2012

Among-Locus Heterogeneity in Genetic Diversity and Divergence in Two Pairs of Duck Species (Genus: Anas)

Kirandeep K. Dhami
Wright State University

AMONG-LOCUS HETEROGENEITY IN GENETIC DIVERSITY AND DIVERGENCE IN TWO PAIRS OF DUCK SPECIES (GENUS: *ANAS)*

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

By

KIRANDEEP K DHAMI

M.S, University of Idaho, 2006

2012

Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

Nov 30, 2012

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Kirandeep K Dhami ENTITLED Among-locus heterogeneity in genetic diversity and divergence in two pairs of duck species (Genus: *Anas)* BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Jeffrey L. Peters, Ph.D.
Dissertation Director

Donald Cipollini, Ph.D.
Director, ES Ph.D. Program

Dr. Andrew Hsu, Ph.D.
Dean, Graduate School

Committee on Final Examination

John O. Stireman III, Ph.D.

Volker Bahn, Ph.D.

Michael Raymer, Ph.D.

Kevin Omland, Ph.D.

ABSTRACT

Kirandeep K Dhami Ph.D., Environmental Sciences Ph.D. Program, Wright State University, 2012. Among locus heterogeneity in genetic diversity and divergence in two pairs of duck species (Genus: *Anas*).

Genetic diversity and divergence at a locus are the result of interactions among the fundamental evolutionary forces of mutation, genetic drift, gene flow and natural selection. Variation in the strength of these forces can cause high heterogeneity in diversity and divergence across the genome. The overall objective of this thesis was to examine the role of population history vs. selection in generating heterogeneity in genetic diversity and differentiation.

In Chapter 1, I examine the role of dispersal behavior in causing genetic differentiation and population structure within and between two morphologically distinct Australian duck species that differ in ecology and life history characteristics. A five-locus nuclear dataset revealed nearly no divergence and similar values of genetic diversity between species. However, as predicted, I found significant population structure in the sedentary chestnut teal (*Anas castanea*) but no structure within the vagile grey teal (*A. gracilis*).

In Chapter 2, a more rigorous examination of differentiation among nineteen autosomal loci also failed to uncover a genetic distinction between these two species. However, DNA sequences from seven loci sampled from the Z-chromosome revealed

strong differentiation between chestnut and grey teal. Furthermore, the most divergent loci are clustered on the shorter *p*-arm of the chromosome, close to the centromere, suggesting this region as an island of differentiation that may have been important in the speciation process. These two species of Australian teal are perhaps the most recently diverged taxa examined to date that reveal a large Z-effect.

In Chapter 3, I quantitatively tested the contribution of gene flow and introgression to the heterogeneity of genetic diversity and differentiation in two deeply divergent taxa, the falcated duck (*A. falcata*) and the gadwall (*A. strepera*). Consistent with previously published mitochondrial DNA analyses, 19 nuclear loci revealed the introgression of nuclear DNA from the falcated duck into the allopatric population of gadwall in North America, but not into the sympatric population in Eurasia. Furthermore, gene flow was insufficient to explain the high heterogeneity in genetic diversity in both species and differentiation between the taxa. Indeed, this heterogeneity failed to fit neutral models of population history, suggesting that selection may be having a pervasive effect throughout the genome.

Overall, this research reveals heterogeneous patterns of diversity and differentiation among nuclear loci in both early and later stages of divergence. Gene flow alone could not explain this heterogeneity, suggesting a prominent role of selection. Substantial divergence at some loci suggests that the strength of divergent selection overrides the homogenizing effects of gene flow and maintains species integrity.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

Firstly, my sincere gratitude to my advisor Dr. Jeffrey L. Peters, who made this work possible. His friendly guidance and expert advice helped me in all the stages of my research work. I could not have asked for a better role model and advisor for my Ph.D. study.

I would also like to thank all my committee members; John Stireman, Volker Bahn, Michael Raymer and Kevin Omland for all the encouragement and support.

I am also thankful to all my friends for making this journey of grad school more fun. Special thanks are due to my parents and sisters for all their unconditional love and support throughout my life.

Last but not the least, I extend my special thanks to the one above all of us, Guru Nanak, for answering my prayers and giving me the strength to accomplish this goal.

AMONG-LOCUS HETEROGENEITY IN GENETIC DIVERSITY AND DIVERGENCE IN TWO PAIRS OF DUCK SPECIES (GENUS: *ANAS)*

During the speciation process, incipient species gradually become more differentiated with time, largely through the independent effects of mutation and genetic drift in each population. Any ongoing gene flow will counter these forces and result in greater homogenization of the genomes, slowing speciation. However, these forces do not necessarily act uniformly across the genome, because natural selection can either inhibit or elevate the rate of divergence for some loci. Two processes can contribute to faster divergence for some loci: differential introgression or divergent selection. In this thesis, I examine the roles of these processes in two pairs of *Anas* spp. of ducks, which differ in their depths of divergence, by sequencing DNA sampled from a genomic transect.

Introgression, or interspecific gene flow, is the penetration of alleles from one species to the gene pool of the other through interspecific mating and subsequent backcrossing of the hybrids into parental populations. This mechanism of exchange of genes between species can be pervasive and important to the maintenance of genetic variation. Geographically, introgression could be restricted by dispersal potential of the species, and genomically it could be restricted by mating preferences, fitness of hybrids and selection. This irregularity in introgression can therefore cause

heterogeneous patterns of inter-specific allele sharing and intra-specific genetic variation. Introgression between species varies on three scales during the process of speciation; temporally, geographically and genomically. Temporally, gene flow can be ongoing between diverging taxa before reproductive isolation or hybrid sterility evolves. Hybrids birds can be fertile for 7–17 million years after divergence and only become completely inviable after 11–55 million years of divergence (Price and Bouvier, 2002). However, the proportion of viable hybrids decreases with an increase in genetic divergence (Mallet, 2007). Geographically, the extent of introgression at a marker strongly depends upon shared breeding ranges between hybridizing taxa or the species ability to disperse. High dispersal capabilities can overcome barriers imposed by geography (that can potentially restrict introgression) and could prevent genetic structuring between diverging taxa. Genomically, introgression across semi-permeable species boundaries can vary among loci for three reasons: sex-biased dispersal, Haldane's rule and selection.

Sex-biased dispersal is an important life history trait that has implications in regulating intraspecific gene flow and genetic structure. Differences in dispersal behavior between the sexes can cause variance in patterns of genetic structure for loci with different modes of inheritance. For example, when males are the dispersing sex, the maternally-inherited mtDNA will be more structured than biparentally-inhertited nuDNA. On the other hand, no or weak genetic structure at mtDNA is expected in the case of female-biased dispersal.

Hybrid viability and fitness can also cause differential introgression among markers with different modes of inheritance. Haldane (1922) proposed a rule that the

heterogametic sex of hybrids has a higher tendency to suffer inviability or sterility, which has been shown empirically for several pairs of species (e.g., Presgraves, 2002; Price and Bouvier, 2002). In female heterogametic taxa (such as birds and butterflies), reduced fitness or inviability of female hybrids restricts introgression at the maternally inherited mtDNA, whereas in male heterogametic taxa (such as mammals and flies), reduced fitness of male hybrids restricts the introgression of the Y-chromosome. These expectations have been supported in a wide array of taxonomic groups (Cianchi *et al*., 2003; Carling and Brumfield, 2008; Storchova *et al*., 2010).

Although introgression has the potential to completely homogenize the genomes of speciating taxa before reproductive isolation evolves, divergent selection can inhibit gene flow at genomic regions that demarcate species (Nosil *et al*., 2009) resulting in elevated divergence in those regions. Such regions of strong differentiation are spread throughout the genome in both autosomal and sex-chromosomes (Ellegren, 2012). However, evidence suggests that these outliers are disproportionately linked to the Z-chromosome in birds, presumably owing to hemizygosity in the heterogametic sex, low recombination, and linkage of traits involved in sexual selection (Backstrom *et al*., 2010), which implies a strong role of selection in inhibiting introgression and driving divergence. Genomic variation in introgression and divergence resulting from selection leads to heterogeneous pattern of differentiation and diversity across the genome.

While the interaction between divergent selection and introgression favors elevated divergence at some loci, positive selection can prevent loss in diversity and

inhibit divergence at other loci by selectively spreading alleles between populations and reducing the loss of alleles through drift (Bachtrog *et al*., 2006; Currat *et al*., 2008). Therefore, selection may favor the introgression of beneficial alleles at some loci but prevent gene flow at other loci, causing heterogeneous diversity across the genome (Tajima, 1989; Charlesworth, 1997; Schluter, 2001; Nordborg *et al*., 2005; Postma and Noordwijk, 2005; Storz and Kelly, 2008). A multi-locus approach could help elucidate the role of various demographic or genomic forces in the among-locus heterogeneity of genetic diversity and differentiation within and between taxa.

In this study, I examine the heterogeneity of genetic diversity and divergence by sampling multiple loci, each linked to a different chromosome (a genomic transect; Peters *et al*. 2012). In chapter 1, I examine the influence of dispersal behavior on spatial patterns of genetic structure and gene flow in two species of Australian ducks that are morphologically, behaviorally, and ecologically well-differentiated, but have nearly zero net-divergence in mtDNA . In chapter 2, I more rigorously examine the differentiation between these two species by adding more loci, and I test for a large contribution of the Z-chromosome in the early stages of speciation. Finally, in chapter 3, I examine a pair of hybridizing species with a deep divergence, and I quantitatively test the contribution of gene flow and introgression to the observed patterns of among-locus heterogeneity.

REFERENCES

Bachtrog D., K. Thornton, A. Clark, and P. Andolfatto. 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the Drosophila yakuba species group. Evolution 60:292-302.

Backstrom N., J. Lindell, Y. Zhang, E. Palkopoulou, A. Qvarnstrom, G. Saetre, and H. Ellegren. 2010. A High-Density Scan of the Z Chromosome in Ficedula Flycatchers Reveals Candidate Loci for Diversifying Selection. Evolution 64:3461-3475.

Carling M. D., and R. T. Brumfield. 2008. Haldane's rule in an avian system- using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the Passerina bunting hybrid zone. Evolution 62:2600-2615.

Charlesworth B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70.

Currat M., M. Ruedi, R. J. Petit, and L. Excoffier. 2008. The hidden side of invasions: Massive introgression by local genes. Evolution 62:1908-1920.

Haldane J. 1922. Sex ratio and unisexual sterility in hybrid animals. Journal of Genetics 12:101-109.

Mallet J., M. Beltran, W. Neukirchen, and M. Linares. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. Bmc Evolutionary Biology 7.

Nordborg M., T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian, H. Zheng, E. Bakker, P. Calabrese, J. Gladstone, R. Goyal, M. Jakobsson, S. Kim, Y. Morozov, B. Padhukasahasram, V. Plagnol, N. Rosenberg, C. Shah, J. Wall, J. Wang, K. Zhao, T. Kalbfleisch, V. Schulz, M. Kreitman, and J. Bergelson. 2005. The pattern of polymorphism in Arabidopsis thaliana. Plos Biology 3:1289-1299.

Nosil P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375-402.

Peters J. L., T. E. Roberts, K. Winker, and K. G. McCracken. 2012. Heterogeneity in Genetic Diversity among Non-Coding Loci Fails to Fit Neutral Coalescent Models of Population History. Plos One 7:e31972.

Postma E., and A. J. van Noordwijk. 2005. Gene flow maintains a large genetic difference in clutch size at a small spatial scale. Nature 433:65-68.

Presgraves D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution 56:1168-1183.

Price T. D., and M. M. Bouvier. 2002. The evolution of F-1 postzygotic incompatibilities in birds. 56:2083-2089.

Schluter D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372-380.

Storchova R., J. Reif, and M. W. Nachman. 2010. Female Heterogamety and Speciation: Reduced Introgression of the Z Chromosome between Two Species of Nightingales. Evolution 64:456-471.

Storz J. F., and J. K. Kelly. 2008. Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: Insights from deer mouse globin genes. Genetics 180:367-379.

