ASPECTS OF THE EVOLUTIONARY HISTORY OF
A PAIR OF FISH SPECIES (ARRIPIDAE: ARRIPIS)
ON EITHER SIDE OF A BIOGEOGRAPHIC BARRIER
IN SOUTHERN AUSTRALIAN SEAS

Glenn Moore
BSc. (Hons.) MSc.

This thesis is presented for the degree of
Doctor of Philosophy
at
Murdoch University
2012
I declare that all sources are acknowledged and that this thesis is my own account of my research and contains, as its main content, work that has not previously been submitted for a degree at any tertiary education institution. The thesis contains no material previously written by another person except where due reference is made in the thesis itself.

___________________
Glenn Moore
For Dad, Mum & Debra

Nothing in biology makes sense except in the light of evolution.

Theodosius Dobzhansky

I am turned into a sort of machine for observing facts
and grinding out conclusions.

Charles Darwin
Abstract

Pairs of closely related species on either side of a barrier provide an opportunity to test evolutionary hypotheses. In southern Australian seas, many species pairs occur on either side of the Bass Strait, which temporarily becomes an isthmus during glacial stages, yet few evolutionary questions have been explored using these species. This thesis used partial nucleotide sequences from mitochondrial cytochrome \( b \) and \( COI \) genes and length-polymorphism in several nuclear introns in one such species pair, \textit{Arrapis trutta} (east of Bass Strait) and \textit{A. truttaceus} (west of Bass Strait), to address some evolutionary questions. Data from the two remaining members of the family Arripidae, \textit{A. xylabion} and \textit{A. georgianus}, were included to provide context for interpreting patterns in the focal species pair. This research found no evidence for population subdivision in each of \textit{A. trutta}, \textit{A. truttaceus} and \textit{A. georgianus} across their Australian distributions. This likely reflects panmixia, maintained by highly migratory life cycles and restricted breeding distributions. The molecular data also supported the currently recognised four-species taxonomy of the family. Phylogenetic reconstructions did not support the view of a sister-species relationship between the east-west species pair of \textit{A. trutta} and \textit{A. truttaceus} but instead suggested that these two species, together with \textit{A. xylabion}, likely arose as a trifurcation (some 700,000 years ago), and so can be considered as a trio of sibling-species. All \textit{Arripis} species were characterised by low genetic diversity and there was evidence for a more recent and/or severe population contraction in \textit{A. truttaceus} compared to \textit{A. trutta}. These different demographic signals support the \textit{a priori} hypothesis that environmental conditions to the west of the Bass Strait were more severe than to the east during glacial stages and, in particular, the last glacial maximum. The results are considered in the context of the historical marine biogeography of temperate Australia.
Contents

ACKNOWLEDGEMENTS v

1.0 GENERAL INTRODUCTION 1
1.1 FAMILY ARRIPIDAE 7
1.1.1 Arrapis trutta Eastern Australian Salmon 9
1.1.1.1 General Biology and Ecology 9
1.1.1.2 Growth and Age 11
1.1.1.3 Life History 11
1.1.2 Arrapis truttaceus Western Australian Salmon 14
1.1.2.1 General Biology and Ecology 14
1.1.2.2 Growth and Age 15
1.1.2.3 Life History 16
1.1.3 Arrapis xylabion Giant Kahawai 18
1.1.3.1 General Biology and Ecology 18
1.1.4 Arrapis georgianus Australian Herring 19
1.1.4.1 General Biology and Ecology 19
1.1.4.2 Growth and Age 20
1.1.4.3 Life History 20
1.2 PREVIOUS GENETIC STUDIES ON ARRIPIS 22
1.3 APPROACH 24
1.3.1 Genetic Markers 25
1.3.1.1 Mitochondrial DNA 25
1.3.1.2 Nuclear DNA 26
1.3.2 Analytical Methods 28
1.4 AIM AND RATIONALE OF THESIS 29
1.5 OUTLINE OF THESIS 30

2.0 GENERAL METHODS 32
2.1 SAMPLING DESIGN 32
2.2 SAMPLE COLLECTION, PROCESSING AND STORAGE 36
2.3 DNA EXTRACTION 36
2.4 MITOCHONDRIAL DNA 37
2.4.1 PCR Amplifications 38
2.4.2 Sequencing 39
2.4.3 Trials involving other mtDNA markers 40
2.5 NUCLEAR DNA 41
2.5.1 PCR Amplifications 41
2.5.2 Fragment Analysis 42
2.5.3 Primer Selection and Optimisation 45
2.5.4 Allele Scoring 46
2.6 NUCLEAR DNA SEQUENCES 52

3.0 POPULATION GENETIC STRUCTURE 54
3.1 INTRODUCTION 54
3.1.1 Aims and Hypothesis 56
3.2 MATERIALS AND METHODS 57
3.2.1 Sampling 57
3.2.2 Genetic Assays 57
3.2.3 Statistical Analyses 58
3.2.3.1 Levels of Polymorphism 58
3.2.3.2 HWE, Linkage Disequilibrium and Cohort Effects 58
3.2.3.3 Population Differentiation 61
3.2.3.4 Pairwise Differentiation 61
3.2.3.5 Overall Differentiation 62
3.2.3.6 Factorial Correspondence Analysis 62
3.2.3.7 Power Analysis 63
3.2.3.8 Temporal Stability 64
3.3 RESULTS
3.3.1 Polymorphism, HWE and Linkage Disequilibrium 64
3.3.2 Cohort Effects 72
3.3.3 Population Differentiation 72
3.3.4 Power Analysis 73
3.3.5 Temporal Stability 73

3.4 DISCUSSION 81
3.4.1 Conclusions 85

4.0 A MOLECULAR APPROACH TO THE TAXONOMY OF AUSTRALIAN SALMON 86
4.1 INTRODUCTION 86
4.1.1 Aim 90

4.2 MATERIALS AND METHODS
4.2.1 Sampling Regime 91
4.2.2 Genetic Assays 93
4.2.3 Data Analyses
4.2.3.1 mtDNA Sequence Characteristics 93
4.2.3.2 Monophyly of mtDNA Sequences using Phylogenetic Analyses 94
   4.2.3.2.1 Choice of outgroup 94
   4.2.3.2.2 Concatenated sequence data 94
   4.2.3.2.3 Substitution saturation 95
   4.2.3.2.4 Substitution model 96
   4.2.3.2.5 Phylogenetic analyses 97
      4.2.3.2.5.1 Distance-based method 97
      4.2.3.2.5.2 Maximum parsimony 98
      4.2.3.2.5.3 Maximum likelihood 99
      4.2.3.2.5.4 Bayesian inference 100
   4.2.3.2.6 Parsimony networks 102
   4.2.3.3 Fixed Differences Between mtDNA Sequences 103
   4.2.3.4 Genetic distances based on mtDNA sequences 103
   4.2.3.5 Fixed Differences in nDNA Genotypes 103
   4.2.3.6 Differentiation in nDNA Allele Frequencies 103
   4.2.3.7 Factorial Correspondence Analysis using Multilocus nDNA Genotypes 104

4.3 RESULTS 104
4.3.1 Mitochondrial DNA 104
   4.3.1.1 Monophyly of mtDNA Sequences using Phylogenetic Analyses 108
   4.3.1.2 Fixed Differences Between mtDNA Sequences 112
   4.3.1.3 Intraspecific versus Interspecific Genetic Distances in mtDNA 112
4.3.2 Nuclear DNA 113
4.4 DISCUSSION 115
4.4.1 Conclusions 118

5.0 MOLECULAR PHYLOGENETICS, SPECIATION AND BIOGEOGRAPHY 119
5.1 INTRODUCTION 119
5.1.1 Objective 120

5.2 MATERIALS AND METHODS
5.2.1 Data 121
5.2.2 Data Analyses
   5.2.2.1 Relationships Among Species of Australian Salmon 122
      5.2.2.1.1 Testing between alternative tree topologies 122
      5.2.2.1.2 Intron length polymorphism 123
   5.2.2.2 Exploring the Phylogenetic Signal 124
      5.2.2.2.1 Testing for a polytomy 125
      5.2.2.2.2 Likelihood mapping 125
      5.2.2.2.3 NeighborNet (Split Networks) 126
   5.2.2.3 Molecular Dating 127
      5.2.2.3.1 Testing for a clock-like rate of substitution 127
      5.2.2.3.2 Divergence rate 128
      5.2.2.3.3 Time to most recent common ancestor (T_{mra}) 128

5.3 RESULTS 129
5.3.1 Relationships Among Species of Australian Salmon 130

iii
5.3.1.1 Support for Alternative Tree Topologies 135
5.3.2 Exploring the Phylogenetic Signal 136
5.3.3 Molecular Dating 139
5.4 DISCUSSION 141
5.4.1 Hard Polytomy in Australian Salmon 141
5.4.2 Pleistocene Divergence 144
5.4.3 Speciation in Australian Salmon 145
5.4.4 Conclusions 147

6.0 GENETIC DIVERSITY AND RECENT EVOLUTIONARY HISTORY 148
6.1 INTRODUCTION 148
6.1.1 Pleistocene Glacial/Interglacial Cycles in Southern Australia 149
6.1.1.1 East of the Bass Strait 149
6.1.1.2 West of the Bass Strait 152
6.1.2 Using Genetic Data to Infer the Evolutionary Histories of *Arripis* Species 153
6.1.3 Aim 153
6.2 MATERIALS AND METHODS 154
6.2.1 Sampling Regime 154
6.2.2 Genetic Assays 155
6.2.3 Genealogical Relationships 155
6.2.4 Data Analyses 156
6.2.4.1 Genetic Diversity 156
6.2.4.2 Comparisons of Genetic Diversity Between Species 158
6.2.4.3 Neutrality of Markers 158
6.2.4.4 Demographic History 159
6.2.4.4.1 Tests for evidence of changes in population size 159
6.2.4.4.1.1 Mitochondrial DNA sequences 159
6.2.4.4.1.2 Nuclear DNA allele frequencies 161
6.2.4.4.2 Estimates of timing of population expansions 162
6.2.4.4.3 Estimates of effective population sizes 163
6.3 RESULTS 163
6.3.1 General Information 163
6.3.1.1 Mitochondrial DNA Sequences 163
6.3.1.2 Nuclear DNA Allele Frequencies 164
6.3.1.3 Genealogical Relationships 166
6.3.2 Genetic Diversity 168
6.3.2.1 General 168
6.3.2.2 Comparisons of Genetic Diversity Between Species 170
6.3.3 Neutrality of Markers 171
6.3.4 Demographic History 173
6.3.4.1 Tests for Evidence of Changes in Population Size 173
6.3.4.1.1 Mitochondrial DNA sequences 173
6.3.4.1.2 Nuclear DNA allele frequencies 175
6.3.4.2 Estimates of Timing of Population Expansions 178
6.4 DISCUSSION 179
6.4.1 Genetic Diversity in *Arripis* 179
6.4.2 Recent Evolutionary History of *Arripis trutta* and *Arripis truttaceus* 182
6.4.3 Comparison of *Arripis truttaceus* and *Arripis georgianus* 185
6.4.4 Recent Evolutionary History of *Arripis xylabion* 186
6.4.5 Conclusions 187

7.0 GENERAL CONCLUSIONS 188
REFERENCES 192
APPENDICES
A.1 mtDNA primers used during primer development 219
A.2 EPIC-PCR primers tested in this study 224
A.3 Technical issues for the use of fragment length polymorphism in EPIC markers 227
Acknowledgements

Firstly, I thank my supervisor, Dr. Jennie Chaplin. Jennie’s support, advice, guidance, encouragement and enthusiasm have been exceptional. Perhaps most importantly, Jennie has taught me to think, to learn and to persevere. Throughout, Professor Ian Potter has provided encouragement and financial support for my project. The Centre for Fish and Fisheries Research and all its members also gave me support and advice.

I enjoyed sharing the lab with my friends Nicole Philips, Ertuğ Sezmiş, Michelle Gardner and Bryn Farmer, and they all provided a great deal of support, advice, help, laughter and an understanding ear when things just weren’t going to plan. I especially thank Ertuğ for suggesting EPIC-PCR as a way to include nuclear data in my project. I am indebted to Masaki Miya, Natural History Museum, Chiba, Japan for providing the complete mitochondrial genome of *Arripis trutta*. I thank Frances Brigg (State Agricultural Biotechnology Centre) who ran most of my sequences and fragment analyses. Frances’ knowledge and problem solving skills rescued my project on many occasions. My friends Dr. Michael Craig, Drs. Roddy and Marie Hale kept me going with words of experience, encouragement and advice, and crucial moments of escape.

This project relied on the assistance and generosity of many people, primarily in providing and/or organising samples: Dumpy Wheatcroft (Albany Seafoods, Albany); Alan & Peta Miles (Selim Fisheries, Dunsborough); Steeg Hoeksema, Ben Chuwen, Dave Fairclough, George Young (Centre for Fish and Fisheries Research); Kyle and Yuki Armstrong (South Australia), Roddy Hale (Lincoln University), Marie Hale (University of Canterbury), Ian Kerr and Sallyann Gudge (Lord Howe Island Marine Park), Kevin Grickage, Rod Grickage and many anonymous recreational fishers. Fellow *Arripis* researchers, John Stewart and Julian Hughes (NSW Department of Primary Industries) kindly sourced all of the *A. trutta* from eastern Australia and prepared the tissue before sending to me.

Mum, Natalie, my extended family, and my dogs Basie and Duke have helped me through with constant love and support. Finally, my biggest support came from my wife Debra who endured the journey with me, pretended to understand my explanations of neutral theory, coalescence and Bayesian inference, and survived being a ‘Ph.D. widow’ with incredible patience and love.
1.0 General Introduction

The marine fauna of temperate Australia is characterised by very high endemism, with most species restricted to the Southern Australian bioregion (Wilson and Allen 1987; Ponder 1995; Edgar 2000). For example, in temperate inshore fishes, 85% of the ~600 species are endemic, with a further 11% restricted to Australian and New Zealand waters (Wilson and Allen 1987; Hutchins 1994). Endemism at the generic level is also remarkably high in several groups, including fishes, molluscs and echinoderms, and probably more for which data are currently limited (Wilson and Allen 1987). Despite awareness of these unique communities, very little is known about the underlying evolutionary histories of the marine fauna of southern Australia and yet this provides the key to the origin of faunal characteristics of the region.

One of the most intriguing faunal characteristics of southern Australian seas is the reported presence of a large number and diverse range of geminate species to the east and west of the Bass Strait. (N.B. Geminate species are defined as “... twin species [and specifically, each other’s closest relatives, i.e. sister-species] - each one representing the other on opposite sides of some form of barrier”; Jordan 1908, p. 75). For example, at least 30 pairs of sister-species of fish alone are thought to be represented on either side of the Bass Strait (Hutchins 1994) and other groups such as macro-algae (Womersley 1981) and a range of invertebrates (Whitley 1932; Bennett and Pope 1953; Dartnall 1974; O’Hara and Poore 2000) are similarly believed to contain large numbers of such species pairs. However, this observation is derived largely from records of the distributions of morpho-species and only rarely has the geminate status of the eastern and western species been confirmed with molecular evidence. Such confirmation is
important because, in some cases (pertaining to other regions), molecular data have shown that members of a presumed geminate are actually most closely related to other species on the same side of the barrier, rather than to each other (e.g. Jones et al. 2003; Craig et al. 2004; Marko and Moran 2009; Miura et al. 2010).

The presence of many geminate species to the east and west of the Bass Strait, if confirmed, provides strong circumstantial support for the view that the Bass Strait is one of the most significant marine biogeographical features of the temperate Australian region. Other evidence includes the occurrence of a range of species and community limits in the general vicinity of this feature (Whitley 1932; Bennett and Pope 1953; Dartnall 1974; Womersley 1981; Hutchins 1994; O’Hara and Poore 2000; Hidas et al. 2007). Speciation in southern Australia has often been linked to repeated glacial cycles that affected the area with substantial fluctuations in climate and sea level (Dartnall 1974; Hutchins 1987; Howard and Prell 1992; Chenoweth et al. 1998; Petit et al. 1999; Edgar 2000; Hewitt 2000; Lambeck et al. 2002). During glacial maxima, sea levels were up to 140 metres lower than present, resulting in the emergence of a land bridge, the Bassian Isthmus, that extended from mainland Australia to southern Tasmania (Wells and Okada 1996; Lambeck et al. 2002). The Bassian Isthmus is widely believed to have subdivided a broad range of taxa and ultimately, in some cases, resulted in geminate species (Figure 1.1). In addition, during glacial events a cold-water front (Subtropical Convergence Zone) moved northwards, causing other large-scale changes to sea temperatures and current flows (Wells and Okada 1996; see Chapter 6), which may also have been pivotal in speciation processes (e.g. Burridge 2000b). It is often assumed that the timing of these speciation events is linked to glaciation events in the
Pleistocene (e.g. Dartnall 1974; Hutchins 1987; Chenoweth et al. 1998; Edgar 2000), although this has rarely been tested (but see Burridge 2000b, for example).

![Figure 1.1 Hypothetical effect of glaciation on the distribution of temperate Australian marine fauna west (grey) and east (black) of the Bass Strait.](image)

A. The Bassian Isthmus land-bridge was exposed by lower sea levels during glaciation (black outline, shown for southern Australia only). The distribution of temperate fauna was also shifted northward by a cold water front. B. Post glacial range shifts resulted in present day distributions of pairs of species and populations divided at the Bass Strait. After Waters et al. (2005).

The evolutionary implications of marine biogeographical barriers are probably best known in relation to the Isthmus of Panama (e.g. Bermingham et al. 1997; Knowlton and Weigt 1998; Williams et al. 2001; Arbogast et al. 2002; Marko 2002; McCartney and Lessios 2002; Lessios 2008; Hurt et al. 2009; Miura et al. 2010). However, unlike this isthmus, which has permanently isolated communities on either side, many barriers in marine systems, e.g. the Bassian Isthmus, the Florida Peninsula, Baja California, the Sunda Shelf, are in the form of a temporary and/or incomplete peninsula, often subject
to repeated inundations and exposures during sea level changes. Studies on the evolutionary implications of these temporary or incomplete barriers are limited (e.g. based on limited taxonomic coverage; see Bowen and Avise 1990; Burridge 2000b; Bernardi et al. 2003; Rocha and Bowen 2008) but have raised some interesting considerations regarding the mode and tempo of marine speciation, as follows.

Firstly, even such temporary barriers are sufficient to cause speciation, at least in some circumstances (e.g. as evidenced by multiple sets of sibling species on either side of the barrier). Secondly, the fact that speciation can occur in the face of incomplete or temporary barriers adds credence to the view that parapatric (rather than allopatric) speciation may be very common in the marine environment (Gavrilets 2003; Rocha and Bowen 2008). Thirdly, the distribution of some taxa around these incomplete biogeographic barriers can be explained by dispersal events, often associated with post-glacial range expansion (e.g. Bernardi and Lape 2005), providing an alternative to a strictly vicariant regional model enforced in studies of the complete barrier of the Isthmus of Panama. Fourthly, these barriers are making an ongoing contribution to speciation (e.g. as evidenced by the presence of taxa at various stages and ages of divergence; see Bowen and Avise 1990; Bernardi et al. 2003). Finally, contrary to expectations, most of the speciation events are estimated to have occurred in the Pliocene or Miocene rather than in the Pleistocene (Rocha and Bowen 2008, although see Rocha et al. 2005; Robertson et al. 2006). Nevertheless, as noted above, studies of the evolutionary implications of temporary/incomplete land-bridges are limited to a few species at a few sites, and so additional studies are required to investigate the generality of these results.
In the context of testing hypotheses about the Bass Strait/Bassian Isthmus, studies that have used molecular data to assess the relationships among putative geminate species specifically, are essentially limited to Burridge (2000b; based on data in Burridge 1999; Burridge and White 2000; Burridge 2000a) and Waters et al. (2004). The former study used mitochondrial DNA sequence data to examine the relationships among species pairs in three genera of cirrhitoid fishes. In two of the three genera, the results were consistent with vicariant speciation associated with changes in oceanographic conditions in southern Australia during glaciation, but at different times, with one pair estimated to have diverged in the Miocene (7.8 – 11.2 Mya) and the other in the Pliocene (2.7 – 3.8 Mya). The third species pair was considered to have diverged by vicariance around the north of Australia (Burridge and White 2000; Burridge 2000b).

In the second example, Waters et al. (2004) used mtDNA and nDNA data to explore relationships among lineages of *Patiriella* (now *Meridiastra*) seastars, hypothesising, *inter alia*, that vicariance associated with the Bassian Isthmus isolating barrier was a major mechanism of speciation in the group. The authors demonstrated the existence of a pair of east-west distributed sister-species that was generally consistent with that hypothesis; they were estimated to have diverged in the Pliocene around 2.1 – 2.4 Mya.

In addition, there is a handful of studies that have used molecular data to investigate the relationships between populations of a single species to the east and the west of the Bass Strait, either incidentally or as a deliberate attempt to examine the role of the Bassian Isthmus (and associated oceanographic changes) in generating allopatric divergence (e.g. Brown 1991; Nurthen *et al.* 1992; Colgan and Paxton 1997; Kassahn *et al.* 2003; Waters and Roy 2003; Dawson 2005; Waters *et al.* 2005; York *et al.* 2008; Ayre *et al.*
of these intraspecific studies are also consistent with the view that the Bass Strait/Bassian Isthmus is an important isolating barrier.

Once the relationship between apparently closely related species on either side of a barrier such as the Bassian Isthmus is established, these species also provide a ‘natural experiment’ to test hypotheses about the evolutionary implications of certain environmental conditions because they represent very similar genomes placed in different environments (Lessios 1998). For example, the evolution of different life-history strategies in allopatric sister-species may be driven by environmental factors such as localised differences in water conditions or habitats (e.g. Cambray 1994; Coleman et al. 2011). Moreover, these ‘natural experiments’ are particularly important in the context of elucidating the effects of glacial cycles on the demographic histories of coastal marine species. In such species, departures from mutation-drift equilibrium are typically explained in terms of glacial effects based on a posteriori explanations of the results (e.g. Avise 2000; Bowen et al. 2001; Beheregaray et al. 2002; Bernardi 2005; López et al. 2010).

To illustrate the importance of studying geminates, there are several investigations that have shown that the genetic signatures of demographic events may be shared between species, but may differ in strength and/or timing, which is related to differences in historical environments. For example, geminate species of blenniid fishes in the Atlantic and Mediterranean basins were both affected by temperature changes during glacial events, but the retention of diversity was greater in one of the pair due to the presence of refugia within its range (Domingues et al. 2008). In another example,
Lessios *et al.* (2001) found evidence for population expansions in trans-isthmian geminate sea urchins, but the date of the start of the expansion in each ocean may have differed by some 40,000 years. Similarly, Muss *et al.* (2001) accounted for contrasting population subdivision in trans-isthmian geminate fishes with reference to dissimilar historical, geological and ocean circulation events in the Pacific and Atlantic basins.

In terms of the Bassian Isthmus, there is considerable paleontological evidence that the effects of glaciation were more severe to the west of this biogeographical feature than to the east (see Chapter 6). Thus far, the only *a priori* test of this assertion used trans-isthmian populations of the kelp *Durvillaea potatorum* (Fraser *et al.* 2009). This work tended to support the idea that glacial conditions were worse on the west however, the data interpretation is predicated on the assumption that the populations to the west of the Bass Strait were extirpated during the last glacial maximum, and is complicated by the potential influence of contemporary gene flow and selection (Fraser *et al.* 2009). A closely related pair of species, rather than conspecific populations, would help to overcome some of these difficulties.

### 1.1 Family Arripidae

This study provides one of the first detailed investigations into the evolutionary history of a fish taxon in the Southern Australian bioregion. The family Arripidae was selected for this investigation for several reasons, as follows. Firstly, there is assumed geminate species with a representative on either side of the Bassian Isthmus, as well as two other congeners that permit results for the geminates to be considered in a broader context (Hutchins 1994). Secondly, the family is endemic to the region of interest (temperate
Australasian seas; Paulin 1993) and so presumably the entire history of the family has been shaped by factors specific to this region. Thirdly, the family has few species (four), which are large, well known and conspicuous (Paulin 1993). Thus, it is almost certain that complete taxonomic coverage of the family, at least in regard to the extant members, is included in this study. Finally, as detailed below, the biology of the family is relatively well known, providing the foundations for interpreting genetic patterns.

In order to interpret the recent evolutionary histories of the *Arripis* species, it is important to understand the biology of each species and accordingly, the following section contains a brief introduction to the family and the first detailed review of published and unpublished research into the biology of all members of the family.

The Arripidae is one of very few fish families that are endemic to temperate Australian and New Zealand seas (Wilson and Allen 1987). This thesis follows the taxonomy of Paulin (1993) and so recognises a single genus *Arripis*, with four species, and uses the nomenclature of Eschmeyer (2011) and Rees *et al.* (2006). Three of the species are large and morphologically similar and collectively referred to as ‘Australian Salmon’, while the remaining species is smaller and commonly referred to as ‘Australian Herring’. In Australian mainland waters, the family is represented by two species of Australian Salmon and the Australian Herring. The Eastern Australian Salmon (*A. trutta*) occurs on the east coast, while the Western Australian Salmon (*A. truttaceus*) and the Australian Herring (*A. georgianus*) occur on the south and west coasts (Figure 1.2). *Arripis trutta* also occurs in New Zealand waters, where it is known as Kahawai. The third Australian Salmon species, the Giant Kahawai (*A. xylabion*) occurs around several south-west Pacific islands and possibly also northern New Zealand (Figure 1.2).
Figure 1.2  Approximate distributions of the four species of Arripidae: *Arripis trutta* (yellow), *A. truttaceus* (pink), *A. xylabion* (orange), and *A. georgianus* (blue).

1.1.1  *Arripis trutta* (Forster 1801)  Eastern Australian Salmon

1.1.1.1  General Biology and Ecology

*Arripis trutta* occurs in temperate coastal waters of eastern Australia, from north of Sydney, New South Wales (occasionally into southern Queensland) to the eastern parts of Victoria and south to Tasmania (Stanley 1978; Paulin 1993; Hutchins and Swainston 1999; Figure 1.2). It is also found around the North Island and northern parts of the South Island of New Zealand (Paulin 1993). Adults form large schools that inhabit exposed beaches and rocky reefs (Stanley 1978), with some evidence that their movements may also include a more oceanic distribution (Malcolm 1960). Adults generally reach up to 89 cm in length and 7 kg in weight (Hutchins and Swainston 1999), although they may grow larger (Duffy and Petherick 1999).
Chapter 1. General Introduction

Dietary studies indicate that adults of *A. trutta* are opportunistic, with around 65% of their diet being fish (especially clupeids) and the rest comprising crustaceans and some molluscs (Baker 1971; Webb 1973b; Stewart et al. 2011). Documented predators of the species are few but include bottlenose dolphins *Tursiops truncatus* (Gales et al. 1992), little penguins *Eudyptula minor* (Montague and Cullen 1987), great cormorants *Phalacrocorax carbo* and pied cormorants *P. varius* (Marchant and Higgins 1990). Anecdotal evidence from fishers indicates that other marine mammals and sharks also target them. Parasites, with potential deleterious effects, are known from *A. trutta* (Hutson et al. 2004; Bolton et al. 2006; Catalano and Hutson 2010; Catalano et al. 2010; 2011).

A commercial fishery targeting *A. trutta* has existed for many decades in eastern Australia and New Zealand, landing some one million fish per year (Scandol et al. 2008; TPC 2009; Anon. 2010; Stewart et al. 2011) and the species has a wide recreational angling appeal, with more than a million individuals caught annually in Australia (Henry and Lyle 2003) and New Zealand (Anon. 2010). The economic importance of this species led to the commencement of monitoring research in the 1940s and considerable resources were committed to tag/recapture and growth studies throughout the species’ range after catches began to fall in the 1950s and 1960s (Fairbridge 1950; Malcolm 1959; Nicholls 1973; Stanley and Malcolm 1977; Stanley 1978; 1988a). *Arripis trutta* is known to have been a reasonable component of historical and contemporary Australian and New Zealand indigenous fish catches (Leach et al. 1996; Anon. 2010; Waddell 2011).
1.1.1.2 Growth and Age

Post-larval growth in *A. trutta* varies slightly across the species distribution (Nicholls 1973; Stewart *et al.* 2011) and age estimates from otolith annuli predict a maximum age of around 12 years in Australia (Stewart *et al.* 2011), and possibly longer in New Zealand (Egglestone 1975). Age at maturity is around 2-4 years (Stanley and Malcolm 1977; Stewart *et al.* 2011).

1.1.1.3 Life History

The life history of *A. trutta* is characterised by a broad adult distribution and a highly dispersive pelagic larval stage followed by a juvenile stage that migrates from nursery areas to the adult distribution (Figure 1.3). Between April and October, most adults of *A. trutta* are concentrated in waters off New South Wales with quiescent gonads (Stanley 1978; Stewart *et al.* 2011). Adult aggregations are highly variable in size and distribution (e.g. Scandol *et al.* 2008) and anecdotal evidence suggests a recent increase in overall size and range (e.g. Miller 2007; Williams 2007). Extensive tagging and

---

**Figure 1.3 (over page)** Generalised map of the distribution and movements of *Arripis trutta* (dashed arrows), *Arripis truttaceus* (black arrows), and *Arripis georgianus* (grey arrows) in Australian waters. ‘I’ indicates presumed larval dispersal, including some local retention; ‘II’ represents direction of movement of juvenile fish; ‘III’ represents movements of sub-adult fish; and ‘IV’ represents adult distribution and movements.
otolith microchemistry studies indicate considerable movement of adult fish with no notable evidence for site-fidelity or fish returning to nursery areas (Stanley 1988a; 1988c; Stewart et al. 2011), however, some non-breeding adults might be resident in Tasmanian waters (Stanley 1978; Hughes et al. 2011).

Spawning is concentrated in the northern parts of the species’ range in summer and early autumn, when the southward flowing East Australian Current and other associated currents are strongest (Nicholls 1973; Stanley and Malcolm 1977; Stewart et al. 2011), enabling transport of pelagic larvae to southern nursery areas. Spawning in *A. trutta* is believed to be concentrated in a relatively small area between Lakes Entrance, Victoria (37°51’S 148°01’E) and Bermagui, New South Wales (36°27’S 150°04’E) with some evidence that spawning time varies geographically (Stanley and Malcolm 1977; Stewart et al. 2011). Spawning is also known from New Zealand waters (Webb 1973a).

Gonad maturation in both sexes commences in October and continues through to March, although peak gonad maturation is in January (Stanley and Malcolm 1977; Stewart et al. 2011). Ovarian development and maturation is not recorded from Tasmanian fish, or most females from Victoria (Stanley and Malcolm 1977; Stewart et al. 2011). The released eggs of *A. trutta* are pelagic and 0.8 - 1.0 mm in diameter with a single oil globule (Bruce et al. 1998). Spermatogenesis take place in larger males in New South Wales and some parts of Victoria (Stanley and Malcolm 1977; Stewart et al. 2011). Fertilisation is external and larvae hatch some 40 hours after fertilisation (Stanley and Malcolm 1977) and are pelagic (Neira et al. 1997). The pelagic larval duration is unknown, but is probably around four to six months (Nicholls 1973; Robertson 1982; Stewart et al. 2011).
Larvae of *A. trutta* are transported south by the East Australian Current and post-settlement and first-year individuals are most abundant in shallow (nursery) areas in Victorian and Tasmanian waters, where they may be broadly sympatric with some post-settlement *A. truttaceus* (Stanley and Malcolm 1977; Stanley 1978; Figure 1.3). Through the juvenile life stage, fish gradually move northward through Victorian waters and into the adult distribution in New South Wales (Stanley 1978). Tagging studies indicate that adult fish make multi-directional movements, including into northern New South Wales (Stanley 1978; Stewart *et al.* 2011).

1.1.2 *Arripis truttaceus* (Cuvier 1829)  
**Western Australian Salmon**

1.1.2.1 **General Biology and Ecology**

*Arripis truttaceus* occurs from north of Perth on the west coast of Australia to the western parts of Victoria and Tasmania (Gomon *et al.* 1994; Hutchins and Swainston 1999; Figure 1.2). As with *A. trutta*, adults form large schools along exposed beaches and rocky reefs (Cappo *et al.* 2000), possibly with some oceanic movements (Malcolm 1960). Juvenile *A. truttaceus* utilise shallow bays and estuaries as nursery areas (Malcolm 1960; Lenanton 1982; McCormack 2000). *Arripis truttaceus* is larger than *A. trutta*, reaching at least 96 cm in length and weighing in excess of 9 kg (Hutchins and Swainston 1999).

Adult *A. truttaceus* are piscivorous, feeding mainly on pelagic clupeoid species (Hoedt and Dimmlitch 1994). Sub-adult fish in Victorian waters eat mostly pilchard *Sardinops*
neopilchardus, anchovy Engraulis australis and sandy sprat Hyperlophus vittatus, but the diet specialises slightly with age and, as fish form larger schools comprising sub-adult fish, they generally follow aggregations of pilchard (Hoedt and Dimmlich 1994). As juveniles, A. truttaceus feed on small benthic fishes, crustaceans and even the seagrass Zostera (Thompson 1957; Robertson 1982), with feeding success enhanced by the formation of small schools (Foster et al. 2001). The predators of A. truttaceus are likely to be the same as those described for A. trutta. Parasites, with potentially deleterious effects, are known from A. truttaceus (Catalano and Hutson 2010; Catalano et al. 2010;2011).

Similar to A. trutta, a commercial fishery targeting A. truttaceus has existed for decades in southern Australia, landing some one million fish per year (Smith and Brown 2007a; DPI 2008; Fowler et al. 2008; TPC 2009). This species is also a prized recreational angling species, with just under a million fish caught annually (Henry and Lyle 2003). As with its eastern congener, falling catch rates dating back to the 1950s led to a plethora of fisheries related research (Fairbridge 1950; Malcolm 1959;1960; Nicholls 1969;1971;1973; Stanley 1978; Kirkwood and Walker 1984; Stanley 1986;1988b;1988c). This has resulted in one of the longest research histories for an Australian marine fish (Cappo et al. 2000).

1.1.2.2 Growth and Age

Arripis truttaceus is thought to grow faster than A. trutta (Fairbridge 1950; Nicholls 1971). Nicholls (1973) found that fish remaining in South Australia do not reach the mean length of most other four year old fish until they are five years old and reported
that growth rates in *A. truttaceus* decrease with increasing distance from the south-west spawning grounds. Additionally, these fish do not show gonad maturation (see below). Age at maturity was estimated at around 4 years (Nicholls 1973), however there is some uncertainty associated with this estimate because it was based on scale annuli, which may not reflect annual growth increments as accurately as otoliths (see Egglestone 1975; Stanley 1980), and the interpretation of the data was often confounded by the fact that *A. truttaceus* and *A. trutta* were considered to be a single species at the time.

### 1.1.2.3 Life History

The life history of *A. truttaceus* is similar to that of *A. trutta* (Figure 1.3). Between July and January, adult *A. truttaceus* concentrate around the Hopetoun and Esperance (~33°49'S 121°52'E) area of Western Australia with quiescent gonads (Malcolm 1960). Adult aggregations are highly variable in size (e.g. Smith and Brown 2007a). Tagging studies indicate considerable movement of adult fish with no evidence for site-fidelity or fish returning to nursery areas (Stanley 1986; 1988b; 1988c). Influences such as the distribution of prey (Hoedt and Dimmlich 1994) and oceanic conditions (Lenanton *et al.* 1991) are thought responsible for these movements.

In contrast to *A. trutta*, spawning in *A. truttaceus* occurs around the south-western corner of Australia in late autumn and early winter when the eastward flowing Leeuwin Current is strongest (Caputi *et al.* 1996), and eastward-directed winds are dominant (Petrusevics and Bye 1995), enabling transport of pelagic larvae to the south-east.
Gonad maturation in *A. truttaceus* commences in February and continues through June, when mature fish are found between Bremer Bay (34°22′S 119°22′E) and Perth in Western Australia (Malcolm 1960). Peak gonad maturation and post-spawning degeneration occurs in April and May, at which time the individuals concentrate in the south-west, from Cape Leeuwin (33°22′S 115°08′E) to Busselton (33°44′S 115°17′E). Malcolm (1960) argued that these observations, along with the fact that this is the only region in which fish in immediate post-spawning condition are landed, is evidence that spawning is centred in the region from Cape Leeuwin to Busselton. During the spawning phase, few adult fish are encountered further east (e.g. Esperance).

Ovarian development and maturation is completely absent in South Australian individuals, and using indirect evidence, Malcolm (1960) extended this status to females from Victoria and Tasmania. Spermatogenesis takes place in larger males in South Australia in phase with individuals in Western Australian, however there is no evidence that milt is ever released and testicular resorption is thought to occur (Malcolm 1960). Although the gametes and larvae of *A. truttaceus* are not specifically known, it is likely that they are very similar to those of *A. trutta* (see above). *Arripis truttaceus* larvae are believed to be transported away from the south-west spawning area toward the south-east by water flow because larval distribution appears to be positively correlated with both mean sea level along the south coast (as an indicator of the Southern Oscillation Index and hence, the Leeuwin Current) and the duration of easterly-directed winds (Lenanton *et al.* 1991; Petrusevics and Bye 1995). Pelagic larval duration is unknown, but is probably around four to six months (Nicholls 1973; Petrusevics and Bye 1995).
Post-settlement and first-year *A. truttaceus* occur along the entire southern coast of Australia but are most abundant in Victorian and Tasmanian waters, where they are broadly sympatric with post-settlement *A. trutta* (Malcolm 1960; Stanley and Malcolm 1977; Figure 1.3). As juveniles and sub-adults, they move away from south-easterly nursery areas to form larger schools (Robertson 1982; Hoedt and Dimmlich 1994), which make large multi-directional movements across their sub-adult range following schools of baitfish over three to four years. Following this immature growth stage, fish undertake a unidirectional pre-spawning/maturation migration to the adult distribution range (Figure 1.3). *Arripis truttaceus* may well have the greatest documented movement from juvenile to adult of any marine fish (Gillanders et al. 2003).

