis.

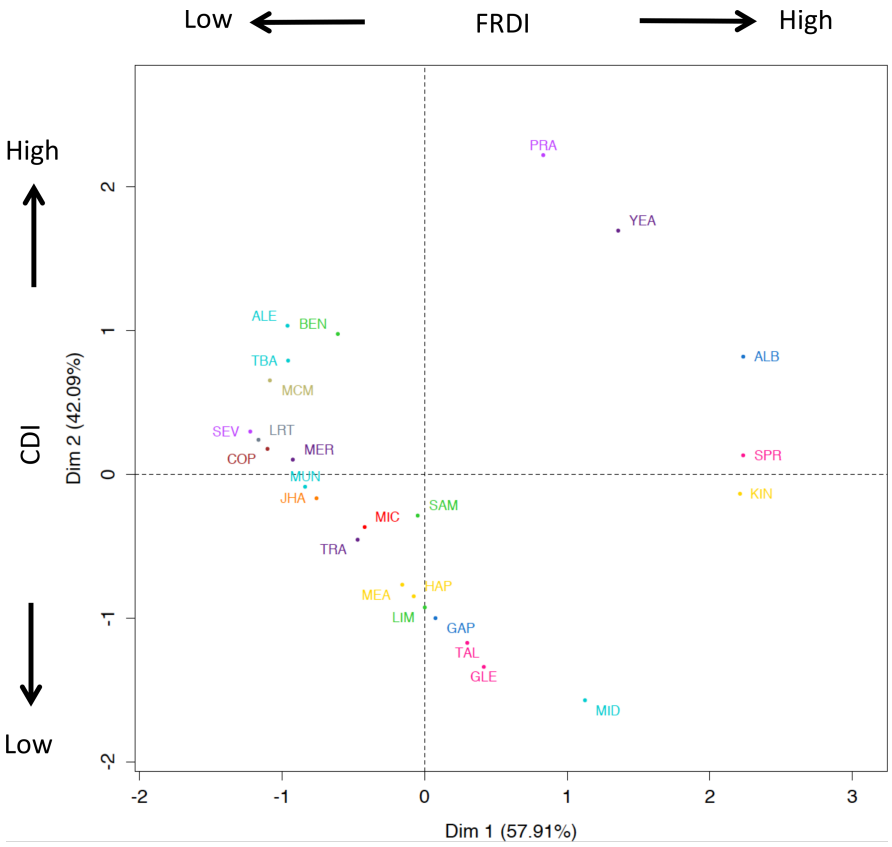
a



b


Low
Elevation
High
High
Strahler stream order
Low
Dim 2 (31.19%)
Dim 1 (52.11%)
ALB
SPR
KIN
TAL
YEA
MEA
JHA
ALE
PRA
HAP
MER
MIC
GLE
LRT
TBA
SAM
MCM
BEN
TRA
MID
MUN
COP
SEV
LIM
GAP

c


Low
FRDI
High
High
CDI
Low
Dim 2 (42.09%)
Dim 1 (57.91%)
PRA
YEA
ALE
BEN
TBA
MCM
ALB
SEV
LRT
MER
COP
SPR
MUN
KIN
JHA
SAM
MIC
TRA
MEA
HAP
LIM
GAP
TAL
GLE
MID

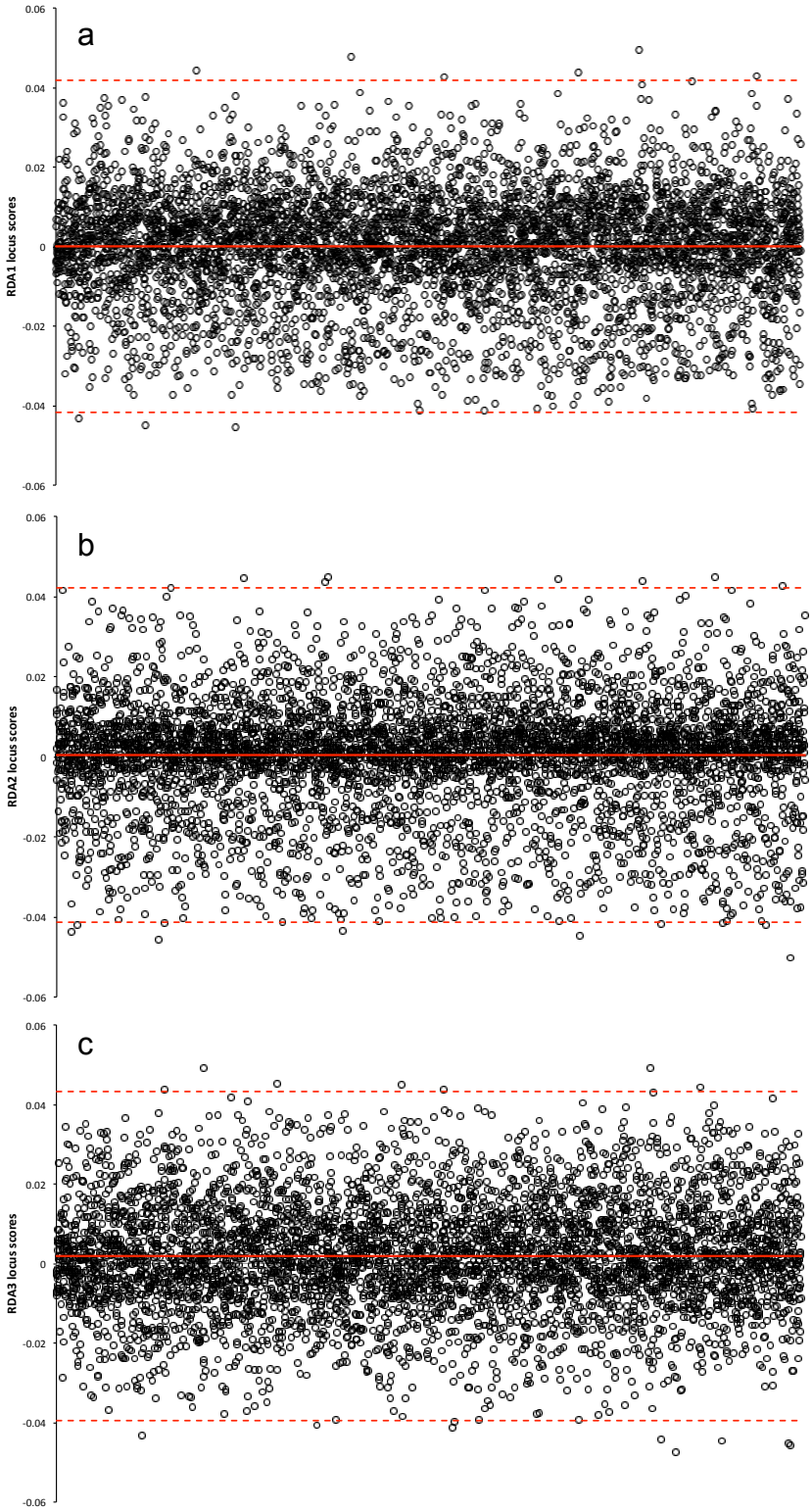
Appendix 10: Variation explained by retained environmental principal components and the correlation between the original variables and each component. Only variables significantly correlated (P<0.05) with each component are shown. Precipitation and disturbance variables shown in bold were analysed as individual variables.

Category	PC	% Variation	Variables	Correlation
Climate	Temp1	42.0%	STRCOLDMTHMIN	0.895
			STRWETQTEMP	0.871
	Temp2	33.3%	STRDRYQTEMP	0.839
			CATDRYQTEMP	0.729
Precipitation	–		**CATDRYQRAIN**	–
			STRWETQRAIN	–
			CATCOLDQRAIN	–
Flow	Flow1	61.4%	CATEROSIVITY	0.949
			RUNPERENIALITY	0.849
			SUBEROSIVITY	0.828
			RUNANNMEAN	0.569
			RUNCVMAXMTH	-0.662
	Flow2	21.7%	RUNANNMEAN	0.762
			RUNCVMAXMTH	0.450
			SUBEROSIVITY	-0.401
Disturbance	–		**FRDI**	–
			CDI	–
Topography	Topo1	52.1%	SUBELEMAX	0.952
			SUBELEMEAN	0.879
			CATELEMEAN	0.729
			CATSLOPE	0.714
			VALLEYSLOPE	0.628
	Topo2	31.2%	STRAHLER	0.942
			CATELEMEAN	0.631
			CATSLOPE	0.447
			VALLEYSLOPE	-0.568

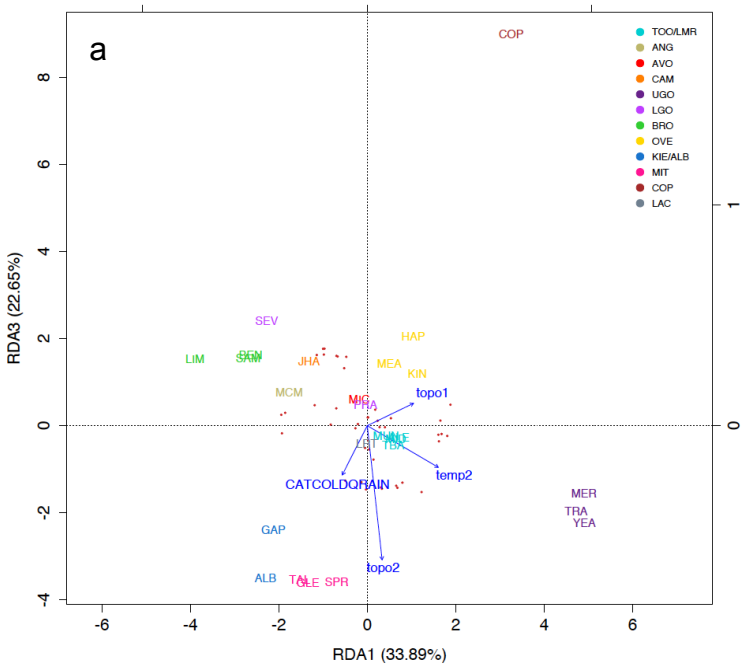
Appendix 11: Number of candidate loci identified for each environmental variable using gINLAnd (log-Bayes factor >15).

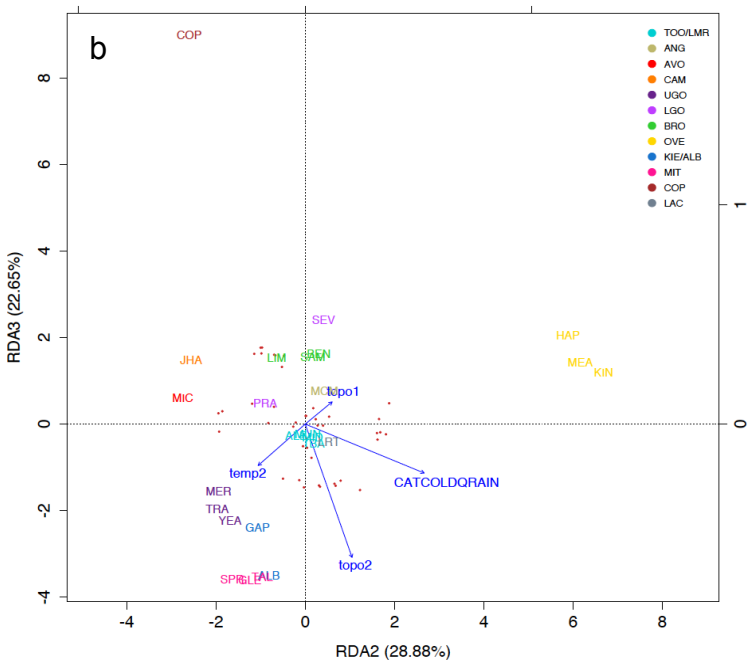
Category	Variable	Loci
Temperature	Temp1	17
	Temp2	41
Precipitation	CATCOLDQRAIN	74
	CATDRYQRAIN	17
	STRWETQRAIN	15
Flow	Flow1	35
	Flow2	4
Disturbance	CDI	35
	FRDI	8
Topography	Topo1	18
	Topo2	9

Appendix 12: Locus scores for the first three axes (a, b, and c respectively) of the partial redundancy analysis. Loci more than three standard deviations (dashed red lines) from the mean locus score (solid red lines) were considered as outliers.



Appendix 13: RDA triplots with a) RDA1 vs. RDA3, and b) RDA2 vs. RDA3. Sampling sites are colour coded according to colours used in Figure 2.1 and plotted based on population scores for each constrained RDA axis. Environmental factors are represented as blue vectors where the length represents the magnitude of the environmental variables contribution in explaining SNP variance. The angle between each environmental vector represents the correlation among environmental variables. Allele frequency vectors for the individual SNPs identified as outliers (more than three SD from the mean locus scores) are indicated by red markers.