Tajima F. 1989. The Effect of Change in Population-Size on Dna Polymorphism. Genetics 123:597-601.

CHAPTER I. MULTILOCUS PHYLOGEOGRAPHY OF AUSTRALIAN TEALS (*ANAS* SPP.): A CASE STUDY OF THE RELATIONSHIP BETWEEN VAGILITY AND GENETIC STRUCTURE

Abstract– Biogeographic barriers potentially restrict gene flow but variation in dispersal or vagility can influence the effectiveness of these barriers among different species and produce characteristic patterns of population genetic structure. The objective of this study was to investigate interspecific and intraspecific genetic structure in two closely related species that differ in several life-history characteristics. The grey teal (*Anas gracilis*) is geographically widespread throughout Australia with a distribution that crosses several recognized biogeographic barriers. The species has high vagility as its extensive movements track broad-scale patterns in rainfall. In contrast, the closely related chestnut teal (*A. castanea*) is endemic to southeastern and southwestern regions of Australia and is more sedentary. I hypothesized that these differences in life-history characteristics would result in more pronounced population structure in the chestnut teal. I sequenced five nuclear loci (nuDNA) for 49 grey teal and 23 chestnut teal and compared results to published mitochondrial DNA (mtDNA) sequences. I used analysis of molecular variance to examine population structure, and applied coalescent-based approaches to estimate demographic parameters. As predicted, chestnut teal were more strongly structured at both mtDNA and nuDNA (Φ_{ST} = 0.163

and 0.054, respectively) than were grey teal ($\Phi_{ST} < 0.0001$ for both sets of loci). Surprisingly, a greater proportion of the total genetic variation was partitioned among populations within species ($\Phi_{SC} = 0.014$ and 0.047 for nuDNA and mtDNA, respectively) than between the two species ($\Phi_{CT} < 0.0001$ for both loci). Coalescent analyses suggested a late Pleistocene divergence between the taxa, but a remarkable deeper divergence between the southeastern and southwestern populations of chestnut teal. I conclude that dispersal potential played a prominent role in the structuring of populations within these species and that divergent selection associated with ecology and life history traits likely contributed to rapid and recent speciation in this pair.

INTRODUCTION

Natural barriers that divide a population and potentially restrict movements among distinct regions (biogeographic barriers) can strongly influence population genetic structure in many taxa, but the effectiveness of these barriers varies between species and taxonomic groups. High vagility in some species can result in gene flow between otherwise disjunct populations or regions. Furthermore, as climates change, present and past rates of exchange across biogeographic barriers likely vary among species depending upon species-specific life-history characteristics. Here, I explore further the role of life-history traits and dispersal in determining genetic structure across biogeographic divides by examining multilocus population structure in two closely related species of Australian ducks (*Anas* spp.) that differ in important life-history characteristics.

Australia is predominantly arid (Hutchinson 2005, Morton *et al*. 2011) with 22 putative biogeographic barriers (Schodde and Mason1999), which have been hypothesized to have genetically structured populations by restricting gene flow (Degnan and Moritz 1992, Driskell *et al*. 2002, Joseph and Wilke 2006; Dolman and Joseph 2012). Phylogeographic studies have documented the role of these barriers in structuring the continent's birds (reviewed in Joseph and Omland 2009, also see Donnellan *et al*. 2009, Dolman and Joseph 2012). However, mobility is a key survival strategy for many species that occupy the arid interior of the continent, and many species move freely across landscapes that may be inaccessible or inhospitable to others. Movements occur in response to weather events (Reside *et al*. 2010), as observed in waterbirds (Roshier *et al*. 2008, Kingsford *et al*. 2010), or as a search

strategy to locate scarce resources, as has been observed in parrots (Blyth and Burbidge 1997, Forshaw 2002), honeyeaters (Keast 1959, Symonds *et al*. 2006) and others (review in Schodde 1982, Symonds and Johnson 2006, Woinarski 2006). Thus, dispersal and its controls, whether physiological, genetic or ecological, are important countervailing forces to population divergence, and I can expect more prominent population structuring in species with low or infrequent movements relative to more vagile species.

In this study, I present a phylogeographic analysis of the two Australian teal species, which are both distributed on either side of the Nullarbor Plain and its fringing xeric woodlands in southern Australia. This composite barrier, coupled with the Great Australian Bight, divides the distributions of many species into disjunct eastern and western populations (Crisp and Cook 2007, Byrne *et al*. 2008). For example, most sedentary bird species of southern Australian mesic habitats such as the musk duck *(Bizura lobata)*, scarlet robin (*Petroica boodang*), and various *Melithreptus* honeyeaters are genetically differentiated across eastern and western Australia (Guay *et al*. 2010, Toon *et al*. 2010, Dolman and Joseph 2012). However, other birds of more xeric habitats (e.g., the singing honeyeater, *Lichenostomus virescens*, spiny-cheeked honeyeater, *Acanthagenys rufogularis*, and black-faced woodswallow, *Artamus cinereus*) have widespread inland distributions that cross several potential barriers and lack pronounced phylogeographic structure between eastern and western populations (Joseph and Wilke 2007). The wider distributions of these birds across xeric habitats likely reflect fewer ecological or physiological

constraints on their distribution, and their vagility promotes gene flow and inhibits genetic divergence.

The grey teal *(Anas gracilis)* is among the most widely distributed ducks in Australasia, whereas the chestnut teal *(Anas castanea)* is endemic to southeastern and southwestern Australia. The breeding distributions of both species are divided into disjunct eastern and western distributions, but grey teal range much more widely (Fig. 1.1). Grey teal are highly vagile and can move thousands of kilometers, including between southeastern and southwestern Australia (Frith 1962, 1963). This species responds to environmental cues associated with rainfall and flooding (Frith 1959, 1962, 1963, Kingsford and Norman 2002, Roshier *et al*. 2008), and satellite tracking has shown that they can move hundreds of kilometers within a few hours and as much as 4000 kilometers in a year (Roshier *et al*. 2006, 2008). In contrast, chestnut teal are more sedentary and are at their highest densities in estuarine habitats (Frith 1967, Norman and Brown 1988). In addition to differences in life-history traits, chestnut teal and grey teal are morphologically differentiated. Male chestnut teal have bright, colorful plumage, whereas both sexes of grey teal have drab-grey plumage similar to, but slightly lighter than, the female chestnut teal (Frith 1967, Johnsgard 1978, Marchant and Higgins 1990). Despite these differences and significant population structure in chestnut teal, the two species are genetically indistinguishable at mitochondrial DNA (mtDNA) having nearly zero net divergence (Sraml *et al*. 1996; Joseph *et al*. 2009). This lack of mitochondrial divergence suggests that grey and chestnut teal either diverged so recently that there has been insufficient time for lineage sorting to result in species-specific lineages (see Omland

et al. 2006), or that these species hybridize at a sufficient rate to prevent lineage sorting. On the basis of mtDNA, neither of these hypotheses could be rejected although the former was favored as more likely (Joseph *et al*. 2009).

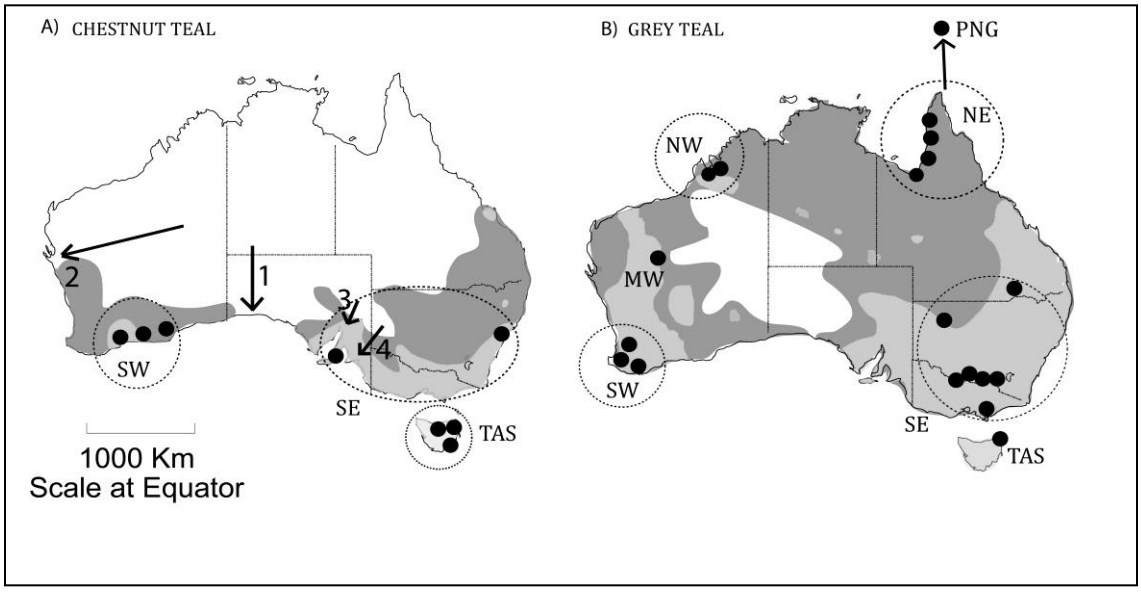

Figure 1.1 Geographic ranges of chestnut teal (A) and grey teal (B). Light shading indicates the known breeding distribution; dark shading indicates the additional non-breeding range. Filled circles mark the sampling locations of each species and the arrows indicate the positions of arid barriers in southern Australia that formed during the Pleistocene (1-Nullarbor, 2-Murchison, 3-Eyrean, 4-Mallee). Open circles enclose the chestnut and grey teal samples sequenced from the southeast (SE), southwest (SW), northeast (NE) and northwest (NW) regions of Australia and Tasmania (TAS). Four grey teals from mid-western (MW) Australia, Papua New Guinea (PNG), and Tasmania (TAS) were not grouped into regions.

The purpose of this study was to apply a multilocus approach to test for genetic structure between eastern and western populations of chestnut and grey teal. I hypothesized that their life-history differences restrict gene flow in chestnut teal relative to grey teal and result in more prominent population structuring in the

former. I evaluated genetic differentiation at six independent markers and applied a coalescent-based approach to infer levels of gene flow within and between these Australian teal species.

MATERIALS AND METHODS

DNA sequencing

The chestnut teal and grey teal were sampled from widespread locations throughout their respective distributions (Fig.1.1). For nuDNA sequencing, the sampling includes all 22 chestnut teal and 33 of the 50 grey teal samples used in Joseph *et al*. (2009) and 16 new grey teal and one chestnut teal. Sampling details are given in Joseph *et al*. (2009) for most samples; new samples included cryofrozen tissue samples (N = 1 chestnut teal) and blood stored in 70% or higher ethanol (N = 16 grey teal). DNA was extracted using a DNeasy Tissue & Blood Kit (Qiagen, Valencia, CA). I also included mtDNA control region sequences obtained from Joseph et al. (2009) and from five additional museum skins of chestnut teal that were collected in southeastern Australia in the 1960s and 1970s for a total of 65 grey teal and 28 chestnut teal.

For each individual, I sequenced five non-coding regions of nuclear DNA: intron 7 from ornithine decarboxylase (ODC1; 324 bp; OD7.F: GCTGTGTGTTTGATATGGGAGT, OD8.R: TGAAGCCAAGTTCAGCCTAA; Peters *et al*., 2008), intron 8 from α-enolase (ENO1; 280 bp; ENO.F: CGCGATGGAAAGTATGACCT, ENO.R: CCAACGCTGCCAGTAAACTT; Peters

et al., 2008), intron 9 from phosphenolpyruvate carboxykinase (PCK1; 324 bp; PCK1-9.F: CAGCCATGAGATCTGAAGCA, PCK1-9.R: TTGAGAGCTGGCTTTCATTG; McCracken *et al*., 2009), intron 7 from fibrinogen beta chain (FGB; 401 bp; FGBF: GTTAGCATTATGAACTGCAAGTAATTG, FGBR: TTTCTTGAATCTGTAGTTAACCTGATG, Peters *et al*., 2012), and intron 11 from the N-methyl-D-aspartate-1-glutamate receptor (GRIN1; 313 bp; GRIN1-11.F: CTGGTGGGGCTGTCTGTG, GRIN1-11.R: ACTTTGAASCGKCCAAATG; McCracken *et al*., 2009). Each intron is linked to a different chromosome in the chicken (*Gallus gallus*) genome (Peters *et al*. 2012), and therefore assumed to be independent. Each locus was amplified using PCR following McCracken *et al*. (2009) and were cleaned using AMPure XP beads following the Agencourt protocol (Beckman Coulter Co.). I sequenced PCR products using the BigDye v. 3.1 Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Automated sequencing was performed on an ABI 3730 at the DNA Sequencing Facility on Science Hill, Yale University, CT.