1.1.3 *Arripis xylabion* (Paulin 1993) Giant Kahawai

1.1.3.1 General Biology and Ecology

*Arripis xylabion* has only recently been recognised (Paulin 1993) and, as a consequence of this and its isolated distribution (see below), little of its biology is known. *Arripis xylabion* is known from Lord Howe Island, Norfolk Island and the Kermadec Islands, with unconfirmed reports from northern New Zealand waters (Paulin 1993; Figure 1.2). It is a coastal pelagic schooling species, whose habits likely resemble those of *A. trutta* (Ian Kerr, Lord Howe Island Marine Park, pers. comm.). *Arripis xylabion* reaches at least 85 cm in length (Paulin 1993) and weighs up to 7 kg (Ian Kerr, pers. comm.). Since the other members of the Arripidae are generally similar to each other in their biology, it is likely that *A. xylabion* will conform to most of the biological attributes of the family, although the movements and dispersal patterns are difficult to predict.
Chapter 1. General Introduction

1.1.4 *Arrpis georgianus* (Valenciennes 1831) Australian Herring

1.1.4.1 General Biology and Ecology

The distribution of *A. georgianus* is broadly sympatric with *A. truttaceus* in most of its range, although it extends further north along the west coast of Australia to around Shark Bay (26°00'S 113°35'E) (Hutchins 1994; Fairclough *et al.* 2000a; Figure 1.2). The accuracy of the identification of specimens reported between eastern Victoria and Sydney is disputed (Hutchins and Swainston 1999; Gomon *et al.* 1994). *Arrpis georgianus* are near-shore pelagic fish, which utilise estuaries more than other members of the family (Hutchins and Swainston 1999; Potter and Hyndes 1999). *Arrpis georgianus* is the smallest member of the family, reaching a maximum size of 41 cm and weighing up to 1 kg (Hutchins and Swainston 1999).

The diet of adult *A. georgianus* is more opportunistic than *A. truttaceus* because, in addition to feeding on fish such as blue sprat *Spratelloides robustus* and sandy sprat *Hyperlophus vittatus*, their diet also commonly includes bivalves, amphipods, shrimps and crabs, as well as insects washed from terrestrial sources (Tregonning *et al.* 1995; pers. obs.). Fish in the 0+ age group rely on amphipods and other crustaceans and bivalves associated with detached macrophytes in the surf zone (Lenanton 1982). As per *A. trutta* and *A. truttaceus*, parasites, with potentially deleterious effects, are known from *A. georgianus* (Catalano and Hutson 2010; Catalano *et al.* 2010;2011).

A commercial fishery targeting *A. georgianus*, which lands 3 – 7 million fish per year, has existed for many decades in southern Australia (Tregonning *et al.* 1995; Smith and
Brown 2007b). This species is also an extremely popular angling species that is considered good eating (Hutchins and Swainston 1999). In fact, *A. georgianus* is the most heavily recreationally fished species in Australia, with around 7 million individuals taken annually (Henry and Lyle 2003). Although fisheries managers have long recognised the importance of research into this species (e.g. Lenanton and Hall 1976; Lenanton 1978), it was not until after catch rates started falling in the early 1990s that most studies were initiated (e.g. Tregonning *et al.* 1995; Fairclough *et al.* 2000a; 2000b; Ayvazian *et al.* 2004).

1.1.4.2  **Growth and Age**

Consistent with the pattern seen in *A. truttaceus*, *A. georgianus* grow more slowly at the eastern end of their distribution compared to the west (Fairclough *et al.* 2000b), however fish do not show gonad maturation until reaching the south-west (see below). On average, female *A. georgianus* have higher growth rates, maximum size, maximum age, and age at maturity compared to males (Lenanton 1978; Fairclough *et al.* 2000a; 2000b). Otolith studies show that *A. georgianus* live up to nine years, and reach sexual maturity by the end of their second year (Fairclough *et al.* 2000b).

1.1.4.3  **Life History**

Tagging studies indicate considerable movement of adults of *A. georgianus*, with no evidence for site-fidelity or for fish returning eastward to nursery areas (see Tregonning *et al.* 1995; Ayvazian *et al.* 2004). Non-spawning adults of *A. georgianus* form schools
along the lower west and south coasts of Australia (Tregonning et al. 1995; Smith and Brown 2007b).

Like the sympatric *A. truttaceus*, spawning in *A. georgianus* coincides with the eastward flowing Leeuwin Current (Caputi et al. 1996) and other eastward-directed winds (Petrusevics and Bye 1995), enabling transport of pelagic larvae to south-eastern nursery areas (Lenanton et al. 1991; Fairclough et al. 2000b; Ayvazian et al. 2004). Gonadosomatic indices for *A. georgianus* in the south-west of Australia show that peak maturation in males occurs around April/May and in females about a month later (Lenanton 1978; Fairclough et al. 2000a). As the spawning season commences, ripe adults merge into large schools (Tregonning et al. 1995; Fairclough et al. 2000a) throughout much, but not all, of their distribution in south-western Australia (between Bremer Bay and the Abrolhos Islands (28°43′S 113°46′E; Fairclough et al. 2000a; Figure 1.3). There is also some evidence that spawning might take place further north of this area (Shark Bay; Fairclough et al. 2000a), but testicular and ovarian development and maturation is completely absent in *A. georgianus* east of Esperance (Lenanton 1978; Fairclough et al. 2000a; 2000b), and spawning does not occur in estuarine-dwelling individuals of *A. georgianus* (S. Hoeksema, Murdoch University, pers. comm.). The released eggs of *A. georgianus* are pelagic and 0.4 - 0.6 mm in diameter (Lenanton 1978). Larvae are similar to those of other members of the family (Neira et al. 1997), and the pelagic larval duration is unknown, but is probably around two to four months (Lenanton et al. 1991; Fairclough et al. 2000b; Ayvazian et al. 2004).

Juvenile *A. georgianus* occur along the entire southern coast of Australia but are most abundant in seagrass beds in South Australian and Victorian waters (Tregonning et al.
1995). While there is some anecdotal evidence for greater local retention of larvae in this species than for others in the family (Tregonning et al. 1995; Fairclough et al. 2000b), it was not supported by genetic data (Ayvazian et al. 2004). Fairclough et al. (2000b) demonstrated a gradual movement of growing fish from nursery areas in the south-east of Australia toward the south-west and predicted that, via this gradual migration, fish enter the adult distribution in their second year of life.

1.2 Previous Genetic Studies on *Arripis*

Only two previous studies have investigated the population genetics of one or more species of *Arripis* and only one of these has been published. Since they are highly relevant to the present research, details of these studies are considered below.

The first study was the PhD research of MacDonald (1980), which focussed on an assessment of the population structure of *A. truttaceus*, but also provided preliminary information about the phylogenetic relationships among *A. trutta*, *A. truttaceus* and *A. georgianus* (*A. xylabion* was not described at that time). Reflecting the era in which it was conducted, MacDonald’s work was based on allozyme markers. Apart from an abstract for a conference (MacDonald 1983), this work was never published. There were three main findings from MacDonald’s (1980) work.

Firstly, the levels of allozyme polymorphism in each of the three species of *Arripis* were generally low (Table 1.1) relative to those in most other marine teleosts (Selander 1976; Ward et al. 1994). MacDonald (1980) argued that the low levels of allozyme diversity in *A. truttaceus*, in particular, were indicative of a population bottleneck.
Secondly, MacDonald (1980) found no evidence of population subdivision in *A. truttaceus* based on samples from nine sites on the south coast between Port Phillip Bay (Melbourne, Victoria) and around Albany (Western Australia), which is consistent with the highly mobile life-style of this species (see Section 1.1.2.3). However, the test had limited power, primarily because it was based on the analysis of the patterns of variation at a single allozyme locus (*EST*), but also because no populations west of Albany were included. MacDonald (1980) did not include population level sampling for *A. trutta* or *A. georgianus*.

Thirdly, in a simple phylogenetic reconstruction, *A. trutta* and *A. truttaceus* were determined to be sister-species and tentatively estimated to have diverged some 0.8 to 1.14 Mya, and were linked to *A. georgianus* from an earlier split, around 5.8 to 6.7 Mya (MacDonald 1980).

The second study was an allozyme-based assessment of the population structure of *A. georgianus* that was conducted by Ayvazian *et al.* (2004), as part of a multi-disciplinary stock assessment. Reflecting the results of MacDonald (1980) for *A. truttaceus*, Ayvazian *et al.* (2004) also found a very low level of allozyme polymorphism in *A. georgianus* and no evidence of population subdivision, based on a test with very limited power (Table 1.1). Specifically, although the test was based on data from 23 allozyme loci, in each case the frequency of the most common allele exceeded 0.95. There is clearly a need to find more polymorphic markers to improve the power of the analyses. Thus far, no studies have used molecular markers to investigate the amount and distribution of genetic variation in any species of *Arripis*. However, the nucleotide sequence of a portion of the *COI* gene of small number of individuals of each *Arripis*
species has been determined as part of a taxonomic barcoding exercise (Fish Barcode of Life Initiative; Ward et al. 2005).

**Table 1.1** Allozyme polymorphism among \( n \) individuals of three species of *Arripis* reported by MacDonald (1980) and in *A. georgianus* by Ayvazian et al. (2004; data indicated by asterisk). The proportion of polymorphic loci was calculated as the number of loci with \( \geq 2 \) alleles. The number of polymorphic loci for which the frequency of the common allele was \( \leq 0.95 \) is also provided. Nei’s (1987) average expected heterozygosity \( (H_E) \) was recalculated from the original data.

<table>
<thead>
<tr>
<th></th>
<th><em>A. trutta</em></th>
<th><em>A. truttaceus</em></th>
<th><em>A. georgianus</em></th>
<th><em>A. georgianus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>103</td>
<td>580</td>
<td>20</td>
<td>646</td>
</tr>
<tr>
<td>Polymorphic loci (number assayed)</td>
<td>0.37 (27)</td>
<td>0.22 (27)</td>
<td>0.23 (26)</td>
<td>0.26 (23)</td>
</tr>
<tr>
<td>Number loci ( \leq 0.95 )</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>( H_E )</td>
<td>0.06</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

### 1.3 Approach

The present study builds upon the previous genetic studies of *Arripis* species (i.e. MacDonald 1980; Ayvazian et al. 2004). It represents a significant step forward from these studies for the following reasons: (i) it jointly considers all four extant species of *Arripis*, although it focuses on *A. trutta* and *A. truttaceus* due to their putative geminate status; (ii) it takes advantage of advances in molecular techniques to directly sample variation at the level of the DNA molecule and so provides a greater amount of information than was possible from the allozyme markers; and (iii) it takes advantage of significant advances in approaches to phylogenetic and phylogeographic research.
Chapter 1. General Introduction

1.3.1 Genetic Markers

The importance of using multiple independent genetic markers to address population and evolutionary questions is widely recognised. This is because different markers have different attributes (e.g. inheritance and mutation rates), and concordance (or not) between markers can provide insights into evolutionary history that cannot be inferred from a single gene (Avise 2009). The present research is based upon two types of genetic markers; nucleotide sequence in mitochondrial DNA and length variation in the alleles of nuclear introns. The reasons for the selection of these markers are explained below.

1.3.1.1 Mitochondrial DNA

Mitochondrial DNA (mtDNA) is a small, circular DNA molecule in the mitochondrion of cells, and has been the mainstay of evolutionary genetics for decades (Brown et al. 1979; Avise 2004). Although debate continues to assess the strengths and weaknesses of the marker (e.g. Zink and Barrowclough 2008; Galtier et al. 2009), mtDNA is generally considered to be an ideal marker for resolving evolutionary questions (such as population subdivision and phylogeography) mainly due to its absence of recombination and uniparental (maternal) inheritance (e.g. Burridge 2000b; Birky 2001; Ward et al. 2005; Percides and Coelho 2006; Mateus et al. 2011). The mtDNA molecule comprises a number of different regions, some of which have different substitution rates and therefore provide the potential to examine processes over different timescales (Ballard and Whitlock 2004). However, the different mtDNA regions are all part of the same molecule and therefore rigidly linked, violating any assumption of independence
(Ballard and Whitlock 2004). Another limitation associated with the use of mtDNA is that such markers allow only the reconstruction of maternal lineages (Avise et al. 1987).

1.3.1.2 Nuclear DNA

Many of the limitations associated with using mtDNA markers to address evolutionary questions can be lessened by also incorporating nuclear DNA (nDNA) markers into the approach (e.g. Ballard and Whitlock 2004; Zink and Barrowclough 2008). This is advantageous for a number of reasons, including that the nDNA markers reflect both male and female contributions to the genetic composition of the population because they have biparental inheritance. Nuclear DNA markers also provide ample scope for repeated independent testing of a hypothesis or outcome because the nuclear genome contains a large number of unlinked loci and all nuclear genes are inherited independently of the mitochondrial genome (Avise 2004).

Microsatellite loci are probably the most commonly used type of nDNA marker in investigations into the population genetics and recent evolutionary history of single species or of closely related species. However, while many of the properties of microsatellite markers make them well suited for this role (reviewed in Li et al. 2002; Ellegren 2004), such use is associated with several drawbacks, including the fact that these markers typically have a narrow taxonomic range and the development of customised markers is relatively expensive and/or labour intensive and requires specialist knowledge and equipment (see Jarne and Lagoda 1996). These problems can be a major issue for studies like the present, which deal with species of low commercial and/or conservation value in a competitive funding environment. Consequently, it
would be a considerable advantage to find markers that exhibit many of the advantageous qualities of microsatellite markers (e.g. co-dominant markers with relatively high amounts of fragment length variation that can be ascribed to a particular locus and easily and unambiguously scored; Li et al. 2002), but which can be amplified using universal primers (and thereby avoid the time and expense of the developmental work of microsatellites).

Length variations in exon-primed, intron-crossing (EPIC) markers offer one such possibility. This technique uses ‘universal’ primers to anneal to a part of the exon of a gene (which should be relatively conserved) and thereby allow for the amplification of the adjacent part of the adjacent intron (which is not translated and therefore potentially more variable) (Palumbi and Baker 1994). The technique was initially used to amplify alleles that were then sequenced (e.g. Lessa 1992; Hare 2001; Atarhouch et al. 2003; Lavoué et al. 2003; Burridge et al. 2006; Morrison et al. 2006; Cooke and Beheregaray 2007), however, more recently, it has been used to examine length variation in the amplified intron sequence (e.g. Berrebi et al. 2005; Atarhouch et al. 2007; Fauvelot et al. 2007; Rolland et al. 2007). However, the number of studies that have used fragment length polymorphism in EPIC markers to address population genetic and/or evolutionary questions is very limited (but see above), and the full potential of this type of marker in these applications is not yet known. Some of the potential limitations associated with the use of EPIC markers in general and length polymorphism in these markers in particular have already been identified, although most of these are not specific to this type of marker (e.g. paralogy; Creer 2007). One of the main issues is the possibility of non-neutrality, given that these markers are a part of a gene and tightly linked to exons, although the empirical data typically do not bear this out (see Friesen
Chapter 1. General Introduction

2000; Creer 2007). Another issue is that the underlying mutational model for these markers is unknown (Friesen 2000; Lynch 2002; Creer 2007). In this regard, it is worth noting that even in the case of microsatellite markers, which are widely used in population and evolutionary studies and the properties of which are relatively well known, the mutational model is not fully understood and likely varies among species and loci (Estoup et al. 2002). Problems stemming from uncertainty about the mutational model of a class of marker can often be circumvented by repeating analyses using contrasting models (e.g. see Cornuet and Luikart 1996).

1.3.2 Analytical Methods

The primary goal of this research was to use genetic data to elucidate aspects of the origins and evolutionary histories of *Arripis* species (see below). The main questions being addressed fall within the fields of population genetics, phylogenetics and phylogeography. Each of these fields has undergone significant advances in recent times due to a combination of increased statistical rigor, associated with a capacity to generate relatively large amounts of genetic data and increased computing power, and significant conceptual gains (e.g. Excoffier et al. 2005; Beheregaray 2008; Avise 2009; Edwards 2009; Knowles 2009). The main advance in each of these fields is probably the ability to develop customised hypotheses (model-based perspective) and test these hypotheses with a relatively high degree of precision (see Knowles 2009; Beaumont et al. 2010). Having said that, there are still limitations in the application of genetic data to ecological and evolutionary questions. For example, the presence of inherently low polymorphism across a species genome still limits the strength of many evolutionary inferences (e.g. Ryman et al. 2006; Palm et al. 2009), although this is less problematic
than it once was (e.g. Brito and Edwards 2008; Edwards 2009; Knowles 2009; Heled and Drummond 2010). Furthermore, phylogenetic reconstructions for so-called ‘rapid radiations’ remain challenging, although this is being increasingly addressed (e.g. Huelsenbeck et al. 2002; Huson and Bryant 2006; Moulton and Huber 2009; Willerslev et al. 2009; Bloomquist et al. 2010; Heled and Drummond 2010).

1.4 Aim and Rationale of Thesis

This thesis examines the distribution of genetic variation within and between species in the family Arripidae and uses the resultant information to test the roles of paleogeographic and paleoclimatic conditions in shaping the evolutionary histories of these species. It focuses on two species, *Arripis trutta* (Eastern Australian Salmon) and *A. truttaceus* (Western Australian Salmon) in particular, as these represent putative geminate species. It also includes information on *A. georgianus* (Australian Herring), which has a similar (west coast) distribution to *A. truttaceus*, and on *A. xylabion* (Giant Kahawai) from Lord Howe Island, which on the basis of morphological evidence (see species account) appears to be very closely related to the two species of Australian Salmon. Information on each of the latter two species, although significant in its own right, is largely used to help place the results for *A. trutta* and *A. truttaceus* into context.

This research is significant for the following reasons. Firstly, it extends existing information about the population structure, taxonomy, relationships, origins and evolutionary history of members of the family Arripidae. Importantly, it is the first genetic study to include all four species, and both mtDNA and nDNA markers. Secondly, it provides one of the first investigations into the relationships (and status) of
putative geminates on the east and west coasts of Australia. Thirdly, it provides one of the first tests of the prediction that species which reside to the west of the Bass Strait may have been more severely impacted by glacial cycles than those that reside to the east. Therefore, this research represents an important early step in the development of knowledge about the origins and evolutionary histories of the biota of the Southern Australian bioregion.

1.5 Outline of Thesis

Chapter 1: General Introduction. The aim of the General Introduction (the present chapter) was to introduce the conceptual framework, provide general background information, state the significance and overall aim of the study and outline the aims of each chapter.

Chapter 2: General Material and Methods. The aim of Chapter 2 was to provide the general details of the methodology. It describes the sampling regime and methodology, as well as the genetic methods and includes an assessment of the reliability of the EPIC markers.

Chapter 3: Population Genetic Structure. The aim of Chapter 3 was to use genetic (intron length polymorphism) data to investigate the population genetic structures of *A. trutta*, *A. truttaceus* and *A. georgianus* effectively over their entire geographic distributions (latter two species) or the entire Australian mainland distributions (former species). This was necessary because an understanding of the underlying genetic
structures of each of these species was essential for interpreting the results of the evolutionary aspects of this study.

Chapter 4: A Molecular Approach to the Taxonomy of Australian Salmon. The aim of Chapter 4 was to investigate whether the current, morphologically-based taxonomy of Australian Salmon is supported by DNA data. It was necessary because it is important to confirm that these taxa represent ‘true’ species, as opposed to conspecific populations that are exchanging genes or have recently done so.

Chapter 5: Molecular Phylogenetics, Speciation and Biogeography. The aim of Chapter 5 was to generate information on the relationships among the species of Australian Salmon and use the results to develop an explanation for the speciation and current biogeography of this group. It focussed on testing the assumption that the east/west species of *A. trutta* and *A. truttaceus* are true geminate species, and applied molecular dating to estimate divergence times between species of Australian Salmon.

Chapter 6: Genetic Diversity and Recent Evolutionary History. The aim of Chapter 6 was to assess and compare the levels and patterns of genetic diversity in the effectively geminate species *A. trutta* and *A. truttaceus* and use the resultant information to infer and compare aspects of the recent evolutionary histories of these species. The chapter tests the prediction that *A. truttaceus*, which resides to the west of the Bass Strait, has been more severely impacted by glacial cycles than *A. trutta*, which resides to the east.

Chapter 7: General Conclusions. The aim of Chapter 7 was to summarise the main findings of each chapter.
2.0 General Methods

2.1 Sampling Design

Sampling was designed to maximise the potential for detecting genetic heterogeneity both within and between *Arripis* species (Page and Holmes 1998). The design of the sampling regime considered the following. The consequences and limitations of each consideration in the context of each question are discussed in the relevant chapters.

**Taxonomic coverage:** all four extant species of *Arripis* were included in the thesis, although the extent of their inclusion varied due to the specific questions that were being asked (see Aim for each chapter), and access to samples (e.g. *A. xylabion* was not included in an assessment of population structure because all individuals were sampled from a single site; see Chapter 3).

**Spatial coverage:** Samples were obtained from virtually either the entire distribution of the species in Australian waters (*A. trutta*) or the entire distribution of the Australian species (*A. truttaceus, A. georgianus*). This strategy was used to reduce the possibility of missing genetic differentiation if, for example, all individuals had been sampled from a single, more localised breeding unit within a subdivided species (see Page and Holmes 1998). Due to a reliance on others for specimens, spatial sampling of *A. xylabion* was not possible (see below).

**Temporal framework:** In order to minimise the potential effects of temporal genetic variations (e.g. Johnson and Black 1984; Hedgecock 1994; Dannewitz *et al.* 2005; Palm
et al. 2009), sampling for each species was generally completed within the same few months (Table 2.1). However, samples of *A. trutta* from New Zealand and samples of *A. truttaceus* and *A. georgianus* from South Australia proved difficult to source and were taken one to two years after the rest of the collections (Table 2.1). Extensive temporal sampling was not undertaken, partly because the initial analysis revealed no evidence of significant spatial genetic heterogeneity. However, a second sample of *A. georgianus* from Fremantle, Western Australia was collected two years after the first, to provide a very limited test for the presence of temporal genetic heterogeneity (see Section 3.2.1).

**Sample size:** The sample sizes for each species from each sampling location are indicated in Table 2.1. A minimum of 30 individuals was sought for each sampling location (e.g. Ryman et al. 2006; Pruett and Winker 2008). However, as described below, the project was largely reliant upon specimens collected by other individuals and this minimum number was not always obtained.

Samples of 29 – 30 *A. trutta* were collected from three locations throughout most of the geographic range of this species in Australian waters (Table 2.1, Figure 2.1). In addition, two specimens of this species were also obtained from near Christchurch, New Zealand. Samples of 13 – 45 *A. truttaceus* were collected from five locations throughout most of the geographic range of this species, which is confined to Australian waters (Table 2.1, Figure 2.1). Samples of 9 – 44 *A. georgianus* were collected from seven locations throughout most of the geographic range of this species, which is also confined to Australian waters (Table 2.1, Figure 2.1). A single sample of 25 *A. xylabion* was collected from Lord Howe Island (Table 2.1, Figure 2.1).
Table 2.1  Collection details for *Arripis* species including approximate coordinates of sampling locations.  See also Figure 2.1.

<table>
<thead>
<tr>
<th>ID (n)</th>
<th>Site</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Collection Date/s</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN (29)</td>
<td>Port Stephens, NSW</td>
<td>32° 40′ 31″ S</td>
<td>152° 12′ 27″ E</td>
<td>09 May 2007</td>
<td>NSW DPI</td>
</tr>
<tr>
<td>EV (30)</td>
<td>Lakes Entrance, Vic.</td>
<td>37° 52′ 48″ S</td>
<td>147° 59′ 24″ E</td>
<td>31 May 2007</td>
<td>NSW DPI</td>
</tr>
<tr>
<td>ET (30)</td>
<td>Flinders Island, Tas.</td>
<td>39° 59′ 49″ S</td>
<td>148° 02′ 28″ E</td>
<td>April-May 2007</td>
<td>NSW DPI</td>
</tr>
<tr>
<td>EZ (2)</td>
<td>Christchurch, NZ</td>
<td>43° 31′ 48″ S</td>
<td>172° 37′ 13″ E</td>
<td>30 January 2009</td>
<td>Commercial</td>
</tr>
<tr>
<td>WA (30)</td>
<td>Augusta (Hamelin Bay), WA</td>
<td>34° 12′ 25″ S</td>
<td>115° 02′ 08″ E</td>
<td>13 March 2007</td>
<td>Commercial</td>
</tr>
<tr>
<td>WY (15)</td>
<td>Yeagarup Beach, WA</td>
<td>34° 26′ 28″ S</td>
<td>115° 37′ 39″ E</td>
<td>29-31 March 2007</td>
<td>Recreational</td>
</tr>
<tr>
<td>WL (39)</td>
<td>Albany, WA</td>
<td>35° 00′ 35″ S</td>
<td>117° 55′ 31″ E</td>
<td>27 March 2007</td>
<td>Commercial</td>
</tr>
<tr>
<td>WW (13)</td>
<td>Wellstead Estuary, WA</td>
<td>34° 23′ 17″ S</td>
<td>119° 23′ 05″ E</td>
<td>March 2007</td>
<td>Murdoch</td>
</tr>
<tr>
<td>WS (45)</td>
<td>The Coorong, SA</td>
<td>35° 56′ 41″ S</td>
<td>139° 25′ 37″ E</td>
<td>05 July 2008</td>
<td>Commercial</td>
</tr>
<tr>
<td>KL (25)</td>
<td>Lord Howe Island, NSW</td>
<td>31° 33′ 10″ S</td>
<td>159° 04′ 50″ E</td>
<td>Aug. – Nov. 2008</td>
<td>Recreational</td>
</tr>
<tr>
<td>HF (44)</td>
<td>Fremantle, WA</td>
<td>32° 02′ 43″ S</td>
<td>115° 44′ 07″ E</td>
<td>06-09 July 2006</td>
<td>Recreational</td>
</tr>
<tr>
<td>HF08 (41)</td>
<td>Fremantle, WA</td>
<td>32° 02′ 43″ S</td>
<td>115° 44′ 07″ E</td>
<td>24-26 April 2008</td>
<td>Recreational</td>
</tr>
<tr>
<td>HC (39)</td>
<td>Canal Rocks, WA</td>
<td>33° 40′ 10″ S</td>
<td>114° 59′ 40″ E</td>
<td>13-14 July 2006</td>
<td>Recreational</td>
</tr>
<tr>
<td>HI (23)</td>
<td>Iwin Inlet, WA</td>
<td>34° 59′ 30″ S</td>
<td>116° 57′ 30″ E</td>
<td>May-August 2006</td>
<td>Murdoch</td>
</tr>
<tr>
<td>HP (36)</td>
<td>Peaceful Bay, WA</td>
<td>35° 02′ 30″ S</td>
<td>116° 55′ 52″ E</td>
<td>May-July 2006</td>
<td>Commercial</td>
</tr>
<tr>
<td>HO (28)</td>
<td>Albany (Oyster Harbour), WA</td>
<td>35° 03′ 12″ S</td>
<td>117° 53′ 28″ E</td>
<td>May-August 2006</td>
<td>Murdoch</td>
</tr>
<tr>
<td>HW (09)</td>
<td>Wellstead Estuary, WA</td>
<td>34° 23′ 17″ S</td>
<td>119° 23′ 05″ E</td>
<td>May-August 2006</td>
<td>Murdoch</td>
</tr>
<tr>
<td>HS (41)</td>
<td>The Coorong, SA</td>
<td>35° 56′ 41″ S</td>
<td>139° 25′ 37″ E</td>
<td>05 July 2008</td>
<td>Commercial</td>
</tr>
</tbody>
</table>
Figure 2.1  Approximate collection locations for samples of *Arripsis* used in this study: *A. trutta* (yellow), *A. truttaceus* (pink), *A. xylabion* (orange), *A. georgianus* (blue). See Table 2.1 for details.
2.2 Sample Collection, Processing and Storage

The *Arripis* samples were sourced from recreational and commercial fishers and from other research projects (at Murdoch University and at New South Wales Department of Primary Industries) between May 2006 and January 2009 (Table 2.1).

For the westerly distributed species (*A. truttaceus* and *A. georgianus*), whole or part fish were placed on ice or into liquid nitrogen as soon as possible after capture before being transported to the laboratory, where they were stored frozen until processed. Muscle tissue was sampled and stored at -20°C. Since they had to be transported by air, tissue samples from *A. trutta* and *A. xylabion* from New South Wales and Lord Howe Island, respectively, were shredded into very small fragments and placed into a 10% DMSO - 5 M NaCl solution (Proebstel *et al.* 1993; Kilpatrick 2002) at a ratio of approximately 1:50 tissue to solution and stored frozen. Upon arrival at Murdoch University, these tissue samples was transferred to a smaller volume (~ 1.5 ml) of the DMSO solution and stored at -20°C. DNA from each of two individuals of *A. trutta* from New Zealand was mailed to Murdoch University after being extracted by Marie Hale (University of Canterbury, Christchurch) using Qiagen extraction columns following the manufacturer’s protocol. They were mailed in 35 µl of TE Buffer (10 mM Tris-HCL [pH 8.0], 1 mM EDTA) and frozen at -20°C on arrival.

2.3 DNA Extraction

Total genomic DNA was isolated from individual tissue samples using a MasterPure™ extraction kit (Epicentre 2003). The *A. trutta* tissue was soaked in 200 µl of autoclaved
water for approximately 1 minute, to remove much of the NaCl storage solution, and blotted dry before extraction. A 5 mg sample of tissue was homogenised in 300 µl of Tissue and Cell Lysis Solution and Proteinase K (50 µg) and incubated at 65°C for 15 minutes. Buffered RNase A (5 µg) was added to the homogenate, which was then incubated at 37°C for a further 30 minutes. While the homogenate was on ice, 175 µl of MPC Protein Precipitation Reagent was added and the mix was centrifuged for 10 minutes at 10,000 x g. The resultant supernatant was transferred to a new tube and DNA was precipitated by centrifuging for 10 minutes at 10,000 x g in 500 µl of isopropanol. The pellet was washed with two 500 µl aliquots of 75% ethanol. Once dry, the pellet was resuspended in 35 µl of TE Buffer (10 mM Tris-HCL [pH 8.0], 1 mM EDTA) and stored at -4°C. The presence (or otherwise) of DNA was confirmed by electrophoresis on a 2% agarose gel, stained with ethidium bromide or SYBR®Safe.

2.4 Mitochondrial DNA

Analysis of mtDNA variation in the four species of *Arripsis* was based on the nucleotide sequence of a 483 bp portion (positions 556 - 1038 in *Arripsis*) of the cytochrome *b* gene and a 654 bp portion of cytochrome oxidase subunit 1 (*COI*) (positions 52 – 705 in *Arripsis*). These gene regions were selected for analysis because they could be amplified reliably from *Arripsis* (unlike the control region; see Section 2.4.3) and are widely used in similar studies (e.g. Burridge and White 2000; Burridge 2000b; Ward *et al.* 2005; Percides and Coelho 2006). The cytochrome *b* partial gene sequences were amplified using *Arripsis* specific primers, which were developed as a part of the present study using an alignment of unpublished sequence (Masaki Miya, Natural History Museum, Chiba, Japan, pers. comm.) and published sequence for *A. trutta* cytochrome *b* (Doiuchi
and Nakabo 2006). The COI partial gene sequences were amplified using the universal primers of the fish barcoding project FISHBOL (Ward et al. 2005) that were already known to work in all species of Arripis (B. Ward, CSIRO, pers. comm.).

2.4.1 PCR Amplifications

Polymerase chain reaction (PCR) in a GeneAmp 9700 Thermal Cycler (Applied Biosystems) was used to amplify the target regions of the mtDNA, with conditions specific to each gene. The cycling conditions for the cytochrome b region were: (i) an initial denaturation phase of 5 minutes at 94° C; (ii) 30 amplification cycles, with each cycle consisting of 30 seconds of denaturation at 94° C, 30 seconds of annealing at 50° C, 30 seconds of extension at 72° C; and (iii) a final 5 minute extension at 72° C.

Each cytochrome b PCR reaction mixture contained 0.1 mM of each of the dNTPs (Promega), 0.25 U of Taq DNA polymerase (Roche), 5 µl Taq Buffer (100 mM Tris-HCl, 15 mM MgCl₂, 500 mM KCl, pH 8.3; Roche), 0.4 µM of each primer (Atr15065L-CytB (5′-TACTCTCCGATCTTTGCCTTCC-3′) and Atr15613H-CytB (5′-AGGATGAGAATAACGAGAAGTAGAGGAA-3′), and approximately 10 ng of DNA template, adjusted to a final volume of 50 µl with PCR-grade H₂O.

The optimised cycling conditions for the COI region were a modified version of those of Ward et al. (2005) and consisted of: (i) an initial denaturation phase of 2 minutes at 95° C; (ii) 30 amplification cycles, with each cycle consisting of 30 seconds of denaturation at 94° C, 30 seconds of annealing at 54° C, 60 seconds of extension at 72° C; and (iii) a final 10 minute extension at 72° C.
Each COI PCR reaction mixture was the same as described for cytochrome \( b \) except the use of 0.2 µM of each primer (\textit{Fish F1} 5'- TCA ACC AAC CAC AAA GAC ATT GGC AC -3' and \textit{Fish R1} 5'- TAG ACT TCT GGG TGG CCA AAG AAT CA -3').

The presence (or otherwise) of PCR product for each reaction was assessed using electrophoresis on a 2% agarose gel, stained with ethidium bromide or SYBR®Safe.

2.4.2 Sequencing

Prior to sequencing, PCR products were cleaned using a QIAquick® PCR Purification Kit (Qiagen 2006). A mixture of 40 µl of a PCR product and 200 µl of Buffer PBI was centrifuged in a spin column for one minute at >17,000 x g. The product was washed with 0.75 ml of Buffer PE and centrifuged for one minute at >17,000 x g. To elute the DNA, 30 µl of Buffer EB was added, and the mixture was left to stand for one minute and then centrifuged at >17,000 x g for one minute.

Approximately 30 ng of clean PCR product was added to 1 µl of PCR grade water, 1 µl (3.2 pmol) of either the forward or reverse primer and 4 µl of Big Dye dye-terminator cycle sequencing ready reaction kit (Applied Biosystems). The amplification conditions were as follows: (i) an initial denaturation phase of 2 minutes at 96° C; and (ii) 25 amplification cycles, with each cycle consisting of 10 seconds of denaturation at 96° C, 5 seconds of annealing at 50° C and 4 minutes of extension at 60° C.
Chapter 2. General Methods

The amplification product was precipitated by adding 1 µl of 3 M sodium acetate (pH 4.6), 1 µl of 125 mM EDTA and 25 µl of 100% ethanol and allowing to stand for 20 minutes before centrifuging for 30 minutes at >17,000 x g and discarding the supernatant. The precipitate was washed with 125 µl of 70% ethanol, centrifuged for 10 minutes at >17,000 x g and air-dried at room temperature.

Sequences were determined using an automated sequencer (ABI Prism 3730) and ABI Prism DNA sequencing analysis software. Heavy and light strands were sequenced for each individual and a consensus sequence generated using the software GENETOOL Lite version 1 (BioTools Inc.).

2.4.3 Trials involving other mtDNA markers

In addition to cytochrome b and COI, attempts were made to PCR-amplify two other regions in the mtDNA - the 16S rRNA gene region and the control region. The attempts were successful for 16S but preliminary data indicated that this gene had the lowest level of variability among the assayed regions and hence was probably less informative for addressing the questions in this thesis, which are concerned with intra-specific variation and variation among closely related species (see Johns and Avise 1998).

Since it is typically the most variable region of the mtDNA (Avise 2000), the control region was initially expected to be the most useful part of the mtDNA for resolving questions about the population structures and phylogeography of Arripis species. Accordingly, attempts were made to amplify this region from each of A. trutta, A. truttaceus and A. georgianus. However, despite a variety of different approaches, the
control region could not be reliably amplified from these species, possibly due to complications from the presence of a large repeat motif. Full details are provided in Appendix A.1.

2.5 Nuclear DNA

The analysis of nuclear DNA variation in each species of *Arripis* was based on patterns of fragment length variation at two to four polymorphic EPIC (exon-primed intron-crossing) amplified loci, depending on the species. The loci were Aldolase B [intron 4] *AldoB4*; Calmodulin [intron 4] *CaM4*; S7 ribosomal protein gene [intron 2] *S72*; and Creatin kinase *CK7* (Appendix A.2, Chapter 3). These loci were chosen from a pool of 28 potential loci (Appendix A.2) because they amplified and exhibited polymorphism in at least one species of *Arripis* (see Section 2.5.3). Data from an additional eight loci, which were monomorphic, at least within each species, were also used in the assessment of the levels of genetic variation within each species and within the family (Appendix A.2; Chapters 4, 5, 6).

2.5.1 PCR Amplifications

PCR reactions were carried out in a GeneAmp 9700 Thermal Cycler (Applied Biosystems), with: (i) an initial denaturation of 5 minutes at 94°C; (ii) 38 amplification cycles, with each cycle consisting of 30 seconds of denaturation at 94°C, 30 seconds of annealing at optimised temperatures (Table 2.2), 30 seconds of extension at 72°C; and (iii) a final 20 minute extension at 72°C (see Section 2.5.4). The only exceptions were: (i) the locus *CK7* in *A. georgianus*, which was amplified using a touch-down PCR of 57
amplification cycles, with annealing starting at 58°C for 60 seconds and decreasing by 1°C and 1 second each cycle; and (ii) the locus CaM4 in all species, which was amplified using a touch-down PCR of 50 amplification cycles, with annealing starting at 65°C for 60 seconds and decreasing by 0.4°C each cycle (Table 2.2). The PCR reaction mixture contained 0.1 mM of each dNTP (Promega), 0.25 U Taq DNA polymerase (Roche), 1.5 µl Taq Buffer (100 mM Tris-HCl, 15 mM MgCl2, 500 mM KCl, pH 8.3; Roche), 0.04-0.06 µM of each primer (Table 2.2), and approximately 10 ng of DNA template, adjusted to a volume of 15 µl with PCR grade H2O. Forward primers were fluorescently labelled with 6-FAM (CaM4), VIC (CK7) or NED (AldoB4, S72).

2.5.2 Fragment Analysis

PCR products were assayed either singularly or multiplexed for one large and one small locus, using optimised volumes for each locus (Table 2.2). The products were precipitated using 1 µl of 3 M sodium acetate (pH 4.6), 1 µl of 125 mM EDTA and 25 µl of 100% ethanol and allowed to stand at room temperature for 13 minutes before centrifuging for 30 minutes at >17,000 x g and discarding the supernatant. The precipitate was washed with 125 µl of 70% ethanol, centrifuged for 10 minutes at >17,000 x g and air dried at room temperature.

