Evolutionary Biology of the Invasive Northern Pacific Seastar, *Asterias amurensis*

by

Mark Richardson
MSc, Bsc (Hons)

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Deakin University

March 2015
I am the author of the thesis entitled


submitted for the degree of Doctor of Philosophy

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1968.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: Mark F. Richardson

Signed: [Signature Redacted by Library]

Date: 8/7/15
I certify the following about the thesis entitled (10 word maximum)

Evolutionary Biology of the Invasive northern Pacific Seastar, Asterias amurensis, submitted for the degree of Doctor of Philosophy

a. I am the creator of all or part of the whole work(s) (including content and layout) and that where reference is made to the work of others, due acknowledgment is given.

b. The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

c. That if the work(s) have been commissioned, sponsored or supported by any organisation, I have fulfilled all of the obligations required by such contract or agreement.

I also certify that any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: Mark F. Richardson

Signed: [Signature Redacted by Library]

Date: 24/03/2015
“The greatest diffusion of my efforts”

Anon
Acknowledgements

Firstly, I would like to thank my supervisors Craig Sherman and Jeff Ross for their enthusiasm, support, patience, and importantly the opportunity they gave me. Thank you also for your mentorship and encouragement throughout, it has been a challenging but rewarding experience and I hope, just the start of some exciting collaborations.

To my fantastic wife, Chelsea, without you this would not have been possible. Your support has been unbelievable. Thank you for motivating me throughout and giving me a push at the right time during the arduous writing up. Thank you so much for picking up all the slack that enabled me to focus on this: I have a lot of making up to do and can’t wait to be home with the family more often now. To Hamish, you put everything in perspective for me. I hope some day you are proud of Dad.

To my family, both abroad and here in Australia, thank you for your continuous encouragement and support. I am grateful for every opportunity you have given me that got me to this point. To Mum, thank you for the proofreading, you are a gem.

To members of the Sherman Lab – Adam Cardilini, Kim Weston, Aimee Hauser, Julia Kent, Tim Smith, Paul York and Dan Lees – thank you for your assistance, advice and companionship (and Adam, endless philosophical debates), you made the experience incredibly enjoyable. Your friendship means a lot to me. To my colleagues – Lee Ann Rollins, Chris Cowled, James Wynne, Kate Buchannan and Rick Shine – thank you for the great advice, discussions
and for comments on this thesis. Thank you also to members of the CIE for their friendship, mentorship and stimulating discussions.

I would like to thank Mieko Komatsu, Mariko Kondo, Koji Akasaka and especially, Miyuki Kaida for facilities and organizational help while in Japan. I am particularly indebted to Kotsuka-san for help in collecting seastars and during experimentation. Thank you also to Masakazu Aoki and members of his Lab, Tatsuya Sakamoto and Kazuhiro Saito for hosting us in Japan and for assisting seastar collections. Thanks are also due to Jim Rooks, David Cahill, Liz McGrath and Rod Watson for providing facilities for temperature treatments and larval rearing in Australia.

Katrina Jantz from the Hawkesbury Institute for the Environment conducted cDNA library preparation and sequencing for data in Chapter 3, rapidly and at a fantastic price: thank you. Sequencing and cDNA library preparation was conducted commercially at Macrogen Inc, South Korea, for data from Chapters 4 and 5.

Research was supported by a Holsworth Wildlife Research Endowment Award (RM23514), funding from the Centre for Integrative Ecology and an International PhD Scholarship (both Deakin University). The collection of adult seastars was conducted under the Victorian and Tasmanian state government noxious aquatic species permits (NP152, NP252, 11103).
# Table of Contents

**Acknowledgements** ........................................................................................................... i  
**Table of Contents** ............................................................................................................ iii  
**Abstract** ............................................................................................................................. 1  
**Chapter 1. Introduction** .................................................................................................. 4  
  1.1 Introduction ....................................................................................................................... 4  
  1.2 Genetic architecture and evolutionary change in invasive species .......... 7  
    1.2.1 Genetic variation ............................................................................................................. 7  
    1.2.2 The influence of bottlenecks and founder events on genetic diversity .......... 8  
    1.2.3 Evolutionary change arising from mutation .......................................................... 10  
    1.2.4 Multiple introductions .............................................................................................. 11  
    1.2.5 Migration between introduced populations ......................................................... 12  
    1.2.6 Phenotypic plasticity and epigenetic change .......................................................... 13  
    1.2.7 Hybridization ............................................................................................................... 15  
    1.2.8 On what traits does selection act and when does it occur during the invasion process? .................................................................................................................. 16  
  1.3 Application of next-generation sequencing to reveal genetic basis of evolutionary change in invasive species ........................................................................................................ 18  
  1.4 A new model, *Asterias amurensis*, to study evolutionary and invasive biology ................................................................. 20  
    1.4.1 Life history .................................................................................................................. 23  
  1.5 PhD aims and objectives .................................................................................................. 25  
  1.6 References ....................................................................................................................... 27  
**Chapter 2. Rapid divergence and maintenance of genetic diversity during range expansion in an invasive marine invertebrate** ....................................................... 40  
  2.1 Introduction ....................................................................................................................... 40  
  2.2 Methods ............................................................................................................................ 44  
    2.2.1 Population genetic sampling and genotyping ..................................................... 44  
    2.2.2 Genetic data analysis ................................................................................................. 46  
    2.2.3 Hydrodynamic modelling ......................................................................................... 48  
    2.2.4 Plankton sampling and *Asterias amurensis* larval identification ................ 49  
  2.3 Results ............................................................................................................................... 51  
    2.3.1 Genetic structure of native and invasive populations ........................................ 51  
    2.3.2 Hydrodynamic modelling ......................................................................................... 60  
    2.3.3 Identification of *A. amurensis* larvae in plankton samples ............................ 61  
  2.4 Discussion ........................................................................................................................ 63  
    2.4.1 Regional genetic structure of *A. amurensis* between the native and invasive ranges ................................................................. 63  
    2.4.2 Local genetic structure and larval dispersal in the invasive range .................. 66  
  2.5 Conclusions ...................................................................................................................... 71  
  2.6 Supplementary Results ................................................................................................... 72  
    2.6.1 Genetic structure of native and invasive populations ........................................ 72  
    2.6.2 Fine-scale genetic structure of invasive populations including additional microsatellite loci ................................................................. 74  
  2.7 References ....................................................................................................................... 76
Chapter 3. De novo assembly and characterization of the invasive northern Pacific seastar transcriptome

3.1 Background ................................................................. 82
3.2 Results and Discussion ............................................... 85
  3.2.1 Sequencing and quality control ........................................ 85
  3.2.2 Transcriptome assembly ................................................ 85
  3.2.3 Functional annotation .................................................. 89
  3.2.4 Transcriptome comparison to the Bat star, Patiria miniata proteins .............. 94
  3.2.5 Gene Ontology (GO) and KEGG annotation ...................... 95
  3.2.6 Identification of coding sequences and protein domains ..................... 97
  3.2.7 Transposable elements .................................................. 99
  3.2.8 Identification of candidate genes for environmental adaptation .......... 100
3.3 Conclusions ................................................................. 101
3.4 Methods ......................................................................... 102
  3.4.1 Sample collection and RNA isolation ..................................... 102
  3.4.2 Sequencing and quality control ........................................... 103
  3.4.3 Digital normalization and de novo assembly ........................... 104
  3.4.4 Functional annotation ..................................................... 106
  3.4.5 Comparative annotation validation to the Bat star, Patiria miniata proteins ... 108
  3.4.6 Identification of candidate genes associated with environmental adaptation ............................................................... 108
3.5 Supplementary Material .................................................. 109
3.6 References ....................................................................... 110

Chapter 4. The molecular basis of a successful invasion: comparative RNA-Seq analysis of native and invasive Asterias amurensis

4.1 Introduction ..................................................................... 116
4.2 Methods .......................................................................... 120
  4.2.1 Sample collection and larval rearing .................................... 120
  4.2.2 RNA isolation ............................................................... 121
  4.2.3 Sequencing and Quality control ....................................... 122
  4.2.4 Read Mapping and expression quantification ....................... 123
  4.2.5 Evaluation of differential expression and diagnostic performance .... 124
  4.2.6 Digital differential gene expression ................................... 125
  4.2.7 Gene ontology enrichment analysis .................................. 127
  4.2.8 Outlier loci analysis .......................................................... 127
4.3 Results ............................................................................. 129
  4.3.1 Sequencing and read mapping .......................................... 129
  4.3.2 Differential expression between native and invasive populations ...... 130
  4.3.3 Putative function of differentially expressed genes .................. 133
  4.3.4 Outlier loci analysis .......................................................... 134
4.4 Discussion ........................................................................ 137
  4.4.1 Functional roles of the most differentially expressed genes ............ 137
  4.4.2 Functional significance of cytochrome p450 expression ............ 139
  4.4.3 Differential selection ...................................................... 141
  4.4.4 Variation in gene expression ............................................. 143
4.5 Conclusions ...................................................................... 146
4.6 Supplementary Results .................................................... 147
  4.6.1 Assessment of technical and diagnostic performance ................ 147
Chapter 5. Comparative RNA-Seq analysis of the temperature response in native and invasive Asterias amurensis

5.1 Introduction ............................................................................................................. 157
5.2 Methods .................................................................................................................. 161
  5.2.1 Larval temperature experiment and sample collection ................................... 161
  5.2.2 RNA isolation ...................................................................................................................... 162
  5.2.3 RNA sequencing and quality control ............................................................... 164
  5.2.4 Read Mapping and expression quantification ...................................................... 164
  5.2.5 Digital differential gene expression .......................................................................... 165
  5.2.6 Gene ontology enrichment analysis ........................................................................ 167
5.3 Results .................................................................................................................. .................. 168
  5.3.1 Larval temperature experiment, sequencing and read mapping ................ 168
  5.3.2 Gene expression response to elevated temperature treatments ............ 169
  5.3.3 Shared trends in gene expression between the native and invasive population ....................................................................................................................................... 172
  5.3.4 Population-specific responses to temperature treatment .................. 173
5.4 Discussion ............................................................................................................... .............. 177
  5.4.1 The transcriptional response of invasive A. amurensis to elevated temperature ................................................................................................................................... 178
  5.4.2 Comparison to the native A. amurensis temperature response .......... 183
  5.4.3 Evolutionary implications ............................................................................................. 184
5.5 Conclusion ............................................................................................................... ............. 187
5.6 Supplementary Material ......................................................................................... 188
  5.6.1 Methods ................................................................................................................................. 188
  5.6.2 Results .................................................................................................................................... 189
5.7 References ............................................................................................................... ............. 195

Chapter 6. Conclusions ............................................................................................... 201
6.2 References ............................................................................................................... 204

Appendix 1
Appendix 2
Abstract

Invasive species provide valuable study systems to evaluate evolutionary and ecological processes occurring during the colonization of new habitats because they represent natural experiments in a contemporary time frame. Exposure to novel environmental conditions experienced upon introduction can generate strong selective pressures, eliciting rapid evolutionary change in morphological, physiological and life history traits. While considerable research has documented the ecological aspect of biological invasions, we still know relatively little about the underlying genetic basis of these often observed adaptive changes and how physiological responses and associated underlying genetic mechanisms facilitate environmental adaptation and invasion success.

In this study, I used the invasive northern Pacific seastar, *Asterias amurensis*, to examine the genetic components of environmental adaptation, which may promote invasion success in Australia. I used a combination of genetic, transcriptomic and experimental techniques to assess: *i*) estimates of diversity and divergence between native and invasive populations, *ii*) how Australian invasive populations are structured, *iii*) larval dispersal for range expansion, *iv*) larval gene expression variation between native and invasive populations, *v*) test for signatures of selection on candidate genes and, *vi*) what genes are potentially involved in how native and invasive populations cope with elevated temperatures to understand the potential for invasive range expansion into warmer waters along the Australian mainland coast.

Through a combination of microsatellite genetic data, hydrodynamic and dispersion modelling and empirical larval detection, I revealed significant
neutral genetic divergence between native and invasive *A. amurensis* populations. The Australian introduction likely has a single origin from Tokyo Bay, Japan, which resulted in the loss of neutral genetic diversity within the invasive range. Range expansion along the Victorian coast is clearly facilitated by natural larval dispersal mechanisms from Port Phillip Bay, which underlies the importance of studying larval traits for further range expansion.

I produced the first transcriptome resource for *A. amurensis* with comprehensive functional annotation in order to examine evolutionary processes affecting dispersal potential of invasive larvae. The assembled *A. amurensis* bipinnaria transcriptome contains 18,319 uniquely annotated protein-coding genes. I reveal ~10,000 orthologs to previously described Bat star proteins while demonstrating a comparable number of genes expressed to sea urchin developmental stages.

Through a comparative RNA-Seq approach I found substantial gene expression divergence between larvae from the native and invasive range, within a common environmental setting. Many important genes with putative roles in immune function and the response to environmental contaminants are differentially expressed between native and invasive populations. These genes may have a role in facilitating adaptation of invasive larvae to Australian environmental conditions. In particular, the up-regulation of cytochrome p450 (*p450*) genes in invasive larvae reveals a response to metabolize environmental xenobiotics not seen in native larvae. I also identified several genes as putatively experiencing differential selection in the invasive range, which are implicated in immune function and developmental processes. This may reflect selection for alleles in invasive larvae that perform better in novel environmental conditions and may indicate rapid local adaptation during the invasion into Australia.
Through the application of a combined temperature experiment and comparative RNA-seq approach, I investigated if Australian *A. amurensis* larvae have greater tolerance to elevated temperatures, a likely limiting factor to future range expansion and invasion success. I found substantial differences in the response to elevated temperatures (17 and 20°C) between native and invasive larvae, with invasive larvae exhibiting a much greater transcriptional response than in natives, suggesting the evolution of increased plasticity or local adaptation. It appears invasive larvae may have evolved greater capacity to cope with higher temperatures during fertilization and larval development, primarily through the activation of genes involved in development, cellular regulation and function, including post-translational protein modifications.

In summary, given the relatively short invasion (~30 generations) data presented here provides new insights into rapid changes in the genetic architecture underlying larval physiological responses to elevated temperatures and the potential for this species to expand its invasive range.
Chapter 1. Introduction

1.1 Introduction

Biological invasions are regarded as one of the most significant threats to global biodiversity (Sala & Knowlton 2006a). Invasive species not only have the potential to drastically alter native ecological systems but also damage human health and economic interests (Vitousek et al. 1997; Sakai et al. 2001). The annual cost of invasive species is in excess of $335 billion, for just the six largest developed nations (Pimentel et al. 2005). Additionally, as a result of global trade and transport, the number of non-native species introduced into new areas is projected to increase (Bossdorf et al. 2005). Estimates suggest that the global total number of species introduced into new geographic areas is approaching half a million (Pimentel et al. 2005).

The threats posed by non-native species is not a contemporary problem; Elton (1958) drew attention to the tremendous damage invasive species can cause not only to ecosystems and biodiversity but also industry. Nonetheless, it was not until the 1970s that biological invasions stimulated widespread research, primarily due to increasing concern about the endangerment and extinction of species (Cox 2004). This proliferation of research has predominantly focussed on the ecological aspects of species invasions (Lodge 1993; Lockwood et al. 2005; Strayer et al. 2006; Phillips et al. 2006; Shine 2010). Baker & Stebbins (1965) first illuminated the importance of an evolutionary approach to study biological invasions by understanding that invasive populations will be genetically dynamic over space and time. Despite this early realisation of the importance of genetic variation and rapid evolutionary change when faced with novel environments, only recently have
attempts been made to understand evolutionary dynamics of biological
invasions. Biological invasions, by definition, involve the introduction,
establishment, spread and proliferation of a species outside its native range
(Estoup & Guillemaud 2010). Accordingly, introduced species are present in
biogeographic areas where they did not evolve and as a result, are exposed to
novel environmental conditions and selection pressures (Sax et al. 2007a;
Prentis et al. 2008). Abiotic factors differing from those in the native range and
changes in biotic interactions of predators, diseases and competitors will select
for adaptations to the new habitat (Cox 2004). As such invasive species can
serve as good models to investigate evolutionary processes (Sax et al. 2007a).

The potential for rapid evolution and adaptation in invasive species will
be dependent on: i) genetic variation in heritable traits, ii) the intensity and
direction of selection acting upon these traits, iii) the effect of drift in founding
populations (i.e. genetic bottlenecks), iv) the accumulation of new mutations
relevant to heritable traits, v) migration and dispersal of invasive populations
and vi) the role of phenotypic plasticity and epigenetic effects (see Lee 2002;

Contemporary studies have documented rapid adaptation during the invasion
process. For example, research has revealed adaptive change during invasions
of Anolis lizard in the Caribbean and North America (Losos et al. 2006; Kolbe et
al. 2012b, 2014; Stuart et al. 2014) and the evolution of increased dispersal
ability in the Cane toad (Rhinella marina) invasion in Australia (Rollins et al. in
press; Phillips et al. 2006; Urban et al. 2007; Alford et al. 2009). Yet relatively
few empirical studies have focussed on, or characterised, the underlying genetic
changes that arise during rapid adaptive evolution during biological invasions
(Lavergne & Molofsky 2007; Whitney & Gabler 2008), although there have been recent efforts to do this (e.g. Rollins et al. in press; Hodgins et al. 2013, 2014).

Here, I aim to provide a summary of the mechanisms and consequences of rapid adaptive evolution during the invasion process, including how this process may be facilitated or constrained by the dynamics of the founding and subsequent invasive populations. I define rapid evolutionary change as ‘a genetic change occurring rapidly enough to have a measurable impact on simultaneous ecological change’ as previously defined by (Carroll et al. 2007); generally these changes will occur over tens of generations. Given recent developments in evolutionary research, I intend to demonstrate the importance of adopting an evolutionary approach to study biological invasions. Furthermore, I will illustrate why it is important to study invasions in this context, not only to minimize their potential impacts, but also gain insights into the mechanisms underlying rapid evolutionary change.
1.2 Genetic architecture and evolutionary change in invasive species

1.2.1 Genetic variation

Central to our understanding of the potential for invasive species to undergo rapid evolutionary change are the underlying genetic attributes that may facilitate invasion success, and how they respond to natural selection. The genetic changes that enable species to adapt to new environments will arise through either, standing genetic variation (allelic variation currently present in a population, as opposed to new variation arising from mutation) or new mutations and recombination (Hartl & Clark 2007; Barrett & Schluter 2008). Current evidence suggests that rapid evolutionary change during the invasion process (change occurring over tens of generations or fewer) should primarily be due to alleles originating from source populations (Hendry et al. 2002; Barrett & Schluter 2008; Prentis et al. 2008). In this case, it is the proportion of genetic diversity present within the founding population that may indicate the potential for evolutionary adaptation. However, it is the level of additive genetic variation (genetic variation in a trait that is due to the additive component of allelic effects) present rather than genetic diversity (polymorphism across all loci, including neutral sites) that responds to natural selection (Lee 2002). As such, additive variation will act as the main substrate for evolutionary change. Several studies have revealed high levels of additive variation within invasive populations for morphological and behavioural traits; in *Drosophila melanogaster* (Lopez-Fanjul & Villaverde 1989) and *Bicyclus anynana* (Saccheri et al. 1996), that may be important for invasion success.

Adaptation arising from standing genetic variation is also likely to be important when considering rapid evolution during invasive range expansion.
(Prentis et al. 2008; Colautti et al. 2010), especially when species spread into new environmentally diverse habitats. The presence of geographical clines across introduced ranges; in life history traits (Huey et al. 2000; Allendorf & Lundquist 2003; Phillips et al. 2006; Urban et al. 2007) and environmental tolerance (Barrett et al. 2011; Hoffmann & Sgro 2011; Pespeni & Palumbi 2013a) illustrate the potential for rapid local adaptation and evolution during this process. For example, the fruit fly, Drosophila subobscura, was introduced into South America in the late 1970s and subsequently spread into North America, where it evolved an adaptive cline in wing size comparable to that exhibited by ‘Old World’ native populations (Huey et al. 2000; Gilchrist et al. 2001, 2004). Consequently, adaptive evolution during an invasive range expansion may arise from the populations intrinsic standing genetic variation. Or, as invasive range expansions generally occur after a lag phase following initial establishment (Mooney & Cleland 2001; Lande 2009), the accumulation of sufficient additive genetic variation may also be responsible for adaptive change (Lee 2002; Blackburn et al. 2011).

1.2.2 The influence of bottlenecks and founder events on genetic diversity

Genetic bottlenecks are often associated with invasive species, as introduced populations are usually founded from a limited number of individuals (Sakai et al. 2001; Frankham 2004; Prentis et al. 2008). Genetic bottlenecks (i.e. the rapid reduction in the number of individuals in a population that leads to a reduction in genetic diversity) and founder effects (genetic changes occurring when a few individuals establish a new population) have commonly been thought to decrease the potential for rapid adaptive evolution due to a perceived reduction
in the level of genetic diversity associated with these stochastic processes (Allendorf & Lundquist 2003; Willi et al. 2006; Slothouber Galbreath et al. 2009). While both these processes are different, they produce an almost identical population genetic signature based on a history of reduced effective population size and changed selective pressures. This may encompass lower allelic diversity and heterozygosity, and different allelic frequencies due to sampling effects associated with the founding of new populations (Hartl & Clark 2007; Peacock et al. 2009). Traditional population genetic theory predicts that a loss in genetic diversity is governed by the growth rate of the population and its effective minimum population size ($N_e$); in the case of introductions, its founding population size (Nei 1975). The lower the $N_e$ and/or growth rate following introduction the higher the potential loss of alleles (particularly those that are rare) through the action of genetic drift (Hartl & Clark 2007). Generally, shifts in neutral or slightly deleterious allele frequencies are expected to have no effect on fitness. Rare alleles, many of which are deleterious, are suggested to have important fitness consequences, particularly those under frequency dependent selection (Dlugosch & Parker 2008a). The magnitude of any potential loss in genetic diversity will be dependent on the extent and duration of the bottleneck, with those of short duration not necessarily leading to a reduction in genetic variation (Nei 1975). Still, even if a population rebounds quickly from a bottleneck, $N_e$ could still be lowered, since it is largely determined by the historical census size of the population (Hartl & Clark 2007). This suggests that bottlenecks may not only limit invasion success but also the potential for rapid adaptive evolution of fitness-related traits in invasive species (Sakai et al. 2001; Frankham 2004; Willi et al. 2006). However, several studies have revealed similar levels of genetic variance in introduced and source
populations (Zeisset & Beebee 2003; Maron et al. 2004; Kliber & Eckert 2005; Chen et al. 2006), higher (Kolbe et al. 2004) or only marginal losses of variation in introduced populations (Holland 2001; Rasner et al. 2004). Further, a growing number of studies have observed the evolution of putatively adaptive traits following introduction (Maron et al. 2004; Lindholm et al. 2005; Phillips et al. 2006; Yonekura et al. 2007; Gordon et al. 2009; Kolbe et al. 2014) and revealed that adaptation is not necessarily limited by reductions of genetic diversity (Stockwell et al. 1996; Stockwell & Weeks 1999; Stockwell & Vinyard 2000; Koskinen et al. 2002a; Dlugosch & Parker 2008c; Rollins et al. 2013).

1.2.3 Evolutionary change arising from mutation

Most mutations are neutral, or nearly neutral and have no effect on fitness, but the few that do affect fitness generally do so negatively and are thus, expected to be purged from populations over time (Keightley & Lynch 2003). Then again, the fate of mutations is different when we consider those arising in stable populations or expanding ones, even with regards to slightly deleterious mutations. For example, new mutations occurring in rapidly expanding populations, such as invasive populations after a bottleneck, are expected to have a much greater probability of becoming established than in stable populations at a demographic equilibrium (Nei 1975; Keightley & Lynch 2003; Cox 2004). On the other hand, empirical studies have found that mutations represent a negligible component of the genetic variation surveyed in founding populations (Bohonak et al. 1998; Bohonak & Roderick 2001; Estoup et al. 2001). The underlying issue in determining the extent to which new mutations contribute to evolutionary change in invasive populations is whether these sampled novel alleles are in situ mutations or rare alleles previously undetected.
in source populations. Revealing this remains extremely difficult due to the inherent problem of obtaining sufficient genetic sampling of source populations and characterizing their complete genetic diversity.

1.2.4 Multiple introductions

It has been observed that multiple introductions are also common to invasions (Novak et al. 1993; Novak & Mack 1995; Novak & Welfley 1997; Huttanus et al. 2011) and it has been postulated that such scenarios may bring together novel genetic combinations and unusually large amounts of genetic variation (Dlugosch & Parker 2008a). For example, a given founding population may be comprised of individuals from multiple differentiated source populations across a species native geographic range. As such, an introduced population may exhibit greater within-population than native among-population diversity levels, as illustrated by Kolbe et al. (2004, 2007). Introductions of the red alga *Polysiphonia harveyi* and green alga *Caulerpa taxifolia* in Europe are similarly derived from multiple source populations, in this case from native Japanese and Australian populations, respectively (McIvor et al. 2001; Meusnier et al. 2004).

This has led to the hypothesis that the influx of genetic variation and consequent emergence of novel allelic combinations, from multiple introduction events, will directly influence the success and adaptive potential of an invasion (Allendorf & Lundquist 2003; Frankham 2004).

Dlugosch & Parker (2008a) conducted a meta-analysis into the role multiple introductions play during the invasion process, where they found an association with increased diversity and allelic variation, but only over long timescales (~100 years). They revealed a U-shaped pattern of genetic diversity over time, where initially diversity declined after initial establishment, but
returned to similar levels present in source populations (Dlugosch & Parker 2008a). During initial establishment, genetic drift and strong selection may contribute to the loss of variation observed (Nei 1975). However, in the long term, larger populations should experience reduced drift and after range expansion become interconnected with other introduced populations; leading to a rise in diversity compared to native populations (Dlugosch & Parker 2008a). Subsequently, most invaders may experience some loss of genetic variation and the benefits associated with multiple introductions may not be realized until long after establishment (Dlugosch & Parker 2008a). By such time gene flow between populations from the same invasive range may have similar effects. Multiple introductions may not, therefore, be ubiquitous to successful invasions or the potential for invasive populations to exhibit adaptive evolution; much will depend on the demographic context of the introduction and subsequent invasion.

1.2.5 Migration between introduced populations

Dispersal patterns between invasive populations may also influence the acquisition and rate of evolutionary change (Münkemüller et al. 2011). Gene flow between independently colonized populations (Stepien et al. 2005) and the combination of short-distance diffusion and long-distance dispersal, known as ‘stratified dispersal’, have been suggested to increase genetic diversity in expanding invasive populations (Tobin & Blackburn 2008; Bronnenhuber et al. 2011; Berthouly-Salazar et al. 2013). A process of repeated founding out of marginal invasive populations into new environments during range expansion has been suggested to shift allele frequencies and influence genetic variation
and fitness-related traits, which may promote evolutionary change in these marginal populations (Quinn et al. 2001; Travis & Dytham 2002; Burton & Travis 2008; Travis et al. 2010). However, a modelling study revealed that evolutionary adaptation was hindered by high migration rates which acted to homogenize genetic differences among populations (Garcia-Ramos & Rodriguez 2002). This has also been observed empirically among morphologically diverged stickleback populations (Hendry & Taylor 2004). Yet, local adaptation has been observed in Atlantic cod populations with high gene flow (Nielsen et al. 2009). Migration between source and marginal populations could act to inhibit local adaptive change, while the founding of isolated populations from marginal sources may act to promote local evolutionary change.

### 1.2.6 Phenotypic plasticity and epigenetic change

Biological invasions generally experience sudden environmental change and the success of an invading species may largely be determined by its ability to immediately respond and adapt to these changes (Lande 2009). A mechanism of environmentally generated variation, phenotypic plasticity (the property of an organism’s genotype to produce multiple phenotypes in different environments) has often been invoked to explain early adaptation of novel environments (Scheiner 2002; Lande 2009; Westley 2011). A number of studies have argued that plasticity enhances the ecological niche breadth of an invasive species as these plastic responses allow it to express advantageous phenotypes in a broad range of environments (Munir et al. 2001; Richards et al. 2005; Donohue et al. 2007). While this mechanism is often viewed separately to an individual’s genetic variation, phenotypic plasticity will still be subject to
evolution by natural selection as it is an emergent ‘property’ of a genotype and therefore governed by genetic variation (Nicoglou 2015). It is therefore likely that genetic variation for plastic traits will also influence the ability of a species to adapt to novel environments. For example, if genetic variation for a plastic trait exists within an introduced population and this plasticity confers a fitness advantage in this novel environment, then we may see the evolution of increased plasticity (Richards et al. 2006). Kolbe et al. (2014) revealed a potential role for adaptive plasticity to temperature in *Anolis sagrei*. Genetic variation for plastic traits has been observed in successful introductions of *Porcellio laevis* (Lardies & Bozinovic 2008), harlequin ladybird (Lombaert et al. 2010) and *Senecio pterophorus* (Caño et al. 2008) and been documented in non-invasive *Wahlenbergia ceracea* (Nicotra et al. 2015). Subsequently, adaptive evolution arising from plasticity may be an important component of the success of an invasive species.

Contemporary theoretical studies have suggested that this environmentally induced variation might also be mediated by non-genetic mechanisms of inheritance, which could influence adaptive evolution (Bonduriansky et al. 2012; Schlichting & Wund 2014). This epigenetic inheritance resulting from the parental phenotype or environment may comprise a variety of proximate mechanisms, such as DNA-methylation and chromatin structure (which mediate gene expression), hormones or behaviours (Jablonka et al. 1995). These epigenetic variant phenotypes can be either i) acquired traits, induced by the environment of the parents or ii) arise spontaneously without environmental effects (e.g. random DNA-methylation patterns) which are then transmitted to offspring (Bonduriansky & Day 2009; Bonduriansky et al. 2012). Salinas & Munch (2012) demonstrated that variation
in an important life history trait (thermal tolerance) is transmitted trans-generationally by non-genetic inheritance in sheepshead minnows. Day & Bonduriansky (2011) showed through the application of evolutionary models that these trans-generational epigenetic effects could interact with genetic variation and influence the response to natural selection. Herrera & Bazaga (2011) studied DNA-methylation variation in the violet, *Viola cazorlensis* and found significant epigenetic variation among individuals, which was related to long-term herbivore damage. While empirical studies revealing epigenetic effects on adaptation are in their infancy, they may play an important role in the adaptation of individuals to new environments and thus, become highly applicable for the study of biological invasions (see Liebl *et al.* 2013).

### 1.2.7 Hybridization

Along with the direct evolutionary responses of invading species described above, there also are important indirect effects upon the genetic variation and evolutionary responses of invading species in relation to the new organisms they may encounter (Huxel 1999; Mooney & Cleland 2001). These indirect effects primarily are inter- or intraspecific hybridization and introgression of invading populations with native or other introduced populations. It has been suggested that this process can generate novel genotypes and limit the potential losses of additive genetic variance associated with founding events through new gene interactions and the removal or masking of unfavourable deleterious alleles (Ellstrand & Schierenbeck 2000; Lee 2002). Prentis *et al.* (2008) predict that the recombination of parental genotypes, forming hybrids, is likely to be an important facet that promotes rapid evolutionary change in invasive plants.
This could promote evolutionary change because recombination may generate novel gene combinations and interactions which, through the action of natural selection, form phenotypes that are suited to the exploitation of novel environments (Prentis et al. 2008). Further, transgressive segregation and adaptive trait introgression (heterosis), or a combination of dominance effects that increase overall heterozygosity, may also promote adaptive change. Examples of adaptive evolution through hybridization are particularly evident in invasive plants; with field mustard (Brassica rapa) populations gaining herbicide resistance from genetically engineered crops (Snow et al. 1999). Also, a recombinant hybrid of two ragwort species (Senacio) is highly invasive in the UK, where neither parent is present in the wild (Abbott et al. 2009).

Hybridization may be primarily important when considering invasions by diseases, parasites and bacteria, as the lateral transfer of genes between species and divergent lineages is common (Marri et al. 2007). However, the literature on this subject is sparse, probably as a direct result of the difficulty in documenting and characterizing these invasions. Hybridization of invasive with native species has also been suggested to result in lower fitness of native species, which potentially threatens their genetic integrity (Abernathy 1994; Huxel 1999). However, this process is unlikely to be unidirectional and the converse is possible, where hybridization may also limit the fitness of invading species.