I aligned and edited the sequences using Sequencher 4.1 (Gene Codes, Ann Arbor, MI) and used the software PHASE 2.1 (Stephens *et al*. 2001, Stephens and Donnelly 2003) to determine the gametic phases of sequences that were heterozygous at more than one nucleotide position. PHASE input files were generated in the program SEQPHASE (Flot 2010).

Population Genetic Analyses

I used a hierarchical analysis of molecular variance (AMOVA) to quantify the partitioning of genetic variation between the species and among different populations using Arlequin vers. 3.5.1.2 (Excoffier and Lischer 2010). To designate populations, I defined five geographic regions: northeast, southeast, southwest, northwest, and Tasmania, and all the individuals of a species from one defined region were pooled to form a single group (Fig. 1.1). Four grey teal individuals did not fit into these groupings (see Fig. 1.1). I calculated Φ_{ST} to test for interspecific and intraspecific genetic differentiation using Arlequin; significance ($P \leq 0.05$) was determined using 10,000 permutations. For this analysis, I used genotypic data from all loci in a single analysis. However, to test the hypothesis that populations of chestnut teal are more differentiated than grey teal, I repeated the AMOVA for each species in separate analyses to obtain an average Φ_{ST} among populations for each locus. I then compared these values between species using a paired t-test that treated each locus as an independent estimate of population structuring.

For each locus, I computed allelic richness (standardized to the smallest sample size of 46) using the RAREFACTION CALCULATOR (University of Alberta, Canada; http://www2.biology.ualberta.ca/jbrzusto/rarefact.php) and nucleotide diversity (the average number of pairwise differences within a species) using Arlequin (Excoffier and Lischer 2010). Each measure of genetic variation was compared between chestnut and grey teal using a paired t-test. I also used Arlequin to calculate locus-specific values for Tajima's *D* (Tajima 1989a, b), a measure of the

relative numbers of high- and low-frequency polymorphisms. A negative *D* indicates an excess of low-frequency polymorphisms, suggesting a population expansion or a selective sweep. In contrast, a positive *D* indicates a paucity of low-frequency polymorphisms, suggesting a bottleneck or the influence of balancing selection.

To test for the independence of nuclear loci, I tested for linkage disequilibrium (LD) among all pairs of loci using a likelihood ratio test (Slatkin and Excoffier 1996) in Arlequin. Significance was tested with 10,000 permutations and 10 starting conditions. Finally, I constructed haplotype networks using the median-joining algorithm in NETWORK ver. 4.1 (Bandelt *et al*. 1999).

Coalescent Analysis

I used the coalescent program IMa2 (Hey 2010) to infer aspects of population history for the two Australian teal. This program uses Bayesian Markov Chain Monte Carlo (MCMC) simulations to estimate six population parameters, scaled to the mutation rate μ, by fitting the data to an isolation-migration model. Demographic parameters include effective population sizes of the ancestral ($\Theta_A = 4N_{eA}\mu$ where N_{eA} is the ancestral effective population size) and each daughter population ($\Theta_1 = 4N_{e1}\mu$ and $\Theta_2 = 4N_{e2}\mu$, where N_{e1} and N_{e2} are the effective population sizes of grey teal and chestnut teal, respectively), time since divergence ($t = T\mu$, where T is the time since divergence in years before present), and effective migration rates ($M_1 = m_1/\mu$, where m_1 is the immigration rate into grey teal from chestnut teal, and $M_2 = m_2/\mu$, where m_2 is the migration rate into chestnut teal).

IMa2 assumes no recombination, and therefore phased data were tested for intra-locus recombination using a four-gamete test in DnaSP vers. 4.10.9 (Rozas *et*

al. 2003). For loci that had evidence of recombination (ENO1, ODC1, and GRIN1), I used the program IMgc to choose a fragment of DNA that was consistent with no recombination by removing individuals and/or base pairs (Woerner et al. 2007). I iteratively adjusted the chromosome weight so that a maximum of 5% of sequences were removed from each data set (i.e., I preferentially truncated sequence length over removing copies).

I ran three models in IMa2. Given the results of the AMOVA and pairwise Φ_{ST} values, I treated grey teal (GT) as a single panmictic population and subdivided chestnut teal into southeastern (CTse; including Tasmanian samples) and southwestern populations (CTsw; see Results). I then conducted each of the three pairwise comparisons in separate analyses (i.e., GT vs. CTse; GT vs. CTsw; CTse vs. CTsw). (Note that I attempted a model that included all three populations, but IMa2 was unstable with these data.) I used an HKY (Hasegawa-Kishino-Yano) model of mutation for FGB, ODC1, and PCK1, and an infinite sites model for GRIN1 and ENO1. I first ran the program to set appropriate upper priors for each parameter. Metropolis coupling was invoked and mixing of Markov chains was assessed from autocorrelations and effective sample sizes (Effective Sample Size >50). I sampled parameters and genealogies every 50 steps and ran the program for more than 5,000,000 steps. To improve mixing, I ran thirty-nine heated chains and a cold chain using a geometric heating scheme. To test for consistent results across multiple runs, I ran IMa2 three times under identical conditions but with different random number seeds; all runs converged on the same stationary distributions.

To calculate biologically informative values from the population parameters estimated in IMa2, I calculated a generation time, *G*, using the equation $G = \alpha + (s/(1 - s)$, where α is the age of maturity and *s* is the expected adult survival rate (Saether *et al*. 2005). I defined α as one year, and *s* as 0.552 based on band-recovery data for grey teal (Halse 1993). These values suggested an average generation time of about two years. I also used the geometric mean of substitution rates among five loci (1.2 $x10^{-9}$ substitutions/site/year) from Peters *et al*. (2008). This calibration is based on the duck-goose divergence and fossil evidence suggesting that the major lineages of anatids likely diverged by the end of the Oligocene (Peters *et al*. 2007; see also Worthy and Lee 2008). Because IMa2 scales all parameters to the mutation rate per locus, I multiplied the per site rate by the average length of fragments included in the analysis (308 bp) for an average rate of 3.7×10^{-7} substitutions/locus/year.

I also tested grey teal, southeast and southwest chestnut teal for evidence of population expansion in LAMARC vers 2.1.8 (Kuhner, 2006) by simultaneously estimating Θ and exponential growth rates, *g* (where, $N_e(t) = N_e(0)exp^{-gt}$, where *Ne(0)* is the current effective population size, and $N_e(t)$ is the effective population size at time *t*). I also jointly estimated recombination rates. I used the Felsenstein 84 model of substitution (ti:tv= 3.0; the average ratio among loci) and ran the program for a burn-in of 2,000,000 generations and sampled a total of 20,000 samples every 1,000 generations. LAMARC scales the estimated parameters to the substitution rate per site (μ).

RESULTS

Genetic Variation

Among the five nuclear loci, chestnut teal and grey teal had comparable nucleotide diversity (Table1.1) ranging between 0.0016 and 0.0164 within grey teal (mean = 0.0077) and 0.0006 and 0.0176 within chestnut teal (mean = 0.0076). Nucleotide diversity did not differ significantly between the two species (paired t-test, t = -0.07, df = 4, P = 0.94). The number of alleles observed was not directly comparable between species, because the sample size of grey teal (N > 90 alleles) was nearly twice the sample size of chestnut teal (N = 46 alleles). Standardizing allelic richness for grey teal to an N of 46, allelic richness did not differ significantly between the two species (t-test, t = 1.03, df = 4, P = 0.35).

Values of Tajima's D for each species did not differ significantly from zero for any locus, except PCK1 in grey teal (P = 0.020; Table 1.1), suggesting that overall, these loci are consistent with neutral evolution. None of the pairs of loci showed any evidence of significant LD within grey or chestnut teal populations (P > 0.12), except that ODC1 and PCK1 showed weak linkage in grey teals (P = 0.04). With twenty pairwise comparisons, I would expect one pair to deviate significantly by chance alone, and therefore, I treat each locus as being independent.

Table 1.1 Measures of genetic diversity and Tajima's *D* within chestnut teal (CT) and grey teal (GT) at five nuclear introns and the mtDNA control region. For grey teal, allelic richness (± StDev) standardized to a sample size of 46 sequences (the sample size for chestnut teal) is given in parentheses.

		Nucleotide diversity		Number of alleles		Tajima's *D*	
Locus	Fragment Length (bp)	CT	GT	CT	GT	CT	GT
FGB	396	0.0042	0.0045	11	14 (10.7± 1.3)	-0.85	-1.12
ENO1	280	0.0040	0.0052	7	11 (9.0±1.0)	-0.78	-0.94
GRIN1	313	0.0110	0.0099	21	23 (16.9±1.7)	0.38	0.60
PCK1	324	0.0006	0.0016	3	7 (6.8± 1.1)	-1.00	-1.63*
ODC	324	0.0164	0.0176	10	20 (15.2±1.5)	-0.06	0.84
mtDNA	609	0.0132	0.0141	20	47 (41.3±1.4)	-0.59	-0.98

*P = 0.020

The allelic networks for each of the five nuclear loci revealed many alleles that were shared between the two species relative to species-specific alleles (Fig. 1.2). In general, only the common alleles in the grey and southeastern chestnut teal were shared with the southwestern chestnut population. Southwestern chestnut teal generally lacked derived alleles.

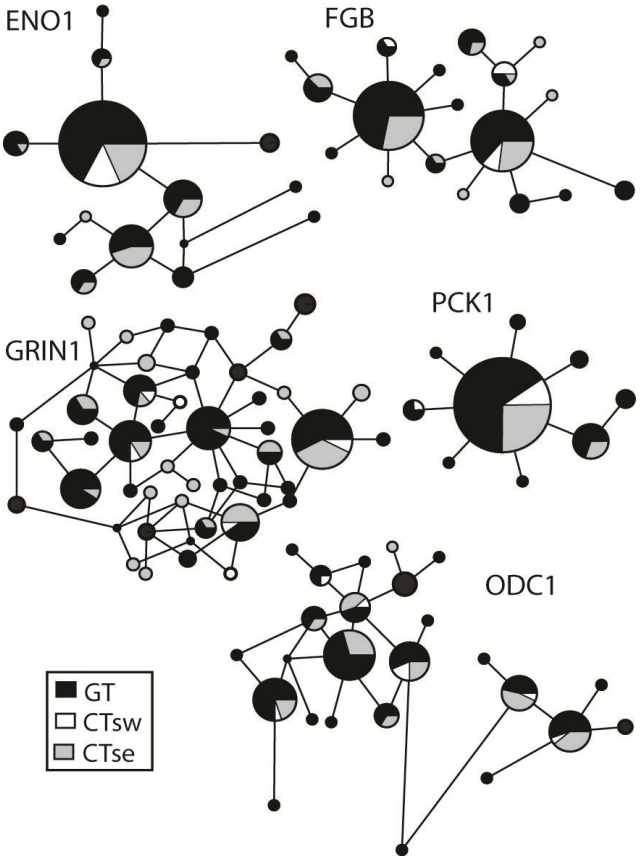

Figure 1.2 Unrooted parsimony networks for five nuclear loci. Given the results of population structure, all four grey teal populations are lumped together (GT, black; N = 42–49 individuals), southeastern and Tasmanian chestnut teals are grouped (CTse, grey; N = 17 individuals), and southwestern chestnut teal are treated separately (CTsw, white, N = 6 individuals). Each circle represents a different allele and the area of the circle is proportional to the sample size for that allele; the length of branches between alleles is proportional to the number of mutations.