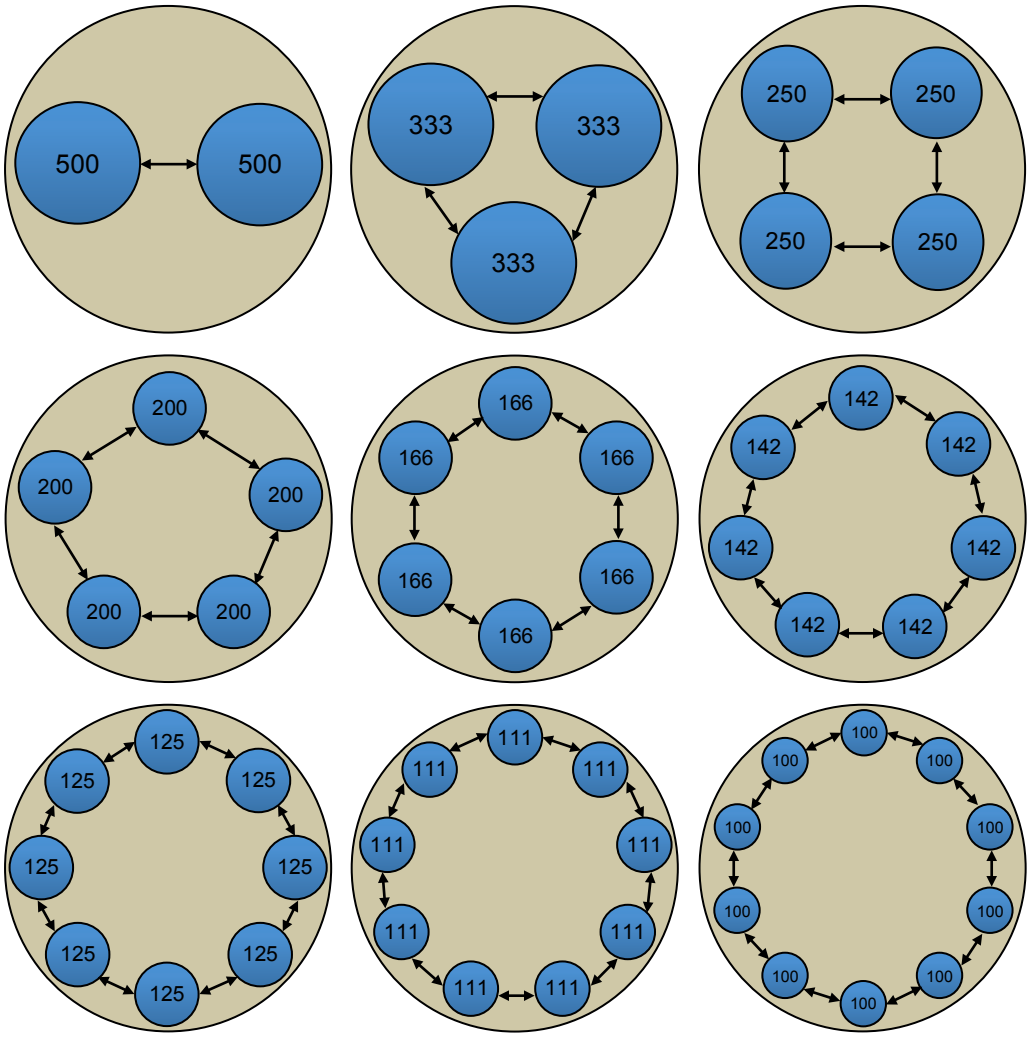




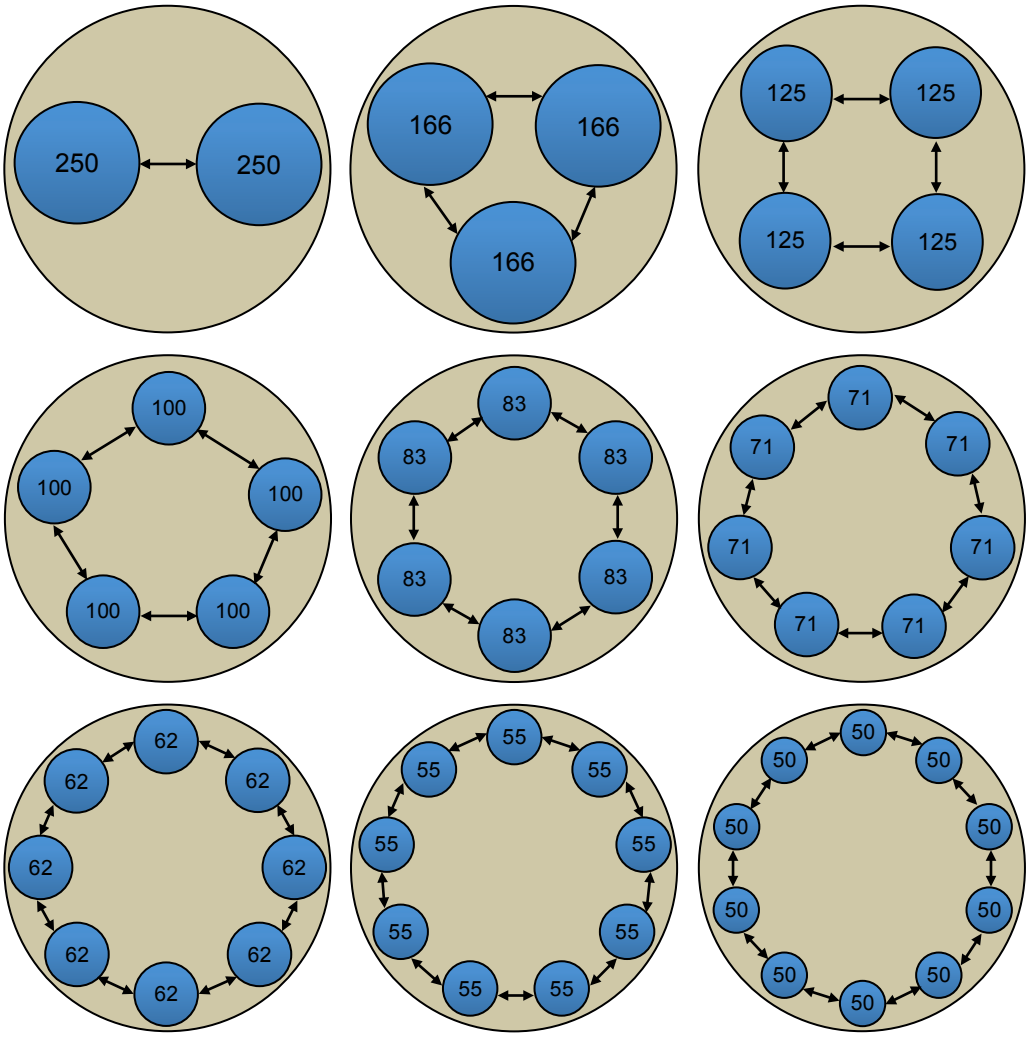
Appendix 14: Binary model matrix describing pairwise comparisons of sampling sites within catchments (0), and among catchments (1).

	TBA	ALE	MID	MUN	MCM	MIC	JHA	MER	TRA	YEA	PRA	SEV	BEN	SAM	LIM	KIN	HAP	MEA	GAP	ALB	SPR	GLE	TAL	COP	LRT
TBA	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ALE	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MID	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MU	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MC	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MIC	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JHA	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MER	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
TRA	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
YEA	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
PRA	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
SEV	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
BEN	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1
SAM	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1
LIM	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1
KIN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1
HAP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1
MEA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1
GAP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
ALB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
SPR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
GLE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
TAL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
COP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
LRT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

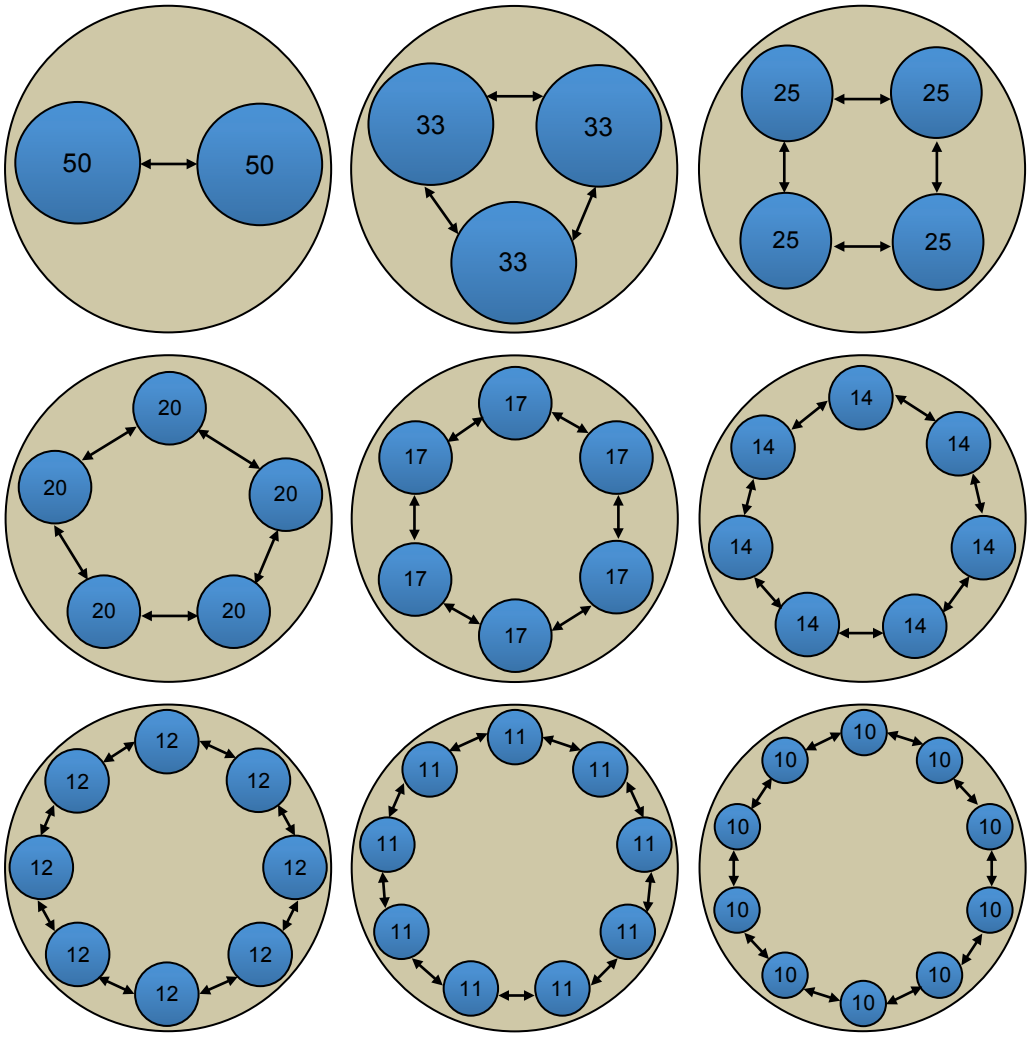
Appendix 15: Schematic representation of simulated headwater catchment metapopulations for *Nannoperca australis*. Nine runs with increasing levels of fragmentation (n=2-10 habitat patches) were completed for each metapopulation. Each simulation was based on a stepping stone population model assuming equal N_e for each sub-population, while maintaining a constant metapopulation N_e of ~1000 to simulate a concurrent reduction in habitat patch size with increasing fragmentation.



Appendix 16: Schematic representation of simulated headwater catchment metapopulations for *Nannoperca australis*. Nine runs with increasing levels of fragmentation (n=2-10 habitat patches) were completed for each metapopulation. Each simulation was based on a stepping stone population model assuming equal N_e for each sub-population, while maintaining a constant metapopulation N_e of ~500 to simulate a concurrent reduction in habitat patch size with increasing fragmentation.



Appendix 17: Schematic representation of simulated headwater catchment metapopulations for *Nannoperca australis*. Nine runs with increasing levels of fragmentation (n=2-10 habitat patches) were completed for each metapopulation. Each simulation was based on a stepping stone population model assuming equal N_e for each sub-population, while maintaining a constant metapopulation N_e of ~100 to simulate a concurrent reduction in habitat patch size with increasing fragmentation.



Appendix 18: Estimates of F_{ST} for 300 generations of simulated metapopulations of N_e=1000 with increasing levels of fragmentation.

	Number of fragments X fragment N_e								
Generation	2x500	3x333	4x250	5x200	6x166	7x142	8x125	9x111	10x100
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.01	0.02	0.02	0.03	0.03	0.04	0.04	0.05	0.05
20	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.10	0.10
30	0.03	0.04	0.06	0.07	0.09	0.10	0.12	0.14	0.14
40	0.04	0.06	0.08	0.10	0.12	0.14	0.15	0.18	0.19
50	0.05	0.07	0.10	0.12	0.15	0.17	0.19	0.21	0.23
60	0.06	0.09	0.11	0.14	0.17	0.20	0.22	0.25	0.27
70	0.07	0.10	0.13	0.16	0.20	0.22	0.26	0.28	0.30
80	0.08	0.11	0.15	0.18	0.22	0.25	0.28	0.31	0.34
90	0.08	0.13	0.16	0.20	0.25	0.28	0.31	0.34	0.37
100	0.09	0.14	0.18	0.22	0.27	0.30	0.34	0.37	0.40
110	0.10	0.15	0.20	0.24	0.29	0.32	0.36	0.40	0.43
120	0.11	0.17	0.21	0.25	0.31	0.35	0.39	0.42	0.45
130	0.12	0.18	0.23	0.27	0.33	0.37	0.41	0.45	0.48
140	0.13	0.19	0.24	0.29	0.35	0.39	0.43	0.47	0.50
150	0.14	0.20	0.26	0.31	0.37	0.41	0.46	0.49	0.52
160	0.14	0.21	0.27	0.32	0.39	0.43	0.47	0.51	0.55
170	0.15	0.23	0.29	0.34	0.41	0.45	0.50	0.53	0.57
180	0.16	0.24	0.30	0.35	0.42	0.47	0.52	0.55	0.59
190	0.17	0.25	0.32	0.37	0.44	0.49	0.53	0.57	0.60
200	0.18	0.26	0.33	0.38	0.46	0.50	0.55	0.59	0.62
210	0.18	0.27	0.34	0.40	0.47	0.52	0.57	0.60	0.64
220	0.19	0.28	0.36	0.41	0.48	0.54	0.59	0.62	0.66
230	0.19	0.29	0.37	0.43	0.50	0.55	0.60	0.63	0.67
240	0.20	0.30	0.38	0.44	0.51	0.57	0.62	0.65	0.68
250	0.21	0.31	0.39	0.45	0.53	0.58	0.63	0.66	0.69
260	0.22	0.32	0.40	0.46	0.54	0.60	0.64	0.67	0.71
270	0.23	0.33	0.41	0.48	0.55	0.61	0.65	0.68	0.72
280	0.24	0.34	0.42	0.49	0.56	0.62	0.67	0.70	0.73
290	0.24	0.34	0.43	0.50	0.58	0.63	0.68	0.71	0.74
300	0.25	0.35	0.45	0.51	0.59	0.64	0.69	0.72	0.75

Appendix 19: Estimates of F_{ST} for 300 generations of simulated metapopulations of N_e=500 with increasing levels of fragmentation.

	Number of fragments X fragment N_e								
Generation	2x250	3x166	4x125	5x100	6x83	7x71	8x62	9x55	10x50
0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.01	0.03	0.04	0.06	0.10	0.08	0.09	0.10	0.11
20	0.03	0.06	0.08	0.11	0.15	0.14	0.16	0.17	0.20
30	0.04	0.09	0.12	0.16	0.20	0.19	0.23	0.25	0.27
40	0.06	0.12	0.16	0.20	0.25	0.24	0.28	0.32	0.35
50	0.07	0.15	0.19	0.23	0.29	0.30	0.34	0.38	0.40
60	0.09	0.17	0.22	0.27	0.33	0.34	0.40	0.42	0.45
70	0.10	0.20	0.25	0.30	0.36	0.38	0.44	0.47	0.50
80	0.11	0.22	0.28	0.34	0.40	0.41	0.47	0.51	0.54
90	0.13	0.25	0.31	0.36	0.42	0.44	0.52	0.55	0.57
100	0.14	0.27	0.34	0.38	0.45	0.48	0.55	0.58	0.61
110	0.15	0.29	0.37	0.41	0.48	0.51	0.58	0.62	0.64
120	0.16	0.31	0.39	0.43	0.50	0.53	0.61	0.65	0.68
130	0.18	0.33	0.41	0.46	0.52	0.56	0.63	0.67	0.70
140	0.19	0.35	0.43	0.48	0.55	0.58	0.65	0.69	0.72
150	0.20	0.37	0.45	0.50	0.57	0.60	0.67	0.71	0.74
160	0.21	0.39	0.47	0.51	0.59	0.63	0.69	0.73	0.76
170	0.22	0.41	0.48	0.53	0.60	0.64	0.71	0.74	0.77
180	0.23	0.42	0.50	0.55	0.62	0.65	0.72	0.75	0.78
190	0.24	0.44	0.52	0.57	0.63	0.66	0.73	0.77	0.79
200	0.25	0.45	0.53	0.59	0.65	0.68	0.75	0.78	0.80
210	0.26	0.47	0.55	0.61	0.66	0.69	0.76	0.79	0.81
220	0.27	0.48	0.56	0.62	0.68	0.72	0.77	0.81	0.81
230	0.28	0.49	0.57	0.63	0.69	0.73	0.79	0.81	0.82
240	0.29	0.51	0.59	0.65	0.70	0.74	0.80	0.82	0.83
250	0.30	0.52	0.60	0.66	0.71	0.75	0.80	0.83	0.84
260	0.31	0.53	0.61	0.67	0.72	0.75	0.81	0.83	0.84
270	0.32	0.54	0.62	0.68	0.73	0.76	0.82	0.84	0.84
280	0.32	0.55	0.64	0.69	0.74	0.77	0.82	0.85	0.85
290	0.33	0.56	0.64	0.70	0.74	0.77	0.82	0.85	0.85
300	0.34	0.58	0.65	0.71	0.75	0.78	0.83	0.86	0.85

Appendix 20: Estimates of F_{ST} for 300 generations of simulated metapopulations of N_e=100 with increasing levels of fragmentation.

	Number of fragments X fragment N_e								
Generation	2x50	3x33	4x25	5x20	6x17	7x14	8x12	9x11	10x10
0	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
10	0.09	0.14	0.20	0.25	0.32	0.33	0.39	0.45	0.44
20	0.15	0.25	0.34	0.41	0.50	0.52	0.62	0.65	0.66
30	0.22	0.35	0.45	0.53	0.63	0.66	0.74	0.77	0.78
40	0.28	0.45	0.56	0.61	0.71	0.75	0.81	0.85	0.84
50	0.35	0.54	0.62	0.68	0.77	0.81	0.85	0.88	0.87
60	0.42	0.62	0.66	0.75	0.82	0.84	0.88	0.90	0.89
70	0.46	0.67	0.72	0.79	0.84	0.87	0.90	0.92	0.92
80	0.52	0.70	0.75	0.81	0.86	0.89	0.91	0.93	0.93
90	0.56	0.71	0.77	0.84	0.87	0.90	0.91	0.93	0.93
100	0.59	0.73	0.80	0.86	0.88	0.90	0.92	0.93	0.93
110	0.60	0.74	0.83	0.86	0.89	0.90	0.92	0.94	0.93
120	0.62	0.75	0.85	0.87	0.88	0.91	0.93	0.93	0.94
130	0.64	0.78	0.86	0.88	0.89	0.91	0.93	0.93	0.94
140	0.66	0.79	0.86	0.88	0.90	0.92	0.94	0.93	0.94
150	0.66	0.79	0.87	0.89	0.91	0.92	0.94	0.94	0.95
160	0.66	0.81	0.87	0.90	0.91	0.93	0.94	0.94	0.95
170	0.68	0.82	0.88	0.90	0.92	0.93	0.94	0.95	0.95
180	0.70	0.83	0.89	0.90	0.91	0.93	0.94	0.94	0.96
190	0.70	0.83	0.89	0.91	0.91	0.93	0.95	0.95	0.96
200	0.72	0.83	0.89	0.91	0.92	0.94	0.94	0.94	0.95
210	0.72	0.84	0.89	0.92	0.92	0.94	0.95	0.94	0.95
220	0.72	0.84	0.89	0.92	0.93	0.93	0.95	0.93	0.95
230	0.74	0.84	0.89	0.93	0.93	0.93	0.95	0.94	0.95
240	0.75	0.85	0.89	0.92	0.93	0.93	0.95	0.94	0.96
250	0.76	0.85	0.89	0.92	0.93	0.94	0.95	0.95	0.95
260	0.76	0.86	0.89	0.92	0.93	0.94	0.95	0.95	0.95
270	0.76	0.86	0.90	0.92	0.93	0.95	0.96	0.96	0.95
280	0.77	0.87	0.90	0.92	0.94	0.95	0.96	0.95	0.96
290	0.78	0.87	0.90	0.92	0.94	0.95	0.96	0.96	0.96
300	0.78	0.87	0.91	0.93	0.94	0.95	0.96	0.96	0.96