1.2.8 On what traits does selection act and when does it occur during the invasion process?

The invasion process is highly dynamic and newly introduced species will experience a multitude of selection pressures at different stages of this process.
I have highlighted the importance of the underlying genetic substrate for selection and revealed how this may facilitate or hinder adaptive evolutionary changes during invasions; but, upon which traits does selection act and at what stages of the invasion process will these evolutionary processes be operating? During initial introduction and establishment, selection may act upon traits associated with physiological tolerances to novel environmental conditions and those that enhance dispersal capacity (Lardies & Bozinovic 2008; Alford et al. 2009; Murray & Phillips 2010). Furthermore, plastic phenotypic responses are most likely to occur during these initial introductory stages, in traits largely associated with, but not limited to, the effects of abiotic factors. After introduced species have established, become invasive and expanded their range limits, evolutionary change may occur in response to selection pressures resulting from the interaction of native communities and the environment (Lee 2002). For example, environmental gradients such as temperature, rainfall or photoperiod might elicit selection pressures (Richards et al. 2006; Runcie et al. 2012; Pespeni & Palumbi 2013; Nicotra et al. 2015). Likewise the effects of resident species, through competition and predation, may lead to evolutionary adaptation (Lee 2002; Cox 2004). These pressures most commonly induce changes in traits associated with morphology, physiology and phenology; encompassing changes in reproductive traits to increase fecundity, their timing to suit environmental conditions, traits important for resource acquisition and usage and those ultimately linked to habitat preference and biotic interactions (Huey et al. 2000; Maron et al. 2004; Phillips et al. 2006; Kolbe et al. 2007b, 2012b; Urban et al. 2007).
1.3 Application of next-generation sequencing to reveal genetic basis of evolutionary change in invasive species

Recently several studies have demonstrated the utility of next-generation sequencing technologies, namely, comparative genomics and transcriptomics to identify the evolutionary genetic changes associated with invasiveness (Wang et al. 2011; Poelchau et al. 2013; Hodgins et al. 2014) and between native and invasive ranges (Hodgins et al. 2013). For example, in the fire ant (Solenopsis invicta) the GP-9 gene encodes an odorant binding protein underlying social behaviour only found in multi-queen, less aggressive and ecologically destructive invasive populations (Krieger & Ross 2002). Additionally, Mueller et al. (2014) revealed polymorphism in the dopamine receptor gene DRD4 only associated with novelty seeking behaviour in introduced populations of yellow-crowned bishops. However, most of these studies have relied on knowledge of the phenotypic traits that are the target of selection or vary between native and invasive populations, something not possible for a large proportion of invasive species. This has lead to an alternative bottom-up approach, where, based on the knowledge of genes functional roles, inferences can be made linking traits to observed genetic changes without prior knowledge of phenotypic variation (Bock et al. 2014). Recent research using this approach has associated gene expression changes in response to thermal and salinity stress to the potential for invasive spread by comparing native and introduced blue mussels (Lockwood et al. 2010; Lockwood & Somero 2011) and revealed gene expression changes that underlie several physiological functions associated with the successful invasive pest, the emerald ash borer (Mittapalli et al. 2010). Further, comparative whole-transcriptomic approaches can be used to characterize the genetic response of a wide range of cellular processes.
underlying physiological capacity (Gracey 2007; Harms et al. 2014) and uncover complex changes in gene regulatory responses and networks, which comprise rapid and versatile ways in which an organism can respond to environmental changes (Todgham & Hofmann 2009). Transcriptome analyses have been especially useful for studying molecular responses of non-model (marine) organisms to environmental stressors including: salinity fluctuations (Lockwood & Somero 2011), osmotic stress and ocean acidification (Todgham & Hofmann 2009; Pespeni et al. 2013; Zhao et al. 2014) and thermal stress (Gracey et al. 2004; Stillman & Tagmount 2009; Franssen et al. 2011; Runcie et al. 2012; De Wit & Palumbi 2013; Harms et al. 2014). Given these advances we can begin to uncover the source of genetic variation underlying adoptions that contribute to the success of some biological invasions and the potential for gene expression levels and regulation to evolve adaptively in response to local environmental conditions; a critical step in fulfilling a challenge within evolutionary biology, linking phenotypes to genotypes.
1.4 A new model, *Asterias amurensis*, to study evolutionary and invasive biology

Marine ecosystems are particularly vulnerable to invasions, with coastal ecosystems among those harbouring the highest proportion of non-native species (Grosholz & Ruiz 1996; Ruiz *et al.* 1997; Roman & Palumbi 2004; Sala & Knowlton 2006a; Roman 2006; Reusch *et al.* 2010). Many marine invasive species are transported as larvae, in large numbers, via ballast water and arrive in new environments freed of the selection pressures of their native ranges (Grosholz 2002; Bax *et al.* 2003, 2006). This high propagule pressure, introduction into novel environments and change in experienced selection pressures provides a substrate for rapid evolutionary changes, making marine invasive species particularly good models to investigate fundamental questions in evolutionary and invasive biology (Geller *et al.* 2010).

The Northern Pacific seastar, *Asterias amurensis*, (Fig. 1) is a benthic marine predator, which has recently established several invasive populations in southern Australian waters. *Asterias amurensis* is native across the northern Pacific, encompassing Russian, Korean and Japanese waters (Fig. 2) (Ward & Andrew 1995; Byrne *et al.* 1997; Matsuoka & Hatanaka 1998; Yamashita *et al.* 2005). The introduction is believed to have occurred through ballast water discharge, with the likely source being a central Japanese population (Ward & Andrew 1995). *Asterias amurensis* has been recognized as one of Australia’s most potentially damaging marine invasive species (Goggin 1998) and has the potential to drastically alter native ecosystems and affect aquaculture industry (Ross *et al.* 2002, 2003). After being introduced into southeast Tasmania, in the 1980s, *A. amurensis* has reached high densities and become the dominant
invertebrate predator in the Derwent estuary (Buttermore et al. 1994; Krieger & Ross 2002; Ross et al. 2006).

![Image of Asterias amurensis](image)

**Figure 1.** Adult and larval (mid Bipinnaria) *Asterias amurensis*

In its native range this species has caused significant damage to commercial fisheries and aquaculture (Goggin 1998), while populations in Tasmania have significantly altered soft-sediment assemblages (Ross et al. 2006). *Asterias amurensis* is reported to have spread along the east and northern coasts of Tasmania and established a large population (estimated to be over 100 million) on the mainland in Port Phillip Bay, Victoria (Goggin 1998; Parry et al. 2004). *Asterias amurensis* is suggested to have expanded its invasive range from Tasmania to the mainland rather rapidly (within 10 years) (Goggin 1998), but
since being introduced into Port Phillip Bay, its spread has until recently been largely contained (a population, totalling 300 individuals was discovered at Andersons Inlet, Victoria and eradicated in 2005. Within its native range A. amurensis occupies a wide thermal range and if invasive populations exhibit similar thermal maxima then it has the potential to expand its geographical range to Sydney in the east and around to Perth on the west coast (Bax et al. 2006; Dunstan & Bax 2007). Recently, two farther populations outside Port Phillip Bay on the mainland have been discovered (Western Port Bay, Victoria and Tidal River, Victoria) (Fig. 3), suggesting that this species may be starting to expand its geographical range along the Australian mainland after an initial lag phase. Has this recent expansion ‘revival’ been facilitated by adaptations to warmer environments accumulated in this lag phase? If so, what may constrain and limit its ultimate expansion? Currently, we have no information on the quantity of genetic variation in invasive A. amurensis populations, or the levels of genetic variation for important traits that may determine its ultimate distribution.
1.4.1 Life history

Several key aspects of the life history of *A. amurensis* are likely to be important in determining its invasion success and establishment in Australia. To date, no studies have specifically explored which traits may have been/or are under selection in Australian waters and which of these traits may be important for
determining its ultimate geographic range within Australia. *Asterias amurensis* reproduces via broadcast spawning where gametes are fertilized externally. In Australia spawning occurs during the winter months (June to September) at temperatures ranging from 10-14°C (Goggin 1998), while in native populations spawning occurs at temperatures ranging 14-17°C (Kashenko 2005a). This species is highly fecund, with individual female seastars potentially producing more than 15 million eggs during the breeding season (Goggin 1998; Parry et al. 2004; Byrne et al. 2006). *Asterias amurensis* gametes develop into gastrulae within 40-50 hours post fertilization and proceed into a long planktonic larval stage, encompassing bipinnaria and brachiolaria larval stages before settlement onto benthic substrates prior to metamorphosis (Davenport & McLoughlin 1993). Gametes can survive at temperatures ranging from a minimum of 5-10°C to 20-23°C (Morris 2002; Kashenko 2005a) while developmental time is also temperature dependent; with larvae remaining in their planktonic stage for 66-91 days at 14°C and 37-44 days at 17°C (Kashenko 2005b) and potentially up to 120 days prior to settlement and development into juvenile seastars (Bruce et al. 1995; Byrne et al. 1997). While the invasion history of this species has received some attention (description of potential invasion route and source population, see Ward & Andrew 1995; Goggin, 1998) it has not been completely resolved. Further, *A. amurensis'* potential to exhibit adaptive evolutionary change, which may facilitate its invasive potential, has not been investigated.
1.5 PhD aims and objectives

The potential for evolutionary adaptation in an invasive species will not only be determined by the novel forces of selection that it experiences, but also the underlying genetic variation present after various demographic effects. The level of standing genetic variation for heritable traits, the interaction of population bottlenecks and multiple introductions, accumulation of new mutations and epigenetic and phenotypic plasticity effects will all determine the ability of an invading species to respond and adapt to its biotic and abiotic environment. The invasive northern Pacific seastar, *A. amurensis*, in Australia provides an excellent model to study these evolutionary mechanisms in contemporary ecological time. As such, the main aims of my thesis will be to: *i*) compare estimates of diversity and divergence between native and invasive populations, *ii*) reveal how Australian invasive populations are structured, test for the presence of bottlenecks and determine effective population size (*Ne*) in invasive and native populations, *iii*) develop a transcriptomic resource for an early larval life history stage which will be used to examine the genetic basis of adaptive change during this invasion, *iv*) compare gene expression profiles between native and invasive populations to identify candidate genes potentially contributing to invasion success, *v*) test for signatures of selection on these candidate genes and then *vi*) investigate what genes are potentially involved in how native and invasive populations cope with elevated temperatures, *vii*) to understand the potential for invasive range expansion into warmer waters along the Australian mainland coast.
The thesis is structured with each chapter written as a separate publishable unit, as such, there is some overlap between the introductions to chapters but I have tried to reduce this redundancy where possible. Also, due to size constraints I have been unable to include some of the supplementary material (RSEM gene expression estimates, edgeR raw and filtered differential expression, sequence homology tables) that will accompany these chapters at publication. To rectify, these supplementary files can be made available electronically.
1.6 References


Bruce B, Sutton C, Lyne V (1995) Laboratory and field studies of the larval distribution and duration of the introduced seastar Asterias amurensis with updated and improved prediction of the species spread based on a larval dispersal. CSIRO Division of Fisheries, Hobart.

Burton OJ, Travis JMJ (2008) The frequency of fitness peak shifts is increased at expanding range margins due to mutation surfing. *Genetics, 179*, 941–950.


Elton CS (1958) *The ecology of invasions by animals and plants*. Methuen, London.


Chapter 2. Rapid divergence and maintenance of genetic diversity during range expansion in an invasive marine invertebrate

Mark F. Richardson, Alastair Hirst, Nathan Bott, Randall S. Lee and Craig DH. Sherman

2.1 Introduction

Introduction and range expansions into new environments may expose species to a wide host of novel selective pressures, including abiotic factors such as local environmental conditions (e.g. Kolbe et al. 2012) and biotic factors encompassing species interactions and community dynamics (Sakai et al. 2001). The success of invasive species facing novel selection pressures will to a large extent be determined by their ability to acclimate or rapidly adapt to these novel conditions (Lee 2002; Bock et al. 2014). This requires the presence of sufficient standing genetic variation in important life history traits, or the accumulation of novel genetic mutations that become rapidly fixed in the invasive populations (Barrett & Schluter 2008; Prentis et al. 2008). Founding events and genetic drift may however, act to deplete a population’s level of genetic diversity and adaptive potential during establishment and spread (Klopfstein et al. 2006; Peacock et al. 2009). Despite theoretical predictions for a loss of genetic diversity, recent research has demonstrated that genetic diversity can be maintained (Dlugosch & Parker 2008a), or even increase (Kolbe et al. 2004) during introduction and that genetic diversity can be preserved...
during range expansion in the introduced range (Berthouly-Salazar et al. 2013).

Further, there is evidence that invasive species can adapt despite low levels of observed genetic diversity (Dlugosch & Parker 2008c; Kolbe et al. 2012a; Rollins et al. 2013). High propagule pressure, multiple introductions from different sources, stratified dispersal and gene flow between independently colonized populations are potential mechanisms by which genetic diversity can be maintained during range expansion (Kolbe et al. 2007a, 2008; Dlugosch & Parker 2008a; Tobin & Blackburn 2008; Bronnenhuber et al. 2011). Further, maintenance or increases of genetic diversity within introduced populations, especially additive genetic variation for life history traits associated with dispersal and colonizing ability, or the accumulation of novel adaptive mutations, may increase the probability of a range expansion being successful (Ramakrishnan et al. 2010; Bock et al. 2014). These studies underscore the utility of a molecular genetics approach to reconstruct invasion history, and evaluate population connectivity and demographic processes underlying range expansion.

Marine environments are among the most heavily invaded, with coastal and estuarine habitats in particular characterized by large numbers of introduced species (e.g. Ruiz et al. 1997; Roman & Palumbi 2004; Molnar et al. 2008; Reusch et al. 2010). Many are introduced through transport in ballast water, fouling on ship hulls and marine equipment, or via intentional introductions arising from aquaculture, fisheries and the aquarium trade (Grosholz 2002). Numerous sessile and benthic marine species reproduce through broadcast spawning and undergo a pelagic larval stage. The associated high levels of fecundity, large number of propagules and long larval duration (from a few days to several months) can facilitate the rapid establishment and

The dispersive life history stages of many marine invaders can result in yearly range expansions over hundreds of kilometres (Lyons & Scheibling 2009). However, larval control over dispersal is often limited to managing their vertical position within the water column (Paris & Cowen 2004), especially during early larval stages. Coastal currents are typically highly variable, with seasonal fluctuations in current direction, temperature and strength likely to affect the distribution of dispersive larvae (Gilg & Hilbish 2003). As such, fluctuations in local ocean currents, their abiotic constituents and organismal life history traits, including reproductive schedule – i.e. spawning seasonality, planktonic larval duration and larval behaviour (Snyder et al. 2014), will be important in determining the extent of invasive marine range expansions. There is a need consequently to integrate information about introduction history, life history, hydrodynamic regimes, dispersal models and empirical genetic data in order to understand marine population connectivity and dispersal (Werner et al. 2007; Pfeiffer-Herbert et al. 2007; Galindo et al. 2010) and predict potential range expansions of invasive marine species (Connolly & Baird 2010).

One of the most successful invaders into Australian waters has been the Northern Pacific seastar, *Asterias amurensis*; a benthic marine predator native to coastal regions of Japan, Russia and the Korean peninsula (Ward & Andrew 1995). First identified in the Derwent River estuary, Tasmania, it spread across the Bass Strait to the Australian mainland and established a large population in Port Phillip Bay, Victoria (Goggin 1998; Parry et al. 2004). Its introduction to Tasmania and spread to the mainland is believed to have occurred via ballast water discharge; potentially originating from a source in central Japan (Ward &
Andrew 1995). Recently, populations have been detected outside the previous range on the Victorian mainland; Andersons Inlet, San Remo (Western Port Bay) and in the Tidal River estuary (Wilsons Promontory National Park) (Fig. 1.). *Asterias amurensis* possess long-lived planktotrophic (feeding) larvae that are capable of remaining in the water column for up to 120 days prior to settlement and development into juvenile seastars (Bruce *et al.* 1995; Byrne *et al.* 1997). Consequently, larval dispersal was implicated in the establishment of *A. amurensis* populations along both the Victorian and Tasmanian coasts (Dunstan & Bax 2007). However, no previous larval surveys have been undertaken outside Port Phillip Bay or the Derwent River estuary and the role of larval dispersal in establishing new populations remains unclear. Further, there has been no genetic analysis regarding the origin of mainland populations or the connectivity between established invasive populations. Determining the extent of larval dispersal in this species will be important for understanding its potential for range expansion, modes of connectivity between established populations and determining the source of recruits into new populations.

Here we integrate data from several sources to reveal: i) the invasion history from native Japanese populations, ii) determine the origin of newly established range edge populations, iii) genetic variation and connectivity between established invasive populations, iv) geographical extent of larval dispersal from major sources and v) the ability of hydrodynamic models to predict larval dispersal and future natural range expansions of this species along the south eastern coast.
2.2 Methods

2.2.1 Population genetic sampling and genotyping

We collected tissue samples from 544 adult individuals from 17 populations across the *A. amurensis* native and invasive ranges (Fig. 1), including two recently established populations; one in Western Port Bay and one in Wilson Promontory National Park, Victoria, Australia. Samples were immediately preserved in 95% ethanol. Genomic DNA was extracted from tube feet using DNeasy kits (Qiagen, USA), according to the manufacturer’s protocol. We used 10 microsatellite loci from Richardson *et al.* (2012) (*AAMR02, AAMR05, AAMR06, AAMR21, AAMR10, AAMR18, AAMR33, AAMR17, AAMR04, AAMR37*) for comparisons between the native and invasive ranges. To gain a better insight into the fine-scale genetic structure in the invasive range we included a further 4 microsatellite loci (*AAMR08, AAMR11, AAMR30, AAMR34*) for Australian samples, see Supplementary results. Multiplexed polymerase chain reactions (PCR) were conducted in 11μL volumes containing; 10ng of genomic DNA; 5μL PCR Master Mix (Qiagen, USA) and 4μL primer multiplex (0.26μM of each forward primer, 0.13μM of reverse primer). PCR products were amplified using Mastercycler ProS thermo cyclers (Eppendorf, USA) according to the following touchdown programme; initial hot start at 94°C for 15min; five cycles of 94°C for 45s, 65°C for 45s, 72°C for 45s; five cycles of 94°C for 45s, 60°C for 45s, 72°C for 45s; 10 cycles of 94°C for 45s, 57°C for 45s, 72°C for 45s; 20 cycles of 94°C for 45s, 55°C for 45s, 72°C for 45s; final elongation at 72°C for 15min. PCR amplicons were electrophoresed using an ABI 3130xl Genetic Analyser, incorporating LIZ 500 size standard (Applied Biosystems). Alleles were scored using GeneMapper, v3.7 (Applied Biosystems).
Figure 1. Maps of population genetic sampling sites from the native range in Japan and the invasive range in Australia. Map of Australia in inset, areas highlighted in red correspond to larger Port Phillip Bay and Derwent Estuary maps. A) Sampling sites in Port Phillip Bay, Victoria, Australia B) Sampling sites in the Derwent Estuary, Tasmania, Australia C) Sampling sites around Japan. D) Genetic sampling sites along the Victorian coast donated by black triangles, plankton sampling sites are donated by grey circles, other places named provide geographical localities relevant to hydrodynamic modelling and plankton sampling.
2.2.2 Genetic data analysis

Microchecker v2.2.3 (Van Oosterhout et al. 2004) was used to test for the presence of null alleles across all loci and revealed the presence of no null alleles in the data set used for the native versus invasive range comparison. Null alleles were detected at AAMR04, AAMR17, AAMR37 in the larger data set used for fine-scale analysis in the invasive range, however these departures were only detected in 1-3 populations and removal did not change the results of the analysis and they were therefore retained for analyses. In order to determine whether each locus assorted independently, we tested each pair-wise combination of loci for linkage disequilibrium (Weir 1979) for each population using GENEPOP v4.2 (Raymond & Rousset 1995). From a total of 783 pairwise tests, only 45 significant inter-locus associations were detected (p <0.05); however, only four of these remained significant after the application of a sequential Bonferroni correction (Rice 1989) and were not consistent across loci or populations indicating they are unlikely to reflect true linkage between loci. We used levels of single locus heterozygosity to test for departures from Hardy Weinberg equilibrium within each population. Of 170 single locus tests across 10 loci we detected 51 significant departures from expected values. However only 22 of these remained significant after a sequential Bonferroni correction of significant levels, 18 of which occurred in Japanese populations. All departures were represented by heterozygous deficits that may reflect the presence of a small number of null alleles in some populations, but there was no discernable pattern across the populations. To compare genetic diversity within and between regions, we calculated the mean number of alleles per locus (NA), the effective number of alleles (Ne) and expected heterozygosity (Nei 1978)
unbiased estimate, $uH_e$). Differences between Australian and Japanese regions were assessed using a GLM in MINITAB v16 (Minitab Inc.).

NeEstimator v2 (Do et al. 2014) was used to calculate contemporary effective population sizes ($Ne$) with the linkage disequilibrium, molecular coancestry and heterozygote excess methods. BOTTLENECK was used to test for evidence of population bottlenecks within introduced populations (Piry et al. 1999). We used the two-phase mutation model that assumes microsatellite loci mutate at a constant rate without respect to their repeat lengths (Di Rienzo et al. 1994), because most microsatellite data sets conform to this mutation model better than the stepwise mutation or infinite allele model (Di Rienzo et al. 1994; Luikart & Cornuet 1998). The two-tailed Wilcoxon test was used to detect bottlenecks and provide a more conservative alpha of 0.025 compared to the one-tailed test with an alpha of 0.5 (Luikart & Cornuet 1998).

We determined levels of population subdivision and patterns of connectivity using hierarchical $F$-statistics (Weir & Cockerham 1984) in Arlequin v3.5.1.2 (Excoffier et al. 2005). As there have been criticisms of $F_{ST}$ we also calculated $G’_{ST}$ based on (Hedrick 2005) to serve as a comparison. Single locus values of $F_{IS}$ gave no clear indication of consistent heterozygosity deficits as would be expected for inbreeding or whalund effects. An Analysis of Molecular Variance (AMOVA) was used to partition this variation between regions ($F_{RT}$), among populations within regions ($F_{SR}$), and the total variation among all populations ($F_{ST}$). Patterns of genetic differentiation within each region were further explored by separately calculating $F_{ST}$ among populations within each region and by calculating pairwise estimates of $F_{ST}$ between all population combinations using the program FSTAT v2.9.3.2 (Goudet 1995). Nei’s (1973) unbiased genetic distance ($D$) was used to examine the genetic
relationship among populations and regions using both a Principal Coordinates Analysis (in Genalex v6.5, Peakall & Smouse 2012) and the construction of an unrooted Neighbour-joining (NJ) tree in PowerMarker v3.25 (Liu & Muse 2005). The robustness of each node in the NJ tree was evaluated by bootstrapping allele frequencies 1000 times. We used STRUCTURE v2.3.4 (Pritchard et al. 2000; Falush et al. 2003), to determine whether multiple genetic groups were present within and between native and invasive regions and the membership of individuals to these genetic groups. We adopted the admixture method with correlated allele frequencies and tested the number of genetic groups ($K$) for each value between: one and 17 for the region comparison, one and seven for the native range, and one and 10 for the invasive range. For each value of $K$ we ran 10 replicates with a burn-in period of 100,000 Markov Chain Monte Carlo steps followed by an additional 500,000 iterations. We implemented the Evanno method (Evanno et al. 2005) for inferring the likely number of genetic groups using STRUCTURE HARVESTER web v0.6.94 (Earl & vonHoldt 2011). We used the CLUMPAK server (Kopelman et al. 2015) to assign group membership to genetic groups, merge runs for each $K$ and build plots.

### 2.2.3 Hydrodynamic modelling

The dispersal potential of *A. amurensis* larvae in coastal waters between Port Phillip Bay and Cape Liptrap was examined using hydrodynamic and dispersion models with Port Phillip Bay designated as a larval source, as described in Hirst et al. (2013). Briefly, the dispersal of buoyant particles, simulating the behaviour of *A. amurensis* larvae, was modelled using an 800 m grid, 8-layer, 3D hydrodynamic and dispersion model (Jenkins et al. 1999; Lee et al. 2012) that
covered the region from Port Phillip Bay to Cape Liptrap and included Western Port Bay. The model was run using data for a typical year (2004) for the period July–October, using observed forcing conditions (spring/neap tides, SW wind events, localized wave behaviour and rainfall) for this period. Hydrodynamic modelling generated information on the circulation patterns at hourly intervals and the dispersion model introduced particles into this flow field at a rate determined by expected larval release (pulse with outgoing tide from Port Phillip Bay) and survival (approximately 100 days), see Hirst et al. (2013). The simulations examined the dispersal and distribution footprint of particles from Port Phillip Bay to Cape Liptrap, and the time it takes for particles to travel from Port Phillip Heads to Cape Liptrap. The latter was used to investigate if the time taken by particles to reach Cape Liptrap via passive dispersal was consistent with the larval duration of *A. amurensis*.

### 2.2.4 Plankton sampling and *Asterias amurensis* larval identification

In order to determine the presence of larvae and validate dispersal patterns predicted by the hydrodynamic model, we sampled zooplankton assemblages during the main spawning period in August and September 2012 (Morris 2002; Dunstan & Bax 2007) from 31 locations. Sites included: four in Western Port Bay; four between Port Phillip Bay and Western Port Bay; ten off the coast of Walkerville and Tidal River; seven in Corner Inlet and off Port Welshpool; one in Tidal River estuary; two in Andersons Inlet; and three in Port Phillip Bay (shown in Fig. 1). Zooplankton assemblages were sampled as described in Hirst et al. (2013). Briefly, a 90 μm-mesh plankton net (mouth diameter 0.48 m; length 3.25 m, with cod-end jar containing 90 μm-mesh windows) was used to
sample zooplankton at each sites using a 5 minute surface tow at 2-3 knots to a
depth of 0.5m (~400m transect). The plankton net was washed and towed
without the cod-end jar for one minute between samples. Plankton samples
were fixed with RNAlater (Qiagen, USA) and stored at < 4°C. Genomic DNA was
extracted using a modified variant of SARDI’s Root Disease Testing Service
(RDTS) commercial DNA extraction method (SARDI, Australia). A modified
genetic probe developed by Bott et al. (2010) was used to measure the quantity
of A. amurensis DNA in zooplankton samples. Extracted DNA was analysed using
quantitative PCR (qPCR) as outlined Hirst et al. (2013). Detection limits for the
A. amurensis qPCR assay are approximately 2 femtograms (fg)/ul of target DNA
(i.e. $1 \times 10^{-15}$ g), substantially less than a single larva allowing for variation in
DNA extraction rates (Bax et al. 2006).
2.3 Results

2.3.1 Genetic structure of native and invasive populations

2.3.1.1 Genetic diversity

We detected relatively high levels of variability across all loci, with the number of alleles per locus varying from 6 to 16 (mean = 11.1 ± 1.14 SE). Levels of genetic diversity were similar among populations within regions; however, there were significant differences between introduced Australian populations and Japanese populations (Table S1, Fig. 2). Australian populations consistently showed lower levels of allelic diversity (NA; $F_{1, 16} = 38.3, P < 0.0001$), effective number of alleles (Ne; $F_{1, 16} = 49.7, P < 0.0001$) and expected heterozygosity (uHe; $F_{1, 16} = 40.3, P < 0.0001$) (Table S1, Fig. 2) compared to the native range. The average number of alleles per locus and expected heterozygosity were similar across invasive populations and we found no significant difference in genetic diversity between Tasmanian and mainland Victorian populations, or within recently established range edge populations (Table S3).
Figure 2. Comparisons of genetic diversity between invasive Australian and native Japanese populations of the sea star *Asterias amurensis*. Error bars represent standard error of the mean. * donates significance at P < 0.001.

2.3.1.2 Patterns of genetic differentiation

Our hierarchical analysis of $F$-statistics revealed significant levels of genetic subdivision between populations and regions. Overall, our global $F_{ST}$ ($\pm$ SE) was 0.170 ± 0.02 and significantly different from zero (P < 0.0001). Our AMOVA analysis revealed that most of this variance was due to differences within populations (83%). However, we still detected significant differentiation between regions ($F_{RT} = 0.146$, 14.5% of the total variation). Levels of genetic differentiation within regions were on average much smaller ($F_{SR} = 0.028$) and accounted for only 2.5% of the total variation. Separate analysis of genetic
differentiation among populations within each region revealed slightly higher levels of genetic differentiation among Japanese populations ($F_{ST} = 0.044; G''_{ST} = 0.261$) compared to Australian populations ($F_{ST} = 0.039; G''_{ST} = 0.081$). Pairwise estimates of $F_{ST}$ and $G''_{ST}$ varied between populations (Tables S2, S3, S5 and S6). However, the greatest amount of genetic differentiation was typically observed between Japanese and Australian populations consistent with the large amount of genetic differentiation revealed among regions in the AMOVA analysis. Pairwise values of $G''_{ST}$ showed very similar results to pairwise $F_{ST}$ values (Supplementary material, Table S2, S3, S5 and S6). Principal Coordinate Analysis (PCoA) illustrates the significant divergence between regions, with little overlap observed between Australian and Japanese populations (Fig. 3). The relationship among populations within and between regions was further explored and visualized with a Neighbour-Joining analysis based on Nei’s (1973) genetic distance (Fig. 4). Japanese and Australian populations showed clear divergence and further structuring among Australian populations was evident with Tasmanian populations forming a distinct clade from mainland Victorian populations (Fig. 4). Fine-scale analysis within the invasive region with an increased number of loci showed consistent results as those obtained with 10 loci, see Supplementary Results.
Figure 3. Principal coordinates analysis based on a genetic distance matrix (Nei 1978), for *Asterias amurensis* samples collected from invasive Australian populations and native Japanese populations.

Figure 4. Unrooted Neighbour Joining tree based on Nei’s (1973) genetic distance between populations of *Asterias amurensis* from Japan and Australia.
A pairwise comparison of the levels of genetic differentiation between Japanese and Australian populations revealed that Australian populations are most similar to samples collected from Yokohama, in Tokyo Bay (Table 1). Within Japanese populations, samples collected from Toyama Bay showed the greatest levels of genetic differentiation to all other locations ($F_{ST}$ ranged from 0.115 to 0.173, Table 1) (within native range pairwise $F_{ST}$ shown in Table S2, $G^{ST}$ shown in Table S3). Within introduced Australian populations, the greatest level of differentiation was observed between Mud Island in Port Phillip Bay (mainland Victoria) and Sandy Bay (Tasmania) with $F_{ST} = 0.102$ (Table S5). Analysis of genetic differentiation among populations within the introduced range revealed higher levels of genetic differentiation among populations in Victoria compared to Tasmania ($F_{ST (VIC)} = 0.028$; $F_{ST (TAS)} = 0.007$, Table S5, see Supplementary results). We found no consistent evidence for genetic bottlenecks within the mainland Australian populations, while three Tasmanian populations did provide significant results for a bottleneck effect (Table 2). Our analysis of effective population sizes shows no consistent estimates across analysis methods; confidence intervals of tests approach infinity (Table 2).
Table 1. Pairwise estimates of genetic differentiation (Weir and Cockerham (1984) $F_{ST}$) between native (Japanese) and invasive (Australian) populations.