The hierarchial AMOVA suggested that the two species were not genetically differentiated at either nuDNA ($\Phi_{CT} < 0.0001$) or mtDNA ($\Phi_{CT} < 0.0001$; Table 1.2). Differences among populations within groups (Φsc), however, explained a significant proportion of the total variation at both mtDNA (Φsc = 4.7%; P = 0.046) and nuDNA (Φsc = 1.4%; P = 0.014). Using an AMOVA for each species separately revealed significant differentiation among chestnut teal populations at both marker types (mtDNA, Φ_{ST} = 0.164, P = 0.001; nuDNA, Φ_{ST} = 0.054, P = 0.003) but not among grey teal populations ($\Phi_{ST} < 0.0001$ for both mtDNA and nuDNA, $P > 0.6$).

Pairwise Φ_{ST} values indicated that all three chestnut teal populations were significantly differentiated at mtDNA (Table 1.3). Likewise, southwestern chestnut teal differed significantly from both southeastern and Tasmanian chestnut teal at nuDNA (Table 1.3), but southeastern and Tasmanian chestnuts did not differ significantly. In contrast, none of the grey teal populations differed significantly at either marker type (Table 1.3). Treating each locus as an independent estimate of population structure, chestnut teal populations were significantly more structured than grey teal populations (paired t-test, $t = 2.65$, $df = 5$, $P = 0.045$; Table 1.2). On the basis of these prominent aspects of substructure, hereafter I treated grey teal as a single population, and I subdivided the chestnut teal into two populations: southeastern (including Tasmania) and southwestern.

Table 1.2 Population differentiation between chestnut teal (CT) and grey teal (GT) and among populations within each species for five nuclear introns and the mtDNA control region.

Locus	Φ_{CT} [1] (CT vs. GT)	Φ_{ST} [1] (within CT)	Φ_{ST} [1] (within GT)
FGB	<0.0001	0.098	<0.0001
ENO1	<0.0001	0.171	<0.0001
GRIN1	0.057	0.012	<0.0001
ODC1	<0.0001	0.016	0.045
PCK1	<0.0001	0.121	0.031
mtDNA	<0.0001	0.164	0.000

[1]Overall *Φ-statistics* for genotypes were $\Phi_{CT} = 0.004$, $\Phi_{ST\,(CT)} = 0.054$, $\Phi_{ST\,(GT)} = -0.003$.

Table 1.3 Pairwise Φ_{ST} for mtDNA (above the diagonal) and nuDNA (below the diagonal) among seven populations of Australian teal; Asterisks indicate significant differentiation between the populations. Population symbols are assigned according to their sampling geographic location; grey teal: northeast, GTne; southeast, GTse; northwest, GTnw; and southwest, GTsw; and chestnut teal: southeast, CTse; southwest, CTsw; Tasmania, CTtas.

	CTtas	CTse	CTsw	GTse	GTne	GTnw	GTsw
CTtas	------	0.1185*	0.2071*	0.0376	0.0280	0.0749	0.0242
CTse	-0.002	-----	0.1867*	0.0576*	0.0354	0.0208	0.0761
CTsw	0.0741*	0.0981*	-----	0.0998*	0.0254	0.0625	0.2678*
GTse	0.0062	0.0132	0.0349*	-----	-0.0364	0.0052	-0.0007
GTne	0.0043	0.0109	0.0630*	-0.0026	------	-0.058	0.0033
GTnw	-0.0039	-0.0110	0.0692*	-0.0034	-0.0145	------	0.0605
GTsw	0.0010	0.0147	0.1197*	0.0037	0.0142	-0.0205	-----

Coalescent Analyses of nuDNA

Results from IMa2 suggested that the effective population size of grey teal was larger than either of the chestnut teal populations (Table 1.4; Fig. 1.3). In the analyses with southeastern and southwestern chestnut teal, N_e of grey teal was estimated to be 1,400,000 individuals (95% highest posterior density, HPD, = 630,000–6,000,000 individuals) and 1,500,000 individuals (95% HPD= 680,000–5,400,000 individuals), respectively. N_e for chestnut teal in the southeast and southwest was approximately 760,000 individuals (95% HPD= 420,000–6,500,000 individuals) and 92,000 individuals (95% HPD= 24,000–536,000 individuals), respectively. At the time of divergence, the ancestral N_e was approximately 680,000 and 620,000 individuals in the comparisons with southeastern and southwestern chestnuts, respectively (GT-CTse; 95% HPD = 360,000–1,100,000 individuals; GT-CTsw 95% HPD = 150,000–1,200,000 individuals). The most likely estimates of divergence times between the two species were similar in both analyses (Table 1.4; Fig. 1.4): 140,000 years ago (95% HPD= 16,000 to 880,000 years ago) in the southeastern comparison and 120,000 years ago (95% HPD= 49,000–2,000,000 years ago) in the southwestern comparison. Migration rates did not differ from zero (i.e., no gene flow) in either direction between grey and southeastern chestnut populations, but I also could not reject a hypothesis of high levels of gene flow (Table 1.4; Fig. 1.3). In contrast, there was evidence of non-zero gene flow into southwestern chestnuts from grey teal, but not in the reverse direction.

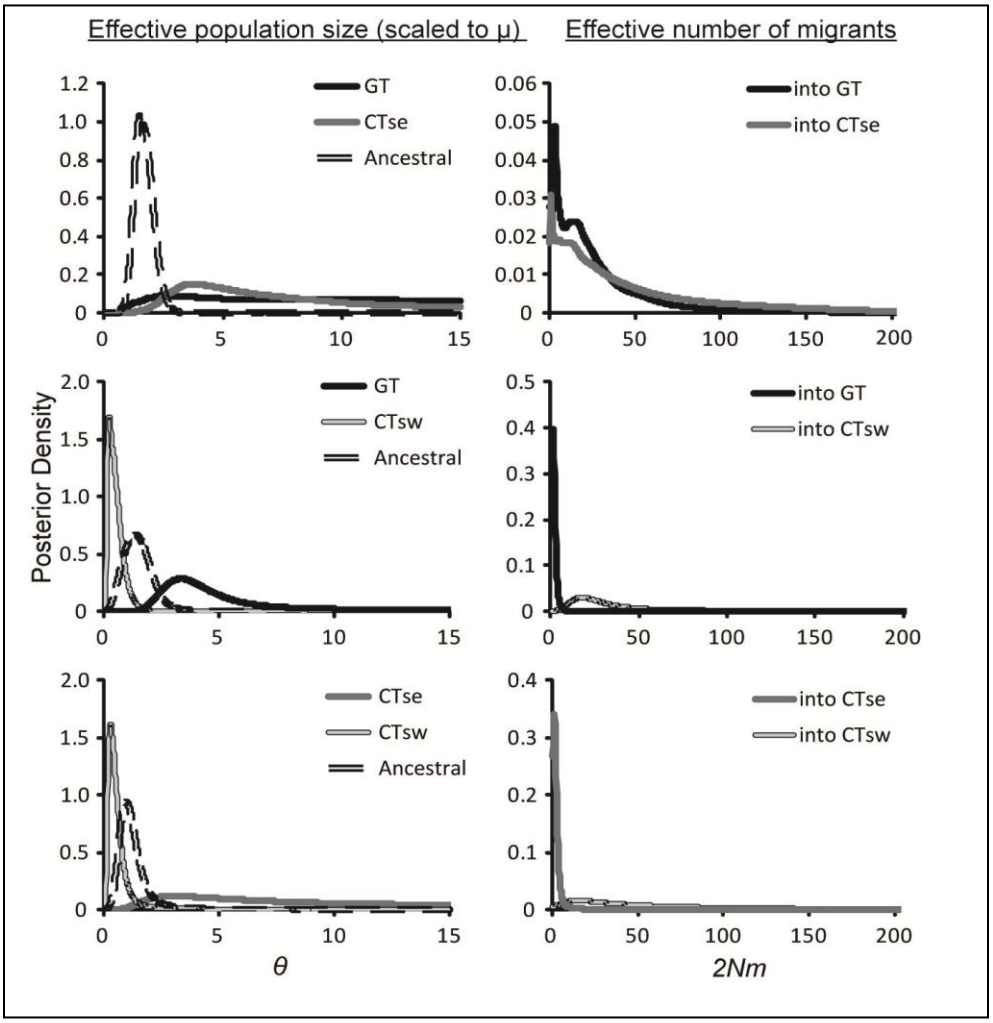

Figure 1.3 Posterior distributions of effective population sizes scaled to the mutation rate ($\theta = 4N_e\mu$; individuals *substitutions/locus/year) and migration rate ($2Nm$, effective number of migrants/generation) estimated in IMa for each of the three pairwise comparisons. Population symbols are as follows: grey teal GT, southeast chestnut teal, and southwest chestnut, CTsw.

The divergence between the southeastern and southwestern chestnut teal populations was deeper than the interspecific divergence in the above comparisons

(Table 1.4; Fig.1.4). Our estimates suggest that these two populations diverged around 260,000 years ago (95% HPD = 97,000–2,000,000 years ago). The estimated N_e for chestnut teal (Table 1.3; Fig. 1.3) in the southwest was approximately 120,000 individuals (95%HPD = 13,600 to 1,960,000 individuals) compared to a much larger N_e of 1,400,000 individuals for the southeast (95%HPD = 550,000 to 6,400,000 individuals). The southeastern chestnut teal was stable or underwent a population expansion since divergence from the southwestern population, whereas the southwestern population has contracted: the most likely estimate of the ancestral N_e for the chestnut teal populations was 430,000 individuals (95% HPD=110,000–1,200,000 individuals). Estimates of migration rates were consistent with no gene flow into southeastern chestnuts, but non-zero gene flow into southwestern chestnuts (Table 1.4; Fig. 1.3).

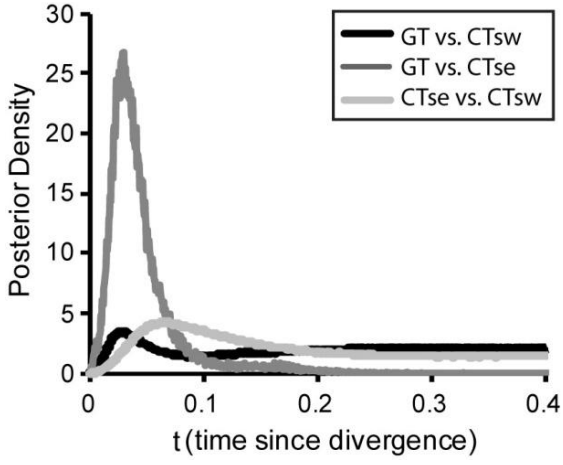

Figure 1.4 Posterior distributions of time since divergence (substitutions/locus) estimated between each pairwise comparison of Australian teals. Population codes are the same as those in Figure 1.3.

Table 1.4 Parameter estimates (peak estimate and 95% highest posterior densities) from IMa. All estimates are scaled to the mutation rate (μ): θ_A is the ancestral population size; θ_1 and θ_2 represent the population sizes of the first and the second population (as given in the heading for each analysis); *t* indicates the divergence time between the populations; $2N_i m_i$ is the effective number of immigrants into population *i*.

Population	θ_A	θ_1	θ_2	t	$2N_1m_1$	$2N_2m_2$
GTand CTse	1.49 (0.78-2.43)	3.18 (1.38-13.24)	1.67 (0.93-14.2)	0.034 (0.004-0.219)	22 (0-118)	2.8 (0-152)
GT and CTsw	1.35 (0.337-2.64)	3.37 (1.5-11.81)	0.22 (0.052-1.16)	0.029 (0.012-0.499)	18 (1.5-113)	1.2 (0-5)
CTse and CTsw	0.95 (0.232-2.57)	3.00 (1.2-14)	0.25 (0.03-4.3)	0.064 (0.024-0.499)	16 (0-142)	1.0 (0-15)

Analyses of population growth in LAMARC suggested negative growth for grey teal (i.e., population decline; $g = -6.25$, 95% HPD= −28.5–67.3) and positive growth for both chestnut teal populations (i.e., population expansion; CTse, $g = 1.64$, 95% HPD = −30–205; CTsw, $g = 6.82$, 95% HPD = −43–356). However, these values of growth were near zero, and I could not reject stable population sizes for any of the three populations. Similarly, average values of Tajima's *D* were near-zero in all three populations (D_{CTse}= -0.066; D_{CTsw}= -0.019; D_{GT} = -0.537; P > 0.05), which is consistent with stable population sizes (Table 1.1).