Appendix 21: PyRAD parameters for ddRAD sequence alignment.

```
==** parameter inputs for pyRAD version 3.0.6  **======================= affected
step ==
./                          ## 1. Working directory
(all)
              ## 2. Loc. of non-demultiplexed files (if not line 18)  (s1)
              ## 3. Loc. of barcode file (if not line 18)            (s1)
vsearch                     ## 4. command (or path) to call vsearch (or usearch)
(s3,s6)
muscle                      ## 5. command (or path) to call muscle
(s3,s7)
                     ## 6. cutsites...     (s1,s2)
6                           ## 7. N processors (parallel)
(all)
7                           ## 8. Mindepth: min coverage for a cluster
(s4,s5)
5                           ## 9. NQual: max # sites with qual < 20 (or see line
20)(s2)
.80                         ## 10. Wclust: clustering threshold as a decimal
(s3,s6)
ddrad                         ## 11. Datatype: rad,gbs,pairgbs,pairddrad,(others:see
docs)(all)
25                           ## 12. MinCov: min samples in a final locus
(s7)
p.60                           ## 13. MaxSH: max inds with shared hetero site
(s7)
ddRAD_SPPYPPout_clust80red                 ## 14. Prefix name for final output (no
spaces)         (s7)
==== optional params below this line ==================================  affected
step ==
                       ## 15.opt.: select subset (prefix* only selector)
(s2-s7)
                       ## 16.opt.: add-on (outgroup) taxa (list or prefix*)
(s6,s7)
                        ## 17.opt.: exclude taxa (list or prefix*)
       (s7)
@./*.fq.gz                          ## 18.opt.: loc. of de-multiplexed data
(s2)
                       ## 19.opt.: maxM: N mismatches in barcodes (def= 1)
(s1)
                       ## 20.opt.: phred Qscore offset (def= 33)
(s2)
                       ## 21.opt.: filter: def=0=NQual 1=NQual+adapters. 2=strict
(s2)
                       ## 22.opt.: a priori E,H (def= 0.001,0.01, if not
estimated) (s5)
                       ## 23.opt.: maxN: max Ns in a cons seq (def=5)
(s5)
                       ## 24.opt.: maxH: max heterozyg. sites in cons seq (def=5)
(s5)
                       ## 25.opt.: ploidy: max alleles in cons seq (def=2;see
docs) (s4,s5)
10                       ## 26.opt.: maxSNPs: (def=100). Paired (def=100,100)
(s7)
6,15                       ## 27.opt.: maxIndels: within-clust,across-clust (def.
3,99) (s3,s7)
                       ## 28.opt.: random number seed (def. 112233)
(s3,s6,s7)
                       ## 29.opt.: trim overhang left,right on final loci,
def(0,0) (s7)
*                       ## 30.opt.: output formats: p,n,a,s,v,u,t,m,k,g,* (see
docs) (s7)
                       ## 31.opt.: maj. base call at depth>x<mindepth
(def.x=mindepth) (s5)
                       ## 32.opt.: keep trimmed reads (def=0). Enter min length.
(s2)
mean+2*SD                           ## 33.opt.: max stack size (int), def=
max(500,mean+2*SD)    (s3)
2                         ## 34.opt.: minDerep: exclude dereps with <= N copies,
def=1 (s3)
                      ## 35.opt.: use hierarchical clustering (def.=0, 1=yes)
(s6)
                       ## 36.opt.: repeat masking (def.=1='dust' method, 0=no)
(s3,s6)
6                       ## 37.opt.: vsearch max threads per job (def.=6; see docs)
(s3,s6)
```

Appendix 22: Custom R script to process results of *EVE* comparative transcriptomics software.

```
# calculate FDR P values from EVE LRT test,  plot results and save
results to .xlsx and .pdf files
# takes as input a *TestLRTs.res result file from EVE
# load xlsx package to write results to excel file
require(xlsx)

# set directory to save output files to
setwd("/Users/chrisbrauer/Analysis/EVE_release/results/final_thesis")

# define EVEresults function
EVEresults <- function(genes, LRT, df, title, rate, output){
  P_diverge <- 1 - pchisq(LRT, df = df)
  P_diverse <- pchisq(LRT, df = df)
  FDR_diverge <- p.adjust(P_diverge, "fdr")
  FDR_diverse <- p.adjust(P_diverse, "fdr")
  EVE_diverge <- as.data.frame(cbind(genes, LRT, P_diverge, FDR_diverge))
  EVE_diverse <- as.data.frame(cbind(genes, LRT, P_diverse, FDR_diverse))
  colnames(EVE_diverge) <- c("Gene", "LRT", "P", "FDR")
  colnames(EVE_diverse) <- c("Gene", "LRT", "P", "FDR")
  EVE_diverge.sub <- subset(EVE_diverge, EVE_diverge$FDR < rate)
  EVE_diverge.sub <- as.data.frame(EVE_diverge.sub)
  colnames(EVE_diverge.sub)<- c("Gene", "LRT", "P", "FDR")
  EVE_diverse.sub <- subset(EVE_diverse, EVE_diverse$FDR < rate)
  EVE_diverse.sub <- as.data.frame(EVE_diverse.sub)
  colnames(EVE_diverse.sub)<- c("Gene", "LRT", "P", "FDR")
  FDR_diverge.len <- length(EVE_diverge.sub[,1])
  EVE_diverge.len <- length(EVE_diverge[,1])
  FDR_diverse.len <- length(EVE_diverse.sub[,1])
  EVE_diverse.len <- length(EVE_diverse[,1])
  print(paste0(FDR_diverge.len, " genes have higher variance among than
within lineages at ", rate, "FDR."))
  print(paste0(FDR_diverse.len, " genes have higher variance within than
among lineages at ", rate, "FDR."))
  diverge_result <- list(EVE_diverge, EVE_diverge.sub)
  diverse_result <- list(EVE_diverse, EVE_diverse.sub)
  wd <- getwd()
  print(paste0("Writing results to ", wd))

  #write results to 2x .csv files if too many genes for .xslx, otherwise
write to 2 sheets of .xlsx
  #Note. I haven't tested exactly how many rows write.xlsx can cope with,
it is likely more than the 5000 I have limited this to.
  if (FDR_diverge.len >5000) {
  write.csv(diverge_result[1],
file=paste0(output,"_EVE_diverge_Pvalues.csv"))
  write.csv(diverge_result[2],
file=paste0(output,"_EVE_diverge_FDRgenes.csv"))

  } else {
  if (FDR_diverge.len >0) {
    write.xlsx(diverge_result[1],
file=paste0(output,"_EVE_diverge_result.xlsx"), sheetName="Gene Pvalues")
    write.xlsx(diverge_result[2],
file=paste0(output,"_EVE_diverge_result.xlsx"), sheetName="FDRgenes",
append=TRUE)
} else {
    write.xlsx(diverge_result[1],
file=paste0(output,"_EVE_diverge_result.xlsx"), sheetName="Gene Pvalues")
  }
}

  if (EVE_diverse.len >5000) {
    write.csv(diverse_result[1],
file=paste0(output,"_EVE_diverse_result_Pvalues.csv"))
```

```
    write.csv(diverse_result[2],
file=paste0(output,"_EVE_diverse_result_FDRgenes.csv"))

  } else {
    if (EVE_diverse.len >0) {
      write.xlsx(diverse_result[1],
file=paste0(output,"_EVE_diverse_result.xlsx"), sheetName="Gene Pvalues")
      write.xlsx(diverse_result[2],
file=paste0(output,"_EVE_diverse_result.xlsx"), sheetName="FDRgenes",
append=TRUE)
    } else {
      write.xlsx(diverse_result[1],
file=paste0(output,"_EVE_diverse_result.xlsx"), sheetName="Gene Pvalues")
    }
  }

  index.lim <- length(EVE_diverge[,1])
  index.lim <- ceiling(index.lim/500)*500
  fig <- plot.default(EVE_diverge$LRT, col=ifelse(EVE_diverge$FDR < rate,
"red", "black"), main = title, xlim = c(0, index.lim), xaxt = "n")
  axis(side=1,at=pretty(seq(0, index.lim, by=1000)),labels=pretty(seq(0,
index.lim, by=1000)))
  pdf(file = paste0(output,"_EVE_diverge_result.pdf"), height = 8.27,
width = 11.69)
  plot.default(EVE_diverge$LRT, col=ifelse(EVE_diverge$FDR < rate, "red",
"black"), main = title, xlim = c(0, index.lim), xaxt = "n")
  axis(side=1,at=pretty(seq(0, index.lim, by=1000)),labels=pretty(seq(0,
index.lim, by=1000)))
  dev.off()
  print("Have a nice day.")
  return(diverge_result)
  return(diverse_result)
  return(fig)
}
```

Appendix 23: Number of Illumina RNA-seq read pairs per sample.

Sample	RAW	Trimmed	Retained
KIN02	17206195	16943343	16034552
KIN03	20473842	20155449	17986078
KIN07	12041370	11864939	11269913
KIN08	14366422	14177906	13405395
KIN09	25204364	24847366	23443402
KIN10a	14182006	13962617	13114174
KIN10b	15471530	15248127	14250972
LIM22	24216163	23883927	22146015
LIM23	20957848	20595925	17760496
LIM25	17820415	17549763	16404591
LIM29	12431410	12254994	11664328
LIM31	6654923	6546730	6142822
MER46	11667003	11483814	10666408
MER48	15555108	15293892	14211602
MER50a	10526891	10365576	9842741
MER50b	9060988	8926589	8485095
MER51	36042279	35515038	33596842
MER53	16588545	16319535	15308229
MER55	5885516	5789189	5409668
SEV34	23762752	23381298	21842245
SEV35	14203198	13985276	12672826
SEV38	15654131	15403471	13677085
SEV39	13253099	13046349	12223131
SEV40a	14944086	14703981	13813820
SEV40b	15930292	15683762	14760080
SEV44	29838764	29408155	27316392
SPBR12	20798302	20525303	18501838
SPBR15	22344392	22033905	20506396
SPBR16a	15507359	15319938	14085089
SPBR16b	17307199	17082622	15941604
SPBR25	17139691	16909199	15491227
SPBR31	13670893	13490868	12566241
SPF1L1	10301534	10137203	9570502
SPF1L2	26109032	25764580	24313867
SPF1M1	19879083	19642279	18665527
SPF1N1	12985562	12778024	11996453
YPBR12	13086886	12921721	12141175
YPBR17	15585089	15383904	14285706
YPBR28	19112234	18814761	17617314
YPBR30	6282669	6175293	5741843
YPBR68a	10855168	10720726	10160099
YPBR68b	12162570	11999286	11415382
YPBR71	10203692	10075835	8945362
YPF1L1	13950565	13738919	12812235
YPF1L2	12948042	12761721	11905306
YPF1M1	20331008	20055838	18590670
YPF1N1	16239723	15986849	15030119
YPF1N2	10715691	10535700	10002290
Average	16,071,990	15,837,323	14,744,482
Total read pairs	771,455,524	760,191,485	707,735,147

Appendix 24: Number of raw reads mapped to the *Nannoperca* de novo transcriptome assembly.

	Reads	%
Proper pairs	769,254,808	64.55
Improper pairs	321,556,874	26.98
Right only	50,750,778	4.26
Left only	50,196,232	4.21
Total aligned reads	1,191,758,692	100.0

Appendix 25: N50 statistics for the *Nannoperca* de novo transcriptome assembly.

	All transcripts	Longest isoform per gene
N10	4613	4296
N20	3555	3026
N30	2850	2119
N40	2299	1417
N50	1810	926
Average length	933.28	605.17

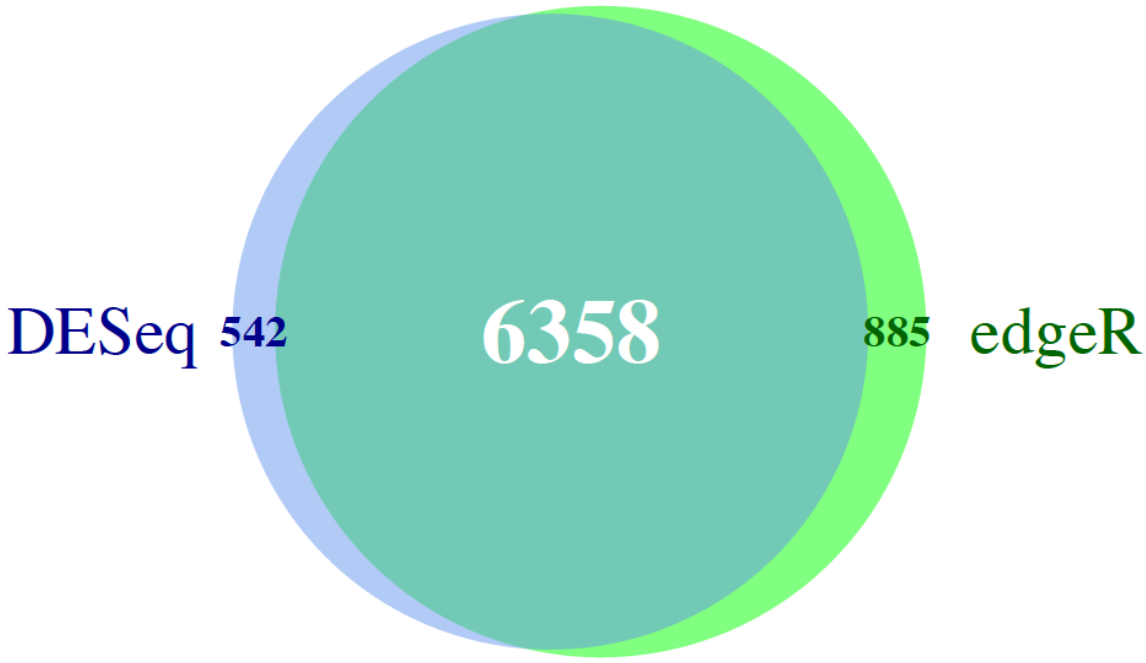
Appendix 26: Number of unique, top-matching SwissProt proteins covered by the Nannoperca de novo transcriptome assembly (*e*-value threshold of $1x10^{-20}$).

% Gene coverage	Count	Cumulative
100	4346	4346
90	1760	6106
80	1706	7812
70	1802	9614
60	2048	11662
50	2193	13855
40	2518	16373
30	2633	19006
20	2624	21630
10	960	22590

Appendix 27: Vertebrate BUSCO database search results for the *Nannoperca* de novo transcriptome assembly.

Complete BUSCOs	1889
Complete and single-copy BUSCOs	1333
Complete and duplicated BUSCOs	556
Fragmented BUSCOs	488
Missing BUSCOs	646
Total BUSCO groups searched	3023

Appendix 28: Comparison of *DESeq* and *edgeR* differential expression analysis results.



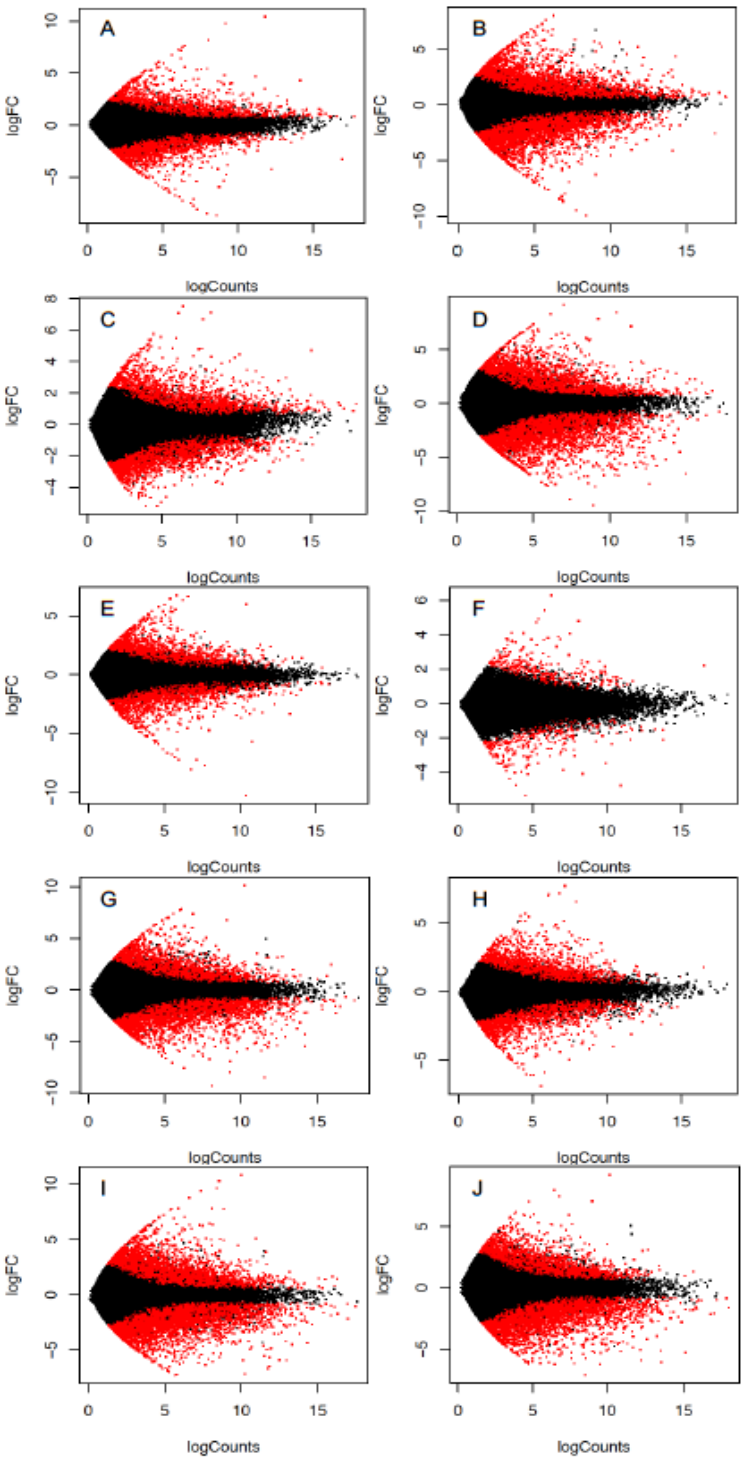
Appendix 29: Number of differentially expresses genes identified by DESeq for *Nannoperca australis*.