<table>
<thead>
<tr>
<th>Mud Island</th>
<th>Mornington</th>
<th>Portarlington</th>
<th>Williamstown</th>
<th>San Remo</th>
<th>Tidal River</th>
<th>Cygnet</th>
<th>Lindisfarne</th>
<th>Ralphs Bay</th>
<th>Sandy Bay</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asamuchi</td>
<td>0.212</td>
<td>0.160</td>
<td>0.171</td>
<td>0.214</td>
<td>0.203</td>
<td>0.192</td>
<td>0.181</td>
<td>0.154</td>
<td>0.129</td>
<td><strong>0.126</strong></td>
</tr>
<tr>
<td>Hokkaido</td>
<td>0.190</td>
<td>0.148</td>
<td>0.147</td>
<td>0.189</td>
<td>0.183</td>
<td>0.175</td>
<td>0.152</td>
<td>0.147</td>
<td>0.137</td>
<td>0.128</td>
</tr>
<tr>
<td>Minami Sanriku</td>
<td>0.196</td>
<td>0.149</td>
<td>0.155</td>
<td>0.195</td>
<td>0.181</td>
<td>0.181</td>
<td>0.159</td>
<td>0.149</td>
<td>0.138</td>
<td>0.129</td>
</tr>
<tr>
<td>Niotojima</td>
<td>0.210</td>
<td>0.161</td>
<td>0.145</td>
<td>0.206</td>
<td>0.199</td>
<td>0.200</td>
<td>0.171</td>
<td>0.168</td>
<td>0.170</td>
<td>0.159</td>
</tr>
<tr>
<td>Yamaguchi</td>
<td>0.213</td>
<td>0.159</td>
<td>0.173</td>
<td>0.214</td>
<td>0.206</td>
<td>0.190</td>
<td>0.187</td>
<td>0.163</td>
<td>0.132</td>
<td>0.135</td>
</tr>
<tr>
<td>Yokohama</td>
<td>0.162</td>
<td><strong>0.120</strong></td>
<td>0.129</td>
<td><strong>0.169</strong></td>
<td><strong>0.156</strong></td>
<td><strong>0.149</strong></td>
<td><strong>0.149</strong></td>
<td><strong>0.142</strong></td>
<td><strong>0.127</strong></td>
<td>0.132</td>
</tr>
<tr>
<td>Toyama Bay</td>
<td>0.279</td>
<td>0.191</td>
<td><strong>0.120</strong></td>
<td>0.285</td>
<td>0.270</td>
<td>0.257</td>
<td>0.237</td>
<td>0.227</td>
<td>0.221</td>
<td>0.210</td>
</tr>
</tbody>
</table>

Values in bold font indicate the smallest $F_{ST}$ between a Japanese and Australian population.
Table 2. Estimates obtained for the presence of bottlenecks and effective population sizes across the invasive range. Population and region with sample sizes in parentheses, the two-phase model (TPM) and 2-tailed Wilcoxon p-value for bottleneck tests. Effective population size estimates from the linkage disequilibrium (LD) method, heterozygote excess (Hex) method and molecular coancestry (MC) method with the 95% confidence interval given.

<table>
<thead>
<tr>
<th>Population</th>
<th>TPM p-value</th>
<th>LD method</th>
<th>Hex method</th>
<th>MC method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ne</td>
<td>95% CI</td>
<td>Ne</td>
<td>95% CI</td>
</tr>
<tr>
<td>Victorian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williamstown</td>
<td>0.85</td>
<td>∞</td>
<td>63.5/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Mornington</td>
<td>0.38</td>
<td>7.7</td>
<td>3/16.3</td>
<td>∞</td>
</tr>
<tr>
<td>Mud Island</td>
<td>1.00</td>
<td>∞</td>
<td>∞/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Portarlington</td>
<td>0.16</td>
<td>16.8</td>
<td>7.2/68.3</td>
<td>∞</td>
</tr>
<tr>
<td>San Remo</td>
<td>0.92</td>
<td>60.9</td>
<td>16.9/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Tidal River</td>
<td>0.23</td>
<td>74.3</td>
<td>24.5/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Tasmanian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindisfarne</td>
<td>0.02*</td>
<td>∞</td>
<td>42.3/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Ralphs Bay</td>
<td>0.00*</td>
<td>75.8</td>
<td>24.5/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Cygnet</td>
<td>0.01*</td>
<td>417</td>
<td>50.6/∞</td>
<td>∞</td>
</tr>
<tr>
<td>Sandy Bay</td>
<td>0.11</td>
<td>1706</td>
<td>23.2/∞</td>
<td>∞</td>
</tr>
</tbody>
</table>

∞ denotes infinity, Ne represents effective population size estimate, * indicates significant bottleneck tests.

The STRUCTURE analysis suggested the presence of two genetic groups when considering samples across both the native and invasive ranges (Fig. 5); one group encompassing native Japanese populations and one covering invasive populations (Fig. 5a). Individuals from mainland invasive populations (Williamstown, Portarlington, Mud Island and Mornington) are similar to Tasmanian populations (Cygnet, Lindisfarne, Sandy Bay, Ralphs Bay) but also show a very small contribution from Japanese populations. Within the native range, six genetic groups were identified (Fig. 5b) with all but Minami sanriku and Asamuchi defining their own groups. Of these groups, those encompassing Yokohama, Minami sanriku, Asamuchi and Yamaguchi were most similar. Within the invasive range we identified three genetic groups (Fig. 5c). Tasmanian populations group together (with only minimal membership proportions with mainland populations) while Victorian populations were separated into two groups, with several Portarlington individuals defining their
own genetic group. The southern Port Phillip Bay populations (Mud Island and Mornington) were most similar to range edge populations, San Remo and Tidal River.
Figure 5. STRUCTURE analysis. Q plots generated for A) A. amurensis samples across both the invasive and native ranges, B) samples from the native range and C) samples from the invasive range. Each individual is represented by a vertical line showing the proportion of membership to an identified genetic cluster. Evanno et al’s (2005) method was used to determine number of genetic groups from maximum value of Delta K, for D) native and invasive range, E) native range and E) invasive range. Genetic group identified in B are shown above with grey lines.
2.3.2 Hydrodynamic modelling

The hydrodynamic and dispersal models indicate buoyant particles released from Port Phillip Bay are predominantly transported east and southeast along the coastline to Cape Liptrap (Fig. 6). The greatest concentration of particles occurs offshore of Port Phillip Heads, the Mornington Peninsula and Phillip Island. The model simulation displays limited exchange between Port Phillip Bay and Western Port Bay and lower particle concentrations for the coastline between Cape Patterson and Cape Liptrap. Hydrodynamic modelling of buoyant particles predicted *A. amurensis* larvae are transported from Port Phillip Heads to Cape Liptrap via prevailing easterly currents in approximately 800 hours (33 days). This translates to a particle velocity of approximately 170 m/hour. At this rate the section of coast between Cape Liptrap and Tidal River (approx. 45 km) can be traversed by buoyant particles in approximately 265 hours — a further 11 days.
2.3.3 Identification of *A. amurensis* larvae in plankton samples

The results of the genetic assays measuring the amount of *A. amurensis* DNA in each plankton sample are shown in Fig. 7. *Asterias amurensis* DNA was detected in 20 of the 31 samples collected in this survey. DNA was detected in all three samples collected in Port Phillip Bay, in three of four samples collected in Western Port Bay and in all four samples collected between Port Phillip Bay and Western Port Bay. DNA was also detected in all samples collected between Cape Liptrap and Shallow Inlet (South Gippsland), and in three of the five samples collected between Norman Island and Oberon Bay offshore of Tidal River. In addition, DNA was recorded in samples collected from Andersons Inlet and in Tidal River estuary. No *A. amurensis* DNA was detected in samples collected in Corner Inlet, off Barry Beach Marine Terminal or in waters adjacent to Port Welshpool.
The quantity of *A. amurensis* DNA detected in the survey varied by several orders of magnitude, indicating high variation in the size and density of larvae collected in the samples. The highest amount of *A. amurensis* DNA was recorded at Cape Schanck (Bass Strait) (25,418 pg) and the lowest amount of DNA (2 pg) was recorded at the eastern site in Western Port Bay. Samples collected offshore of Port Phillip Heads contained > 10,000 pg of *A. amurensis* DNA, whereas samples collected from Cape Liptrap to Port Welshpool (with the exception of Waratah Bay where 1002 pg was recorded) contained <100 pg. The plankton sample collected from the Tidal River estuary contained only 94 pg of DNA suggesting that only small or low densities of *A. amurensis* larvae were present.

**Figure 7.** Bubble plot showing the quantity of *Asterias* DNA (pg/5 minute plankton tow) recorded at each location between Port Phillip Bay and Port Welshpool. Figure modified from Hirst *et al.* (2013).
2.4 Discussion

Our results demonstrate that there has been substantial divergence between the native and invasive *A. amurensis* populations. Introduction into Australia likely occurred from the central Japanese area of Tokyo Bay and has resulted in the subsequent loss of genetic diversity within the invasive range (Table 1, Fig. 5A). We suggest two possible scenarios for the Australian introduction: i) introduction into Tasmania and subsequent spread to Port Phillip Bay on the Victorian mainland, or ii) there could have been two separate introductions into Australia; one into Tasmania and one into Port Phillip Bay, both originating from the same native source population. We detect differentiation between invasive populations consistent with limited gene flow between Tasmanian and mainland populations. Range expansion along the Victorian coast is clearly supported by hydrodynamic, genetic analysis and plankton sampling, and is likely to be occurring via larval dispersal from the major source population (Port Phillip Bay) with no evidence of reduced diversity in recently established populations.

2.4.1 Regional genetic structure of *A. amurensis* between the native and invasive ranges

Theory suggests introduced populations founded from a few individuals should exhibit low levels of genetic variation. The observed distribution of genetic variation between the native and invasive regions revealed significantly less genetic diversity among the invasive samples than the native samples (Table 1, Fig. 2), although invasive populations were not genetically depauperate. Further, we observed marginally greater genetic differentiation within the native range ($F_{ST} = 0.044$) than for the invasive range ($F_{ST} = 0.039$), although the
estimates are close. Likewise, AMOVA and PCoA indicate substantial and significant differentiation between the native and invasive regions with 14.5% of the total observed genetic variation attributable to differences among regions. Further, STRUCTURE analysis indicated two genetic groups exist between the native and invasive ranges, with only small representation of genetic groups across both regions. Together, these data suggest some divergence between the ranges since the initial introduction to Tasmania ~30 years ago. Despite the lower levels of genetic diversity observed within the invasive populations, consistent with a founder or bottleneck effect, our analysis of a bottleneck signature found no such evidence in the mainland Australian populations, but some evidence for Tasmanian populations (Table 2). The lack of any bottleneck signature (i.e. heterozygote excess versus the long term expected heterozygosity) in mainland populations may result from high levels of fecundity and effective population sizes. However, recent studies evaluating the ability of bottleneck tests to reliably detect declines in effective population size have indicated that the amount of pre-bottleneck genetic diversity, the timing and duration of the bottleneck and ongoing immigration can also influence and obscure genetic signals of population declines (Cornuet & Luikart 1996; Garza & Williamson 2001; Williamson-Natesan 2005; Peery et al. 2012). Alternatively, the signature observed could reflect two separate introductions from the same source in the native range, where those founding the Tasmanian invasion experienced a bottleneck and those founding Port Phillip Bay contained a greater level of source population genetic diversity. No ongoing natural gene flow is expected between native Japanese and invasive Australian ranges yet there is clearly some very small similarity between the ranges (Fig 5), probably as a result of invasive populations originating from Japan. Given the
substantial differences that exist between the native and invasive regions recent

gene flow seems unlikely. As such, the action of genetic drift, selection and

mutation within the native and invasive ranges is likely to have contributed to

the observed divergence between the ranges. The results present here are

consistent with losses of genetic diversity detected during other invasive

species introductions (reviewed in Rollins et al. 2013).

Our analyses identified Yokohama, Tokyo Bay as most similar to the

majority of Australian populations, indicating potential introduction from this

area. This analysis primarily reflects the genetic similarity between Australian

populations and Yokohama in Japan and does not confirm that it is the true

source of the introduction. However, this similarity is consistent with previous

allozyme genetic work that indicated the likely origin of the Tasmanian

introduction as being from a central Japanese population, which also implicated

Tokyo Bay as a potential source (Ward & Andrew 1995). We cautiously suggest

that the introduction is most likely to have come for this central Japanese

coastline. Tasmanian populations are marginally more similar to Tokyo Bay on

average than invasive populations from mainland Victoria (although they are

still similar), suggesting a single native source origin of the Australian

introduction into Tasmania and subsequent spread to the Victorian mainland, as

suggested previously (Goggin 1998). Further work utilizing mitochondrial DNA

markers should be conducted to test for the presence of multiple introductions

in the Australian invasion, as they can provide useful insights compared to

analyses with microsatellite markers (Rollins et al. 2011).

Samples within the native populations exhibited moderate levels of

genetic differentiation and grouped into between three and six distinct genetic

groups (Fig. 4, Fig. 5), which appear to correlate somewhat with the geographic
distribution of the samples, while the STRUCTURE analysis also indicates several genetic groups in the native range. There is also clear evidence that the Japanese populations have mixed ancestry from multiple genetic groups, likely a result of gene flow between these geographic regions. Previously, Yamashita et al. (2005) identified four genetic subpopulations around coastal Japan but did not include samples from central Honshu. Their designation of genetic groups shows marked similarity to the genetic data presented here. They identify a western Japan Sea group (seen in our data as a group encompassing Toyama Bay and Niotojima), a Hokkaido group (Hokkaido also groups out singularly in our data), northern Pacific coast group (seen in our data as a group encompassing Minami sanriku and Asamuchi) and south western group (Yamaguchi in our data). Furthermore, results presented here suggests the existence of a central subpopulation (Yokohama in our data) that is different from the closest northern Pacific group (Minami sanriku, Asamuchi), as also seen in Ward & Andrew (1995). Our data do indicate consistent gene flow between the northern cluster of native populations: Asamuschi and Minami sanriku. Given the proximity to each other and the long pelagic planktotrophic larval stages of *A. amurensis*, gene flow between populations is not entirely unfeasible, although transport by commercial shipping ballast water also clearly remains a possibility.

2.4.2 Local genetic structure and larval dispersal in the invasive range

The levels of genetic diversity observed on a fine-scale in the introduced range are not significantly different between Tasmania and Victoria (Table S1). Also, recently established expansion populations are not genetically depauperate,
suggesting these populations have captured a significant proportion of the variation from within the source population. Consequently, range edge samples may not be genetically constrained for future range expansion. Measures of pairwise genetic differentiation, AMOVA and STRUCTURE analysis all indicate small but still significant differentiation between Tasmanian and Victorian populations. These indicate that if any dispersal is occurring; either natural or anthropogenic, it is insufficient to prevent the formation of population structure between the Tasmanian and Victorian regions. Despite this significant differentiation between the regions, the differences are small, which suggests differentiation may only have begun recently. This observed pattern could also suggest a common native source origin, where the isolation between the invasive regions has not been long enough for the signature of shared ancestry to be obscured. Furthermore, samples across the introduced range are more similar to each other than any native samples (within region compared to native range), lending further support to a shared native origin hypothesis.

Previous studies have suggested that the Victorian populations in Port Phillip Bay originated from Tasmania, ~10 years after establishment in the Derwent estuary via ballast water transfer (Goggin 1998). While our data suggests this could have been the case, it also provides contradictory evidence for this hypothesis indicating that they could be a separate introduction from the same native source population (most likely still through ballast water transfer). In particular, the signature of a bottleneck in Tasmanian but not Port Phillip Bay populations, indicates they may originate from the same source but the levels of genetic diversity in founding recruits to each was different. It also seems unlikely that an expansion and founding of a bottlenecked population would not provide the same signature in the subsequent populations.
Tasmanian samples exhibit low levels of genetic differentiation indicating relatively high levels of population connectivity throughout the Derwent and Huon river estuaries. Interestingly, we find greater genetic differentiation among Victorian populations than Tasmanian. In particular, Williamstown is less similar to southern Port Phillip Bay (Portarlington, Mud Island and Mornington) and range edge populations. The currents in Port Phillip Bay form two counter-cycling gyres, separating the Bay along a northwest/southeast direction (Walker 1999), with further fine-scale modelling indicating several other small-scale eddies (O’Leary et al. 1999). This may facilitate the local retention of propagules and result in the limited structuring observed within Port Phillip Bay.

Recent studies have focused on the genetic signature during invasive range expansions (Darling & Folino-Rorem 2009; Ramakrishnan et al. 2010) and revealed the importance of larval dispersal mechanisms to this process (Bronnenhuber et al. 2011). Population genetic analysis presented here indicates that the range edge samples from San Remo and Tidal River are most similar to those from Port Phillip Bay, specifically the southern Port Phillip Bay population of Mud Island. Range edge samples exhibit consistent levels of genetic diversity compared to those from Port Phillip Bay, indicating Port Phillip Bay as the most likely source of recruits for range edge populations with high enough propagule pressure to avoid potential founder effects. San Remo samples also show similarity to the Tasmanian population from Lindisfarne; we cannot confirm a separate introduction from Tasmania, but due to high similarity to other Victorian populations, this seems unlikely. Also, based on genetic data alone we are unable to discount the potential for range edge
recruits to come from an undetected offshore population with a similar genetic signature to Port Phillip Bay.

Previous studies have revealed the utility of hydrodynamic and dispersion models for estimating larval dispersal routes in invasive species (Gilg & Hilbish 2003; Connolly & Baird 2010). Our hydrodynamic and dispersion modelling indicates range edge recruits are likely to have originated from Port Phillip Bay, via larval dispersal, supporting our population genetic inferences. The model indicates larvae are likely to travel eastwards with the prevailing currents and could reach the Tidal River estuary in ~44 days, well within the estimated larval duration of *A. amurensis* (Bruce et al. 1995; Byrne et al. 1997). The model also reveals the limited exchange of larvae from Port Phillip Bay to Western Port Bay, with larval density past Western Port Bay decreasing with distance eastwards. All predictions from the model are consistent with local natural diffusive dispersal of larvae from Port Phillip Bay.

While hydrodynamic and dispersion models provided additional evidence for the origin and mechanism of range expansion, they remain predictions. As such, we conducted extensive planktonic sampling throughout the region predicted to contain larvae and past the Wilsons promontory peninsula in order to identify *A. amurensis* larvae in the water column. We detected the presence of *A. amurensis* in plankton samples throughout the predicted larval dispersal range (Fig. 7) with lower quantities of DNA recorded as distance from Port Phillip Bay increased. This is consistent with a dilution effect on larval density as they travel further from the source in Port Phillip Bay. Also Western Port Bay and Tidal River do not appear to act as major sources of larvae into the Bass Strait, which is unsurprising given the size and recent origins of these populations. Interestingly, the model predicts only a small influx
of larvae into Western Port Bay, yet plankton sampling indicates high larval density. This discordance potentially arises from the high tidal turnover in Western Port Bay. As such the model indicates the presence of only a few larvae in the bay all the time (both incoming and outgoing tides accounted for). However, if you sample on the incoming tide, or early outgoing tide a high density of larvae could be detected due to their high density offshore from Western Port Bay. The absence of *A. amurensis* larval DNA in samples from Corner Inlet suggests that the Wilsons promontory peninsula may act as a biogeographical barrier to further larval dispersal eastwards. Modelling of winter currents within the Bass Strait supports this because currents flow predominantly southward from Wilsons promontory towards northern Tasmania, rather than around the peninsula (Greer *et al.* 2008). However, further work is needed to model the potential dispersal path of larvae from Tidal River. Together our data implicates a natural larval dispersal mechanism of range edge population establishment from the nearest large source of recruits, Port Phillip Bay. We must note however, that while the pattern is consistent with larval dispersal, we are unable to disentangle potential human-mediated transport of individuals from local sources although this remains unlikely.
2.5 Conclusions

Unravelling the invasion history and predominant mechanism for post-establishment range expansion has important implications for management practices and evolutionary biologists. Our data confirm the utility of an integrative analysis to address these fundamental questions. We demonstrate significant genetic divergence between the native Japanese and invasive Australian ranges of *A. amurensis*. Invasive *A. amurensis* population within Australia likely have an origin from the central Japanese area of Tokyo Bay. We see small levels of genetic structure among Japanese populations, identifying six potential subpopulations, however there is clearly some admixture between them. We reveal small but significant structure between Tasmanian and mainland Victoria and within Port Phillip Bay. Yet this differentiation is also small and there is clear evidence they share a common origin and that isolation is only recent. Further analysis and modelling explains high density of larvae eastwards from Port Phillip Bay. Given this high larval density it is interesting why we do not see greater recruitment of *A. amurensis* larvae along the Victorian coast. Recruitment of larvae will depend on a number of factors, including density, but also their ability to survive and adapt to a suite of novel environmental conditions. As such, further work should investigate the potential of larvae to cope with the different environmental conditions they may experience as they expand their range. Finally, our observations from population genetic analysis, hydrodynamic modelling and larval surveys are consistent with the establishment of range edge populations (Tidal River and San Remo) mediated by larval dispersal from Port Phillip Bay.
2.6 Supplementary Results

2.6.1 Genetic structure of native and invasive populations

**Table S1.** The number of samples, location and estimates of genetic diversity for the sea star *Asterias amurensis* in Japan (native) and Australia (invasive).

<table>
<thead>
<tr>
<th>Region</th>
<th>Population</th>
<th>N</th>
<th>NA</th>
<th>NE</th>
<th>Ho</th>
<th>uHe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia (VIC)</td>
<td>Mud Island</td>
<td>39</td>
<td>4.80</td>
<td>2.57</td>
<td>0.44</td>
<td>0.53</td>
</tr>
<tr>
<td>Australia (VIC)</td>
<td>Mornington</td>
<td>31</td>
<td>4.70</td>
<td>2.89</td>
<td>0.49</td>
<td>0.60</td>
</tr>
<tr>
<td>Australia (VIC)</td>
<td>Portarlington</td>
<td>24</td>
<td>5.10</td>
<td>3.50</td>
<td>0.47</td>
<td>0.66</td>
</tr>
<tr>
<td>Australia (VIC)</td>
<td>Williamstown</td>
<td>32</td>
<td>4.40</td>
<td>2.39</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>Australia (VIC)</td>
<td>San Remo</td>
<td>17</td>
<td>4.10</td>
<td>2.52</td>
<td>0.51</td>
<td>0.53</td>
</tr>
<tr>
<td>Australia (VIC)</td>
<td>Tidal River</td>
<td>35</td>
<td>4.50</td>
<td>2.89</td>
<td>0.50</td>
<td>0.55</td>
</tr>
<tr>
<td>Australia (TAS)</td>
<td>Cygnet</td>
<td>37</td>
<td>4.00</td>
<td>2.84</td>
<td>0.52</td>
<td>0.60</td>
</tr>
<tr>
<td>Australia (TAS)</td>
<td>Lindisfarne</td>
<td>35</td>
<td>4.50</td>
<td>3.00</td>
<td>0.57</td>
<td>0.61</td>
</tr>
<tr>
<td>Australia (TAS)</td>
<td>Ralphs Bay</td>
<td>37</td>
<td>4.30</td>
<td>2.97</td>
<td>0.58</td>
<td>0.61</td>
</tr>
<tr>
<td>Australia (TAS)</td>
<td>Sandy Bay</td>
<td>31</td>
<td>4.10</td>
<td>2.81</td>
<td>0.55</td>
<td>0.61</td>
</tr>
<tr>
<td>Japan</td>
<td>Asamuchi</td>
<td>38</td>
<td>7.70</td>
<td>4.29</td>
<td>0.41</td>
<td>0.68</td>
</tr>
<tr>
<td>Japan</td>
<td>Hokkaido</td>
<td>38</td>
<td>8.00</td>
<td>4.10</td>
<td>0.45</td>
<td>0.68</td>
</tr>
<tr>
<td>Japan</td>
<td>Minami sanriku</td>
<td>39</td>
<td>8.20</td>
<td>4.13</td>
<td>0.48</td>
<td>0.69</td>
</tr>
<tr>
<td>Japan</td>
<td>Niotojima</td>
<td>39</td>
<td>7.10</td>
<td>3.83</td>
<td>0.59</td>
<td>0.71</td>
</tr>
<tr>
<td>Japan</td>
<td>Yamaguchi</td>
<td>31</td>
<td>7.20</td>
<td>4.24</td>
<td>0.50</td>
<td>0.72</td>
</tr>
<tr>
<td>Japan</td>
<td>Yokohama</td>
<td>34</td>
<td>7.20</td>
<td>3.97</td>
<td>0.59</td>
<td>0.70</td>
</tr>
<tr>
<td>Japan</td>
<td>Toyama Bay</td>
<td>7</td>
<td>4.30</td>
<td>3.26</td>
<td>0.39</td>
<td>0.71</td>
</tr>
<tr>
<td>Mean Australia</td>
<td></td>
<td>31.80</td>
<td>4.45</td>
<td>2.84</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>Mean Japan</td>
<td></td>
<td>36.50</td>
<td>7.10</td>
<td>3.97</td>
<td>0.49</td>
<td>0.70</td>
</tr>
</tbody>
</table>
Table S2. Pairwise estimates of genetic differentiation (Weir and Cockerham (1984) FST) between native (Japanese) populations.

<table>
<thead>
<tr>
<th></th>
<th>Asamuchi</th>
<th>Hokkaido</th>
<th>Minami sanriku</th>
<th>Niotojima</th>
<th>Yamaguchi</th>
<th>Yokohama</th>
<th>Toyama Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asamuchi</td>
<td>0.070</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hokkaido</td>
<td>0.049</td>
<td>0.047</td>
<td>0.047</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minami sanriku</td>
<td>0.078</td>
<td>0.052</td>
<td>0.051</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Niotojima</td>
<td>0.098</td>
<td>0.084</td>
<td>0.078</td>
<td>0.108</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yamaguchi</td>
<td>0.020</td>
<td>0.002</td>
<td>0.026</td>
<td>0.044</td>
<td>0.067</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Yokohama</td>
<td>0.105</td>
<td>0.147</td>
<td>0.166</td>
<td>0.210</td>
<td>0.200</td>
<td>0.190</td>
<td></td>
</tr>
<tr>
<td>Toyama Bay</td>
<td>0.166</td>
<td>0.183</td>
<td>0.161</td>
<td>0.210</td>
<td>0.200</td>
<td>0.190</td>
<td></td>
</tr>
</tbody>
</table>

Table S3. Pairwise estimates of genetic differentiation (Hedrick (2005) G'*) between native (Japanese) populations.

<table>
<thead>
<tr>
<th></th>
<th>Asamuchi</th>
<th>Hokkaido</th>
<th>Minami sanriku</th>
<th>Niotojima</th>
<th>Yamaguchi</th>
<th>Yokohama</th>
<th>Toyama Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asamuchi</td>
<td>0.155</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hokkaido</td>
<td>0.166</td>
<td>0.183</td>
<td>0.161</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minami sanriku</td>
<td>0.112</td>
<td>0.210</td>
<td>0.200</td>
<td>0.169</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Niotojima</td>
<td>0.051</td>
<td>0.100</td>
<td>0.092</td>
<td>0.123</td>
<td>0.082</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Yamaguchi</td>
<td>0.568</td>
<td>0.532</td>
<td>0.614</td>
<td>0.412</td>
<td>0.552</td>
<td>0.541</td>
<td></td>
</tr>
<tr>
<td>Yokohama</td>
<td>0.557</td>
<td>0.557</td>
<td>0.557</td>
<td>0.557</td>
<td>0.557</td>
<td>0.557</td>
<td></td>
</tr>
<tr>
<td>Toyama Bay</td>
<td>0.568</td>
<td>0.532</td>
<td>0.614</td>
<td>0.412</td>
<td>0.552</td>
<td>0.541</td>
<td>-</td>
</tr>
</tbody>
</table>
2.6.2 Fine-scale genetic structure of invasive populations including additional microsatellite loci

2.6.2.1 Genetic diversity and Hardy Weinberg equilibrium

We detected moderate to high levels of variability across all 14 loci with the number of alleles per locus varying from 4 to 11 (mean = 6.5 ± 0.59 SE). The average number of alleles per locus and expected heterozygosity were similar across invasive populations and we found no significant difference in genetic diversity between Tasmanian and mainland populations, or within recently established range edge populations (Table S2).

Table S4. The number of samples, location and estimates of genetic diversity for the sea star *Asterias amurensis* in Australia (invasive) for the expanded data set encompassing an extra 4 microsatellite markers.

<table>
<thead>
<tr>
<th>Region</th>
<th>Pop</th>
<th>N</th>
<th>Na</th>
<th>Ho</th>
<th>He</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIC</td>
<td>MID</td>
<td>38</td>
<td>4.43</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td>VIC</td>
<td>MT</td>
<td>28</td>
<td>4.29</td>
<td>0.46</td>
<td>0.55</td>
</tr>
<tr>
<td>VIC</td>
<td>PA</td>
<td>23</td>
<td>4.57</td>
<td>0.45</td>
<td>0.59</td>
</tr>
<tr>
<td>VIC</td>
<td>WTP</td>
<td>29</td>
<td>4.50</td>
<td>0.49</td>
<td>0.51</td>
</tr>
<tr>
<td>VIC (range edge)</td>
<td>SRO</td>
<td>16</td>
<td>3.93</td>
<td>0.43</td>
<td>0.49</td>
</tr>
<tr>
<td>VIC (range edge)</td>
<td>TR</td>
<td>32</td>
<td>4.14</td>
<td>0.46</td>
<td>0.52</td>
</tr>
<tr>
<td>Tas</td>
<td>CYG</td>
<td>35</td>
<td>3.86</td>
<td>0.49</td>
<td>0.56</td>
</tr>
<tr>
<td>Tas</td>
<td>LNF</td>
<td>33</td>
<td>4.14</td>
<td>0.51</td>
<td>0.55</td>
</tr>
<tr>
<td>Tas</td>
<td>RYB</td>
<td>33</td>
<td>4.21</td>
<td>0.53</td>
<td>0.56</td>
</tr>
<tr>
<td>Tas</td>
<td>SYB</td>
<td>27</td>
<td>3.93</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>29.30</td>
<td>4.20</td>
<td>0.47</td>
<td>0.54</td>
</tr>
<tr>
<td>SE</td>
<td></td>
<td>0.59</td>
<td>0.14</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*MID= Mud Island, MT= Mornington, PA=Portarlington, WTP=Williamstown, SRO=San Remo, TR=Tidal River, CYG=Cygnet, LNF-Lindisfarne, RYB=Ralphs Bay, SYB= Sandy Bay.*

2.6.2.2 Patterns of genetic differentiation

We detected small but significant amounts of genetic differentiation both among invasive populations and regions (Victoria, Tasmania). Overall, our global estimate of genetic differentiation was small but significantly different from zero ($F_{ST} = 0.033, 95\% CI 0.022$ to $0.044$). Our AMOVA analysis revealed that most of this variance was due to differences within populations (95.96%).
However, we still detected significant differentiation between regions ($F_{RT} = 0.0284$, 2.83% of the total variation) and among populations within a region ($F_{SR} = 0.0125$, 1.21% of total variation). Pairwise estimates of $F_{ST}$ between populations are provided in Table S3.

**Table S5.** Pairwise estimates of genetic differentiation (Weir and Cockerham (1984) $F_{ST}$) between invasive (Australian) populations for the expanded data set with an additional 4 microsatellites.

<table>
<thead>
<tr>
<th></th>
<th>MID</th>
<th>MT</th>
<th>PA</th>
<th>WTP</th>
<th>SRO</th>
<th>TR</th>
<th>CYG</th>
<th>LNF</th>
<th>RYB</th>
<th>SYB</th>
</tr>
</thead>
<tbody>
<tr>
<td>MID</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>0.0040</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.0613</td>
<td>0.0239</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTP</td>
<td>0.0108</td>
<td>0.0124</td>
<td>0.0725</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRO</td>
<td>0.0069</td>
<td>0.0186</td>
<td>0.0756</td>
<td>0.0204</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>0.0096</td>
<td>0.0209</td>
<td>0.0574</td>
<td>0.0290</td>
<td>0.0247</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYG</td>
<td>0.0557</td>
<td>0.0413</td>
<td>0.0837</td>
<td>0.0436</td>
<td>0.0295</td>
<td>0.0549</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNF</td>
<td>0.0308</td>
<td>0.0224</td>
<td>0.0620</td>
<td>0.0346</td>
<td>0.0099</td>
<td>0.0354</td>
<td>0.0088</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RYB</td>
<td>0.0262</td>
<td>0.0266</td>
<td>0.0658</td>
<td>0.0267</td>
<td>0.0139</td>
<td>0.0269</td>
<td>0.0092</td>
<td>0.0032</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SYB</td>
<td>0.0583</td>
<td>0.0416</td>
<td>0.0835</td>
<td>0.0446</td>
<td>0.0182</td>
<td>0.0605</td>
<td>0.0070</td>
<td>0.0009</td>
<td>0.0107</td>
<td>-</td>
</tr>
</tbody>
</table>

Bold font indicates smallest values between range expansion populations and potential similar source populations. MID = Mud Island, MT = Mornington, PA = Portarlington, WTP = Williamstown, SRO = San Remo, TR = Tidal River, CYG = Cygnet, LNF = Lindisfarne, RYB = Ralphs Bay, SYB = Sandy Bay.

**Table S6.** Pairwise estimates of genetic differentiation (Hedrick (2005) $G''_{ST}$) between invasive (Australian) populations for the expanded data set with an additional 4 microsatellites.