DISCUSSION

Intraspecific Genetic Structure

Multi-locus analyses detected differences in population structure between the two species of Australian teal. The chestnut teal was significantly structured between eastern and western Australia, but the grey teal populations were consistent with a hypothesis of panmixia. These differences in genetic structure between the two species suggest dissimilar levels of gene flow between their eastern and western ranges, which is consistent with the predictions based on their respective life-history characteristics (reviewed in Marchant and Higgins 1990).

Occasional long distance movements in the vagile grey teal can facilitate gene flow across xeric habitats. Although the grey teal has a disjunct breeding range, their expanded non-breeding range allows for complex spatio-temporal patterns of mixing of flocks from different populations, and thus movements between regions for breeding. For example, grey teal banded near the north coast of the Northern Territory have been recovered in both the southeastern and southwestern breeding ranges (Frith 1962). Furthermore, from only two release points the movement paths of 23 grey teal fitted with satellite-tags showed displacement across much of eastern Australia within two years (Roshier *et al*. 2006, 2008). That these movements included long flights across deserts during a period of extreme drought highlights the mobility and vagility of this species. In contrast, the mainly, or solely, short-distance movements of the more sedentary chestnut teal make it likely that eastern and

western populations remain segregated. The results presented here suggest that differences in vagility and movement ecology have a prominent effect on inter-population differentiation (see also Miller *et al*. 2011) and highlights the likely importance of non-breeding movements on inter-connectivity among widely dispersed populations.

Along with the other recognised barriers in arid southern Australia (i.e., Murchison, Eyrean and Mallee), the Nullarbor Plain (and its fringing xeric woodlands) was an important barrier to dispersal and gene flow in birds, and other vertebrates, during the Pleistocene (Ford 1987). Several studies support east-west population splits here and at other barriers that date to this time period (reviewed in Byrne *et al*. 2008, also see Donnellan *et al*. 2009, Joseph *et al*. 2009, Guay *et al*. 2010, Toon *et al*. 2010, Murphy *et al*. 2011). Our multilocus data suggest that eastern and western populations of chestnut teal diverged during the late Pleistocene, specifically around 260,000 ybp (95% HPD = 97,000–2,000,000 years ago). This is consistent with several other bird species in Australia with discontinuous east-west distributions having their population divergences dated to this same time period (~100,000-200,000 ybp, *Malurus splendens, Climacteris rufus* and possibly *Phylidonyris novaehollandiae,* and *Gliciphila melanops,* Dolman and Joseph 2012).

Inter-specific Variation

The five locus nuDNA analyses of the two Australian teal species presented here support and significantly expand previous mtDNA results (Sraml *et al*. 1996; Joseph

et al. 2009). Both species shared alleles and had no significant differences in allele frequencies and no fixed differences at any locus, a pattern consistent with rampant polyphyly (Omland *et al*. 2006). These species began diverging around 128,000 ybp (95% HPD= 30,000–1,450,000), which is remarkably similar to the estimate from mtDNA of 103,000 ybp (95% HPD= 70,000–165,000 ybp) (Joseph *et al*. 2009) and coincides with the last interglacial period when the climate was similar to the present. The Pleistocene arid cycles in Australia are hypothesized to have influenced the distributions and genetic structure of the Australian biota generally (Byrne *et al*. 2008, 2011), and Australian birds specifically (Toon *et al*. 2007, Kearns *et al*. 2010). The estimated timing of divergence between these Australian teal species is consistent with this hypothesis.

In addition to recent divergence, hybridization and introgression can cause sequences to be shared across taxa (Funk and Omland 2003). Hybridization can continue for millions of years after divergence (e.g., Kronforst 2008) and can sometimes reverse the speciation process (Grant *et al*. 2004; Webb *et al*., 2011). Our coalescent analysis supported the hypothesis of no gene flow between grey teal and southeastern chestnut teal, but I also could not reject the possibility of high levels of gene flow between the species (see similar inference from mtDNA, Joseph *et al*. 2009). However, there is some evidence of non-zero gene flow from the grey teal into the southwest chestnut teal population (Fig. 1.3; the 95% HPD does not include zero). Therefore, recent divergence and introgression both likely contributed to the lack of differentiation between these species, but given the results from both mtDNA

and nuDNA, a rapid divergence in morphology and behavior is strongly supported as the preferred hypothesis.

Selection can maintain the morphological distinction between species in spite of gene flow (Senar *et al.* 2006), and ecologically based divergent selection and sexual selection can intensify this divergence (Schluter and Conte 2009, Rundle and Nosil 2005). Several attributes of the teal pair make ecological speciation a likely hypothesis for this system. These species have broadly overlapping ranges but use different niches; the chestnut teal is more of an estuarine and coastal species, whereas grey teal more often use inland, freshwater habitats. In addition, the color distinction in male chestnut teal suggests a potential role of strong sexual selection in mating patterns. However, testing for gene flow, which is an important parameter in models of ecological speciation (Schluter 1996, Ogden and Thorpe 2002, Rundle and Nosil 2005), is complicated by the indistinguishable genetic make-up between the species given our current markers. A larger sampling of loci or even genomic scans will be necessary to test this hypothesis, as these methods could detect regions under selection that might have contributed to this divergence (Beaumont 2004, Bonin *et al*. 2006, Egan *et al*. 2008). Under the hypothesis of ecological speciation, I predict that some loci (e.g., loci associated with male-plumage characters, female mate-choice, and environmental factors) will be more strongly structured than the putatively neutral loci examined in this study.

Demographic Analyses

The coalescent estimates of effective population size of the teal species did not completely match expectations. The estimated population size of grey teal is about 1.7 million and perhaps as high as >5 million at times (Marchant and Higgins 1990), whereas the population size of southeast and southwest chestnut teal populations have been estimated at 300,000 and 20,000, respectively (Johnsgard1978). IMa2 estimates of effective population size for grey teal (N_e=1,500,000) were close to the census size, but the estimates for chestnut teal populations (CTse, N_e = 1,300,000; CTsw, N_e = 90,000) were larger than census sizes. Multilocus analyses in LAMARC 2.1.8 (Kuhner, 2006) were consistent with stable population sizes for all three populations suggesting that long term changes in population sizes do not explain this disparity. The disparity between genetic estimates of N_e and census size could result from a number of factors, including complex historical demography, inaccurate mutation rate calibrations, balancing selection, or hybridization with a third species, any of which can violate assumptions of the models used to analyze the data and bias results (Becquet and Przeworski 2009, Strasburg and Rieseberg 2010, Peters *et al.* 2012). It is not clear from our data how model violations contributed to our results, but in another study of *Anas* ducks that examined these same loci, heterogeneity in genetic diversity was considerably higher than expected given the inferred neutral, population model (Peters *et al.* 2012). Indeed, different loci gave very different estimates of N_e in that study, and given the variance in nucleotide diversity for these teal (Table 1.1), the same is probably true in this study. Understanding how this heterogeneity contributed to our coalescent models will help us better understand the

history of population demography, divergence, and gene flow between grey and chestnut teals.

Conclusion

In conclusion, intraspecific nuDNA analyses among populations revealed that the highly vagile grey teal is genetically homogeneous across its range in Australia, but the sedentary chestnut teal is significantly structured between eastern and western Australia at both mtDNA and nuDNA markers. Thus, population structure reflects the predicted effects of differences in life history characteristics, specifically dispersal capacity. However, the two species are not genetically differentiated at any of the five nuclear loci, which is consistent with the lack of differentiation observed in mtDNA (Joseph *et al*. 2009). Among all loci, these teal species share many alleles and have similar levels of genetic diversity despite clear morphological and behavioral differences. Genetically, these two species, which are for the most part easily diagnosable phenotypically, are indistinguishable, at least for the markers and sample sizes examined in this study. I suggest a role of ecological speciation and/or sexual selection in driving the divergence between these species, and I predict that a thorough sampling of the genome will uncover loci important to maintaining the taxonomic distinctiveness of these species. Regardless, this multilocus study of grey and chestnut teals reveals a compelling case of rapid morphological divergence between two very closely related species.

REFERENCES

Bandelt, H. J., Forster, P. and Rohl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. – Mol. Biol. Evol. 16: 37–48.

Beaumont, M. A. and Balding, D. J. 2004. Identifying adaptive genetic divergence among populations from genome scans. – Mol. Ecol. 13: 969–980.

Becquet, C. and Przeworski, M. 2009. Learning about modes of speciation by computational approaches. – Evolution 63: 2547–2562.

Blyth, J. and Burbidge, A. H. (1997) What do we know about the Princess parrot *Polytelis alexandrae*? – Eclectus 3: 26–29.

Bonin, A., Taberlet, P., Miaud, C. and Pompanon, F. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). – Mol. Biol. Evol. 23: 773–783.

Byrne, M., Yeates, D. K., Joseph, L., Kearney, M., Bowler, J., Williams, M. A. J., Cooper, S., Donnellan, S. C., Keogh, J. S., Leys, R., Melville, J., Murphy, D.J., Porch, N. and Wyrwoll, H. 2008. Birth of a biome: Insights into the assembly and maintenance of the Australian arid zone biota. – Mol. Ecol. 17: 4398–4417.

Byrne, M., Steane, D. A., Joseph, L., Yeates, D. K., Jordan, G. J., Crayn, D., Aplin, K., Cantrill, D. J., Cook, L. G., Crisp, M. D., Scott Keogh, J., Melville, J., Moritz, C., Porch, N., Kale Sniderman, J.M., Sunnucks, P. and Weston, P.H. 2011. Decline of a biome: Evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. – J. Biogeogr. 38: 1635–1656.

Crisp, M. D. and Cook, L. G. 2007. A congruent molecular signature of vicariance across multiple plant lineages. – Mol. Phylogenet. Evol. 43: 1106–1117.

Degnan, S. and Moritz, C. 1992. Phylogeography of mitochondrial DNA in two species of white-eyes in Australia. – Auk 109: 800–811.

Dolman, G. and Joseph, L. 2012. A species assemblage approach to comparative phylogeography of birds in southern Australia. – Ecol. Evol. 2: 354–369.

Donnellan, S. C., Armstrong, J., Pickett, M., Milne, T., Baulderstone, J., Hollfelder, T. and Bertozzi, T. 2009. Systematic and conservation implications of mitochondrial DNA diversity in emu-wrens, *Stipiturus* (Aves: Maluridae). – Emu 109: 143–152.

Driskell, A. C., Pruett-Jones, S., Tarvin, K. A. and Hagevik, S. 2002. Evolutionary relationships among blue- and black-plumaged populations of the white-winged fairy-wren (*Malurus leucopterus*). – Aust. J. Zool. 50: 581–595.

Egan, S. P., Nosil, P. and Funk, D. J. 2008. Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. – Evolution 62: 1162–1181.

Excoffier, L. and Lischer, H. E. L. 2010. *Arlequin* suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. – Mol Ecol Resour. 10: 564–567.

Flot, J-F. 2010. SEQPHASE: A web tool for interconverting phase input/output files and fasta sequence alignments. – Mol. Ecol. Resour. 10: 162–166.

Ford, J. 1987. Hybrid zones in Australian birds. – Emu. 87: 158–178.

Forshaw, J.M. 2002. Australian parrots. – Robina Press, Queensland.

Frith, H.J. 1967. Waterfowl in Australia. – Angus and Robertson, Sydney.