DESeq	KING	LIMA	MERTON	SEVEN	SPBR
KING	0				
LIMA	852	0			
MERTON	2238	737	0		
SEVEN	668	144	625	0	
SPBR	2484	1529	2817	1738	0

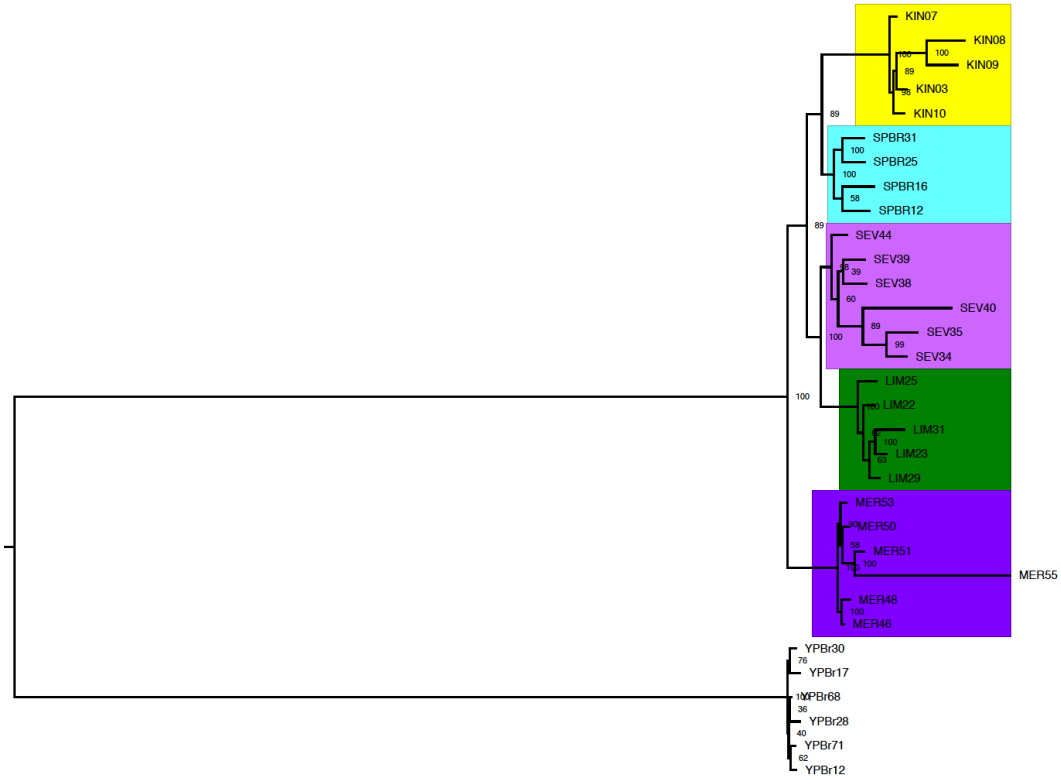
Appendix 30: Number of differentially expresses genes identified by edgeR for *Nannoperca australis*.

edgeR	KING	LIMA	MERTON	SEVEN	SPBR
KING	0				
LIMA	879	0			
MERTON	2547	820	0		
SEVEN	656	156	701	0	
SPBR	2434	1447	2987	1555	0

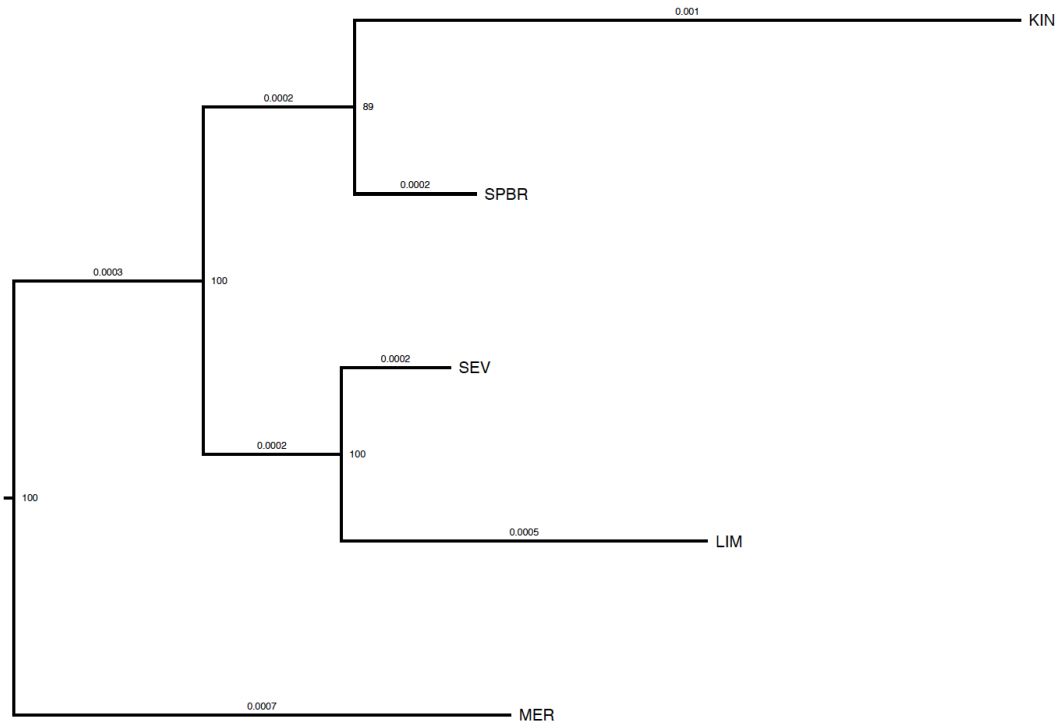
Appendix 31: Differential expression results for pairwise comparisons of *Nannoperca australis*. Plots depict $\log_2$ fold change in expression versus $\log_2$ mean expression counts for A) KIN–LIM, B) KIN–MER, C) KIN–SEV, D) KIN–SPBR, E) LIM–MER, F) LIM–SEV, G) LIM–SPBR, H) MER–SEV, I) MER–SPBR and J) SEV–SPBR. Genes identified as DE (FDR 5%) are highlighted in red.



Appendix 32: Maximum likelihood tree based on ddRAD sequences for *Nannoperca australis* (out-group *N. obscura*). Numbers in nodes are ML bootstrapping scores. Colours are based on those used in Figure 4.1.



Appendix 33: Maximum likelihood tree used as input phylogeny for the Expression Variance and Evolution model (*EVE*) gene expression analysis. Numbers in nodes are ML bootstrapping scores and branch labels are relative branch lengths.



Appendix 34: Functional annotation enrichment analysis of gene ontology terms assigned to 24 genes identified as showing high expression plasticity within lineages of *Nannoperca australis*. In total 107 terms were significant (P<0.05), however no terms remained significant at a FDR of 10%. DE ratio is the ratio of number of DE genes with a given GO term to the number of genes with that term in the entire catalogue. BP, MF and CC refer to Biological Process, Molecular Function and Cellular Component.

Category	DE ratio	*P* value	FDR	GO term
GO:0001824	2:21	0.0001	1	BP blastocyst development
GO:0003950	2:53	0.0006	1	MF NAD+ ADP-ribosyltransferase activity
GO:0021691	1:1	0.0007	1	BP cerebellar Purkinje cell layer maturation
GO:0070037	1:1	0.0007	1	MF rRNA (pseudouridine) methyltransferase activity
GO:0006997	2:64	0.0008	1	BP nucleus organization
GO:0007000	1:2	0.0013	1	BP nucleolus organization
GO:0017126	1:2	0.0013	1	BP nucleologenesis
GO:0016763	2:102	0.0020	1	MF transferase activity, transferring pentosyl groups
GO:0031616	1:4	0.0026	1	CC spindle pole centrosome
GO:0051903	1:5	0.0033	1	MF S-(hydroxymethyl)glutathione dehydrogenase activity
GO:0008330	1:5	0.0033	1	MF protein tyrosine/threonine phosphatase activity
GO:0031167	1:7	0.0046	1	BP rRNA methylation
GO:0000154	1:8	0.0052	1	BP rRNA modification
GO:0005635	2:173	0.0057	1	CC nuclear envelope
GO:0004735	1:9	0.0059	1	MF pyrroline-5-carboxylate reductase activity
GO:0005721	1:9	0.0059	1	CC pericentric heterochromatin
GO:0004022	1:9	0.0059	1	MF alcohol dehydrogenase (NAD) activity
GO:0031967	2:185	0.0065	1	CC organelle envelope
GO:0031975	2:185	0.0065	1	CC envelope
GO:2000279	1:10	0.0065	1	BP negative regulation of DNA biosynthetic process
GO:0006069	1:10	0.0065	1	BP ethanol oxidation
GO:0071695	1:11	0.0072	1	BP anatomical structure maturation
GO:0000059	1:11	0.0072	1	BP protein import into nucleus, docking
GO:0004726	1:12	0.0078	1	MF non-membrane spanning protein activity
GO:0008649	1:13	0.0085	1	MF rRNA methyltransferase activity
GO:0035970	1:13	0.0085	1	BP peptidyl-threonine dephosphorylation
GO:0051447	1:13	0.0085	1	BP negative regulation of meiotic cell cycle
GO:0055129	1:14	0.0091	1	BP L-proline biosynthetic process
GO:0090266	1:15	0.0098	1	BP regulation of mitotic cell cycle spindle assembly
GO:1903504	1:15	0.0098	1	BP regulation of mitotic spindle checkpoint
GO:0044615	1:15	0.0098	1	CC nuclear pore nuclear basket
GO:0007077	1:15	0.0098	1	BP mitotic nuclear envelope disassembly
GO:0030397	1:15	0.0098	1	BP membrane disassembly
GO:0051081	1:15	0.0098	1	BP nuclear envelope disassembly
GO:0031080	1:16	0.0104	1	CC nuclear pore outer ring
GO:0005315	1:16	0.0104	1	MF inorganic phosphate transmembrane transporter activity
GO:0006067	1:16	0.0104	1	BP ethanol metabolic process
GO:0030126	1:17	0.0111	1	CC COPI vesicle coat
GO:0033574	1:17	0.0111	1	BP response to testosterone
GO:0006561	1:17	0.0111	1	BP proline biosynthetic process
GO:0071850	1:18	0.0117	1	BP mitotic cell cycle arrest
GO:0006979	2:252	0.0118	1	BP response to oxidative stress
GO:0006890	1:19	0.0124	1	BP retrograde vesicle-mediated transport, Golgi to ER
GO:0090231	1:20	0.0130	1	BP regulation of spindle checkpoint
GO:0051292	1:20	0.0130	1	BP nuclear pore complex assembly

GO:0042405	1:21	0.0137	1	CC nuclear inclusion body
GO:0001706	1:21	0.0137	1	BP endoderm formation
GO:0034399	1:21	0.0137	1	CC nuclear periphery
GO:0017056	1:21	0.0137	1	MF structural constituent of nuclear pore
GO:0000188	1:21	0.0137	1	BP inactivation of MAPK activity
GO:0006817	1:21	0.0137	1	BP phosphate ion transport
GO:0006999	1:22	0.0143	1	BP nuclear pore organization
GO:0006560	1:22	0.0143	1	BP proline metabolic process
GO:2000278	1:22	0.0143	1	BP regulation of DNA biosynthetic process
GO:0051057	1:23	0.0149	1	BP regulation of small GTPase mediated signal transduction
GO:0006471	1:24	0.0156	1	BP protein ADP-ribosylation
GO:0051881	1:24	0.0156	1	BP regulation of mitochondrial membrane potential
GO:0006470	2:300	0.0163	1	BP protein dephosphorylation
GO:0046931	1:26	0.0169	1	BP pore complex assembly
GO:0004725	2:309	0.0173	1	MF protein tyrosine phosphatase activity
GO:0001510	1:28	0.0181	1	BP RNA methylation
GO:0042578	3:881	0.0192	1	MF phosphoric ester hydrolase activity
GO:0017017	1:30	0.0195	1	MF MAP kinase tyrosine/serine/threonine activity
GO:0033549	1:31	0.0201	1	MF MAP kinase phosphatase activity
GO:0051445	1:32	0.0208	1	BP regulation of meiotic cell cycle
GO:0006998	1:36	0.0233	1	BP nuclear envelope organization
GO:0002102	1:37	0.0239	1	CC podosome
GO:0016646	1:37	0.0240	1	MF oxidoreductase activity
GO:1901976	1:41	0.0265	1	BP regulation of cell cycle checkpoint
GO:0070373	1:43	0.0278	1	BP negative regulation of ERK1 and ERK2 cascade
GO:0016311	2:409	0.0291	1	BP dephosphorylation
GO:0051146	1:46	0.0297	1	BP striated muscle cell differentiation
GO:1901677	1:48	0.0310	1	MF phosphate transmembrane transporter activity
GO:0006996	5:2800	0.0323	1	BP organelle organization
GO:0034308	1:52	0.0335	1	BP primary alcohol metabolic process
GO:0004721	2:446	0.0341	1	MF phosphoprotein phosphatase activity
GO:0008173	1:54	0.0348	1	MF RNA methyltransferase activity
GO:0015758	1:55	0.0354	1	BP glucose transport
GO:0009084	1:55	0.0354	1	BP glutamine family amino acid biosynthetic process
GO:0043473	1:56	0.0358	1	BP pigmentation
GO:0000723	1:56	0.0360	1	BP telomere maintenance
GO:0006939	1:56	0.0361	1	BP smooth muscle contraction
GO:0032200	1:57	0.0367	1	BP telomere organization
GO:0016645	1:60	0.0385	1	MF oxidoreductase activity
GO:0022402	3:1165	0.0394	1	BP cell cycle process
GO:0051297	1:62	0.0398	1	BP centrosome organization
GO:0004435	1:62	0.0398	1	MF phosphatidylinositol phospholipase C activity
GO:0004629	1:62	0.0398	1	MF phospholipase C activity
GO:0043407	1:63	0.0403	1	BP negative regulation of MAP kinase activity
GO:0042246	1:64	0.0410	1	BP tissue regeneration
GO:0019843	1:65	0.0417	1	MF rRNA binding
GO:0008645	1:65	0.0417	1	BP hexose transport
GO:0016234	1:65	0.0417	1	CC inclusion body
GO:0016757	2:509	0.0433	1	MF transferase activity, transferring glycosyl groups
GO:0030120	1:68	0.0436	1	CC vesicle coat
GO:0032355	1:70	0.0449	1	BP response to estradiol
GO:0030131	1:71	0.0454	1	CC clathrin adaptor complex
GO:0001704	1:71	0.0455	1	BP formation of primary germ layer
GO:0015749	1:71	0.0455	1	BP monosaccharide transport
GO:0043408	2:524	0.0457	1	BP regulation of MAPK cascade
GO:0008344	1:73	0.0467	1	BP adult locomotory behavior
GO:0042542	1:74	0.0473	1	BP response to hydrogen peroxide
GO:0051384	1:74	0.0473	1	BP response to glucocorticoid
GO:1901654	1:74	0.0474	1	BP response to ketone
GO:0005643	1:75	0.0479	1	CC nuclear pore
GO:0030119	1:77	0.0492	1	CC AP-type membrane coat adaptor complex

GO:0010827	1:77	0.0492	1	BP regulation of glucose transport

Appendix 35: Functional annotation enrichment analysis of gene ontology terms assigned to 165 genes identified as showing divergent expression profiles among lineages of *Nannoperca australis*. In total 347 terms were significant (P<0.05), with 10 terms remaining significant at a FDR of 10% (highlighted in bold). DE ratio is the ratio of number of DE genes with a given GO term to the number of genes with that term in the entire catalogue. BP, MF and CC refer to Biological Process, Molecular Function and Cellular Component.