<table>
<thead>
<tr>
<th></th>
<th>MID</th>
<th>MT</th>
<th>PA</th>
<th>WTP</th>
<th>SRO</th>
<th>TR</th>
<th>CYG</th>
<th>LNF</th>
<th>RYB</th>
<th>SYB</th>
</tr>
</thead>
<tbody>
<tr>
<td>MID</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>0.0144</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.1437</td>
<td>0.0675</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTP</td>
<td>0.0262</td>
<td>0.0310</td>
<td>0.1689</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRO</td>
<td>0.0226</td>
<td>0.0508</td>
<td>0.1862</td>
<td>0.0479</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>0.0230</td>
<td>0.0512</td>
<td>0.1374</td>
<td>0.0642</td>
<td>0.0596</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYG</td>
<td>0.1264</td>
<td>0.1013</td>
<td>0.2098</td>
<td>0.1002</td>
<td>0.0746</td>
<td>0.1265</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNF</td>
<td>0.0706</td>
<td>0.0562</td>
<td>0.1546</td>
<td>0.0777</td>
<td>0.0281</td>
<td>0.0812</td>
<td>0.0245</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RYB</td>
<td>0.0612</td>
<td>0.0663</td>
<td>0.1665</td>
<td>0.0606</td>
<td>0.0369</td>
<td>0.0633</td>
<td>0.0253</td>
<td>0.0106</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SYB</td>
<td>0.1306</td>
<td>0.1020</td>
<td>0.2098</td>
<td>0.0996</td>
<td>0.0478</td>
<td>0.1638</td>
<td>0.0218</td>
<td>0.0063</td>
<td>0.0286</td>
<td>-</td>
</tr>
</tbody>
</table>

Bold font indicates smallest values between range expansion populations and potential similar source populations. MID = Mud Island, MT = Mornington, PA = Portarlington, WTP = Williamstown, SRO = San Remo, TR = Tidal River, CYG = Cygnet, LNF = Lindisfarne, RYB = Ralphs Bay, SYB = Sandy Bay.
2.7 References


Bruce B, Sutton C, Lyne V (1995) *Laboratory and field studies of the larval distribution and duration of the introduced seastar Asterias amurensis with updated and improved prediction of the species spread based on a larval dispersal*. CSIRO Division of Fisheries, Hobart.


Chapter 3. *De novo* assembly and characterization of the invasive northern Pacific seastar transcriptome

This chapter represents a modified version of that in review at *PLOS one* and is formatted in accordance with the journal's specifications.

Mark F. Richardson and Craig DH. Sherman

### 3.1 Background

Invasive species generally live in ecosystems where they have short evolutionary histories and experience novel environmental conditions (Prentis *et al.* 2008; Vandepitte *et al.* 2014), which may result in strong selection on morphological and physiological traits (Keller & Taylor 2008). Research has documented that adaptive change in response to novel environments is common during the invasion process (Dlugosch & Parker 2008; Rollins *et al.* 2013; Kolbe *et al.* 2014). Yet, the source of genetic or epigenetic variation underlying adaptive change during the invasion process remains largely uncharacterised (Prentis *et al.* 2008; Bock *et al.* 2014), which has occurred in part, due to a lack of genomic information.

Recently, with the reduction in the cost of next generation sequencing technologies we can now generate large quantities of genomic data in a short time, which is particularly valuable for studies on non-model species (Martin & Wang 2011). Accordingly, we have seen several genomic resources created for
invasive species over the past few years (Smith et al. 2011; Wang et al. 2011, 2012; Ometto et al. 2013; Ioannidis et al. 2014). Transcriptome analyses in particular, are useful for studying the molecular basis of responses to different environmental conditions. For example, thermal and salinity stress elicit diverged transcriptomic responses between two species of blue mussel (genus *Mytilus*), that may explain the invasive status of one and not the other (Lockwood et al. 2010; Lockwood & Somero 2011). Additionally, transcriptome resources have helped reveal substantial shifts in the expression of metabolism and cellular repair genes which may contribute to the increased dispersal ability of invasion front cane toads (*Rhinella marina*) (Rollins et al. 2015). Gene expression data can therefore provide valuable information to understand important evolutionary processes in invasion biology, especially because it links observable genetic changes to functional roles.

Marine ecosystems are particularly vulnerable to invasions, with coastal habitats among those harbouring the highest proportion of non-native species (Grosholz 2002). Arguably, one of the most successful invaders into Australian coastal waters over the past ~30 years is the northern Pacific seastar (*Asterias amurensis*). *A. amurensis*, is a benthic marine predator that has the potential to drastically alter native ecosystems and affect aquaculture industries (Ross et al. 2002, 2006). In the national priority pests report, *A. amurensis* is ranked among the most potentially damaging invasive species in Australia (Goggin 1998). After its introduction into southeast Tasmania in the early 1980s it spread northwards and established a large mainland population in Port Phillip Bay, Victoria, which was discovered in 1995. Recently, three further populations outside Port Phillip Bay have been discovered (Inverloch, San Remo, and Tidal River, all Victoria), suggesting that this species is currently undergoing a range
expansion. Consequently, invasive A. amurensis populations provide an exciting opportunity to investigate the evolutionary response to novel environmental conditions and the underlying genetic basis of important processes in invasion ecology.

Here we report the sequencing of the A. amurensis bipinnaria larval transcriptome by RNA-Seq and the subsequent de novo assembly to produce a comprehensive set of reference contigs for gene discovery and gene expression studies. A. amurensis possess long-lived planktrophic larvae that are capable of remaining in the water column for up to 112 days before settlement and development into juvenile seastars (Bruce et al. 1995). This early life history stage is highly dispersive and more susceptible to changes in environmental conditions than adults (Kashenko 2005). As such, the early larval stages are likely to be strongly influenced by novel selection pressures. The work presented here represents the first transcriptomic resource for this species. This resource will provide a valuable public dataset for future studies on the genetic basis of invasion and for comparisons to previously characterised echinoderm transcriptomes. Identification of a list of candidate genes that might respond to several environmental stressors, previously seen to be important in other marine invasions (Lockwood et al. 2010; Lockwood & Somero 2011), can serve as a genetic resource to investigate ecological and evolutionary processes during the invasion of this species.
3.2 Results and Discussion

3.2.1 Sequencing and quality control

A cDNA library was constructed for the mid-bipinnaria larval stage of *A. amurensis*. We selected this developmental stage (10 days post fertilization) for two reasons: *i*) to avoid an overrepresentation of early developmental genes within the RNA sample and *ii*) to capture information on genes responding to environmental conditions experienced during the larval dispersive stage.

Sequencing generated 58,776,662 pairs of 100 bp paired-end reads (11.7 Gbp). Quality control resulted in the removal of 26.9% of raw sequence reads leaving 35,078,206 pairs and 15,761,360 orphan reads, from which $1.28 \times 10^9$ bases were removed during GC-content bias trimming. The second quality control phase corrected potential assembly-confounding systematic sequence read errors present in Illumina HiSeq-2000 data (Minoche *et al.* 2011; Allhoff *et al.* 2013), which can improve the accuracy of assemblies (Macmanes & Eisen 2013). This second phase identified and corrected 3,408,050 potential base call errors, accounting for 0.18% of bases in the digitally normalized read set. An error correction rate of 0.18% of bases is very small, so to examine the efficacy of error correction prior to *de novo* assembly and annotation we ran these procedures on both the error-corrected and original read sets.

3.2.2 Transcriptome assembly

To reduce the time and computational power needed to assemble the transcriptome, we adopted a strategy that combines a digital normalization (Brown *et al.* 2012) step prior to assembly. The digital normalization strategy reduced both the error-corrected and original read data sets by 74.2%, resulting
in 8,063,870 paired and 6,057,372 orphan reads that were used for assembly.

An additive multiple $k$-mer approach with Velvet and Oases generated 713,013 transcripts (> 100 bp) with N50 of 1907 bp for the error-corrected assembly and 725,467 transcripts (> 100 bp) with N50 of 2000 bp for the original data set assembly. Digital normalization followed by assembly using Velvet and Oases has been shown to generate comparable results to assemblies with Trinity, while requiring substantially less computing resources (Tulin et al. 2013).

| Table 1. Assembly summary statistics of both transcriptome assemblies for Asterias amurensis |
|-----------------------------------------------|-----------------------------------------------|
| Error corrected assembly                      | Original assembly                              |
| Number of contigs                              | 115,691                                       | 123,388                                       |
| Total contig length (bp)                       | 1.60E+06                                      | 1.78E+06                                      |
| Mean contig length (bp)                        | 1,383                                         | 1,443                                         |
| Median contig length (bp)                      | 954                                           | 976                                           |
| Minimum contig length (bp)                     | 200                                           | 200                                           |
| Maximum contig length                          | 26,819                                        | 27,497                                        |
| N50 (bp)                                       | 2,081                                         | 2,229                                         |
| Contigs <300 bp (%)                            | 9.68                                          | 9.66                                          |
| Alignment rate (%)                             | 94.05                                         | 93.82                                         |
| Discordant mappings (%)                        | 4.90                                          | 4.99                                          |

Both assemblies exhibited redundancy so we used CD-HIT-EST to merge duplicate transcripts and retain the longest possible transcripts. Only transcripts >200bp and coverage >5× were kept. Filtering assemblies by length and coverage in this way has been demonstrated to effectively clean non-reference transcriptome assemblies (Cahais et al. 2012). In total, 115,654 contigs with a N50 of 2,081 bp were generated for the error-corrected assembly and 123,388 contigs with a N50 of 2,229 bp for the original assembly (Table 1.).

Our assemblies show comparable summary statistics to other published non-model transcriptome assemblies in terms of N50, mean and median contig lengths (Du et al. 2012; Nourisson et al. 2014; Arthofer et al. 2015). Contigs
from the error-corrected and original assemblies had total lengths of $1.6 \times 10^6$ and $1.78 \times 10^6$, respectively. They exhibited similar characteristics in terms of mean and median lengths (error-corrected; mean = 1,383 bp, median = 954 bp; original; mean = 1,443 bp, median = 976 bp) and had a similar proportion of short (<300 bp) contigs. High levels of successful mapping to both assemblies’ contigs were observed (Table 1). However, mapping to the error-corrected contigs resulted in fewer discordant mappings (where one of a read pair maps to a contig but the other does not). This may be the result of the error-correction step producing fewer misassemblies, spurious contigs and less partial gene fragments. Distributions of contig length and average base coverage for both assemblies are shown in Figure 1. Both assemblies again, exhibit similarity in the distribution of contig lengths, with the error-corrected assembly generating a larger proportion of contigs <2000 bp, while the original assembly had fractionally more contigs > 2000 bp. Yet, the error-corrected assembly produced contigs with higher average per-base coverage (original, mean = 30.58, SE ± 0.985; error-corrected, mean = 33.14, SE ± 1.020).
Figure 1. Length and coverage distributions of assembled contigs. (A) Contig length (bp) distribution for the error-corrected (EC) and original (O) datasets. (B) Contig coverage, calculated as average per base coverage across a contig, for the error-corrected (EC) and original (O) datasets.

We evaluated the representation of conserved Core Eukaryotic Genes (CEGs) from the CEGMA KOG database (http://korflab.ucdavis.edu/datasets/cegma/) (Parra et al. 2007) to assess the completeness of the two assemblies. We found that 457 of the CEGs had at least one hit in both of the error-corrected and original assemblies. A total of 397 and 394 CEGs, for the error-corrected and original respectively, had successful alignments (Table 2.). Of these, fewer full-length alignments were reported for the original (366) than error-corrected
Additionally, fewer potential nonsense alignments were reported for the error-corrected assembly (28), as compared to 35 in the original. This shows the transcriptome assembly strategies we adopted were able to successfully assemble a majority of contigs that represent conserved core Eukaryotic genes. While the error-corrected read set produced a better quality assembly in terms of CEG representation (0.66 % improvement in CEGs identified; 1.75 % improvement in full-length CEG assemblies and 1.53 % fewer discordant mappings) the magnitude of difference between the assembly strategies was small. This suggests adopting an error-correction strategy before de novo assembly may not always generate substantially better de novo assemblies and should be assessed on a species by species basis. However, with transcriptome assembly of non-model species, the goal is generally to build the most comprehensive set of genes for use in further experimental work. Consequently, even marginal improvements in generating more full-length gene assemblies may be beneficial. As such, contigs from the error-corrected assembly were used for all subsequent analysis.

<table>
<thead>
<tr>
<th></th>
<th>Error corrected assembly</th>
<th>Original assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOGs with hits</td>
<td>457</td>
<td>457</td>
</tr>
<tr>
<td>Successfully aligned</td>
<td>397</td>
<td>394</td>
</tr>
<tr>
<td>Full-length alignments</td>
<td>374</td>
<td>366</td>
</tr>
<tr>
<td>Potential nonsense</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>alignments</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**3.2.3 Functional annotation**

Homology-based functional annotation was carried out on the error-corrected set of 115,654 contigs utilizing BLASTX searches against the NCBI non-redundant protein (NR), Swiss-Prot and TrEMBL databases. A total of 41,663
contigs had a match to a known protein within the three databases, covering 36% of all contigs (Table 3.). Of these, we were able to map Gene Ontology (GO) terms to 87.8% of matches, comprising 31.6% of all contigs. The largest annotated contig was 26,819 bp, which corresponds to the axonemal dynein heavy chain, a motor protein that causes sliding of microtubules in cilia and flagella. This discovery is unsurprising; the larvae are extensively ciliated and they play a role in both feeding and movement during this developmental stage (Kashenko 2005). The 73,988 (64%) contigs that did not produce a BLASTX match to a known protein are predominantly shorter (mean 964.9 bp; median 720 bp) than those annotated (mean 2,042 bp and median 1,572 bp, respectively). The group of unannotated contigs still likely contains some biologically relevant contigs that code for novel proteins and polyadenylated non-coding RNAs without similar sequences within the databases. However, our annotation rate is within the range reported (20-40%) for several other de novo transcriptome assemblies in non-model species (Vera et al. 2008; Wang et al. 2010; Hou et al. 2011; Du et al. 2012). While we successfully annotated > 41,000 contigs, this will be an over-representation of the true number of expressed A. amurensis genes. This likely occurs due to annotating contigs separately that: i) belong to the same multi-domain containing genes, ii) are fragments of the same gene and iii) constitute separately assembled allelic variants and isoforms of the same gene. We estimate 11,355 genes are expressed in A. amurensis bipinnaria larvae (based on the number of unique annotated genes), which is comparable to gene expression levels reported for several developmental stages (~11,500) in S. purpuratus (Tu et al. 2014).
Table 3. Summary of BLASTX annotations

<table>
<thead>
<tr>
<th></th>
<th>Number of contigs</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>With BLASTX match</td>
<td>41,703</td>
<td>36</td>
</tr>
<tr>
<td>- with GO annotation</td>
<td>36,504</td>
<td>31.6</td>
</tr>
<tr>
<td>- without GO annotation</td>
<td>5,199</td>
<td>4.5</td>
</tr>
<tr>
<td>Without BLASTX match</td>
<td>73,988</td>
<td>64</td>
</tr>
</tbody>
</table>

The frequency distribution of top hit E-values shows that 45.1% of annotated contigs exhibit strong homology with a matched database protein (E-value < $1.0 \times 10^{-50}$), while the majority of sequence matches (54.9%) had an E-value range of $1.0 \times 10^{-50} – 1.0 \times 10^{-3}$ (Figure 2a). Of the annotated contigs 25.4% (10,583) had a percentage similarity > 60% to a matched database protein, while 69.1% (28,814) had a similarity of > 40 % and 30.8% (12,839) had a similarity between 20 – 40% (Figure 2b.). Although our contigs exhibit a high proportion of strong matches (E-value < $1.0 \times 10^{-50}; 45.1$ %), a smaller percentage of matches (25.4 %) cover the majority of the contig sequence. This likely arises through BLAST matches to sequences sharing short, highly conserved functional domains having statistically stronger matches to our sequences. Consequently, the similarity between some matches to sequences of different species may not represent true orthology. A filtered species list is proposed to be better able to reconcile interspecific contig homology as only longer alignments with a high sequence similarity are retained (Chu et al. 2012). As such, we filtered the species-hit distribution to remove lower similarity BLASTX matches, retaining only those with similarity of > 60 % and where a contig contains > 100 amino acid residues ($n = 8,544$). This filtered species hit distribution shows the most represented species, with 47.9 % of top hits, being the purple sea urchin *Strongylocentrotus purpuratus*, which has the most extensive genomic information for echinoderms. The next most represented species is an acorn worm, *Saccoglossus kowalevskii* (13.5 %), which belongs to the Hemichordata
phylum and is closely related to the Echinodermata. A further 4.8 % of top hits come from other species belonging to the Echinodermata, with the most represented of these being the sea stars, *Patiria pectinifera* and *P. miniata*. Only a small number of top hits are to previously described *A. amurensis* proteins (28 in total), although this is expected due to the limited genomic resources for this species. This filtered species list of top hits reveals strong homology between our *A. amurensis* assembled contigs and Echinodermata proteins. While 52.7% of annotated sequences match echinoderm proteins the proportion to closely related Asteroids is small (< 4.8%) and is most likely due to insufficient sequences from phylogenetically closely related species in the searched databases (Du *et al.* 2012). Furthermore, the BLASTX annotation procedure is biased by the completeness of genome annotations for each respective genome within the searched databases (Shaw *et al.* 2012). Thus the majority of our BLAST hits are to *S. purpuratus* sequences and not closely related Asteroidia. These issues are an inherent problem with this method of annotating sequences to the available protein databases, although this approach is still used extensively (Papenfuss *et al.* 2012; Liu *et al.* 2013; Tulin *et al.* 2013; Rollins *et al.* 2015) and often represents the best available annotation method for non-model species where there are little or no genomic resources for closely related species. However, the Echinobase database contains sequence data for several echinoderm species, including another seastar and we used this data to identify Asteroid orthologs and examine our gene annotations (see below).
Figure 2. BLASTX annotation results. (A) Distribution of E-values for BLASTX top hits for each contig with a cutoff E-value of < 1.0 x 10^{-3}. (B) Similarity distribution based on the percentage (%) match of the BLASTX top hits and each query contig. (C) Filtered species distribution of BLASTX top hits where hits have a >60% similarity to query sequences containing >100 amino acid residues. 'Other' represents the grouping of species with low numbers of hits to query contigs together.
3.2.4 Comparative annotation validation to the Bat star, *Patiria miniata* proteins

The Bat star, *Patiria miniata* represents the closest echinoderm species for which extensive genomic and transcriptomic data is available (Cameron et al. 2009), allowing direct comparison to the *A. amurensis* sequences produced here. Reciprocal BLAST searches revealed 99.2 % (41,365) of assembled *A. amurensis* proteins had significant matches to 47 % (14,032) of the *P. miniata* proteins (based on gene models and transcriptome conformation) and 92.9 % (27,692) of *P. miniata* proteins had significant matches to 30.1 % (12,535) of *A. amurensis* proteins. In total, 9,739 best matches were common to both BLAST searches, representing putative orthologs between the two species (Online supplementary material). This is potentially an underestimate of the actual number of orthologs, as our data set contains both full and partial protein sequences that map to several *P. miniata* proteins. We annotated 2,423 *A. amurensis* contigs whose corresponding *P. miniata* ortholog did not have annotation information, while only 125 *P. miniata* proteins had annotation information when the *A. amurensis* ortholog did not. To determine the accuracy of our annotation pipeline we manually compared the annotations of 200 randomly selected putative orthologs that both had annotation information. 89.5 % (179) of our *A. amurensis* annotations were positives (i.e. matched correctly with those of the *P. miniata* orthologs), while 10.5 % (21) exhibited discrepancies. The high number of positive annotations reveals the efficacy of our annotation methods and validity of our dataset for future studies. The discrepancy in annotations may represent mis-annotation due to BLAST matches against short protein domains or be due to differences in gene nomenclature. For example, we annotate an *A. amurensis* protein as Serine
incorporator 5, while the \textit{P. miniata} ortholg is identified as a Serine
incorporator 3. Such mis-annotations are not unusual from electronic
annotation pipelines and can only be resolved through further manual curation.

3.2.5 Gene Ontology (GO) and KEGG annotation

To functionally categorize the \textit{A. amurensis} contigs, we mapped the associated
GO terms to the 41,663 contigs that had BLAST matches. In total, 258,322 GO
terms were mapped to 36,505 annotated contigs. GO terms are divided into
three GO categories, biological process, molecular function and cellular
component, each containing 7,144; 2,704 and 1,091 unique GO terms,
respectively. The top 10 GO assignments for each of the three categories are
detailed in Figure 3. The top represented GO terms for biological process were
transcription (2,346), regulation of transcription (1,423) and proteolysis
(1,128). For molecular function the top represented terms are from binding
domains; ATP binding (4,226), zinc ion binding (3,012) and metal ion binding
(2,598). Lastly, the top cellular component GO terms were, integral to
membrane (6,927), cytoplasm (6,338) and nucleus (6,294). We used the KEGG
Automatic Annotation Server (KASS) to provide KEGG Orthology (KO)
annotations to the annotated contigs. This resulted in 5,533 unique KO
annotations to 24,929 contigs. The top 10 represented KO annotations are
provided in (Figure 4) with the most represented being the KRAB-domain
containing zinc finger protein (208), Notch (193) and DNAH: dynein heavy
chain, axonemal (165).
Figure 3. Gene Ontology (GO) annotations. The top 10 represented GO terms for each of the GO categories: Biological Process, Molecular Function and Cellular Component. GO functional annotations are derived from similarity to the protein databases (Swiss-Prot, TrEMBL and NCBI's non-redundant database).
3.2.6 Identification of coding sequences and protein domains

Following the homology-based BLAST annotation process, 73,988 contigs (64 % of all contigs) did not have a significant match to a protein in any of the three databases. However, it is likely that some of these contigs are derived from protein-coding genes, representing novel *A. amurensis* mRNAs. These contigs may have failed to get a significant BLAST hit due to a truncated coding sequence (CDS) or their relatively short overall length compared to the annotated contigs, potentially arising from incomplete assembly. To identify unannotated potential protein-coding genes we predicted open reading frames (ORFs) and extracted amino acid sequences for the unannotated contigs. This revealed 3,176 contigs (4.29 %) that contained putative ORFs of > 100 amino acids in length. To further evaluate the quality of our annotated contigs we
performed the same ORF searching on the previously annotated contigs. Of the annotated contigs, 36,175 (86.7%) contained a putative CDS larger than 100 amino acids, giving a combined total of 39,351 predicted proteins in the error-corrected assembled contigs. This set of predicted proteins contains 18,319 unique proteins with homology-based annotations. This is larger than our estimate for unique genes expressed (11,355) with the redundancy attributable to isoform and allele specific assembly during *de novo* assembly and potential separate assembly and annotation of multi-domain proteins.

To provide further functional information, the translated ORFs were searched against the Pfam database to identify conserved protein domains. In total, 91,083 protein domains were identified, representing 4,762 unique domains. The top represented domains (Table 4.) were the Zinc finger, C2H2 type, Ankyrin repeat domain and Epidermal growth factor-like (EGF) domains. The zinc finger C2H2 domain is an ubiquitous interacting domain, reported to be involved in sequence-specific DNA binding, RNA binding, as well as mediating protein interactions (Brayer & Segal 2008). This method of identifying functional roles is also prone to the problems associated with BLAST searches, i.e. preferentially identifying short sequence matches, and electronic annotation discussed previously, see (Salzberg 2007). As such, it should only be considered a preliminary analysis of putative function.

**Table 4.** The top 10 represented Pfam domains from the protein domain annotations

<table>
<thead>
<tr>
<th>Pfam domain</th>
<th>Number of contigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>zfC2H2: Zinc finger, C2H2 type</td>
<td>5,363</td>
</tr>
<tr>
<td>Ank: Ankyrin repeat domains</td>
<td>4,886</td>
</tr>
<tr>
<td>EGF: Epidermal growth factor-like domains</td>
<td>3,447</td>
</tr>
<tr>
<td>LRR: Leucine-rich repeat motifs</td>
<td>3,029</td>
</tr>
<tr>
<td>TPR: Tetratricoptide-like repeats</td>
<td>2,847</td>
</tr>
<tr>
<td>EHand: helix-loop-helix structural domains</td>
<td>2,085</td>
</tr>
<tr>
<td>WD40: WD40 repeat containing domain</td>
<td>1,980</td>
</tr>
<tr>
<td>RRM: RNA recognition motif</td>
<td>1,771</td>
</tr>
<tr>
<td>Pkinase: Protein kinases</td>
<td>1,710</td>
</tr>
<tr>
<td>Ig: Immunoglobulin domain</td>
<td>1,291</td>
</tr>
</tbody>
</table>
3.2.7 Transposable elements

To further explore the unannotated contigs (73,988), we assessed the representation of repeating elements including retroelements and DNA transposons in the assembled *A. amurensis* transcriptome. Such transposable elements (TE) are proposed to have important roles in the adaptive process of invasive species in response to different environments, either through the maintenance of genetic variation or contribution to phenotypic plasticity (Schrader *et al.* 2014; Stapley *et al.* 2015). Additionally, mounting evidence indicates TEs are under selection following environmental stress (both abiotic and biotic) and that TE activity may have facilitated adaptation across many taxa (Oliver *et al.* 2013; Casacuberta & González 2013; Barrón *et al.* 2014; Mateo *et al.* 2014).

In our data, retroelements constitute the majority (annotated, 78%; unannotated, 72%) of TEs compared to DNA transposons (annotated, 20%; unannotated, 27%). Both sets of contigs exhibit similar representation of retroelement classes (Table S1), however, retroelements are proportionately less abundant in the unannotated than annotated set of contigs, despite fewer retroelements overall reported for the annotated set (Table S1). TEs are much less abundant in *A. amurensis* (~0.34%) than the sea urchin *Evechinus chloroticus* transcriptome (~2-3%) (Gillard *et al.* 2014). The representation of TE classes is similar between and *A. amurensis* and *E. chloroticus* although there are differences in DNA transposon diversity, particularly PiggyBac and Tourist/Harbinger, which show opposite abundances. The estimates of the number and diversity of TEs present within the *A. amurensis* transcriptome presented here can serve as a useful start for further studies investigating a potential role of TEs during the *A. amurensis* invasion and for comparisons to
other invasive species TE diversity estimates. TEs identified from RNA-Seq data may be particularly important as they are likely to include TEs close to genes and regulatory regions, which are more likely to be involved in rapid adaptation (Stapley *et al.* 2015).

### 3.2.8 Identification of candidate genes for environmental adaptation

To identify genes that may be involved in environmental adaptation in the invasive range we searched the annotated contigs’ GO terms for: ‘response to heat’, ‘response to cold’, ‘response to stress’, ‘response to salt’, ‘osmotic stress’ and ‘oxygen binding’. Previous research has shown that environmental perturbations have elicited gene expression responses in genes linked to these GO categories (Lockwood *et al.* 2010; Lockwood & Somero 2011; Logan & Somero 2011; Harms *et al.* 2014). In total, we identified 150 genes (Online supplementary material) that will serve as *a priori* candidates to investigate how populations of *A. amurensis* have adapted to novel environmental conditions across their native and invasive range.
3.3 Conclusions

Using high throughput paired-end sequencing of RNA extracted from mid-bipinnaria larvae followed by de novo assembly we derived a dataset comprising 115,654 contigs from the *A. amurensis* transcriptome. Of these, we predicted 39,351 proteins and functionally annotated 36,175 through significant matches to three protein databases. These proteins were assigned Gene Ontology and Kegg Orthology terms and annotated with Pfam protein domains to provide additional information. Overall, we identify and provide functional information for 18,319 unique proteins, comprising at least 11,355 expressed genes, with the remainder likely constituting gene isoforms and allelic variants. Of the annotated genes at least 9,739 are orthologs to *P. miniata* proteins. This data allowed the construction of a list of candidate genes that might respond to changing environmental conditions experienced during the dispersive phase in this species and will form the basis for further investigation. The relatively recent invasive history and contemporary range expansion of *A. amurensis* provides exciting opportunities to study the genetic basis of evolutionary adaptation during the invasion process. The construction of this larval transcriptome can serve as a genetic resource to investigate interesting questions in regards to ecological and evolutionary processes, such as the genetic and plastic basis of rapid adaptation and evolution occurring during the invasive range expansion.
3.4 Methods

3.4.1 Sample collection and RNA isolation

*Asterias amurensis* adults were collected from Williamstown, Victoria, Australia in July 2012. Adults were individually rinsed with UV-treated 1 μm filtered seawater in order to remove potential gamete cross-contamination and then induced to spawn by injecting with 1ml 10⁻⁵ M 1-methyladenine in filtered seawater into the coelom, as described in (Byrne 1992). Males and females were spawned dry in separate containers; gametes rinsed and then re-suspended in 50 ml filtered seawater. The concentration of sperm for each male was determined from three replicate counts using an improved Neubauer haemocytometer and sperm standardized to 1 × 10⁶ sperm ml⁻¹. Egg concentrations were assessed from three replicate counts using a Beckman multisizer™ 3 Coulter counter and standardized to 1 × 10⁴ eggs ml⁻¹. Artificial fertilization was carried out in a total volume of 100 ml filtered seawater at 14 °C, using 10,000 eggs from a single female and 100,000 sperm from a single male (sperm:egg ratio of 10:1). Gametes were left for 2 hours to fertilize at 14 °C. Embryos were transferred into 1.5L containers (density of 5 larvae per ml) and cultured at 14 °C for 10 days. Developing larvae were fed an algal diet of cultured *Chetocherous muleri* at 50,000 cells ml⁻¹.

Cultured larvae were removed at the mid-bipinnaria larval stage (10 days post fertilization) and larval aliquots (approximately 2,000 individuals) were transferred to an 1.5 ml tube, gently spun to a pellet and the supernatant removed. The larvae pellet was immediately stored in Trizol Reagent (Invitrogen, USA), homogenized, flash frozen in liquid nitrogen and then transferred to a -80 °C freezer for storage. Total RNA was extracted from this pooled sample of whole full-sib larvae using Trizol reagent according to the
manufacturer’s instructions. Total RNA was further purified using an RNeasy
spin column (Qiagen, USA) and the quality and quantity of total RNA measured
using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc,
USA).

3.4.2 Sequencing and quality control

Sequencing and cDNA library preparation was conducted commercially at the
Hawkesbury Institute for the Environment (University of Western Sydney,
Australia). Briefly, 1ug of total RNA was used to construct a single polyA cDNA
library using the Illumina TruSeq RNA protocol with the size selection step
selecting for 200bp fragments. The amplified cDNA library was sequenced on
one flow cell lane of the Illumina HiSeq-2000 platform, generating 100bp
paired-end reads. Raw sequence reads were generated using the standard
Illumina pipeline, exported in FASTQ format and deposited at the NCBI short
read archive (SRA) under the Bioproject accession number [SRR1642063].

The raw sequence reads were filtered for quality in order to generate a
high quality dataset for de novo assembly. Quality control steps were performed
with the FASTX-Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/) and
FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). First,
raw reads containing adaptor contamination were discarded. Second, reads
were filtered based on quality scores (Phred) and reads were discarded if 100%
of bases in the read did not have a minimum Phred score of 20. Next, we
computed the GC content distribution for all reads in the dataset. Random
hexamer priming is known to introduce a GC content bias in the first 13 bases of
Illumina RNA-Seq reads (Hansen et al. 2010). This bias might cause an
imbalance in read coverage that persists through the assembly process,
potentially affecting the quality of the assembly (Tulin et al. 2013). As our reads exhibited this uneven base content, we removed this bias by trimming the initial 15 bases from the reads.

For the successful implementation and generation of an accurate *de novo* assembly, the quality of the reads is paramount. While the quality control steps above can remove many assembly-confounding errors, certain specific sequence motifs can produce false positive base calling errors in Illumina HiSeq 2000 data (Minoche et al. 2011; Allhoff et al. 2013). To remove these systematic sequence read errors we utilized the Reptile v1.1 error correction pipeline (http://aluru-sun.ece.iastate.edu/doku.php?id=reptile) (Yang et al. 2010). An initial optimization run was conducted to determine the configuration for error correction, as some Reptile parameters are dependent upon the data being used. A final run was conducted with the following parameters: \( kmerLen = 14 \), \( T_{\expGoodCnt} = 8 \), \( T_{\text{card}} = 1 \), \( MaxBadQPerKmer = 6 \), \( Qlb = 67 \). Error corrected sequences were generated and used in the subsequent assembly. To assess the effect this step had on assembly quality we ran all subsequent assemblies and analyses on both the error-corrected and non error-corrected read sets.