Frith, H. J. 1963. Movements and mortality rates of the black duck and grey teal in south-eastern Australia. – CSIRO Wildl. Res. 8: 119–131.

Frith, H. J. 1962. Movements of the grey teal, *Anas gibberifrons* Muller (Anatidae). – CSIRO Wildl. Res. 7: 50–70.

Frith, H. J. 1959. The ecology of wild ducks in inland New South Wales. – CSIRO Wildl. Res. 4: 108–130.

Funk, D. J. and Omland, K. E. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. – Annu. Rev. Ecol. Evol. Syst. 34: 397–423.

Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F. and Petren, K. 2004. Convergent evolution of darwin's finches caused by introgressive hybridization and selection. – Evolution 58: 1588–1599.

Guay, P. J., Chesser, R. T., Mulder, R. A., Afton, A. D., Paton, D. C. and McCracken, K. G. 2010. East-west genetic differentiation in musk ducks (*Biziura lobata*) of Australia suggests late pleistocene divergence at the Nullarbor Plain. – Conserv. Genet. 11: 2105–2120.

Halse, S., James, I., Fitzgerald, P., Diepeveen, D. and Munro, D. 1993. Survival and hunting mortality of pacific black ducks and gray teal. – J. Wildl. Manage. 57: 42–48.

Hey, J. 2010. Isolation with migration models for more than two populations. – Mol. Biol. Evol. 27: 905–920.

Hutchinson, M. F., McIntyre, S., Hobbs, R. J., Stein, J. L., Garnett, S. and Kinloch, J. 2005. Integrating a global agro-climatic classification with bioregional boundaries in Australia. – Global Ecol. Biogeogr. 14: 197–212.

Johnsgard, P. A. 1978. Ducks, geese and swans of the world. – University of Nebraska Press, Lincoln.

Joseph, L. and Omland, K. E. 2009. Phylogeography: Its development and impact in Australo-Papuan ornithology with special reference to paraphyly in Australian birds. – Emu 109: 1–23.

Joseph, L. and Wilke, T. 2007. Lack of phylogeographic structure in three widespread Australian birds reinforces emerging challenges in Australian historical biogeography. – J. Biogeogr. 34: 612–624.

Joseph, L. and Wilke, T. 2006. Molecular resolution of population history, systematics and historical biogeography of the Australian ringneck parrots *Barnardius*: Are we there yet? – Emu 106: 49–62.

Joseph, L., Adcock, G. J., Linde, C., Omland, K. E., Heinsohn, R., Chesser, R. T. and Roshier, D. 2009. A tangled tale of two teal: Population history of the grey *Anas gracilis* and chestnut teal *A.castanea* of Australia. – J. Avian Biol. 40: 430-439.

Kearns, A. M., Joseph, L. and Cook, L. G. 2010. The impact of pleistocene changes of climate and landscape on australian birds: A test using the pied butcherbird (*Cracticus nigrogularis*). – Emu 110: 285–295.

Keast, J. A. 1959. Australian birds: their zoogeography and adaptations to an arid continent. Biogeography and ecology in Australia (ed. by A. Keast, R. L. Crocker, and C. S. Christian), pp. 89–114. Junk, The Hague.

Kingsford, R. T. and Norman, F. I. 2002. Australian waterbirds - products of the continent's ecology. – Emu 102: 47–69.

Kingsford, R. T., Roshier, D. A. and Porter, J. L. 2010. Australian waterbirds - time and space travellers in dynamic desert landscapes. – Mar. Freshwater Res. 61: 875–84.

Kronforst, M. R. 2008. Gene flow persists millions of years after speciation in *Heliconius* butterflies. – BMC Evol Biol. 8: 98.

Kuhner M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768-770.

Marchant, S. and Higgins, P. J. 1990. Handbook of Australian new zealand and antarctic birds. Oxford Univ. Press, Melbourne.

McCracken, K. G., Bulgarella, M., Johnson, K. P., Kuhner, M. K., Trucco, J., Valqui, T. H., Wilson, R. E., and Peters, J. L. 2009. Gene flow in the face of countervailing selection: Adaptation to high-altitude hypoxia in the beta A hemoglobin subunit of yellow-billed pintails in the Andes. – Mol. Biol. Evol. 26: 815–827.

Miller, E. J., Eldridge, M. D. B., Morris, K. D., Zenger, K. R. and Herbert, C.A. 2011. Genetic consequences of isolation: Island tammar wallaby (*Macropus eugenii*) populations and the conservation of threatened species. – Conserv. Genet. 12: 1619–1631.

Morton, S. R., Stafford Smith, D. M., Dickman, C. R., Dunkerley, D. L., Friedel, M. H., McAllister, R. R. J., Reid, J. R. W., Roshier, D. A., Smith, M. A., Walsh, F. J., Wardle, G. M., Watson, I. W. and Westoby, M. 2011. A fresh framework for the ecology of arid Australia. – J. Arid. Environ.75: 313–329.

Murphy, S. A., Joseph, L., Burbidge, A. H. and Austin, J. 2011. A cryptic and critically endangered species revealed by mitochondrial DNA analyses: The western ground parrot. – Conserv. Genet. 12: 595–600.

Norman, F. I. and Brown, R. S. 1988. Aspects of the distribution and abundance of chestnut teal in southeastern Australia. – Emu 88: 70–80.

Ogden, R. and Thorpe, R. S. 2002. Molecular evidence for ecological speciation in tropical habitats. – Proc. Natl. Acad. Sci. U. S. A. 99: 13612–13615.

Omland, K. E., Baker, J. M. and Peters, J. L. 2006. Genetic signatures of intermediate divergence: Population history of old and new world Holarctic ravens (*Corvus corax*). – Mol. Ecol. 15: 795–808.

Peters, J. L., Zhuravlev, Y., Fefelov, I., Logie, A., and Omland, K. E. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). – Evolution 61: 1992–2006.

Peters, J. L., Zhuravlev, Y. N., Fefelov, I., Humphries, E. M. and Omland, K. E. 2008. Multilocus phylogeography of a Holarctic duck: Colonization of North America from Eurasia by gadwall (*Anas strepera*). – Evolution 62: 1469–1483.

Peters, J. L., Roberts, T. E., Winker, K. and McCracken, K. G. 2012. Heterogeneous genetic diversity among non-coding loci fails to fit neutral coalescent models of population history. – PLoS One.7: e31972

Reside, A. E., VanDerWal, J. J., Kutt, A. S. and Perkins, G. C. 2010. Weather, not climate, defines distributions of vagile bird species. – PLoS One 5: e13569.

Roshier, D. A., Klomp, N. I. and Asmus, M. 2006. Movements of a nomadic waterfowl, grey teal *Anas gracilis*, across inland Australia- results from satellite telemetry spanning fifteen months. – Ardea 94: 461–475.

Roshier, D., Asmus, M. and Klaassen, M. 2008.What drives long-distance movements in the nomadic grey teal *Anas gracilis* in Australia? – Ibis 150: 474–484.

Roshier, D. A., R. Heinsohn, G. J. Adcock, P. Beerli, and L. Joseph. 2012. Biogeographic models of gene flow in two waterfowl of the Australo-Papuan tropics. – Ecology and Evolution 2: 2803–2814.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. – Bioinformatics 19: 2496–2497.

Rundle, H. D. and Nosil, P. 2005. Ecological speciation. – Ecology Letters 8: 336–352.

Saether, B., Lande, R., Engen, S., Weimerskirch, H., Lillegard, M., Altwegg, R., Becker, P. H., Bregnballe, T., Brommer, J. E., McCleery, R. H., Merila, J., Nyholm, E., Rendell, W., Robertson, R.R., Tryjanowski, P. and Visser, E.

2005. Generation time and temporal scaling of bird population dynamics. – Nature 436: 99–102.

Schluter, D. (1996) Ecological speciation in postglacial fishes. Philos. Trans. R . Soc. Lond. B Biol. Sci. 351: 807–814.

Schluter, D. and Conte, G. L. 2009. Genetics and ecological speciation. – Proc. Natl. Acad. Sci. U. S. A. 106: 9955–9962.

Schodde, R. 1982. Origin, adaptation and evolution of birds in arid Australia. *Evolution of the flora and fauna of arid Australia* (ed. by W.R. Barker and P.J.M. Greenslade), pp. 191–224 – Peacock Publications, Frewville.

Schodde R. and Mason, I. 1999. The Directory of Australian Birds. Passerines – CSIRO Publishing.

Senar, J. C., Borras, A., Cabrera, J., Cabrera, T. and Bjorklund, M. 2006. Local differentiation in the presence of gene flow in the citril finch *Serinus citrinella*. – Biol. Lett. 2: 85–87.

Slatkin, M, Excoffier, L. 1996. Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm – Heredity 76:377-383.

Sraml, M., Christidis, L., Easteal, S., Horn, P., and Collet, C. 1996. Molecular relationships within Australasian waterfowl (Anseriformes). – Austr. J. Zool. 44:47-58

Stephens, M. and Donnelly, P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. – Am. J. Hum. Genet. 73: 1162–1169.

Stephens, M., Smith, N. J. and Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. – Am. J. Hum. Genet. 68: 978–989.

Strasburg, J. L. and Rieseberg, L. H. 2010. How robust are "isolation with migration" analyses to violations of the IM model? A simulation study. – Mol. Biol. Evol. 27: 297–310.

Symonds, M. R. E and Johnson, C. N. 2006. Determinants of local abundance in a major radiation of Australian passerines (Aves: Meliphagoidea). – J. Biogeogr. 33: 794–802.

Symonds, M. R. E., Christidis, L., and Johnson, C. N. 2006. Latitudinal gradients in abundance, and the causes of rarity in the tropics: a test using Australian honeyeaters (Aves: Meliphagidae). – Oecologia 149: 406–417.

Tajima, F. 1989a. The effect of change in population-size on DNA polymorphism. – Genetics 123:597-601.

Tajima, F. 1989b. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. – Genetics 123:585-595.

Toon, A., Hughes, J. M. and Joseph, L. 2010. Multilocus analysis of honeyeaters (Aves: Meliphagidae) highlights spatio-temporal heterogeneity in the influence of biogeographic barriers in the Australian monsoonal zone. – Mol. Ecol. 19: 2980–2994.

Toon, A., Mather, P. B., Baker, A. M., Durrant, K. L. and Hughes, J. M. 2007. Pleistocene refugia in an arid landscape: Analysis of a widely distributed Australian passerine. – Mol. Ecol. 16: 2525–2541.

Webb, W.C., Marzluff, J.M., Omland, K.E. 2011. Random interbreeding between cryptic lineages of the common raven: Evidence for speciation in reverse. Mol. Ecol. 20:2390-2402.

Woerner, A. E., Cox, M. P. and Hammer, M. F. 2007. Recombination-filtered genomic datasets by information maximization. – Bioinformatics.23: 1851–1853.

Woinarski, J. C. Z. 2006. Predictors of nomadism in Australian birds: a reanalysis of Allen and Saunders (2002). – Ecosystems 9: 689–693.

Worthy, T. H., and Lee, M. S. Y. 2008. Affinities of Miocene waterfowl (Anatidae: Manuherikia, Dunstanetta and Miotadorna) from the St Bathans Fauna, New Zealand. – Palaeontology 51: 677-708.

CHAPTER II. SIGNATURES OF A LARGE Z-EFFECT IN THE EARLIEST STAGES OF DIVERGENCE IN AUSTRALASIAN TEALS

Abstract. –Two recently diverged species of Australian ducks, the grey teal (*Anas gracilis*) and chestnut teal (*A. castanea*), are well-differentiated in male plumage, ecology, and dispersal behaviors, yet they are genetically indistinguishable at mitochondrial DNA (mtDNA) and nuclear DNA (nuDNA) loci. Given a strong role of the sex-chromosomes in speciation, I hypothesized a pronounced divergence on the Z-chromosome between these two taxa despite the possibility of gene flow. I sequenced a genomic transect of seventeen autosomal and seven Z-linked loci in 49 grey teal and 23 chestnut teal from Australia. Consistent with the proposed hypothesis, the two species are genetically indistinguishable at autosomal loci (mean Φ_{ST} = 0.014), but significantly divergent in six of seven z-linked loci (mean Φ_{ST} = 0.274). Furthermore, these species were particularly divergent on the shorter arm of the chromosome (mean Φ_{ST} = 0.428) compared to the longer arm (mean Φ_{ST} = 0.068). Simulations under the neutral demographic histories over-predicted the differentiation in autosomal DNA but under-predicted the divergence in Z-linked loci, suggesting differences in the evolutionary histories of classes DNA. In particular, divergence on the short arm did not overlap with the expected neutral values. I conclude that the Z-chromosome contains a genomic island of

differentiation that has played a prominent role in the speciation of grey and chestnut teal, and I discuss the possible influence of sexual and/or ecological speciation.