Category	DE ratio	*P* value	FDR	GO term
GO:0004523	**8:138**	**1.67E-09**	**2.72E-05**	**MF RNA-DNA hybrid ribonuclease activity**
GO:0016891	**8:161**	**2.11E-08**	**1.72E-04**	**MF endoribonuclease activity**
GO:0004521	**8:180**	**9.08E-08**	**4.92E-04**	**MF endoribonuclease activity**
GO:0016893	**8:222**	**3.47E-07**	**1.41E-03**	**MF endonuclease activity**
GO:0004540	**8:221**	**1.12E-06**	**3.66E-03**	**MF ribonuclease activity**
GO:0004190	**7:259**	**1.52E-05**	**3.53E-02**	**MF aspartic-type endopeptidase activity**
GO:0070001	**7:259**	**1.52E-05**	**3.53E-02**	**MF aspartic-type peptidase activity**
GO:0006979	**8:252**	**3.15E-05**	**6.41E-02**	**BP response to oxidative stress**
GO:0015074	**9:618**	**5.25E-05**	**9.00E-02**	**BP DNA integration**
GO:0039660	**3:39**	**5.53E-05**	**9.00E-02**	**MF structural constituent of virion**
GO:0020002	3:44	8.97E-05	1.33E-01	CC host cell plasma membrane
GO:0004519	8:473	1.17E-04	1.59E-01	MF endonuclease activity
GO:0033644	3:49	1.65E-04	2.07E-01	CC host cell membrane
GO:0044218	3:53	0.0002	0.2464	CC other organism cell membrane
GO:0044279	3:53	0.0002	0.2464	CC other organism membrane
GO:0009881	3:27	0.0003	0.2475	MF photoreceptor activity
GO:0004175	13:1011	0.0003	0.2475	MF endopeptidase activity
GO:0006369	3:27	0.0004	0.3559	BP termination of RNA polymerase II transcription
GO:0042542	4:74	0.0007	0.5387	BP response to hydrogen peroxide
GO:0031213	2:9	0.0007	0.5387	CC RSF complex
GO:0003964	5:317	0.0009	0.7064	MF RNA-directed DNA polymerase activity
GO:0033643	3:78	0.0010	0.7094	CC host cell part
GO:0019013	3:65	0.0012	0.8469	CC viral nucleocapsid
GO:0044217	3:83	0.0013	0.8633	CC other organism part
GO:0004518	8:595	0.0013	0.8776	MF nuclease activity
GO:0052695	2:15	0.0015	0.9393	BP cellular glucuronidation
GO:0034599	4:98	0.0017	0.9925	BP cellular response to oxidative stress
GO:0006353	3:52	0.0018	0.9925	BP DNA-templated transcription, termination
GO:0003913	2:12	0.0019	0.9925	MF DNA photolyase activity
GO:0006063	2:16	0.0019	0.9925	BP uronic acid metabolic process
GO:0019585	2:16	0.0019	0.9925	BP glucuronate metabolic process
GO:0016712	3:83	0.0020	0.9925	MF oxidoreductase activity, reduction of molecular oxygen
GO:0044423	3:73	0.0025	1	CC virion part
GO:0000302	4:116	0.0026	1	BP response to reactive oxygen species
GO:0030036	8:526	0.0035	1	BP actin cytoskeleton organization
GO:0070011	14:1422	0.0039	1	MF peptidase activity, acting on L-amino acid peptides
GO:0065007	78:13863	0.0041	1	BP biological regulation
GO:0031124	3:62	0.0043	1	BP mRNA 3'-end processing
GO:0048264	2:21	0.0045	1	BP determination of ventral identity
GO:0034061	5:383	0.0047	1	MF DNA polymerase activity
GO:0006805	3:71	0.0048	1	BP xenobiotic metabolic process
GO:0006711	1:3	0.0050	1	BP estrogen catabolic process
GO:0042447	1:3	0.0050	1	BP hormone catabolic process
GO:0006351	20:2449	0.0053	1	BP transcription, DNA-templated
GO:0097659	20:2449	0.0053	1	BP nucleic acid-templated transcription

GO:0035098	2:24	0.0053	1	CC ESC/E(Z) complex
GO:0008233	14:1476	0.0055	1	MF peptidase activity
GO:0051252	29:3974	0.0056	1	BP regulation of RNA metabolic process
GO:0043149	2:20	0.0058	1	BP stress fiber assembly
GO:0006355	28:3793	0.0060	1	BP regulation of transcription, DNA-templated
GO:0061036	2:28	0.0061	1	BP positive regulation of cartilage development
GO:0050789	74:13107	0.0061	1	BP regulation of biological process
GO:0042752	3:76	0.0062	1	BP regulation of circadian rhythm
GO:0097524	1:1	0.0062	1	CC sperm plasma membrane
GO:1903506	28:3809	0.0062	1	BP regulation of nucleic acid-templated transcription
GO:0042611	2:70	0.0063	1	CC MHC protein complex
GO:2001141	28:3815	0.0064	1	BP regulation of RNA biosynthetic process
GO:1901362	25:3373	0.0065	1	BP organic cyclic compound biosynthetic process
GO:0033882	1:1	0.0066	1	MF choloyl-CoA hydrolase activity
GO:0048385	2:29	0.0067	1	BP regulation of retinoic acid receptor signaling pathway
GO:0031010	2:24	0.0067	1	CC ISWI-type complex
GO:0030029	8:609	0.0069	1	BP actin filament-based process
GO:0006771	1:1	0.0071	1	BP riboflavin metabolic process
GO:0008531	1:1	0.0071	1	MF riboflavin kinase activity
GO:0009231	1:1	0.0071	1	BP riboflavin biosynthetic process
GO:0009398	1:1	0.0071	1	BP FMN biosynthetic process
GO:0033860	1:1	0.0071	1	BP regulation of NAD(P)H oxidase activity
GO:0033864	1:1	0.0071	1	BP positive regulation of NAD(P)H oxidase activity
GO:0046444	1:1	0.0071	1	BP FMN metabolic process
GO:0008150	133:26649	0.0072	1	BP biological_process
GO:0018298	2:27	0.0073	1	BP protein-chromophore linkage
GO:2000736	3:85	0.0073	1	BP regulation of stem cell differentiation
GO:0009649	2:30	0.0077	1	BP entrainment of circadian clock
GO:0001502	2:32	0.0080	1	BP cartilage condensation
GO:0098743	2:32	0.0080	1	BP cell aggregation
GO:0021502	1:1	0.0086	1	BP neural fold elevation formation
GO:0034465	1:1	0.0086	1	BP response to carbon monoxide
GO:0061419	1:1	0.0086	1	BP positive regulation of transcription in response to hypoxia
GO:0071245	1:1	0.0086	1	BP cellular response to carbon monoxide
GO:0071279	1:1	0.0086	1	BP cellular response to cobalt ion
GO:0008210	2:29	0.0087	1	BP estrogen metabolic process
GO:0004864	2:30	0.0087	1	MF protein phosphatase inhibitor activity
GO:0034654	23:3098	0.0088	1	BP nucleobase-containing compound biosynthetic process
GO:0042074	3:79	0.0089	1	BP cell migration involved in gastrulation
GO:0018931	1:2	0.0091	1	BP naphthalene metabolic process
GO:0018979	1:2	0.0091	1	BP trichloroethylene metabolic process
GO:0090420	1:2	0.0091	1	BP naphthalene-containing compound metabolic process
GO:0052815	1:3	0.0093	1	MF medium-chain acyl-CoA hydrolase activity
GO:0015020	2:39	0.0093	1	MF glucuronosyltransferase activity
GO:0072347	1:2	0.0096	1	BP response to anesthetic
GO:2000112	29:4126	0.0098	1	BP regulation of cellular biosynthetic process
GO:0032774	20:2600	0.0100	1	BP RNA biosynthetic process
GO:0018467	1:2	0.0103	1	MF formaldehyde dehydrogenase activity
GO:0061030	1:2	0.0105	1	BP epithelial cell differentiation
GO:0043620	2:33	0.0109	1	BP regulation of transcription in response to stress
GO:0033333	2:24	0.0111	1	BP fin development
GO:0031123	3:91	0.0114	1	BP RNA 3'-end processing
GO:0019212	2:33	0.0114	1	MF phosphatase inhibitor activity
GO:0004497	4:228	0.0117	1	MF monooxygenase activity
GO:0010556	29:4204	0.0121	1	BP regulation of macromolecule biosynthetic process
GO:0016779	6:541	0.0122	1	MF nucleotidyltransferase activity
GO:0019043	2:76	0.0126	1	BP establishment of viral latency
GO:0075713	2:76	0.0126	1	BP establishment of integrated proviral latency
GO:0070244	1:3	0.0129	1	BP negative regulation of thymocyte apoptotic process
GO:1901360	40:6579	0.0129	1	BP organic cyclic compound metabolic process
GO:0042065	1:1	0.0130	1	BP glial cell growth

GO:0042066	1:1	0.0130	1	BP perineurial glial growth
GO:0008088	2:42	0.0131	1	BP axon cargo transport
GO:0015117	1:2	0.0134	1	MF thiosulfate transmembrane transporter activity
GO:0015140	1:2	0.0134	1	MF malate transmembrane transporter activity
GO:0015709	1:2	0.0134	1	BP thiosulfate transport
GO:0006259	13:1685	0.0134	1	BP DNA metabolic process
GO:1902459	1:3	0.0138	1	BP positive regulation of stem cell population maintenance
GO:0019219	29:4263	0.0138	1	BP regulation of nucleobase compound metabolic process
GO:0019438	23:3226	0.0138	1	BP aromatic compound biosynthetic process
GO:0015671	1:5	0.0143	1	BP oxygen transport
GO:0005797	1:4	0.0144	1	CC Golgi medial cisterna
GO:0042726	1:2	0.0145	1	BP flavin-containing compound metabolic process
GO:0042727	1:2	0.0145	1	BP flavin-containing compound biosynthetic process
GO:0018130	23:3231	0.0146	1	BP heterocycle biosynthetic process
GO:0001501	3:125	0.0155	1	BP skeletal system development
GO:0015129	1:3	0.0155	1	MF lactate transmembrane transporter activity
GO:1901475	1:3	0.0155	1	BP pyruvate transmembrane transport
GO:0002248	1:2	0.0157	1	BP connective tissue inflammatory response, wound healing
GO:0034645	23:3331	0.0157	1	BP cellular macromolecule biosynthetic process
GO:0010990	1:3	0.0160	1	BP regulation of SMAD protein complex assembly
GO:0010991	1:3	0.0160	1	BP negative regulation of SMAD protein complex assembly
GO:0008670	1:4	0.0161	1	MF 2,4-dienoyl-CoA reductase (NADPH) activity
GO:0016757	7:509	0.0162	1	MF transferase activity, transferring glycosyl groups
GO:0071250	1:2	0.0164	1	BP cellular response to nitrite
GO:0080033	1:2	0.0164	1	BP response to nitrite
GO:2000286	1:7	0.0166	1	BP receptor internalization involved in signaling pathway
GO:0042694	1:4	0.0167	1	BP muscle cell fate specification
GO:0031326	29:4336	0.0168	1	BP regulation of cellular biosynthetic process
GO:0000138	1:5	0.0172	1	CC Golgi trans cisterna
GO:0033764	2:42	0.0172	1	MF steroid dehydrogenase activity
GO:2000738	2:49	0.0173	1	BP positive regulation of stem cell differentiation
GO:0014021	1:2	0.0173	1	BP secondary neural tube formation
GO:0060574	1:2	0.0175	1	BP intestinal epithelial cell maturation
GO:0000377	3:96	0.0176	1	BP RNA splicing
GO:0000398	3:96	0.0176	1	BP mRNA splicing, via spliceosome
GO:0001708	2:49	0.0179	1	BP cell fate specification
GO:0036157	1:2	0.0179	1	CC outer dynein arm
GO:0015727	1:5	0.0181	1	BP lactate transport
GO:0035873	1:5	0.0181	1	BP lactate transmembrane transport
GO:0006957	2:93	0.0182	1	BP complement activation, alternative pathway
GO:0061035	2:55	0.0183	1	BP regulation of cartilage development
GO:1900037	1:4	0.0183	1	BP regulation of cellular response to hypoxia
GO:0009889	29:4369	0.0185	1	BP regulation of biosynthetic process
GO:0019418	1:5	0.0186	1	BP sulfide oxidation
GO:0070221	1:5	0.0186	1	BP sulfide oxidation, using sulfide:quinone oxidoreductase
GO:0000103	1:3	0.0186	1	BP sulfate assimilation
GO:0004020	1:3	0.0186	1	MF adenylylsulfate kinase activity
GO:0004779	1:3	0.0186	1	MF sulfate adenylyltransferase activity
GO:0004781	1:3	0.0186	1	MF sulfate adenylyltransferase (ATP) activity
GO:0015141	1:3	0.0188	1	MF succinate transmembrane transporter activity
GO:0010870	1:2	0.0189	1	BP positive regulation of receptor biosynthetic process
GO:0006310	6:681	0.0190	1	BP DNA recombination
GO:0044271	23:3295	0.0190	1	BP cellular nitrogen compound biosynthetic process
GO:0070330	2:75	0.0192	1	MF aromatase activity
GO:0035097	3:124	0.0192	1	CC histone methyltransferase complex
GO:0043565	10:1083	0.0194	1	MF sequence-specific DNA binding
GO:0046847	2:40	0.0194	1	BP filopodium assembly
GO:0002069	1:3	0.0197	1	BP columnar/cuboidal epithelial cell maturation
GO:0071228	1:4	0.0202	1	BP cellular response to tumor cell
GO:0036158	1:4	0.0202	1	BP outer dynein arm assembly
GO:0019882	3:169	0.0204	1	BP antigen processing and presentation

GO:0031415	1:5	0.0205	1	CC NatA complex
GO:0020037	4:256	0.0205	1	MF heme binding
GO:0042196	1:4	0.0206	1	BP chlorinated hydrocarbon metabolic process
GO:0042197	1:4	0.0206	1	BP halogenated hydrocarbon metabolic process
GO:0000375	3:102	0.0206	1	BP RNA splicing, via transesterification reactions
GO:0001922	1:4	0.0209	1	BP B-1 B cell homeostasis
GO:0021538	1:5	0.0210	1	BP epithalamus development
GO:0043704	1:5	0.0210	1	BP photoreceptor cell fate specification
GO:0090304	31:4977	0.0215	1	BP nucleic acid metabolic process
GO:0005911	6:527	0.0218	1	CC cell-cell junction
GO:0046294	1:6	0.0219	1	BP formaldehyde catabolic process
GO:0051903	1:5	0.0221	1	MF S-(hydroxymethyl) glutathione dehydrogenase activity
GO:0005576	11:1332	0.0224	1	CC extracellular region
GO:0048736	2:53	0.0228	1	BP appendage development
GO:0060992	1:6	0.0228	1	BP response to fungicide
GO:0070243	1:6	0.0228	1	BP regulation of thymocyte apoptotic process
GO:0051171	29:4462	0.0229	1	BP regulation of nitrogen compound metabolic process
GO:0046292	1:7	0.0232	1	BP formaldehyde metabolic process
GO:0038033	1:3	0.0232	1	BP positive regulation of endothelial chemotaxis pathway
GO:1901727	1:3	0.0232	1	BP positive regulation of histone deacetylase activity
GO:0001071	12:1414	0.0233	1	MF nucleic acid binding transcription factor activity
GO:0003700	12:1414	0.0233	1	MF transcription factor activity, DNA binding
GO:0097070	1:5	0.0233	1	BP ductus arteriosus closure
GO:0003887	3:185	0.0233	1	MF DNA-directed DNA polymerase activity
GO:0036003	1:8	0.0235	1	BP positive regulation of transcription in response to stress
GO:0009058	32:5122	0.0236	1	BP biosynthetic process
GO:0042157	3:112	0.0237	1	BP lipoprotein metabolic process
GO:0050833	1:5	0.0238	1	MF pyruvate transmembrane transporter activity
GO:0010468	29:4459	0.0240	1	BP regulation of gene expression
GO:0018119	1:7	0.0240	1	BP peptidyl-cysteine S-nitrosylation
GO:0016229	2:54	0.0241	1	MF steroid dehydrogenase activity
GO:0010035	5:387	0.0241	1	BP response to inorganic substance
GO:0006139	36:6031	0.0243	1	BP nucleobase-containing compound metabolic process
GO:0021797	1:3	0.0243	1	BP forebrain anterior/posterior pattern specification
GO:1901725	1:4	0.0245	1	BP regulation of histone deacetylase activity
GO:0032364	1:5	0.0246	1	BP oxygen homeostasis
GO:0033483	1:5	0.0246	1	BP gas homeostasis
GO:2000370	1:4	0.0247	1	BP positive regulation of clathrin-mediated endocytosis
GO:0070060	1:3	0.0249	1	BP 'de novo' actin filament nucleation
GO:0004601	2:56	0.0249	1	MF peroxidase activity
GO:0090285	1:4	0.0249	1	BP negative regulation of protein glycosylation in Golgi
GO:0007567	1:5	0.0250	1	BP parturition
GO:0016938	1:3	0.0251	1	CC kinesin I complex
GO:0046906	4:271	0.0253	1	MF tetrapyrrole binding
GO:1901576	31:4972	0.0253	1	BP organic substance biosynthetic process
GO:0071439	1:5	0.0254	1	CC clathrin complex
GO:0006089	1:5	0.0255	1	BP lactate metabolic process
GO:0030260	2:89	0.0256	1	BP entry into host cell
GO:0044409	2:89	0.0256	1	BP entry into host
GO:0046718	2:89	0.0256	1	BP viral entry into host cell
GO:0051806	2:89	0.0256	1	BP symbiotic interaction with other organism
GO:0051828	2:89	0.0256	1	BP symbiotic interaction with other organism
GO:0052126	2:89	0.0256	1	BP movement in host environment
GO:0052192	2:89	0.0256	1	BP movement in environment of other organism
GO:0005034	1:2	0.0260	1	MF osmosensor activity
GO:0030054	11:1498	0.0260	1	CC cell junction
GO:0001667	4:223	0.0265	1	BP ameboidal-type cell migration
GO:0007010	10:1084	0.0267	1	BP cytoskeleton organization
GO:0006725	37:6290	0.0267	1	BP cellular aromatic compound metabolic process
GO:0000137	1:4	0.0275	1	CC Golgi cis cisterna
GO:0034708	3:140	0.0278	1	CC methyltransferase complex