### 3.4.3 Digital normalization and *de novo* assembly

Our data exhibit very high sequence coverage, so in order to reduce computing power and the time needed for the *de novo* assembly we conducted digital normalization, which reduces the total number of reads to be assembled. Furthermore, assemblies generated with more than 60 million reads can lead to the accumulation of errors in highly expressed genes (Francis et al. 2013). Digital normalization preferentially removes high abundance reads (reducing
redundancy) while retaining read complexity and preserving low abundance reads (Brown et al. 2012); it requires both the khmer (git://github.com/ged-lab/khmer.git) and screed software packages (git://github.com/ged-lab/screed.git). As recommended, we followed the single-pass digital normalization pipeline using normalize-by-median.py and –C 20, -k 20 and –x 4e9 parameters. These reduced read sets were assembled using Velvet v1.2.10 (https://github.com/dzerbino/velvet) (Zerbino & Birney 2008) and Oases v0.2.8 (https://github.com/dzerbino/oases) (Schulz et al. 2012). We adopted an additive multiple k-mer approach (Surget-Groba & Montoya-Burgos 2010), where k-mers ranged from 27 to 75 with a step of 4, so as to maximize contiguity in assembling highly expressed transcripts at high k-mers and sensitivity at low k-mers to assemble lowly expressed transcripts. Subsequently, these multiple k-mer assemblies were merged with another pass through Velvet and Oases at a k-mer of 27; only transcripts >100bp were kept. As anticipated, duplicate transcripts were present in the merged assemblies as a result of identical transcripts being produced at different k-mers. We used CD-HIT-EST v4.5.4 (http://weizhong-lab.ucsd.edu/cd-hit/) (Li & Godzik 2006; Fu et al. 2012) to remove this redundancy (by matching sequences at the 95% level) and retain the longest possible transcripts (now termed contigs); at this stage we filtered out contigs <200bp.

To assess our assemblies we mapped the read set pre digital normalization back to the assembled contigs using BWA v0.7.7 (http://bio-bwa.sourceforge.net) (Li & Durbin 2009). We removed potentially spurious and uninformative contigs when each contig had an average read coverage of less than 5 ×, the majority of which were short, <500 bp. This generated a reduced set of contigs for both the error-corrected and original assemblies that were
used in the following annotation pipeline. Contig statistics were computed with in house scripts. Finally, we used the python script KogBlaster.py v1.5 (https://bitbucket.org/beroe/mbari-public.git) (Francis et al. 2013) to search the assembled contigs against the CEGMA KOG database (http://korflab.ucdavis.edu/datasets/cegma/) (Parra et al. 2007) and report completeness of the Core Eukaryotic Genes.

3.4.4 Functional annotation

Functional annotation was carried out following a method described in (De Wit et al. 2012). The reduced set of contigs from the error corrected assembly was searched against the NCBI non-redundant protein database (NR) and UniProts’ Swiss-Prot and TrEMBL databases with BLASTX (Camacho et al. 2009) using an E-value cutoff of $1.0 \times 10^{-3}$; only the top 20 hits per query sequence were returned. We filtered matches to the NR database further to return only informative top hits by excluding hits to ‘predicted’ and ‘unknown’ proteins to enable more accurate mapping of Gene Ontology (GO) terms (Ashburner et al. 2000). Top hits described as ‘predicted’ were kept for the species distribution to better represent the homology between sequences. GO terms were assigned to contigs, to infer functional annotations, based on the best hit from the databases with the following preference (Swiss-Prot, TrEMBL, and NR). We combined BLAST matches from the 3 databases, functional descriptions and associated GO terms into a master annotation metatable, using custom python scripts adapted from (De Wit et al. 2012). To avoid a representation of algal genes within the final transcriptome dataset (potentially arising through the assembly of genes expressed by sequenced gut contents) we removed any sequences that only had
predominant hits to plant species or identified as constituents of:
photosynthesis, Chloroplasts, Chlorophyll or the Calvin cycle during the
annotation procedures.

The KEGG Automatic Annotation Server (KAAS) v1.6a
(http://www.genome.jp/tools/kaas) (Moriya et al. 2007) was used to annotate
contigs with Kegg Orthology (KO) codes (Kanehisa & Goto 2000). RepeatMasker
v4.0.3 (http://www.repeatmasker.org) was used to search for repeating
elements using the (22-4-2013) version of the RepBase database
(http://www.girinst.org) (Jurka et al. 2005). RepeatMasker searches DNA
sequences for interspersed repeats, including retroelements and DNA
transposons and also reports simple repeats such as microsatellites. We ran
RepeatMasker with default settings and the \(-q\), quick search option with the
species parameter set to echinoderms.

We identified candidate coding regions within assembled contigs by
searching for ORFs containing the longest stretch of uninterrupted sequences
between a start and stop codon, using TransDecoder r20131117
(http://transdecoder.sourceforge.net) with default options. This enables the
further identification of informative functional contigs, even when they do not
provide a significant match in the annotation process. Here, we consider a full-
length contig to be those that show a complete CDS and at least partial 5’ and 3’
UTR sequences. The start and stop codons are used to define the boundary
between the CDS and 5’ and 3’ UTRs. Contigs were considered to be partial CDSs
if they contained, only a start or stop codon and a combination of 5’ or 3’ UTRs,
or an uninterrupted chain of >100 amino acids and no start or stop codon. The
CDS from the contigs were transcribed into proteins and searched against the
Pfam databases (http://pfam.xfam.org) (Finn et al. 2010) to identify conserved
protein domains using HMMER v 2.3.2 (http://hmmer.janelia.org) (Eddy 2011), with an E-value of $1.0 \times 10^{-5}$. Contigs remaining without annotations or predicted CDS were further clustered with CD-HIT-EST at 90% similarity to compile a less redundant set of unannotated contigs which may represent novel A. amurensis sequences.

3.4.5 Comparison to the Bat star, Patiria miniata proteins

The protein sequences of P. miniata were downloaded from (http://spbase.org/) (Cameron et al. 2009). We performed reciprocal BLAST searches to identify putative orthologous genes following a method described (Du et al. 2012). Briefly, assembled A. amurensis contigs containing a CDS were compared to P. miniata protein using BLASTX. We then used tBLASTX to compare the P. miniata proteins to A. amurensis contigs used in the previous search. We retained only the best hit with an E-value cutoff $> 1.0 \times 10^{-3}$ and pairs of orthologous sequences were identified based on the reciprocal best matches. We randomly selected 200 reciprocal best hits, where both orthologs had annotation information and these were used to assess the efficacy of our annotation methods.

3.4.6 Identification of candidate genes associated with environmental adaptation

### 3.5 Supplementary Material

**Table S1. Summary of repeating elements**

<table>
<thead>
<tr>
<th></th>
<th>Number of elements</th>
<th>Percentage of total sequence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Retroelements</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINEs:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Penelope</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>LINES:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- CRE/SLACS</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>- L2/CR1/Rex</td>
<td>146 (80)</td>
<td>0.04 (0.01)</td>
</tr>
<tr>
<td>- R1/LOA/Jockey</td>
<td>9 (3)</td>
<td>~0 (~0)</td>
</tr>
<tr>
<td>- R2/R4/NeSL</td>
<td>1 (0)</td>
<td>~0 (0)</td>
</tr>
<tr>
<td>- RTE/Bov-B</td>
<td>49 (13)</td>
<td>0.01 (~0)</td>
</tr>
<tr>
<td>- L1/CIN4</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>LTR elements:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- BEL/Pao</td>
<td>13 (1)</td>
<td>0.1 (~0)</td>
</tr>
<tr>
<td>- Ty1/Copia</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>- Gypsy/DIRSI</td>
<td>95 (38)</td>
<td>0.05 (~0)</td>
</tr>
<tr>
<td>- Retroviral</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>DNA transposons</strong></td>
<td>186 (377)</td>
<td>0.03 (0.05)</td>
</tr>
<tr>
<td>- hobo-Activator</td>
<td>27 (65)</td>
<td>~0 (0.01)</td>
</tr>
<tr>
<td>- Tc1-Is630-Pogo</td>
<td>14 (56)</td>
<td>~0 (0.01)</td>
</tr>
<tr>
<td>- En-Spm</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>- MuDR-IS905</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>- PiggyBac</td>
<td>0 (1)</td>
<td>0 (~0)</td>
</tr>
<tr>
<td>- Tourist/Harbinger</td>
<td>80 (192)</td>
<td>0.01 (0.02)</td>
</tr>
<tr>
<td>- Other (Mirage, P-element, Transib)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Unclassified</td>
<td>21 (10)</td>
<td>~0 (~0)</td>
</tr>
<tr>
<td><strong>Total interspersed repeats</strong></td>
<td>951 (1,385)</td>
<td>0.18 (0.16)</td>
</tr>
</tbody>
</table>

|                        |                    |                                  |
| Small RNA              | 44 (13)            | 0.03 (~0)                        |
| Satellites             | 0 (0)              | 0 (0)                            |
| Simple repeats         | 21,714 (40,512)    | 1.39 (3.25)                      |
| Low complexity         | 3,237 (4,777)      | 0.54 (1.03)                      |
| **Total**              | 25,946 (46,687)    | 2.13 (4.44)                      |

~ Denotes approximate value. Values in parentheses correspond to those from the set of assembled contigs for which there were no matches within any of the protein databases. Values not in parentheses are derived from the set of contigs with matches in the protein databases.
3.6 References


Bruce B, Sutton C, Lyne V (1995) *Laboratory and field studies of the larval distribution and duration of the introduced seastar Asterias amurensis with updated and improved prediction of the species spread based on a larval dispersal*. CSIRO Division of Fisheries, Hobart.


Chapter 4. The molecular basis of a successful invasion: comparative RNA-Seq analysis of native and invasive *Asterias amurensis*

Mark F. Richardson and Craig DH. Sherman

### 4.1 Introduction

Invasive species provide valuable study systems to evaluate evolutionary and ecological processes occurring during the colonization of new habitats because they represent natural experiments in a contemporary time frame (Prentis & Pavasovic 2013; Bock *et al.* 2014). Exposure to novel environmental conditions experienced upon introduction can generate strong selective pressures, eliciting rapid evolutionary change in morphological, physiological and life history traits (Phillips *et al.* 2006; Dlugosch & Parker 2008c; Alford *et al.* 2009). Considerable research has focused on the ecological aspects during invasions (see Lowry *et al.* 2012) and in recent years much effort has been directed at revealing underlying evolutionary processes responsible for adaptive changes often seen in invasive populations (reviewed in Rollins *et al.* 2013). These studies have primarily informed us of the changing features of the organisms themselves (Gilchrist *et al.* 2001, 2004), sources and distributions of genetic variation during invasions (Tsutsui *et al.* 2000; Koskinen *et al.* 2002b; Roman & Palumbi 2004; Kolbe *et al.* 2007a) and the potential genetic processes underlying these changes (reviewed in Lee 2002; Bock *et al.* 2014). With the recent advances in next-generation sequencing and the increased availability of genomic
information, we are making steps to uncover the genetic basis of adaptive change during invasions, including signatures of selection and gene regulatory processes (see Hodgins et al. 2013), although this is still in its infancy and has largely focused on plants. Nonetheless, given these advances we can begin to uncover the source of genetic variation underlying adaptations that contribute to the success of some biological invasions and the potential for gene expression levels and regulation to evolve adaptively in response to local environmental conditions.

Recently, several comparative genomic and transcriptomic approaches have been used to identify the genetic changes associated with invasiveness (Rollins et al. in press; Wang et al. 2011; Poelchau et al. 2013; Hodgins et al. 2014) and between native and invasive ranges (Hodgins et al. 2013). However, no common set of unique traits or genetic mechanism that promote invasiveness has been identified (see Weinig et al. 2007). Several studies have underscored the importance of particular genes underlying adaptive phenotypic traits in invasive species. For example, in the fire ant (Solenopsis invicta) the GP-9 gene encodes an odorant binding protein underlying social behaviour only found in multi-queen, less aggressive and ecologically destructive invasive populations (Krieger & Ross 2002). Additionally, Mueller et al. (2014) revealed polymorphism in the dopamine receptor gene DRD4 only associated with novelty seeking behaviour of introduced yellow-crowned bishops. These examples have relied on knowledge of the phenotypic traits that are the target of selection or vary between native and invasive populations, something not possible for the majority of invasive species. Consequently, a bottom-up approach has been utilized, where, based on the knowledge of genes functional roles, inferences can be made linking traits to observed genetic changes without
prior knowledge of phenotypic variation in traits underlying invasion (Bock et al. 2014). Recent research using this approach has associated gene expression changes in response to thermal and salinity stress to the potential for invasive spread by comparing native and introduced blue mussels (Lockwood et al. 2010; Lockwood & Somero 2011). Further, Mittapalli et al. (2010) revealed gene expression changes between tissues that underlie several physiological functions associated with the invasive emerald ash borer.

The northern Pacific seastar, *Asterias amurensis*, is recognized as one of Australia’s most potentially damaging marine invasive species (Goggin 1998). Our work and previous studies have identified the most likely source of this introduction as being from a central Japanese population near Tokyo Bay (Chapter 2). This species established a large population on the Australian mainland in Port Phillip Bay, Victoria, in the mid 1990s (Goggin 1998; Parry et al. 2004) and in recent years has established a number of populations along the Victorian coast. *A. amurensis* possess a long planktonic larval stage (~40 – 100+ days) providing the opportunity for long distance dispersal and establishment of new populations in the invasive range (Chapter 3). However, the larval stage is also recognized as the most vulnerable life history stage, with dispersing larvae more susceptible to environmental changes than adults (Kashenko 2005a). We hypothesize that selection on the larval stage is likely to be particularly strong and may lead to rapid changes in larval traits. As such, *A. amurensis* larvae provide an exciting opportunity to examine selection at the molecular level and identify genes that may be important for successful contemporary invasion. Here, we use a comparative transcriptomics approach to assess differences in gene expression between invasive Australian and native Japanese populations and test for signatures of selection on candidate genes.
We discuss the functional roles for several genes identified as differentially expressed or under putative selection in order to determine the molecular basis of local adaption that may contribute to the success of this invasive species.
4.2 Methods

4.2.1 Sample collection and larval rearing

*Asterias amurensis* were collected from Port Phillip Bay (Williamstown, Victoria, Australia, 37.8610° S, 144.8850° E) in July 2013, and from Tokyo Bay (Hakkeijima Island, Yokohama, Japan, 35.3374° N, 139.6473° E) in March 2013. Adults were spawned and used for artificial fertilisations within 24 hrs of collection using the protocol described in Chapter 3. Briefly, adult *A. amurensis* were individually rinsed with UV-treated 1 μm filtered seawater to remove potential gamete cross-contamination. They were induced to spawn by injecting with 1 ml 10⁻⁵ M 1-methyladenine in filtered seawater injected into the coelom. Males and females spawned dry in separate containers; gametes were rinsed and re-suspended in 50 ml filtered seawater. The concentration of sperm for each male was determined using an improved Neubauer haemocytometer (from three replicate counts) and standardized to 1 × 10⁶ sperm ml⁻¹. Female egg concentrations were assessed from three replicate counts using a Beckman multisizer™ 3 Coulter counter and standardized to 1 x 10⁴ eggs ml⁻¹. Artificial fertilization was carried out in 100 ml filtered seawater at 14°C, using 10,000 eggs from a single female and 100,000 sperm from a single male (sperm:egg ratio 10:1). For each population six replicate male × female (M × F) crossings were conducted, where each individual was used once. This generated six biological replicates of full-sibling larvae for each population. Gametes were left for 2 hours to fertilize at 14 °C. Embryos were then washed with filtered seawater, transferred into 1L containers (density of 5 larvae per ml) and cultured at 14 °C for 10 days. We chose this temperature as it represents the shared natural spawning temperature experienced across both ranges (see Chapter 1). These consisted of two technical replicates for each biological
replicate, totalling 12 samples per population. Developing larvae were maintained at constant common garden conditions in both Japan and Australia. The difference in experimental month represents the equivalent winter spawning period between the northern and southern hemispheres only. Larvae were cultured in UV-treated 1 μm filtered seawater, maintained at a constant 14 °C, 12:12h light:dark cycle and fed a standard algal diet of cultured *Chetocherous muleri* culture at 50,000 cells ml$^{-1}$.

### 4.2.2 RNA isolation

Cultured larvae were collected at the mid-bipinnaria larval stage (10 days post fertilization) in order to avoid a possible overrepresentation of early expressed developmental genes in the total RNA sample. Larvae aliquots (approximately 2,000 individuals) from each cross (including experimental technical replicates) were transferred to separate 1.5ml tubes, gently spun to a pellet and the supernatant removed. The larvae pellet was immediately stored in RNAlater (Qiagen, USA), incubated at 4 °C for 24 hrs and then transferred to a -80 °C freezer for storage. Total RNA was extracted from each sample of whole full-sibling larvae using an RNeasy spin column (Qiagen, USA) and Qiashredder (Qiagen, USA) pre-extraction, according to the manufacturers protocol. The quality and quantity of total RNA for each sample was measured using an Agilent 2100 Bioanalyser (Agilent Technologies, USA) to ensure that >1 ug of total RNA was extracted (range 1.1 – 1.6 ug) and sample had high RNA Integrity Numbers (RIN values: range 7.2 – 9.8). RNA from the two experimental technical replicates per sample was combined, resulting in six samples per population. We combined the technical replicates for each of the larval rearing
and RNA extraction protocols to try and minimize the bias any single protocol might have introduced before sequencing.

4.2.3 Sequencing and Quality control

To each 1 ug of total RNA from each sample we added 2 ul of a 1:100 dilution of either mix 1 or mix 2 of the External RNA Controls Consortium (ERCC) spike-in control RNA (Life Technologies, USA) (see Appendix 2 for sample and ERCC control mix details). These external RNA spike-in controls serve as positive and negative controls for differential gene expression analysis and allow the assessment of technical performance and method validation (Munro et al. 2014). Sequencing and cDNA library preparation was conducted commercially at Macrogen, South Korea. Briefly, total RNA was used to construct a single cDNA library using the Illumina TruSeq RNA poly A protocol (Illumina Inc, USA); with the size selection step selecting for ~280 bp fragments. The amplified cDNA library was sequenced on two flow cell lanes of the Illumina HiSeq-2500 platform, generating 101 bp paired-end reads. Samples for both populations were randomly assigned to each flow cell lane to minimize technical bias during sequencing. Raw sequence reads were generated using the standard Illumina pipeline and exported in Fastq format.

Quality control steps were performed with cutadapt v1.3 (Martin 2011) and FASTX-Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Raw reads containing adaptor sequences were trimmed then reads were filtered based on quality scores (Phred), with reads discarded if 95% of bases across the read did not have a minimum Phred score of 30. We computed the GC content distribution for all reads in the dataset and removed the common GC bias (see Hansen et al. 2010) in the initial 13 bases of the reads. One of the samples from
Port Phillip Bay (P314) contained a large proportion of adaptor sequences and low quality reads and was excluded from all further analyses.

4.2.4 Read Mapping and expression quantification

De novo transcriptomes often suffer from redundancy and contain multiple contigs per unigene (here considered a complete single gene mRNA consensus sequence) resulting from isoform or allele specific assembly, or the result of uneven transcript coverage. Mapping reads to references that contain this redundancy may provide inaccurate gene expression quantification as reads can map to several contigs. As such, we built a reference that contained a subset of the de novo assembled contigs produced in Chapter 3 (where we only included the longest assembled contig for a given unigene) and the ERCC RNA spike-in sequences, totalling 18,411 unique transcripts. Accordingly, this reference was restricted to genes that had functional annotations or identifiable amino acid coding sequences, primarily because down-stream analysis of gene function is restricted to those genes with functional information. Reads from each sample were mapped to this reference using RSEM v1.2.14 (Li & Dewey 2011) and Bowtie2 v2.1.0 (Langmead & Salzberg 2012), using the default parameters, which quantified expression on a per-gene basis as raw read counts. RSEM does not combine both paired-end and single-end read data into gene expression read counts, which our datasets contained after adaptor and quality filtering. Therefore, we only considered paired-end reads when generating read counts. Trials where we considered all reads as single-ended resulted in similar read counts but had higher levels of mutli-mapping reads (data not shown).
4.2.5 Evaluation of differential expression and diagnostic performance

To evaluate the technical performance of the differential expression experiment, expression values of individual ERCC transcript ratios (mix 1 and mix 2) were analysed with erccdashboard (Munro et al. 2014). This program does not compare both ERCC mix types across populations simultaneously. We therefore chose to examine experimental performance between the populations using samples with mix1 from the native population and samples with mix 2 from the invasive population as this aligned with our predicted gene expression differences. This package enables the examination of diagnostic performance with Receiver Operator Characteristic (ROC) curves and Area Under the Curve (AUC) statistics; assessment of the experiments lower limit of differential expression detection (LODR) and expression ratio variability and bias. These measures are based on the intrinsic ERCC transcript abundances that were added to the samples before library preparation and these external controls are essential to understand the validity of digital gene expression experiments (Munro et al. 2014). ROC curves and the corresponding AUC statistic are based on the discrimination of ERCC true positive and negative control expression values and relate to our ability to call true differentially expressed genes as differentially expressed. An AUC of 1 represents prefect ability, while an AUC of 0.5 represents no reliability when calling differentially expressed genes. LODR estimates assess diagnostic performance as a function of abundance relative to a given fold-change (i.e. how much of a signal is needed to have confidence that a given fold-change in expression will be detected as significantly differentially expressed; see Munro et al. (2014) for more details. RSEM generated raw read counts for the ERCC sequences and reference genes were used as input for
erccdashboard when gene CPM > 1 (see below for explanation) with the following parameters: erccmix "RatioPair", erccdilution 1/100, spikeVol 2, totalRNAmass 1. These parameters result from the amount, type and dilution of the ERCC ratio mixes added to each sample (Supplementary Table S2). As per the recommendation of Munro et al. (2014), we chose a high false discovery rate (FDR) (Benjamini & Hochberg 1995) threshold (FDR= 0.1) because only a few differentially expressed genes were used to estimate parameters. We utilized the diagnostic performance metrics to set our acceptable FDR and $p$-values for calling differentially expressed genes.

### 4.2.6 Digital differential gene expression

Differentially expressed genes were identified with the BioConductor tool edgeR (Chen et al. 2014), which requires the R (R Development Core Team 2011) programming language. We adopted the classic approach for the comparison of two groups (native and invasive). Firstly, RSEM quantified expression counts were filtered for lowly expressed genes. Genes were removed when they had < 1 count per million reads (CPM) in $n = 5$ or fewer libraries (where $n$ is the smallest number of replicates per treatment). These genes are removed as their expression value is too low to be detected as significant and their inclusion results in a reduction of power to detect differential expressed genes (Anders et al. 2013). We conducted multi-dimensional scaling (MDS) analysis (similar to principal coordinates analysis) based on distances between samples (in terms of expression estimates) to visualise the variation within and among populations before differential expression testing. MDS analysis allows the representation of the dimensionality in expression for clustering population
samples without the assumption of multivariate normality. CPM filtered counts were then TMM-normalized (Trimmed mean on M-values) (Robinson & Oshlack 2010), which accounts for library size and expression bias between replicates. The quantile-adjusted conditional maximum likelihood method (Hansen et al. 2012b) was used to estimate dispersions and the exact test was used to call differentially expressed genes corrected for false discovery rate $p < 0.05$. We utilized scripts from the trinity project (Haas et al. 2013) for further filtering of differentially expressed genes, including calling differentially expressed genes identified with edgeR based on ERCC derived $p$-values, minimum reliably detectible fold-change (fold change was determined based on an invasive/native comparison) and for building MA plots and heatmaps. We filtered differentially expressed genes keeping those with a minimum absolute log$_2$ fold-change $> 1$ (which represents a two fold-change) and when their $p$-value $< 0.01$. This provides a conservative cut off for evaluating significant differential gene expression between the native and invasive range as the values chosen are based upon the intrinsic gene expression differences between the regions and are supported by the addition of true and false positives for the expression ranges provided by the ERCC controls (see Munro et al. 2014; Rollins et al. in press). TMM-normalized read counts were converted to fragments per kilobase of exon per million fragments mapped (FPKM), to allow for better clustering and representation of genes on heatmaps. Genes were considered differentially expressed when the minimum absolute log$_2$ fold-change was $> 1$ (which represents a 2 fold-change) and their $p$-value $< 0.01$. 
4.2.7 Gene ontology enrichment analysis

We used enrichment analysis to identify whether functional categories associated with an *a priori* suite of genes (those from the reference set) contain more differentially expressed genes for each population than expected by chance. RNA-seq data sets, however, are generally characterized by transcript length heterogeneity, which can bias enrichment analysis (Young *et al.* 2014). For example, longer transcripts can generate higher read counts, which increase the probability of the gene being called as differentially expressed. Categories containing longer genes on average are therefore more likely to contain differentially expressed genes. To account for this we adopted the goseq method (Young *et al.* 2014), which is available as a Bioconductor R package, and takes length bias into account. We extracted the Gene Ontology (GO) categories (Biological Process, Molecular Function and Cellular Component) for genes from the *de novo* assembly Annotation Table (see Chapter 3). Each set of GO categories for differentially expressed genes from the native and invasive population comparison were assessed for enrichment compared to the reference set of GO categories. Enrichment was determined using a FDR corrected *p*-value < 0.05.

4.2.8 Outlier loci analysis

For the identification of SNPs we used the Genome Analysis Toolkit (DePristo *et al.* 2011) and followed the Broad Institutes best practices workflow for calling SNPs in RNA-Seq data (http://gatkforums.broadinstitute.org/discussion/3891/calling-variants-in-rnaseq). Briefly, we mapped reads from each individual family cross (ie. each
individual larval pool) to our reference transcriptome using STAR v2.1.4 (Dobin et al. 2013) and the 2-pass method, as this has been shown to provide better specificity in SNP calling for RNA-Seq data (Engström et al. 2013). We omitted the base recalibration step, as this requires known variant sites and ran the HaplotypeCaller function with the following parameters, _nct 4, recoverDanglingHeads, dontUseSoftClippedBases, stand_call_conf 20, stand_emit_conf 20_. We only called SNPs that were biallelic despite there being a possible four alleles in the larval pool (2 from each of the male and female parents) as it is difficult to determine accurately which SNP would be real compared to a sequencing artefact in the case of four possibilities. The list of polymorphic sites was filtered to retain sites only if they had high quality SNP genotypes across all individuals (phred-scale threshold 20). Individuals were grouped according to population. The list of polymorphic SNPs was further filtered (as outlined in De Wit et al. 2014) by removing sites with a minor allele frequency < 5%, those not in Hardy-Weinberg equilibrium (_p_ = 0.05) or contained within an annotated CDS. To scan the transcriptome for _F_{ST} outlier_ SNPs we used the Lositan selection workbench (Beaumont & Nichols 1996), using the forced neutral mean _F_{ST}, 250,000 simulations and a 95% confidence interval. We further filtered the resulting list of outlier SNPs, only retaining genes where all of the SNPs were outliers and exhibited the same putative selection regime. Gene ontology (GO) enrichment analysis was performed using the method previously described, against the same set of reference genes used. The top 100 represented GO terms associated with these genes were plotted using REVIGO (Supek et al. 2011) with the SimRel semantic clustering of similar GO functions when using the whole UniProt database to source GO functional annotations.
## 4.3 Results

### 4.3.1 Sequencing and read mapping

Sequencing of cDNA libraries yielded between 62.2 and 109.4 million reads per sample (Table 1). After adaptor removal and quality filtering of reads, we obtained between 31 and 87.5 million reads per sample that contained both paired reads (Table 1). All reads that passed this filtering step had average Phred scores $> Q30$, which corresponds to a base call accuracy of $> 99.9\%$. Between 68.4 and 72.8 % of reads successfully mapped and were properly paired (where both paired-end reads map to the same gene) to the 18,411 reference genes, including the 92 ERCC RNA control sequences. Counts per million mapped reads (CPM) for all of the sequenced samples ranged from a minimum of 0 in all samples, to a maximum of 1,053,152 in the native population (replicate J114). After filtering of genes that had CPM $< 1$ in 5 samples, we quantified expression (in read counts) for the remaining 13,907 genes from the reference transcriptome, to be used in differential expression analysis. Assessment of experimental performance and diagnostic power from the ERCC ratio analysis revealed only a small likelihood of measurement error and good power to detect differential gene expression, at the 2- and 4-fold levels (Supplementary Results, Fig. S1). Based on the performance metric obtained we set a minimum fold-change threshold of 2 and a $p$-value threshold of $p < 0.01$ for filtering endogenous differential expression between the native and invasive populations in subsequent analysis.
<table>
<thead>
<tr>
<th></th>
<th>Total Reads</th>
<th>Reads after filtering</th>
<th>Mapped reads</th>
<th>Mapped %</th>
<th>Min RC</th>
<th>Max RC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J114</td>
<td>109,405,130</td>
<td>87,467,780</td>
<td>60,105,750</td>
<td>68.72</td>
<td>0</td>
<td>1,053,152</td>
</tr>
<tr>
<td>J214</td>
<td>100,798,756</td>
<td>73,824,320</td>
<td>50,498,646</td>
<td>68.40</td>
<td>0</td>
<td>551,393</td>
</tr>
<tr>
<td>J314</td>
<td>71,677,612</td>
<td>53,544,428</td>
<td>36,634,926</td>
<td>68.42</td>
<td>0</td>
<td>420,031</td>
</tr>
<tr>
<td>J414</td>
<td>63,708,076</td>
<td>49,247,200</td>
<td>34,327,816</td>
<td>69.71</td>
<td>0</td>
<td>419,310</td>
</tr>
<tr>
<td>J514</td>
<td>64,302,226</td>
<td>47,796,498</td>
<td>33,544,960</td>
<td>70.18</td>
<td>0</td>
<td>358,497</td>
</tr>
<tr>
<td>J614</td>
<td>70,788,870</td>
<td>57,255,084</td>
<td>39,088,136</td>
<td>68.27</td>
<td>0</td>
<td>465,429</td>
</tr>
<tr>
<td>Invasive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P114</td>
<td>68,544,216</td>
<td>31,026,208</td>
<td>21,496,272</td>
<td>69.28</td>
<td>0</td>
<td>359,644</td>
</tr>
<tr>
<td>P214</td>
<td>67,838,726</td>
<td>51,255,528</td>
<td>37,302,312</td>
<td>72.78</td>
<td>0</td>
<td>437,198</td>
</tr>
<tr>
<td>P414</td>
<td>64,702,424</td>
<td>48,189,934</td>
<td>34,450,026</td>
<td>71.46</td>
<td>0</td>
<td>424,820</td>
</tr>
<tr>
<td>P514</td>
<td>67,372,218</td>
<td>49,555,502</td>
<td>35,094,666</td>
<td>70.82</td>
<td>0</td>
<td>383,471</td>
</tr>
<tr>
<td>P614</td>
<td>62,247,820</td>
<td>44,472,982</td>
<td>31,834,726</td>
<td>71.58</td>
<td>0</td>
<td>332,463</td>
</tr>
</tbody>
</table>

### 4.3.2 Differential expression between native and invasive populations

Pair-wise comparisons of the invasive versus the native population gene expression measurements identified 2,488 significantly differentially expressed genes after FDR correction. When filtered based on our LODR metrics \((p < 0.01, \text{fold-change} > 2)\) from the ERCC ‘spike ins’, we retained 347 significant differently expressed genes. Of these, 149 had the greatest expression (were up-regulated) in the invasive population, while 198 were down-regulated in the invasive population (Fig. 2.). The 10 most significant differentially expressed genes, are presented in Table 2. Multidimensional scaling analysis revealed overall variation in gene expression between the populations was large (Biological coefficient of variation, BCV = 0.330, Supplementary Results Fig. S2). With variation in expression levels within the native source population (BCV = 0.318) greater than that exhibited by the invasive population (BCV = 0.194).
Overall gene expression was more similar among individuals from the same population than between individuals from different populations. Differentially expressed genes exhibited greater consistency in expression levels for the up-regulated genes across invasive population replicates compared to down-regulated genes (Fig. 3).

**Figure 2.** MA plot of differentially expressed genes identified between the native and invasive population. For each gene the log2 fold-change (y-axis) is plotted against the average log2 expression (x-axis) in counts per million mapped reads. Each dot represents a transcript and significant differential expression is indicated as red dots with at most 0.05% FDR. The blue line denotes a 2-fold-change in expression.
Table 2. Differential gene expression between native and invasive populations. Gene name, \( p \)-value for difference in expression and log fold-change are given.