INTRODUCTION

Divergent selection can have an important role in speciation, especially when species pairs diverge in the face of gene flow (Saint-Laurent *et al*., 2003; Emilianov *et al*., 2004; Hendry *et al*., 2004; McCracken *et al*., 2009a,b; Nosil *et al*., 2009; Nosil and Feder, 2012) or in response to local adaptation (Peichel *et al*., 2001; Grant and Grant, 2006). Although gene flow tends to homogenize the genomes of hybridizing species, divergent selection on locally adapted alleles can prevent or inhibit gene flow for some loci, causing some genomic regions to be exceptionally differentiated (outlier loci; Nosil *et al*., 2009). Those outlier regions help maintain species integrity despite the homogenizing force of gene flow in other regions and create a pattern of heterogeneous genetic divergence. However, it is not necessary for divergent selection alone to limit introgression; other factors such as sex-biased dispersal can partially restrict introgression among markers with different modes of inheritance (Scribner *et al*., 2001; Crochet *et al*., 2003; Carling and Brumfield, 2008, Carling *et al*., 2010; Peters *et al*., 2012a). The goal of this study is to test for heterogeneous divergence and gene flow between different classes of markers during the early stages of divergence and speciation.

Genomic heterogeneity is an important component of speciation in allopatry or sympatry when the divergence of neutral and non-neutral loci proceeds at different evolutionary rates. Whereas genetic homogeneity is expected between hybridizing taxa, some loci can be strongly divergent despite gene flow (Turner and Hahn, 2007; Turner and Hahn, 2010). These outlier loci constitute "islands of divergence" and are composed of those loci that are under selection or tightly linked to loci under

selection (Nosil and Feder, 2012). When neutral alleles are freely exchanged between taxa, the level of differentiation at "islands of divergence" should be a function of selective strength. Thus, heterogeneity in genetic divergence may be higher in sympatric and parapatric conditions with gene flow than in strictly allopatric conditions.

Different marker types, such as mitochondrial DNA (mtDNA), autosomal DNA (auDNA) and sex-linked loci (zDNA) can vary in their potential to introgress for different reasons, such as inviability of hybrids of the heterogametic sex and sex-biased dispersal (Carling and Brumfield, 2008). Haldane's rule predicts inviability or sterility of the heterogametic sex (Haldane 1922), which has been demonstrated empirically for several birds (Sattler and Braun, 2000; Kirby *et al*., 2004). This reduction in viability of the heterogametic sex could restrict introgression at markers exclusively inherited by the heterogametic sex. For example, reduced introgression of mtDNA relative to nuclear DNA (nuDNA) is anticipated in female heterogametic taxa. Similarly, sex-biased dispersal favors gene flow at markers inherited by the dispersive sex while it restricts the introgression at markers inherited by the non-dispersive sex. Female philopatry and male-biased dispersal could generate greater genetic structure at mtDNA by restraining its movement between populations or species (Sattler and Braun, 2000; Helbig *et al*., 2001; Saetre *et al.*, 2001; Crochet *et al*., 2003; Saetre *et al.*, 2003 Peters *et al*. 2012a, b). Therefore, levels of introgression at a marker can result from the combination of both selection and sex-biased dispersal. Likewise, heterogeneity in the level of introgression could be expected between auDNA and sex-linked DNA (zDNA in female heterogametic taxa).

Dominance theory posits that sterility or inviability of the heterogametic hybrid (ZW) is controlled by recessive Z-linked alleles that have greater accumulative effects in the hemizygous state (Turelli and Orr, 1995), which inhibits zDNA introgression while not influencing auDNA introgression. In support of dominance theory, several studies have reported reduced introgression in sex-linked loci relative to autosomal loci (Hagen & Scriber, 1989; Tucker *et al*., 1992; Sætre *et al*., 2003; Carling & Brumfield, 2008). Additionally, the restrictions on introgression could also result from the ecological based selection between the divergent populations that eventually evolve to different species.

Ecological speciation refers to the evolution of reproductive incompatibility between diverging populations in response to variable ecological adaptations; specifically, different populations are adapted to different ecological niches (Schluter, 2009, Schluter *et al*., 2010). The divergence could begin with selection on either standing genetic variation (which is generally a faster route to speciation; Barrett *et al*., 2008) or by the occurrence of new mutations. While auDNA contributes >90% of standing variation in the whole genome, the Z-chromosome could be more vital in speciation due to its higher substitution rate and faster evolution (Axelsson *et al*., 2004). Although auDNA could be more important than zDNA in ecological adaptation from standing genetic variation, zDNA could be more important in the alternative route to speciation via mutations. Either way, advantageous mutations could sweep to fixation in populations adapting to new ecological niches.

The Z-chromosome could have an important role in speciation because of its higher mutation rate and rapid fixation of alleles. Higher mutation rate in zDNA arises as male gametes undergo significantly more rounds of replication than female gametes, and Z spends two thirds of its time in the male germline (Miyata *et al*., 1987; Hurst and Ellegren, 1998). Rapid fixation of any new beneficial mutations in zDNA could result from to the role of selection (Borge *et al*., 2005; Storchova *et al*., 2010; Backstrom and Vali, 2011) or lower effective population size and faster sorting rate of Z-linked loci (Mank *et al*., 2007; Vicoso and Charlesworth, 2009). These properties can make zDNA diversify more rapidly than auDNA, potentially resulting in zDNA being more likely to contain "islands of divergence". Several studies have found genes involved in reproductive isolation (pre-zygotic or post-zygotic) mapping to sex-chromosomes (Orr and Coyne, 1989; Presgraves, 2002; Saether *et al*., 2007). Additionally, there is over-representation of Z-linked genes among those genes that have accelerated evolution in birds (i.e larger Z-effect; Ellegren, 2009). Although a faster and larger Z-effect has been studied in the later stages of speciation (Borge *et al*., 2005; Storchova *et al*., 2010; Elgvin *et al*., 2011; Backstrom and Vali, 2011; Hogner *et al*., 2012) the contribution of Z in the early stages of speciation has not been well-studied. The over-reaching goal of this study is to examine the genetics of speciation in a pair of taxa that have undergone a recent divergence with rapid speciation.

Study taxa

The grey teal (*Anas gracilis*) and chestnut teal (*A.castanea*) are recently diverged taxa that have undergone extensive morphological, ecological, and behavioral

divergence. Chestnut teal are strongly sexually dichromatic, with males having bright, colorful plumage and females having dull, grey-brown plumage. In contrast, grey teal are sexually monochromatic; both sexes have grey-brown plumage similar to that of the chestnut teal female, but lighter. Grey teal preferentially use inland, fresh-water habitats throughout the Australian continent and tend toward nomadism, quickly moving long distances to newly formed freshwater lakes after heavy rainfall (Frith, 1959, 1962, 1963; Roshier *et al.*, 2006, 2008). In contrast, the chestnut teal is a coastal species that is most numerous in brackish waters and is restricted to southern Australia with a disjunct distribution between southeastern and southwestern Australia partitioned by the dry deserts of central Australia. The chestnut teal differs from the grey teal in that it is mostly sedentary with some short-distance movements (Frith, 1967; Norman and Brown, 1988). The two species have overlapping breeding ranges (Fig. 1.3) and hybridize in captivity (Frith, 1963), although hybridization has not been documented in the wild. Despite these morphological, behavioral, and ecological differences, the species are genetically indistinguishable in mtDNA (Joseph *et al.*, 2009) and five autosomal loci (Chapter 1). In both markers, none of the genetic variation was explained by differences between the two species and more variation was partitioned among populations of chestnut teal than between species (Joseph *et al.*, 2009; Chapter 1). Extensive introgression between the taxa or recent and rapid divergence could explain the observed pattern of genetic similarities between the species. However, coalescent analysis on five nuclear loci could not resolve the contribution of gene flow to this genetic similarity between the grey and chestnut teal (Joseph *et al.*, 2009; Chapter 1).

These species are genetically more similar to each other than expected given their morphological and ecological divergence. Therefore, this pair of taxa provides an excellent model system for examining heterogeneous differentiation during the earliest stages of divergence accompanied by rapid speciation.

In this study, I examine genetic differentiation at seventeen autosomal and seven sex-linked markers to address the following questions in Australian Teal; (1) Is there evidence of genetic divergence from a genomic transect of auDNA? (2) Do different classes of nuclear markers (auDNA and zDNA) differ in the pattern of introgression (i.e., is there a larger Z-effect during the early stages of divergent)?

MATERIALS AND METHODS

DNA sequencing

I sequenced seventeen autosomal and seven sex-linked loci using the 23 chestnut teal and 49 grey teal samples described in Chapter 1. Each autosomal intron is located on a different chromosome in the chicken (*Gallus gallus*) genome (Peters *et al*., 2012c). For zDNA, I chose seven loci distributed across the entire chromosome (Jacobsen *et al*., 2009; Backstrom *et al*., 2010; Peters *et al*., 2007). Published sequences from other birds were aligned with the mallard (*Anas plathyrhynchos*) genome (Kraus *et al*., 2011), and new primers were developed to specifically target duck zDNA. Each locus was amplified using PCR at an annealing temperature of 58°C using the

primers and protocols given in Peters *et al*. (2012c). Sequences for newly designed primers for zDNA are provided in table 2.1. PCR products were cleaned using AMPure XP beads following the Agencourt protocol (Beckman Coulter Co.). Sequencing of PCR products was done using the BigDye v. 3.1 Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Automated sequencing was performed on an ABI 3730 at the DNA Sequencing Facility on Science Hill, Yale University, CT. We aligned and edited the sequences in Sequencher 4.1 (Gene Codes, Ann Arbor, MI) and determined the gametic phases of sequences using the software PHASE 2.1 (Stephens *et al*.2001, Stephens and Donnelly, 2003). PHASE input files were generated in the program SEQPHASE (Flot, 2010).

Diversity, Divergence and Neutrality

Arlequin vers. 3.5.1.2 (Excoffier and Lischer, 2010) was used to quantify genetic diversity (nucleotide diversity, π, the mean number of pairwise differences among alleles within a population) and interspecies differentiation (Φ_{ST}, the proportion of nucleotide diversity explained by differences among populations) for each locus. I used RAREFACTION CALCULATOR (University of Alberta, Canada) to compute allelic richness standardized to the smallest sample size of 46 in auDNA and 44 in zDNA. I constructed haplotype networks using the median-joining algorithm in NETWORK ver. 4.1 (Bandelt *et al*.1999). I estimated Tajima's *D* (a measure of the difference in the estimates of θ based on number of segregating sites and the average number of pairwise nucleotide differences between random sequences) to test the neutrality of auDNA and zDNA (Tajima, 1989) in Arlequin vers. 3.5.1.2. (Excoffier

and Lischer, 2010). I also tested for linkage disequilibrium in Arlequin. For z-linked loci, males and females were tested independently, because the gametic phases were known for females but not males.