Following resuspension in 15 µl of formamide, each PCR product was loaded into a 96-well plate and fragment length was determined with an ABI 3730 Automated Sequencer (Applied Biosystems) using 0.09 µl of LIZ1200 (Applied Biosystems) size standard per
well. Allele peaks were manually scored to one decimal place using the software GENEMARKER version 1.71 (SoftGenetics Inc.).

To the best of my knowledge, this is the first study to use capillary electrophoresis to detect and score fragment length variation at EPIC amplified intron loci; previously only agarose or denaturing polyacrylamide gels have been used for such. Capillary electrophoresis, and associated genotyping software, was used on the expectation that it would offer a superior resolution to agarose (Giribet and Wheeler 1999; Sánchez-Pérez *et al.* 2006) and is faster and less toxic than polyacrylamide (Sánchez-Pérez *et al.* 2006).
Table 2.2 PCR and plate loading conditions for nuclear intron loci assays in the four species of *Arripis*. In each case, the optimised annealing temperature and primer concentration used in PCR reactions is given, along with the volume of PCR product loaded onto plates for fragment analysis.

<table>
<thead>
<tr>
<th>Locus</th>
<th><em>A. georgianus</em></th>
<th><em>A. trutta</em></th>
<th><em>A. truttaceus</em></th>
<th><em>A. xylabion</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anneal. temp. (°C)</td>
<td>Primer conc. (µM)</td>
<td>Load vol. (µl)</td>
<td>Anneal. temp. (°C)</td>
</tr>
<tr>
<td>AldoB4</td>
<td>57</td>
<td>0.04</td>
<td>0.5</td>
<td>56</td>
</tr>
<tr>
<td>CaM4</td>
<td>65-45*</td>
<td>0.04</td>
<td>1.0</td>
<td>65-45*</td>
</tr>
<tr>
<td>CK7</td>
<td>58-41*</td>
<td>0.04</td>
<td>2.0</td>
<td>53</td>
</tr>
<tr>
<td>S72</td>
<td>64</td>
<td>0.06</td>
<td>2.5</td>
<td>62</td>
</tr>
</tbody>
</table>

* Touch-down PCR – see text.
2.5.3 Primer Selection and Optimisation

As noted earlier, the EPIC primers were selected from a pool of published primers. There were three main phases to this selection process, as described below.

Firstly, 28 sets of published EPIC primers were selected from the literature with preference given to loci with reported length polymorphism, cross-species amplification and reliability (Appendix A.2). Primers were taken from studies including crustaceans, marine and freshwater fishes and cetaceans (see Appendix A.2).

Secondly, preliminary trials showed that 13 of the 28 sets of published primer could be used to amplify a product in representatives of each species of *Arripis* using touchdown PCR conditions and these 13 loci were therefore selected for further testing. These tests trialled various conditions to determine, for each of the 13 loci, the optimum PCR conditions, including: (i) cycle number; (ii) annealing temperature; (iii) extension times; and (iv) primer concentration.

Thirdly, following the screening of ten individuals of each species, four of the 13 loci (*AldoB4, CaM4, S72 and CK7*) were found to be polymorphic in at least one species (Appendix A.2) and were screened for length polymorphism as described above. Some further development of the assays conditions for each of these four loci was required to determine the optimal volume of PCR product to be loaded onto the plates for fragment length analysis (Table 2.2).
2.5.4 Allele Scoring

Three methods were employed to reduce the potential for errors in scoring genotypes at the intron loci.

Firstly, both positive controls (assays for individuals whose genotype was already known) and negative controls (PCR assays with no added DNA template) were included on most fragment analysis plates. These provided a means for detecting, respectively, inconsistencies in allele scoring across assays (see below) and the presence of any foreign DNA in the assays.

Secondly, a long final extension time (20 minutes) was used to PCR-amplify the alleles at all loci. This was done to maximise the potential of incorporating a “poly-A” on all amplified PCR products and thereby make it easier to distinguish between real alleles and those with an added poly-A (see Figure 2.2). NB: Non-templated nucleotide additions by the Taq DNA polymerase (usually a single adenosine to the 3’ end of the fluorescently labelled strand; e.g. Smith et al. 1995; Anon. 2004) are well known in PCR-based studies particularly when short extension times are used (Clark 1988).

Thirdly, all genotypes were scored twice (blindly). Where a discrepancy was detected, the chromatogram was re-examined, and if the score remained ambiguous, the individual was re-assayed by independent PCR and genotyping.
Figure 2.2  Examples of the effect of final extension time (0, 5, 15, 20 minutes) on poly-A addition to the 177 bp allele of *AldoB4* in *Arripis georgianus*. The x-axis represents fragment length in bp. Longer extension time resulted in a greater proportion of alleles exhibiting the poly-A extension as indicated by the increasing height of the second peak. See text for detailed explanation.
Throughout this study, allele scoring at the intron loci was based on a conservative allele binning approach such that all scores within 20 decimals (where 10 decimals = 1 bp) of another for each of the two larger loci, CK7 (maximum allele size = 984 bp) and S72 (maximum allele size = 931 bp), and within 10 decimals for each of the smaller loci, AldoB4 (maximum allele size = 191 bp) and CaM4 (maximum allele size = 504 bp), were pooled into a single allele class (Figures 2.3 to 2.6, which for simplicity present the data for each locus for only a single species, but the details were similar for the other species). Each allele class (bin) was named by the modal integer and bins were consistent for each locus across all species. This is believed to have minimised the chance of scoring PCR or fragment analysis artefacts of the same allele as a different allele (see below). Although it is possible that this method sometimes resulted in the pooling of different alleles into the same allele class, this type of error is likely to have less significant implications for the data interpretation than the reverse, i.e. it will lead to an underestimate of the extent of genetic diversity and population structure rather than an overestimate.

**Figure 2.3**  Frequency histogram of fragment sizes for the locus AldoB4 in all assayed Arripis trutta. Bars and arrows indicate allele bins, labelled with the modal integer (see text).
Figure 2.4  Frequency histogram of fragment sizes for the locus CaM4 in all assayed Arripis georgianus. Bars and arrows indicate allele bins, labelled with the modal integer (see text).

Figure 2.5  Frequency histogram of fragment sizes for the locus CK7 in all assayed Arripis trutta. Bars and arrows indicate allele bins, labelled with the modal integer (see text).
Figure 2.6 Partial frequency histogram of fragment sizes for the locus S72 in all assayed *Arrapis trutta*. Bars and arrows indicate allele bins, labelled with the modal integer (see text). Note that for scaling reasons, the x-axis has been abbreviated at the origin and thus a single occurrence of each of two alleles (816 & 869) is not shown.

 Allele binning, which is common place in studies of fragment length variation at intron loci (e.g. Gomulski *et al.* 1998; Berrebi *et al.* 2005; Hubert *et al.* 2006; Atarhouch *et al.* 2007), was necessary because of variability among repeats of the same individual (Table 2.3). Bins encompassing 20 decimals were used for large loci (CK7 and S72), because repeat assays indicated intra-individual variability may be up to 1.5 bp and the discrepancies would therefore sometimes have been scored as a different allele if allele bins had not been used (Table 2.3). Although repeat assays at the smaller loci (*AldoB4* and *CaM4*) indicated that mis-scoring due to intra-individual variability was unlikely, there was still variability – less than 0.5 bp (Table 2.3). Consequently, alleles within 10 decimals were binned in these small loci. Variability among repeat assays may be due to PCR and/or electrophoretic factors (see Appendix A.3).
Binning of fragment length alleles at EPIC amplified intron loci is usually based on a subjective approach with alleles placed into arbitrary, fixed size classes, usually five (e.g. Hubert et al. 2006; Atarhouch et al. 2007), ten (e.g. Berrebi et al. 2005) or more (e.g. Gomulski et al. 1998) bases in size. Sometimes bins are required due to the inaccuracies of ‘by eye’ scoring of gels (e.g. Atarhouch et al. 2007), whereas sometimes binning is used at the post-scoring stage to address ambiguities in the data (e.g. Hubert et al. 2006). The method of allele binning in this study, which is based on gaps in the frequency distribution of the scoring of the raw alleles, is thought to be more meaningful than the arbitrary allele size classes used in previous studies (see also Appendix A.3).

Table 2.3  (over page) Results of tests of variability in raw scoring (i.e. without binning) of intron loci allele lengths in repeat assays in Arripis. The tests were based on (two) repeat assays for n individuals. Some tests were done using different PCR reactions, with the products loaded onto the same or different fragment analysis plates, whereas other tests were done using the same PCR reaction, with the product loaded onto the same or different plates. Mean, standard error and maximum errors (given as integer) are the difference between scores (in bp) for the two repeats. The percentage of repeats where a different allele would have been scored if allele bins had not been used is given as ‘% different’. For AldoB4, CK7 and S72, data are pooled across all alleles for all three Australian species. In CaM4, only data for A. georgianus were used (because this was the only species for which this locus was polymorphic).
### Chapter 2. General Methods

<table>
<thead>
<tr>
<th>Allele</th>
<th>Different PCR</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Same plate</td>
<td>Different plate</td>
<td>Same plate</td>
<td>Different plate</td>
<td>Same plate</td>
<td>Different plate</td>
<td>Same plate</td>
<td>Different plate</td>
<td>Same plate</td>
<td>Different plate</td>
<td>Same plate</td>
<td>Different plate</td>
</tr>
<tr>
<td>AldoB4</td>
<td>1 12 0.03 ± 0.02 0.2 0.0</td>
<td>2 12 0.05 ± 0.02 0.2 0.0 0.64 0.05 0.05 0.06 0.01 0.01 0.01 0.01</td>
<td>1 12 0.1 0.0 0.2</td>
<td>2 39 0.09 ± 0.02 0.2</td>
<td>1 39 0.03 0.0</td>
<td>2 39 0.02 ± 0.02 0.0 0.0 0.0</td>
<td>1 64 0.06 0.0</td>
<td>2 64 0.05 0.0</td>
<td>1 6 0.06 0.0 0.0 0.0</td>
<td>2 6 0.01 0.0 0.0 0.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| CaM4   | 6 0.03 ± 0.03 0.2 | 6 0.1 ± 0.03 0.2 | 2 0.11 ± 0.02 0.2 | 2 0.04 ± 0.02 0.0 0.0 0.0 | 0.07 0.07 0.1 0.1 | 0.04 0.02 0.00 0.00 | 0.0 | 0.0 | 0.0 0.0 0.0 | 0.0 0.0 0.0 0.0 |

| CK7    | 49 0.27 ± 0.02 0.9 | 49 0.32 ± 0.03 0.9 | 75 0.41 ± 0.04 0.9 | 75 0.47 ± 0.04 0.9 | 0.24 0.27 0.29 0.29 | 0.02 0.02 0.08 0.08 | 0.0 | 0.0 | 0.0 0.0 0.0 | 0.0 0.0 0.0 0.0 |

| S72    | 45 0.15 ± 0.01 0.5 | 45 0.22 ± 0.02 0.5 | 26 0.23 ± 0.03 0.5 | 26 0.29 ± 0.04 0.5 | 0.22 0.26 0.21 0.23 | 0.02 0.03 0.04 0.04 | 0.0 | 0.0 | 0.0 0.0 0.0 | 0.0 0.0 0.0 0.0 |

|        | 10 0.5 0.9 | 7.8 0.1 0.6 | 26.9 0.5 0.8 | 26.9 0.9 0.6 | 18.8 1.6 9.5 | 18.8 1.6 9.5 | 0.0 | 0.0 | 0.0 0.0 0.0 | 0.0 0.0 0.0 0.0 |

### 2.6 Nuclear DNA Sequences

An attempt was made to determine the nucleotide sequence of some EPIC markers in order to provide additional sequence data (i.e. in addition to the mtDNA data), as has been successfully conducted in other similar studies (e.g. Lessa 1992; Hare et al. 2002; Lavoué et al. 2003; Teske et al. 2004; Morrison et al. 2006). This attempt was based on simultaneous sequencing of both of the alleles at a diploid locus and recognising...
heterozygotes as two peaks of nearly equal intensity at the same base position in the chromatograms (Creer 2007). These sites may then be scored using standard IUPAC ambiguity codes. When successful, this method avoids the time and costs associated with the cloning and multiple sequencing required to independently ascertain the sequence of both alleles (Lavoué et al. 2003; Teske et al. 2004; Morrison et al. 2006), although the method has some potential disadvantages (e.g. Hare 2001; Creer 2007).

The most significant problem arises when heterozygotes contain alleles of different length. Insertions and deletions (indels) within alleles will result in unreadable chromatograms as the two superimposed sequences will be out of phase and distinguishing between them becomes impossible (Mallarino et al. 2005; Creer 2007).

The intron loci trialled for direct sequencing were those loci that could be amplified in *Arripis* and for which no length variants were detected (*AldoB1, Am2B-1, GnRH3-1, GnRH3-2, GnRH3-3, Gpd2, Mlc-2-c, S71; see Appendix A.2*). PCR was first used to amplify DNA product with conditions previously optimised for each gene (data not shown). Of the eight loci tested, four gave indecipherable chromatograms, which were probably caused by numerous indels, the sum of which maintained length monomorphism among alleles, but still caused the sequences to be out of phase. The other four loci (*AldoB1, GnRH3-1, GnRH3-2, GnRH3-3*) were then amplified using the same touch-down PCR conditions described for *CaM4* in Section 2.5.1. Each PCR reaction mixture was the same as that described in Section 2.4.1 (see Appendix A.2 for primer sequences), and cleaning, sequencing and alignment was as described in Section 2.4.2. However, the amount of sequence polymorphism at each of the four loci was extremely low and provided little utility for the purposes of this thesis.
3.0 Population Genetic Structure

3.1 Introduction

As outlined in Chapter 1, the primary goal of this thesis was to examine the distribution of genetic variation within and between *Arripis* species and to use the resultant information to infer aspects of the origins and evolutionary histories of these species. However, before using this information to address evolutionary questions, it is important to have an understanding of the population genetic structure of each species and associated patterns of contemporary gene flow. This is important because the extent and patterns of contemporary gene flow will have a major bearing on the distribution of genetic variation within each species. As a part of this, an understanding of the patterns of contemporary gene flow can also be used to identify the most appropriate study units (e.g. genetically differentiated populations or the entire species) for the evolutionary questions. Hence, this chapter uses length variation among alleles at nuclear loci to investigate the population genetic structure of each of *A. trutta*, *A. truttaceus* and *A. georgianus* in Australian waters. Regardless of the thesis-wide perspective, a study of the population genetic structures of these species is significant in its own right in the context of the current interest in the extent, patterns and determinants of contemporary gene flow in marine systems (e.g. Andre´ et al. 2011; Toonen et al. 2011).

As described in Chapter 1, tagging and otolith microchemistry studies indicate that each of *A. trutta*, *A. truttaceus* and *A. georgianus* have complex life histories involving a very high dispersal potential (see Figure 1.3). Specifically, in all species: (i) pelagic
larvae are passively transported over potentially vast distances; (ii) post-settlement and
sub-adult fish progressively return to the respective adult distributions over several
years; and (iii) adults of all species make seasonal and spawning movements within
their adult distributions (see species accounts in Chapter 1). In fact, based on this life-
history, few coastal marine fishes appear to share the dispersal potential of *Arripis*

On the basis of possessing a seemingly high dispersal potential at every life stage, it is
tempting to conclude that the species of *Arripis* provide some rare examples of coastal
pelagic species that are panmictic (random mating) over their entire range (e.g. Waples
1987; Ward *et al.* 1994; Siegel *et al.* 2003). However, although, in the past, panmixia
has been assumed for many coastal pelagic fishes, most studies of species with a
seemingly highly dispersive life-style have, in fact, detected the presence of population
genetic structure (e.g. Gold *et al.* 2002; Knutsen *et al.* 2003; Rohfrisch and Borsa 2005;
Ruzzante *et al.* 2006a; Atarhouch *et al.* 2007; Chairi *et al.* 2007; Marko *et al.* 2007;
Chlaida *et al.* 2008; Wang *et al.* 2008, but see Hoolihan *et al.* 2006; Cárdenas *et al.*
2009; Palm *et al.* 2009). This emphasises the importance of compiling detailed
information before jumping to conclusions. Testing the assumption of panmixia is
especially important in exploited species, such as *Arripis*, for which fisheries-
management decisions are (or should be) based on an understanding of connectivity and
dispersal (Sale *et al.* 2005; Ruzzante *et al.* 2006a).

As established in Chapter 1, there is a need for analysis of the population genetic
structure of *Arripis* species based on DNA markers. Certainly, previous genetic studies
on these species have been based on allozyme markers with extremely low levels of
polymorphism (MacDonald 1980; Ayvazian et al. 2004). In addition, only the genetic results for *A. georgianus* have been published (Ayvazian et al. 2004).

3.1.1 Aims and Hypothesis

Length polymorphism in several nuclear introns was used to examine the genetic structure of each of *A. trutta*, *A. truttaceus* and *A. georgianus* in Australian waters, and test the following hypothesis.

*Each of the three Arripis species is genetically homogenous across its Australian distribution.*

This is predicted on the basis that each of these species undergoes extensive movements at all stages of life and the potential for mixing is high.

The reasons for using length polymorphism in introns for this analysis are explained in Section 1.3.1.2. Ideally, it would have been useful to complement the nDNA data with information from a mtDNA marker (e.g. Chow and Takeyama 2000; Haney et al. 2007), however, while mtDNA sequence data were collected for these species, the levels of polymorphism were not adequate to test for population structure (see Chapters 4, 5 and 6) and it was not possible to reliably amplify the potentially more variable control region (see Chapter 2).
3.2 Materials and Methods

3.2.1 Sampling

The results of this chapter are based upon the analysis of samples of: (i) *A. trutta* from three sampling sites (NSW, Victoria, Tasmania); (ii) *A. truttaceus* from five sampling sites ranging from Augusta, Western Australia to The Coorong, South Australia; and (iii) *A. georgianus* from seven sampling sites ranging from Fremantle, Western Australia to The Coorong, South Australia (see Chapter 2, Figure 2.1 for details). For *A. trutta*, the number of individuals per sample ranged from 29 to 30 (total number of individuals = 89), while the range for *A. truttaceus* was 9 to 39 (total number of individuals = 130) and that for *A. georgianus* was 9 to 44 (total number of individuals = 220) (see Chapter 2 for details). While most sites were only sampled once, a sample of *A. georgianus* was collected from one site (Fremantle) once in 2006 and again in 2008 in order to test for temporal changes at a site. Individuals of *A. georgianus* were measured for total length in order to provide an estimation of age or if only a head was available, total length was estimated from the relationship between head length and total length. Most samples included fish of various sizes and thus contained fish of different ages (i.e. multiple cohorts).

3.2.2 Genetic Assays

The results of this chapter are based on the analysis of length polymorphism from two to four EPIC amplified intron loci per species, depending on the species. In particular, the results for *A. trutta* are based on length variation at the loci AldoB4, S72, CK7; those
for *A. truttaceus* are based on the loci *AldoB4, S72*; and those for *A. georgianus* are based on the loci *AldoB4, CaM4, S72, CK7* (see Chapter 2). The target loci for each species were selected for analysis on the basis that they could be reliably amplified from that species and that they exhibited scorable length polymorphism, where the frequency of the most common allele was less than 0.95 in pooled samples for a species (see Chapter 2 for details of locus selection). The alleles at each locus were PCR-amplified and screened and scored using the methods described in Chapter 2. The alleles at each locus were binned into size classes as explained in Chapter 2 (Section 2.5.4). The frequencies of (binned) alleles were calculated from the raw genotype data.

3.2.3 Statistical Analyses

3.2.3.1 Levels of Polymorphism

For each species, the number of alleles per locus (*Nₐ*), observed heterozygosity (*Hₒ*) and expected unbiased gene diversity (expected heterozygosity, *Hₑ*, Nei 1987) were calculated for each sample and over all samples.

3.2.3.2 HWE, Linkage Disequilibrium and Cohort Effects

Exact tests (Guo and Thompson 1992) were used to assess the statistical significance of differences between the observed numbers of homozygotes and heterozygotes and the numbers expected under Hardy-Weinberg Equilibrium (HWE) conditions at each intron locus in each sample of *Arriplis*. This was done to test for the existence of patterns of non-random mating (see Nei 1987) in the sampled assemblages of *Arriplis* and/or the
presence of PCR artefacts such as null alleles and large allele drop-out in the data set (see Pompanon et al. 2005). The exact tests used the Markov chain method to estimate the exact probability of a Type I error, as implemented by GENEPOP version 4.0 (Raymond and Rousset 1995a). The iteration parameters for these, and all subsequent tests conducted with GENEPOP, were 10,000 dememorisation steps, 1,000 batches and 10,000 iterations per batch. Exact probability tests are not considered to be biased by small sample sizes or low frequencies of alleles or genotypes (Raymond and Rousset 1995a) and so are appropriate for the EPIC data of this research. A sequential Bonferroni correction (Rice 1989) was applied to control for group-wide Type I error.

Exact probability tests were also used to test for evidence of non-random associations of genotypes between loci (i.e. for departures from linkage disequilibrium) in the samples of *Arripis*, using GENEPOP. Exact probability tests were conducted for all pairwise combinations of loci in each sample of each species. A global test for all samples within a species was also conducted using an exact chi-square test with Fisher’s combined probability method in GENEPOP (Raymond and Rousset 1995b).

Since the presence of multiple age-class cohorts could potentially affect the results through events such as annual variations in recruitment (e.g. Johnson and Black 1984; Hedgecock 1994), a limited test for the effects of cohort on genetic variation was conducted using the *A. georgianus* data. This was done by assigning individuals of *A. georgianus* to one of four year classes using the length-age relationship developed for this species by Fairclough et al. (2000b) (Table 3.1) and then using a hierarchical Analysis of Molecular Variance (AMOVA) to investigate how genetic variation was partitioned within and among the cohorts from the different collection sites. The
AMOVA was conducted using ARLEQUIN version 3.1 (Excoffier et al. 2005), with statistical significance determined by a non-parametric approach over 10,000 permutations. *Arripis georgianus* was selected for this analysis because the sampled individuals of this species exhibited the widest range of sizes and a calibrated method of estimating age groups of individuals was available (Fairclough et al. 2000b).

**Table 3.1** Sample size for each cohort within each spatial sample of *Arripis georgianus*. Individuals were assigned to cohorts based on the length-age relationship of Fairclough et al. (2000b). Collection sites are ordered from the most northerly to the most south-easterly (see Table 2.1 for details).

<table>
<thead>
<tr>
<th>Cohort</th>
<th>HF</th>
<th>HC</th>
<th>HP</th>
<th>HI</th>
<th>HO</th>
<th>HW</th>
<th>HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2004</td>
<td>17</td>
<td>11</td>
<td>30</td>
<td>17</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2006</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
</tbody>
</table>

On the basis that the hierarchical AMOVA indicated that there was no cohort effect (see Results, Section 3.3.2), the rest of the analyses for each species proceeded using pooled cohorts. Note that the age-based nature of migration in these fishes means that it is difficult to sample all cohorts at a single site (see Table 3.1), and as a consequence the AMOVA (for *A. georgianus*) unavoidably had many low cell values, which might reduce its power, although the *p*-values were far from significant.
3.2.3.3  Population Differentiation

Several approaches were used to assess population genetic structure in each of *A. trutta*, *A. truttaceus* and *A. georgianus*, as detailed below. The 2008 sample of *A. georgianus* was not included in any of the spatial analyses, except as described in Section 3.2.3.8.

3.2.3.4  Pairwise Differentiation

$F_{ST}$, the standardised genetic variance, was used as a measure of the proportion of the total allele frequency variation due to differences between samples for each species of *Arripis*. If genetic heterogeneity exists among samples, $F_{ST}$ values will be significantly greater than zero. Pairwise $F_{ST}$ tests between samples within each species were conducted with the permutation approach in FSTAT version 2.9.3.2 (Goudet 1995), with 10,000 permutations assuming samples in HWE. Using the method of Weir & Cockerham (1984), $F_{ST}$ values represented weighted averages of variation across loci.

There is increasing recognition of the importance of standardising $F_{ST}$ according to the level of polymorphism detected within a species in order to compare across species and markers (e.g. Hedrick 2005; Meirmans 2007), but given that the $F_{ST}$ values for *Arripis* were effectively zero or nearly so (see Results), standardisation was not applicable in this case.

The pairwise relationships among the samples, as measured by pairwise values of $F_{ST}$, of each species was summarised in a multi-dimensional scaling (MDS) plot that was constructed with 9,999 restarts in PRIMER version 5.2.9 (Clarke and Gorley 2002).
3.2.3.5 **Overall Differentiation**

Single-locus $F_{ST}$ estimates were made using FSTAT, as described in Section 3.2.3.4. The statistical significance of differences in the allele frequencies at each locus between all samples of a species was determined as the probability of obtaining, by chance, a value as large as or larger than the observed $F_{ST}$.

Differentiation over all loci between spatial samples was tested using a re-sampling permutation test utilising 10,000 permutations with the software GENETIX version 4.05.2 (Belkhir *et al.* 2004) and generating a probability estimate by counting the number of times a randomly permuted $F_{ST} \geq$ actual $F_{ST}$.

3.2.3.6 **Factorial Correspondence Analysis**

The factorial correspondence analysis (FCA) of GENETIX was used to search for structure in the data. In this analysis, composite axes were constructed using allelic data so that inter-individual differences were maximised and the result visualised by plotting onto a two-dimensional chart. Guinand (1996) showed that the values of inertia along each axis can be regarded as combinations of single locus $F_{ST}$ linear values, thus FCA is a valid method to examine heterogeneity using $F_{ST}$. It differs from other methods by using individuals (rather than sub-populations) as the unit of analysis.
3.2.3.7 Power Analysis

A retrospective power analysis was conducted with 1000 runs using the method described by Theisen et al. (2008) and the software POWSIM version 4.0 (Ryman and Palm 2006). The actual data (i.e. number of populations and their sample sizes, number of alleles at each locus and their observed frequencies) were used to generate 1000 random populations that were then allowed to drift to various levels of genetic divergence (quantified by $F_{ST}$ and determined by a combination of effective population size, $N_e$, and generations of drift, $t$). These simulated populations were sampled, and tested against the null hypothesis of genetic homogeneity among populations using a chi-square approach and Fisher’s exact test. The percentage of significant outcomes ($p < 0.05$) from this resampling was interpreted as the power of the data to detect the defined level of genetic divergence. For each simulation in POWSIM, $t$ was altered, while $N_e$ was held constant at 1000. The simulations were repeated until the lowest level of genetic divergence (i.e. lowest value of $t$) that could be detected among the sampled populations of each species with a power $\geq 95\%$ was identified. The software cannot accept different sample sizes for different loci within a sample, so the analysis used a reduced dataset by removing individuals with missing data. Therefore, this method provides an underestimate of the true power of the data. Sample sizes for the power analysis were as follows: (see Table 2.1 for sample codes): *A. trutta* EN (28), EV (27), ET (27); *A. truttaceus* WA (27), WY (15), WL (38), WW (8), WS (31); *A. georgianus* HF (30), HC (35) HP (36), HI (21), HO (25), HW (6), HS (35).
3.2.3.8  **Temporal Stability**

Extensive temporal sampling was not undertaken, partly because preliminary analyses revealed no evidence of significant spatial genetic heterogeneity in any of the three study species. Nevertheless, two samples of *A. georgianus* collected from the same site (Fremantle, Western Australia) two years apart were compared, providing a very limited test of the extent of any temporal heterogeneity. Three approaches were used to test for temporal stability. Firstly, the values of $F_{ST}$ between the sample collected from Fremantle in 2008 and each of the seven samples of *A. georgianus*, including one from Fremantle, collected in 2006 were plotted in a MDS, as described above. Secondly, an AMOVA between just the two temporal samples (HF and HF08) was conducted. By nesting the two samples into a single group, this design directly tests for temporal differences (e.g. Hale *et al.* 2001). Thirdly, the *A. georgianus* samples collected from Fremantle in 2008 were added to the FCA, and graphed as described above.

### 3.3  Results

3.3.1  Polymorphism, HWE and Linkage Disequilibrium

Length polymorphism at four EPIC amplified intron loci (*AldoB4, S72, CK7, CaM4*) was used to investigate the population genetic structure of *A. georgianus*. The first three of these loci also exhibited length polymorphism in the samples of *A. trutta* and so were also used to investigate the population genetic structure of this species. Similarly, the first two loci exhibited length polymorphism in the samples of *A. truttaceus* and
were used for this species. Based on all samples for all species combined, the alleles at the locus *AldoB4* ranged from 170 to 191 bp, while those at the locus *CaM4* ranged from 469 - 504 bp and those of *S72* and *CK7* ranged between 816 – 931 bp and 844 - 984 bp, respectively (Table 3.2).

The overall level of polymorphism was greatest in the samples of *A. georgianus* (four polymorphic loci, with a total of 32 different alleles), followed by those of *A. trutta* (three polymorphic loci, with a total of 20 different alleles), while those of *A. truttaceus* contained only two polymorphic loci, with 9 different alleles (Table 3.2). Mean expected heterozygosity per sample per locus, when the locus was polymorphic in the sample, was typically low to moderate, for example it ranged from 0.083 (*CaM4* locus in the HI sample of *A. georgianus*) to 0.655 (*S72* locus in the EN sample of *A. trutta*) (Table 3.2).

There were no statistically significant, non-random associations of genotypes at different loci for the samples of each of the three species (once a Bonferroni correction was applied); thus the patterns of variation at the different intron loci are assumed to be independent of each other (Tables 3.3, 3.4, 3.5). The genotype frequencies at each locus in each sample of each of the three species were generally in accordance with those expected under HWE. Specifically, of 49 tests, only a single departure from HWE was detected (*CK7* locus in the HO sample of *A. georgianus*; Table 3.2) and this was not statistically significant following Bonferroni correction.
Table 3.2  (over page) Allele frequencies at each of four intron loci for samples of $n$ individuals of three species of *Arripis*. $N_A$ is the number of alleles for each sample and for the species (in parentheses). $H_E$ and $H_O$ are expected and observed heterozygosity, respectively. HWE($p$) is the probability of a Type 1 error in exact tests for departures from Hardy-Weinberg equilibrium expectations. A dash (-) indicates the allele was not recorded. Collection sites are ordered from the most northerly to the most south-easterly (see Table 2.1 for details).
<table>
<thead>
<tr>
<th></th>
<th>Arrips trutta</th>
<th></th>
<th></th>
<th>Arrips truttaceus</th>
<th></th>
<th></th>
<th>Arrips georgianus</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EN EV ET</td>
<td>WA WY WL WW WS</td>
<td>HF HC HP HI HO HW HS HF08</td>
<td>n</td>
<td>29 30 30</td>
<td>30 15 39 9 37</td>
<td>44 39 36 23 28 9 41 41</td>
<td></td>
</tr>
<tr>
<td>AldoB4 locus</td>
<td></td>
<td></td>
<td></td>
<td>N&lt;sub&gt;A&lt;/sub&gt;</td>
<td>5 3 3 (6)</td>
<td>2 2 2 2 2 2 (2)</td>
<td>5 2 4 4 3 4 4 5 (5)</td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>0.000 0.000 0.017</td>
<td>- - - - - -</td>
<td></td>
<td>0.114 0.064 0.153 0.152 0.089 0.056 0.061 0.073</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172</td>
<td>0.017 0.017 0.000</td>
<td>- - - - - -</td>
<td></td>
<td>0.011 0.000 0.000 0.000 0.018 0.000 0.012 0.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>- - - - - -</td>
<td></td>
<td></td>
<td>0.023 0.000 0.014 0.043 0.000 0.000 0.012 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>0.741 0.783 0.800</td>
<td>0.600 0.733 0.615 0.778 0.743</td>
<td></td>
<td>0.818 0.936 0.819 0.783 0.893 0.833 0.915 0.902</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>182</td>
<td>0.034 0.000 0.000</td>
<td>0.400 0.267 0.385 0.222 0.257</td>
<td></td>
<td>0.034 0.000 0.014 0.022 0.000 0.111 0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>0.034 0.000 0.000</td>
<td>- - - - - -</td>
<td></td>
<td>- - - - - - - - - - - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>191</td>
<td>0.172 0.200 0.183</td>
<td>- - - - - -</td>
<td></td>
<td>- - - - - - - - - - - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H&lt;sub&gt;E&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td>0.419 0.347 0.326</td>
<td>0.480 0.391 0.474 0.345 0.382</td>
<td></td>
<td>0.316 0.120 0.305 0.361 0.194 0.291 0.159 0.180</td>
<td></td>
</tr>
<tr>
<td>H&lt;sub&gt;O&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td>0.414 0.433 0.333</td>
<td>0.667 0.400 0.513 0.222 0.405</td>
<td></td>
<td>0.295 0.128 0.250 0.348 0.214 0.333 0.171 0.171</td>
<td></td>
</tr>
<tr>
<td>HWE(p)</td>
<td></td>
<td></td>
<td></td>
<td>0.33 0.45 1.00</td>
<td>0.06 1.00 0.74 0.34 1.00</td>
<td></td>
<td>0.50 1.00 0.43 0.71 1.00 1.00 1.00 0.16</td>
<td></td>
</tr>
<tr>
<td>CaM4 locus</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>3 3 4</td>
<td>2 2 2 2 2 2</td>
<td>42 39 36 23 28 8 40 41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N&lt;sub&gt;A&lt;/sub&gt;</td>
<td>1 1 1 (1)</td>
<td>1 1 1 1 1 1 (1)</td>
<td>5 5 3 3 3 2 6 5 (8)</td>
<td></td>
</tr>
<tr>
<td>469</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.000 0.013 0.000 0.000 0.018 0.000 0.000 0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>477</td>
<td>1.000 1.000 1.000</td>
<td>1.000 1.000 1.000 1.000 1.000</td>
<td></td>
<td>0.000 0.000 0.000 0.000 0.000 0.000 0.013 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>480</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.000 0.013 0.000 0.000 0.000 0.000 0.013 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>483</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.012 0.013 0.000 0.022 0.000 0.000 0.013 0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>486</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.060 0.038 0.042 0.022 0.107 0.063 0.075 0.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>491</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.905 0.923 0.944 0.957 0.875 0.938 0.875 0.939</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>498</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.012 0.000 0.000 0.000 0.000 0.000 0.000 0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td>0.012 0.000 0.014 0.000 0.000 0.000 0.013 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H&lt;sub&gt;E&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td>- - - - - - - - - - -</td>
<td></td>
<td>0.177 0.146 0.107 0.083 0.223 0.116 0.228 0.117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H&lt;sub&gt;O&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td>- - - - - - - - - - -</td>
<td></td>
<td>0.190 0.154 0.111 0.087 0.179 0.125 0.200 0.122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWE(p)</td>
<td></td>
<td></td>
<td></td>
<td>- - - - - - - - - - -</td>
<td></td>
<td>1.00 1.00 1.00 1.00 0.35 - 0.48 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>arrripis trutta</td>
<td>arrripis truttaceus</td>
<td>arrripis georgianus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EN  EV  ET</td>
<td>WA  WY  WL  WW  WS</td>
<td>HF  HC  HP  HI  HO  HW  HS  HF08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CK7 locus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>28  28  29</td>
<td>8  9  8  12  10</td>
<td>42  38  35  23  26  8  39  38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_A</td>
<td>3  4  3 (4)</td>
<td>1  1  1  1  1 (1)</td>
<td>6  7  4  3  2  1  6  6  (11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>844</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.000 0.000 0.000 0.000 0.000 0.000 0.026 0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>848</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.012 0.026 0.014 0.022 0.000 0.000 0.051 0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>923</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.000 0.013 0.000 0.000 0.000 0.000 0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>931</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.000 0.000 0.000 0.000 0.000 0.000 0.013 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>936</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.012 0.000 0.000 0.000 0.000 0.000 0.000 0.013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>939</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.000 0.013 0.000 0.000 0.000 0.000 0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>943</td>
<td>0.089 0.018 0.052</td>
<td>- - - - - -</td>
<td>0.060 0.013 0.071 0.087 0.077 0.000 0.026 0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>948</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.060 0.013 0.029 0.000 0.000 0.000 0.038 0.053</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>952</td>
<td>0.839 0.893 0.845</td>
<td>1.000 1.000 1.000 1.000 1.000</td>
<td>0.845 0.908 0.886 0.891 0.923 1.000 0.846 0.829</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>961</td>
<td>0.071 0.071 0.103</td>
<td>- - - - - -</td>
<td>0.012 0.000 0.000 0.000 0.000 0.000 0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>976</td>
<td>- - - -</td>
<td>- - - - - -</td>
<td>0.000 0.013 0.000 0.000 0.000 0.000 0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>984</td>
<td>0.000 0.018 0.000</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td>- - - - - -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H_E</td>
<td>0.283 0.197 0.273</td>
<td>- - - - - -</td>
<td>0.278 0.174 0.209 0.198 0.142 - 0.279 0.306</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H_O</td>
<td>0.321 0.214 0.321</td>
<td>- - - - - -</td>
<td>0.214 0.184 0.229 0.217 0.154 - 0.231 0.237</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HWE(p)</td>
<td>1.00 1.00 1.00</td>
<td>- - - - - -</td>
<td>0.20 1.00 1.00 1.00 1.00 - 0.13 0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Arripis trutta</strong></td>
<td></td>
<td><strong>Arripis truttaceus</strong></td>
<td></td>
<td><strong>Arripis georgianus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
<td>----------</td>
<td>------------------------</td>
<td>----------</td>
<td>------------------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>EN</strong></td>
<td><strong>EV</strong></td>
<td><strong>ET</strong></td>
<td><strong>WA</strong></td>
<td><strong>WY</strong></td>
<td><strong>WL</strong></td>
<td><strong>WW</strong></td>
<td><strong>WS</strong></td>
</tr>
<tr>
<td><strong>S72 locus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>29</td>
<td>25</td>
<td>28</td>
<td>27</td>
<td>15</td>
<td>38</td>
<td>13</td>
<td>45</td>
</tr>
<tr>
<td><strong>N_A</strong></td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>(5)</td>
</tr>
<tr>
<td><strong>816</strong></td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>869</strong></td>
<td>0.000</td>
<td>0.020</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>871</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>875</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>883</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>889</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>892</strong></td>
<td>0.000</td>
<td>0.020</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.022</td>
</tr>
<tr>
<td><strong>895</strong></td>
<td>0.000</td>
<td>0.020</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>900</strong></td>
<td>0.224</td>
<td>0.100</td>
<td>0.107</td>
<td>0.111</td>
<td>0.233</td>
<td>0.066</td>
<td>0.038</td>
<td>0.178</td>
</tr>
<tr>
<td><strong>914</strong></td>
<td>0.500</td>
<td>0.580</td>
<td>0.625</td>
<td>0.722</td>
<td>0.600</td>
<td>0.697</td>
<td>0.731</td>
<td>0.600</td>
</tr>
<tr>
<td><strong>919</strong></td>
<td>0.207</td>
<td>0.240</td>
<td>0.250</td>
<td>0.148</td>
<td>0.167</td>
<td>0.211</td>
<td>0.192</td>
<td>0.167</td>
</tr>
<tr>
<td><strong>924</strong></td>
<td>0.034</td>
<td>0.020</td>
<td>0.018</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>931</strong></td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.019</td>
<td>0.000</td>
<td>0.026</td>
<td>0.038</td>
<td>0.033</td>
</tr>
<tr>
<td><strong>H_E</strong></td>
<td>0.655</td>
<td>0.594</td>
<td>0.535</td>
<td>0.444</td>
<td>0.558</td>
<td>0.465</td>
<td>0.426</td>
<td>0.579</td>
</tr>
<tr>
<td><strong>H_O</strong></td>
<td>0.552</td>
<td>0.480</td>
<td>0.679</td>
<td>0.333</td>
<td>0.467</td>
<td>0.474</td>
<td>0.462</td>
<td>0.667</td>
</tr>
<tr>
<td><strong>HWE(p)</strong></td>
<td>0.31</td>
<td>0.31</td>
<td>0.51</td>
<td>0.12</td>
<td>0.56</td>
<td>1.00</td>
<td>0.43</td>
<td>0.24</td>
</tr>
</tbody>
</table>

* not significant after Bonferroni correction.
Table 3.3  Results of exact tests for linkage disequilibrium for all pairwise combinations of three intron loci for each sample of *Arripis trutta* and for all of these samples combined. The results for the pairwise comparisons are presented in terms of the probability of Type I error, as determined by exact tests. The results for the global test are exact probabilities from a chi-square test ($\chi^2$), with associated degrees of freedom (df), using Fisher’s combined probability method. Sample codes are as per Table 2.1.