<table>
<thead>
<tr>
<th>Gene</th>
<th>( p )-value</th>
<th>Fold-change (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome b-245 light chain (\textit{CYBA/p22-phox})</td>
<td>6.46E-91</td>
<td>11.78</td>
</tr>
<tr>
<td>Type II AFP (\textit{AFPII})</td>
<td>2.78E-49</td>
<td>7.95</td>
</tr>
<tr>
<td>Actin (\textit{ACT}) *</td>
<td>1.31E-48</td>
<td>-12.47</td>
</tr>
<tr>
<td>Uncharacterized protein</td>
<td>2.30E-41</td>
<td>2.89</td>
</tr>
<tr>
<td>Pyridoxine 5’-phosphate oxidase (\textit{PNPOx})</td>
<td>3.70E-35</td>
<td>3.65</td>
</tr>
<tr>
<td>LON peptidase N terminal domain and ring-finger protein 1 (\textit{LONPRF1})</td>
<td>9.42E-27</td>
<td>2.52</td>
</tr>
<tr>
<td>Proteosome subunit beta type-5 (\textit{PSMB5})</td>
<td>1.32E-26</td>
<td>5.10</td>
</tr>
<tr>
<td>Cytochrome p450 3A12 (\textit{CYP3A12})</td>
<td>2.76E-26</td>
<td>3.50</td>
</tr>
<tr>
<td>Cytochrome p450 3A41 (\textit{CYP3A41})</td>
<td>3.23E-24</td>
<td>3.36</td>
</tr>
<tr>
<td>Pleckstrin homology domain containing family g member 4b (\textit{PLEKHG4B})</td>
<td>4.92E-24</td>
<td>9.23</td>
</tr>
</tbody>
</table>

* denotes gene down-regulated, all other genes are up-regulated in the invasive population. Abbreviation of gene name is give in parentheses.

Figure 3. Heatmap of differentially expressed genes between the native and invasive population. Genes are clustered by similarity in per-gene gene expression measurements and by replicate across-gene gene expression. Invasive replicates marked with the red bar and native with the blue bar. Expression values are coloured by magnitude, higher expression in red.
4.3.3 Putative function of differentially expressed genes

We were able to assign GO classifications to 290 of the 347 significant differentially expressed genes between the invasive and native *A. amurensis* populations (see Supplementary Table S2.). Of the genes down-regulated in the invasive population, the most highly represented GO biological process terms were “proteolysis” and “transcription”, with 15 and 9 genes annotated in each, respectively. “ATP binding” (25) and “zinc ion binding” (18) were the most abundant molecular function terms. The most represented biological process terms in the invasive population up-regulated genes were “proteolysis” (7) and “protein transport: (5), whereas, the molecular function terms most represented were “calcium ion binding” (14) and “metal ion binding” (10). These molecular function terms are linked GO categories at different hierarchal levels within the ontology and as such, we have the same genes annotated with multiple terms. When we performed GO enrichment analysis for genes significantly different between the native and invasive population we found four terms were significantly over-represented (Table 3), with a FDR corrected $p$-value < 0.05. This over-represented gene set contains genes relating to positive regulation of cell adhesion, aromatase activity, heme binding and extracellular space.
Table 3. Gene ontology (GO) terms enriched in differentially expressed genes.

<table>
<thead>
<tr>
<th>GO term ID</th>
<th>Description</th>
<th># DE in term</th>
<th># in term</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0020037</td>
<td>heme binding (MF)</td>
<td>13</td>
<td>119</td>
<td>0.031</td>
</tr>
<tr>
<td>GO:0005615</td>
<td>extracellular space (CC)</td>
<td>19</td>
<td>224</td>
<td>0.031</td>
</tr>
<tr>
<td>GO:0045785</td>
<td>positive regulation of cell adhesion (BP)*</td>
<td>4</td>
<td>7</td>
<td>0.033</td>
</tr>
<tr>
<td>GO:0070330</td>
<td>aromatase activity (MF)</td>
<td>7</td>
<td>34</td>
<td>0.033</td>
</tr>
</tbody>
</table>

Benjamini-Hochberg corrected p-values. All terms contain genes that are both up and down-regulated in the invasive population (Bold font indicates whether more terms are over-represented in up- vs. down-regulated genes in the invasive population, *same number of genes up- and down-regulated), DE denotes differential expression, ontology in parentheses; BP, biological process; MF, molecular function; CC, cellular component.

4.3.4 Outlier loci analysis

We reliably identified 27,173 SNP sites in coding sequences for which there was genotype information across all individuals after our hard filtering on minor allele frequencies, SNP quality and conformity to Hardy-Weinberg equilibrium. 25,003 SNPs remained after filtering out transcripts with < 2 SNP sites. FST outlier analysis indicated 4,962 SNPs as significantly more divergent between our populations than likely from chance alone (at the 95% confidence interval). After further filtering of these SNP to only include transcripts with annotations and where all SNPs in a transcript exhibited the same putative selection signature, we retained SNPs across 157 genes. Lositan indicated 17 genes potentially exhibiting balancing selection (low FST relative to within-population variation) while 140 genes may be experiencing positive selection (high FST relative to within-population variation). Enrichment analysis of the GO terms for these genes revealed no significant over representation of any functional roles. We clustered the most represented GO terms with REVIGO to try and gain an insight into the functional processes the genes fulfil. Clustering revealed the presence of a large number of functional roles for the regulation of immune
system processes, phosphorylation and fundamental developmental processes (Fig. 4).
Figure 4. REVIGO scatterplot of the most represented GO categories for genes containing Fst outlier SNPs in the invasive population. GO categories are clustered by semantic similarity (simRel). Darker circles represent uncorrected enrichment values, whereas circle size denotes frequency and breadth of the GO categories.
4.4 Discussion

While considerable research has focused on the ecological and evolutionary factors contributing to successful biological invasions, investigations into the genomic basis of invasiveness and population-specific gene regulatory evolution are still in their infancy. Here, we identify candidate genes for invasion success based on gene expression differences between the previously identified, most likely *A. amurensis* native source population and a naturally expanding invasive population. Specifically, we identify 347 genes with significant differences in gene expression between the populations and 165 genes that exhibit evidence for the role of directional selection in the invasive population. In particular we identify genes with putative roles in immune function and response to environmental contaminants, which may underlie rapid adaptation of the invasive population in response to novel environmental conditions.

4.4.1 Functional roles of the most differentially expressed genes

Our comparison of differential expression between native and invasive populations allows us to identify the potential for gene expression levels to have evolved adaptively in response to local environmental conditions during the invasion process. We found substantial differences in gene expression between the two populations. Nine of the ten most significant differentially expressed genes are up-regulated in the invasive population, while one is down-regulated (discussed last). These genes are primarily involved in amino acid metabolism, immune response, oxidative metabolism and proteolysis. Cytochrome b-245 light chain (*CYBA/p22(phox)*) is a critical component of phagocyte activity and
its production of superoxide is instrumental in the killing of ingested microbes (Segal et al. 1992; Wientjes & Segal 1995; Ueno et al. 2005). A null allele of the gene increases susceptibility to infection and reduced inflammatory responses in mice (Pollock et al. 1995). Proteosome subunit beta type-5 (PSMB5) is involved in the formation of the proteasome, an essential part of immune surveillance mechanisms (Kloetzel 2001). Cytochrome p450 3A12 and 3A41 (CYP3A12, CYP3A41) are members of the monooxygenase system that is involved in the metabolism of xenobiotics and steroid production in echinoderms (den Besten 1998). Pyridoxine 5′-phosphate oxidase (PNPOx) catalyses the oxidation of Pyridoxal 5′-phosphate (PLP), which is an important cofactor for over 100 cellular enzymes (di Salvo et al. 2003); primary roles are in amino acid metabolism, steroid-receptor interaction and the regulation of immune function (Allgood et al. 1993; di Salvo et al. 2003). Pleckstrin homology domain containing family g member 4b (PLEKHG4B) has a potential role in the regulation of intercellular signalling, including signal transduction pathways (Shaw 1996). A homolog to Type II AFP (AFPII), an antifreeze glycoprotein that protects cellular processes from freezing damage (Liu et al. 2007) was also identified. LON peptidase N terminal domain and ring-finger protein 1 (LONPRF1) is a member of the LON peptidase family which regulate protein quality control and metabolism, gene expression and stress response in mitochondria (Nie et al. 2013; Strauss et al. 2015). The only gene significantly down-regulated in the top ten was Actin (ACT), which is the most abundant eukaryotic protein and plays a critical role in the cell, from the cytoskeleton to the regulation of transcription (Dominguez & Holmes 2011). At least half of these genes (all up-regulated in the invasive population) have a putative role in immune function, signalling or response and detoxification. This suggests that in
the invasive population, these genes may have an important role in dealing with exposure to novel environmental factors, both biotic and abiotic, not experienced by the native population. However, further experimental work will be needed to confirm the functional role of these genes in *A. amurensis*.

**4.4.2 Functional significance of cytochrome p450 expression**

We found four GO categories were significantly enriched (over-represented); two of which relate to the molecular function terms *heme-binding* and *aromatase activity*, comprising a total of 13 genes, nine of these genes being Cytochrome p450s. Cytochrome p450s (p450s) are one of the largest, versatile and most diverse protein superfamilies (Rewitz *et al.* 2006), with roles in metabolism of endogenous signal molecules (including sterols) and xenobiotics (see Werck-Reichhart & Feyereisen 2000; Rewitz *et al.* 2006). Previous work has revealed the up-regulation of p450s in a number of invasive species (Qiu *et al.* 2009; Mittapalli *et al.* 2010; Wang *et al.* 2011; Ioannidis *et al.* 2014; Ye *et al.* 2014), indicating they may be key contributors of invasion success in these species. Further, Hodgins *et al.* (2013) found differential expression of p450s between the native and invasive populations of the common ragweed (*Ambrosia artemisiifolia*), although different p450s were both up- and down-regulated across the populations. Similarly, we show the up-regulation of six and the down-regulation of three p450s in the invasive population. Having up- and down-regulated genes implicated in the same functional role seems initially contradictory. However, marine invertebrates and echinoderms in particular, exhibit a vast array of p450 gene families that are involved in different processes (see Snyder 2000; Rewitz *et al.* 2006). For example, *CYP1A* homologs
play a central role in the metabolism of Polychlorinated biphenyls (a xenobiotic) in the seastar *A. rubens* (Everaarts *et al.* 1994; Danis *et al.* 2004); moreover, induction of p450 activity in *A. rubens* shows specificity towards the type of contaminant (den Besten *et al.* 1991; den Besten 1998). However, p450s in marine invertebrates are also involved in steroid biosynthesis (Osada *et al.* 2004; Rewitz *et al.* 2006). Levels of p450 activity vary seasonally in *A. rubens*, with increased activity during the onset of the reproductive cycle, when an increase in steroid synthesis is also detected (Voogt *et al.* 1991; den Besten *et al.* 1993; den Besten 1998). Furthermore, there is a complex interplay between p450s, with the metabolism of xenobiotics resulting in a reduction in steroid hydroxylation rates (den Besten *et al.* 1991, 1993) and induction of specific p450s can suppress others (Guengerich 1988). Unfortunately, we are unable to distinguish the exact role of the expressed CYP genes here, as the whole protein family remains relatively uncharacterized in Asteroids, although *CYP3A* homologs in other marine invertebrates have been implicated in xenobiotic metabolism (Oberdörster *et al.* 1998; Snyder 2000) suggesting a similar functional role here. CYP expression in the invasive population could also be indicative of transgenerational expression effects. We are primarily interested in the difference between the native and invasive populations, as such; epigenetic control of expression is a potential important difference that could contribute to the comparative differences seen. Our data is unable to disentangle this but it is something which should be done in the future.

*Asterias amurensis* populations across the introduced range are associated with human degraded habitats, many of which experienced high levels of historical pollution (Morris 2002). Moreover, *A. amurensis* is a benthic predator more likely to be exposed to pollutants contained within the benthos
and its food sources. Our pattern of p450 expression shows similarity to that reported during exposure to xenobiotics in *A. rubens* and other invasive species. Combined with the invasive occurrence and life history, regulation of p450s enzyme activity may play a central role in the adaption of invasive *A. amurensis* to local environmental conditions, potentially impacting reproductive development. The next critical steps are, to confirm the function of the candidate genes presented here and uncover the relative importance of these mechanisms with regards to seasonal reproductive cycles during this successful invasion.

### 4.4.3 Differential selection

We detected a high number of SNPs putatively under differential selection in the invasive population, inferred through $F_{ST}$ outlier analysis. This type of analysis is prone to the reporting of false positives (Narum & Hess 2011; De Wit *et al.* 2014). As such, we adopted a conservative approach and only analysed SNPs that showed the same signature as all other SNPs present within a gene. We identified 157 genes most likely to be under selection and discuss the functional role of the most represented GO categories and the potential consequences to local adaptation during the invasion process.

Annotating genes with GO terms allows us to extract and characterize their functional roles. However there are also several challenges to this approach, particularly in non-model species. GO annotations are generally inferred through electronic annotation and sequence homology based searches. While the functional role of these inferred annotations is often correct and there is generally conserved function of orthologous sequences, shifts in functions
across species can occur (Gharib & Robinson-Rechavi 2011; Thomas et al. 2012). Furthermore, functional annotations in existing databases can be hindered by: imprecise or incorrect annotations, lags in annotating the databases or biased annotations due to a focus on model species (Khatri & Drăghici 2005; Thomas et al. 2012). Detailed functional analysis and annotation of identified genes remains a significant challenge and is the major hurdle in identifying the molecular basis of ecologically important traits (Hodgins et al. 2013). Taking these provisions into account, functional analysis with GO terms remains an important method for unravelling the potential functional roles of important genes under selection between the putative native source and expanding invasive population.

The most represented GO term was phosphorylation, which is involved in regulating the activity of enzymes through the addition of a phosphate group during a range of cellular processes (Cohen 1988, 2002), the maintenance of which may be fundamental during exposure to new environments. This suggests a potential adaptive role for these genes during the invasion of A. amurensis into Australia, primarily because these alleles are not present at high frequency within the putative native source population.

Adaptation to new environments can occur through two often-independent (but not always) modes – changes in alleles between generations, and changes in expression of loci during an individual’s lifetime. Genes involved in the regulation of immune system processes are highly represented in our GO clustering and suggested to be under positive selection. Yet, this gene is not differentially expressed between the populations, indicating its relative expression may be at ‘normal’ developmental levels in both the invasive and native populations. Gamma-glutamyltranspeptidase 1 (GGT, GGTP) contributes
to cysteine homeostasis, intracellular redox systems and inflammation (West et al. 2013). In humans, increased GGTP activity can be used as a biomarker for several diseases (West et al. 2013). GGTP activity has been previously identified in the developing larvae of the seastar *Pisaster ochraceous* (Sulakhe et al. 1990), where activity increases during the course of larval development. Protein phosphatase 1B (*PP1B*) is a major eukaryotic serine/threonine phosphate that regulates numerous cellular functions, including the antiviral response (Cohen 2002; Zhao et al. 2012). Increased *PP1B* expression inhibits a viral response in humans (Zhao et al. 2012), while *PP1* presence can inhibit oocyte maturation in the seastars *A. rubens* and *Marthasterias glacialis* (Meijer et al. 1986). Again, this gene is not differentially expressed between the populations, suggesting local adaptation of this gene's allelic variant in the invasive population. Selection may be acting to alter the function of these genes in response to either novel infectious agents or developmental challenges in the invasive population. Further work is needed to elucidate the functional role and significance of the SNPs to protein configuration before we can resolve the significance of these changes to the invasions success. However, given the relatively short timescale of the *A. amurensis* introduction into Australia (~30 generations), these data provide exciting preliminary evidence for rapid adaptation during the course of the invasion.

### 4.4.4 Variation in gene expression

Previously, we have shown that most neutral DNA sequence variation is partitioned within *A. amurensis* populations rather than among them, across both the native and introduced ranges (Chapter 2). Here we also found that a
substantial amount of gene expression variation is also within population variation, but the majority of this variation was due to population differences. Because we maintained a consistent environment and development time across the experiment, this pattern is likely to reflect true gene expression variation between our populations, suggesting substantial divergence of gene expression between the populations. Our experiment was designed to assess within population and native source versus expanding invasive population differences in gene expression. Thus in order to maximize biological replication we did not use cDNA library preparation or next-generation sequencing (NGS) technical replicates. However, we did have technical replicates during experimentation and RNA extraction, which were pooled prior to NGS steps, potentially minimizing the effect of technical variation on our interpretations. Further, diagnostic performance analysis from the addition of ERCC transcripts revealed no significant bias between the fractions of mRNA or large technical variation within conditions, suggesting good technical replication of library preparations and NGS.

Previous studies have revealed the relative importance of neutral genetic drift and directional selection to differences in gene expression among populations (Whitehead & Crawford 2006a, 2006b). Our study presents gene expression differences between the putative native source and expanding invasive population that may reflect the action of drift or divergent selection. While previous ecological work has provided evidence for rapid adaptation during invasion based on clinal variation in phenotypes across environments (Phillips et al. 2006), given the paucity of direct phenotypic information regarding *A. amurensis* larval traits across native and introduced environments, we were are unable to link phenotypic trait comparisons to observed gene
expression differences. However, our study has revealed many genes that might
underlie larval phenotypic differences between the native and introduced range,
especially in terms of physiological responses, and further investigation into
these is warranted. Introduced populations of *A. amurensis* exhibit less neutral
genetic variation than native populations (Chapter 2). Likewise, the distribution
of sample gene expression variation was smaller for the invasive population
compared to the putative native source. This pattern may indicate that founder
events and potential population bottlenecks are contributing to the observed
differences in gene expression among our populations. The small reduction but
relative consistency of gene expression variation within our populations could
reflect the demographic population history or represent a relative maintenance
of genetic variation in gene expression for specific genes, allowing for the action
of selection on traits important to local adaptation in the invasive population.
4.5 Conclusions

Our study revealed that many important genes with putative roles in immune function and response to environmental contaminants have constituent differences in expression between the native source and an expanding invasive population. We suggest a potential role of these genes in the rapid adaptation of the invasive population to novel environmental conditions. Similar responses have been seen in other invasive species, which suggests there may be common responses to invasion at the molecular level. Genes indicated as experiencing differential selective pressures are also implicated in immune function and developmental processes. This could allow for the selection of alleles in the invasive population that perform better in novel environmental conditions and may indicate the presence of rapid local adaptation during invasion of *A. amurensis* in Australia.
4.6 Supplementary Results

4.6.1 Assessment of technical and diagnostic performance

Diagnostic and technical performance was assessed by comparing the expression estimates of ERCC transcripts between both populations (see Methods for more details). We were reliably able to detect gene expression differences across a dynamic range of $2^{15}$ (Fig. 1a). This is less than the range of $2^{20}$ for which the ERCC ratios are designed to detect, meaning we are unable to reliably quantify ERCC expression at low abundances. This smaller range might be the result of the experimental sequencing depth after our quality filtering or the CPM based filtering which filters out genes expressed at low levels, including ERCC control sequences. Comparison between the ERCC ratios and endogenous gene expression values showed no significant bias in the mRNA fraction between samples ($\log (r_m) = 0.033 \pm 0.084$ S.E; Fig. 1c). Diagnostic performance is assessed through two measures, ROC curves (Fig. 1b.) and LODR estimates (Fig. 1d.). Our AUC of 1 represents a perfect score for diagnostic performance indicating our experiment has sufficient power to detect differentially expressed genes between our populations. Our ability to detect differential expression improves with an increase in the fold-change as demonstrated with lower LODR estimates for the 4 fold-change compared to the 2 fold (14 and 110 average read counts, respectively). The LODR estimates suggest poor performance when it comes to detecting a small 1.5 fold-change in expression. Together, these LODR estimates indicate we could confidently call differentially expressed genes when there is a 2 fold or greater change in expression and average read counts are above 80. As such,
Figure S1. ERCC technical and diagnostic plots produced by erccdashboard. Each sample type contained n=5 biological replicates. A) Signal-abundance plot, points are coloured by ratio sub-pool, shape represents the sample type; error bars denote the standard deviations of the replicates. B) ROC curves and AUC statistics for each group of true-positive ERCC controls (detected = number of controls used, spiked = the total included in the ERCC control mixture. C) MA plot of ERCC ratio measurement variability and bias. Coloured data points represent the mean ratio measurement per ERCC transcript, error bars the standard deviation of the replicates ratios, and filled circles are ERCC ratios above the LODR estimates. Grey points denote endogenous transcript gene expression measurements. Nominal ERCC ratios for each sub-pool are annotated with coloured solid lines, dashed lines represent the adjusted ratios based on the estimate of mRNA fraction differences between the samples, rm. D) LODR estimates are indicated by coloured arrows for each fold-change that crosses the threshold p-value, the black dashed line denotes the threshold p-value derived for the chosen FDR. LODR results and bootstrap confidence interval are provided in the table below the plot.
Figure S2. Multidimensional scaling plot of the variation of per-sample expression across the native and invasive range. Sample designation beginning with P from invasive range, those beginning with J from native range. Distances between samples correspond to leading-log-fold-change between all combinations of sample pairs. Leading-log-fold dimensions 1 and 2 are shown as these represent the most variation, leading-log-fold-change calculated as the average (mean-square-root) of the largest absolute log-fold-change between sample pairs.
4.7 References


Danis B, Goriely S, Dubois P et al. (2004) Contrasting effects of coplanar versus non-coplanar PCB congeners on immunomodulation and CYP1A levels (determined using an adapted ELISA method) in the common sea star Asterias rubens L. *Aquatic toxicology (Amsterdam, Netherlands), 69*, 371–83.


gonadal development of Japanese scallop, Patinopecten yessoensis: gonadal
profile and immunolocalization of P450 aromatase and estrogen.
Comparative biochemistry and physiology. Part B, Biochemistry & molecular
biology, 139, 123–8.

Asterias amurensis in Port Phillip Bay 1999-2003 (M and freshwater
systems report series, Ed.).


sequencing reveals complex mechanisms of diapause preparation in the
invasive mosquito, Aedes albopictus. Proceedings of the Royal Society B:
Biological Sciences, 280, 20130143.

Pollock JD, Williams DA, Gifford MA et al. (1995) Mouse model of X-linked
chronic granulomatous disease, an inherited defect in phagocyte

Molecular ecology, 22, 2366–8.

resistance in the cytochrome P450 genes of invasive and native Bemisia
tabaci (Hemiptera: Aleyrodidae) populations in China. Insect Science, 16,
237–245.

R Development Core Team R (2011) R: A Language and Environment for
Computing, 1, 409.

invertebrate cytochrome P450: emerging insights from vertebrate and
insects analogies. Comparative biochemistry and physiology. Toxicology &


Rollins L a, Moles AT, Lam S et al. (2013) High genetic diversity is not essential

Rollins L, Richardson MF, Shine R (in Press) A genetic perspective on rapid
evolution in cane toads (Rhinella marina). Molecular ecology.

Roman J, Palumbi SR (2004) A global invader at home: population structure of

Segal A, West I, Wientjes FB *et al.* (1992) Cytochrome b-245 is a flavocytochrome containing FAD and the NADPH-binding site of the microbicidal oxidase of phagocytes.


Young MD, Wakefield MJ, Smyth GK, Oshlack A (2014) goseq : Gene Ontology testing for RNA-seq datasets Reading data.

Chapter 5. Comparative RNA-Seq analysis of the temperature response in native and invasive *Asterias amurensis*

Mark F. Richardson and Craig DH. Sherman

5.1 Introduction

Invasive species can have devastating ecological and economic impacts (Sakai *et al.* 2001; Sax *et al.* 2007b), with marine ecosystems in particular exhibiting a high prevalence of successful invasions (Grosholz 2002; Reusch *et al.* 2010). Invasive species live in ecosystems where they have short evolutionary histories and experience novel environmental conditions (Prentis *et al.* 2008; Vandepitte *et al.* 2014), which may result in strong selection on morphological and physiological traits (Keller & Taylor 2008). This may involve rapid evolutionary responses (e.g. Shine 2012), adaptive phenotypic plasticity (e.g. Kolbe *et al.* 2012) (i.e. those individuals with a greater breadth of responses to an environmental change, Pigliucci 2005; Pigliucci *et al.* 2006; Lande 2015) or both. Several studies have shown that rapid evolutionary change in response to novel environmental conditions is common during the invasion process (Dlugosch & Parker 2008; Rollins *et al.* 2013; Rollins *et al.* in press). However, we still know little about how physiological responses and their underlying mechanisms facilitate environmental adaptation and invasion success.

The application of next-generation sequencing technology to these questions provides exciting opportunities to understand the mechanism of
physiological responses to different environments. For example, comparative whole-transcriptomic approaches can be used to characterize the genetic response (both expressed genes sequence variation and expression changes themselves) of a wide range of cellular processes underlying physiological capacity and response to common environmental conditions (Gracey 2007; Harms et al. 2014). These analyses also have the potential to uncover complex changes in gene regulatory responses and networks, which comprise rapid and versatile ways in which an organism can respond to environmental changes (Todgham & Hofmann 2009). Further, gene expression differences between environmental treatments may reveal the molecular mechanisms conferring physiological plasticity (Gracey 2007). Transcriptome analyses have been especially useful for studying molecular responses of non-model (marine) organisms to environmental stressors including; salinity fluctuations (Lockwood & Somero 2011), osmotic stress and ocean acidification (Todgham & Hofmann 2009; Pespeni et al. 2013a; Zhao et al. 2014) and thermal stress (Gracey et al. 2004; Stillman & Tagmount 2009; Franssen et al. 2011; Runcie et al. 2012; De Wit & Palumbi 2013; Harms et al. 2014). In particular, Lockwood et al. (2010) revealed a diverged gene expression response to thermal stress between two blue mussel species (Mytilus trossulus and M. galloprovincialis), which may help to explain the invasive status of one and not the other.

The northern Pacific seastar, Asterias amurensis, is an invasive benthic marine predator in southern Australia. Previous work suggests the most likely origin of this invasion is from a central Japanese population (Chapter 2; Ward & Andrew 1995). In native populations, reproduction can occur over a broad thermal range from 5-10°C to 23-25°C (Morris 2002; Kashenko 2005c), suggesting that we risk seeing this species spread its current reproductive range
into warmer Australian waters. *Asterias amurensis* possess a long planktonic larval stage, the length of which is also temperature-dependent, 66-91 days at 14°C and 37-44 days at 17°C (Goggin 1998; Kashenko 2005b). The larval stage is the primary mode of dispersal in this species and is responsible for the recent range expansion along the southern Australian coast (Chapter 2). The ability of invasive larval stages to tolerate elevated temperatures is therefore likely to be a key determinant of the success of any future range expansion in this species. The larval stage may be the most vulnerable life history stage, with larvae susceptible to a narrower range of environmental conditions than are adults (Morris 2002; Kashenko 2005a; 2005b). Hence, strong selection on the larval stage may lead to rapid changes in the genetic architecture or the expression of certain genes underlying larval physiological responses to elevated temperature conditions. Understanding the molecular mechanism of temperature-induced physiological changes will clarify the potential for this species to expand its invasive range.

Here, we used comparative gene expression analysis to evaluate the molecular response of *A. amurensis* planktonic larvae to elevated temperatures in native and invasive populations. Invasive populations are likely to have experienced selection during the establishment process (Chapter 4), potentially altering the range of larval physiological responses to temperature. So, identifying the genetic basis of physiological responses to elevated temperatures will be crucial for understanding the capacity for this species to expand its geographical range into warmer waters. By comparing the genetic response to elevated temperatures between invasive and native populations we can uncover whether the genetic response to increased temperature is conserved between the populations or whether selection has altered the genetic
architecture of invasive larvae. Here, we describe and interpret transcriptomic responses to elevated temperatures in native and invasive *A. amurensis*. 
5.2 Methods

5.2.1 Larval temperature experiment and sample collection

Because the present study focuses on mechanisms involved in physiological responses to elevated temperature, we selected temperatures based upon those potentially experienced across the native distribution of this species and those temperatures that dispersing larvae are likely to face during the expansion of the invasive range. Spawning occurs during the winter months in both ranges, native: Jan-Mar 14-17°C; invasive: Jun-Sept, 10-14°C (Goggin 1998; Kashenko 2005c). We selected the common spawning temperature of 14°C as the control and two elevated temperatures (intermediate, 17°C and high, 20°C) to represent reproductive thermal maxima.

We collected Asterias amurensis adults from Williamstown, Victoria, Australia in July 2013 (6 males, 6 females) and Hakkeijima Island, Yokohama, Japan (6 males, 6 females) in March 2013. The spawning protocol followed an extended design based on Chapter 4. Experiments were conducted within 24 h of adult capture. Briefly, adult A. amurensis were individually rinsed with UV-treated 1 μm filtered seawater to remove potential gamete cross-contamination. A solution of 1 ml 10⁻⁵ M 1-methyladenine, in filtered seawater, was injected into the coelom of each individual to induce spawning. Males and females were spawned dry in separate containers and gametes rinsed and re-suspended in 50 ml filtered 14°C seawater. The concentration of sperm for each male was standardized to 1 × 10⁶ sperm ml⁻¹ and female egg concentrations were standardized to 1 x 10⁴ eggs ml⁻¹. Artificial fertilizations were carried out in 100 ml filtered seawater at each of three temperature treatments (14°C, 17°C and 20°C), using 10,000 eggs from a single female and 100,000 sperm from a single male (sperm:egg ratio of 10:1). For both native and invasive populations, we
collected gametes from six males (M) and six females (F). Fertilizations of these gametes were conducted across the three temperatures, for each of the six (M x F) crosses where each individual was only used in one cross. By using the same six M x F crosses across temperature treatments, we could look at changes in gene expression across the same set of full siblings (i.e. we kept the same genetic background) and partition variation in gene expression attributable to temperature, between populations and differences between individual crosses. Gametes were left for 2 hours to fertilize at 14°C, 17°C and 20 °C. Embryos were then transferred into 1L containers (~ density of 5 larvae per ml) and cultured at each of the three temperature treatments for 10 days. These consisted of two replicates for each cross at each temperature, totalling 36 samples per population. Larvae were cultured at constant common garden conditions in both Japan and Australia. The difference in experimental month represents the different winter spawning period between the northern and southern hemispheres only. Larvae were cultured in UV-treated 1 μm filtered seawater, maintained at constant experimental temperatures (14°C, 17°C and 20 °C), 12:12h light:dark cycle and fed a standard algal diet of cultured *Chetocherus muleri* culture at 50,000 cells ml⁻¹. While it is inevitable we will have sequenced some algal RNA, the reference transcriptome only contains *A. amurensis* genes, so its presence will not alter expression estimates (assuming there is not strong homology between *A. amurensis* and *C. muleri* genes, which is unlikely).

### 5.2.2 RNA isolation

Cultured larvae were removed at the mid-bipinnaria larval stage (10 days post fertilization) to avoid a possible overrepresentation of early-expressed
developmental genes in the total RNA sample. Further, as temperature has a
direct effect on larval development (Chapter 1) we chose the middle of the
bipinnaria stage as: larvae feed during this developmental stage (by day ~3-4)
(Kashenko, 2005a), are all at the same developmental stage (in all temperature
treatments by day 4) and we did not want any possible change in larval stage
induced by temperature treatments (larvae reach the next developmental stage,
brachiolaria, at ~26-28 days at 17°C) (Kashenko, 2005a) to confound gene
expression differences between treatments. The choice of this development
stage and timing of larval gene expression measurement allows direct
comparison between treatments at a larval stage potentially under selective
pressure. RNA isolation followed methods described in Chapter 4. Briefly, larvae
aliquots (approximately 2,000 individuals) from each sample (including
technical replicates) were transferred to separate 1.5ml tubes, spun to a pellet
and immediately stored in RNALater (Qiagen, USA). Samples were incubated at
4°C for 24hrs and then transferred to a -80°C freezer for storage. RNeasy spin
columns (Qiagen, USA) with a Qiashredder (Qiagen, USA) pre-extraction were
used to extract total RNA according to the manufacturer’s protocol. Quality and
quantity of total RNA was assessed for each sample using an Agilent 2100
Bioanalyser (Agilent Technologies, USA) to ensure that >1 ug of total RNA was
extracted (range 1.13 – 1.93 ug) and samples had high RNA Integrity Numbers
(RIN values: range 7.2 – 9.8). RNA from the two technical replicates per sample
was combined, providing a technical replicate of each of the RNA extraction
protocols.
5.2.3 RNA sequencing and quality control

We added 2 ul of a 1:100 dilution of either mix 1 or mix 2 of the External RNA Controls Consortium (ERCC) spike-in control RNA (Invitrogen/Life Technologies, USA) to 1 ug of total RNA from each sample (see Appendix 2). A discussion on the benefit/use of these controls is provided in Chapter 4.