GO:0090283	1:5	0.0281	1	BP regulation of protein glycosylation in Golgi
GO:0004022	1:9	0.0282	1	MF alcohol dehydrogenase (NAD) activity
GO:2001053	1:7	0.0282	1	BP regulation of mesenchymal cell apoptotic process
GO:2001054	1:7	0.0282	1	BP negative regulation of cell apoptotic process
GO:0019276	1:8	0.0283	1	BP UDP-N-acetylgalactosamine metabolic process
GO:0009059	23:3517	0.0283	1	BP macromolecule biosynthetic process
GO:0019896	1:7	0.0284	1	BP axon transport of mitochondrion
GO:0051216	3:133	0.0285	1	BP cartilage development
GO:0046548	1:11	0.0288	1	BP retinal rod cell development
GO:0032355	2:70	0.0289	1	BP response to estradiol
GO:0010573	1:5	0.0289	1	BP vascular endothelial growth factor production
GO:0043170	56:9805	0.0290	1	BP macromolecule metabolic process
GO:0044282	5:378	0.0294	1	BP small molecule catabolic process
GO:0003947	1:3	0.0295	1	MF o-glycan biosynthesis
GO:0017014	1:9	0.0295	1	BP protein nitrosylation
GO:0030866	2:64	0.0296	1	BP cortical actin cytoskeleton organization
GO:0051409	1:5	0.0296	1	BP response to nitrosative stress
GO:0004021	1:8	0.0297	1	MF L-alanine:2-oxoglutarate aminotransferase activity
GO:0047635	1:8	0.0297	1	MF alanine-oxo-acid transaminase activity
GO:0031032	3:144	0.0300	1	BP actomyosin structure organization
GO:0051541	1:4	0.0300	1	BP elastin metabolic process
GO:0097411	1:6	0.0301	1	BP hypoxia-inducible factor-1alpha signaling pathway
GO:0061298	1:8	0.0304	1	BP retina vasculature development in camera-type eye
GO:0060394	1:6	0.0306	1	BP negative regulation of protein phosphorylation
GO:0017034	1:7	0.0306	1	MF Rap guanyl-nucleotide exchange factor activity
GO:0004596	1:7	0.0307	1	MF peptide alpha-N-acetyltransferase activity
GO:0016684	2:61	0.0311	1	MF oxidoreductase activity, acting on peroxide as acceptor
GO:0006366	4:267	0.0312	1	BP transcription from RNA polymerase II promoter
GO:0090045	1:5	0.0312	1	BP positive regulation of deacetylase activity
GO:0006848	1:6	0.0316	1	BP pyruvate transport
GO:0038089	1:5	0.0318	1	BP positive regulation of cell migration
GO:0046581	1:7	0.0320	1	CC intercellular canaliculus
GO:0005575	128:27241	0.0321	1	CC cellular_component
GO:0030865	2:65	0.0323	1	BP cortical cytoskeleton organization
GO:0002070	1:9	0.0326	1	BP epithelial cell maturation
GO:0010717	2:55	0.0333	1	BP regulation of epithelial to mesenchymal transition
GO:0044723	9:931	0.0335	1	BP single-organism carbohydrate metabolic process
GO:0070493	1:6	0.0335	1	BP thrombin receptor signaling pathway
GO:0002347	1:9	0.0343	1	BP response to tumor cell
GO:0048665	1:10	0.0344	1	BP neuron fate specification
GO:0047484	1:4	0.0344	1	BP regulation of response to osmotic stress
GO:0045760	1:6	0.0347	1	BP positive regulation of action potential
GO:0031519	2:74	0.0347	1	CC PcG protein complex
GO:0050427	1:4	0.0348	1	BP 3'-phosphoadenosine 5'-phosphosulfate metabolism
GO:0050428	1:4	0.0348	1	BP 3'-phosphoadenosine 5'-phosphosulfate biosynthesis
GO:0044428	20:2746	0.0352	1	CC nuclear part
GO:0034035	1:5	0.0358	1	BP purine ribonucleoside bisphosphate metabolic process
GO:0034036	1:5	0.0358	1	BP purine ribonucleoside bisphosphate biosynthetic process
GO:0032352	1:8	0.0359	1	BP positive regulation of hormone metabolic process
GO:0046886	1:8	0.0359	1	BP positive regulation of hormone biosynthetic process
GO:0006069	1:10	0.0363	1	BP ethanol oxidation
GO:0045906	1:8	0.0363	1	BP negative regulation of vasoconstriction
GO:0004857	5:422	0.0364	1	MF enzyme inhibitor activity
GO:0032320	4:232	0.0364	1	none
GO:0015114	1:9	0.0365	1	MF phosphate ion transmembrane transporter activity
GO:0071257	1:6	0.0365	1	BP cellular response to electrical stimulus
GO:0070654	1:6	0.0367	1	BP sensory epithelium regeneration
GO:1990399	1:6	0.0367	1	BP epithelium regeneration
GO:0060255	34:5639	0.0370	1	BP regulation of macromolecule metabolic process
GO:0010996	1:6	0.0371	1	BP response to auditory stimulus
GO:0006958	2:144	0.0373	1	BP complement activation, classical pathway

GO:0022617	2:75	0.0373	1	BP extracellular matrix disassembly
GO:0016290	1:8	0.0375	1	MF palmitoyl-CoA hydrolase activity
GO:0004514	1:5	0.0377	1	MF nicotinate-nucleotide diphosphorylase activity
GO:0034212	1:8	0.0383	1	MF peptide N-acetyltransferase activity
GO:0002092	1:15	0.0388	1	BP positive regulation of receptor internalization
GO:0010008	5:403	0.0390	1	CC endosome membrane
GO:0015232	1:7	0.0395	1	MF heme transporter activity
GO:0060051	1:7	0.0396	1	BP negative regulation of protein glycosylation
GO:0006406	2:48	0.0397	1	BP mRNA export from nucleus
GO:0070358	1:9	0.0397	1	BP actin polymerization-dependent cell motility
GO:0044710	37:6275	0.0403	1	BP single-organism metabolic process
GO:0071704	70:13000	0.0409	1	BP organic substance metabolic process
GO:0046483	36:6266	0.0413	1	BP heterocycle metabolic process
GO:0002504	2:75	0.0413	1	BP antigen processing of antigen via MHC class II
GO:0072091	2:82	0.0416	1	BP regulation of stem cell proliferation
GO:0060027	2:53	0.0417	1	BP convergent extension involved in gastrulation
GO:0042851	1:12	0.0423	1	BP L-alanine metabolic process
GO:0042853	1:12	0.0423	1	BP L-alanine catabolic process
GO:0004222	4:291	0.0427	1	MF metalloendopeptidase activity
GO:0030010	2:83	0.0430	1	BP establishment of cell polarity
GO:0019222	39:6695	0.0431	1	BP regulation of metabolic process
GO:0051017	2:77	0.0431	1	BP actin filament bundle assembly
GO:0061572	2:77	0.0431	1	BP actin filament bundle organization
GO:0004114	2:75	0.0432	1	MF 3',5'-cyclic-nucleotide phosphodiesterase activity
GO:0006524	1:13	0.0433	1	BP alanine catabolic process
GO:0009080	1:13	0.0433	1	BP pyruvate family amino acid catabolic process
GO:0034641	37:6459	0.0435	1	BP cellular nitrogen compound metabolic process
GO:0070286	1:10	0.0435	1	BP axonemal dynein complex assembly
GO:0006405	2:52	0.0438	1	BP RNA export from nucleus
GO:0006522	1:14	0.0444	1	BP alanine metabolic process
GO:0009078	1:14	0.0444	1	BP pyruvate family amino acid metabolic process
GO:0044249	29:4813	0.0450	1	BP cellular biosynthetic process
GO:0004112	2:76	0.0452	1	MF cyclic-nucleotide phosphodiesterase activity
GO:0038084	1:7	0.0453	1	BP vascular endothelial growth factor signaling pathway
GO:0044440	5:419	0.0455	1	CC endosomal part
GO:0070233	1:10	0.0458	1	BP negative regulation of T cell apoptotic process
GO:0042613	1:25	0.0461	1	CC MHC class II protein complex
GO:0060644	1:7	0.0462	1	BP mammary gland epithelial cell differentiation
GO:0030742	1:7	0.0465	1	MF GTP-dependent protein binding
GO:0031012	5:505	0.0468	1	CC extracellular matrix
GO:0006067	1:16	0.0469	1	BP ethanol metabolic process
GO:0001527	1:7	0.0473	1	CC microfibril
GO:0046976	1:6	0.0474	1	MF histone methyltransferase activity (H3-K27 specific)
GO:0098900	1:8	0.0474	1	BP regulation of action potential
GO:0043619	1:16	0.0477	1	BP regulation of transcription in response to oxidative stress
GO:0010561	1:8	0.0479	1	BP negative regulation of glycoprotein biosynthetic process
GO:0035089	1:11	0.0481	1	BP establishment of apical/basal cell polarity
GO:0035774	1:12	0.0481	1	BP regulation of insulin secretion
GO:0048251	1:13	0.0481	1	BP elastic fiber assembly
GO:0016404	1:16	0.0488	1	MF 15-hydroxyprostaglandin dehydrogenase (NAD+) activity
GO:0003206	1:15	0.0491	1	BP cardiac chamber morphogenesis
GO:0003208	1:15	0.0491	1	BP cardiac ventricle morphogenesis
GO:0080090	33:5573	0.0492	1	BP regulation of primary metabolic process
GO:0015556	1:11	0.0495	1	MF C4-dicarboxylate transmembrane transporter activity
GO:0070402	1:16	0.0495	1	MF NADPH binding
GO:0035360	1:7	0.0495	1	BP regulation of peroxisome receptor signaling pathway
GO:0048318	1:10	0.0496	1	BP axial mesoderm development
GO:0042541	1:8	0.0497	1	BP hemoglobin biosynthetic process

References

Aitken SN, Whitlock MC (2013) Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics* **44**, 367–388.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews: Genetics* **11**, 697–709.

Alvarez M, Schrey AW, Richards CL (2015) Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular Ecology* **24**, 710–725.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81–92.

Arthington AH, Pusey BJ (2003) Flow restoration and protection in Australian rivers. *River Research and Applications* **19**, 377–395.

Attard C, Brauer C, van Zoelen J*, et al.* (2016a) Multi-generational evaluation of genetic diversity and parentage in captive southern pygmy perch (*Nannoperca australis*). *Conservation Genetics*, doi:10.1007/s10592–10016–10873–y.

Attard C, Möller L, Sasaki M*, et al.* (2016b) A novel holistic framework for genetic-based captive-breeding and reintroduction programs. *Conservation Biology*, doi:10.1111/cobi.12699.

Avise J (2010) Perspective: conservation genetics enters the genomics era. *Conservation Genetics* **11**, 665–669.

Baillon L, Pierron F, Coudret R*, et al.* (2015) Transcriptome profile analysis reveals specific signatures of pollutants in Atlantic eels. *Ecotoxicology* **24**, 71–84.

Balcombe SR, Sheldon F, Capon SJ*, et al.* (2011) Climate-change threats to native fish in degraded rivers and floodplains of the Murray-Darling Basin, Australia. *Marine and Freshwater Research* **62**, 1099–1114.

Balkenhol N, Landguth EL (2011) Simulation modelling in landscape genetics: on the need to go further. *Molecular Ecology* **20**, 667–670.

Barnosky AD, Matzke N, Tomiya S*, et al.* (2011) Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**, 38–44.

Barson N, Cable J, Van Oosterhout C (2009) Population genetic analysis of microsatellite variation of guppies (*Poecilia reticulata*) in Trinidad and Tobago: evidence for a dynamic source–sink metapopulation structure, founder events and population bottlenecks. *Journal of Evolutionary Biology* **22**, 485–497.

Bateman A, Coin L, Durbin R*, et al.* (2004) The Pfam protein families database. *Nucleic Acids Research* **32**, D138–D141.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**, 1619–1626.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters* **15**, 365–377.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* **7**, 431–452.

Bice C, Whiterod N, Wilson P, Zampatti B, Hammer M (2012) *The Critical Fish Habitat Project: Reintroductions of Threatened Fish Species in the Coorong, Lower Lakes and Murray Mouth Region in 2011/12* SARDI Aquatic Sciences.

Bijlsma R, Loeschcke V (2012) Genetic erosion impedes adaptive responses to stressful environments. *Evolutionary Applications* **5**, 117–129.

Blanquart F, Gandon S, Nuismer S (2012) The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology* **25**, 1351–1363.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

Bond N, Thomson J, Reich P, Stein J (2011) Using species distribution models to infer potential climate change-induced range shifts of freshwater fish in south-eastern Australia. *Marine and Freshwater Research* **62**, 1043–1061.

Botstein D, Cherry JM, Ashburner M*, et al.* (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29.

Bourret V, Dionne M, Bernatchez L (2014) Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular Ecology* **23**, 4444–4457.

Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L (2013) Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene–environment interactions driving local adaptation. *Evolution* **67**, 3469–3487.

Bozinovic G, Oleksiak MF (2011) Genomic approaches with natural fish populations from polluted environments. *Environmental Toxicology and Chemistry* **30**, 283–289.

Bracamonte SE, Smith S, Hammer M*, et al.* (2015) Characterization of MHC class IIB for four endangered Australian freshwater fishes obtained from ecologically divergent populations. *Fish & Shellfish Immunology* **46**, 468–476.

Bradshaw WE, Holzapfel CM (2006) Evolutionary response to rapid climate change. *Science(Washington)* **312**, 1477–1478.

Brauer C, Hammer M, Beheregaray L (2016) Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology*, DOI:10.1111/mec.13830.

Brauer CJ, Unmack PJ, Hammer MP, Adams M, Beheregaray LB (2013) Catchment-scale conservation units identified for the threatened Yarra pygmy perch (*Nannoperca obscura*) in highly modified river systems. *PLoS ONE* **8**, e82953.

Brown AR, Owen SF, Peters J*, et al.* (2015) Climate change and pollution speed declines in zebrafish populations. *Proceedings of the National Academy of Sciences* **112**, E1237–E1246.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249–256.

Buckley DB, Klaassen CD (2007) Tissue-and gender-specific mRNA expression of UDP-glucuronosyltransferases (UGTs) in mice. *Drug Metabolism and Disposition* **35**, 121–127.

Burkhead NM (2012) Extinction rates in North American freshwater fishes, 1900–2010. *Bioscience* **62**, 798–808.

Burnett KG, Bain LJ, Baldwin WS*, et al.* (2007) Fundulus as the premier teleost model in environmental biology: Opportunities for new insights using genomics. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **2**, 257–286.

Cahill AE, Aiello-Lammens ME, Fisher-Reid MC*, et al.* (2012) How does climate change cause extinction? *Proceedings of the Royal Society of London B: Biological Sciences*, doi:10.1098/rspb.2012.1890.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171–182.

Charmantier A, McCleery RH, Cole LR*, et al.* (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800–803.

Chevin L-M, Gallet R, Gomulkiewicz R, Holt RD, Fellous S (2013) Phenotypic plasticity in evolutionary rescue experiments. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**.

Chevin L-M, Lande R, Mace GM (2010) Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. *PLoS Biology* **8**, e1000357.

Chevin LM, Lande R (2011) Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology* **24**, 1462–1476.

Chiew F, Teng J, Kirono D*, et al.* (2008) *Climate Data for Hydrologic Scenario Modelling Across the Murray-Darling Basin: A Report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project* CSIRO Canberra, Australia.

Chong Z, Ruan J, Wu C-I (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics* **28**, 2732–2737.