<table>
<thead>
<tr>
<th>Locus Pair</th>
<th>samples</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>EN</td>
<td>EV</td>
<td>ET</td>
<td></td>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
</tr>
<tr>
<td>CK7 S72</td>
<td></td>
<td>0.19</td>
<td>0.50</td>
<td>0.58</td>
<td></td>
<td></td>
<td>5.833</td>
<td>6</td>
</tr>
<tr>
<td>CK7 AldoB4</td>
<td></td>
<td>0.32</td>
<td>0.52</td>
<td>0.93</td>
<td></td>
<td></td>
<td>3.752</td>
<td>6</td>
</tr>
<tr>
<td>S72 AldoB4</td>
<td></td>
<td>0.77</td>
<td>0.14</td>
<td>0.31</td>
<td></td>
<td></td>
<td>6.842</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3.4  Results of exact tests for linkage disequilibrium for all pairwise combinations of two intron loci for each sample of *Arripis truttaceus* and for all of these samples combined. The results for the pairwise comparisons are presented in terms of the probability of Type I error, as determined by exact tests. The results for the global test are exact probabilities from a chi-square test ($\chi^2$), with associated degrees of freedom (df), using Fisher’s combined probability method. Sample codes are as per Table 2.1.

<table>
<thead>
<tr>
<th>Locus Pair</th>
<th>samples</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>WA</td>
<td>WY</td>
<td>WL</td>
<td>WW</td>
<td>WS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S72 AldoB4</td>
<td></td>
<td>0.80</td>
<td>0.82</td>
<td>0.75</td>
<td>0.11</td>
<td>0.70</td>
<td>6.631</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 3.5  Results of exact tests for linkage disequilibrium for all pairwise combinations of four intron loci for each sample of *Arripis georgianus* and for all of these samples combined. The results for the pairwise comparisons are presented in terms of the probability of Type I error as determined by exact tests. The results for the global test are exact probabilities from a chi-square test ($\chi^2$), with associated degrees of freedom (df), using Fisher’s combined probability method. Sample codes are as per Table 2.1.

<table>
<thead>
<tr>
<th>Locus Pair</th>
<th>samples</th>
<th>global</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HF</td>
<td>HC</td>
</tr>
<tr>
<td>CK7</td>
<td>0.04</td>
<td>0.57</td>
</tr>
<tr>
<td>S72</td>
<td>0.46</td>
<td>0.22</td>
</tr>
<tr>
<td>AldoB4</td>
<td>0.56</td>
<td>0.11</td>
</tr>
<tr>
<td>CaM4</td>
<td>0.85</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>AldoB4</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>CaM4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Cohort Effects

As mentioned in Section 3.2.3.2, genetic variation among cohorts was negligible. The hierarchical AMOVA, which included individuals grouped into cohorts within each spatial sample of *A. georgianus*, found that almost all of the variance in the data (99.35%) occurred within cohort groups, although $F_{ST}$ was effectively zero ($F_{ST} = 0.006, p = 0.33$). Furthermore, variance due to differences between cohorts within a spatial sample ($F_{SC} = 0.002$) was less than the variance between spatial samples ($F_{CT} = 0.004$), although both were effectively zero ($p = 0.26$ and 0.48, respectively).

3.3.3 Population Differentiation

The values of $F_{ST}$ between each pair of samples within each of the three study species were invariably low (e.g. effectively zero for *A. trutta*, $F_{ST} < 0.018$ for *A. truttaceus*, and $F_{ST} < 0.030$ for *A. georgianus*) and none were significantly different from zero (Tables 3.6, 3.7, 3.8). For each species, the distribution of samples in a MDS plot, which is based on pairwise $F_{ST}$ values, revealed no clear evidence of a relationship between the genetic composition of samples and the location of the source population (Figures 3.1, 3.2, 3.3).

Within each species, single locus estimates of (global) $F_{ST}$ were consistently low (Table 3.9), reflecting the fact that the allele frequencies at each locus were generally similar in all of the spatial samples of a particular species (Table 3.2). Accordingly the (global) multi-locus values of $F_{ST}$ for each species were also low (e.g. $\leq 0.005$) and not significantly different from zero (Table 3.9).
Factorial correspondence analysis revealed little to no clustering of individuals in any species, with clouds of data points from each spatial sample generally overlapping (Figures 3.4, 3.5, 3.6). In all cases, no spatial patterns were discernable and the distribution of the data was not strongly explained by either axis (Figures 3.4, 3.5, 3.6).

3.3.4 Power Analysis

Power analyses suggested that genetic divergence would be detected with $\geq 95\%$ confidence, at least at $F_{ST} \geq 0.034$ for $A.\ trutta$ ($N_e = 1000$, $t = 70$), $F_{ST} \geq 0.039$ for $A.\ truttaceus$ ($N_e = 1000$, $t = 80$) and $F_{ST} \geq 0.009$ for $A.\ georgianus$ ($N_e = 1000$, $t = 20$).

3.3.5 Temporal Stability

No pairwise tests involving the temporal samples of $A.\ georgianus$ from Fremantle were significant (Table 3.8), and the multi-dimensional (MDS) distance between the two temporal samples was equal to or greater than the distances between most of the other spatial samples (Figure 3.3). AMOVA detected no significant difference between the samples of $A.\ georgianus$ collected from Fremantle two years apart ($F_{ST} = 0.012$, $p = 0.12$), and the cloud of individuals from the two samples in the FCA generally overlapped (Figure 3.6).
**Table 3.6** Results of tests for differentiation between pairs of samples of *Arripis trutta* presented as pairwise $F_{ST}$ values (below diagonal) and the probability of obtaining by chance, a value as large or larger than the observed $F_{ST}$ (above diagonal, italicised) using a permutation approach. Sample codes are as per Table 2.1.

<table>
<thead>
<tr>
<th></th>
<th>EN</th>
<th>EV</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>0.27</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>EV</td>
<td>-0.004</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>ET</td>
<td>-0.000</td>
<td>-0.013</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.1** Multi-dimensional scaling (MDS) plot of pairwise $F_{ST}$ values between samples of *Arripis trutta*, as presented in Table 3.6. Sample codes are as per Table 2.1. The stress value is an inverse measure of the degree of correspondence between the distances among points implied by MDS and the matrix input.
Table 3.7  Results of tests for differentiation between pairs of samples of *Arripis truttaceus* presented as pairwise $F_{ST}$ values (below diagonal) and the probability of obtaining by chance, a value as large or larger than the observed $F_{ST}$ (above diagonal, italicised) using a permutation approach. Sample codes are as per Table 2.1.

<table>
<thead>
<tr>
<th></th>
<th>WA</th>
<th>WY</th>
<th>WL</th>
<th>WW</th>
<th>WS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
<td>0.40</td>
<td>0.83</td>
<td>0.49</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>WY</td>
<td>0.009</td>
<td></td>
<td>0.11</td>
<td>0.22</td>
<td>0.79</td>
</tr>
<tr>
<td>WL</td>
<td>-0.011</td>
<td>0.012</td>
<td></td>
<td>0.76</td>
<td>0.08</td>
</tr>
<tr>
<td>WW</td>
<td>0.008</td>
<td>-0.011</td>
<td>0.000</td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td>WS</td>
<td>0.018</td>
<td>-0.019</td>
<td>0.018</td>
<td>-0.006</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.2  Multi-dimensional scaling (MDS) plot of pairwise $F_{ST}$ values between samples of *Arripis truttaceus*, as presented in Table 3.7. Sample codes are as per Table 2.1. The stress value is an inverse measure of the degree of correspondence between the distances among points implied by MDS and the matrix input.
Table 3.8  Results of tests for differentiation between pairs of samples of *Arripis georgianus* presented as pairwise $F_{ST}$ values (below diagonal) and the probability of obtaining by chance, a value as large or larger than the observed $F_{ST}$ (above diagonal, italicised) using a permutation approach. Sample codes are as per Table 2.1.

<table>
<thead>
<tr>
<th></th>
<th>HF</th>
<th>HC</th>
<th>HP</th>
<th>HI</th>
<th>HO</th>
<th>HW</th>
<th>HS</th>
<th>HF08*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>0.35</td>
<td>0.90</td>
<td>0.34</td>
<td>0.37</td>
<td>0.87</td>
<td>0.72</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>HC</td>
<td>0.003</td>
<td>0.39</td>
<td>0.14</td>
<td>0.38</td>
<td>0.61</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP</td>
<td>0.005</td>
<td>0.012</td>
<td></td>
<td>0.92</td>
<td>0.68</td>
<td>0.24</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>HI</td>
<td>0.010</td>
<td>0.019</td>
<td>-0.013</td>
<td>0.30</td>
<td>0.07</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HO</td>
<td>0.010</td>
<td>0.005</td>
<td>-0.005</td>
<td>-0.000</td>
<td>0.19</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HW</td>
<td>-0.016</td>
<td>0.000</td>
<td>0.029</td>
<td>0.030</td>
<td>0.026</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>0.004</td>
<td>-0.002</td>
<td>0.002</td>
<td>0.009</td>
<td>-0.007</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF08*</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*HF08 has only been compared to HF (see text).*

Figure 3.3  Multi-dimensional scaling (MDS) plot of pairwise $F_{ST}$ values between samples of *Arripis georgianus*, as presented in Table 3.8. Sample codes are as per Table 2.1. The stress value is an inverse measure of the degree of correspondence between the distances among points implied by MDS and the matrix input. HF is the sample collected from Fremantle at the same time as the other samples (i.e. 2006) while HF08 is the sample collected from Fremantle in 2008.
Table 3.9 Results of tests of differentiation in each Arripis species. (i) Single-locus $F_{ST}$ values across all samples for each species and the probability of obtaining, by chance, a value as large as or larger than the observed $F_{ST}$ from 10,000 permutations. A dash (-) indicates that a single allele was recorded in the species. (ii) Multi-locus $F_{ST}$ values from permutation test and the number of times a randomly permuted $F_{ST} \geq$ actual $F_{ST}$. Sample sizes are as reported in Table 3.2.

<table>
<thead>
<tr>
<th></th>
<th>A. trutta $F_{ST}$</th>
<th>p-value</th>
<th>A. truttaceus $F_{ST}$</th>
<th>p-value</th>
<th>A. georgianus $F_{ST}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Single-locus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AldoB4</td>
<td>-0.010</td>
<td>0.34</td>
<td>0.011</td>
<td>0.23</td>
<td>0.007</td>
<td>0.14</td>
</tr>
<tr>
<td>CaM4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.004</td>
<td>0.98</td>
</tr>
<tr>
<td>CK7</td>
<td>-0.006</td>
<td>0.52</td>
<td>-</td>
<td>-</td>
<td>-0.009</td>
<td>0.26</td>
</tr>
<tr>
<td>S72</td>
<td>-0.002</td>
<td>0.63</td>
<td>0.001</td>
<td>0.42</td>
<td>0.005</td>
<td>0.48</td>
</tr>
<tr>
<td>(ii) Permutation</td>
<td>-0.005</td>
<td>0.73</td>
<td>0.005</td>
<td>0.22</td>
<td>0.005</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Figure 3.4  Factorial correspondence analysis (FCA) illustrating the lack of strong relationship between the multi-locus genotypes and collection site of individuals of *Arripis trutta*. The total variation explained by each axis is given as a percentage. Sample codes are as per Table 2.1. For scaling purposes, a single outlier has been shifted and its correct position indicated by an arrow and coordinates. Note, many points overlap and are obscured.
Figure 3.5  Factorial correspondence analysis (FCA) illustrating the lack of strong relationship between the multi-locus genotypes and collection site of individuals of *Arripis truttaceus*. The total variation explained by each axis is given as a percentage. Sample codes are as per Table 2.1. For scaling purposes, outliers have been shifted and their correct position indicated by an arrow and coordinates. Note, many points overlap and are obscured.
Factorial correspondence analysis (FCA) illustrating the lack of strong relationship between the multi-locus genotypes and collection site of individuals of *Arripis georgianus*. The total variation explained by each axis is given as a percentage. Sample codes are as per Table 2.1. For scaling purposes, outliers have been shifted and their correct position indicated by an arrow and coordinates. Note, many points overlap and are obscured.
3.4 Discussion

This research did not reveal any evidence of genetic structuring in any of *A. trutta*, *A. truttaceus* and *A. georgianus* in Australian waters. For one or more species, this could be due to: (i) inadequate geographic sampling; (ii) errors in the genetic data; (iii) the effects of selection; and/or (iv) the presence of genetic structuring that was too subtle to be apparent given the loci/sample size combinations. Alternatively, and most likely, it could reflect the presence of sufficient gene flow across the entire range of the species to prevent genetic structuring. Each of these alternatives is considered in turn below.

Point (i) – inadequate geographic sampling. The results were based on samples obtained from virtually either the entire distribution of the species in Australian waters (*A. trutta*) or the entire distribution of the Australian species (*A. truttaceus*, *A. georgianus*). Thus, it seems highly unlikely that any genetic differentiation in any of the species in Australian waters was missed, as could have happened if, for example, all individuals had been sampled from a single, more localised breeding unit within a subdivided species (see Page and Holmes 1998).

Point (ii) – errors in the genetic data. Inaccurate results due to errors in the production and/or scoring of the genetic data seem unlikely for the following reasons. Firstly, most methodological errors are likely to produce heterogeneity rather than homogeneity (Pompanon *et al.* 2005), although the allele binning could have masked some variation (see Chapter 2). Secondly, the patterns of variation at each locus in each sample for each species were virtually always in accordance with those expected under HWE conditions. Thus, there was no evidence to suggest that a range of PCR artefacts, such
as null alleles and/or large allele drop out, which typically generate homozygote excesses (Pompanon et al. 2005), have had a major impact on the data sets. Thirdly, the patterns of polymorphism and genetic variation were consistent across multiple loci, suggesting there were no locus-specific errors. Finally, none of the positive or negative controls revealed any evidence of contamination of samples with non-target DNA or of scoring inconsistencies (beyond the boundaries of the allele bins).

Point (iii) – selection. The absence of genetic subdivision in the arripids is unlikely to be reflecting the effects of selection because the patterns of genetic diversity, genetic variance ($F_{ST}$) and heterozygosity were similar across multiple loci. Instead, selection is expected to act on an individual locus rather than the whole genome (Hare 2001).

Point (iv) – subtle genetic structure. A reliance on few loci and/or markers with low levels of polymorphism will limit the power of statistical tests (Ryman et al. 2006), and therefore the ability to detect genetic structuring, if it exists. In this regard, it is relevant that the detected amount of polymorphism at the intron loci in each species of Arripis was greater than that of earlier allozyme studies. In addition, power analyses showed that the minimum power of the present data to detect heterogeneity would be sufficient to detect genetic heterogeneity equivalent to an $F_{ST}$ value of 0.009 to 0.039, depending on the species (see Section 3.3.4). Realistically, the only way to increase the power of the test for the presence of genetic structure in Arripis is to increase the amount of genetic information generated for each species. For example, power analysis simulations suggest that, in order to increase power by an order of magnitude using the current genetic markers, sample sizes would need to be increased at least six- to seven-fold (i.e. up to 700 – 1300 individuals per species; data not shown).
The ability to generate large amounts of genetic data for species, like those of *Arripis*, whose genome structures are largely unknown is becoming easier and cheaper (e.g. Santana et al. 2009). However, given the apparent extreme mobility of these fish and the limitations of genetic data in producing accurate information about high gene-flow situations, combined with the apparently low levels of genome-wide polymorphism in the representatives of this family (see Chapter 6), other non-genetic techniques (such as otolith microchemistry and tagging) may be potentially more useful in identifying sub-groups, if they exist (see Point v, below).

**Point (v) – no genetic structure.** All of the available evidence suggests that each of the three study species of *Arripis* is genetically homogeneous throughout its entire distribution (*A. truttaceus* and *A. georgianus*), or entire Australian distribution (*A. trutta*). Neither the EPIC data of the current study nor the allozyme data of previous studies (i.e. *A. truttaceus*, MacDonald 1980; *A. georgianus*, Ayvazian et al. 2004) have revealed evidence of the presence of genetic structure in any of the study species. This is consistent with the observations of extensive movements in each species based on tagging (e.g. Stanley 1986; 1988a; 1988b; 1988c; Ayvazian et al. 2004; Stewart et al. 2011) and otolith microchemistry (Ayvazian et al. 2004; Hughes et al. 2011). These rigorous studies strongly suggest that the species of *Arripis* are highly vagile and do not spend their entire life in a single water body.

The apparent absence of genetic structure combined with the mobility suggested by tagging and otolith microchemistry data raises the possibility that one or more of the *Arripis* species are actually panmictic over their entire distribution (Australian species) or entire Australian distribution (*A. trutta*). Indeed, genetic data alone cannot be used to
prove or disprove a hypothesis of panmixia because very limited amounts of gene flow can homogenise the genetic composition of what are effectively independent breeding units in a species (Shaklee and Currens 2003). The European Eel (*Anguilla anguilla*) provides an excellent example of the difficulties of establishing whether or not a marine species is panmictic or even genetically homogeneous over its range. This eel shows a life-history consistent with high gene flow in that all adults (which are normally dispersed along the coasts of Europe and northern Africa) migrate to form a single breeding unit (in the Sargasso Sea), followed by larval and post-spawning adult dispersal. For decades, based on life-history information as described above, the European Eel was considered as a classic example of a panmictic species, and early studies did not find any evidence of genetic structuring across this species entire range (e.g. DeLigny and Pantelouris 1973; Lintas *et al.* 1998). However, three subsequent studies reported evidence of genetic structuring, including isolation-by-distance (Daemen *et al.* 2001; Wirth and Bernatchez 2001; Maes and Volckaert 2002). More recently, the results of carefully controlled studies have reinstated the original idea that panmixia is the most likely model and suggested that the structure detected in some of the previous studies resulted from temporal variation in recruitment and sampling (Dannewitz *et al.* 2005; Palm *et al.* 2009; Als *et al.* 2011).

The results of this study add to a small list of pelagic coastal species that have been found to be genetically homogeneous and possibly panmictic over a vast distance (e.g. Hoolihan *et al.* 2006; Cárdenas *et al.* 2009; Palm *et al.* 2009). In the case of each *Arripsis* species, this high gene flow life-history is closely coincident (and possibly linked) with oceanographic conditions where larval dispersal is passively driven by
winds and currents, followed by active movements of immature and adult fish that seasonally converge on a spatially restricted breeding ground (see Introduction).

Finally, the distribution of *A. trutta* extends beyond Australian waters into New Zealand. Despite repeated attempts to obtain samples from New Zealand, only two specimens from near Christchurch (43° 32’ S, 172° 37’ E) could be sourced. The alleles present in these two specimens (EZ01: 178/191, 919/919, 961/961; EZ02: 178/178, 914/914, 943/952 for loci *AldoB4*, *CK7* and *S72*, respectively) were among the common alleles in the samples from eastern Australia. While this might suggest the presence of recent (in an evolutionary sense) connections between the *A. trutta* in Australia and New Zealand, obviously the sample size is inadequate to draw any conclusions about the extent of any mixing. However, such a connection might be expected based on water movement of the East Australian Current, which delivers a branch of water (and thus possibly larvae) from the spawning area of *A. trutta* in New South Wales directly to the west coast of New Zealand (Bostock *et al.* 2006; see Chapter 5).

### 3.4.1 Conclusions

This study provides the strongest evidence to date that each of *Arripsis trutta*, *A. truttaceus* and *A. georgianus* is genetically homogenous in Australian waters, although the results are to a certain extent limited by modest sample sizes and a small number of polymorphic loci. In fact, the combination of genetic and life-history data suggest that each species of *Arripsis* has considerable gene flow throughout these waters and might, in fact, comprise a single, panmictic population. On the basis of these results, it is valid to treat each species as a single population in analyses in subsequent chapters.
4.0 A Molecular Approach to the Taxonomy of Australian Salmon

4.1 Introduction

As discussed in Chapter 1, one of the main objectives of this thesis was to use molecular data to investigate the evolutionary relationships of the putative geminate species of *A. trutta* and *A. truttaceus*. However, it is relevant to first establish whether genetic data are consistent with the current view in the literature (e.g. Paulin 1993; Gomon *et al.* 2008), that these two taxa represent ‘true’ species (as opposed to conspecific populations that are exchanging genes or have recently done so). Since the taxonomic status of these two taxa is linked with that of the third species of Australian Salmon, *A. xylabion*, the status of this latter taxon is also a critical part of this issue.

The current view that each of the three types of Australian Salmon – *A. trutta*, *A. truttaceus* and *A. xylabion* – are separate species is based on morphological evidence. In particular, *A. truttaceus* differs from each of the other species in having fewer lower gill rakers and from *A. xylabion* (but not *A. trutta*) by having fewer upper gill rakers and a longer upper lobe of the caudal fin (Table 4.1; Figure 4.1). The length of the upper lobe of the caudal fin is also longer in *A. xylabion* than in *A. trutta* (Table 4.1; Figure 4.1). There are no other (known) diagnostic differences between *A. trutta* and *A. xylabion*, or between *A. trutta* and *A. truttaceus* (Table 4.1; Figure 4.1). Note the Australian Herring *A. georgianus* is distinguished from the three Australian Salmon by its much smaller maximum length and higher gill raker count (Table 4.1). As implied
above, the extent of the morphological differences between the Australian Salmon is limited and subtle. Consequently, the taxonomic status of the three species has changed through time, where various workers have recognised the Australian Salmon as a single species, two to three subspecies, or, two to three full species (see Paulin 1993).

Table 4.1  Selected morphological and meristic measures of $n$ individuals of the four species of Arrpis. Data from Paulin (1993). Also see Figure 4.1.

<table>
<thead>
<tr>
<th></th>
<th>$A. trutta$</th>
<th>$A. truttaceus$</th>
<th>$A. xylabion$</th>
<th>$A. georgianus$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n^*$</td>
<td>34</td>
<td>20</td>
<td>32</td>
<td>7</td>
</tr>
<tr>
<td>Maximum length</td>
<td>890 mm</td>
<td>960 mm</td>
<td>850 mm</td>
<td>410 mm</td>
</tr>
<tr>
<td>Gill raker count†</td>
<td>9-13 + 20-24</td>
<td>7-11 + 16-17</td>
<td>12-17 + 20-25</td>
<td>16-18 + 28-32</td>
</tr>
<tr>
<td>Dorsal fin rays#</td>
<td>IX, 15-16</td>
<td>IX, 15-16</td>
<td>IX, 15-16</td>
<td>IX, 13-14</td>
</tr>
<tr>
<td>Anal fin rays§</td>
<td>III, 9-10</td>
<td>III, 9-10</td>
<td>III, 9-10</td>
<td>III, 10</td>
</tr>
<tr>
<td>Pectoral fin rays</td>
<td>16-18</td>
<td>16-18</td>
<td>16-18</td>
<td>16-18</td>
</tr>
<tr>
<td>Lateral line scales</td>
<td>49-53</td>
<td>49-56</td>
<td>50-55</td>
<td>54-59</td>
</tr>
<tr>
<td>Vertebrae$§$</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Caudal lobe$^*$</td>
<td>28.2%</td>
<td>26.2%</td>
<td>34.1%</td>
<td>29.3%</td>
</tr>
</tbody>
</table>

Modern approaches to species delimitation typically include genetic evidence (along with morphological evidence) because morphological data are not always sufficient for species recognition. For example, some species are polytypic with considerable morphological variation (e.g. Hull et al. 2010), while, on the other hand, cryptic species may be morphologically indistinguishable (e.g. Ross et al. 2010). Having said this, it is
Figure 4.1  Graphs showing the range of measurements for the three diagnostic characters in Australian Salmon – *Arripis trutta* (black), *A. truttaceus* (grey), *A. xylabion* (stippled). Bars show the range of measurements in each species for: **A.** number of upper gill rakers; **B.** number of lower gill rakers; and **C.** the length of the upper caudal lobe as a percentage of standard length (mean shown as vertical bar). Data from Paulin (1993). See also Table 4.1.
Chapter 4. Molecular Taxonomy

also important to recognise that genetic data do not invariably allow for the unambiguous resolution of species boundaries due to a range of problems, such as incomplete lineage sorting (e.g. Avise and Ball 1990; Knowles and Carstens 2007), hybridisation and introgression (e.g. Mallet and Willmott 2003) and homoplasy (e.g. Xia 2009). In addition, a gene tree might not reflect the species tree, especially when only a single gene is used (e.g. Mallet and Willmott 2003).

Genetic (and morphological) approaches to species delimitation are to a certain extent plagued by the controversies surrounding the appropriate conceptual definition of a species and the operational criteria for recognising the species (Sites and Marshall 2004). Rather than overtly committing to a particular species concept, many recent studies in molecular taxonomy, including the present chapter, have adopted part of the approach of de Queiroz (1998) and others (e.g. Mayden 1999; Hey 2006a), whereby any of the properties acquired by lineages during divergence (e.g. reciprocal monophyly, fixed allelic differences) are potentially informative and multiple pieces of supportive evidence are required.

Monophyly, or some other form of genetic exclusivity (where all members of a species form a group, to the exclusion of all members of other species), is probably the most widely used genetic criterion for species delimitation (reviewed in Sites and Marshall 2004). This criterion is broadly acceptable because lineages are expected to reach this condition only after a relatively long period of isolation (lineage sorting) (see Avise and Ball 1990; Knowles and Carstens 2007). In this sense, genetic exclusivity provides a relatively conservative test of species boundaries. In addition, it is applicable to a broad range of taxa (Knowles and Carstens 2007). Monophyly is often tested using mtDNA
gene trees, partly because a mtDNA locus will generally reach monophyly faster than a nDNA locus due to it smaller effective population size (Avise and Ball 1990). However, phylogenetic reconstruction based on a single locus is potentially misleading and there is consensus that genetic methods of species delimitation should incorporate loci from both mitochondrial and nuclear genomes (e.g. Sites and Marshall 2004; de Queiroz 2007; Knowles and Carstens 2007). Other types of genetic properties that are commonly used to distinguish species include fixed allelic differences or differences in allele frequencies (e.g. Paine et al. 2007) and/or the presence of interspecific differences that are substantially greater than intraspecific differences (e.g. Ward 2009).

The only previous attempt to use genetic data to address taxonomic questions within the Arripidae is in the PhD research of MacDonald (1980). This research used allozyme markers to address, inter alia, the species status of the east and west coast populations of Australian Salmon. The results supported the view that these populations represent different species because fixed allelic differences were found at two of 27 loci. However, MacDonald’s work: (i) is unpublished; (ii) did not include specimens from the ‘population’ to which the name *A. xylabion* is now applied (it pre-dated Paulin’s (1993) revision of the family); and (iii) was restricted to allozyme markers. Hence, there is a need for an analysis of the taxonomic status of the group that includes all three species of Australian Salmon and uses additional genetic markers.

4.1.1 Aim

The present chapter investigates whether the current, morphologically-based taxonomy of Australian Salmon (described above) is supported by DNA data. The molecular data
will be considered to support the current taxonomy of *Arripis* if there is reciprocal monophyly of species haplotypes in the mtDNA trees and if this monophyly is accompanied by other lines of supportive evidence (e.g. fixed allelic differences and/or significant allele frequency differences), especially in the nuclear genome. In addition, the interspecific genetic distances are expected to be substantially greater than intraspecific ones.

### 4.2 Materials and Methods

#### 4.2.1 Sampling Regime

The results of this chapter are based upon the analysis of samples of: *A. trutta* from four sampling sites (New South Wales, Victoria, Tasmania, New Zealand); *A. truttaceus* from five sampling sites ranging from Augusta, Western Australia to The Coorong, South Australia; *A. xylabion* from Lord Howe Island; and the outgroup species *A. georgianus* from six sampling sites ranging from Fremantle, Western Australia to The Coorong, South Australia (see Chapter 2 for details). The sample sizes for each of the sampled mtDNA and intron loci for each sampling site for each species are provided in Table 4.2. Given no evidence of population structure in any species (Chapter 3), individuals from different sites were pooled into a single sample for each species for the phylogenetic analyses. This included two individuals of *A. trutta* from New Zealand, which were pooled with the Australian samples of this species on the basis that the haplotypes and alleles in these fish were among the most common in *A. trutta* from Australia (Chapter 3, Section 3.4). Details of the sampling sites and the collection methods were provided in Chapter 2.
Table 4.2  Sample sizes for each gene used in the phylogenetic analyses of *Arripis*. The total (pooled) number of individuals used in the analyses is provided (bold), along with the sample sizes for each sampling site for each species of Australian Salmon (ingroup). Sample site abbreviations are the same as those used throughout the thesis and are defined in Chapter 2. To simplify presentation, only total sample sizes are given for the outgroup taxon *A. georgianus* (details of geographic sampling strategy are given in Chapter 2).

<table>
<thead>
<tr>
<th>Locus</th>
<th><em>Arripis trutta</em></th>
<th><em>Arripis truttaceus</em></th>
<th><em>Arripis xylabion</em></th>
<th><em>Arripis georgianus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EN</td>
<td>EV</td>
<td>ET</td>
<td>EZ</td>
</tr>
<tr>
<td>Mitochondrial Loci cytochrome <em>b</em></td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>COI</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Nuclear Loci</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>AldoB4</em></td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td><em>CaM4</em> *</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><em>CK7</em></td>
<td>28</td>
<td>28</td>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td><em>S72</em></td>
<td>29</td>
<td>25</td>
<td>28</td>
<td>2</td>
</tr>
</tbody>
</table>

* Data for *CaM4* are included here, foreshadowing their use in analyses in Chapter 5.
4.2.2 Genetic Assays

The mitochondrial DNA data used in this chapter were in the form of the nucleotide sequence of a 483 bp portion of the cytochrome \( b \) gene and a 654 bp portion of the cytochrome oxidase sub-unit 1 gene (\( COI \)) in \( Arripsi \). These data were generated using the methods described in Chapter 2. For each gene, all sequences were the same length (i.e. with no insertions or deletions) in all sampled individuals and there were no missing data (i.e. unresolved nucleotides). The nuclear DNA data used in this chapter were in the form of the frequency of length variants (alleles) in the introns of three genes, selected on the basis that they were polymorphic in at least one of the three species of Australian Salmon (see Table 3.2). These nDNA data were generated using the methods described in Chapter 2 and were also used to assess the population structures of each of these species in Australian waters, as described in Chapter 3.

4.2.3 Data Analyses

4.2.3.1 mtDNA Sequence Characteristics

The general characteristics of the sequences of each mitochondrial region were assessed using MEGA version 4 (Tamura \textit{et al.} 2007). Firstly, the average composition of the nucleotide sequence was calculated as a percentage of each base (A, T, C, G). Secondly, for each gene, the number of polymorphic sites along the aligned sequences was counted, along with the number of parsimony informative sites. Finally, the transition/transversion ratio was determined.
4.2.3.2  **Monophyly of mtDNA Sequences using Phylogenetic Analyses**

Phylogenetic approaches were used to assess monophyly of the mtDNA sequences of each species. Before providing details on these analyses, it is first necessary to address four issues that underlie the phylogenetic analyses.

4.2.3.2.1  **Choice of outgroup**

*Arrips georgianus* was chosen as the outgroup taxon to provide a reference point for resolving the relationships among the Australian Salmon (see Maddison *et al.* 1984). It was chosen because it is the only other extant member of the family (Paulin 1993), and diverged prior to divergences within the ingroup (MacDonald 1980). Specifically, concatenated sequence data (see Section 4.2.3.3.2) from a single individual *A. georgianus* (HC01, from Canal Rocks, Western Australia) were included in the mtDNA-based analyses, while the allele frequency data at the intron loci were from the individuals of *A. georgianus* assayed as a part of the population genetic structure analysis in Chapter 3.

4.2.3.2.2  **Concatenated sequence data**

The cytochrome *b* and *COI* sequences of each individual were concatenated. Concatenation was used here because the concatenated sequences were expected to provide a better phylogenetic signal by increasing the number of segregating sites (e.g. Cunningham 1997a; 1997b; Ballard *et al.* 1998; Gadagkar *et al.* 2005). To assess the validity of combining the data for the two regions of the mtDNA a partition-
homogeneity test (or incongruence length difference test; Farris et al. 1994; 1995), was implemented in PAUP* using a branch and bound search over 1000 replicates. This test evaluated whether the two gene regions contained similar phylogenetic signals by measuring whether the likelihood (of a tree) increased when loci were combined, compared to random data partitions. Since there was no incongruence in the signal between the cytochrome b and COI data ($p = 1.00$), the sequences were concatenated without partitions (total 1137 bp) for phylogenetic analyses. Throughout this chapter, the concatenated mtDNA sequence of an individual is referred to as its haplotype.

4.2.3.2.3 Substitution saturation

Tests for substitution saturation in the concatenated mtDNA sequence data were conducted among all haplotypes in all species because multiple substitutions at a site, especially backward substitutions may create a ‘saturated’ site and determining the relatedness between sequences becomes problematic due to homoplasy (Nei and Kumar 2000; Xia et al. 2003). Using the entropy method in the software DAMBE (Xia et al. 2003), an index ($I_{ss}$) was calculated whereby the mean entropy (uncertainty due to randomness) among sites was compared against the expected entropy under full sequence saturation (Xia 2009). A two-tailed $t$-test was used to determine whether the $I_{ss}$ was significantly smaller than the critical index value ($I_{ss,c}$) at which sequences began to fail to recover the tree (Xia 2009). The results did not reveal any evidence of saturation among concatenated mtDNA sequences of the four species of *Arripis* ($I_{ss} = 0.05 < I_{ss,c} = 0.77; p = 0.00$).
4.2.3.2.4 Substitution model

The substitution model of Tamura & Nei (1993) with a gamma distribution correction, G (Uzzell and Corbin 1971), was determined to be the most appropriate for use in the phylogenetic analyses with the concatenated mtDNA sequence data. This model (TrN + G) was selected using concatenated mtDNA sequence data for all haplotypes for all species using the Akaike Information Criterion (AIC; Table 4.3) and the Bayesian Information Criterion (BIC; results not shown), as implemented in JMODELTEST version 0.1.1 (Guindon and Gascuel 2003; Posada 2008). Tamura & Nei’s (1993) substitution model allows for different transversion and transition rates, and also makes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Selected</td>
<td><strong>TrN + G</strong></td>
</tr>
<tr>
<td>Freq [A]</td>
<td>0.2441</td>
</tr>
<tr>
<td>Freq [C]</td>
<td>0.3024</td>
</tr>
<tr>
<td>Freq [G]</td>
<td>0.1616</td>
</tr>
<tr>
<td>Freq [T]</td>
<td>0.2920</td>
</tr>
<tr>
<td>R [AC]</td>
<td>2.4791</td>
</tr>
<tr>
<td>R [AG]</td>
<td>24.0650</td>
</tr>
<tr>
<td>R [AT]</td>
<td>1.4160</td>
</tr>
<tr>
<td>R [CG]</td>
<td>0.2171</td>
</tr>
<tr>
<td>R [CT]</td>
<td>14.7458</td>
</tr>
<tr>
<td>R [GT]</td>
<td>1.0000</td>
</tr>
<tr>
<td>G</td>
<td>0.2718</td>
</tr>
</tbody>
</table>

Table 4.3 Output of JMODELTEST showing the values of relevant parameters for the Tamura-Nei (with gamma correction) substitution model based on Akaike Information Criterion (AIC) on the concatenated mtDNA data set using all haplotypes for all species of *Arripis*. The frequency of each of the bases (Freq), values for the rate matrix (R) and the gamma shape correction value (G) are provided.
a distinction for transition rates between purines and between pyrimidines, while the gamma distribution corrections account for substitution rate variation among codon positions (Uzzell and Corbin 1971). In phylogenetic analysis, it is important to apply an appropriate substitution model because otherwise various changes within a sequence, such as backward or parallel substitutions and unequal rates of substitution, may complicate the interpretation of homology (Nei and Kumar 2000).

4.2.3.2.5 Phylogenetic analyses

Having established the preliminary parameters and assumptions above, the phylogenetic relationships between the three species of Australian Salmon, based on the 1137 bp concatenated mtDNA sequences were examined by four methods: (i) distance (phenetic); (ii) maximum parsimony; (iii) maximum likelihood; and (iv) Bayesian inference (Felsenstein 1988; Nei and Kumar 2000). Although used to assess monophyly in the present chapter, all methods ultimately represent phylogenetic relationships in a tree (and are used for such in Chapter 5). Branch tips of the tree represent taxa and the patterns of branching within the tree represent relationships between the taxa (Swofford et al. 1996). All methods were included because each has strengths and weaknesses, as discussed below.