Sequencing and cDNA library preparation was conducted commercially at Macrogen, South Korea. We used the previously described sequencing data of 14°C ‘ambient’ reared larvae from Chapter 4 as the control treatment data. For the 17°C and 20°C treatment, sequencing methods were identical and were conducted as a continuation of the same sequencing run. Details of cDNA library construction and size selection step (~ 280 bp fragments) are provided in Chapter 4. Amplified cDNA libraries for the 17°C and 20°C treatments were sequenced across a further four flow cell lanes of the Illumina HiSeq-2500 platform, generating 100bp paired-end reads (six flow cells lanes for all treatments). Samples for both populations were randomly assigned to each flow cell lane to minimize technical bias during sequencing. Quality control steps followed those used in Chapter 4 and were performed with cutadapt v1.3 (Martin 2011) and FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Briefly, this involved removal of adaptor sequences, filtering based on quality scores (95% of bases in a read with min Q > 30) and removal of a GC content bias in the first 13 bases the RNA-Seq reads.

5.2.4 Read Mapping and expression quantification

We used the reference transcriptome generated in Chapter 3, which contained only the longest unique versions of genes identified during the de novo assembly
(Chapter 4) for read alignment. This provided the best available annotated reference set for gene expression quantification in this species. Reads from each sample were mapped to the reference transcriptome individually using RSEM v1.2.14 (Li & Dewey 2011) and Bowtie2 v2.1.0 (Langmead & Salzberg 2012), using the default parameters. This allowed us to quantify expression on a per-gene basis as the number of raw counts. We only considered paired-end reads when generating read counts, as previously described in Chapter 4. Technical performance of the experiment was analysed using erccdashboard (Munro et al. 2014) and the expression values of the individual ERCC transcript ratios. Brief methods on implementation for this experiment and the parameters used are provided in the Supplementary methods.

5.2.5 Digital differential gene expression

Differentially expressed genes were identified with the BioConductor tool edgeR (Chen et al. 2014), in R (R Development Core Team 2011). Before testing for differential expression we performed a multidimensional scaling analysis (MDS) to illustrate the contribution of population and experimental temperature to variation in gene expression and search for outlier replicates. Given the paired nature of the experiment and that variation in gene expression was largely attributable to individuals and population of origin (see results), we performed differential expression analysis in the populations separately and blocked by samples baseline expression (expression at 14°C) before comparing the response to the elevated temperature treatments. We adopted the Generalized Linear Model (GLM) approach (Chen et al. 2014) for the comparison of the three temperature treatments (14°C, 17°C and 20°C) across the native and invasive populations. RSEM quantified expression counts were filtered for lowly
expressed genes. Genes were removed when they had < 1 count per million reads (CPM) in \( n = 4 \) or fewer libraries (where \( n \) is the smallest number of replicates per temperature treatment within a population). These genes are removed as they have too low an expression to be detected as significant and their inclusion results in a reduction of power to detect differential expressed genes (Anders et al. 2013). CPM filtered counts were TMM-normalized (Trimmed mean on M-values) (Robinson & Oshlack 2010), to account for library size between replicates. We calculated the dispersions for the data then fitted the additive GLM that adjusted for the baseline gene expression differences. Differential gene expression was tested between all combinations of the treatments per population using likelihood ratio tests with a 5% false discovery rate (FDR).

We used scripts from the Trinity project (Hansen et al. 2012a) for filtering of differentially expressed genes based on ERCC derived \( p \)-values, minimum reliably detectible fold-change and for building heatmaps. Based on the ERCC performance metrics, differentially expressed genes from the invasive population were kept when the minimum absolute \( \log_2 \) fold-change was >1 (which represents a two fold-change) and their \( p \)-value <0.01. Likewise, based on ERCC measures, we only considered differentially expressed genes in the native population when the minimum \( \log_2 \) fold-change was > 2 (a four fold-change) and \( p \)-values <0.05. These values are higher than those for the invasive population as the smaller number of differentially expressed genes within the native samples contributes to reduced power to detect low levels of differential expression. This provides the most conservative cut off for evaluating significant differential gene expression without removing all differential
expression. The values for this filtering are based upon the intrinsic gene expression differences between the regions and are supported by the addition of true and false positives for the expression ranges provided by the ERCC controls, as such, they will likely keep true differences in gene expression between the treatments.

5.2.6 Gene ontology enrichment analysis

Enrichment analysis allows us to identify whether functional categories associated with an a priori suite of genes contain more differentially expressed genes for each population than expected by chance. We used goseq (Young et al. 2014), which takes length bias into account, to test for enrichment (over-representation) of GO terms between the temperature treatments. We extracted the Gene Ontology categories (Biological Process, Molecular Function and Cellular Component) for the genes used in the reference transcriptome from the de novo assembly Annotation Table (Chapter 3). For each population the set of GO terms for differentially expressed genes between the temperature treatments were assessed for enrichment compared to the reference set of GO categories. Enrichment was determined by a Benjamini-Hochberg corrected p-value < 0.05.
5.3 Results

5.3.1 Larval temperature experiment, sequencing and read mapping

Sequencing of cDNA libraries yielded between 57.5 and 105.4 million reads per sample (Supplementary Results; Table S1). After adapter trimming and quality filtering we obtained between 87,467,780 and 26,357,854 million reads per sample that still contained both paired reads (Table S1.). Between 72.78 % and 63.64 % of reads were properly mapped (i.e. when both pairs map to the same reference gene) to the 18,411 reference genes and 92 ERCC transcripts. RNA extracted from replicate (P3) for control treatments and another (P1) at 17°C contained a large proportion of adaptor sequences and grouped as outliers in the MDS analysis. We removed P1 at 17°C and all P3 samples from the analysis as we could not establish a P3 baseline level of gene expression and did not want to introduce a potential bias. Assessment of experimental performance and diagnostic power revealed we have high power to detect real endogenous differential gene expression, at the two- and four-fold levels, in both the native and invasive populations (Supplementary Results, Fig. S1; Fig. S2). Based on these results we chose to filter differentially expressed genes, including only those with a greater than two fold-change in expression and \( p \)-value < 0.05. These criterions account for the different ability to reliably call differentially expressed gene between the populations and limits the analysis to those genes most significantly differentially expressed between the temperature treatments.

We were unable to reliably or consistently rear larvae to day 10 in the native population for the 20°C temperature treatment, with only two population replicates running to full term (J120 and J620), despite all population replicates surviving at the control (14°C) and 17°C temperature treatments. Gene expression measurements for the native 20°C treatment come from these two
surviving replicates (J120, J620). None of the failing population replicates exhibited high bacterial load in either of their technical replicates, both of which failed.

### 5.3.2 Gene expression response to elevated temperature treatments

Both the native and invasive populations showed discernable shifts in gene expression in response to the temperature treatments. The MDS illustrated that samples grouped according to their population of origin in terms of gene expression variation, regardless of temperature treatment (Biological coefficient of variation, BCV 0.26; Fig. S3). Within populations, temperature treatments elicited a varied response in gene expression. However samples largely grouped close together even across the different temperature treatments, apart from in the native population, which exhibited greater variance across temperatures than that observed in the invasive population (native BCV 0.30; invasive BCV 0.20).

<table>
<thead>
<tr>
<th>Population</th>
<th>Treatment</th>
<th># up-regulated</th>
<th># down-regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive</td>
<td>17°C</td>
<td>215</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>20°C</td>
<td>85</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>20°C compared to 17°C</td>
<td>82</td>
<td>186</td>
</tr>
<tr>
<td>Native</td>
<td>17°C</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>20°C</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>20°C compared to 17°C</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Both populations exhibited similar signatures of differential expression, however, the magnitude of the response was greater in the invasive population compared with the native population (Fig. 1; Fig. 2). In the invasive population,
483 unique genes showed differential expression across the elevated temperature treatments (Fig. 1). At 17°C, 215 genes were up- and 50 down-regulated compared to the control (Table 1.). While at 20°C, 85 were up- and 45 were down-regulated compared to the control. Across both of these elevated temperatures, 64 genes were commonly differentially expressed (54 up-regulated in both, 10 down-regulated in both). Comparison between the 17°C and 20°C treatments revealed 82 up- and 186 down-regulated genes, of which one was also differentially expressed between the elevated temperatures and the control. In the native population, 42 unique genes were differently expressed between the temperature treatments, encompassing 19 up- and 8 down-regulated between 17°C and the control, and 10 up- and 7 down-regulated between 20°C and the control (Table 1.; Fig. 2). Comparison across both elevated temperatures revealed the common differential expression of only 3 genes relative to the control (2 up- and 1 down-regulated, Table 2.). Between the 17°C and 20°C elevated temperature treatments 3 genes were up- and 1 was down-regulated, none of which showed differential expression in the elevated temperatures compared to the control. Of the genes showing common differential expression to both the elevated temperatures compared to the control, only 1 gene, *Ubox domain containing protein*, exhibited differential expression across both populations, however the pattern was not consistent across populations and treatments (i.e. both up- and down-regulated).
Figure 1. Heatmap of differentially expressed genes between the temperature treatments (14, 17 and 20°C) in the invasive population. Genes are clustered by similarity in mean centred gene expression measurements. Expression values are coloured by magnitude, higher expression in red.
5.3.3 Shared trends in gene expression between the native and invasive population

Only four genes showed the same response to the temperature treatments in both the native and invasive populations (Table 2.). Three genes were up-regulated at the 17°C treatment and one was up-regulated at the 20°C treatment compared to the control in both populations. We found no evidence of shared down-regulation of genes for either of these temperature treatments across populations. There were no differentially expressed genes for the 17°C to 20°C comparison that were consistent across native and invasive populations. The GO
terms associated with these commonly expressed genes reveal roles in lignin
catabolic processes (GO:0046274), ATP binding (GO:0005524) and apoptotic
processes (GO:0006915).

Table 2. Genes showing the same expression response to the temperature treatments
for both the native and invasive populations.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Associated GO terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptor-interacting serine-threonine kinase 2-like</td>
<td>GO:0005524, GO:0004672</td>
</tr>
<tr>
<td>Serine/threonine-protein kinase</td>
<td>GO:0005737, GO:0005634, GO:0005524,</td>
</tr>
<tr>
<td></td>
<td>GO:0030145, GO:0004674, GO:0043024,</td>
</tr>
<tr>
<td></td>
<td>GO:0006915, GO:0007049, GO:0008283,</td>
</tr>
<tr>
<td></td>
<td>GO:0043066, GO:0043433, GO:0031659,</td>
</tr>
<tr>
<td></td>
<td>GO:0046777</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>na</td>
</tr>
<tr>
<td>Laccase 1 isoform G</td>
<td>GO:0005576, GO:0005507, GO:0052716,</td>
</tr>
<tr>
<td></td>
<td>GO:0046274</td>
</tr>
</tbody>
</table>

5.3.4 Population-specific responses to temperature treatment

Of the 525 unique genes differentially expressed across both native and invasive
populations and across temperature treatments, 521 showed population
specific responses to the elevated temperature treatments. In the invasive
population (483 genes differentially expressed), the largest changes in
expression for the 17°C treatment were attributable to genes involved in
protein complex formation (GO:0043234), DNA-dependent transcription
(GO:0006351), and developmental growth (GO:0048589) (Table 3.).

Enrichment analysis revealed key functional GO terms were over-represented in
this treatment (Table 4.) with several of the genes liked to roles in cellular
defence and stress responses. The elevated 20°C treatment showed a GO term
enrichment for genes associated with NAD+ ADP-ribosyltransferase activity, as
also seen in the 17°C treatment (Table 4.). The largest changes in expression between the 20°C treatment and the control resulted from a down-regulation of genes involved in cell-cell signalling (GO:0007267) and Rab guanyl-nucleotide exchange factor activity (GO:0017112), and the up-regulation of genes involved in cellular iron ion homeostasis (GO:0006879) and gene silencing by RNA (GO:0031047)(Table 3.). Comparison of gene expression values for the 17°C to the 20°C treatment showed the most significant change in expression was also from the down-regulation of genes involved in Rab guanyl-nucleotide exchange factor activity (GO:0017112) and sequence-specific DNA binding transcription factor activity (GO:0003700), and the up-regulation of genes involved in developmental growth (GO:0048589) and stress responses. The GO enrichment analysis for this comparison revealed the over-representation of eight GO terms not identified in the previous comparisons of elevated temperatures to the control (Table 4).

Within the native population (42 unique genes differentially expressed), the largest changes in gene expression for the 17°C treatment was for genes involved in: DNA-dependent DNA replication (GO:0006261), glycoside catabolic process (GO:0016139), nutrient reservoir activity (GO:0045735) and translation (GO:0006412)(Table 3.). The elevated 20°C treatment when compared to the control revealed the greatest change in expression was for up-regulated genes with functional roles in DNA-dependent DNA replication (GO:0006261), protein neddylation (GO:0045116) and proteolysis (GO:0006508), and the down-regulation of genes involved in several development processes including: gonad development (GO:0008406) and collagen fibril organization (GO:0030199)(Table 3.). Comparison of gene expression values for the 17°C to the 20°C treatment showed the most
significant change in expression was for the up-regulation of genes with functional roles in proteolysis (GO:0006508) and protein neddylation (GO:0045116), and the down-regulation of a gene involved in DNA-dependent transcription (GO:0006351) (Table 3.). Of all the temperature treatment comparisons for the native population only one GO term, alanine-glyoxylate transaminase activity, was significantly over-represented for genes differentially expressed between the 20°C treatment and the control (Table 4.).

Of the genes showing common differential expression between the two elevated temperature treatments, several show population specific responses. For genes differentially expressed at the 17°C treatment, a Ubox domain containing protein is up-regulated in the invasive population, while down-regulated in the native population. Conversely, leucyl-tRNA synthetase is down-regulated in the invasive and up-regulated in the native population. For the 20°C treatment, when compared to the control, the same Ubox domain containing protein is up-regulated in the invasive and down-regulated in the native population.
Table 3. Genes exhibiting the largest change in expression between the temperature treatments. Gene name, log fold-change and differential expression p-value are given.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Log FC</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invasive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>17°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein kintoun-like</td>
<td>7.28</td>
<td>9.61E-13</td>
</tr>
<tr>
<td>Stress-induced-phosphoprotein 1</td>
<td>6.38</td>
<td>4.21E-05</td>
</tr>
<tr>
<td>myosin heavy chain</td>
<td>-6.80</td>
<td>6.23E-06</td>
</tr>
<tr>
<td>Stimulated by retinoic acid gene 6 protein homolog</td>
<td>-6.31</td>
<td>5.95E-05</td>
</tr>
<tr>
<td>DNA (cytosine-5)-methyltransferase 1</td>
<td>-6.09</td>
<td>1.81E-04</td>
</tr>
<tr>
<td><strong>20°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>protein kintoun-like</td>
<td>7.26</td>
<td>1.16E-13</td>
</tr>
<tr>
<td>Melanotransferrin</td>
<td>6.59</td>
<td>1.33E-05</td>
</tr>
<tr>
<td>CCR4-NOT transcription complex subunit 11</td>
<td>5.54</td>
<td>1.47E-04</td>
</tr>
<tr>
<td>Putative uncharacterized protein</td>
<td>-6.74</td>
<td>1.17E-05</td>
</tr>
<tr>
<td>DENN domain-containing protein 2C</td>
<td>-6.54</td>
<td>2.86E-06</td>
</tr>
<tr>
<td><strong>20°C compared to 17°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulated by retinoic acid gene 6 protein homolog</td>
<td>7.14</td>
<td>8.36E-06</td>
</tr>
<tr>
<td>arginine N-methyltransferase</td>
<td>5.92</td>
<td>1.64E-04</td>
</tr>
<tr>
<td>myosin heavy chain</td>
<td>5.36</td>
<td>6.88E-04</td>
</tr>
<tr>
<td>DNA-binding protein D-ETS-4</td>
<td>-5.87</td>
<td>1.11E-03</td>
</tr>
<tr>
<td>DENN domain-containing protein 2C</td>
<td>-5.78</td>
<td>1.11E-04</td>
</tr>
<tr>
<td><strong>Native</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>17°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative uncharacterized protein</td>
<td>5.03</td>
<td>4.26E-05</td>
</tr>
<tr>
<td>DNA topoisomerase 2</td>
<td>4.13</td>
<td>8.91E-05</td>
</tr>
<tr>
<td>Alpha-galactosidase A</td>
<td>-3.66</td>
<td>5.43E-05</td>
</tr>
<tr>
<td>Vitellogenin-1</td>
<td>-3.03</td>
<td>1.77E-06</td>
</tr>
<tr>
<td>60S ribosomal protein L17</td>
<td>-3.00</td>
<td>1.62E-05</td>
</tr>
<tr>
<td><strong>20°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA topoisomerase 2</td>
<td>10.43</td>
<td>4.16E-08</td>
</tr>
<tr>
<td>Acylamino-acid-releasing enzyme</td>
<td>8.93</td>
<td>1.74E-09</td>
</tr>
<tr>
<td>NEDD8-activating enzyme E1 catalytic subunit</td>
<td>7.57</td>
<td>1.89E-07</td>
</tr>
<tr>
<td>sal-like protein 3-like</td>
<td>-10.03</td>
<td>6.90E-06</td>
</tr>
<tr>
<td>Homeobox protein Mohawk</td>
<td>-8.58</td>
<td>1.57E-08</td>
</tr>
<tr>
<td><strong>20°C compared to 17°C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uncharacterized protein</td>
<td>7.73</td>
<td>3.28E-06</td>
</tr>
<tr>
<td>Acylamino-acid-releasing enzyme</td>
<td>7.47</td>
<td>3.73E-07</td>
</tr>
<tr>
<td>NEDD8-activating enzyme E1 catalytic subunit</td>
<td>6.65</td>
<td>4.75E-06</td>
</tr>
<tr>
<td>Transcription factor AP-1</td>
<td>-2.04</td>
<td>1.14E-05</td>
</tr>
</tbody>
</table>
Table 4. Gene ontology (GO) terms enriched in differentially expressed genes between the temperature treatments.

<table>
<thead>
<tr>
<th>GO term</th>
<th>Description</th>
<th># DE in term</th>
<th># in term</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invasive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0045087</td>
<td>nonspecific immune response (BP)+</td>
<td>14</td>
<td>132</td>
<td>0.002</td>
</tr>
<tr>
<td>GO:0003950</td>
<td>NAD+ ADP-ribosyltransferase activity (MF)+</td>
<td>7</td>
<td>26</td>
<td>0.002</td>
</tr>
<tr>
<td>GO:0051607</td>
<td>antiviral response (BP)+</td>
<td>8</td>
<td>49</td>
<td>0.011</td>
</tr>
<tr>
<td>20°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0003950</td>
<td>NAD+ ADP-ribosyltransferase activity (MF)+</td>
<td>5</td>
<td>26</td>
<td>0.028</td>
</tr>
<tr>
<td><strong>17°C to 20°C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0016491</td>
<td>oxidoreductase activity (MF)*</td>
<td>12</td>
<td>122</td>
<td>0.015</td>
</tr>
<tr>
<td>GO:0034081</td>
<td>polyketide synthase complex (CC)-</td>
<td>3</td>
<td>3</td>
<td>0.015</td>
</tr>
<tr>
<td>GO:0071770</td>
<td>DIM/DIP cell wall layer assembly (BP)-</td>
<td>3</td>
<td>3</td>
<td>0.015</td>
</tr>
<tr>
<td>GO:0097040</td>
<td>phthiocerol biosynthetic process (BP)-</td>
<td>3</td>
<td>3</td>
<td>0.015</td>
</tr>
<tr>
<td>GO:0097041</td>
<td>phenolic phthiocerol biosynthetic process (BP)-</td>
<td>3</td>
<td>3</td>
<td>0.015</td>
</tr>
<tr>
<td>GO:0031177</td>
<td>phosphopantetheine binding (MF)-</td>
<td>4</td>
<td>9</td>
<td>0.024</td>
</tr>
<tr>
<td>GO:004499</td>
<td>N,N-dimethylaniline monooxygenase activity (MF)-</td>
<td>4</td>
<td>9</td>
<td>0.024</td>
</tr>
<tr>
<td>GO:004315</td>
<td>3-oxoacyl-synthase activity (MF)-</td>
<td>3</td>
<td>4</td>
<td>0.038</td>
</tr>
<tr>
<td><strong>Native</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0008453</td>
<td>Alanine-glyoxylate transamine activity (MF)-</td>
<td>2</td>
<td>4</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Benjamini-Hochberg corrected p-values. + only contains genes up-regulated – only contains down-regulated genes, * both up and down-regulated genes contained, DE denotes differential expression, ontology in parentheses; BP, biological process; MF, molecular function; CC, cellular component.

5.4 Discussion

The ability of invasive species to establish and become widespread within an introduced range will to a large extent be determined by: i) the level of phenotypic plasticity in traits that allow them to acclimate to novel environmental conditions experienced in the invasive range, ii) selection acting on standing genetic variation that may alter baseline gene expression and regulatory pathways and potentially, iii) selection for wider plastic responses to environmental conditions. As such, there is potential for evolved gene regulatory differences between the native and invasive populations as well as
differences in the level of individual plasticity to elevated temperature. Ultimately, an organism’s ability to regulate cellular processes and physiological responses when encountering different environmental conditions will allow it to maintain essential cellular functions, avoiding cellular damage and cell death. Here, we compared the transcriptomic response of native and invasive *A. amurensis* larvae to two elevated temperatures. Uncovering the gene expression response underlying acclimation to elevated temperatures is crucial for understanding the capacity of this species to expand its invasive range into warmer waters. We find large differences in the transcriptomic response between elevated temperature treatments and populations. Our results suggest that disparities between the populations’ physiological responses may underlie differences in the ability of invasive and native *A. amurensis* to cope with elevated temperature stress.

### 5.4.1 The transcriptional response of invasive *A. amurensis* to elevated temperature

Our data demonstrate that, in response to the two elevated temperature treatments, the invasive *A. amurensis* population exhibits a much greater response in terms of differential gene expression (483 differential genes) than that exhibited by the native population (42 differential genes), with only four genes showing common expression patterns. This may suggest a signature of either the evolution of increased plasticity in the invasive population in response to elevated temperature, or local adaptation from standing genetic variation. Within each population, we see similarities in the response to both the elevated temperatures for genes involved in the maintenance of cellular
functions and homeostasis, although there remain substantial differences between the elevated temperatures.

In the invasive population, we show an over-representation of genes involved in NAD+ ADP-ribosyltransferase activity, which is an important mechanism in the posttranslational modification of proteins regulating many cellular processes (Frye 1999). Specifically, Poly [ADP-ribose] polymerase 14, 12 and a TCDD-inducible poly [ADP-ribose] polymerase are up-regulated at both 17°C and 20°C. Poly [ADP-ribose] polymerases (PARPs) are DNA nick sensors that regulate DNA repair, cell death, chromatin functions and genomic stability (Herceg & Wang 2001; Strosznajder et al. 2005); activation of PARPs is one of the early DNA damage responses to environmental and endogenous genotoxic agents (de Murcia & de Murcia 1994). Similarly, a Kintoun-like protein (Ktu) is up-regulated at both elevated temperatures. Ktu is the first cytoplasmic component of dynein assembly and is essential for the formation of motile cilia (Roy 2009; Kobayashi & Takeda 2012). A lack of Ktu causes immotile cilia and has been implicated in several human diseases (Roy 2009; Kobayashi & Takeda 2012). Seastar bipinnaria larvae have two loops of motile ciliated bands and associated neurons enabling larvae to feed and swim in sensory response to environmental cues (Chee & Byrne 1999; Strathmann & Grünbaum 2006; Elia et al. 2009; Yankura et al. 2013). These common changes in gene expression between the elevated temperatures could suggest invasive A. amurensis possess a molecular mechanism to protect against DNA damage and maintain essential organismal functions. As this response is not seen in the native population it could suggest that the elevated temperature elicit a damage control mechanism to maintain cellular function in the invasive population, or
alternatively, that invasive *A. amurensis* may have evolved greater resilience to elevated temperatures.

In the invasive population, several changes in gene expression between the two elevated temperatures indicate that the magnitude of the physiological response evoked may be different. At 17°C we see an up-regulation of stress-induced-phosphoprotein 1 (*STIP1*), a cochaperone that organizes other molecular chaperones such as heat shock proteins (*HSPs*) (Tsai *et al.* 2012), which is not up-regulated at 20°C. Further, *HSP70*, which is potentially associated with *STIP1*, shows the same pattern. Elevated *HSP70* expression has previously been associated with the temperature stress response of several marine species (Lang *et al.* 2009; Dong & Dong 2008), including an invasive blue mussel, *M. galloprovinciallis* (Lockwood *et al.* 2010) and the purple sea urchin, *Strongylocentrotus purpuratus* (Runcie *et al.* 2012). This is an interesting but confusing result; the up-regulation of only one *HSP* indicates that the 17°C treatment does not elicit a response requiring widespread *HSP* cellular protection. Given this minimal response at 17°C we would expect to see an increase in *HSP* expression at 20°C due to elevated temperature stress. The 20°C treatment may elicit a pressure high enough to surpass the threshold for *HSP* expression in invasive *A. amurensis*. Or, invasive *A. amurensis* developing at 20°C may be more robust than at 17°C due to the expression of genes not induced at 17°C. Further work will be needed to define the boundaries of *HSP* expression in *A. amurensis*.

We also find genes that may have non-specific immune and cellular regulation roles up-regulated at 17°C and not at 20°C, including E3 ubiquitin protein ligase and Serine/threonine-protein kinases. Serine/threonine-protein kinases are involved in cell proliferation and cell death (apoptosis) and could be
involved in the regulation of apoptosis due to cellular stress (Cross et al. 2000). Ubiquitin ligases and degradation genes have also been identified in the temperature stress acclimation response of blue mussels (Lockwood et al. 2010) and fish (Logan & Somero 2011). Together, these genes, the presence of DNA repair machinery (PARPs), molecular cochaperones (STIP1) and chaperones (HSP70) at 17°C may indicate activation of the cellular stress response (CSR) to temperature. However the minimal CSR, conserved in eukaryotes, implicates activation of several genes we do not see (see Kültz 2005 for review). The cnidarian *Anthopleura elegantissima* when exposed to heat stress also did not show widespread activation of genes previously identified in this minimal stress response (Richier et al. 2008), indicating the minimal CSR is likely to be induced at varying thresholds across taxa.

At 20°C compared to 17°C we find an over-representation of genes involved in oxidoreductase activity (GO:0016491) among others (Table 4.). Similarly, genes involved in oxidoreductase activity are also up-regulated during *S. purpuratus* development in response to elevated temperatures (Runcie et al. 2012). Several of the genes belonging to this GO term also belong to others in the enriched list. Ten of these genes were identified as polyketide synthase (*pks*) like genes and are down-regulated at 20°C. *Pks* are an essential component of calcium carbonate biomineralization in echinoderms (Hojo et al. 2015) and therefore essential to further development of larvae into juvenile seastars with calcareous skeletons. Pespeni et al. (2013b) revealed potentially adaptive differences in the regulation of biomineralization genes across the range of *S. purpuratus* populations that span different thermal conditions. However, the developmental role of this gene expression still needs to be examined in *A. amurensis* bipinnaria larvae.
Genes involved in posttranslational modifications are up-regulated at 20°C and not at 17°C. Arginine N-methyltransferase catalyses the methylation of arginine residues and is implicated in signal transduction, RNA transport, mRNA splicing and transcription (Frankel et al. 2002). The CCR4-NOT transcription complex is involved in the deadenylation of mRNAs, a fundamental aspect of gene regulation and the modification of mRNA function (Tucker et al. 2001; Fabian et al. 2011; Bhandari et al. 2014). The down-regulation of genes involved in future developmental growth and up-regulation of genes involved in modifying gene function, transcription and potential splicing could indicate that 20°C generates substantial pressure for ongoing organismal survival. The 20°C treatment may be eliciting a greater stress response revealing gene expression to maintain cellular processes and homeostasis not seen at 17°C. Given invasive larvae were able to develop and survive at 20°C, the underlying differences in expression between temperatures may explain an enhanced physiological response to this higher temperature resulting from increased plasticity or local adaptation. This could provide a molecular mechanism that enables the spread of larvae into warmer waters along the Australian coast. The difference in responses to temperature treatments also suggests a complex effect of environmental temperature on the regulation of gene expression. Given the relatively small amounts of neutral genetic diversity in the invasive range, further work could examine the role of epigenetic modifications that modulate gene expression in invasive A. amurensis exposed to elevated temperatures. Additionally, the viability and fitness of larvae through to later stages is something that should be explored further.
5.4.2 Comparison to the native *A. amurensis* temperature response

Previously, we have shown that there are large differences in baseline gene expression profiles between invasive and native populations (Chapter 4). As such, it is not surprising that the responses to thermal stress between the invasive and native populations also differ. Larvae from the invasive population may differ physiologically from larvae from the native population due to selective pressures experienced during the establishment period and we have previously shown that there are significant differences in baseline gene expression between the populations (Chapter 4). Given fertilization and development were carried out experimentally in common environments, the disparity in response to elevated temperatures should underlie any evolved physiological differences between the two populations, while commonalities support conserved physiological responses or maintained phenotypic plasticity. Both populations exhibit co-expression of only four genes in response to the elevated temperatures, including two Serine/threonine-protein kinases and laccase 1 isoform G. As previously indicated, Serine/threonine-protein kinases are involved in cellular regulation, proliferation, signalling and apoptosis (Cross *et al.* 2000). Laccases are involved in the biological degradation of lignin, which is found in the cell walls of some algae (Martone *et al.* 2009; Karp *et al.* 2012).

The common expression profile of laccase 1 isoform G across both populations may constitute a specific response to digestion of the algal food stock, *Chetocherous muleri*, or the maintenance of mechanisms to digest algae at elevated temperatures. The expression of these genes could represent a shared response to maintain effective regulation of important cellular processes at elevated temperatures, including digestion.
In the native population, gene expression changes in response to elevated temperatures are substantially different to the invasive, both in terms of the magnitude of differential expression and the genes underlying the response. For example, we find an up-regulation of DNA topoisomerase 2 for both the elevated temperatures in the native population and not in the invasive. DNA topoisomerase 2 (TOP2) is involved in DNA damage sensing and repair in response to cellular stress (Li et al. 1999; Kültz 2005) and has been implicated in thermal adaptation (Lopez-Garcia & Forterre 1999; López-García 1999; López-García & Forterre 2000). This could suggest invasive *A. amurensis* are unable to detect temperature-induced DNA damage to the same extent as do native larvae, i.e. they are not coping as well with elevated temperatures, and going into (unsuccessful) damage control. Alternatively, the discrepancy may arise because invasive *A. amurensis* larvae do not experience as much DNA damage as do the native population, suggesting greater resistance to thermally induced DNA damage. These hypotheses will need to be tested experimentally to examine the specific effect of changed expression of each gene on fitness, and the physiological pathways involved.

### 5.4.3 Evolutionary implications

The comparative approach we adopted has provided a unique opportunity to gain insights into the transcriptional response between native and invasive populations of *A. amurensis* to elevated temperatures during fertilization and larval development. Much of the temperature-induced transcriptome differs between the populations, with invasive *A. amurensis* exhibiting a much greater response in terms of maintaining essential cellular functions. This could
indicate that *A. amurensis* experience greater thermal stress in the invasive range or they have increased capacity to regulate physiological responses to thermal stress than the native population. Regardless, the discrepancy in gene expression between the populations suggests differences in either the molecular targets or regulation of gene networks by elevated temperatures, potentially through a mechanism of adaptation or evolution of increased plasticity. For example, differences between the two populations may arise from subtle changes in regulatory mechanisms. Differences in a few high-level gene regulators can elicit a large change in gene expression across many genes through shared regulatory elements, rather than requiring the independent evolution of many gene specific regulatory regions; as has been suggested for gene expression shifts across an environmental gradient in the purple sea urchin (Pespeni *et al.* 2013b).