Cole T, Hammer M, Unmack P*, et al.* (2016) Range-wide fragmentation in a threatened fish associated with post-European settlement modification in the Murray-Darling Basin, Australia. *Conservation Genetics*, doi:10.1007/s10592–10016–10868–10598.

Coleman RA, Weeks AR, Hoffmann AA (2013) Balancing genetic uniqueness and genetic variation in determining conservation and translocation strategies: a comprehensive case study of threatened dwarf galaxias, *Galaxiella pusilla* (Mack) (Pisces: Galaxiidae). *Molecular Ecology* **22**, 1820–1835.

Conesa A, Götz S, García-Gómez JM*, et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.

Cook BD, Bunn SE, Hughes JM (2007) Molecular genetic and stable isotope signatures reveal complementary patterns of population connectivity in the regionally vulnerable southern pygmy perch (*Nannoperca australis*). *Biological Conservation* **138**, 60–72.

Cooke GM, Chao NL, Beheregaray LB (2012a) Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. *Molecular Ecology* **21**, 2410–2427.

Cooke GM, Chao NL, Beheregaray LB (2012b) Natural selection in the water: freshwater invasion and adaptation by water colour in the Amazonian pufferfish. *Journal of Evolutionary Biology* **25**, 1305–1320.

Cooke GM, Landguth EL, Beheregaray LB (2014) Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. *Evolution* **68**, 1947–1960.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423.

Crisci JL, Dean MD, Ralph P (2016) Adaptation in isolated populations: when does it happen and when can we tell? *Molecular Ecology*, DOI: 10.1111/mec.13729.

Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* **15**, 49–62.

Crook DA, Reich P, Bond NR*, et al.* (2010) Using biological information to support proactive strategies for managing freshwater fish during drought. *Marine and Freshwater Research* **61**, 379–387.

Cunningham M, Moritz C (1998) Genetic effects of forest fragmentation on a rainforest restricted lizard (Scincidae: Gnypetoscincus queenslandiae). *Biological Conservation* **83**, 19–30.

Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.

Davis J, O'Grady AP, Dale A*, et al.* (2015) When trends intersect: the challenge of protecting freshwater ecosystems under multiple land use and hydrological intensification scenarios. *Science of the Total Environment* **534**, 65–78.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673–679.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science* **332**, 53–58.

Dayan DI, Crawford DL, Oleksiak MF (2015) Phenotypic plasticity in gene expression contributes to divergence of locally adapted populations of Fundulus heteroclitus. *Molecular Ecology* **24**, 3345–3359.

DeBiasse MB, Kelly MW (2016) Plastic and Evolved Responses to Global Change: What Can We Learn from Comparative Transcriptomics? *Journal of Heredity* **107**, 71–81.

DeWitt TJ, Sih A, Wilson DS (1998) Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* **13**, 77–81.

DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* **56**, 461–473.

Do C, Waples RS, Peel D*, et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (*N*e) from genetic data. *Molecular Ecology Resources* **14**, 209–214.

Donders TH, Haberle SG, Hope G, Wagner F, Visscher H (2007) Pollen evidence for the transition of the Eastern Australian climate system from the post-glacial to the present-day ENSO mode. *Quaternary Science Reviews* **26**, 1621–1637.

Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* **22**, 1–20.

Driscoll DA, Hardy CM (2005) Dispersal and phylogeography of the agamid lizard Amphibolurus nobbi in fragmented and continuous habitat. *Molecular Ecology* **14**, 1613–1629.

Dudgeon D, Arthington AH, Gessner MO*, et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* **81**, 163–182.

Dunn CW, Luo X, Wu Z (2013) Phylogenetic Analysis of Gene Expression. *Integrative and Comparative Biology* **53**, 847–856.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* **19**, 3746–3759.

Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, btu121.

Edmands S (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology* **16**, 463–475.

Epperson BK, McRae BH, Scribner K*, et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology* **19**, 3549–3564.

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology* **24**, 6021–6040.

Epps CW, Palsbøll PJ, Wehausen JD*, et al.* (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters* **8**, 1029–1038.

ESRI (2012) ArcMap 10.1. Environmental Systems Research Institute, Inc., Redlands.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285–298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Fagan WF (2002) Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology* **83**, 3243–3249.

Faulks LK, Gilligan DM, Beheregaray LB (2010a) Evolution and maintenance of divergent lineages in an endangered freshwater fish, *Macquaria australasica*. *Conservation Genetics* **11**, 921–934.

Faulks LK, Gilligan DM, Beheregaray LB (2010b) Islands of water in a sea of dry land: hydrological regime predicts genetic diversity and dispersal in a widespread fish from Australia's arid zone, the golden perch (*Macquaria ambigua*). *Molecular Ecology* **19**, 4723–4737.

Faulks LK, Gilligan DM, Beheregaray LB (2011) The role of anthropogenic vs. natural in-stream structures in determining connectivity and genetic diversity in an endangered freshwater fish, Macquarie perch (*Macquaria australasica*). *Evolutionary Applications* **4**, 589–601.

Ferchaud A-L, Hansen MM (2016) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: three-spine sticklebacks in divergent environments. *Molecular Ecology* **25**, 238–259.

Finkel T, Holbrook NJ (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247.

Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography* **16**, 265–280.

Foley JA, DeFries R, Asner GP*, et al.* (2005) Global Consequences of Land Use. *Science* **309**, 570–574.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**, 875–891.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics* **180**, 977–993.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2015) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology* **25**, 104–120.

Fountain T, Nieminen M, Sirén J, Wong SC, Hanski I (2016) Predictable allele frequency changes due to habitat fragmentation in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences* **113**, 2678-2683.

Fourcade Y, Chaput-Bardy A, Secondi J, Fleurant C, Lemaire C (2013) Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology* **22**, 2065–2073.

Frankham R (1995) Conservation Genetics. *Annual Review of Genetics* **29**, 305–327.

Frankham R (2005) Genetics and extinction. *Biological Conservation* **126**, 131–140.

Frankham R (2010) Where are we in conservation genetics and where do we need to go? *Conservation Genetics* **11**, 661–663.

Frankham R (2015) Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology* **24**, 2610–2618.

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation* **170**, 56–63.

Fraser DJ, Debes PV, Bernatchez L, Hutchings JA (2014) Population size, habitat fragmentation, and the nature of adaptive variation in a stream fish. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, doi.org/10.1098/rspb.2014.0370.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**, 1687–1699.

Funk WC, Lovich RE, Hohenlohe PA*, et al.* (2016) Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Molecular Ecology* **25**, 2176–2194.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489–496.

Garner BA, Hand BK, Amish SJ*, et al.* (2015) Genomics in conservation: case studies and bridging the gap between data and application. *Trends in Ecology & Evolution* **31**, 81–83.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Gavin M, Brooke D, Howe P (1996) *Environmental hazard assessment: Naphthalene* Great Britain, Department of the Environment, Chemical and Biotechnology Division.

Geoscience Australia (2011) *National surface water information.* http://www.ga.gov.au/topographic-mapping/national-surface-water-information.html

Ghalambor CK, Hoke KL, Ruell EW*, et al.* (2015) Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*.

Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in Genetics* **21**, 616–623.

Gienapp P, Teplitsky C, Alho JS, Mills JA, Merila J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology* **17**, 167–178.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**, 184–186.

Gouskov A, Reyes M, Wirthner-Bitterlin L, Vorburger C (2016) Fish population genetic structure shaped by hydroelectric power plants in the upper Rhine catchment. *Evolutionary Applications* **9**, 394–408.

Grabherr MG, Haas BJ, Yassour M*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652.

Guillot G, Vitalis R, le Rouzic A, Gautier M (2014) Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spatial Statistics* **8**, 145–155.

Haas BJ, Papanicolaou A, Yassour M*, et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494–1512.

Hammer MP, Bice CM, Hall A*, et al.* (2013) Freshwater fish conservation in the face of critical water shortages in the southern Murray–Darling Basin, Australia. *Marine and Freshwater Research* **64**, 807–821.

Hand BK, Muhlfeld CC, Wade AA*, et al.* (2015) Climate variables explain neutral and adaptive variation within salmonid metapopulations: the importance of replication in landscape genetics. *Molecular Ecology* **25**, 689–705.

Hansen MM, Limborg MT, Ferchaud A-L, Pujolar J-M (2014) The effects of Medieval dams on genetic divergence and demographic history in brown trout populations. *BMC Evolutionary Biology* **14**, 1.

Hanski I (1998) Metapopulation dynamics. *Nature* **396**, 41–49.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* **7**, 1008–1025.

Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology* **24**, 5573–5595.

Hedrick P (2005) 'Genetic restoration': a more comprehensive perspective than ‘genetic rescue'. *Trends in Ecology & Evolution* **20**, 109–109.

Hedrick PW (1995) Gene flow and genetic restoration: the Florida panther as a case study. *Conservation Biology* **9**, 996–1007.

Hendry AP, Farrugia TJ, Kinnison MT (2008) Human influences on rates of phenotypic change in wild animal populations. *Molecular Ecology* **17**, 20–29.

Hermes-Lima M, Zenteno-Savín T (2002) Animal response to drastic changes in oxygen availability and physiological oxidative stress. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **133**, 537–556.

Hermisson J, Pennings PS (2005) Soft Sweeps. *Genetics* **169**, 2335–2352.

Hoban S, Bertorelle G, Gaggiotti OE (2012) Computer simulations: tools for population and evolutionary genetics. *Nature Reviews: Genetics* **13**, 110–122.

Hoffmann AA, Sgro CM (2011) Climate change and evolutionary adaptation. *Nature* **470**, 479–485.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews: Genetics* **9**, 421–432.

Horreo JL, Martinez JL, Ayllon F*, et al.* (2011) Impact of habitat fragmentation on the genetics of populations in dendritic landscapes. *Freshwater Biology* **56**, 2567–2579.

Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. *Nature Reviews Genetics* **11**, 855–866.

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Huey JA, Baker AM, Hughes JM (2008) The effect of landscape processes upon gene flow and genetic diversity in an Australian freshwater fish, *Neosilurus hyrtlii*. *Freshwater Biology* **53**, 1393–1408.

Hughes JM, Schmidt DJ, Finn DS (2009) Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. *Bioscience* **59**, 573–583.

Humphries P (1995) Life history, food and habitat of southern pygmy perch, Nannoperca australis, in the Macquarie River, Tasmania. *Marine and Freshwater Research* **46**, 1159–1169.

Jackson RB, Carpenter SR, Dahm CN*, et al.* (2001) Water in a changing world. *Ecological Applications* **11**, 1027–1045.

Joost S, Bonin A, Bruford MW*, et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955–3969.

Kalinowski ST, Meeuwig MH, Narum SR, Taper ML (2008) Stream trees: a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. *Canadian Journal of Fisheries and Aquatic Sciences* **65**, 2752–2760.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the american statistical association* **90**, 773–795.

Keller I, Nentwig W, LargiadÈR CR (2004) Recent habitat fragmentation due to roads can lead to significant genetic differentiation in an abundant flightless ground beetle. *Molecular Ecology* **13**, 2983–2994.

Kennard MJ, Pusey BJ, Olden JD*, et al.* (2010) Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology* **55**, 171–193.

Kershaw P, Moss P, Van Der Kaars S (2003) Causes and consequences of long-term climatic variability on the Australian continent. *Freshwater Biology* **48**, 1274–1283.

Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology* **85**, 1049–1064.

Khaitovich P, Weiss G, Lachmann M*, et al.* (2004) A Neutral Model of Transcriptome Evolution. *PLoS Biology* **2**, e132.

Kidd KA, Blanchfield PJ, Mills KH*, et al.* (2007) Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National Academy of Sciences* **104**, 8897–8901.

Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary fisherian life-history evolution in small salmonid populations. *Nature* **419**, 826–830.

Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567–580.

Lande R (1993) Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *American Naturalist*, 911–927.

Lande R (1998) Anthropogenic, ecological and genetic factors in extinction and conservation. *Researches on Population Ecology* **40**, 259–269.

Landguth EL, Cushman SA, Schwartz MK*, et al.* (2010) Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* **19**, 4179–4191.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, 1.

Laporte M, Pavey SA, Rougeux C*, et al.* (2015) RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology* **25**, 219–237.

Lasky JR, Des Marais DL, McKay J*, et al.* (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology* **21**, 5512–5529.

Laurance WF, Dell B, Turton SM*, et al.* (2011) The 10 Australian ecosystems most vulnerable to tipping points. *Biological Conservation* **144**, 1472–1480.

Lázaro-Nogal A, Matesanz S, Godoy A*, et al.* (2015) Environmental heterogeneity leads to higher plasticity in dry-edge populations of a semi-arid Chilean shrub: insights into climate change responses. *Journal of Ecology* **103**, 338–350.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* **21**, 1548–1566.

Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software* **25**, 1–18.

Leder EH, McCairns RS, Leinonen T*, et al.* (2015) The evolution and adaptive potential of transcriptional variation in sticklebacks—signatures of selection and widespread heritability. *Molecular Biology and Evolution* **32**, 674–689.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195.

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*.

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595.

Li W, Jaroszewski L, Godzik A (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**, 282–283.

Liermann CR, Nilsson C, Robertson J, Ng RY (2012) Implications of Dam Obstruction for Global Freshwater Fish Diversity. *Bioscience* **62**, 539–548.

Limborg MT, Helyar SJ, De Bruyn M*, et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology* **21**, 3686–3703.

Lintermans M (2007) *Fishes of the Murray-Darling Basin: an introductory guide* Murray-Darling Basin Commission Canberra.

Llewellyn L (1974) Spawning, development and distribution of the southern pigmy perch *Nannoperca australis australis* Gunther from inland waters in eastern Australia. *Marine and Freshwater Research* **25**, 121–149.

Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O (2009) Joint effects of inbreeding and local adaptation on the evolution of genetic load after fragmentation. *Conservation Biology* **23**, 1618–1627.

Losos JB, Arnold SJ, Bejerano G*, et al.* (2013) Evolutionary biology for the 21st century. *PLoS Biology* **11**, e1001466.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* **23**, 2178–2192.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 1.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews: Genetics* **4**, 981–994.

Manel S, Joost S, Epperson BK*, et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**, 3760–3772.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189–197.

Martínez-Arias R, Calafell F, Mateu E*, et al.* (2001) Sequence variability of a human pseudogene. *Genome Research* **11**, 1071–1085.

Martins EP, Diniz-Filho JA, Housworth EA (2002) Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution* **56**, 1–13.

Matala AP, Ackerman MW, Campbell MR, Narum SR (2014) Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. *Evolutionary Applications* **7**, 682–701.

Matthews WJ, Marsh-Matthews E (2007) Extirpation of red shiner in direct tributaries of Lake Texoma (Oklahoma-Texas): a cautionary case history from a fragmented river-reservoir system. *Transactions of the American Fisheries Society* **136**, 1041–1062.

McCairns RJS, Smith S, Sasaki M, Bernatchez L, Beheregaray LB (2016) The adaptive potential of subtropical rainbowfish in the face of climate change: heritability and heritable plasticity for the expression of candidate genes. *Evolutionary Applications*, n/a–n/a.

McCraney WT, Goldsmith G, Jacobs DK, Kinziger AP (2010) Rampant drift in artificially fragmented populations of the endangered tidewater goby (Eucyclogobius newberryi). *Molecular Ecology* **19**, 3315–3327.

McGuigan K, Chenoweth SF, Blows MW (2005) Phenotypic divergence along lines of genetic variance. *The American Naturalist* **165**, 32–43.

McGuigan K, Franklin CE, Moritz C, Blows MW (2003) Adaptation of rainbow fish to lake and stream habitats. *Evolution* **57**, 104–118.

McGuigan K, Nishimura N, Currey M, Hurwit D, Cresko WA (2011) Cryptic genetic variation and body size evolution in threespine stickleback. *Evolution* **65**, 1203–1211.

McKenzie-Smith F, Tiller D, Allen D (1994) Organochlorine pesticide residues in water and sediments from the Ovens and King rivers, north-east Victoria, Australia. *Archives of Environmental Contamination and Toxicology* **26**, 483–490.

McLaren S, Wallace MW, Gallagher SJ*, et al.* (2011) Palaeogeographic, climatic and tectonic change in southeastern Australia: the Late Neogene evolution of the Murray Basin. *Quaternary Science Reviews* **30**, 1086–1111.

Meech R, Mackenzie PI (1997) Structure and function of uridine diphosphate glucuronosyltransferases. *Clinical and Experimental Pharmacology and Physiology* **24**, 907–915.

Meirmans PG (2011) Marlin, software to create, run, and analyse spatially realistic simulations. *Molecular Ecology Resources* **11**, 146–150.

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology* **24**, 3223–3231.

Meirmans PG, Van Tienderen PH (2004) GenoType and GenoDive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792–794.

Meldgaard T, Nielsen EE, Loeschcke V (2003) Fragmentation by weirs in a riverine system: A study of genetic variation in time and space among populations of European grayling (Thymallus thymallus) in a Danish river system. *Conservation Genetics* **4**, 735–747.