4.2.3.2.5.1 Distance-based method

A matrix of pairwise TrN + G distances between haplotypes of all species of Australian Salmon was generated in PAUP*. From this matrix, a 50% majority rule consensus neighbour-joining tree, based on 1000 bootstrap replicates, was constructed with A.
georgianus as the outgroup. Although phenetic/distance-based measures have largely been superseded (Felsenstein 1988; Rohlf et al. 1990; Li and Gouy 1991; Nei 1991; Hillis et al. 1994), a neighbour-joining analysis was included in this study because it is believed to produce reliable results when sequences are similar (which is the case for the Arrapis data) and because it is based on very few assumptions (Saitou and Nei 1987; Swofford et al. 1996; Nei and Kumar 2000).

Support for neighbour-joining trees was assessed by bootstrapping, which provided the frequency (or, bootstrap value) with which a particular node appeared in 1000 replicates using a subset of characters (nucleotides) from the original data matrix (Felsenstein 1985). Thus, this method resampled characters rather than taxa (Felsenstein 1988). Since bootstrapping generated numerous resampled trees, a majority rules consensus tree was constructed to include nodes that were present in ≥ 50% of all generated trees (Felsenstein 1988).

4.2.3.2.5.2 Maximum parsimony

Maximum parsimony analyses were conducted using unweighted parsimony as implemented in the software PAUP* with the three Australian Salmon as the ingroup and A. georgianus as the outgroup. A maximum parsimony consensus tree, based on 1000 bootstrap replicates, was generated by a heuristic search over 1000 replicates and tree bisection and reconnection branch swapping (see below).

Although maximum parsimony analysis makes use of informative characters only; in this case, derived nucleotide substitutions that were shared among a particular subset of
Arripis haplotypes (i.e. synapomorphies; Hennig 1965; Fitch 1977; Swofford et al. 1996), it was included in this study because it is free from many of the assumptions of other methods such as a particular substitution model (Nei and Kumar 2000) (although not entirely assumption-free; e.g. branch lengths are assumed to be equal, Lewis 2001). Also, when sequence divergence is low (as in the Arripis data), the rate of nucleotide substitution is approximately constant and, when a large number of nucleotides is examined, maximum parsimony methods often perform better than distance methods (Sourdis and Nei 1988; Nei 1991).

In this analysis, the length of each possible tree topology was calculated as the minimum number of substitutions over all sites, and the topology with the shortest tree length was considered the best, or most parsimonious tree (Nei and Kumar 2000). The tree bisection and reconnection (TBR) method was used to generate possible tree topologies because it is the most extensive method available in PAUP* (Swofford and Sullivan 2009) and among the most commonly used true heuristic searches.

4.2.3.2.5.3 Maximum likelihood

Maximum likelihood analyses (Felsenstein 1973; 1981), were conducted using the TrN + G model, as implemented in the software PAUP*, with the three Australian Salmon as the ingroup and A. georgianus as the outgroup. Values for base frequencies, rate matrix and gamma shape values were specified from the JMODELTEST output (Table 4.3). A maximum likelihood consensus tree, based on 1000 bootstrap replicates, was generated by a heuristic search and TBR branch swapping.
Maximum likelihood was used in this study because it is a consistent estimator of tree topologies with a lower variance than other methods (Felsenstein 1973; Swofford et al. 1996; Nei and Kumar 2000). This method also utilized the whole sequence, including singletons, rather than just ‘informative’ sites (as in maximum parsimony). In addition, the maximum likelihood analysis explicitly allowed for heterogeneity in branch lengths whereas that based on maximum parsimony implicitly assumed equal branch lengths (Lewis 2001).

Maximum likelihood analysis was used to find the tree topology with the highest likelihood of observing the given data (Nei and Kumar 2000). Starting with a random unrooted tree topology (based on the data), the tree topology was changed incrementally and the likelihood (average likelihood of each nucleotide arising independently at each position in the sequence for each taxon; Felsenstein 1981) for the new tree calculated. When the likelihood of the new tree was greater, the change was accepted and the new tree became the basis for further changes (Felsenstein 1973).

4.2.3.2.5.4 Bayesian inference

A 50% majority-rule consensus tree was constructed using Bayesian inference, as implemented in the software MRBAYES version 3.1.2 (Ronquist et al. 2009), with the three Australian Salmon as the ingroup and _A. georgianus_ as the outgroup. Using the general time reversible model with gamma corrections (GTR + G; Rodriguez et al. 1990), default Markov Chain Monte Carlo (MCMC) settings (one cold and two heated chains, heating temperature = 0.2) were run for one million generations with a sampling frequency of 100 generations. The first 25% of samples and trees were discarded as
burn-in as recommended (Ronquist et al. 2005). Support for the consensus tree was summarised by the proportion of credible trees exhibiting the most common node splits (i.e. posterior probabilities; Ronquist et al. 2009). The MCMC process (see below) visits trees in proportion to their posterior probability, and so this value is a natural measure of nodal support, i.e. a true probability tree (Lewis 2001; Ronquist et al. 2009).

Bayesian inference was included in addition to the maximum likelihood method because it approaches the question from a different direction - by determining the probability of the hypothesis given the data, rather than the probability for the data given the hypothesis (Nei and Kumar 2000; Lewis 2001). Thus, it is the probability of $x$ occurring, conditioned on $y$ (Huelsenbeck et al. 2002; Ronquist et al. 2009), where the posterior probability of a tree, $x$, is related to the likelihood of the observed data which is determined by the prior probability of the tree, $y$ (such as the GTR + G model of substitution).

The posterior probability is proportional to the product of the prior probability and the likelihood (Ronquist et al. 2009). The strength of the estimate of prior probability is largely dependent on the substitution model (Lewis 2001; Ronquist et al. 2009). The GTR + G model, rather than the TrN + G model, was used in this analysis because: i) the AIC scores for the two models were almost identical; and ii) MRBAYES does not incorporate the TrN model, and the GTR is the most parameterised alternative it offers and is the recommended default model (Tavare 1986; Ronquist et al. 2005).

Unlike the previous methods, a single optimal tree is not possible in Bayesian analyses. Instead it used a MCMC simulation to sample trees in proportion to their likelihood and
generate a sample of credible trees that was used to create a consensus tree (Huelsenbeck et al. 2002). In a chain-like fashion, MCMC searched for peaks on the likelihood surface, where each point on the surface represented a likelihood value for a given hypothetical tree. By randomly moving across the likelihood surface and accepting the move if likelihood improved, chains ascended likelihood peaks to identify localised optima (Ronquist et al. 2009). To cover a greater area of the likelihood surface, the Metropolis coupled-MCMC algorithm was used, with several chains simultaneously searching and randomly switching positions (Ronquist et al. 2005).

4.2.3.2.6 Parsimony networks

A parsimony (haplotype) network was constructed with default settings (95% confidence) in the software TCS version 1.21 (Clement et al. 2000) using the concatenated mtDNA sequence data. This method evaluated the minimum number of substitutions required to connect parsimoniously any two haplotypes. A network was constructed by firstly linking sequences with the smallest number of differences and then connecting to sequences with increasing numbers of differences. The original concept of parsimony networks (Templeton et al. 1992; 1995) was rooted in intraspecific phylogeographic inference, particularly nested-clade analysis (NCA; Templeton 1998), much of which is now out of favour (e.g. Knowles 2008; Petit 2008). Despite this, parsimony networks provide a useful method for displaying relationships between haplotypes of a single species and also for interpreting interspecific relationships when the taxa are closely related (Templeton 2004).
4.2.3.3  *Fixed Differences Between mtDNA Sequences*

Fixed nucleotide sites among aligned concatenated mtDNA sequences were identified using MEGA. In this chapter, a fixed site is defined as a site with a nucleotide common to all sequences in one species, but with a different nucleotide(s) at the site in the other species of Australian Salmon involved in the comparison (Nei and Kumar 2000).

4.2.3.4  *Genetic distances based on mtDNA sequences*

A matrix of pairwise $D_A$ distances (Nei *et al.* 1983) based on TrN + G substitution model between the different mtDNA haplotypes was generated in PAUP* version 4.0b10 (Swofford 2005). The corrected chord distance $D_A$ was used because it is not biased by low-frequency haplotypes and low sample size (Nei *et al.* 1983).

4.2.3.5  *Fixed Differences in nDNA Genotypes*

A table of nDNA allele frequencies (Table 3.2) was assessed for fixed allele differences, defined as an allele present at a locus in one species, but not found at that locus in either of the other species of Australian Salmon (Nei and Kumar 2000).

4.2.3.6  *Differentiation in nDNA Allele Frequencies*

Exact probability tests of differentiation ($G$-tests) were used to test for the presence of statistically significant allele frequency differentiation at each intron locus between each pair of species of Australian Salmon. These tests used a Markov chain, as implemented...
in GENEPOP (Raymond and Rousset 1995a), to estimate the exact probability of a Type I error. This type of test was used because it is not biased by rare alleles (Raymond and Rousset 1995b), which were widespread in the *Arripis* data (see Results). In addition, using GENEPOP, a Fisher’s exact test was used to test for evidence of differentiation across loci for each pairwise combination of the three species of Australian Salmon.

4.2.3.7  *Factorial Correspondence Analysis using Multilocus nDNA Genotypes*

A factorial correspondence analysis (FCA) was performed on the multilocus genotypes of individuals using the three intron loci that were polymorphic in at least one of the species of Australian Salmon (*AldoB4*, *CK7*, *S72*). The FCA was conducted in GENETIX, as described in Section 3.2.3.6.

4.3  *Results*

4.3.1  *Mitochondrial DNA*

The average base composition of the mtDNA sequences was similar for all species of Australian Salmon (Table 4.4). It was slightly biased towards adenine and thymine, as has been found for the mtDNA in other teleosts (Saccone et al. 1999; Ward et al. 2005). Alignment of the cytochrome *b* sequences from the different individuals of the three species of *Arripis* revealed 16 haplotypes and a relatively low level of variability, with 17 of the total of 483 sites variable among species (Tables 4.4, 4.5). Of the 17 variable
Table 4.4  Selected characteristics of the 483 bp partial cytochrome \(b\) sequences and 654 bp partial \(COI\) sequences in three species of Australian Salmon. For each gene region, the number of individuals assayed, the number of haplotypes, the average base composition of the sequence (expressed as percentage), the number of polymorphic sites (i.e. where nucleotide differs between sequences), the transition /transversion (\(Ts / Tv\)) ratio, and the codon positions at which the polymorphic sites are found, are presented. Data incorporating sequences from all species are indicated in bold. See Table 4.5 for full details of polymorphic sites.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Species</th>
<th>Number sequenced</th>
<th>Number of haplotypes</th>
<th>% T</th>
<th>% C</th>
<th>% A</th>
<th>% G</th>
<th>Number of polymorphic sites</th>
<th>Ts / Tv ratio</th>
<th>Codon position ((1 / 2 / 3))</th>
</tr>
</thead>
<tbody>
<tr>
<td>cytochrome (b)</td>
<td>(A.\ trutta)</td>
<td>25</td>
<td>11</td>
<td>28.1</td>
<td>32.9</td>
<td>25.3</td>
<td>13.7</td>
<td>11</td>
<td>10 / 1</td>
<td>2 / 1 / 8</td>
</tr>
<tr>
<td></td>
<td>(A.\ truttaceus)</td>
<td>36</td>
<td>4</td>
<td>28.0</td>
<td>33.1</td>
<td>25.5</td>
<td>13.5</td>
<td>3</td>
<td>3 / 0</td>
<td>0 / 0 / 3</td>
</tr>
<tr>
<td></td>
<td>(A.\ xylabion)</td>
<td>11</td>
<td>1</td>
<td>28.0</td>
<td>33.1</td>
<td>25.7</td>
<td>13.3</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>72</td>
<td>16</td>
<td>28.0</td>
<td>33.1</td>
<td>25.4</td>
<td>13.5</td>
<td>17</td>
<td>16 / 1</td>
<td>2 / 1 / 14</td>
</tr>
</tbody>
</table>

| \(COI\)    | \(A.\ trutta\)    | 25               | 8                    | 31.1 | 27.2 | 24.9 | 16.8 | 8                           | 7 / 1          | 0 / 0 / 8                    |
|            | \(A.\ truttaceus\) | 36               | 2                    | 30.9 | 27.4 | 24.9 | 16.8 | 1                           | 1 / 0          | 0 / 0 / 1                    |
|            | \(A.\ xylabion\)  | 11               | 2                    | 30.7 | 27.5 | 25.0 | 16.8 | 1                           | 1 / 0          | 0 / 0 / 1                    |
|            | Total            | 72               | 11                   | 30.9 | 27.3 | 24.9 | 16.8 | 15                          | 14 / 1         | 0 / 0 / 14                   |
sites, almost all (i.e. 16/17) showed transitional changes and most were at third codon positions (Tables 4.4, 4.5), as would be expected for comparisons between closely related species (Nei and Kumar 2000). Likewise, the cytochrome oxidase sub-unit 1 (COI) gene, which was sequenced for all of the same individuals, revealed a total of 11 haplotypes (Tables 4.4, 4.5). The level of variability in these sequences was lower than that of cytochrome \( b \), with 15 of the total of 654 sites variable among species (Tables 4.4, 4.5). All of the polymorphic sites showed transitional changes at third codon positions (Tables 4.4, 4.5). The concatenated mtDNA sequence was 1137 base pairs long and resulted in 14 \( A. \ trutta \) haplotypes, five \( A. \ truttaceus \) haplotypes, and two \( A. \ xylabion \) haplotypes (Table 4.5). Alignment of the 72 concatenated mtDNA sequences from the different individuals of the three species of \textit{Arrips} revealed 32 polymorphic sites among species (Table 4.5).

\textbf{Table 4.5} \hspace{1cm} (\textit{over page}) The location and distribution of 32 polymorphic sites in the concatenated mtDNA sequences (483 bp portion of the cytochrome \( b \) gene and 654 bp portion of the COI gene) in three species of Australian Salmon. These polymorphic sites revealed a total of 21 haplotypes: 14 in \( A. \ trutta \); 5 in \( A. \ truttaceus \); and 2 in \( A. \ xylabion \) (Roman numerals). A dot indicates that the nucleotide at that position in a haplotype is the same as that indicated for Haplotype I. Fixed differences for each species, when present, are indicated in bold. Sample sizes (\( n \)) and the frequency of each haplotype (\( f \)) for each species are also provided. See Table 4.4 for a summary of selected characteristics of the sequences.
<table>
<thead>
<tr>
<th>Haplotype Number</th>
<th>Chromosome &quot;p&quot;</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**A. trutta (n = 25)**

<table>
<thead>
<tr>
<th>Codon position</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.48</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>II</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XI</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XII</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIII</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIV</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**A. trutta (n = 36)**

<table>
<thead>
<tr>
<th>Codon position</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>XV</td>
<td>0.89</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
<td>C</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XVI</td>
<td>0.03</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
<td>C</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>XVII</td>
<td>0.03</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td>C</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>XVIII</td>
<td>0.03</td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>XIX</td>
<td>0.03</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>G</td>
</tr>
</tbody>
</table>

**A. xylabion (n = 11)**

<table>
<thead>
<tr>
<th>Codon position</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>A</td>
<td></td>
<td>G</td>
<td></td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>XXI</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>A</td>
<td></td>
<td></td>
<td>C</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>

**A. xylabion (n = 11)**

<table>
<thead>
<tr>
<th>Codon position</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>A</td>
<td></td>
<td>G</td>
<td></td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>XXI</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>A</td>
<td></td>
<td></td>
<td>C</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>
4.3.1.1 Monophyly of mtDNA Sequences using Phylogenetic Analyses

Three of the four phylogenetic tree building methods using concatenated mtDNA sequences produced monophyletic terminal clusters. Each cluster included only the haplotypes of a single species, with nodal support above 70% in each case (Figure 4.2). The exception was maximum likelihood analysis in which a monophyletic clade for the two *A. xylabion* haplotypes was present in only 35% of bootstrap replicates (and therefore not represented in the 50% majority consensus tree in Figure 4.2C). In addition, the parsimony network showed complete monophyly of the concatenated mtDNA haplotypes of each species (Figure 4.3), with at least five mutational steps between the haplotypes of the different species (Figure 4.3). Similar results were found with each method when each gene was analysed separately (full data not shown).

---

**Figure 4.2** (over page) Majority consensus phylogenies of Australian Salmon inferred from concatenated mtDNA sequence haplotypes based on four methods. **A.** Neighbour joining (NJ) method using distances based on the TrN + G model. **B.** Maximum parsimony (MP; bootstrap values above line), utilising a heuristic search with TBR branch swapping. Bayesian inference (BI), using the GTR + G model generated the same tree (posterior probabilities below line). **C.** Maximum likelihood (ML), using the TrN + G model with a heuristic search and TBR branch swapping. Clades are colour coded by species. Haplotypes are numbered as in Table 4.5. Node support values were derived from 1000 bootstrap replicates (NJ, MP, ML) or 15,000 sampled post burn-in trees (BI). In each tree, the congener *Arripsis georgianus* was the outgroup.
Figure 4.3  (over page) Parsimony network for the mtDNA haplotypes from the concatenated sequence data (cytochrome b and COI) of the three species of Australian Salmon obtained by the criterion of parsimony with the program TCS version 1.21 (Clement et al. 2000). Each line in the network shows a single mutational change. A haplotype is shown by a coloured circle, the surface area of which is proportional to the frequency of the haplotype. Intermediate, unsampled haplotypes are indicated by a small black dot. Similar networks are shown for each of the genes separately (inset). In each case, *Arripis trutta* is represented by yellow; *A. truttaceus* by pink; and *A. xylabion* by orange.
4.3.1.2 Fixed Differences Between mtDNA Sequences

The concatenated mtDNA sequences of all three species of Australian Salmon showed fixed differences at multiple sites, in pairwise comparisons with each of the other species (Tables 4.5, 4.6). In addition, five sites in *A. trutta* and four sites in *A. truttaceus* were monomorphic within the species but had a nucleotide that was not present at that site in the sequences any other species (Tables 4.5, 4.6).

**Table 4.6** Number of fixed site differences in the 1137 bp concatenated mtDNA sequences of each species of Australian Salmon. A fixed site had a nucleotide common to all sequences in a species, but a different nucleotide(s) in the other species involved in the comparison. Data are presented as fixed differences in each species (bold) relative to the other two species combined and fixed differences in pairwise comparisons between two species.

<table>
<thead>
<tr>
<th>Fixed Differences</th>
<th><em>A. trutta</em></th>
<th><em>A. truttaceus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. trutta</em></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><em>A. truttaceus</em></td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td><em>A. xylabion</em></td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

4.3.1.3 Intraspecific versus Interspecific Genetic Distances in mtDNA

Using concatenated mtDNA sequences, all pairwise interspecific distances, $D_A$, were low in absolute terms, but were nevertheless between 4 and 12 times greater than the respective intraspecific $D_A$ genetic distances (Table 4.7).
Chapter 4. Molecular Taxonomy

Table 4.7  Average pairwise $D_A$ genetic distances (and standard error) among $n$
different mtDNA haplotypes within each *Arripis* species (in italics along the diagonal)
and between each species, expressed as a percentage.

<table>
<thead>
<tr>
<th></th>
<th>$A. trutta$</th>
<th>$A. truttaceus$</th>
<th>$A. xylabion$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. trutta</em> ($n = 14$)</td>
<td>0.29 ± 0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>A. truttaceus</em> ($n = 5$)</td>
<td>1.30 ± 0.01</td>
<td>0.14 ± 0.01</td>
<td></td>
</tr>
<tr>
<td><em>A. xylabion</em> ($n = 2$)</td>
<td>1.06 ± 0.02</td>
<td>0.67 ± 0.02</td>
<td>0.09</td>
</tr>
</tbody>
</table>

4.3.2  Nuclear DNA

Details of the alleles and allele frequencies at each of the three intron loci for each of *A. trutta*, *A. truttaceus* and *A. georgianus* were provided in Chapter 3 (Section 3.3.1). Briefly, there was a total of 19 alleles in *A. trutta* and eight in *A. truttaceus*. In addition, four alleles were found in *A. xylabion* (Table 4.8).

No fixed allelic differences were identified at any nDNA locus, instead the most frequent allele at each locus in each species was the same, and the alleles present in each of *A. truttaceus* and *A. xylabion* were a subset of those in *A. trutta* (Table 4.8). However, the results of the exact tests indicate that the frequencies of alleles in pairwise comparisons of the species samples were invariably significantly different for each locus and for all loci combined (Table 4.8).

Factorial correspondence analysis based on multilocus genotypes revealed a tendency for individuals from each species to cluster together, although there was overlap among all three species (Figure 4.4). The individuals with overlapping genotypes represented every sampling site for each species in similar proportions (data not shown).
Table 4.8 Allele frequencies at each of three intron loci for samples of three species of *Arripis* and probability values for exact tests. The number of alleles sampled (n) is provided and a dash (-) indicates that the allele was not recorded in the species. The probability (p) ± standard error were determined by exact tests against a null hypothesis that allele frequencies for species in the respective pairwise comparison were the same. Pairwise comparisons, using combined data from all three loci, are provided as chi-square tests ($\chi^2$) with the relevant degrees of freedom (df).

<table>
<thead>
<tr>
<th>Locus/Allele</th>
<th>A. trutta</th>
<th>A. truttaceus</th>
<th>A. xylabion</th>
<th>A. trutta/ A. truttaceus</th>
<th>A. trutta/ A. xylabion</th>
<th>A. truttaceus/ A. xylabion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AldoB4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>170</td>
<td>172</td>
<td>178</td>
<td>182</td>
<td>185</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.011</td>
<td>0.775</td>
<td>0.011</td>
<td>0.187</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>0.673</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td></td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.000$</td>
<td>$p = 0.000$</td>
<td>$p = 0.000$</td>
<td>$\pm 0.0000$</td>
<td>$\pm 0.0000$</td>
<td>$\pm 0.0000$</td>
</tr>
<tr>
<td><strong>CK7</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>943</td>
<td>952</td>
<td>961</td>
<td>986</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.057</td>
<td>0.845</td>
<td>0.092</td>
<td>0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>1.000</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.000$</td>
<td>$p = 0.007$</td>
<td>$\pm 0.0000$</td>
<td>$\pm 0.0000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>S72</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>816</td>
<td>869</td>
<td>892</td>
<td>895</td>
<td>900</td>
<td>914</td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.143</td>
<td>0.565</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>0.007</td>
<td>-</td>
<td>0.127</td>
<td>0.663</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>0.762</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = 0.011$</td>
<td>$p = 0.019$</td>
<td>$p = 0.009$</td>
<td>$\pm 0.0002$</td>
<td>$\pm 0.0003$</td>
<td>$\pm 0.0001$</td>
</tr>
<tr>
<td><strong>All three loci</strong></td>
<td>$\chi^2 = \infty$</td>
<td>$\chi^2 = 35.027$</td>
<td>$\chi^2 = \infty$</td>
<td>$df = 6$</td>
<td>$df = 6$</td>
<td>$df = 4$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.000$</td>
<td>$p = 0.000$</td>
<td>$p = 0.000$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.4  Factorial correspondence analysis (FCA) of multi-locus genotypes for the three species of Australian Salmon. The total variation explained by each axis is given as a percentage. *Arripis trutta* is represented by squares (yellow); *A. truttaceus* by diamonds (pink); and *A. xylabion* by circles (orange). Note, many points overlap and are obscured.

4.4  Discussion

With multiple lines of evidence, the molecular data support the current morphologically-based taxonomy of Australian Salmon (Paulin 1993). Specifically, there was reciprocal monophyly in the mtDNA gene trees, accompanied by fixed allelic differences in mtDNA sequences in all pairwise comparisons and evidence of differences in allele frequencies in the nuclear genome.

Monophyly in the mtDNA gene trees probably represents the best single piece of evidence of the species status of the three Australian Salmon. This is because this
evidence suggests that these taxa have been isolated for sufficient time for lineage sorting of these genes to have reached completion (see Avise and Ball 1990; Baum and Shaw 1995; Knowles and Carstens 2007). However, tree-based approaches using reciprocal monophyly for species delimitation can be limited by the fact that even highly supported clusters may not belong to a distinctive taxonomic unit, but result from chance alone (Rosenberg 2007). In order to address this, a probability test of taxonomic distinctiveness based on coalescence (a test of the null hypothesis that monophyly is a chance outcome of random branching of lineages within a species; Rosenberg 2007) was applied to each of the species of Australian Salmon. The results of these tests indicated that random branching patterns were strongly rejected as the cause of monophyly in any species ($p \leq 10^{-13}$ in all three species; data not shown). On this basis, reciprocal monophyly of the three species of Australian Salmon in the mtDNA trees almost certainly arose by evolutionary processes and provides strong support for the view that they represent true species.

Monophyly in the mtDNA sequences is also reflected in the presence of multiple fixed nucleotide differences in pairwise comparisons of the species of Australian Salmon. Such fixed differences are expected to accumulate through time, and Hey (1991) has shown that fixed differences are so unlikely to appear within an interbreeding population that the mere presence of them can be considered statistically significant.

Monophyly in the mtDNA trees, which were based on what is effectively a single locus, was not accompanied by genetic exclusivity in the nDNA data. This may reflect the point in time in the process of divergence and the fact that lineage sorting at the intron loci is incomplete. This discrepancy is likely to have arisen because the rate of genetic
Chapter 4. Molecular Taxonomy

drift at a mtDNA locus is expected to be faster than that at a nDNA locus due to a lower effective population size in the former (Brown et al. 1979; Avise 1986; see Chapter 5). Although there was no overlap at all with mtDNA markers and it is a reasonable expectation for lineage sorting at the EPIC markers to proceed more slowly due to a larger effective population size, it is important to acknowledge the difficulty in separating ancestral polymorphism from interspecific gene flow resulting from hybridisation (introgression), especially in recently diverged species (e.g. Mallet 2005; Baack and Rieseberg 2007; Roberts et al. 2010). Nevertheless, there was evidence of allele frequency differences in the nDNA, and so evidence of genetic divergence was not limited to a single (i.e. mtDNA) genome.

As expected, the interspecific distances in the three *Arripis* species were substantially greater than intraspecific distances, and is taken here as part of the accumulated molecular evidence supporting the current taxonomy of *Arripis*. Although fixed cut-off thresholds may be useful as a ‘first step’ in delimiting species (e.g. Hebert et al. 2004; Tavares and Baker 2008; Ward 2009), it is generally recognised that a standard numerical threshold cannot be universally applied (Ferguson 2002; Moritz and Cicero 2004; Meyer and Paulay 2005; Ward et al. 2005). However, interspecific genetic distances smaller than that found in *Arripis* in this study have been reported between closely related species of fishes and birds (e.g. Johnson and Cicero 2004; Radchenko et al. 2008; Ward 2009).

The genetic differences between the species of Australian Salmon, although not exceptionally high, contrast with the subtle morphological differences, where few diagnostic characters exist (Paulin 1993; see Section 4.1). Paulin (1993) provided
detailed morphological data for some 24 characters, of which only three differed sufficiently between the species of Australian Salmon to be useful for phenetic diagnosis. Thus, in all species of *Arripis*, the rate of morphological change appears to be constrained, relative to the rate of molecular evolution. In general, the process of substitution within the genome is continuous but, for many reasons, is disassociated with morphological change (Bromham *et al.* 2002), which may progress much slower (e.g. Inoue *et al.* 2005) or faster (e.g. Bromham and Penny 2003). The reasons for morphological stasis in *Arripis* are unknown, but may reflect functional, mechanical and physiological limitations imposed by a pelagic habit (see Barton 2007).

4.4.1 Conclusions

The molecular data in this chapter, which include information from both mitochondrial and nuclear loci, support the *Arripis* taxonomy of Paulin (1993). Given that ‘populations’ can be of varying evolutionary ages and form a continuum of genetic differentiation of which ‘species’ are a part (Goldstein *et al.* 2000; Mallet 2007), the decision of where the population/species boundary lies remains somewhat arbitrary (Avise 2000; Hey 2006b). Nonetheless, the three species clearly represent independently evolving lineages (*sensu* de Queiroz 2007) and as such, Paulin’s (1993) “educated nomenclatural judgements” (Avise 2000, p. 308) using morphological data are now supported by multiple lines of molecular evidence. Having established molecular support for the current taxonomy of Australian Salmon, the next step is to investigate the relationships among these species. This investigation forms the basis of Chapter 5.
5.0 Molecular Phylogenetics, Speciation and Biogeography

5.1 Introduction

This study has already demonstrated that: (i) each of *A. trutta* and *A. truttaceus* (and *A. georgianus*) is likely represented by a single population throughout its entire distribution in Australia (Chapter 3); and (ii) that these eastern (*A. trutta*) and western (*A. truttaceus*) populations likely represent different species (Chapter 4). However, the evolutionary relationships between these two species and a third species of Australian Salmon (*A. xylabion*) remain to be determined. The need to establish these relationships is inextricably linked with the development of hypotheses explaining speciation and biogeography in this group in particular and the region in general. The relationships are also important in the context of comparing the recent evolutionary histories of these closely related species, which is explored in Chapter 6.

As established in Chapter 1, *A. trutta* and *A. truttaceus* have long been considered as one example among many geminates (i.e. sister-species on either side of a barrier) in temperate Australian seas (e.g. Hutchins 1994). This is consistent with the view that the Bass Strait is a major contributor to speciation in the region, with the emergence of a land-bridge across the Bass Strait during glacial events, leading to subdivision of many faunal groups (Whitley 1932; Bennett and Pope 1953; Dartnall 1974; Womersley 1981; Hutchins 1994; O’Hara and Poore 2000). One of the most effective ways to test the (hitherto largely assumed) role of vicariance as a speciation process in this region is to assess the validity of putative geminate relationships. The presence of many east/west geminates should provide strong support for a hypothesis of vicariance (see Bowen and
Chapter 5. Molecular Phylogenetics

Avise 1990; Lessios 2008; Rocha and Bowen 2008). However, relationships between putative geminates should not simply be assumed. This is because genetic studies have shown that, in general, presumed geminates are not always each other’s closest relatives and that their divergence may not be associated with the emergence of a barrier (e.g. Jones et al. 2003; Craig et al. 2004; Marko and Moran 2009; Miura et al. 2010). The presence of a species pair across a barrier that are not geminates does not necessarily preclude a conclusion of vicariance, especially if the species are very closely related (see Johnson and Cicero 2004; Spencer et al. 2007), but it may require a more complex explanation.

The Bassian Isthmus is an ephemeral barrier that has emerged repeatedly, raising the possibility that vicariant speciation in the region has been ongoing (see Waters et al. 2004; Rocha et al. 2005; Robertson et al. 2006; Teske et al. 2007; Corrigan and Beheregaray 2009). This can be tested by estimating the timing of speciation in a range of geminate taxa in the region, with the expectation of finding a range of divergence times among them. In this way, applying a ‘molecular clock’ (e.g. Nei and Kumar 2000; Arbogast et al. 2002; Bromham and Penny 2003) to estimate divergence times in Arripis is an important first step in assessing the degree of concordance (or not) of divergence times among species.

5.1.1 Objective

The main objective of this chapter was to generate information on the relationships among the species of Australian Salmon and use the results to develop an explanation for the speciation and current biogeography of this group. This objective will be
addressed by firstly using molecular data to establish a phylogeny for the Australian Salmon and test whether the genetic evidence supports the view that *A. trutta* and *A. truttaceus* are sister-species. Secondly, the present chapter will use a molecular dating approach to estimate the age of species divergence events within the Australian Salmon.

5.2 Materials and Methods

5.2.1 Data

The mtDNA data used in this chapter were the same as those used in Chapter 4, namely a 1137 bp concatenated sequence consisting of a 483 bp portion of the cytochrome *b* gene and a 654 bp portion of the cytochrome oxidase sub-unit 1 gene (*COI*) in *Arripis*. The nDNA data used in this chapter were in the form of the frequency of length variants (alleles) in the introns of four genes – the three loci used in Chapter 4 (*AldoB4, CK7, S72*) and an additional locus (*CaM4*). Although *CaM4* was monomorphic among Australian Salmon (and thus, uninformative for determining species status in Chapter 4), it was polymorphic when the outgroup taxon (*A. georgianus*) was included. This locus was therefore included in this analysis because it provided greater character polarity for the calculation of a distance matrix (see Section 5.2.2.1.2). Both the mtDNA and nDNA data were generated using the methods described in Chapter 2. Total sample sizes for each species from samples across the entire range are provided in Table 4.2 and Table 6.2. As described in Section 4.2.3.2.1, *A. georgianus* was used as an outgroup for phylogenetic analyses.
Chapter 5. Molecular Phylogenetics

5.2.2 Data Analyses

5.2.2.1 Relationships Among Species of Australian Salmon

The summary statistics and underlying assumptions (e.g. substitution model, sequence saturation) for the concatenated mtDNA sequence data have already been determined and described in Chapter 4 (Section 4.3.1). Neighbour-joining (NJ), maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) trees and a parsimony (haplotype) network were generated for these data as described in Chapter 4, and have already been discussed in the context of questions about species status (see Section 4.2.3.2.5). However, these trees and the network are also presented in this chapter because they also include information about the relationships among species of Australian Salmon.

5.2.2.1.1 Testing between alternative tree topologies

Three hypothetical tree topologies are possible for the three species of Australian Salmon. Specifically, the sister-species may be either: (i) *A. trutta/A. truttaceus*; (ii) *A. trutta/A. xylabio*; or (iii) *A. truttaceus/A. xylabion*. The likelihoods of these competing topologies were compared by two different topology tests, as follows.

The first topology test was the Shimodaira and Hasegawa (1999) (SH) test, which used a resampling estimated log-likelihood (RELL, Kishino *et al.* 1990) method to compare between *a posteriori* selected trees. The null hypothesis for this test was that all trees were equally good explanations of the data. Thus each of four competing tree
topologies (i.e. the three hypothetical resolved trees and a polytomy) was constrained in PAUP* version 4.0b10 (Swofford 2005), and site-wise log likelihoods calculated (see Section 4.2.3.2.5.3). The tree with the highest likelihood was identified (in this case it was the polytomy), against which each of the other trees was compared in CONSEL version 0.1i (Shimodaira and Hasegawa 2001). Confidence in each of the trees, relative to the tree with the highest likelihood, was provided as a probability value, assessed as a one-sided test (Schmidt 2009). In both this and the following test, the null hypothesis was rejected if \( p < 0.05 \).

The second topology test was the approximately unbiased (AU) test (Shimodaira 2002), based on a multi-scale RELL method. It was used here because it corrects for tree selection bias that might affect the SH method (see Shimodaira 2002). The null hypothesis for this test was that the expected likelihood for any one tree was greater or equal to that expected for all trees (Shimodaira 2002). The multi-scale bootstrap method generated replicates for many different sequence lengths, some of which were smaller and some longer than the actual sequence, and site wise log-likelihoods were calculated for each replicate as described above. Probability estimates for the AU test were calculated as the change in bootstrap support across replicates as the length of the sequence changed through the RELL method (Shimodaira 2002).

5.2.2.1.2 Intron length polymorphism

The intron allele frequency data for four loci were also used to explore the relationships among the Australian Salmon species. Specifically, a matrix of pairwise \( D_A \) values (corrected chord distance; Nei et al. 1983) using allele frequency data of four intron loci
(see Section 5.2.1) in all species of *Arripis* was generated in DISPAN (Ota 1993). From this matrix, a 50% majority rule consensus neighbour-joining tree, based on 10,000 bootstrap replicates, was constructed with *A. georgianus* as an outgroup.

The corrected chord distance ($D_A$) was used because it is the most efficient distance measure in obtaining the correct topology under the assumption of Kimura and Crow’s (1964) infinite alleles model (Nei *et al.* 1983; Nei and Takezaki 1994), which is considered the most appropriate model for allele frequency data (Cornuet and Luikart 1996). Assuming the mutation-drift balance is maintained, $D_A$ is expected to increase linearly with evolutionary time under the infinite alleles model (Takezaki and Nei 1996). Furthermore, $D_A$ is not biased by low-frequency alleles and low sample size (Nei *et al.* 1983). Neighbour-joining (Saitou and Nei 1987) was used to construct the nDNA trees because it is the most broadly applicable in this context (Nei 1991), and performs better than other methods (e.g. UPGMA) when some populations have been subject to bottlenecks (Nei and Takezaki 1994).

### 5.2.2.2 Exploring the Phylogenetic Signal

Given that an *Arripis* phylogeny was not consistently or reliably resolved (see Results), methods were applied to the mtDNA data to determine if the phylogeny was more likely to be bifurcating (tree-like) or multifurcating (polytomy) or reticulated (network-like) and to highlight possible conflicts in the bifurcating trees.
Section 5.2.2.1  Testing for a polytomy

A zero-length branch test (Slowinski 2001) was used to test whether resolving the tree was a significant improvement relative to leaving the relationship as a polytomy. This test was conducted with PAUP* using the concatenated mtDNA sequence data, where each of the three possible resolved tree topologies were treated as equally likely and thus analysed separately. One by one, branches in the tree were collapsed to zero-length and the new tree compared to the original using a likelihood ratio test (LRT; Swofford and Sullivan 2009). The LRT determined the probability that the zero-length tree was significantly worse than the starting tree (Slowinski 2001; Swofford and Sullivan 2009).

Section 5.2.2.2  Likelihood mapping

A likelihood map was used to test whether the mtDNA data contained a tree-like signal (Schmidt and von Haeseler 2009) using TREE-PUZZLE version 5.2 (Schmidt et al. 2002) from all possible quartets under the TrN + G substitution model.

Maximum likelihood was used to derive posterior probabilities, $p$, of all possible quartets for each tree topology (a quartet is the smallest set of taxa for which more than one unrooted topology exists; Strimmer and von Haeseler 1996). The relative position of all $p$-values was plotted within an equilateral triangle, of which each apex represented one of the three possible tree topologies for that quartet. The resulting distribution of points within seven regions of the triangle (Figure 5.1) gives an indication of whether the phylogenetic signal is tree-like (points clustered near the apices), partly resolved network-like (points clustered along the edges) or, star-like (points clustered in the
centre). Schmidt and von Haeseler (2009) suggested that if 20% - 30% of points occur in the network-like and/or star-like areas then the phylogeny is unlikely to represent a bifurcating tree.