The relatively few commonalities in gene expression between populations indicate a minimal conserved response to elevated temperatures between the native and invasive range. This is in direct contrast to the large conserved temperature induced stress responses between native and invasive mussel species (Lockwood *et al.* 2010), although a diverse response to temperature was observed in *A. elegantissima* (Richter *et al.* 2009). However, there are similarities in the functional roles of genes activated between the populations (regulation of transcription, DNA replication, homeostasis) despite the underlying genes being different. Clearly, no singular process dominates the thermal response. Interestingly, native larvae exhibited less ability to survive at the high temperature treatment (20°C), despite the same larval families surviving at lower temperatures. As we found no high bacterial load or contamination it is unlikely that this reflects a technical issue with the
experiment. As such the difference in survival could be explained by more
limited potential within the native population for larvae to respond and develop
at higher temperatures. The greater survival of larvae at 20°C in the invasive
population may therefore reflect the action of selection and local adaptation,
and hints that the invasive populations may have evolved a greater capacity to
cope with thermal stress or higher thermal maxima. Interestingly, these
differences arise despite lower variation in gene expression diversity across the
invasive range temperature treatments (see Fig. S2 and Chapter 4) suggesting
that the invasive population may have acquired these adaptations after
experiencing a bottleneck in gene expression diversity.
5.5 Conclusion

Our results suggest *A. amurensis* exhibit substantially different responses to elevated temperature between the native and invasive ranges. Elevated temperatures elicit a greater transcriptional response in invasive *A. amurensis* larvae, activating important genes for development, cellular regulation and function being activated. Up-regulation of those genes potentially comprises a genetic basis of environmental adaptation. Much of the gene expression is involved is post-translational protein modification and understanding the function of these changes on larval fitness and the physiological pathways involved will be an important next step in unravelling the full extent of the observed molecular responses. For invasive *A. amurensis* to expand their range within Australia, larvae will have to disperse and survive in warmer waters. Given the relatively short invasion history (~30 generations) our data provide new insights into rapid changes in the genetic architecture underlying larval physiological responses to elevated temperatures and the potential for this species to expand its invasive range.
5.6 Supplementary Material

5.6.1 Methods

5.6.1.1 Evaluation of experimental diagnostic performance

Technical performance of the experiment was analysed using erccdashboard (Munro et al. 2014) and the expression values of the individual ERCC transcript ratios. The use of these controls for this experiment has been described in Chapter 4. The importance of these external controls and the analysis methods of erccdashboard enabling the examination of diagnostic performance can be found in Munro et al. (2014). A particular strength of the addition of ERCC ‘spike ins’ is that it enables the assessment of reproducibility and comparison of experimental performance between experiments. The larval temperature experiment had to be conducted at different times in Japan and Australia as *A. amurensis* spawn in each respective hemisphere’s winter. As such, addition of the ERCC transcripts allows us to compare and evaluate the performance of the experiment in both the native and invasive populations separately. Accordingly, these diagnostic performance metrics were calculated separately for each population but included each of the three temperature treatments. We utilized the diagnostic performance metrics to set our acceptable false discovery rates and *p*-values for calling differentially expressed genes within each population.
### 5.6.2 Results

**Table 1.** Summary of Illumina sequencing, filtering and mapping results for both the native and invasive populations. Samples used in each treatment, Total reads obtained from sequencing and the percentage mapped.

<table>
<thead>
<tr>
<th></th>
<th>Total Reads</th>
<th>Reads after filtering</th>
<th>Mapped reads</th>
<th>Mapped %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Native</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(14°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J114</td>
<td>109,405,130</td>
<td>87,467,780</td>
<td>60,105,750</td>
<td>68.72</td>
</tr>
<tr>
<td>J214</td>
<td>100,798,756</td>
<td>73,824,320</td>
<td>50,498,646</td>
<td>68.40</td>
</tr>
<tr>
<td>J314</td>
<td>71,677,612</td>
<td>53,544,428</td>
<td>36,634,926</td>
<td>68.42</td>
</tr>
<tr>
<td>J414</td>
<td>63,708,076</td>
<td>49,247,200</td>
<td>34,327,816</td>
<td>69.71</td>
</tr>
<tr>
<td>J514</td>
<td>64,302,226</td>
<td>47,796,498</td>
<td>33,544,960</td>
<td>70.18</td>
</tr>
<tr>
<td>J614</td>
<td>70,788,870</td>
<td>57,255,084</td>
<td>39,088,136</td>
<td>68.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J117</td>
<td>91,443,432</td>
<td>70,945,244</td>
<td>47,829,242</td>
<td>67.42</td>
</tr>
<tr>
<td>J217</td>
<td>95,589,886</td>
<td>70,489,440</td>
<td>47,714,608</td>
<td>67.69</td>
</tr>
<tr>
<td>J317</td>
<td>69,268,648</td>
<td>48,069,614</td>
<td>30,590,504</td>
<td>63.64</td>
</tr>
<tr>
<td>J417</td>
<td>68,459,716</td>
<td>50,951,582</td>
<td>35,311,650</td>
<td>69.30</td>
</tr>
<tr>
<td>J517</td>
<td>63,284,482</td>
<td>46,975,852</td>
<td>32,903,472</td>
<td>70.04</td>
</tr>
<tr>
<td>J617</td>
<td>66,389,668</td>
<td>48,633,440</td>
<td>32,482,408</td>
<td>66.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J120</td>
<td>102,019,792</td>
<td>76,210,396</td>
<td>52,783,500</td>
<td>69.26</td>
</tr>
<tr>
<td>J620</td>
<td>64,935,072</td>
<td>50,243,664</td>
<td>33,262,276</td>
<td>66.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Invasive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(14°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P114</td>
<td>68,544,216</td>
<td>31,026,208</td>
<td>21,496,272</td>
<td>69.28</td>
</tr>
<tr>
<td>P214</td>
<td>67,838,726</td>
<td>51,255,528</td>
<td>37,302,312</td>
<td>72.78</td>
</tr>
<tr>
<td>P414</td>
<td>64,702,424</td>
<td>48,189,934</td>
<td>34,450,026</td>
<td>71.46</td>
</tr>
<tr>
<td>P514</td>
<td>67,372,218</td>
<td>49,555,502</td>
<td>35,094,666</td>
<td>70.82</td>
</tr>
<tr>
<td>P614</td>
<td>62,247,820</td>
<td>44,472,982</td>
<td>31,834,726</td>
<td>71.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P217</td>
<td>64,853,860</td>
<td>48,033,988</td>
<td>34,300,983</td>
<td>71.41</td>
</tr>
<tr>
<td>P417</td>
<td>60,129,084</td>
<td>41,488,970</td>
<td>28,659,212</td>
<td>69.08</td>
</tr>
<tr>
<td>P517</td>
<td>67,334,720</td>
<td>28,378,812</td>
<td>19,764,716</td>
<td>69.65</td>
</tr>
<tr>
<td>P617</td>
<td>62,170,376</td>
<td>44,360,548</td>
<td>31,450,902</td>
<td>70.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20°C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P120</td>
<td>69,613,382</td>
<td>33,950,332</td>
<td>23,284,946</td>
<td>68.59</td>
</tr>
<tr>
<td>P220</td>
<td>69,812,906</td>
<td>26,357,854</td>
<td>18,962,936</td>
<td>71.94</td>
</tr>
<tr>
<td>P420</td>
<td>60,781,866</td>
<td>39,728,424</td>
<td>26,985,192</td>
<td>67.92</td>
</tr>
<tr>
<td>P520</td>
<td>67,721,918</td>
<td>47,114,880</td>
<td>33,066,142</td>
<td>70.18</td>
</tr>
<tr>
<td>P620</td>
<td>68,244,810</td>
<td>50,234,116</td>
<td>35,683,584</td>
<td>71.03</td>
</tr>
</tbody>
</table>
5.6.2.1 Assessment of technical and diagnostic performance

We assessed the diagnostic capabilities of the experiment by comparing the expression estimates of the reference genes and ERCC transcripts between the temperature treatments for each population separately (see Methods). We are reliably able to detect gene expression differences across a dynamic range of just under $2^{20}$ for both the invasive and native population (Figs. S1a and S2a, respectively). Both populations exhibit this pattern consistently, indicating we can detect and have sufficient evidence to quantify expression across the range for which the experiment was designed. Comparison between the ERCC ratio expression vs. average expression indicates a reasonably small and consistent variation in the ERCC ratio measurements as a function of dynamic range for the invasive ($\log(r_m) -0.28 \pm 0.16$ weighted SE, Fig. S1c) and native ($\log(r_m) -0.22 \pm 0.27$ weighted SE, Fig. S2c) populations. This suggests no large bias in the mRNA fractions between the samples, across temperature treatments or between the experiments in the native and invasive populations exists, indicating good technical reproducibility. Likewise, the ROC analysis showed our data and analysis had high power to detect real endogenous differential gene expression, at the 2- and 4-fold levels, in both the native and invasive populations; indicated by AUC statistics of 1 (Fig. S1b; Fig. S2b, respectively). The native population showed reduced power to detect a 1.5 fold-change in gene expression (AUC 0.88) compared to the invasive population (AUC 0.99). In the invasive population, our diagnostic power increases with an increased fold-change. The LODR analysis indicates that we can reliably detect expression differences as true positives for the entire designed ratios fold-changes, at a threshold p-value < 0.1, and 2- and 4-fold-changes at p-value 0.05 (Fig. S1d). Based on these metrics we set a minimum fold-change threshold of 2 and a p-value threshold of
0.05 for filtering endogenous differential expression in the invasive population. For the native population we were unable to obtain a LODR estimate and these approach infinity for all of the designed ratios fold-changes (Fig. S2d). This most likely arises from there being much less variation in endogenous gene expression across the native population temperature treatments. As such, we decided to use the same filtering metric as for the invasive population, primarily because more stringent filters would have removed all signature of differential expression in the native population.
Figure S1. ERCC technical and diagnostic plots produced by erccdashboard for the invasive population larval temperature experiment. A) Signal-abundance plot, points are coloured by ratio sub-pool, shape represents the sample type, error bars denote the standard deviations of the replicates. B) ROC curves and AUC statistics for each group of true-positive ERCC controls (detected = number of controls used, spiked = the total included in the ERCC control mixture). C) MA plot of ERCC ratio measurement variability and bias. Coloured data points represent the mean ratio measurement per ERCC transcript, error bars the standard deviation of the replicates ratios, and filled circles are ERCC ratios above the LODR estimates. Grey points denote endogenous transcript gene expression measurements. Nominal ERCC ratios for each sub-pool are annotated with coloured solid lines, dashed lines represent the adjusted ratios based on the estimate of mRNA fraction differences between the samples, rm. D) LODR estimates are indicated by coloured arrows for each fold-change that crosses the threshold p-value, the black dashed line denotes the threshold p-value derived for the chosen FDR. LODR results and bootstrap confidence interval are provided in the table below the plot.
Figure S2. ERCC technical and diagnostic plots produced by erccdashboard for the native population larval temperature experiment. A) Signal-abundance plot, points are coloured by ratio sub-pool, shape represents the sample type, error bars denote the standard deviations of the replicates. B) ROC curves and AUC statistics for each group of true-positive ERCC controls (detected = number of controls used, spiked = the total included in the ERCC control mixture). C) MA plot of ERCC ratio measurement variability and bias. Coloured data points represent the mean ratio measurement per ERCC transcript, error bars the standard deviation of the replicates ratios, and filled circles are ERCC ratios above the LODR estimates. Grey points denote endogenous transcript gene expression measurements. Nominal ERCC ratios for each sub-pool are annotated with coloured solid lines, dashed lines represent the adjusted ratios based on the estimate of mRNA fraction differences between the samples, \( r_m \). D) LODR estimates are indicated by coloured arrows for each fold-change that crosses the threshold p-value, the black dashed line denotes the threshold p-value derived for the chosen FDR. LODR results and bootstrap confidence interval are provided in the table below the plot.
Figure S3. Multidimensional scaling plot of the variation of per-sample expression across the native and invasive range and temperature treatments. Sample designation beginning with P from invasive range, those beginning with J from native range. The last 2 digits in sample names correspond to the temperature treatment. Distances between samples correspond to leading-log-fold-change between all combinations of sample pairs. Leading-log-fold dimensions 1 and 2 are shown as these represent the most variation, leading-log-fold-change calculated as the average (mean-square-root) of the largest absolute log-fold-change between sample pairs.
5.7 References


Fabian MR, Cieplak MK, Frank F et al. (2011) miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nature structural & molecular biology, 18, 1211–7.


Goggin LC (1998) Technical Report 15: Proceedings of a meeting on the biology and management of the introduced seastar, Asterias amurensis, in Australian waters (Center for research on marine pests) CSIRO.


Young MD, Wakefield MJ, Smyth GK, Oshlack A (2014) goseq: Gene Ontology testing for RNA-seq datasets Reading data.

Chapter 6. Conclusions

The ecological aspects of biological invasions have been studied extensively and in recent years many studies have informed us of the underlying evolutionary processes responsible for adaptive change often seen in introduced species (Lee 2002; Rollins et al. 2013; Bock et al. 2014). However, what is critically lacking is uncovering the underlying genetic basis of adaptive change in phenotypic traits that are relevant to successful invasions.

In this thesis, I initially revealed significant neutral genetic divergence between native and invasive *A. amurensis* populations (Chapter 2). The Australian introduction most likely has a single origin from Tokyo Bay, which resulted in the loss of some neutral genetic diversity within the invasive range, confirming previous results (Ward & Andrew 1995). Range expansion along the Victorian coast is likely to be facilitated by natural larval dispersal mechanisms from the major source population in Port Phillip Bay. The novel integrative method used provides compelling support for this and underlies the importance of larval traits for further range expansion.

I produced the first transcriptome resource for *A. amurensis* in order to examine evolutionary processes affecting dispersive larvae (Chapter 3). I identified protein-coding genes and associated functional information that allows us to understand the genetic architecture linked to physiological functions and roles. The *A. amurensis* bipinnaria transcriptome assembled contains 18,319 uniquely annotated protein coding genes and shows strong orthology to previous described Bat star proteins while demonstrating a comparable number of genes expressed to sea urchin developmental stages.
I found substantial gene expression divergence between larvae from the likely native source population and the invasive population providing recruits to the invasive range edge (Chapter 4), within a common environmental setting. Interestingly, the variation in gene expression measurements showed a similar pattern to that of neutral genetic diversity; lower in the invasive range (Chapter 2). Many important genes with putative roles in immune function and the response to environmental contaminants have constituent differences in expression between native and invasive larvae. These genes may have a role in facilitating adaptation of invasive larvae to Australian environmental conditions. In particular, the up-regulation of cytochrome p450 (p450) genes in invasive larvae reveals a response to metabolize environmental xenobiotics. Similar responses have been seen in other invasive species (Qiu et al. 2009; Mittapalli et al. 2010; Hodgins et al. 2013; Ioannidis et al. 2014; Ye et al. 2014), which suggests there may be some common responses to invasion at the molecular level. I also found intriguing results where several p450s were also down-regulated in the invasive range, demonstrating the interplay between their role in steroid synthesis and xenobiotic metabolism (as previously described in other seastars, den Besten 1998; Danis et al. 2004; Rewitz et al. 2006) that warrants further exploration. I also identified several genes as putatively experiencing different selection regimes (both balancing and positive) in the invasive range, which are implicated in immune function and developmental processes. This could result from the selection of alleles in invasive larvae that perform better in novel environmental conditions and may indicate the presence of rapid local adaptation during the invasion into Australia.
For invasive *A. amurensis* to expand their range within Australia, larvae will have to disperse and survive in warmer waters (Dunstan & Bax 2007). I found substantial differences in the gene expression response to elevated temperatures (17 and 20°C) between native and invasive larvae (Chapter 5), with invasive larvae exhibiting a much greater transcriptional response than in natives. In invasive larvae, important genes for development, cellular regulation and function are activated, many of which are involved in post-translational protein modifications. The discrepancy in the gene expression response to elevated temperatures between the ranges could suggest differences in either the molecular targets or regulation of gene networks, potentially through a mechanism of adaptation or selection for increased plasticity in invasive larvae. It appears invasive larvae may have evolved greater capacity to cope with higher temperatures during fertilization and larval development, with cellular physiological responses not constrained by invasion history. As such, invasive larvae from Port Phillip Bay may have the potential to disperse and survive at much higher temperatures. These hypotheses should form the basis for future experiments to evaluate the identified gene expression changes on larval fitness.

Given the relatively short invasion (~30 generations) our data together provide new insights into rapid changes in both the genetic architecture and expression of genes underlying larval physiological responses to elevated temperatures and the potential for this species to expand its invasive range.
6.2 References


Appendix 1

Development of novel microsatellite markers for the invasive Northern Pacific seastar, Asterias amurensis

Mark F. Richardson · AnnaLise M. Stanley · Craig D. H. Sherman

DOI 10.1007/s12686-011-9539-8

Abstract The Northern Pacific seastar, Asterias amurensis, is a benthic marine predator, which has recently established several invasive populations in Australian waters. To investigate population structure, diversity and patterns of connectivity, we isolated and characterised 27 microsatellite loci and tested their polymorphism based on 46 individuals from two invasive populations. The mean allelic richness was 4.33; observed heterozygosity was 0.42, while the percentage of polymorphic loci was 92.6%. The polymorphic markers will prove useful in the assessment of population genetic parameters, in both invasive and native A. amurensis populations.

Keywords Northern Pacific seastar · Asterias amurensis · Microsatellites · Invasive species · Genetic structure · Gene flow

The impact of invasive species upon native biodiversity, community structure and ecological processes is widely recognised and considered a major threat to global species loss (Vitousek et al. 1997; Sakai et al. 2001; Molnar et al. 2008). Marine ecosystems are particularly vulnerable to invasions, with coastal ecosystems among those harbouring the highest proportion of non-native species (Grosholz 2002; Reusch et al. 2010). The success of an invading species may be influenced by its genetic architecture as well as physiological tolerance (Lee 2002). As such, revealing the underlying population genetic variation in invasive populations becomes important. The Northern Pacific seastar, Asterias amurensis, is recognised as one of Australia’s most potentially damaging marine invasive species (Goggin 1998). After being introduced into Hobart, Tasmania, in the 1980s, this species has now spread along the east and northern coasts of Tasmania and established a population on the mainland in Port Phillip Bay, Victoria (Goggin 1998; Perry et al. 2004). It is a voracious predator that has the potential to drastically alter native ecosystems and affect aquaculture industry (Roes et al. 2003, 2006). Here we developed a panel of microsatellite markers to examine key population parameters including population structure, diversity and patterns of connectivity. These markers will also provide an important tool kit that will allow managing authorities to identify the source of any new introductions and monitor range expansions.

Genomic DNA was isolated from a single individual from Williams town, Port Phillip Bay, Victoria, Australia, from a combination of gonad tissue and tube feet using ISOLATE Genomic DNA kits (Bioline, USA), following the manufacturers animal tissue protocol. We developed microsatellites following the methodology of (Gardner et al. 2011). Briefly, we sequenced one-sixteenth of a plate using the GS-FLX 454 platform (Roche, Germany), providing 29,109 sequenced reads (totalling 9,861,161 bp) between 23 and 593 bp. A total of 612 assembled contigs (ranging from 100 to 6,402 bp) were analysed for microsatellite repeats (di-, tri-, tetra-, penta- and hexanucleotide) over eight repeats long using MSATCOMMANDER, v0.8.1 (Faircloth 2008). We identified 48 potential microsatellite loci; primer pairs were designed using Primer3, where we excluded short sequence repeats from primer designs. Multiplexes consisting of four loci were designed where forward primers had a fluorescent dye associated tag added (FAM-GCCTCCCTCGCGCCA; NED-GCCTTGCGCGCGCGCCA).
## Appendix 2

**Table 1.** ERCC ‘spike in’ mixes added to each sample and their use for technical performance analysis in thesis chapters.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Ercc mix</th>
<th>Chapter 4</th>
<th>Chapter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Invasive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P114</td>
<td>14°C</td>
<td>2</td>
<td>*</td>
</tr>
<tr>
<td>P214</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P314</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P414</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P514</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P614</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P117</td>
<td>17°C</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P217</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P317</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P417</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P517</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P617</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P120</td>
<td>20°C</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>P220</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P320</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P420</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P520</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P620</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Native** |          |           |           |
| J114       | 14°C     | 1         |           |
| J214       | 1        |           |           |
| J314       | 1        |           |           |
| J414       | 2        |           |           |
| J514       | 2        |           |           |
| J614       | 2        |           |           |
| J117       | 17°C     | 2         |           |
| J217       | 2        |           |           |
| J317       | 2        |           |           |
| J417       | 1        |           |           |
| J517       | 1        |           |           |
| J617       | 1        |           |           |
| J120       | 20°C     | 1         |           |
| J620       | 2        |           |           |
## AUTHORSHIP STATEMENT, Chapter 2

### 1. Details of publication and executive author

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrating genetic analysis, hydrodynamic modeling and plankton sampling reveals a natural dispersal mechanism for range expansion in an invasive marine invertebrate.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of executive author</th>
<th>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</th>
<th>Email or phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Richardson</td>
<td>LES</td>
<td>0413005431</td>
</tr>
</tbody>
</table>

### 2. Inclusion of publication in a thesis

<table>
<thead>
<tr>
<th>Is it intended to include this publication in a higher degree by research (HDR) thesis?</th>
<th>Yes</th>
<th>If Yes, please complete Section 3 If No, go straight to Section 4.</th>
</tr>
</thead>
</table>

### 3. HDR thesis author’s declaration

<table>
<thead>
<tr>
<th>Name of HDR thesis author if different from above. (If the same, write “as above”)</th>
<th>School/Institute/Division if based at Deakin</th>
<th>Thesis title</th>
</tr>
</thead>
<tbody>
<tr>
<td>As above</td>
<td></td>
<td>Evolutionary Biology of the Invasive northern Pacific Seastar, <em>Asterias amurensis</em></td>
</tr>
</tbody>
</table>

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

I contributed to the design of the genetic study and genetic methodology. I conducted sampling for microsatellite analysis across both native and invasive ranges, microsatellite lab work and data analysis. I interpreted all analysis and wrote the manuscript.

*I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.*

<table>
<thead>
<tr>
<th>Signature and date</th>
<th>Signature Redacted by Library</th>
</tr>
</thead>
<tbody>
<tr>
<td>24/03/2015</td>
<td></td>
</tr>
</tbody>
</table>
4. Description of all author contributions

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alastair Hirst, Deakin University</td>
<td>Experimental design and plankton sample collection</td>
</tr>
<tr>
<td>Craig Sherman, Deakin University</td>
<td>Experimental design, field work collections in Japan, preliminary data analysis and comments on the manuscript</td>
</tr>
<tr>
<td>Nathan Bott,</td>
<td>Genetic analysis of plankton samples</td>
</tr>
<tr>
<td>Randall Lee,</td>
<td>Hydrodynamic modelling</td>
</tr>
</tbody>
</table>

5. Author Declarations

I agree to be named as one of the authors of this work, and confirm:

i. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,

ii. that there are no other authors according to these criteria,

iii. that the description in Section 4 of my contribution(s) to this publication is accurate,

iv. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

v. consent to the incorporation of the publication into the candidate’s HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alastair Hirst</td>
<td>Signature Redacted by Library</td>
<td>23/02/2015</td>
</tr>
<tr>
<td>Craig Sherman</td>
<td></td>
<td>23/03/2015</td>
</tr>
<tr>
<td>Nathan Bott</td>
<td></td>
<td>23/02/2015</td>
</tr>
</tbody>
</table>
6. Other contributor declarations

*I agree to be named as a non-author contributor to this work.*

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author.*

7. Data storage

The original data for this project are stored in the following locations. (The locations must be within an appropriate institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin University, permission for this must be given by the Head of Academic Unit within which the executive author is based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple</td>
<td>Deakin University</td>
<td></td>
<td>Craig Sherman Alastair Hirst</td>
</tr>
</tbody>
</table>

*This form must be retained by the executive author, within the school or institute in which they are based.*

*If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.*
# AUTHORSHIP STATEMENT, Chapter 3

## 1. Details of publication and executive author

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>De novo</em> assembly and characterization of the invasive Northern Pacific Seastar transcriptome</td>
<td>In review, BMC genomics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of executive author</th>
<th>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</th>
<th>Email or phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Richardson</td>
<td>LES</td>
<td>0413005431</td>
</tr>
</tbody>
</table>

## 2. Inclusion of publication in a thesis

<table>
<thead>
<tr>
<th>Is it intended to include this publication in a higher degree by research (HDR) thesis?</th>
<th>Yes</th>
<th>If Yes, please complete Section 3 If No, go straight to Section 4.</th>
</tr>
</thead>
</table>

## 3. HDR thesis author’s declaration

<table>
<thead>
<tr>
<th>Name of HDR thesis author if different from above. (If the same, write “as above”)</th>
<th>School/Institute/Division if based at Deakin</th>
<th>Thesis title</th>
</tr>
</thead>
<tbody>
<tr>
<td>As above</td>
<td></td>
<td><em>Evolutionary Biology of the Invasive northern Pacific Seastar, Asterias amurensis</em></td>
</tr>
</tbody>
</table>

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

I contributed to conception and design of study. I conducted experiments and sampling. I conducted all lab work and bioinformatics analysis. I interpreted all analysis and wrote the manuscript.

*I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.*

<table>
<thead>
<tr>
<th>Signature and date</th>
<th>24/03/2015</th>
</tr>
</thead>
</table>

## 4. Description of all author contributions

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Sherman, Deakin University</td>
<td>Conception of project, contribution to design, contribution to manuscript revisions.</td>
</tr>
</tbody>
</table>
5. Author Declarations
I agree to be named as one of the authors of this work, and confirm:
vi. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,
vii. that there are no other authors according to these criteria,
viii. that the description in Section 4 of my contribution(s) to this publication is accurate,
ix. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further
x. consent to the incorporation of the publication into the candidate’s HDR thesis
submitted to Deakin University and, if the higher degree is awarded, the subsequent
publication of the thesis by the university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Sherman</td>
<td>Signature Redacted by Library</td>
<td>23/03/2015</td>
</tr>
</tbody>
</table>

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
</table>

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage

The original data for this project are stored in the following locations. (The locations must be within an appropriate institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin University, permission for this must be given by the Head of Academic Unit within which the executive author is based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple</td>
<td>Deakin University</td>
<td>Mark Richardson Craig Sherman</td>
<td></td>
</tr>
</tbody>
</table>

This form must be retained by the executive author, within the school or institute in which they are based.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.
# AUTHORSHIP STATEMENT, Chapter 4

## 1. Details of publication and executive author

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td>The molecular basis of a successful invasion: comparative RNA-Seq analysis of native and invasive Asterias amurensis</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of executive author</th>
<th>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</th>
<th>Email or phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Richardson</td>
<td>LES</td>
<td>0413005431</td>
</tr>
</tbody>
</table>

## 2. Inclusion of publication in a thesis

<table>
<thead>
<tr>
<th>Is it intended to include this publication in a higher degree by research (HDR) thesis?</th>
<th>Yes</th>
<th>If Yes, please complete Section 3 If No, go straight to Section 4.</th>
</tr>
</thead>
</table>

## 3. HDR thesis author’s declaration

<table>
<thead>
<tr>
<th>Name of HDR thesis author if different from above. (If the same, write “as above”)</th>
<th>School/Institute/Division if based at Deakin</th>
<th>Thesis title</th>
</tr>
</thead>
<tbody>
<tr>
<td>As above</td>
<td></td>
<td>Evolutionary Biology of the Invasive northern Pacific Seastar, Asterias amurensis</td>
</tr>
</tbody>
</table>

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

I contributed to conception and design of the study. I conducted all experiments and sampling. I conducted all lab work and bioinformatics analysis. I interpreted all analysis and wrote the manuscript

*I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.*

<table>
<thead>
<tr>
<th>Signature and date</th>
<th>24/03/2015</th>
</tr>
</thead>
</table>

Signature Redacted by Library

## 4. Description of all author contributions

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Sherman, Deakin University</td>
<td>Conception of project, contribution to design, contribution to manuscript revisions.</td>
</tr>
</tbody>
</table>
5. Author Declarations
I agree to be named as one of the authors of this work, and confirm:
xvi. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,
xx. that there are no other authors according to these criteria,
xxx. that the description in Section 4 of my contribution(s) to this publication is accurate,
xxxv. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further
xxxv. consent to the incorporation of the publication into the candidate’s HDR thesis
submitted to Deakin University and, if the higher degree is awarded, the subsequent
publication of the thesis by the university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Sherman</td>
<td></td>
<td>23/03/2015</td>
</tr>
</tbody>
</table>

Signature Redacted by Library

6. Other contributor declarations
I agree to be named as a non-author contributor to this work.

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
</table>

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author

7. Data storage
The original data for this project are stored in the following locations. (The locations must be within an appropriate institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin University, permission for this must be given by the Head of Academic Unit within which the executive author is based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple</td>
<td>Deakin University</td>
<td>Mark Richardson Craig Sherman</td>
<td></td>
</tr>
</tbody>
</table>

This form must be retained by the executive author, within the school or institute in which they are based.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.
# AUTHORSHIP STATEMENT, Chapter 5

## 1. Details of publication and executive author

<table>
<thead>
<tr>
<th>Title of Publication</th>
<th>Publication details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparative RNA-seq analysis of the temperature response in native and invasive <em>Asterias amurensis</em></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name of executive author</th>
<th>School/Institute/Division if based at Deakin; Organisation and address if non-Deakin</th>
<th>Email or phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Richardson</td>
<td>LES</td>
<td>0413005431</td>
</tr>
</tbody>
</table>

## 2. Inclusion of publication in a thesis

<table>
<thead>
<tr>
<th>Is it intended to include this publication in a higher degree by research (HDR) thesis?</th>
<th>Yes</th>
<th>If Yes, please complete Section 3 If No, go straight to Section 4.</th>
</tr>
</thead>
</table>

## 3. HDR thesis author’s declaration

<table>
<thead>
<tr>
<th>Name of HDR thesis author if different from above. (If the same, write “as above”)</th>
<th>School/Institute/Division if based at Deakin</th>
<th>Thesis title</th>
</tr>
</thead>
<tbody>
<tr>
<td>As above</td>
<td>Evolitional Biology of the Invasive northern Pacific Seastar, <em>Asterias amurensis</em></td>
<td></td>
</tr>
</tbody>
</table>

If there are multiple authors, give a full description of HDR thesis author’s contribution to the publication (for example, how much did you contribute to the conception of the project, the design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)

I contributed to conception and design of the study. I conducted all experiments and sampling. I conducted all lab work and bioinformatics analysis. I interpreted all analysis and wrote the manuscript.

*I declare that the above is an accurate description of my contribution to this paper, and the contributions of other authors are as described below.*

<table>
<thead>
<tr>
<th>Signature and date</th>
</tr>
</thead>
<tbody>
<tr>
<td>24/03/2015</td>
</tr>
</tbody>
</table>

## 4. Description of all author contributions

<table>
<thead>
<tr>
<th>Name and affiliation of author</th>
<th>Contribution(s) (for example, conception of the project, design of methodology or experimental protocol, data collection, analysis, drafting the manuscript, revising it critically for important intellectual content, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Sherman, Deakin University</td>
<td>Conception of project, contribution to design, contribution to manuscript revisions.</td>
</tr>
</tbody>
</table>
5. Author Declarations
I agree to be named as one of the authors of this work, and confirm:

xvi. that I have met the authorship criteria set out in the Deakin University Research Conduct Policy,

xvii. that there are no other authors according to these criteria,

xviii. that the description in Section 4 of my contribution(s) to this publication is accurate,

xix. that the data on which these findings are based are stored as set out in Section 7 below.

If this work is to form part of an HDR thesis as described in Sections 2 and 3, I further

xx. consent to the incorporation of the publication into the candidate’s HDR thesis submitted to Deakin University and, if the higher degree is awarded, the subsequent publication of the thesis by the university (subject to relevant Copyright provisions).

<table>
<thead>
<tr>
<th>Name of author</th>
<th>Signature*</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Sherman</td>
<td></td>
<td>23/03/2015</td>
</tr>
</tbody>
</table>

Signature Redacted by Library

6. Other contributor declarations

I agree to be named as a non-author contributor to this work.

<table>
<thead>
<tr>
<th>Name and affiliation of contributor</th>
<th>Contribution</th>
<th>Signature* and date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* If an author or contributor is unavailable or otherwise unable to sign the statement of authorship, the Head of Academic Unit may sign on their behalf, noting the reason for their unavailability, provided there is no evidence to suggest that the person would object to being named as author.

7. Data storage

The original data for this project are stored in the following locations. (The locations must be within an appropriate institutional setting. If the executive author is a Deakin staff member and data are stored outside Deakin University, permission for this must be given by the Head of Academic Unit within which the executive author is based.)

<table>
<thead>
<tr>
<th>Data format</th>
<th>Storage Location</th>
<th>Date lodged</th>
<th>Name of custodian if other than the executive author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple</td>
<td>Deakin University</td>
<td>Mark Richardson, Craig Sherman</td>
<td></td>
</tr>
</tbody>
</table>

This form must be retained by the executive author, within the school or institute in which they are based.

If the publication is to be included as part of an HDR thesis, a copy of this form must be included in the thesis with the publication.