Mergeay J, Santamaria L (2012) Evolution and Biodiversity: the evolutionary basis of biodiversity and its potential for adaptation to global change. *Evolutionary Applications* **5**, 103–106.

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications* **7**, 1–14.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* **28**, 659–669.

Mommsen TP (2004) Salmon spawning migration and muscle protein metabolism: the August Krogh principle at work. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **139**, 383–400.

Moore J-S, Bourret V, Dionne M*, et al.* (2014) Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology* **23**, 5680–5697.

Morrissey MB, de Kerckhove DT (2009) The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *The American Naturalist* **174**, 875–889.

Morrongiello JR, Beatty SJ, Bennett JC*, et al.* (2011a) Climate change and its implications for Australia's freshwater fish. *Marine and Freshwater Research* **62**, 1082–1098.

Morrongiello JR, Bond NR, Crook DA, Wong B (2011b) Eucalyptus leachate inhibits reproduction in a freshwater fish. *Freshwater Biology* **56**, 1736–1745.

Morrongiello JR, Bond NR, Crook DA, Wong B (2013) Intraspecific variation in the growth and survival of juvenile fish exposed to *Eucalyptus* leachate. *Ecology and Evolution* **3**, 3855–3867.

Morrongiello JR, Bond NR, Crook DA, Wong BBM (2010) Nuptial coloration varies with ambient light environment in a freshwater fish. *Journal of Evolutionary Biology* **23**, 2718–2725.

Morrongiello JR, Bond NR, Crook DA, Wong BBM (2012) Spatial variation in egg size and egg number reflects trade-offs and bet-hedging in a freshwater fish. *Journal of Animal Ecology* **81**, 806–817.

Murray–Darling Basin Authority (2013) *Murray-Darling Basin Weir Information System.* http://asdd.ga.gov.au/asdd/profileinfo/anzlic-jurisdic.xml - anzlic-jurisdic

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**, 2841–2847.

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics* **24**, 1552–1553.

Nilsson C, Reidy CA, Dynesius M, Revenga C (2005) Fragmentation and Flow Regulation of the World's Large River Systems. *Science* **308**, 405–408.

Nilsson-Örtman V, Stoks R, De Block M, Johansson F (2012) Generalists and specialists along a latitudinal transect: patterns of thermal adaptation in six species of damselflies. *Ecology* **93**, 1340–1352.

Oksanen J, Blanchet FG, Kindt R*, et al.* (2015) Package 'vegan'. *Community ecology package, version*, 2.3–0.

Olden JD, Hogan ZS, Zanden MJV (2007) Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world's freshwater and marine fishes. *Global Ecology and Biogeography* **16**, 694–701.

Oleksiak MF (2008) Changes in gene expression due to chronic exposure to environmental pollutants. *Aquatic Toxicology* **90**, 161–171.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics* **32**, 261–266.

Oleksiak MF, Roach JL, Crawford DL (2005) Natural variation in cardiac metabolism and gene expression in Fundulus heteroclitus. *Nature Genetics* **37**, 67–72.

Olofsson H, Ripa J, Jonzén N (2009) Bet-hedging as an evolutionary game: the trade-off between egg size and number. *Proceedings of the Royal Society of London B: Biological Sciences*, doi:10.1098/rspb.2009.0500.

Osborne MJ, Perkin JS, Gido KB, Turner TF (2014) Comparative riverscape genetics reveals reservoirs of genetic diversity for conservation and restoration of Great Plains fishes. *Molecular Ecology* **23**, 5663–5679.

Ouborg NJ, Angeloni F, Vergeer P (2010) An essay on the necessity and feasibility of conservation genomics. *Conservation Genetics* **11**, 643–653.

Palmer MA, Reidy Liermann CA, Nilsson C*, et al.* (2008) Climate change and the world's river basins: anticipating management options. *Frontiers in Ecology and the Environment* **6**, 81–89.

Papakostas S, Vøllestad LA, Bruneaux M*, et al.* (2014) Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nat Commun* **5**.

Paschke M, Bernasconi G, Schmid B, Williams R (2003) Population size and identity influence the reaction norm of the rare, endemic plant cochlearia bavarica across a gradient of environmental stress. *Evolution* **57**, 496–508.

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology* **22**, 925–946.

Pavey SA, Bernatchez L, Aubin-Horth N, Landry CR (2012) What is needed for next-generation ecological and evolutionary genomics? *Trends in Ecology & Evolution* **27**, 673–678.

Paz-Vinas I, Blanchet S (2015) Dendritic connectivity shapes spatial patterns of genetic diversity: a simulation-based study. *Journal of Evolutionary Biology* **28**, 986–994.

Peakall R, Smouse P (2012) GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research, an update. *Bioinformatics* **28**, 2537–2539.

Pearce L (2015) Surveys, monitoring and conservation status of Southern Pygmy Perch (*Nannoperca australis*) within Blakney and Pudman Creeks. NSW Department of Primary Industries.

Pels S (1964) The present and ancestral Murray River system. *Australian Geographical Studies* **2**, 111–119.

Perry GLW, Bond NR (2009) Spatially explicit modeling of habitat dynamics and fish population persistence in an intermittent lowland stream. *Ecological Applications* **19**, 731–746.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786.

Peterson B, Weber J, Kay E, Fisher H, Hoekstra H (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135.

Pettigrove V (1989) Larval mouthpart deformities in Procladius paludicola Skuse (Diptera: Chironomidae) from the Murray and Darling rivers, Australia. *Hydrobiologia* **179**, 111–117.

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. *Journal of Heredity* **90**, 502–503.

Pittock J, Finlayson CM (2011) Australia's Murray–Darling Basin: freshwater ecosystem conservation options in an era of climate change. *Marine and Freshwater Research* **62**, 232–243.

Poff NL, Allan JD, Bain MB*, et al.* (1997) The Natural Flow Regime. *Bioscience* **47**, 769–784.

Poff NL, Olden JD, Merritt DM, Pepin DM (2007) Homogenization of regional river dynamics by dams and global biodiversity implications. *Proceedings of the National Academy of Sciences* **104**, 5732–5737.

Pollino CA, Georgiades E, Holdway DA (2009) Physiological changes in reproductively active rainbowfish (Melanotaenia fluviatilis) following exposure to naphthalene. *Ecotoxicology and Environmental Safety* **72**, 1265–1270.

Pollino CA, Holdway DA (2002) Toxicity Testing of Crude Oil and Related Compounds Using Early Life Stages of the Crimson-Spotted Rainbowfish (Melanotaenia fluviatilis). *Ecotoxicology and Environmental Safety* **52**, 180–189.

Primmer CR, Papakostas S, Leder EH, Davis MJ, Ragan MA (2013) Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. *Molecular Ecology* **22**, 3216–3241.

Pritchard JK, Di Rienzo A (2010) Adaptation–not by sweeps alone. *Nature Reviews Genetics* **11**, 665–667.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology* **20**, 208–215.

Pujolar JM, Marino IA, Milan M*, et al.* (2012) Surviving in a toxic world: transcriptomics and gene expression profiling in response to environmental pollution in the critically endangered European eel. *BMC Genomics* **13**, 1.

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431.

Quintero I, Wiens JJ (2013) Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. *Ecology Letters* **16**, 1095–1103.

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabergh C, Airaksinen S, Soitamo A*, et al.* (2000) Tissue-specific expression of zebrafish (Danio rerio) heat shock factor 1 mRNAs in response to heat stress. *Journal of Experimental Biology* **203**, 1817–1824.

Raeymaekers JAM, Konijnendijk N, Larmuseau MHD*, et al.* (2014) A gene with major phenotypic effects as a target for selection vs. homogenizing gene flow. *Molecular Ecology* **23**, 162–181.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–589.

Réale D, McAdam AG, Boutin S, Berteaux D (2003) Genetic and plastic responses of a northern mammal to climate change. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 591–596.

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? a meta-analysis. *Evolution* **55**, 1095–1103.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**, 4348–4370.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.

Rohlfs RV, Nielsen R (2015) Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Systematic Biology* **64**, 695–708.

Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics* **13**, 505–516.

Saddlier S, Koehn JD, Hammer MP (2013) Let's not forget the small fishes – conservation of two threatened species of pygmy perch in south-eastern Australia. *Marine and Freshwater Research* **64**, 874–886.

Sætre CLC, Coleiro C, Austad M*, et al.* (2017) Rapid adaptive phenotypic change following colonization of a newly restored habitat. *Nature Communications* **8**, 14159.

Sala OE, Stuart Chapin F, III*, et al.* (2000) Global Biodiversity Scenarios for the Year 2100. *Science* **287**, 1770–1774.

Sasaki M, Hammer MP, Unmack PJ, Adams M, Beheregaray LB (2016) Population genetics of a widely distributed small freshwater fish with varying conservation concerns: the southern purple-spotted gudgeon, *Mogurnda adspersa*. *Conservation Genetics*, 1–15.

Schoville SD, Bonin A, François O*, et al.* (2012) Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics* **43**, 23–43.

Sgrò CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* **4**, 326–337.

Shafer ABA, Wolf JBW, Alves PC*, et al.* (2015a) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**, 78–87.

Shafer ABA, Wolf JBW, Alves PC*, et al.* (2015b) Reply to Garner *et al*. *Trends in Ecology & Evolution* **31**, 83–84.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, btv351.

Skovrind M, Olsen MT, Vieira FG*, et al.* (2016) Genomic population structure of freshwater-resident and anadromous ide (*Leuciscus idus*) in north-western Europe. *Ecology and Evolution* **6**, 1064–1074.

Slaninova A, Smutna M, Modra H, Svobodova Z (2009) REVIEWS Oxidative stress in fish induced by pesticides. *Neuroendocrinology Letters* **30**, 2.

Smith L, Oseid D (1972) Effects of hydrogen sulfide on fish eggs and fry. *Water Research* **6**, 711–720.

Smith S, Bernatchez L, Beheregaray L (2013) RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics* **14**, 375.

Smith S, Brauer C, Sasaki M*, et al.* (*in preparation*) Latitudinal variation in adaptive resilience in an ecologically important aquatic species: a genomic test of the climatic variability hypothesis.

Sork VL, Waits L (2010) Contributions of landscape genetics – approaches, insights, and future potential. *Molecular Ecology* **19**, 3489–3495.

Soulé ME (1985) What is conservation biology? A new synthetic discipline addresses the dynamics and problems of perturbed species, communities, and ecosystems. *Bioscience* **35**, 727–734.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

Steane DA, Potts BM, McLean E*, et al.* (2014) Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology* **23**, 2500–2513.

Stein J, Hutchinson M, Stein J (2014) A new stream and nested catchment framework for Australia. *Hydrology and Earth System Sciences* **18**, 1917–1933.

Stein JL, Stein JA, Nix HA (2002) Spatial analysis of anthropogenic river disturbance at regional and continental scales: identifying the wild rivers of Australia. *Landscape and Urban Planning* **60**, 1–25.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution* **18**, 94–101.

Storfer A, Murphy MA, Evans JS*, et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128–142.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology* **19**, 3496–3514.

Strayer DL, Dudgeon D (2010) Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society* **29**, 344–358.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800.

Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution* **19**, 489–496.

Tao T (2014) Standalone BLAST Setup for Unix. Bethesda (MD): National Center for Biotechnology Information BLAST® Help [Internet].

Thomaz AT, Christie MR, Knowles LL (2016) The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution* **70**, 731–739.

Thuiller W (2007) Biodiversity: Climate change and the ecologist. *Nature* **448**, 550–552.

Torterotot J-B, Perrier C, Bergeron NE, Bernatchez L (2014) Influence of Forest Road Culverts and Waterfalls on the Fine-Scale Distribution of Brook Trout Genetic Diversity in a Boreal Watershed. *Transactions of the American Fisheries Society* **143**, 1577–1591.

Uebbing S, Künstner A, Mäkinen H*, et al.* (2016) Divergence in gene expression within and between two closely related flycatcher species. *Molecular Ecology* **25**, 2015–2028.

UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Research* **43**, 204–212.

Unmack PJ, Hammer MP, Adams M, Dowling TE (2011) A Phylogenetic Analysis of Pygmy Perches (Teleostei: Percichthyidae) with an Assessment of the Major Historical Influences on Aquatic Biogeography in Southern Australia. *Systematic Biology* **60**, 797–812.

Unmack PJ, Hammer MP, Adams M, Johnson JB, Dowling TE (2013) The role of continental shelf width in determining freshwater phylogeographic patterns in south-eastern Australian pygmy perches (Teleostei: Percichthyidae). *Molecular Ecology* **22**, 1683–1699.

Vajda AM, Kumar A, Woods M*, et al.* (2015) Integrated assessment of wastewater treatment plant effluent estrogenicity in the Upper Murray River, Australia, using the native Murray rainbowfish (Melanotaenia fluviatilis). *Environmental Toxicology and Chemistry* **34**, 1078–1087.

Valladares F, Matesanz S, Guilhaumon F*, et al.* (2014) The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecology Letters* **17**, 1351–1364.

van de Pol M, Osmond HL, Cockburn A (2012) Fluctuations in population composition dampen the impact of phenotypic plasticity on trait dynamics in superb fairy-wrens. *Journal of Animal Ecology* **81**, 411–422.

Visser ME (2008) Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proceedings of the Royal Society of London B: Biological Sciences* **275**, 649–659.

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of earth's ecosystems. *Science* **277**, 494–499.

Vorosmarty CJ, McIntyre PB, Gessner MO*, et al.* (2010) Global threats to human water security and river biodiversity. *Nature* **467**, 555–561.

Walther G-R, Post E, Convey P*, et al.* (2002) Ecological responses to recent climate change. *Nature* **416**, 389–395.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology* **19**, 2605–2608.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* **67**, 3403–3411.

Wang PA, Stenvik J, Larsen R, Mæhre H, Olsen RL (2007) Cathepsin D from Atlantic cod (Gadus morhua L.) liver. Isolation and comparative studies. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **147**, 504–511.

Wang Z, Gerstein M, Snyder M (2009) Innovation: RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews: Genetics* **10**, 57–63.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary *N*e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* **3**, 244–262.

Waples RS, Pess GR, Beechie T (2008) Evolutionary history of Pacific salmon in dynamic environments. *Evolutionary Applications* **1**, 189–206.

Webb PW (1982) Locomotor patterns in the evolution of actinopterygian fishes. *American Zoologist* **22**, 329–342.

Webster MS, Colton MA, Darling ES*, et al.* (2017) Who Should Pick the Winners of Climate Change? *Trends in Ecology & Evolution*.

Wedderburn S, Hammer M, Bice C (2012) Shifts in small-bodied fish assemblages resulting from drought-induced water level recession in terminating lakes of the Murray-Darling Basin, Australia. *Hydrobiologia* **691**, 35–46.

Weeks AR, Sgro CM, Young AG*, et al.* (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* **4**, 709–725.

Weeks AR, Stoklosa J, Hoffmann AA (2016) Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. *Frontiers in Zoology* **13**, 1.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Weir BS, Hill W (2002) Estimating F-statistics. *Annual Review of Genetics* **36**, 721–750.

Whitehead A (2012) Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. *The Journal of Experimental Biology* **215**, 884–891.

Whitehead A, Galvez F, Zhang S, Williams LM, Oleksiak MF (2011) Functional genomics of physiological plasticity and local adaptation in killifish. *Journal of Heredity* **102**, 499–511.

Whitehead A, Triant DA, Champlin D, Nacci D (2010) Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular Ecology* **19**, 5186–5203.

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. *Trends in Ecology & Evolution* **30**, 42–49.

Whiterod NS, Zukowski S, Asmus M, Gilligan D, Miller AD (2016) Genetic analyses reveal limited dispersal and recovery potential in the large freshwater crayfish Euastacus armatus from the southern Murray–Darling Basin. *Marine and Freshwater Research*.

Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of fst. *The American Naturalist* **186**, S24–36.

Willi Y, Hoffmann AA (2012) Microgeographic adaptation linked to forest fragmentation and habitat quality in the tropical fruit fly *Drosophila birchii*. *Oikos* **121**, 1627–1637.

Willi Y, Van Buskirk J, Schmid B, Fischer M (2007) Genetic isolation of fragmented populations is exacerbated by drift and selection. *Journal of Evolutionary Biology* **20**, 534–542.

Wood JLA, Belmar-Lucero S, Hutchings JA, Fraser DJ (2014) Relationship of habitat variability to population size in a stream fish. *Ecological Applications* **24**, 1085–1100.

Wood JLA, Fraser DJ (2015) Similar plastic responses to elevated temperature among different-sized brook trout populations. *Ecology* **96**, 1010–1019.

Wood JLA, Tezel D, Joyal D, Fraser DJ (2015) Population size is weakly related to quantitative genetic variation and trait differentiation in a stream fish. *Evolution* **69**, 2303–2318.

Wood JLA, Yates MC, Fraser DJ (2016) Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications* **9**, 640–657.

Yeomans KA, Golder PA (1982) The Guttman-Kaiser criterion as a predictor of the number of common factors. *The Statistician* **31**, 221–229.

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**, 1.