**Figure 5.1** The seven regions within a likelihood map. The three apices (areas 1, 2, 3) represent areas supporting strictly bifurcating trees. The sides (areas 4, 5, 6) represent areas where the decision between alternative trees is only partly resolved and the relationship might be network-like. The centre (area 7) represents points where all trees are equally supported, suggesting a star-like signal.

5.2.2.2.3 *NeighborNet (Split Networks)*

A NeighborNet splits graph was used to provide an assessment of possible conflict among competing bifurcating branching patterns in a phylogeny. The splits graph was generated using SPLITSTREE4 version 4.10 (Huson 1998; Huson and Bryant 2006), based on uncorrected $P$-distances (= Hamming) with the equal-angle algorithm. Support for edges (see below) was determined as the number of times an edge was supported in 10,000 bootstrap replicates.
NeighborNet was used in this study because it uses neighbour-joining (Saitou and Nei 1987) and least squares estimates to avoid the conservative biases of traditional split decomposition methods (Bandelt and Dress 1992; Bryant and Moulton 2004; Moulton and Huber 2009), and is one of the best network methods for assessing phylogenies that are poorly resolved by traditional methods (e.g. Kennedy et al. 2005; Morrison 2005; Wolf et al. 2007; Leaché et al. 2009; Wägele et al. 2009; Willerslev et al. 2009).

Splits are pairs of sub-trees, or bipartitions, created by the removal of a branch from a full tree and a collection of splits equates to the branches of a phylogenetic tree if every pair of splits in the collection is compatible (Moulton and Huber 2009). A NeighborNet splits graph for the Arripis data was generated as a network of compatible splits, where each branch (or edge) reflected a split with a length equal to the weight of the split (Bryant and Moulton 2004). Incompatible splits were represented as reticulations or boxes and illustrated that there were several possible paths between taxa. Thus, each split in the splits graph corresponded to a collection of parallel edges, all with the same length. Empirical support was provided by bootstrapping (Huson and Bryant 2006).

5.2.2.3 Molecular Dating

5.2.2.3.1 Testing for a clock-like rate of substitution

In order to determine the appropriate method for estimating divergence times, it was necessary to test for a uniform rate of nucleotide substitution in the concatenated mtDNA sequences of all Arripis lineages. This was done using a likelihood ratio test (LRT) for the assumption of a molecular clock (Felsenstein 1981; Huelsenbeck and
Chapter 5. Molecular Phylogenetics

5.2.2.3.2 Divergence rate

In this chapter, a rate of 2% sequence divergence per million years between lineages (Brown et al. 1979) was used in order to estimate divergence times between the species of *Arripsis*. This rate was used because it is widely applied to mitochondrial genes (including cytochrome *b* and *COI*) as a ‘standard’ rate when no other accurate estimates (from fossils or related extant taxa) are possible (e.g. Burridge 2002; Theisen et al. 2008; Drew et al. 2009), as was the case for *Arripsis* (see Discussion).

5.2.2.3.3 Time to most recent common ancestor (*T_{mrca}* )

Molecular dating was performed on the concatenated mtDNA sequence data using Bayesian Markov Chain Monte Carlo (MCMC) resampling with the program BEAST version 1.4.8 (Drummond and Rambaut 2007). Based on the results of the tests for a uniform rate of substitution across all lineages (see Section 5.2.2.3.1), a strict clock was
specified. A Yule speciation process was specified as a tree prior along with model parameters derived from JMODELTEST (see Table 4.3). The MCMC analyses were run for 10 million generations sampling every 100 generations with the first 10% discarded as burn-in. Given limited support for one topology over another (see Results), each of the three possible tree topologies (representing three different sister-species relationships among the Australian Salmon) was analysed separately. In each analysis, a topology was constrained so the MCMC search was limited to trees satisfying a particular sister-species relationship between the three species of Australian Salmon. Results were inspected using TRACER version 1.4.1 (Rambaut and Drummond 2007) to confirm that stationarity and adequate effective sample sizes had been obtained for all parameters indicating that the analysis had run for a sufficient number of generations. Confidence was provided by 95% highest posterior density (HPD) intervals, which is a credible set of trees containing 95% of the sampled values (Drummond and Rambaut 2007).

5.3 Results

As discussed in Chapter 4, the nucleotide composition of a 1137 bp concatenated mtDNA sequence was determined for 25, 36 and 11 individuals of *A. trutta*, *A. truttaceus* and *A. xylabion*, respectively, and revealed the presence of 14, five and two haplotypes, respectively, and 32 polymorphic sites among species (Table 4.4). *Arripis trutta* sequences contained five fixed sites, *A. truttaceus* four and no completely fixed differences were found in *A. xylabion* (Table 4.5), although there were fixed differences in this species in pairwise comparisons (see Section 4.3.1).
5.3.1 Relationships Among Species of Australian Salmon

The phylogenetic trees and the parsimony network did not provide consistent support for a sister-species relationship between *A. trutta* and *A. truttaceus* (Figures 5.2, 5.3). For the mtDNA data, all phylogenetic (tree-building) methods resulted in either unresolved topologies or resolved topologies with very low support for interspecific relationships (Figure 5.2; Table 5.1). In three of the four methods (maximum parsimony, maximum likelihood, Bayesian inference), the relationship was represented as a trifurcating polytomy with each species arising from a single node (Figure 5.2B, C). Only neighbour-joining reconstructed a resolved consensus tree (with *A. trutta* and *A. truttaceus* as sister-species), although bootstrap support was very low (Figure 5.2A). Note that similar results were also found using each gene region independently in that, even if a tree was resolved, bootstrap support was low (Table 5.1, details not shown).

The neighbour-joining phylogenetic tree constructed using the nDNA allele frequency data indicated a sister-species relationship between *A. truttaceus* and *A. xylabion*, however, bootstrap support was very low (Figure 5.4).
Figure 5.2  (over page) Majority consensus phylogenies of Australian Salmon inferred from concatenated mtDNA sequence haplotypes based on four methods.  **A.** Neighbour joining (NJ) method using distances based on the TrN + G model.  **B.** Maximum parsimony (MP; bootstrap values above line), utilising a heuristic search with TBR branch swapping.  Bayesian inference (BI), using the GTR + G model generated the same tree (posterior probabilities below line).  **C.** Maximum likelihood (ML), using the TrN + G model with a heuristic search and TBR branch swapping.  Clades are colour coded by species.  Haplotypes are numbered as in Table 4.5.  Node support values were derived from 1000 bootstrap replicates (NJ, MP, ML) or 15,000 sampled post burn-in trees (BI).  In each tree, the congener *Arripis georgianus* was the outgroup.
Figure 5.3  Parsimony network for the mtDNA haplotypes from the concatenated sequence data (cytochrome *b* and *COI*) of the three species of Australian Salmon obtained by the criterion of parsimony with the program TCS version 1.21 (Clement *et al.* 2000). Each line in the network shows a single mutational change. A haplotype is shown by a coloured circle, the surface area of which is proportional to the frequency of the haplotype. Intermediate, unsampled haplotypes are indicated by a small black dot. Similar networks are shown for each of the genes separately (inset). In each case, *Arripis trutta* is represented by yellow; *A. truttaceus* by pink; and *A. xylabion* by orange.
Table 5.1  Summary of sister-species relationships between the three species of Australian Salmon generated by four phylogenetic methods with the concatenated mtDNA sequence data. Methods are neighbour-joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). Where a relationship was resolved, the sister-species are listed with either bootstrap or Bayesian support in parentheses. Unresolved relationships are indicated as a polytomy. See Figure 5.2.

<table>
<thead>
<tr>
<th></th>
<th>concatenated</th>
<th>cytochrome b</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>A. trutta</em></td>
<td><em>A. trutta</em></td>
<td><em>A. trutta</em></td>
</tr>
<tr>
<td></td>
<td><em>A. truttaceus</em></td>
<td><em>A. truttaceus</em> (52)</td>
<td><em>A. trutta</em></td>
</tr>
<tr>
<td></td>
<td><em>A. truttaceus</em></td>
<td><em>A. xylabion</em> (52)</td>
<td><em>A. xylabion</em></td>
</tr>
<tr>
<td>NJ</td>
<td>polytomy</td>
<td><em>A. truttaceus</em></td>
<td>polytomy</td>
</tr>
<tr>
<td></td>
<td><em>A. xylabion</em></td>
<td><em>A. xylabion</em> (59)</td>
<td>polytomy</td>
</tr>
<tr>
<td>MP</td>
<td>polytomy</td>
<td><em>A. truttaceus</em></td>
<td>polytomy</td>
</tr>
<tr>
<td></td>
<td><em>A. xylabion</em></td>
<td><em>A. xylabion</em> (55)</td>
<td>polytomy</td>
</tr>
<tr>
<td>ML</td>
<td>polytomy</td>
<td><em>A. truttaceus</em></td>
<td>polytomy</td>
</tr>
<tr>
<td></td>
<td><em>A. xylabion</em></td>
<td><em>A. xylabion</em> (0.69)</td>
<td>polytomy</td>
</tr>
<tr>
<td>BI</td>
<td>polytomy</td>
<td><em>A. truttaceus</em></td>
<td>polytomy</td>
</tr>
</tbody>
</table>

Figure 5.4  Majority consensus phylogeny of Australian Salmon inferred from a $D_A$ distance matrix of allele frequency data from length polymorphism in four intron loci, constructed by neighbour-joining. Node support value is derived from 1000 bootstrap replicates. Connections to the outgroup have been shortened for scaling purposes (dotted line).
5.3.1.1 **Support for Alternative Tree Topologies**

The results of both the SH and AU topology tests, which were applied to the mtDNA data in view of the absence of compelling support for any particular tree topology, indicated that the likelihood of all possible resolved tree topologies were not significantly different from each other or from a polytomy (Table 5.2).

### Table 5.2  Probability values derived from two topology tests for support of each of the possible resolved tree topologies among the three species of Australian Salmon using the concatenated mtDNA sequence data. Probability for the SH test was determined by comparing the constrained tree against the highest likelihood tree (a polytomy; see Section 5.2.2.1.1). Probability for the AU test equates to multi-scale bootstrap support for the null hypothesis ($H_0$). Null hypotheses are rejected when $p < 0.05$. Note that each topology was an independent test.

<table>
<thead>
<tr>
<th></th>
<th>(A.) <em>muta</em></th>
<th>(A.) <em>trutta</em></th>
<th>(A.) <em>spp</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Shimodaira &amp; Hasegawa (SH)*</td>
<td>0.828</td>
<td>0.172</td>
<td>0.172</td>
</tr>
<tr>
<td>Approximately Unbiased (AU)†</td>
<td>0.826</td>
<td>0.174</td>
<td>0.176</td>
</tr>
</tbody>
</table>

* $H_0$ (SH) = all trees are equally likely
† $H_0$ (AU) = likelihood for constrained tree is greater or equal to that expected for all bootstrap trees
5.3.2 Exploring the Phylogenetic Signal

The three approaches that were used to further explore the phylogenetic signal in the mtDNA data (i.e. zero-length branch test, likelihood mapping, NeighborNet), failed to find support for a bifurcating tree.

The zero-length branch tests were unable to reject the null hypothesis of a polytomy (i.e. a topology with zero-length branches) for any of the three possible bifurcating tree topologies. Specifically, none of the trees with bifurcating topologies generated a tree with an improved likelihood over a polytomy (all \( p \)-values = 1.00).

Likelihood mapping also indicated that it is not possible to find a resolved bifurcating tree with the mtDNA data. Specifically, based on the guidelines of Schmidt and von Haeseler (2009) (see Section 5.2.2.2.2), the analysis revealed considerable evidence for a non-bifurcating signal (areas 4 + 5 + 6 + 7 \( \approx \) 49\% of points; Figure 5.5) with only weak support for a tree-like signal (areas 1 + 2 + 3 \( \approx \) 51\% of points; Figure 5.5).

In the NeighborNet splits graph, there was considerable bootstrap support for each of the species ‘clades’ (81\% - 99\%) and hence a high least squares fit (99.9) (Figure 5.6). However, the graph contained conflicting signals at the point of divergence of the three species of Australian Salmon, as indicated by the presence of parallelograms, such that the maximum level of bootstrap support for any of the alternative pathways was only 34.4\% (Figure 5.6).
Figure 5.5  Likelihood map for concatenated mtDNA sequence data for the three species of Australian Salmon.  **A.** Likelihood map indicating the percentage of points contained within each of seven regions.  **B.** Point map showing relative position of posterior probabilities used to generate the percentage values in A.  The apices represent support for bifurcating trees; the edges, support for network-like trees and; the centre, support for star-like trees (see Section 5.2.2.2.2).

Figure 5.6  *(over page)* NeighborNet splits graph based on uncorrected P-distances with the rooted equal-angle algorithm on the concatenated mtDNA sequence data for the three species of Australian Salmon.  *Arripsis georgianus* was the outgroup root (not shown).  Haplotypes are indicated by Roman numerals as given in Table 4.5.  Parallel edges (branches) represent alternative pathways between taxa.  Bootstrap support is indicated for the edges leading to the ‘clades’ representing each species (all > 80%) and for the edges indicating alternative basal pathways between the three species (all < 35%; enlarged inset).  *P*-distances are indicated by the scale bar.
5.3.3 Molecular Dating

Phylogenetic trees based on the assumption of a molecular clock indicated that branch lengths between the Australian Salmon species were very short, irrespective of which of the three possible topologies was enforced, and all topologies had similar posterior likelihood probabilities (Table 5.3). The mean estimates of $T_{\text{mrca}}$ for the divergence of each of the three Australian Salmon were very similar, regardless of the tree topology, and suggested a Pleistocene divergence some 647,000 – 704,000 years ago (see Node ‘a’ in Table 5.3). Confidence intervals of these estimates ranged from 249,000 years to 1,060,000 years and overlapped across all tree topologies and also between nodes (Table 5.3), although they all inevitably place the divergence in the mid to late Pleistocene.
Table 5.3  Estimates of the time to most recent common ancestor ($T_{\text{mrca}}$) for each of the three possible tree topologies for Australian Salmon based on Bayesian resampling of the concatenated mtDNA sequence data using a divergence rate of 2% per million years. $T_{\text{mrca}}$ estimates and confidence (95% highest posterior density interval; in parentheses) are given in millions of years before present. Node ‘a’ and ‘b’ refer to the respective node annotated on each pictogram. The estimated tree height is equivalent to an estimate time since divergence from the outgroup, *Arripis georgianus*. The Bayesian posterior likelihood for each possible tree topology is provided.

<table>
<thead>
<tr>
<th>Posterior likelihood</th>
<th>Tree height</th>
<th>$T_{\text{mrca}}$ (Node a)</th>
<th>$T_{\text{mrca}}$ (Node b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2259.309</td>
<td>7.830</td>
<td>0.675 (0.451 – 0.945)</td>
<td>0.669 (0.430 – 0.888)</td>
</tr>
<tr>
<td>-2303.929</td>
<td>13.516</td>
<td>0.704 (0.441 – 0.968)</td>
<td>0.668 (0.383 – 0.854)</td>
</tr>
<tr>
<td>-2259.483</td>
<td>8.002</td>
<td>0.647 (0.493 – 1.060)</td>
<td>0.507 (0.249 – 0.750)</td>
</tr>
</tbody>
</table>
Chapter 5. Molecular Phylogenetics

5.4 Discussion

The molecular data in this chapter provided no definitive support for the geminate status of the *Arrapis* species on the east and west coasts of Australia, *A. trutta* and *A. truttaceus*, respectively. In fact, the evidence is consistent with a trifurcating polytomy where the three species of Australian Salmon (the two mentioned above and *A. xylabion*) arose at, or nearly at, the same time.

5.4.1 Hard Polytomy in Australian Salmon

The molecular phylogeny of the three species of Australian Salmon is probably a ‘hard’ polytomy (e.g. Coddington and Scharff 1996), and likely reflects a radiation, in which the three species arose from a common ancestor at, or almost at, the same time. However, distinguishing a hard polytomy from a soft polytomy (reflecting systematic errors or low resolution data) is difficult. Although the default has often been to assume a soft polytomy, this rejects a polytomy as a valid and likely phylogenetic scenario (Coddington and Scharff 1996; Page and Holmes 1998) and examples of hard (or effectively hard) polytomies with short internal branch lengths are being increasingly recognised (e.g. Poe and Chubb 2004; Barth *et al.* 2008).

The polytomy might represent a true trifurcating radiation, in which three Australian Salmon species arose at precisely the same time. More likely, however, is that the lineages diverged over a short time period, relative to the rate of character evolution, and so internal branches did not have time to accumulate mutations (e.g. Walsh *et al.* 1999; Braun and Kimball 2001; Whitfield and Lockhart 2007). If this is the case, there
probably is insufficient evolutionary information in the genome to be able to reliably reconstruct the precise order of divergences, in which case the polytomy is ‘effectively hard’ (see Hoelzer and Melnick 1994; Page and Holmes 1998). This is supported by a maximum-likelihood power analysis (Walsh et al. 1999), which indicated that, based on the detected levels of polymorphism, even sequencing the entire *Arripis* mtDNA genome would permit detection of two divergences only if they occurred at least 30,000 years apart (data not shown). The same power analysis method also indicated that >117,000 bp (more than six times the length of the entire mtDNA) would be required to discriminate between divergences that occurred within 10,000 years of each other.

On the assumption that the mtDNA gene tree is a hard polytomy, and is consistent with the species tree (e.g. Maddison 1997), the results of this chapter infer that *A. trutta* and *A. truttaceus* are not, by definition, true geminates (*sensu* Jordan 1908). While this has potential implications for our understanding of speciation in southern Australia (see below), the eastern and western species of Australian Salmon may nevertheless be considered as ‘sibling’ species (Jordan 1908; Johnson and Cicero 2004). As such, they still provide an excellent opportunity for comparing the recent evolutionary histories of species on the east and west coasts of Australia (see Chapter 6).

Despite compelling evidence for a hard polytomy, it remains difficult to completely rule out a soft polytomy. In theory, a soft polytomy could have been caused by low levels of polymorphism in the mtDNA sequences, a systematic error in the data (e.g. from the concatenation), and/or inadequate sampling (see Coddington and Scharff 1996; Slowinski 2001). The amount of polymorphism in the cytochrome *b* and *COI* sequences of *Arripis* was low, despite these being among the most phylogenetically
useful and widely used (see Avise 2004; Hebert et al. 2004; see also Chapters 1 and 2).

To a certain extent, this problem was resolved by concatenating the two types of sequence. Although concatenation of sequences can lead to errors in phylogenetic analyses (e.g. Huelsenbeck et al. 1996), this is unlikely to be a problem in the present study because the combining of the cytochrome $b$ and $COI$ sequences was validated by the partition-homogeneity test (see Section 4.2.3.2.2). Inadequate sampling also does not appear to have been a major problem in the present study. This is because the analyses included at least 25 individuals of each of $A. trutta$ and $A. truttaceus$ from across the entire geographic range of the species. The sampling range for $A. xylabion$ was more limited as it was not possible to obtain samples of this species from any site other than Lord Howe Island. Furthermore, for all three species of Australian Salmon, sample size increases during the production of the mtDNA data only resulted in detection of more rare haplotypes, one or two mutational steps away from the common haplotype (see parsimony networks), which offer little phylogenetic utility (see Nei and Kumar 2000). Increasing the number of genes sampled (both mtDNA and nDNA) is probably required to provide greater confidence in a hard polytomy (e.g. Kliman et al. 2000; Waddell et al. 2000; Cooper et al. 2001; Kraus et al. 2006; Willerslev et al. 2009). A genome-wide, species tree approach (Maddison 1997; Brito and Edwards 2008; Avise 2009; Edwards 2009) would provide the ultimate genetic test of the relationships in $Arripsis$, however if the divergences were close in evolutionary time as predicted from the current data set, this probably would not fundamentally change the outcome – i.e. an effectively hard polytomy.
5.4.2 Pleistocene Divergence

Regardless of the relationships among the three species of Australian Salmon, the mtDNA results of the present chapter indicate that these species probably diverged from the most recent common ancestor in the mid to late Pleistocene, some 700,000 years ago. These estimates of divergence time are similar to allozyme-based estimates for *A. trutta* and *A. truttaceus* (MacDonald 1980) in that the divergence is placed in the mid to late Pleistocene and the confidence intervals overlap, although the allozyme estimates are slightly earlier (0.8 – 1.14 Mya). Although the divergence estimates of the present study are based on the standard mtDNA substitution rate of 2%/Myr (which might not be ideal), the consistency of divergence estimates between the two studies suggests that this rate provides a reasonable approximation for *Arripsis* (see also Grant and Bowen 1998; Avise 2000; Burridge 2002; Theisen et al. 2008; Weir and Schluter 2008; Drew et al. 2009). Regardless, in the case of the Arripidae, it is impossible to obtain precise data on the rate of molecular evolution because there are no fossil records for the family (M. Siversson, Western Australian Museum, pers. comm.) and even the phylogenetic position of the family is uncertain (K. Carpenter, Old Dominion University, pers. comm.) (see Bromham and Penny 2003). Finally, while the data suggest a mid to late Pleistocene divergence, it is important to recognise that the estimated divergence times for *Arripsis* are crude (e.g. have large confidence intervals) such that it is impossible to directly link them to specific glacial events within the Pleistocene.

In the present study, even the most conservative estimate of divergence time for the three Australian Salmon (1.05 Mya; see Table 5.3: HDP, Node ‘a’), places the split in the mid Pleistocene. Notwithstanding the difficulties of direct comparisons across
studies, this is later than available estimates for other geminate marine species in temperate Australia. For example, using a similar type of data, Burridge (2000b) estimated that two pairs of teleost sister-species diverged well into the Miocene (7.8 – 11.2 Mya) and another in the Pliocene (2.7 – 3.8 Mya), while divergences between east/west sister-species of sharks (Corrigan and Beheregaray 2009) and sea-stars (Waters et al. 2004) were calculated as near the onset of glaciation around the Pliocene/early Pleistocene (1.0 – 3.9 Mya, and 2.1 – 2.4 Mya, respectively). At the other extreme, the presence of divergent (conspecific) populations across the Bass Strait in a range of taxa provides circumstantial evidence of much more recent divergence (e.g. Dawson 2005; Waters et al. 2005; York et al. 2008; Fraser et al. 2009). It is likely that at least some of these east/west populations will eventually evolve into separate species (or be recognised as such; see Spencer et al. 2007), and thus become true geminates. The apparently wide range of divergence times for pairs of marine taxa on either side of the Bass Strait is consistent with the view that the (ephemeral) Bassian Isthmus has acted as an isolating barrier on multiple occasions, both prior to and throughout the Pleistocene and into the present. Similar conclusions have been reached for other temporary marine barriers (Bowen and Avise 1990; Bernardi et al. 2003).

5.4.3 Speciation in Australian Salmon

While a simple hypothesis invoking the Bass Strait as an isolating mechanism among sister-species in temperate Australian marine fauna may be appealing (see Section 5.1), the results of this chapter suggest that the Bassian Isthmus might only be one contributing factor to species radiations in the region. As sibling species (see above), the relationship between *A. trutta* and *A. truttaceus* is still consistent with speciation by
vicariance associated with the Bassian Isthmus and oceanographic conditions (see Chapter 1). Specifically, the emergence of a continuous land bridge across the Bass Strait, in conjunction with colder sea temperatures (Wells and Okada 1996; see also Chapter 6), could be expected to have vicariantly isolated populations and caused them to move northwards to escape the cold water. Northward movement of individuals on the west of the Bassian Isthmus would inevitably have taken these individuals on a considerable journey around the south-west corner and onto the west coast of Australia. This most likely would have involved adult fish since the west to east direction of prevailing winds and currents at that time suggests that passive dispersal by larvae from east to west was/is very unlikely (CLIMAP 1976; Prell et al. 1980; Petrusevics and Bye 1995). A similar scenario was proposed for two other east/west sister-species of fishes (in the genera *Chironemus* and *Nemadactylus*; Burridge 2000a;2000b).

In contrast to the species on the east and west coasts of Australia, the presence of another sibling-species of *Arripis* at Lord Howe Island and nearby islands is probably best explained as allopatric speciation by peripheral isolation. During glacial periods, in addition to the presence of a land bridge across the Bass Strait, current flows would have likely facilitated the dispersal of larvae of *Arripis* (and other species, e.g. Waters and Roy 2003) from the east coast of Australia to the islands north of New Zealand. Specifically, eddies diverging from the mainland were thought to be positioned further north, thus a branch of water, carrying larval *Arripis*, could have reached Lord Howe Island (31.5° S; Bostock et al. 2006; Woodroffe et al. 2006). Geomorphological cores around Lord Howe Island show that the current tropical community is Holocene in origin (< 7000 years old) and the island was dominated by temperate marine communities throughout much of the Pleistocene (Woodroffe et al. 2006), indicating the
availability of suitable conditions when the species of *Arripis* are predicted to have diverged. In a related way, this hypothesis also provides an explanation for why *A. trutta* is not found on Lord Howe Island, but is found in New Zealand. In present conditions, the East Australian Current diverges from the mainland at around 32° S (Bostock *et al.* 2006) and flows to the south of Lord Howe Island, thus larval *A. trutta* are transported to New Zealand, rather than to Lord Howe Island. This is reflected in contemporary dispersal to Lord Howe Island, which is found in tropical taxa with sources north of 32° S (e.g. mangroves, Duke *et al.* 1998; corals, Ayre and Hughes 2004; sea turtles, Boyle *et al.* 2009), rather than temperate fauna such as *A. trutta*.

5.4.4 Conclusions

Phylogenetic relationships among the three species of Australian Salmon were not fully resolved despite the use of a range of analytical approaches. It is suggested that this is because the species arose within a very short timeframe and consequently probably represent an effectively hard polytomy. As such, a ‘resolved’ bifurcating phylogeny may not exist and by definition, true ‘sister’ species may not exist. Rather, the three species might best be considered ‘sibling’ species that appear to have diverged in the mid to late Pleistocene, which is later than the divergence times estimated for other geminate species in the same region. The available evidence is consistent with the view that glaciation may have caused a species radiation in *Arripis* by two different modes – vicariance, leading to the east and west coast species (*A. trutta* and *A. truttaceus*, respectively) and dispersal to the east (*A. xylabion*). This scenario also provides for a relatively fast radiation, since a single glacial cycle would simultaneously result in the conditions that could lead to both mechanisms.
6.0 Genetic Diversity and Recent Evolutionary History

6.1 Introduction

As explained in Chapter 1, the presence of putative geminate species in southern Australia provides an opportunity to explore the evolutionary processes affecting the fauna in the region. Although the east/west species pair of *A. trutta* and *A. truttaceus* were not confirmed as true sister-species, it is nevertheless likely that they are part of a trio of sibling-species that arose nearly simultaneously (see Chapter 5). Consequently, this pair can effectively be regarded as geminates in such explorations. In particular, these two species can be used to test whether glacial cycles and associated changes in environmental conditions have impacted more severely on species distributed on the western side of the Bass Strait compared to the eastern side (e.g. Fraser *et al.* 2009).

As discussed in Chapter 5, the species of Australian Salmon (*A. trutta, A. truttaceus, A. xylabion*) likely arose during the mid to late Pleistocene (~ 700,000 years ago). Since then, the climate and oceanography of southern Australia (and other locations) has experienced marked fluctuations through numerous glacial cycles (Jansen *et al.* 2007). These glacial cycles have had a significant impact on coastal marine environments in the region and, associated with this, can be expected to have had a significant impact on the demographies and distributions of the resident flora and fauna (e.g. Sotka *et al.* 2005; Janko *et al.* 2007; Fraser *et al.* 2009). Interestingly, a range of historical environmental data indicate that the environmental changes accompanying glacial cycles have been more severe on the western side of the Bass Strait compared to the eastern side, as is explained below. Consequently, it can also be expected that the
demography and distribution of species such as \textit{A. truttaceus} on the western side will potentially be more severely impacted by glacial cycles than species such as \textit{A. trutta} on the eastern side.

Although there have been numerous glacial cycles since the mid to late Pleistocene, the Last Glacial Maximum (LGM; \textasciitilde20,000 years ago) is the best known. As a result, the details of the LGM are often used as a proxy for interpreting earlier events (Paulo \textit{et al.} 2001; Jansen \textit{et al.} 2007), despite the fact that some aspects (e.g. magnitude and duration) of the LGM might have been different. In the context of genetic studies, the LGM is particularly important because its effects, if strong enough, may erase signals from earlier events (Hewitt 2000).

6.1.1 Pleistocene Glacial/Interglacial Cycles in Southern Australia

6.1.1.1 East of the Bass Strait

Both sea temperatures and current flows on the east coast of Australia differed only slightly between glacial and interglacial stages (CLIMAP 1976; Lawrence and Herbert 2005; Hostetler \textit{et al.} 2006). For example, during the LGM, the water in the Tasman Sea was probably only substantially colder in the southern-most parts (CLIMAP 1976; Hostetler \textit{et al.} 2006). Consequently, for example, glacial and interglacial foraminiferal assemblages on the east coast were similar, differing mostly in subtle latitudinal shifts (Bostock \textit{et al.} 2006). The similarity of water temperatures between glacial and interglacial stages was largely due to two factors, as follows.
Firstly, during the LGM, east of the Bassian/Tasmanian peninsula, the position of the Subtropical Convergence Zone (which is the interface between cool southerly waters and warmer, tropically derived waters) traced the southern boundary of the Tasman Sea between Tasmania and New Zealand. This is only slightly north of the present (interglacial) position.

Secondly, the southward flow of warm water via the East Australian Current, which dominates the patterns of present-day water movement down the east coast, continued through the LGM, albeit reduced in strength and slightly cooler (CLIMAP 1976; Bostock et al. 2006; Hostetler et al. 2006). However, the point where the East Australian Current diverts from the mainland towards New Zealand (the Tasman Front) was pushed a few degrees northwards (Bostock et al. 2006; see also Chapter 5). Nevertheless, throughout the LGM, an extension of the East Australian Current continued south of the Tasman Front to Tasmania (Bostock et al. 2006; Figure 6.1).

**Figure 6.1** *(over page)* Circulation features of southern Australia (modified from Wells and Okada 1996). **A.** Present (interglacial) conditions: Subtropical Convergence Zone is positioned south of the continent. The south flowing Leeuwin Current carries warm water down the west coast and along the south coast. The cool, offshore Western Australian Current flows northwards. **B.** Reconstruction of Late Quaternary glacial conditions: Polar ice-sheet expanded, lowering sea levels and exposing the Bassian Isthmus. The Subtropical Convergence Zone shifted northwards. The Leeuwin Current ceased and the Western Australian Current dominated on the west coast.
6.1.1.2 West of the Bass Strait

In contrast to the eastern side of the Bass Strait, both sea temperatures and current flows on the western side differed markedly between glacial and interglacial stages (CLIMAP 1976; Lawrence and Herbert 2005; Hostetler et al. 2006). For example, sea temperatures were likely to have been significantly lower than the present - around 4°C less on the west coast and up to 10°C lower on the south coast. This was largely due to three factors, as follows.

Firstly, during the LGM, the Subtropical Convergence Zone shifted northwards by several degrees (due to an expansion of the Antarctic ice-shelf) and was juxtaposed with the southern Australian coastline from the south-west corner to the exposed Bassian Isthmus (CLIMAP 1976; Howard and Prell 1992; Wells and Okada 1996; Figure 6.1).

Secondly, circulation patterns in the southern Indian Ocean were deflected off the Subtropical Convergence Zone causing an exaggerated flow of cool water to the south-west of Australia, which was diverted northwards by the continental land mass (CLIMAP 1976; Prell et al. 1980). This flow of cool water corresponds to the Western Australian Current, which, while weak during interglacial stages (including the present), dominated the circulation along the west coast of Australia during glacial stages (Belyaeva and Burmistrova 1992; Figure 6.1).

Thirdly, and partly as a consequence of the strengthened northward flow of the Western Australian Current, the Leeuwin Current, which dominates the patterns of present-day winter water movement down the west coast and along the south coast (Godfrey and
Ridgway 1985), completely ceased during glacial stages (Wells and Wells 1994; Gingele et al. 2001; Figure 6.1).

6.1.2 Using Genetic Data to Infer the Evolutionary Histories of *Arripis* Species

Even when fossil and sub-fossil records exist, it is typically difficult to obtain good information about the evolutionary history of a species and, in most cases, genetic data provide the only option for investigating demographic history, as is the case for *Arripis*. Genetic approaches in this regard are part of the rapidly expanding field of phylogeography (e.g. Avise 2000;2009) and hinge on the fact that patterns in genetic data, especially departures from mutation-drift equilibrium, can provide clues about the recent demographic history of a species, such as whether the size of a population has been stable, growing or contracting during certain time periods (Watterson 1984; Grant and Bowen 1998; Avise 2000; Knowles 2009). Nevertheless, it is important to recognise that the interpretation of genetic data in this context is not always straightforward, and is often dependent on restrictive assumptions (e.g. see Avise 2000; Hare 2001).

6.1.3 Aim

The main aim of this chapter was to assess and compare the levels and patterns of genetic diversity in the effectively geminate species *A. trutta* and *A. truttaceus*. The resultant information was used to infer and compare aspects of recent evolutionary histories of these species. In particular, this chapter tests the prediction that *A. truttaceus*, which resides to the west of the Bass Strait, may have been more severely
impacted by glacial cycles than *A. trutta*, which resides to the east. Finally, data on *A. georgianus*, which is sympatric with *A. truttaceus*, and on *A. xylabion* on Lord Howe Island, are also presented to provide context for the results for Eastern and Western Australian Salmon.

6.2 Materials and Methods

6.2.1 Sampling Regime

The results of this chapter are based upon the analysis of samples of *A. trutta*, *A. truttaceus* and *A. georgianus* from across their Australian distributions, and of *A. xylabion* from Lord Howe Island. Since one of the main aims of this chapter was to explore the effects of glaciation on populations to the east and west of the Bass Strait in Australian waters, and since the relationship between the Australian and New Zealand assemblages of *A. trutta* is not clear, the two individuals of this species from New Zealand were not included. [NB: although, the exclusion of these individuals had a negligible effect on all results; data not shown]. The sample sizes for each of the sampled mtDNA and intron loci for each sampling site for each species are provided in Chapter 4 (Table 4.2). Given no evidence of population structure in any species in Australian waters (Chapter 3), all samples were pooled into a single sample for each species as described in Chapter 4.
6.2.2 Genetic Assays

The mitochondrial DNA data used in this chapter are in the form of the nucleotide sequence of a 483 bp portion of the cytochrome b gene in Arripis, generated as described in Chapter 2. The cytochrome b region was selected for mtDNA analyses because it contained a higher level of polymorphism than the COI region. Concatenated sequences (i.e. cytochrome b plus COI) were not used in this chapter because, while concatenation may improve phylogenetic resolution between species (see Section 4.2.3.2.2), they may impose a greater risk of error when used for demographic inference (Brito and Edwards 2008; Edwards 2009; Knowles 2009).

The nuclear DNA data used in this chapter are in the form of the frequency of length variants (alleles) in the introns of four genes (AldoB4, S72, CaM4 and CK7). These loci were selected for study because, of the 12 intron loci that could be reliably amplified in Arripis (see Chapters 2 and 3), only these four loci had scorable length polymorphism in one or more species of Arripis. However, one measure of genetic diversity (the proportion of polymorphic loci) was based on data from all 12 of the scorable loci (see Chapter 2). Alleles were binned into size classes as explained in Section 2.5.4. The nDNA data were generated using the methods described in Chapter 2.

6.2.3 Genealogical Relationships

The evolutionary relationships among the cytochrome b haplotypes of A. trutta, A. truttaceus and A. xylabion are discussed in Chapters 4 and 5 in the context of species relationships. The associated parsimony networks (excluding the two New Zealand
specimens) are repeated in this chapter, along with the equivalent parsimony network from a sample of 30 individuals of *A. georgianus* (presented for the first time), because they can also provide circumstantial evidence in support of demographic inferences (e.g. Avise 2000; Mateus *et al.* 2011).

6.2.4 Data Analyses

6.2.4.1 Genetic Diversity

For the mtDNA data, genetic diversity was measured in three ways: (i) the standardised number of haplotypes; (ii) haplotype diversity; and (iii) nucleotide diversity.

The standardised number of haplotypes ($SN_H$) was calculated using POPTOOLS (www.cse.csiro.au/CDG/poptools). Since the number of haplotypes observed in a sample strongly depends on the size of the sample, estimates of the number of haplotypes were standardised according to sample size in order to compare between species. The standardising procedure was based around the number of individuals present in the smallest sample (i.e. 23 for *A. trutta*). Thus, for each of *A. truttaceus* and *A. georgianus*, 23 individuals were shuffled (i.e. sampled without replacement) over 1000 times. This method was used because simulations have shown that sampling without replacement is one of the most robust and precise methods for standardising haplotype numbers (Leberg 2002).

Estimates of haplotype diversity (and standard errors) were calculated using ARLEQUIN version 3.1 (Excoffier *et al.* 2005). Haplotype diversity ($h$) is the
probability that two haplotypes, randomly chosen from a sample, are different (Nei 1987). Estimates of nucleotide diversity were also calculated using ARLEQUIN. Nucleotide diversity ($\pi$) is the probability that two homologous nucleotides, randomly selected from a sample, are different (Nei 1987). It was estimated according to the number of substitutions between two sequences, using the substitution model TrN (see Section 4.2.3.2.4 for details of the choice of model) with a gamma correction (G) of 0.02 in *A. trutta* and 97.03 in both *A. truttaceus* and *A. georgianus*.

For the nDNA data, genetic diversity was measured in three ways: (i) proportion of polymorphic loci; (ii) standardised number of alleles; and (iii) average expected heterozygosity per polymorphic locus. The proportion of polymorphic loci ($P_0$) was calculated as the proportion of the 12 intron loci where the common allele occurred with a frequency less than 0.95.

The standardised number of alleles ($S_{N_{A}}$) was calculated for each polymorphic locus for each species and then the values were summed over all loci to provide a standardised estimate of the total number of alleles for each species. As described above, for each locus, samples for *A. truttaceus* and *A. georgianus* were standardised to the number of *A. trutta* sampled (the species with the smallest $n$), by shuffling over 1000 times in POPTOOLS.

Expected heterozygosity ($H_e$) in all species was averaged for four loci (*AldoB4, CaM4, CK7, S72*). Average expected heterozygosity and its variance was calculated using the method of Nei (1987), as implemented in ARLEQUIN.
6.2.4.2 Comparisons of Genetic Diversity Between Species

For each of the five sampled loci, each species was ranked from highest to lowest according to the amount of diversity present and trends in the data were inspected in order to assess whether the rankings of the species were consistent across loci. This was done separately for values of: (i) the standardised number of haplotypes/alleles; and (ii) haplotype diversity/expected heterozygosity. Simple ranks were used (e.g. Kendall 1970) and an average of the ranks was used when values were tied between two or more species (e.g. Legendre 2005). It was not possible to assess the statistical significance of these results because, for the relevant test, Kendall's coefficient of concordance for ranks test ($W$; Kendall 1970), there were too few polymorphic loci in Arripis (see Legendre 2005; Chang 2009).

6.2.4.3 Neutrality of Markers

The neutrality of the markers in each species of Arripis was assessed because whether or not the markers are under the influence of selection has a major influence on the interpretation of the results. Kendall's coefficient of concordance for ranks test ($W$) was used to test for evidence of neutrality via concordance in the patterns of variation at different loci, based on the theoretical assumption that, if loci are essentially selectively neutral, then concordant patterns of genetic diversity should be detected across all loci (Nei 1987; Avise 2004). Two characteristics (standardised number of haplotypes/alleles and haplotype diversity/expected heterozygosity) were used to test for concordance between the five loci (cytochrome b and four introns) across all four species of Arripis.
Kendall’s $W$ was calculated using the online program STATTOOLS (Chang 2009) and the significance assessed using a table of critical values (Siegel and Castellan 1988).

### 6.2.4.4 Demographic History

Three general approaches were used to examine aspects of the recent demographic history of the four species of *Arripis*. Firstly, evidence was sought for changes in population size using both mtDNA and nDNA loci. Secondly, where evidence for population expansion was detected in the mtDNA, the date of the start of the expansion was estimated. Finally, estimates of effective population sizes prior to the expansion were made.

#### 6.2.4.4.1 Tests for evidence of changes in population size

##### 6.2.4.4.1.1 Mitochondrial DNA sequences

Several approaches, all of which are based around expectations for mutation-drift equilibrium, were applied to the *Arripis* cytochrome $b$ sequence data to test for evidence of changes in population sizes (Ramos-Onsins and Rozas 2002). The tests were chosen on the basis that they: (i) are widely used; (ii) are the most appropriate; and (iii) include a representative from each of three classes of test (Ramos-Onsins and Rozas 2002), as described below. Except where indicated, the statistical significance of each was assessed by randomly sampling the data 10,000 times in DnaSP version 5.00.04 (Rozas and Rozas 1999), assuming selective neutrality and population equilibrium.
Class I – The presence of an excess of singleton mutations at segregating sites, indicating recent population expansion, was tested by Tajima’s (1989) $D$-test and Ramos-Onsins and Ramos’ (2002) $R_2$-test. Both tests were included because the former is based solely on the number of singletons, while the latter compares that number to the mean number of nucleotide differences between sequences. A negative value of $D$ or a low value of $R_2$ would each signify an excess of low frequency polymorphisms, indicating population size expansion and/or directional selection.

Class II – An excess of rare haplotypes relative to that expected based on pairwise nucleotide differences ($\pi$) was used to test for population expansion using Fu’s (1997) $F_s$-test. A negative value of $F_s$ would indicate an excess of rare mutations and hence, population size expansion and/or directional selection.

Class III – Mismatch distributions (Slatkin and Hudson 1991; Rogers and Harpending 1992) of the cytochrome $b$ sequence data were generated for each species and parameters derived from these were used to obtain, by simulation, expected values under a model of: (i) population expansion assuming $\theta_1 = \infty$ (i.e. after the growth event); and (ii) stable population size (Rozas and Rozas 1999). The statistical significance of any departures between the observed data and the population expansion model was assessed using the sum of squared differences ($SSD$) and bootstrap approach, as implemented in ARLEQUIN. In this approach, the observed value of $SSD$ was compared to the distribution of $SSD$ values expected for population expansion, generated via 10,000 bootstrap simulations. The $p$-value of the $SSD$ statistic was calculated as the proportion of simulated (expected) $SSD$ values larger than the observed value; a small $p$-value therefore provides evidence for departure from a model of
population expansion (Excoffier and Schneider 1999). Another statistic based on the mismatch distribution, raggedness, \( rg \) (Harpending 1994) was calculated for each species of \textit{Arripsis} using DnaSP. By providing a measure of the ‘smoothness’ of the mismatch distribution, this test differentiates between unimodal (‘smooth’) curves in expanding populations and multimodal (‘ragged’) curves in stable populations that have accumulated many mutations. Raggedness is an \textit{ad hoc} test (Harpending 1994), which differs from the SSD method by not testing against a specific model (i.e. expansion). A mismatch distribution could not be generated for the monomorphic \textit{A. xylabion}.

6.2.4.1.2 Nuclear DNA allele frequencies

The intron data (using polymorphic loci only) were used to test for changes in population size with the software BOTTLENECK version 1.2.02 (Cornuet and Luikart 1996) over 10,000 replicate simulations for each locus. This method used the theoretical expectation that the number of alleles in a population would be reduced more than the heterozygosity during a population size reduction due to the preferential loss of rare alleles (Nei \textit{et al.} 1975; Maruyama and Fuerst 1985). The expected heterozygosity of each locus \( H_E \); based on observed allele frequencies, \textit{sensu} Nei 1987) was compared to a distribution of heterozygosity expected for a population at mutation-drift equilibrium with the observed number of alleles \( H_{\text{Exp}} \). This distribution was obtained by coalescent simulation under the infinite-allele model (IAM; Kimura and Crow 1964) and then again for the stepwise mutation model (SMM; Ohta and Kimura 1973). The results were summarised using the equation \( H_E - H_{\text{Exp}} \), where a positive value suggests a population bottleneck, whereas a negative value suggests a population expansion (Luikart and Cornuet 1998). Although standard nuclear data such as intron length
polymorphism probably conform more closely to the IAM (Cornuet and Luikart 1996), tests using both IAM and SMM were conducted because they are thought to represent two extremes of mutation models (Chakraborty and Jin 1992), and one major assumption of the IAM is that of non-homology of alleles, and in the present study there is no way to confirm that this was not violated. It was not possible to assess the statistical significance of these results because, as described in Section 6.2.4.2, the recommended approach (Wilcoxon sign–rank test; Cornuet and Luikart 1996) could not be applied to four loci.

6.2.4.4.2 Estimates of timing of population expansions

For each of the species where evidence of population expansion was detected in the mtDNA, the age since the start of the expansion (Harpending 1994) was estimated using two methods which, despite some possible limitations (see Arbogast et al. 2002), are the main approaches used in this context (e.g. Sotka et al. 2005; Janko et al. 2007; Lohman et al. 2008; López et al. 2010). Firstly, the crest of the mismatch distribution curve (see Section 6.2.4.4.1.1), expressed as tau, $\tau$, was estimated with the software DnaSP with $\theta_1$ set as infinite (Rogers 1995). Secondly, the coalescent simulation method of Wakeley and Hey (1997) was used to estimate $\tau$ in the program SITES (Hey and Wakeley 1997). In both cases, the time since expansion, $t$, was then calculated by $\tau = 2ut$ (Rozas and Rozas 1999; Excoffier et al. 2005), where $u$ is the substitution rate per site per year - in this case $4.83 \times 10^{-6}$ (1%/site/Myr x length of sequence; Rogers and Harpending 1992, see Section 5.2.2.3.2).
6.2.4.4.3  Estimates of effective population sizes

In the species where expansion was detected, the effective female population size $N_f(0)$ at the start of the expansion was estimated. This was estimated using $\theta = N_f \mu$, and dividing by the generation time (4.0 years, based on age of sexual maturity for both \textit{A. trutta} and \textit{A. truttaceus}; see Chapter 1). Mean values of $\theta_0$ (i.e. time = zero) along with 95% confidence estimates, were generated by parametric bootstrap in ARLEQUIN.

6.3  Results

6.3.1  General Information

6.3.1.1  Mitochondrial DNA Sequences

A 483 bp portion of the cytochrome \textit{b} gene was sequenced for each of 30 individuals of: \textit{A. georgianus}, 23 \textit{A. trutta} and 36 \textit{A. truttaceus} from across their Australian distributions; and 11 \textit{A. xylabion} from a single locality, and revealed a total of 18 haplotypes (Table 6.1). The average base composition of the sampled portion of the cytochrome \textit{b} gene was generally similar in all four species of \textit{Arripis}, although the AT content of the sequence in the \textit{A. georgianus} samples was slightly less than that in each of the three species of Australian Salmon (Table 6.1). As detailed in Section 4.3.1, there were generally few segregating sites, and substitutions were exclusively synonymous transitional changes at third codon positions in \textit{A. truttaceus} and \textit{A. georgianus}, while a single transversion and (rare) first and second codon substitutions were also found in \textit{A. trutta} (Table 6.1).
Up to 89 individuals of *A. trutta*, 143 *A. truttaceus*, 25 *A. xylabion*, and 261 *A. georgianus* were assayed for length polymorphism at the four polymorphic intron loci (full details are provided in Table 3.1 and Table 4.2). Locus *AldoB4* had the smallest alleles (all < 200 bp) and locus *CK7* had the largest (up to 984 bp). The total number of alleles at each locus for all species combined ranged from seven at *AldoB4* to 13 at *S72* (see Chapter 3; Table 3.2).

Overall polymorphism was greatest in the samples of *A. georgianus* (four polymorphic loci; total alleles = 32), followed by *A. trutta* (three polymorphic loci; total alleles = 20), *A. truttaceus* (two polymorphic loci; total alleles = 9), and *A. xylabion* (one polymorphic locus; total alleles = 5) (but see standardised results in Section 6.3.2.1). The patterns of variation at the different intron loci were assumed to be independent of each other based on the results of linkage tests (see Chapter 3). The genotype frequencies at each locus in each species were generally in accordance with those expected under Hardy Weinberg expectations (see Chapter 3).
Table 6.1  Selected characteristics of the 483 bp partial cytochrome *b* sequences in the species of *Arripis*. The number of individuals assayed, the number of haplotypes, the average base composition of the sequence (expressed as percentage), the number of polymorphic sites (i.e. where a nucleotide at a site differs between sequences), the transition/transversion (Ts/Tv) ratio, and the codon positions at which the polymorphic sites were found, are presented. See Table 4.5 for full details of polymorphic sites.

<table>
<thead>
<tr>
<th>Species</th>
<th>Number sequenced</th>
<th>Number of haplotypes</th>
<th>% T</th>
<th>% C</th>
<th>% A</th>
<th>% G</th>
<th>Number of polymorphic sites</th>
<th>Ts/Tv ratio</th>
<th>Codon position (1 / 2 / 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. georgianus</em></td>
<td>30</td>
<td>3</td>
<td>28.8</td>
<td>33.7</td>
<td>23.0</td>
<td>14.5</td>
<td>2</td>
<td>2 / 0</td>
<td>0 / 0 / 2</td>
</tr>
<tr>
<td><em>A. trutta</em></td>
<td>23</td>
<td>10</td>
<td>28.1</td>
<td>32.9</td>
<td>25.3</td>
<td>13.7</td>
<td>10</td>
<td>9 / 1</td>
<td>2 / 1 / 7</td>
</tr>
<tr>
<td><em>A. truttaceus</em></td>
<td>36</td>
<td>4</td>
<td>28.0</td>
<td>33.1</td>
<td>25.5</td>
<td>13.5</td>
<td>3</td>
<td>3 / 0</td>
<td>0 / 0 / 3</td>
</tr>
<tr>
<td><em>A. xylabion</em></td>
<td>11</td>
<td>1</td>
<td>28.0</td>
<td>33.1</td>
<td>25.7</td>
<td>13.3</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
6.3.1.3 Genealogical Relationships

Before presenting the genetic diversity results, it is important to consider the relationships between conspecific cytochrome \( b \) haplotypes. For each of the four species, the parsimony networks were centred on a common haplotype (Figure 6.2). In *A. trutta* numerous singletons (haplotypes represented by a single occurrence), which differed from the common haplotype by one or two mutational steps, were present, forming a classic star-shape pattern (Figure 6.2). This pattern provides strong circumstantial evidence of the accumulation of new mutations during a population expansion (Avise 2000). Similarly, the parsimony network of *A. truttaceus* showed a similar star-shape, with a few singletons that were all one mutational step from the common haplotype (Figure 6.2), which is consistent with a very recent population expansion. All sequenced individuals of *A. xylabion* had identical haplotypes. In *A. georgianus*, there was one central and two ‘peripheral’ haplotypes, which, although relatively uncommon, were represented by 23% and 7% of individuals (Figure 6.2). Hence, for this species, the network provided no clear evidence of an expansion, although the peripheral haplotypes still differed from the common haplotype by only a single mutational step.
Figure 6.2  Parsimony networks for the cytochrome b haplotypes for each species of *Arripis*. Each line in the network shows a single mutational change. A haplotype is shown by a circle. The surface area of the circle is proportional to the frequency of each haplotype. Black dots indicate missing haplotypes, which are necessary to link all observed haplotypes present in the network. Sample sizes are indicated in parentheses.
6.3.2 Genetic Diversity

6.3.2.1 General

Both the standardised number of haplotypes and haplotype diversity were noticeably higher in *A. trutta* ($SN_H = 10; h = 0.64$) than in *A. truttaceus* ($SN_H = 3.0; h = 0.16$) (Table 6.2). Haplotype diversity was non-existent in *A. xylabion* and moderate in *A. georgianus* ($SN_H = 2.9; h = 0.47$) (Table 6.2). The level of nucleotide diversity in the partial cytochrome *b* sequences was very low in each of the four *Arrapis* species ($\pi \leq 0.002$), and especially so in *A. truttaceus* and *A. xylabion* (Table 6.2).

### Table 6.2 Measures of mtDNA genetic diversity based on the nucleotide sequence of a 483 bp fragment of the cytochrome *b* gene in *Arrapis* species. Calculation of standardised number of haplotypes ($SN_H$), haplotype diversity and nucleotide diversity is described in Section 6.2.4.1. Standard errors are provided in parentheses. The monomorphic *A. xylabion* was not included in the calculations of $SN_H$.

<table>
<thead>
<tr>
<th>Statistic</th>
<th><em>A. trutta</em></th>
<th><em>A. truttaceus</em></th>
<th><em>A. xylabion</em></th>
<th><em>A. georgianus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>23</td>
<td>36</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>No. haplotypes</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>$SN_H$</td>
<td>10.0</td>
<td>3.0</td>
<td>-</td>
<td>2.9</td>
</tr>
<tr>
<td>Haplotype Diversity ($h$)</td>
<td>0.640 (±0.014)</td>
<td>0.162 (±0.082)</td>
<td>0</td>
<td>0.467 (±0.087)</td>
</tr>
<tr>
<td>Nucleotide Diversity ($\pi$)</td>
<td>0.0020 (±0.0020)</td>
<td>0.0003 (±0.0005)</td>
<td>0</td>
<td>0.0010 (±0.0009)</td>
</tr>
</tbody>
</table>
Chapter 6. Recent Evolutionary History

The proportion of polymorphic intron loci was low, ranging from 4/12 ($P_O = 0.33$) in *A. georgianus* to only a single locus in *A. xylabion* (Table 6.3). The standardised number of alleles across the four polymorphic intron loci ranged from 20 in *A. trutta* to five in *A. xylabion* (Table 6.3). Similar standardised values were obtained by a rarefaction approach (Szpiech et al. 2008; results not shown). Average expected heterozygosity ranged from moderate in *A. trutta*, *A. truttaceus* and *A. georgianus* ($H_E = 0.239 - 0.312$) to low in *A. xylabion* ($H_E = 0.093$) (Table 6.3). Consistent with the mtDNA data, *A. trutta* exhibited the highest $H_E$, although the standard deviations of the $H_E$ estimates were considerable (Table 6.3).

Table 6.3  Measures of nDNA genetic diversity in $n$ individuals of *Arripis* species. $P_O$ is the proportion of polymorphic loci (of 12), $N_A$ is the uncorrected number of alleles, $SN_A$ is the standardised number of alleles, and $H_E$ is average expected heterozygosity. Standard deviations are provided in parentheses. All estimates except $P_O$ were based on only the four loci that were polymorphic in at least one species.

<table>
<thead>
<tr>
<th>Statistic</th>
<th><em>A. trutta</em></th>
<th><em>A. truttaceus</em></th>
<th><em>A. xylabion</em></th>
<th><em>A. georgianus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>89</td>
<td>143</td>
<td>25</td>
<td>261</td>
</tr>
<tr>
<td>$P_O$</td>
<td>0.25</td>
<td>0.17</td>
<td>0.08</td>
<td>0.33</td>
</tr>
<tr>
<td>$N_A$</td>
<td>20</td>
<td>9</td>
<td>5</td>
<td>32</td>
</tr>
<tr>
<td>$SN_A$</td>
<td>20</td>
<td>8.6</td>
<td>5</td>
<td>15.2</td>
</tr>
<tr>
<td>$H_E$</td>
<td>0.312</td>
<td>0.239</td>
<td>0.093</td>
<td>0.249</td>
</tr>
</tbody>
</table>

($\pm 0.208$) ($\pm 0.241$) ($\pm 0.161$) ($\pm 0.077$)
6.3.2.2 Comparisons of Genetic Diversity Between Species

As implied above, the patterns in the mtDNA and nDNA data consistently indicated that *A. trutta* and *A. georgianus* had the highest levels of diversity, followed by *A. truttaceus* and then, *A. xylabion*, with very low levels of diversity (Tables 6.4, 6.5). An exception was that the number of cytochrome *b* haplotypes in *A. georgianus* was less than what might be expected based on the number of intron alleles and intron diversity (Table 6.4; Figure 6.3). This exception held for the number of haplotypes, but not for haplotype diversity (Table 6.5; Figure 6.3).

### Table 6.4

<table>
<thead>
<tr>
<th>Locus</th>
<th><em>A. trutta</em> (Rank)</th>
<th><em>A. truttaceus</em> (Rank)</th>
<th><em>A. xylabion</em> (Rank)</th>
<th><em>A. georgianus</em> (Rank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cytochrome <em>b</em></td>
<td>10 (1)</td>
<td>3.1 (2)</td>
<td>1 (4)</td>
<td>3 (3)</td>
</tr>
<tr>
<td>AldoB4</td>
<td>6 (1)</td>
<td>2 (3)</td>
<td>1 (4)</td>
<td>4.7 (2)</td>
</tr>
<tr>
<td>CaM4</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>8 (1)</td>
</tr>
<tr>
<td>CK7</td>
<td>4 (2)</td>
<td>1 (3.5)</td>
<td>1 (3.5)</td>
<td>6.8 (1)</td>
</tr>
<tr>
<td>S72</td>
<td>9 (1)</td>
<td>4.6 (3)</td>
<td>2 (4)</td>
<td>5.2 (2)</td>
</tr>
</tbody>
</table>
Table 6.5  Haplotype diversity (cytochrome b) and expected heterozygosity (introns) for each locus in the species of *Arripis*. Species are ranked (parentheses) from highest to lowest for each locus. Mean rank was used for ties (see Section 6.2.4.2).

<table>
<thead>
<tr>
<th>Locus</th>
<th>A. trutta</th>
<th>A. truttaceus</th>
<th>A. xylabion</th>
<th>A. georgianus</th>
</tr>
</thead>
<tbody>
<tr>
<td>cytochrome b</td>
<td>0.64 (1)</td>
<td>0.16 (3)</td>
<td>0.00 (4)</td>
<td>0.47 (2)</td>
</tr>
<tr>
<td><em>AldoB4</em></td>
<td>0.37 (2)</td>
<td>0.44 (1)</td>
<td>0.00 (4)</td>
<td>0.24 (3)</td>
</tr>
<tr>
<td><em>CaM4</em></td>
<td>0.00 (3)</td>
<td>0.00 (3)</td>
<td>0.00 (3)</td>
<td>0.65 (1)</td>
</tr>
<tr>
<td><em>CK7</em></td>
<td>0.25 (1)</td>
<td>0.00 (3.5)</td>
<td>0.00 (3.5)</td>
<td>0.23 (2)</td>
</tr>
<tr>
<td><em>S72</em></td>
<td>0.61 (1)</td>
<td>0.51 (2)</td>
<td>0.36 (4)</td>
<td>0.37 (3)</td>
</tr>
</tbody>
</table>

6.3.3  Neutrality of Markers

Each locus ranked the diversity of the species, relative to each other, in more or less the same order (Tables 6.4 and 6.5) for both haplotype/allele number ($W = 0.64, p < 0.01$) and haplotype diversity/expected heterozygosity ($W = 0.47, p < 0.05$). This is consistent with the view that the patterns of diversity at each locus in each species are (more or less) selectively neutral and relate to factors controlling diversity at neutral loci. Therefore, unless otherwise stated, the remainder of the results are interpreted on the assumption that the patterns of variation at the sampled loci are selectively neutral.
Figure 6.3  Relationship between cytochrome $b$ and intron locus diversity for each of: *Arripis trutta* (yellow), *A. truttaceus* (pink), *A. xylabion* (orange) and *A. georgianus* (blue).  

A. standardised number of haplotypes against standardised number of alleles.  

B. haplotype diversity against expected heterozygosity. Intron loci are identified by shape: *AldoB4* (♦), *CaM4* (■), *CK7* (▲), *S72* (●). Graphics have been added to illustrate where the data points for *A. georgianus* are outliers (circled) compared to an approximate trend line for the three species of Australian Salmon.
6.3.4 Demographic History

Three aspects of demographic history were considered: (i) changes in population size; (ii) the evolutionary age of any such changes; and (iii) the effective population size at the start of such changes. The results of each are presented below.

6.3.4.1 Tests for Evidence of Changes in Population Size

6.3.4.1.1 Mitochondrial DNA sequences

The patterns of cytochrome \( b \) sequence variation for \( A. trutta \) and, to a lesser degree \( A. truttaceus \), showed evidence of departures from mutation-drift equilibrium and in particular of a population expansion, as follows. Firstly, the values of both Tajima’s \( D \) and Fu’s \( F_s \) were negative (and statistically significant; Table 6.6) for the samples of both species, reflecting the excess of rare haplotypes (Section 6.3.1.3). Secondly, the values of \( R_2 \) for both species were low, as expected under rapid population growth, although bootstrapping for this test indicated a significant disparity from mutation-drift equilibrium only in \( A. trutta \) (Table 6.6). Thirdly, the observed mismatch distribution of \( A. trutta \) corresponded closely to that predicted under a model of population expansion (Figure 6.4) and did not differ significantly from it (Table 6.6). Although a history of population expansion equally could not be rejected for \( A. truttaceus \) (Table 6.6), the extremely low level of polymorphism in this species generated curves that were almost identical for models of either population expansion or a stable population (Figure 6.4). This is typical of species with very low variability and/or those recovering from either a very severe or very recent bottleneck (Harpending 1994; Ramos-Onsins and Rozas
2002). Fourthly, a low raggedness index with significant bootstrap support indicated an expanding population in *A. trutta*, but was not repeated in *A. truttaceus* (Table 6.6).

The results of the mtDNA-based analyses did not provide strong evidence of a population expansion in *A. georgianus* although, based on the mismatch SSD, a model of population expansion could not be rejected (Table 6.6; Figure 6.4).

**Table 6.6** Tests for evidence of changes in population size in *Arripis* species.

Test results and associated probabilities (*p*) are presented for Tajima’s *D*, Fu’s *Fs*, Ramos-Onsins and Rozas *R*₂, Harpending’s raggedness *rg*, and Mismatch SSD.

<table>
<thead>
<tr>
<th>Statistic</th>
<th><em>A. trutta</em></th>
<th><em>A. truttaceus</em></th>
<th><em>A. xylabion</em></th>
<th><em>A. georgianus</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D</em></td>
<td>-2.093</td>
<td>-1.723</td>
<td>-</td>
<td>-0.024</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.001</td>
<td>0.020</td>
<td>-</td>
<td>0.450</td>
</tr>
<tr>
<td><em>Fs</em></td>
<td>-7.793</td>
<td>-3.695</td>
<td>-</td>
<td>0.058</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.000</td>
<td>0.001</td>
<td>-</td>
<td>0.501</td>
</tr>
<tr>
<td><em>R</em>₂</td>
<td>0.058</td>
<td>0.092</td>
<td>-</td>
<td>0.125</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.000</td>
<td>0.214</td>
<td>-</td>
<td>0.401</td>
</tr>
<tr>
<td><em>rg</em></td>
<td>0.042</td>
<td>0.487</td>
<td>-</td>
<td>0.173</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.022</td>
<td>0.593</td>
<td>-</td>
<td>0.236</td>
</tr>
<tr>
<td>SSD</td>
<td>0.0020</td>
<td>0.0007</td>
<td>-</td>
<td>0.0540</td>
</tr>
<tr>
<td><em>p</em></td>
<td>0.694</td>
<td>0.380</td>
<td>-</td>
<td>0.460</td>
</tr>
</tbody>
</table>
6.3.4.1.2 Nuclear DNA allele frequencies

The patterns of genetic diversity at the intron loci for *A. trutta* revealed a heterozygosity deficiency at each polymorphic locus (three), relative to predictions for populations in mutation-drift equilibrium under each mutational model (IAM and SMM), and therefore suggest a population expansion (Table 6.7). In contrast, the patterns for *A. truttaceus* indicated an excess of heterozygosity at each polymorphic locus (two), for all comparisons except for the S72 locus under the SMM, thus providing some evidence of a relatively recent population bottleneck (Table 6.7). As explained in Section 6.2.4.4.1.2, it was not possible to test the significance of these differences.

The heterozygosity deficiency at all loci also provides evidence of a population expansion in the Australian Herring *A. georgianus* (Table 6.7). Little can be drawn from the excess of heterozygosity at the single locus examined in the other species of Australian Salmon *A. xylabion* (Table 6.7).
Figure 6.4  Mismatch distributions for A. *Arripis trutta*, B. *Arripis truttaceus* and C. *Arripis georgianus*, indicating the observed frequency of nucleotide differences between pairs of individuals in the sample (columns). The solid line represents the expected distribution under a model of population expansion and the dotted line represents the distribution expected in a stable population. In *A. truttaceus*, expectations for the two models were effectively the same and so both are represented by the single solid line.
Table 6.7  Expected heterozygosity of the actual data ($H_E$) and heterozygosity expected at mutation-drift equilibrium given the actual number of alleles under the infinite alleles (IAM), $H_{ExpI}$, and stepwise mutation (SMM), $H_{ExpS}$, models for each polymorphic intron locus in each species of *Arripis*. Expected values and standard deviation (s.d.) under each model were calculated by 10,000 replications using BOTTLENECK. The prediction of heterozygote deficiency ($H_{def}$) or excess ($H_{exc}$) was based on the direction of the equation ($H_E - H_{Exp}$).

<table>
<thead>
<tr>
<th>Locus</th>
<th>$H_E$</th>
<th>$H_{ExpI}$</th>
<th>s.d.</th>
<th>predict</th>
<th>$H_{ExpS}$</th>
<th>s.d.</th>
<th>predict</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Arripis trutta</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AldoB4</td>
<td>0.366</td>
<td>0.532</td>
<td>0.162</td>
<td>$H_{def}$</td>
<td>0.716</td>
<td>0.070</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td>CK7</td>
<td>0.254</td>
<td>0.397</td>
<td>0.183</td>
<td>$H_{def}$</td>
<td>0.575</td>
<td>0.110</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td>S72</td>
<td>0.606</td>
<td>0.668</td>
<td>0.123</td>
<td>$H_{def}$</td>
<td>0.816</td>
<td>0.040</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td><em>Arripis truttaceus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AldoB4</td>
<td>0.442</td>
<td>0.156</td>
<td>0.166</td>
<td>$H_{exc}$</td>
<td>0.198</td>
<td>0.170</td>
<td>$H_{exc}$</td>
</tr>
<tr>
<td>S72</td>
<td>0.514</td>
<td>0.446</td>
<td>0.181</td>
<td>$H_{exc}$</td>
<td>0.650</td>
<td>0.090</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td><em>Arripis xylabion</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S72</td>
<td>0.372</td>
<td>0.227</td>
<td>0.163</td>
<td>$H_{exc}$</td>
<td>0.270</td>
<td>0.164</td>
<td>$H_{exc}$</td>
</tr>
<tr>
<td><em>Arripis georgianus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AldoB4</td>
<td>0.234</td>
<td>0.414</td>
<td>0.191</td>
<td>$H_{def}$</td>
<td>0.641</td>
<td>0.091</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td>CaM4</td>
<td>0.158</td>
<td>0.566</td>
<td>0.161</td>
<td>$H_{def}$</td>
<td>0.778</td>
<td>0.052</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td>CK7</td>
<td>0.230</td>
<td>0.663</td>
<td>0.131</td>
<td>$H_{def}$</td>
<td>0.842</td>
<td>0.033</td>
<td>$H_{def}$</td>
</tr>
<tr>
<td>S72</td>
<td>0.371</td>
<td>0.570</td>
<td>0.159</td>
<td>$H_{def}$</td>
<td>0.780</td>
<td>0.051</td>
<td>$H_{def}$</td>
</tr>
</tbody>
</table>
Since strong mtDNA evidence of population expansion was found only in the species of primary interest (*A. trutta* and *A. truttaceus*), estimates of the time since the start of the expansion, and the effective female population size at that time, were limited to these two species. Estimates based on both mismatch and coalescent methods indicated that the detected population expansion of *A. trutta* commenced in excess of 100,000 years ago, but only 17,000 years ago in *A. truttaceus*, although the confidence intervals for the estimates for each species overlapped (Table 6.8). Associated estimates suggested that the size of the female population at the start of the expansion was higher in *A. trutta* than in *A. truttaceus* (Table 6.8).

**Table 6.8** Estimates of the time since the start of population expansion and female effective population size at the start of expansion ($N_f(0)$) in *Arripis trutta* and *A. truttaceus*. Estimates are based on mismatch distributions and coalescent simulations. Mutational time, $\tau$, and theta, $\theta_0$, were estimated as described in Section 6.2.4.4.2. The 95% confidence intervals derived from bootstrapping are provided for the mismatch method.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\tau$</th>
<th>Time (yrs)</th>
<th>$\theta_0$</th>
<th>$N_f(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. trutta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mismatch</td>
<td>1.028</td>
<td>106,418</td>
<td>0.04</td>
<td>2070</td>
</tr>
<tr>
<td>(95% C.I.)</td>
<td>(0.316 - 1.739)</td>
<td>(32,712 – 180,020)</td>
<td>(0 - 0.241)</td>
<td>(0 – 12,474)</td>
</tr>
<tr>
<td>coalescent</td>
<td>1.051</td>
<td>108,799</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>A. truttaceus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mismatch</td>
<td>0.167</td>
<td>17,288</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(95% C.I.)</td>
<td>(0.000 - 0.729)</td>
<td>(0 – 75,466)</td>
<td>(0 - 0.005)</td>
<td>(0 - 259)</td>
</tr>
<tr>
<td>coalescent</td>
<td>0.167</td>
<td>17,288</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
6.4 Discussion

All classes of genetic marker studied so far, i.e. the cytochrome \( b \) and EPIC markers of this study and the allozyme data of MacDonald (1980) and Ayvazian et al. (2004), indicate that the levels of genetic diversity in all extant representatives of the family Arripidae are relatively low in comparison with the levels typically detected in other abundant marine teleosts (e.g. Grant and Bowen 1998; Hassan et al. 2002). These markers also suggest that the amount of diversity varies more or less predictably among different members of the Arripidae, with species in order from highest diversity to lowest being \( A. trutta \), \( A. georgianus \), \( A. truttaceus \) and \( A. xylabion \) (notwithstanding the limited spatial sampling that was conducted for this last species). The above two patterns of diversity are likely to be linked strongly to demographic stochasticity in these species, all of which may be particularly susceptible to the effects of glacial cycles (see below). Overall the mtDNA and nDNA data indicate that the populations of Australian Salmon east and west of the Bass Strait are not in mutation-drift equilibrium and suggest that \( A. truttaceus \) has undergone a more recent and/or more significant population expansion compared to \( A. trutta \). This latter finding is consistent with the \textit{a priori} prediction that \( A. truttaceus \) has been more severely impacted by glacial cycles than \( A. trutta \) by virtue of its distribution to the west of Bass Strait (see Section 6.1.1).

6.4.1 Genetic Diversity in \textit{Arrpis}

While it is difficult to compare across studies, the available evidence indicates that the levels of genetic diversity in the extant species of the Arripidae are relatively low compared to those typically seen in other abundant marine teleosts. This is despite the
fact that data for each of three of the four arripid species are based on samples from across the entire Australian distribution of the species. For example, in data reported for the cytochrome \( b \), ND4/5 or cytochrome oxidase regions of the mtDNA for 31 marine teleosts in Table 4 of Grant and Bowen (1998), 23 (74\%) and 22 (71\%) species had values of haplotype diversity and nucleotide diversity, respectively, that were higher than those of \textit{A. trutta} (the most diverse of the Arripidae in this respect). In addition, in data reported for intron length polymorphism for 13 marine teleosts in Hassan \textit{et al.} (2002), 9 (69\%) species had a higher proportion of polymorphic loci than \textit{A. georgianus} (the most diverse of the Arripidae in this respect). Furthermore, Parker \textit{et al.} (1998) reported that the average proportion of polymorphic allozyme loci in 183 species of fishes was 21\%, which is considerably higher than the highest value recorded in a species of \textit{Arripis} (\textit{A. trutta} with 3/27 (11\%); MacDonald 1980). The fact that the results for each \textit{Arripis} species were fairly consistent over all classes of marker studied so far suggests that the low levels of diversity in each species are genome-wide. However, more intensive sampling of the genome, e.g. via techniques, such as amplified fragment length polymorphism, that simultaneously amplify a large number of loci (e.g. Fischer \textit{et al.} 2011), is required to confirm this.

The exact reason for the generally low levels of genetic diversity in the Australian arripids is not clear, but in general terms must reflect either a low rate of molecular evolution and/or a relatively high rate of loss of genetic variation, which could be driven by either genetic drift and/or natural selection, in each of these species (see Amos and Harwood 1998). The rate of molecular evolution in the family Arripidae is not known and cannot readily be determined (see Section 5.2.2.3.2), however there is no compelling reason to believe that this rate is dramatically different to that in other
teleosts. Genome-wide reductions in diversity are often attributed to a small effective population size associated with either a persistently small population size and/or a population bottleneck(s) (e.g. Watterson 1984; Avise 2000; Campos et al. 2010), although Amos and Harwood (1998) have argued that this explanation is rarely sufficient on its own and that selective sweeps may be more important than generally recognised. Regardless, a small effective population size provides a plausible explanation for the low levels of genetic diversity in each Arripis species, particularly if one invokes recurrent bottlenecks associated with Pleistocene glacial events and their effects on coastal environments.

If the low levels of genetic diversity in each Arripis species are mainly due to Pleistocene bottlenecks, then this implies that these species are more susceptible to the effects of glaciation than many other teleosts that occur in temperate coastal waters (because many such species do not show such low diversity). Although highly mobile, the Arripis species could nevertheless be particularly susceptible to the effects of glacial events if, for example, current flows that maintain critical linkages between larval, juvenile and adult distributions (see Gillanders et al. 2003) are disrupted (see Chapter 1 for the details of the unusually complex life-cycle of these species). It is also possible that the Arripis species have particularly narrow thermal tolerances and/or range of spawning temperature. While there is no specific evidence to support or refute this idea, it is relevant that, where known, spawning in each species is highly temporally and spatially restricted (see Chapter 1) and temperature is likely to be an important cue in the formation of the spawning aggregations (Stewart et al. 2011).
6.4.2 Recent Evolutionary History of *Arripis trutta* and *Arripis truttaceus*

Both the mtDNA data (e.g. ‘star’ phylogenies, neutrality tests, mismatch distributions) and the EPIC data (e.g. heterozygosity deficiency) provide evidence of population expansion in *A. trutta*, with the expansion reflected in the mtDNA dated to around 100,000 years ago. In contrast, while the mtDNA data also suggest a population expansion in *A. truttaceus* (as per *A. trutta*), but starting considerably more recently than *A. trutta* (around 20,000 years ago), the intron loci (e.g. heterozygosity excess) suggest a contraction. The two different classes of markers suggest different demographic scenarios (expansion and contraction) for *A. truttaceus* probably because they are each measuring demographic changes over different time-scales due to different rates of mutation (e.g. Eytan and Hellberg 2010). In particular, since the mutation rate of a vertebrate mtDNA locus is typically around ten times faster than that of a nDNA locus (e.g. Brown *et al.* 1979, but see Ballard and Whitlock 2004), it is likely that *A. truttaceus* experienced an expansion following a contraction. In the case of *A. trutta*, assuming that this species also underwent a contraction before the detected expansion(s), evidence of the contraction may no longer be apparent in either the EPIC or mtDNA data if the contraction occurred a relatively long time ago.

The results of this chapter indicate that *A. truttaceus* has experienced more dramatic and/or recent population expansion than *A. trutta*, as outlined below.

Firstly, the levels of genetic diversity are consistently higher in *A. trutta* compared to *A. truttaceus*, although the difference could not be tested statistically based on the data from the four polymorphic loci of this study. However, by also including data for five
polymorphic allozyme loci from MacDonald (1980), a Wilcoxon sign-rank test (Gehring 1978) demonstrated that *A. trutta* had a significantly higher diversity than *A. truttaceus* for both the standardised number of haplotypes/alleles ($p = 0.004$) and haplotype diversity/expected heterozygosity ($p = 0.019$) (data not shown).

Secondly, while the mismatch distributions of the mtDNA sequence data (and associated neutrality tests) provide evidence of population expansions in both *A. trutta* and *A. truttaceus*, the estimated time of the start of the expansion in the latter species coincides with the end of the peak LGM, some 20,000 years ago (see Lambeck *et al.* 2002), which is considerably more recent than that estimated for the former (~ 100,000 years ago).

Thirdly, the results of Cornuet and Luikart’s (1996) bottleneck tests using intron loci imply a population expansion in *A. trutta* and a population bottleneck in *A. truttaceus*, although they are based on only a small number of loci with no statistical testing. Note that, although a heterozygosity deficit was suggested at one nDNA locus in *A. truttaceus* under the SMM, Cornuet and Luikart (1996) have shown that, under this model, some loci may show such a deficiency even during a bottleneck. Furthermore, the SMM is the model considered least likely to apply to the intron length frequency data (see Section 6.2.4.4.1.2). Adding the allozyme data (of MacDonald 1980, as above) to the bottleneck tests was considered, but there were few polymorphic loci, with very few alleles and extremely low $H_E$, rendering their addition to the analyses effectively meaningless or worse, erroneous (see Cornuet and Luikart 1996).
The apparent differences in the demographic histories of *A. trutta* and *A. truttaceus* are probably related to distributional differences between the species (e.g. Colson and Hughes 2007; Janko et al. 2007; McGaughran et al. 2010) rather than to biological or ecological differences (e.g. Fauvelot et al. 2003; Hickerson and Cunningham 2005). This is because the ecological and biological differences between these two species are in fact minimal (see Chapter 1). On the other hand, the results are consistent with expected differences in the impacts of glacial events on *A. trutta* (lesser) and *A. truttaceus* (greater) by virtue of their respective distributions to the east and west of the Bassian Isthmus (see Section 6.1.1).

The above conclusion that *A. truttaceus* was more severely impacted by the LGM (and/or other glacial events) than *A. trutta* is dependent on several assumptions. This includes the assumption that the rate of mutation is similar in both species (rather than higher in *A. trutta*), which seems reasonable given that they are closely related species (see Avise 2000; Arbogast et al. 2002). The conclusion is also partly based on the results of the bottleneck tests, which ideally should include a much larger number of loci (Cornuet and Luikart 1996). Similarly, the dating of the demographic changes is dependent upon accurate knowledge of the rate of molecular evolution in the Arripidae (which is not known, see Section 5.2.2.3.2), although it is noteworthy that the estimate of the start of the expansion in *A. truttaceus* coincides with the peak of the LGM. Furthermore, this conclusion also assumes that the results reflect aspects of the demographic histories of each species rather than a differential capture of polymorphism from their most recent common ancestor, and is supported by the estimates that date the demographic changes to more than 500,000 years after speciation (see Chapter 5). Finally, linking demographic stochasticity to Pleistocene glacial events
is necessarily based on inferential evidence (e.g. Avise 2000; Bowen et al. 2001; Beheregaray et al. 2002; Bernardi 2005; Crandall et al. 2008; López et al. 2010), although one of the strengths of this study is that it used sibling species to test a priori predictions about the impacts of glacial events. Despite these assumptions, multiple aspects of the data set are nevertheless consistent with the above conclusion.

This is the second study to provide genetic evidence in support of the view that populations to the west of the Bass Strait have been more severely impacted by glacial events than comparable populations to the east. The first study compared eastern and western populations of the Bull Kelp Durvillaea potatorum (Fraser et al. 2009). In addition, other single-species studies have found higher levels of genetic diversity in eastern populations compared to their western counterparts in a gastropod (Brown 1991) and a fish (Colgan and Paxton 1997), although these studies were descriptive rather than providing explicit tests regarding the differential impact of glaciation events on the east and west side of the Bass Strait. The present study is unique in that it has compared eastern and western sibling species and has therefore potentially provided a longer-term perspective.

6.4.3 Comparison of Arripis truttaceus and Arripis georgianus

In contrast to A. truttaceus, with which it is broadly co-distributed, the Australian Herring A. georgianus did not show clear evidence of a population contraction followed by an expansion linked to the LGM. This implies that the impact of the LGM was not as great on A. georgianus as for A. truttaceus. Different responses to the LGM in these two species may be linked to subtle, but important, differences in their distributions. In
particular, the present distribution of *A. georgianus* extends much further up the west coast of Australia than that of *A. truttaceus* (see Chapter 1) and the temperature effects of the LGM were probably less severe along the west coast than on the south coast where the latter species mainly resides (see Section 6.1.1; Figure 6.1). In addition, *A. georgianus* may differ in some biological attributes (see Chapter 1) that enhance resilience to glacial changes (see Kaustuv et al. 2001; Wares and Cunningham 2001; Rockwood 2006). For example, the diet of *A. georgianus* is much more general than that of *A. truttaceus* and resilience to glacial changes is thought to be higher among diet-generalists (Hickerson and Cunningham 2005).

### 6.4.4 Recent Evolutionary History of *Arripis xylabion*

The sample of the third species of Australian Salmon, *A. xylabion* from Lord Howe Island had almost no genetic diversity, although it is not clear if more genetic diversity is present elsewhere in the species distribution. This lack of diversity prohibited a detailed analysis of the recent evolutionary history of this species. The low diversity at Lord Howe Island may reflect a: (i) persistently small population size; and/or (ii) severe bottleneck (or series of bottlenecks); and/or (iii) founding of the population by a small number of individuals. Although it is not possible to distinguish among these alternatives, it has been proposed that this species arose via peripheral isolation (founder effect) in the mid to late Pleistocene (Chapter 5) and thus may have started with relatively low levels of genetic diversity. It is also likely that the population of this species at Lord Howe Island may have experienced repeated bottlenecks in response to subsequent glacial events and/or other factors (see Bostock et al. 2006; Woodroffe et al. 2006). For example, the fringing coral reefs of the island are Holocene in origin.
implying that the fauna was influenced by the LGM and the subsequent Holocene marine transgression (see Woodroffe et al. 2006).

6.4.5 Conclusions

This study provides one of the first direct tests of the prediction that the populations to the west of the Bass Strait may have been more severely impacted by glacial cycles, relative to their east coast counterparts. The evidence was consistent with this prediction. In particular, the evidence suggests that *A. truttaceus*, which resides to the west of the Bass Strait, may have experienced a severe population bottleneck during the LGM followed by an expansion commencing some 20,000 years ago. In contrast, the population of *A. trutta*, which resides to the east of the Bass Strait, appears to have been largely unaffected by the LGM and has been expanding over the past 100,000 years or more. This general finding needs to be confirmed by examining other species, and the results for *A. georgianus* illustrate the value of conducting such comparisons using comparable populations (conspecific or sibling species) in order to minimise the effect of confounding factors.
Chapter 7. General Conclusions

7.0 General Conclusions

The purpose of this final chapter is to summarise the main conclusions of each of the data chapters. This study used the patterns of variation in a 483 bp fragment of the cytochrome \( b \) gene and a 654 bp fragment of the \( COI \) gene in the mtDNA, and at four nuclear intron loci to investigate aspects of the population genetic structure, phylogenetic relationships and recent evolutionary history of four species of \( Arripsis \).

The results of this study indicate the following:

1. Fragment length polymorphism at EPIC amplified intron loci is generally an effective approach for generating nuclear data using universal primers and avoiding the high costs associated with the development of other nDNA markers (e.g. microsatellites). Despite the use of capillary electrophoresis, binning of alleles into size classes was required in order to account for scoring variabilities and this may have masked some genetic variation or underestimated some measures of genetic diversity. This study found that EPIC primers that amplify shorter fragments (< 500 bp) should be favoured because they give highly consistent results, are polymorphic and the resultant data require considerably less binning. It also worth noting that the advent of next generation sequencing, which was not readily available during the planning phase of this research, has made the development of novel microsatellite markers more affordable (Santana et al. 2009; Metzker 2010).

2. There was no evidence for spatial genetic heterogeneity across almost the entire distributions of any of the three species of \( Arripsis \) in Australian waters – \( A. trutta \), \( A. truttaceus \) and \( A. georgianus \).
3. On the basis of knowledge of their life history, each species of *Arripis* probably has considerable gene flow across its distribution and each might comprise a single, panmictic population in Australian waters, with genetic homogeneity promoted by: (i) an extensive pelagic planktonic larval stage; (ii) extensive mixing by migrating juveniles and sub-adults; (iii) highly vagile and dispersive adults; and (iv) a relatively restricted spawning area.

4. The current four-species taxonomy of *Arripis* based on morphological data (i.e. Paulin 1993), was supported by molecular data on the basis that there were fixed nucleotide differences in mtDNA sequences (at least in two species) accompanied by significant differentiation in nDNA allele frequencies, interspecific genetic distances that were substantially greater than intraspecific distances and reciprocal monophyly inferred by all phylogenetic methods.

5. There was no definitive phylogenetic support for the putative geminate status of the species east and west of the Bass Strait (i.e. *A. trutta* and *A. truttaceus*, respectively). There was considerable evidence for a trifurcating polytomy where the three species of Australian Salmon (*A. trutta*, *A. truttaceus*, *A. xylabion*) arose at, or nearly at, the same time, although inferences may be limited by the resolving power of the data.

6. Regardless of the relationship between the Australian Salmon, the three species were estimated to have diverged from the most recent common ancestor in the mid to late Pleistocene, around 650,000 to 700,000 years ago.
7. Relative to other marine fishes, all four species of *Arripis* appear to exhibit generally low levels of genetic diversity in both the mtDNA and at nuclear intron loci. This is consistent with the low diversity reported for allozyme markers by MacDonald (1980).

8. Among the three closely related Australian Salmon species there was a consistent trend across all loci where *A. trutta* showed the highest genetic diversity, *A. truttaceus* had very low levels of diversity and *A. xylabion* almost none. The Australian Herring *A. georgianus* also showed a high diversity, relative to other *Arripis*, in all measures except the number of cytochrome *b* haplotypes.

9. The species of Australian Salmon to the west of the Bass Strait, *A. truttaceus*, may have been more severely impacted by glacial cycles, relative to *A. trutta* on the east coast. The evidence suggests that *A. truttaceus* may have experienced a severe population bottleneck during the Last Glacial Maximum followed by an expansion commencing some 20,000 years ago. In contrast, by virtue of its distribution, the population of *A. trutta* appears to have been largely unaffected by the LGM and has been expanding over the past 100,000 years or more.

10. A concordant response to the Last Glacial Maximum was not detected between the two westerly distributed species, *A. truttaceus* and *A. georgianus*. The genetic patterns in *A. georgianus* suggest that this species may have survived the LGM better than its sympatric congener *A. truttaceus* due to a combination of more westerly distribution and possible biological differences.
11. In summary, this study finds evidence of different evolutionary histories for species with east/west allopatric distributions across southern Australia. There are many putative geminate species in the region and this study should encourage many more similar studies. As suggested by Burridge (2000b), such studies should examine genetic variation in taxa across different distributional ranges, habitat requirements, life histories and dispersal capabilities. This study also adds weight to the importance of the Bass Strait as an example of a significant temperate biogeographical barrier and its role in shaping the temperate marine fauna of the Australian region. Given that the Bass Strait is now well established as a genetic isolating barrier (e.g. Burridge 1999; 2000b; Waters and Roy 2003; Waters et al. 2004; Dawson 2005; York et al. 2008; this study), future studies could focus on addressing the paucity of genetic data comparing species-level population and recent evolutionary histories. The many putative geminate species in the region provide an excellent opportunity for comparative studies in this context, especially for explicit hypothesis testing.
References


Bennett, I. and E.C. Pope (1953). Intertidal zonation of the exposed rocky shores of Victoria, together with a rearrangement of the biogeographical provinces of
References


References


References


References


References


Eshmeyer, W.N. (2011). *Catalog of Fishes (electronic version).*


References


References


References


Mateus, C.S., P.R. Almeida, B.R. Quintella and M.J. Alves (2011). MtDNA markers reveal the existence of allopatric evolutionary lineages in the threatened lampreys
References


Lampetra fluviatilis (L.) and Lampetra planeri (Bloch) in the Iberian glacial refugium Conservation Genetics: doi 10.1007/s10592-011-0209-x.


References


References


References


References


Waddell, V., M.C. Milinkovitch, M. Berube and M.J. Stanhope (2000). Molecular phylogenetic examination of the Delphinoidea trichotomy: congruent evidence from three nuclear loci indicates that Porpoises (Phocoenidae) share a more recent


References


Appendix A.1  Characteristics and amplification of mtDNA control region in *Arripis*.

As stated in Chapter 2 (Section 2.4.3), attempts to find reliable primers for the PCR amplification of the control region in the mtDNA of *Arripis* were unsuccessful despite the fact that a range of primer pairs and approaches was used. An overview of the characteristics and amplification of the mtDNA control region in *Arripis* is provided below.

Nucleotide sequence information was obtained from the control region of some individuals of *A. trutta* and *A. truttaceus* but not *A. georgianus* (using combinations of primer pairs *Atr15614L, Atr15621L* with *Atr16229H, Atr16295H, Atr16311H*; Table A.1). However, less than 60% of PCR amplification attempts resulted in interpretable sequences and no single set of primers was successful in all of those individuals from which the control region was amplified. Furthermore, despite extensive optimisation, double bands and smears (in the post-PCR gel) could not be eliminated in any of the samples that did amplify. No primers amplified any part of the control region sequence in *A. georgianus*. No attempt was made to amplify the control region in *A. xylabion*.

As implied above, mtDNA sequences of inconsistent length and position were generated for *A. trutta* and *A. truttaceus*. In addition, the presence a 55 bp variable number of tandem repeats (VNTR) was discovered at the 5′ end of the d-loop region of the control region (VNTR motif: 5′-ATG TTG ACC ATA CAT ATA TGC TTT AAA TAC ATA TAT GTA TAA TCA CCA TAA ATT T-3′). In *A. trutta*, between five and eight copies of the VNTR were detected among 13 sequences, and the last repeat was always 49 bp long (the last 6 bp had been deleted; an imperfect repeat at the 3′ end is a well known
phenomenon in VNTR – see Fumigalli et al. 1996). In *A. truttaceus*, between six and nine copies of the VNTR were detected among eight sequences, and all repeats were of equal length (i.e. no deletions). As is typical for VNTR (Lee et al. 1995; Lunt et al. 1998), the repeat sequence was AT rich (78%).

The incidence of VNTR in the mtDNA of vertebrates, including teleosts, appears to be quite common (e.g. Lee et al. 1995; Chen et al. 1998; Nesbø et al. 1998; Kong et al. 2007), and is usually attributed to slipped-strand replication derived from repair mechanisms and secondary structures that induce misalignment prior to elongation and replication of the d-loop strand (Lee et al. 1995; Fumigalli et al. 1996). Alignment of the repeat motif of *Arripis* with its reverse compliment revealed that most of the sequence was complementary, suggesting slipped-strand replication might be the source of VNTR in this genus (Gemmell et al. 1996).

VNTR have been useful markers in a range of vertebrates, using both sequence based or allele-based (repeat number) analyses (Broughton and Dowling 1997; Chen et al. 1998; Lunt et al. 1998; Nesbø et al. 1998; Ravago et al. 2002; Roques et al. 2004; Ketmaier and Bernardini 2005; Takagi et al. 2006; Kong et al. 2007). However, several hurdles still remain. For example, the presence of VNTR in the control (d-loop) region can cause problems during enzyme reading in the PCR reaction making PCR failure a common problem even when trying to amplify larger sequences of which the VNTR is only a small part (e.g. Lunt et al. 1998). Secondly, while the number of repeat copies might be scorable (as alleles) by counting bands on an electrophoretic gel, multiple heterogeneous motifs provide uninterpretable sequences downstream of the VNTR (due to multiple peaks on a chromatogram); this was common in the *Arripis* chromatograms.
A third serious problem arises from both unreliable PCR and heteroplasy; repeated PCR-amplifications of the same individual may result in different estimates of the number of repeat motifs (Lunt et al. 1998), as was the case in the assays of *Arripis*.

Ultimately, in this study, the control region, and its VNTR component in particular, were rejected as a useful genetic marker for the species of *Arripis* for the following reasons. Firstly, as described above, at least two technical hurdles regarding the use of VNTR are yet to be resolved (slippage and heteroplasy) and both appear likely to be affecting the *Arripis* sequence. Secondly, despite targeting priming sites inside and outside of the control region (Table A.1), amplification of PCR product with interpretable sequence in *A. trutta* and *A. truttaceus* was only possible in a few individuals, and not at all possible in *A. georgianus*. To avoid these problems, some workers have avoided VNTR by targeting primer sites within Conserved Sequence Blocks (CSB), especially CSB-D (e.g. Lee et al. 1995), however, this was also unsuccessful in *Arripis* (see below).

Many primer combinations were used in an attempt to PCR-amplify the mtDNA control region. Three different approaches were used to find reliable primers, namely: universal primers, customised primers and primer walking.

The first approach applied universal fish primers (Kocher et al. 1989; Meyer et al. 1990; Table A.1) that have been used to amplify partial-fragments of the d-loop of the control region in a range of other perciform species (e.g. Bearham 2004), but without success in *Arripis*. 
In a second approach, attempts were made to develop customised primers to amplify the control region in two ways. Firstly, a complete unpublished mtDNA genome sequence for a single individual of *A. trutta* (provided by M. Miya, Natural History Museum, Chiba, Japan) was used to design a range of *A. trutta* specific primers in: (i) the tRNA genes at each end of the control region; (ii) the conserved region (CSB) in the middle of the control region; (iii) the position of universal d-loop fish primers; (iv) cytochrome *b*; and (v) *16S* rRNA (Table A.1). As described above, some PCR-amplifications were achieved, but were not reliable and were complicated by the discovery of the VNTR. Secondly, using the BLAST function of GenBank (Benson *et al.* 2007), species for which parts of the mtDNA were similar to *Arripis* were identified. Sequence alignments involving these identified species were used to design primers based in the conserved parts of the genes that flank the control region (tRNA Threonine, tRNA Proline, tRNA Phenylalanine; Table A.1), although none successfully amplified any PCR product.

In the third approach, alignments, as described above, were used to design primers in the cytochrome *b* and *12S* rRNA genes and fill in the ~5 kb gap between (spanning the control region) using ‘primer walking’ (Gromek and Kaczorowski 2005; Table A.1). This failed to amplify any PCR product, even when high-fidelity long-range polymerase was used (Phusion™ Flash; Finnzymes Oy).
Table A.1  MtDNA primers tested for PCR-amplification of the control region in this study. The region/gene in which the primer was situated is indicated. Sequences use IUPAC codes for nucleotides.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Region</th>
<th>Strand</th>
<th>Sequence (5′–3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Universal Primers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L15926</td>
<td>tRNA&lt;sub&gt;Proline&lt;/sub&gt;</td>
<td>forward</td>
<td>AACTCTCACCCCTAGCTCCCAAAG</td>
</tr>
<tr>
<td>H16498</td>
<td>control region</td>
<td>reverse</td>
<td>CCTGAAGTAGGAACCAGATG</td>
</tr>
<tr>
<td><strong>Primers developed from indirect sequence alignment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Thr F</td>
<td>tRNA&lt;sub&gt;Thr&lt;/sub&gt;</td>
<td>forward</td>
<td>TAARCCGRAGYCGGAGGTT</td>
</tr>
<tr>
<td>tRNA-Pro F</td>
<td>tRNA&lt;sub&gt;Pro&lt;/sub&gt;</td>
<td>forward</td>
<td>TTTARTTTAGAATYCTRGCTTTTG</td>
</tr>
<tr>
<td>tRNA-Phe R</td>
<td>tRNA&lt;sub&gt;Phe&lt;/sub&gt;</td>
<td>reverse</td>
<td>TCTTAACAKCTTCAGTGTTATGCT</td>
</tr>
<tr>
<td><strong>Primers developed from <em>Arripis trutta</em> sequence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atr15806L</td>
<td>tRNA&lt;sub&gt;Proline&lt;/sub&gt;</td>
<td>forward</td>
<td>ATTCTACAACATTAACTATCCTTT</td>
</tr>
<tr>
<td>Atr15828L</td>
<td>tRNA&lt;sub&gt;Proline&lt;/sub&gt;</td>
<td>forward</td>
<td>CTTGCAGATAACTATTTGGACA</td>
</tr>
<tr>
<td>Atr16229H</td>
<td>control region</td>
<td>reverse</td>
<td>GATAGGTTCATAGGGTACACTT</td>
</tr>
<tr>
<td>Atr16295H</td>
<td>control region</td>
<td>reverse</td>
<td>CCTGATGTTGCGTTCTTATAC</td>
</tr>
<tr>
<td>Atr17270H-Phe</td>
<td>tRNA&lt;sub&gt;Phe&lt;/sub&gt;</td>
<td>reverse</td>
<td>GCTCATCCTAACTACCTCAGTG</td>
</tr>
<tr>
<td>Atr15778L</td>
<td>tRNA&lt;sub&gt;Proline&lt;/sub&gt;</td>
<td>forward</td>
<td>AACTCTACCCTAGCTCCCA</td>
</tr>
<tr>
<td>Atr16311H</td>
<td>control region</td>
<td>reverse</td>
<td>CATTAAGATAAATACCTGAT</td>
</tr>
<tr>
<td>Atr16495H</td>
<td>control region</td>
<td>reverse</td>
<td>TTGGAGTGCAACCTAGGGCAT</td>
</tr>
<tr>
<td>Atr16746H</td>
<td>control region</td>
<td>reverse</td>
<td>AGTTTTGCGGGGGGTAGG</td>
</tr>
<tr>
<td>Atr15065L-Cytb</td>
<td>cytochrome &lt;i&gt;b&lt;/i&gt;</td>
<td>forward</td>
<td>TACTCTCCCGATTCTTGTGCCTCC</td>
</tr>
<tr>
<td>Atr01250L-16S</td>
<td>16S rRNA</td>
<td>forward</td>
<td>GACAGAAAAAGAACTATTTGGAGCT</td>
</tr>
<tr>
<td>Atr01612H-16S</td>
<td>16S rRNA</td>
<td>reverse</td>
<td>CTTTCATCATCTCCTGTTTTC</td>
</tr>
<tr>
<td>Atr15737L</td>
<td>tRNA&lt;sub&gt;Thr&lt;/sub&gt;</td>
<td>forward</td>
<td>CGTTCAACCTCCCTTCA</td>
</tr>
<tr>
<td>Ageo15737L</td>
<td>tRNA&lt;sub&gt;Thr&lt;/sub&gt;</td>
<td>forward</td>
<td>CGGTGAATACCTGCATACTAA</td>
</tr>
<tr>
<td>Atr15614L</td>
<td>control region</td>
<td>forward</td>
<td>CCTCTACTTCTGTTATCTTCATCCT</td>
</tr>
<tr>
<td>Atr15621L</td>
<td>control region</td>
<td>forward</td>
<td>TTTCTGTATTCTCCTATCC</td>
</tr>
<tr>
<td>Atr16383H</td>
<td>control region</td>
<td>reverse</td>
<td>GGAACCTATGCGAGGAATA</td>
</tr>
<tr>
<td>Atr16739H</td>
<td>control region</td>
<td>reverse</td>
<td>GGGGGGTAGGGGGGGGT</td>
</tr>
<tr>
<td><strong>Primers developed for primer walking</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Arripis</em> Cytb</td>
<td>cytochrome &lt;i&gt;b&lt;/i&gt;</td>
<td>forward</td>
<td>TCCGTATGATGAAACTTC</td>
</tr>
<tr>
<td><em>Atrutt</em> Cytb</td>
<td>cytochrome &lt;i&gt;b&lt;/i&gt;</td>
<td>forward</td>
<td>CGAGAAACAAGGTTTAGGATGA</td>
</tr>
<tr>
<td><em>Arripis</em> 12S</td>
<td>12S rRNA</td>
<td>reverse</td>
<td>TCTCGGTGTAAGGGA</td>
</tr>
<tr>
<td><em>Ageo</em> 12S</td>
<td>12S rRNA</td>
<td>reverse</td>
<td>CTTCTCGGTGTAAGGGAATG</td>
</tr>
</tbody>
</table>
Appendix A.2  EPIC-PCR primers tested in this study. Genes in bold were polymorphic in at least one species and subsequently assayed. Genes marked with an asterisk (*) amplified, but were monomorphic within each species. Sequences use IUPAC codes for nucleotides.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Intron Name</th>
<th>Primer Name</th>
<th>Strand</th>
<th>Primer Sequence (5′ – 3′)</th>
<th>Reference/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actin (intron 1)</td>
<td>Act1</td>
<td>PmActI-F</td>
<td>forward</td>
<td>CCATACCTTCTACAATGAGCTCCG</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PmActI-R</td>
<td>reverse</td>
<td>GACCAGAGGCCTAGAGGCAGGC</td>
<td></td>
</tr>
<tr>
<td>Actin (intron 2)</td>
<td>Act2</td>
<td>Act-2-F</td>
<td>forward</td>
<td>GCATAACCCTCGTAGATGGGAC</td>
<td>2, 3, 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Act-2-R</td>
<td>reverse</td>
<td>ATCTGGACCCACCTTCTACAA</td>
<td></td>
</tr>
<tr>
<td>Aldolase B (intron 1) *</td>
<td>AldoB1</td>
<td>AldoB1-1F</td>
<td>forward</td>
<td>GCTCCAGGAAAGGGAATCTCGGC</td>
<td>6, 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AldoB1-2R</td>
<td>reverse</td>
<td>CTGCCAGGAAAGGGAATCTCGGC</td>
<td></td>
</tr>
<tr>
<td>Aldolase B (intron 2)</td>
<td>AldoB2</td>
<td>AldoB2F</td>
<td>forward</td>
<td>TCAGGGCATTGTCGTCG</td>
<td>2, 6, 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aldexon3R</td>
<td>reverse</td>
<td>CTGTTCCATTAGACCAGC</td>
<td></td>
</tr>
<tr>
<td>Aldolase B (intron 4)</td>
<td>AldoB4</td>
<td>Aldo5F</td>
<td>forward</td>
<td>GCCAGATATGCGAGCTCTGCC</td>
<td>6, 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Aldo3.1R</td>
<td>reverse</td>
<td>GGGTCCATCAGGAGATCTCTGCC</td>
<td></td>
</tr>
<tr>
<td>Aldolase C (intron 1)</td>
<td>AldoC1</td>
<td>AldoC1F</td>
<td>forward</td>
<td>CTTGCATCTCCAGGAGGTAC</td>
<td>2, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AldoC2R</td>
<td>reverse</td>
<td>TTCACCAGATCAATAC</td>
<td></td>
</tr>
<tr>
<td>Alpha amylase (intron 1) *</td>
<td>Am2B-1</td>
<td>Am2B1F</td>
<td>forward</td>
<td>CTTTCATCTCCAGGAGGTAC</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Am2B2R</td>
<td>reverse</td>
<td>TTCACCAGATCAATAC</td>
<td></td>
</tr>
<tr>
<td>Alpha amylase (intron 2)</td>
<td>Am2B-2</td>
<td>Am2B2F</td>
<td>forward</td>
<td>GGGGATGGGATGTCTTACACC</td>
<td>6, 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Am2B3R</td>
<td>reverse</td>
<td>AGGCCCTCCAGGAGGTCTCAG</td>
<td></td>
</tr>
<tr>
<td>β-globin (intron 1)</td>
<td>β-globin</td>
<td>β-1</td>
<td>forward</td>
<td>GTTGGTTGGAGGCCCTGGGAG</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-2</td>
<td>reverse</td>
<td>CCTGAGGTGTCTGAGGATCCATGCA</td>
<td></td>
</tr>
<tr>
<td>Calmodulin (intron 3)</td>
<td>CaM3</td>
<td>Cam-3-F</td>
<td>forward</td>
<td>TGGCGGAGCTGAGGCCTCCG</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cam-3-R</td>
<td>reverse</td>
<td>TGGAGGAGGCAGGCTCCG</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix A.2 continued.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Intron Name</th>
<th>Primer Name</th>
<th>Strand</th>
<th>Primer Sequence (5′ – 3′)</th>
<th>Reference/s</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Calmodulin (intron 4)</strong></td>
<td>CaM4</td>
<td>CalMex4F</td>
<td>forward</td>
<td>CTGACCATATGTCGCCAGAAA GTCACGTTCCCTCCCCAGGTT</td>
<td>3, 10, 11, 12</td>
</tr>
<tr>
<td><strong>Creatin kinase (intron 7)</strong></td>
<td>CK7</td>
<td>CK7F/CK8R</td>
<td>forward/reverse</td>
<td>AAGAGGGTCTTTGACAGGTTCTGC TTCTCTGGATCAGACGCTCCACC</td>
<td>6, 7, 13</td>
</tr>
<tr>
<td><strong>Elongation factor -1α (intron 6)</strong></td>
<td>EF1a6</td>
<td>EF16cF/EF17bR</td>
<td>forward/reverse</td>
<td>GAAACACTGGCTGAGAC AATGCGACATCTCCAGACT</td>
<td>11</td>
</tr>
<tr>
<td><strong>Elongation factor (intron 2)</strong></td>
<td>EF2</td>
<td>EF2-F/EF2-R</td>
<td>forward/reverse</td>
<td>GTCTTCTGCAGGACAAACTG GAGGCGAGGTTCTCTGACTCCAGC</td>
<td>1</td>
</tr>
<tr>
<td><strong>Glucose-6-phosphate dehydrogenase (intron 4)</strong></td>
<td>G6PD</td>
<td>G6Pdx4F/G6Pdx5R</td>
<td>forward/reverse</td>
<td>GACGAGACGTATTTGTGAG GACCGGTGAAGAGGCGGT</td>
<td>14</td>
</tr>
<tr>
<td><strong>Growth hormone (intron 1)</strong></td>
<td>GH1</td>
<td>SG1/FIS</td>
<td>forward/reverse</td>
<td>AGACCTGAACCACATGG AGGTGTTGAACCTGAGAC</td>
<td>15</td>
</tr>
<tr>
<td><strong>Growth hormone (intron 2)</strong></td>
<td>GH2</td>
<td>GH2F/GH3R</td>
<td>forward/reverse</td>
<td>AGGGTTCCTCCATGGCCATGCAGC TGTGGTGTGAGCGCTGGT</td>
<td>6, 7</td>
</tr>
<tr>
<td><strong>Growth hormone (intron 5)</strong></td>
<td>GH5</td>
<td>GH5F/GH6R</td>
<td>forward/reverse</td>
<td>AGCCAACTGAGCAGGC TGCCACTGTCAGATAAGTCTCC</td>
<td>6, 7, 8</td>
</tr>
<tr>
<td>**Gonadotropin-releasing hormone 3 (intron 1) * **</td>
<td>GnRH3-1</td>
<td>GnRH1F/GnRH1R</td>
<td>forward/reverse</td>
<td>AATGCGACCATGCTAAAGAGG CGACATCAGTCTGTTGC</td>
<td>6, 7, 8</td>
</tr>
<tr>
<td>**Gonadotropin-releasing hormone 3 (intron 2) * **</td>
<td>GnRH3-2</td>
<td>GnRH2F/GnRH2R</td>
<td>forward/reverse</td>
<td>AGAGATGGTGGAGAGCAGT AGAGGACACACTTCTGTCAC</td>
<td>6, 7, 16</td>
</tr>
<tr>
<td>Gene</td>
<td>Intron Name</td>
<td>Primer Name</td>
<td>Strand</td>
<td>Primer Sequence (5′ – 3′)</td>
<td>Reference/s</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------</td>
<td>--------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase (intron 2) *</td>
<td>Gpd2</td>
<td>GPD2F</td>
<td>forward</td>
<td>GCCATCAATGACCCCTTCATCG</td>
<td>6, 7, 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GPD3R</td>
<td>reverse</td>
<td>TTGACCTCACCCCTTGAAGCGGCG</td>
<td></td>
</tr>
<tr>
<td>Major histocompatibility complex class II antigen (intron 1)</td>
<td>MhcII</td>
<td>Mhc1F</td>
<td>forward</td>
<td>ACTCTAAATCTGGAGTACATGC</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mhc2R</td>
<td>reverse</td>
<td>CAGGAGATCTCTCTCAGCC</td>
<td></td>
</tr>
<tr>
<td>Myosin light chain B (intron 2) * †</td>
<td>Mlc-2-c</td>
<td>Mlc-2-F</td>
<td>forward</td>
<td>CTCGGGGTTCAGGACCTTGA</td>
<td>2, 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mlc-2-R</td>
<td>reverse</td>
<td>CTTCAAGGTAGCACCTCATG</td>
<td></td>
</tr>
<tr>
<td>Opsin (intron 1)</td>
<td>Ops1</td>
<td>Ops-1-F</td>
<td>forward</td>
<td>GCTCATGGGCGCTGAGACCACAA</td>
<td>5, 8, 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ops-1-R</td>
<td>reverse</td>
<td>CCTGCTCAACCTGAGCATGCC</td>
<td></td>
</tr>
<tr>
<td>S7 ribosomal protein (intron 1) *</td>
<td>S71</td>
<td>S7RPEX1F</td>
<td>forward</td>
<td>TGGCCTCTCTCTTGCGTCA</td>
<td>8, 17, 18, 19,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S7RPEX2R</td>
<td>reverse</td>
<td>AACTCGTCTGTTTCGCG</td>
<td>20, 21, 22,23</td>
</tr>
<tr>
<td>S7 ribosomal protein (intron 2)</td>
<td>S72</td>
<td>S7RPEX2F</td>
<td>forward</td>
<td>AGCGCCAAATACTGAGCC</td>
<td>17, 18, 21,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S7RPEX3R</td>
<td>reverse</td>
<td>GCTTCTACAGTTAGTCC</td>
<td>22, 23</td>
</tr>
<tr>
<td>Alpha tropomyosin (intron 1)</td>
<td>Tr1</td>
<td>Tr1F</td>
<td>forward</td>
<td>AGGGAAACAGAGGATAGGCTGTG</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tr1R</td>
<td>reverse</td>
<td>TCTCAGCTTCCCTACGAGCTGT</td>
<td></td>
</tr>
</tbody>
</table>

Appendix A.3  Technical issues regarding the use of fragment length polymorphism at intron loci.

The present study used nDNA markers in the form of fragment length polymorphism at intron loci. This type of nDNA marker was used because it provided many of the advantageous qualities of microsatellite markers (e.g. locus-specific, co-dominance, relatively high amounts of fragment length polymorphism; e.g. Li et al. 2002), but which could be amplified using universal primers (and thereby avoid the time and expense of developing customised microsatellite primers; see Chapter 2). Since this is one of few studies to use such markers to address population and evolutionary questions, and the first to use automated electrophoretic scoring methods with this type of marker, it is worthwhile to briefly consider some of the technical issues associated with the production of the intron data.

As discussed in Chapter 2, there were some ambiguities associated with determining the sizes of the alleles, particularly larger ones, at the intron loci. These ambiguities were manifested in two ways. Firstly, the raw estimates of the sizes of one or both alleles at a locus in repeat assays of the same individuals were sometimes different (by up to 1.5 bp; see Section 2.5.4). Secondly, the frequencies of raw estimates of the sizes of an allele at a locus tended to follow a normal distribution (Figures A.1, A.2), which was interpreted as reflecting minor scoring errors of a single allele. It is likely that aspects of both the PCR and post-PCR assays contributed to these ambiguities, as considered below.
Figure A.1  Example of the frequency distribution of the raw estimates of the fragment sizes around the 178 bp allele of locus *AldoB4* in all assayed *Arripis trutta*.

Figure A.2  Example of the frequency distribution of the raw estimates of the fragment sizes around the 900 bp allele of locus *S72* in all assayed *Arripis trutta*. The modal integer is 900.
PCR Assays

Due to limitations in the replication fidelity of the *Taq* polymerase (e.g. Tindall and Kunkel 1988), there may have been small differences in the sequence compositions in the amplicons of identical alleles in the present study (e.g. Taberlet *et al.* 1996; Davidson *et al.* 2003; Shinde *et al.* 2003). In absolute terms, the potential for these errors is obviously greater in longer fragments. This in turn, could have affected the electrophoretic mobility of the amplicons (see below).

Post-PCR Assays

Intron ‘alleles’ of the same length (number of base pairs) do not necessarily have identical sequence compositions and even a slight difference in composition can result in minor disparities in electrophoretic mobility (Anon. 2004). This effect is expected to be greater in longer fragments because the absolute number of mutations (and hence, differential electrophoretic migration) is positively correlated with sequence length (i.e. number of nucleotides) (e.g. Salamon *et al.* 1998; Sefc *et al.* 2003; Stres 2006). This is consistent with the observation that the longer DNA fragments of *CK7* and *S72* in this study had the lowest scoring repeatability (see Chapter 2).

The size standard used in this study (LIZ1200) showed evidence of decreasing accuracy with increasing fragment length (Figure A.3), which might be expected to result in less accurate scoring of fragment lengths for longer alleles (see Pereira *et al.* 2001; Symonds and Lloyd 2004, among others).

In general, the reliability of post-PCR assays may also have been affected by slight differences in assay reagents and/or conditions (see Pompanon *et al.* 2005), as suggested
Appendix A.3

by the fact that the disparity between repeat assays in the present study was typically greatest when the repeats were on different plates (see Chapter 2).

![Figure A.3](image)

**Figure A.3** Peak quality of LIZ1200 size standard in relation to fragment size. Peaks for smaller fragments are stronger, narrower and sharper than those of longer fragments.

**Other Issues**

In the present study, there was no evidence of contamination (negative controls were clear), inconsistent poly-A addition (see Section 2.5.4), or of electrophoretic artefacts in the assays. Consequently, although each of these factors has been reported to contribute to ambiguities in the scoring of length variants in other studies (e.g. Clark 1988; Smith *et al.* 1995; Fernando *et al.* 2001; Pompanon *et al.* 2005), none of them appears to have been a significant issue for the intron data in the present study.

**Conclusions.** Despite some ambiguities with the scoring of the alleles, particularly of the larger alleles, this study has demonstrated that fragment length polymorphism at intron loci can be used to investigate population and evolutionary questions in species for which genome data are not available and funds are limited. As per other studies (e.g. Gomulski *et al.* 1998; Berrebi *et al.* 2005; Hubert *et al.* 2006; Atarhouch *et al.* 2007), these problems were addressed by binning (see Chapter 2). However, the results of this study also suggest that it would be better to focus on loci with small to medium
sized alleles (< 500 bp) in order to reduce the potential for scoring problems (see Chapter 2). This also has the advantage of requiring a smaller size standard, which is cheaper than the LIZ1200 (which is required for larger loci). However, such a focus will potentially make it more difficult to find highly polymorphic loci, since the level of polymorphism typically increases with locus size (see Hare 2001; Vekemans et al. 2002), which could be a major problem in species with very low levels of polymorphism (like species of *Arripis*). Regardless, as with any genetic study, genotyping errors can never be completely eliminated because neither the assays nor human factors are 100% infallible (Broquet and Petit 2003; Bonin *et al.* 2004; Hoffman and Amos 2005; Pompanon *et al.* 2005).