Demography

I applied Bayesian Markov Chain Monte Carlo (MCMC) simulations in the coalescent program IMa2 (Hey, 2010) to estimate demographic parameters, scaled to the mean mutation rate per locus (μ), under an isolation-migration model for the autosomal loci and Z-linked loci in separate analyses. These demographic parameters included i) effective population sizes of the ancestral ($\Theta_A = 4N_{eA}\mu$ where N_{eA} is the ancestral effective population size)and each daughter population ($\Theta_1 = 4N_{e1}\mu$ and $\Theta_2 = 4N_{e2}\mu$, where N_{e1} and N_{e2} are the effective population sizes of grey teal and chestnut teal, respectively), ii) time since divergence ($t = T\mu$, where T is time since divergence in years before present), and iii) migration rates ($M_1 = m_1/\mu$, where m_1 is the immigration rate into grey teal from chestnut teal, and $M_2 = m_2/\mu$, where m_2 is the migration rate into chestnut teal). I defined inheritance scalars as 1.0 for autosomal loci and 0.75 for Z-linked loci to reflect differences in modes of inheritance; by doing so, demographic parameters estimated from the two marker types were on the same scale (that of autosomal DNA) and directly comparable.

We estimated the number of recombination events for each locus using the four-gamete test in DnaSP vers. 4.10.9 (Rozas *et al*., 2003) and removed the recombinant sites using the program IMgc (Woerner *et al*., 2007). We chose the recombination-free blocks by iteratively adjusting the chromosome weight so that a

maximum of 5% of sequences were removed from the data set (i.e., we preferentially truncated sequence length over removing copies).

We first ran IMa2 to set appropriate upper priors for each parameter. We ran a cold chain and 39 heated chains using a geometric heating scheme to invoke metropolis coupling and improve mixing. We assessed the mixing of Markov chains from the autocorrelations and effective sample sizes (ESS >50). We sampled parameters and genealogies every 100 steps and ran the program for more than $3x10^7$ steps after a burn-in of 500,000 steps. On average, we saved 300,000 genealogies in both the autosomal and Z-linked data runs. To test for consistent results across multiple runs, we ran IMa2 three times under identical conditions but with different random number seeds; all runs converged on the same stationary distributions.

Coalescent simulations

I simulated genetic divergence and diversity of the taxa for both auDNA and zDNA under the assumptions of the neutral demographic history estimated from the alternative marker type in IMa2. Specifically, expected zDNA diversity and divergence was simulated under the model inferred from auDNA (parameters were rescaled by a factor of 0.75 to account for differences in models of inheritance—the effective population size of zDNA is 0.75 of that of auDNA because of the hemizygosity of females), and expectations for auDNA were simulated under the zDNA model (because an inheritance scalar of 0.75 was defined for zDNA in IMa2, no adjustments to the parameters were required). Using the program MS (Hudson, 2002), I simulated 1000 datasets/locus for each model following the methods described in Peters *et al*. (2012c). To account for uncertainty in the models, one-

thousand estimates of each demographic parameter were randomly selected from the coalescent probability distributions of each parameter obtained from IMa2 (θ_{GT}, θ_{CT}, θ_A, t_0, $\theta_m M_{GT}$, $\theta_m M_{CT}$, $\theta_m M_A$,). I also included locus-specific recombination rates and mutation rates in the analysis. Recombination rates for each locus (C/μ, where C is recombination rate/generation and μ is the mutation rate/site/generation) was estimated in LAMARC using a bayesian approach (Kuhner, 2006). An evolutionary model F84 was used with a transition to tranversion ratio of 3.0 and a posterior density curve was generated with 20,000 genealogies from a complete set of $2x10^7$ sampled genealogies following a burn in of 2,000,000 generations. A random sample of 1,000 recombination rates for each locus were chosen from the posterior probability distribution generated in LAMARC , which were randomly combined with the 1000 estimates of demographic histories . Similarly, a random sample of 1000 mutation rates for each locus was chosen from the probability curves generated based on a deep Anseriformes phylogeny (Peters *et al*. 2012c). I used relative mutation rates among loci; the mean rate was scaled to 1.0. Among Z-loci, relative mutation rates were only available for CHD1z, which were applied to all Z-loci. In the simulations, all parameters were rescaled to the effective population size of grey teal. Each dataset for auDNA and zDNA contained 144 and 122 sequences, respectively, to mimic sample sizes for empirical data. In total, I simulated 17,000 datasets for auDNA and 7,000 datasets for zDNA (1,000 datasets per locus).

Goodness-of- fit test

The fit of the data to the models was assessed using a goodness-of-fit test for each marker type . I compared locus-specific empirical values of π, Φ_{ST}, and Tajima's D

with the posterior predictive probabilities generated from the simulated sequence data; these summary statistics were calculated using ms.output (Peters *et al.*, 2012c). For each class of DNA, expected mean values for seventeen auDNA and seven zDNA loci were compared to the mean empirical values. I rejected the null hypothesis of neutrality if the empirical values were outside the expected 95% confidence interval simulated under neutrality. I also performed locus-specific goodness-of-fit tests by comparing empirical values with the 1,000 values simulated for each locus.

RESULTS

Genetic Variation

Intraspecifically, sex-linked loci had less genetic variation than autosomal loci (Table 2.1). Chestnut teal had significantly fewer alleles at Z-linked loci than autosomal loci (mean $R_{CT\text{-}A}$ = 12.0 $\pm$ 8.91 StDev, $R_{CT\text{-}Z}$ =4.0$\pm$3.53 StDev; Mann-Whitney U-test, U = 22.5, df = 22, P = 0.020; subscripts CT and GT refer to chestnut teal and grey teal, respectively, and subscripts A and Z refer to autosomal and Z-linked loci, respectively). A similar trend was also observed for the grey teal, although the results were not significant ($R_{GT\text{-}A}$ = 14.0$\pm$9.92 StDev, $R_{GT\text{-}Z}$ = 6.0$\pm$3.17 StDev; Mann-Whitney U-test, U = 32.0, df = 22, P = 0.081). Autosomal nucleotide diversity was significantly larger (almost five times to that found for z-loci) in grey teal ($\pi_{GT\text{-}A}$ = 0.0094, $\pi_{GT\text{-}Z}$ = 0.0016; Mann-Whitney U-test, U = 22.0, df = 22, P =

0.018), and a similar trend was observed in the chestnut teal ($\pi_{CT\text{-}A} = 0.0093$, $\pi_{CT\text{-}Z} = 0.0016$; Mann-Whitney U-test, $U = 33.0$, $df = 22$, $P = 0.098$).

Inter-specifically, allelic richness in both autosomes and the Z-chromosome was significantly higher in grey teal than in chestnut teal after controlling for differences in sample sizes using rarefaction (Table 2.1; auDNA, paired t-test, $t = 2.11$, $df = 16$, $P = 0.01$; zDNA, paired t-test, $t = 2.44$, $df = 6$, $P = 0.01$). Overall, shared polymorphisms appeared relatively smaller on the Z-linked loci relative to the autosomal loci as the network analyses revealed (Fig. 2.6 a & b). However, chestnut teal and grey teal did not differ significantly in nucleotide diversity in either autosomes or the Z-chromosome (auDNA, paired t-test, $t = 2.11$, df = 16, $P = 0.79$; zDNA, paired t-test, $t = 2.44$, df = 6, $P = 0.97$).

Neutrality

Mean Tajima's *D* was negative for both grey teal and chestnut teal at both markers (Table 2.1; $D_{GT\text{-}A} = -0.62$, $D_{CT\text{-}A} = -0.67$, $D_{GT\text{-}Z} = -1.548$, $D_{CT\text{-}Z} = -0.32$). Intraspecifically, Tajima's *D* was significantly more negative in zDNA than auDNA in grey teal (Wilcoxon rank sum test, $z = 2.57$, $df = 22$, $P = 0.010$), but not chestnut teal (Wilcoxon rank sum test, $z = 0.48$, $df = 22$, $P = 0.62$; Table 2.2). Interspecifically, the comparisons revealed that Tajima's *D* was significantly more negative in the grey teal relative to the chestnut teal for zDNA (paired t-test, $t = 1.94$, df = 16, $P = 0.033$) but not auDNA (paired t-test, $t = 1.79$, df = 6, $P = 0.760$). Overall, Tajima's *D* was significantly negative at eight autosomal loci and five Z-loci in grey teal and six autosomal loci and one Z-locus in chestnut teal (Table 2.1).

Genetic Differentiation

The two teals were on average twenty times more differentiated in zDNA than auDNA (Table 2.1), a significant difference (Wilcoxon rank sum test, $z = 2.53$, $df = 22$, $P = 0.011$). In autosomal loci, grey and chestnut teal were significantly differentiated at only two (GHRL & GRIN1) of the 17 loci (mean $\Phi_{ST} = 0.014$). Conversely, the species were significantly differentiated on the Z-chromosome at six of the seven loci (mean $\Phi_{ST} = 0.274$; Table 2.1). The maximum differentiation recorded among all twenty-four loci was the Z-linked ADAMTS6 ($\Phi_{ST} = 0.852$) that was about sixty times greater than the average recorded at autosomal loci. Among sex-linked loci, the two taxa showed insignificant genetic differentiation only at ALDOB ($\Phi_{ST} = 0.0$). Additionally, the divergence was almost six times greater on the short arm of the Z-chromosome, especially near the centromere (Itoh *et al.*, 2011) relative to the long arm ($\Phi_{ST} = 0.423$ and $\Phi_{ST} = 0.068$, respectively; Fig. 2.1).

Table 2.1 Nucleotide diversity (π), interspecific differentiation (Φ_{ST}), and Tajima's *D* for seventeen autosomal and seven Z-linked markers in the grey teal (GT) and chestnut teal (CT). Numbers within the parentheses refer to allelic richness standardized to a sample size of 46 for auDNA and 44 for zDNA.

	$\pi_{(GT)}$	$\pi_{(CT)}$	Φ_{ST}	Tajima's D_{GT}	Tajima's D_{CT}
Autosomal Loci					
CRYAB	0.0001 (1.9)	0.0002 (2)	0.006	-1.38*	-0.85
PCK1	0.0016 (6.8)	0.0006 (3)	<0.001	-1.63*	-1.00
LDHB	0.0006 (3.8)	0.0006 (5)	<0.001	-1.50*	-1.81*
FAST	0.0010 (4.2)	0.0010 (3)	<0.001	-1.82*	-1.55*
MSTN	0.0019 (3.7)	0.0010 (3)	0.020	-0.39	-1.34*
SOAT1	0.0022 (6.8)	0.0013 (5)	0.003	-1.55*	-1.57*
ENO1	0.0052 (8.9)	0.0040 (7)	<0.001	-0.95*	-0.78
FGB	0.0045 (10.6)	0.0042 (11)	0.011	-1.12*	-0.85
GH1	0.0051 (17.2)	0.0061 (14)	<0.001	-0.95	-1.64*
GHRL	0.0152 (24.3)	0.0106 (20)	0.050*	-0.63*	-1.43*
GRIN1	0.0099 (16.9)	0.0114 (21)	0.056*	0.60*	0.38
CPD	0.0120 (16.4)	0.0135 (15)	<0.001	-1.19	-0.61
ANXA11	0.0148 (5.6)	0.0160 (5)	<0.001	-0.33	0.44
ODC1	0.0164 (15.1)	0.0176 (10)	0.005	0.83*	1.50
LCAT	0.0176 (29.9)	0.0181 (25)	<0.001	-0.40	-0.41
NCL	0.0249 (35.3)	0.0252 (30)	0.009	1.05	1.25
SAA	0.0270 (23.0)	0.0269 (22)	<0.001	1.63	1.44
Mean autosomal	0.0094 (13.5)	0.0093 (11.8)	0.014	-0.57	-0.52
Z-chromosome					
ALDOB	0.0004 (4.8)	0.0001 (3)	<0.001	-1.95*	-1.47*
MUSK	0.0003 (3.1)	0.0009 (3)	0.148*	-1.41*	-0.32
ADAMTS6	0.0002 (3.8)	0.0010 (2)	0.852*	-1.94*	0.08
CHD1Z	0.0021 (4.8)	0.0013 (3)	0.057*	-0.80	-0.22
BRM	0.0010 (4.8)	0.0016 (2)	0.308*	-1.91*	1.60
BRIX	0.0028 (8.4)	0.0017 (4)	0.537*	-2.09*	-1.15
ATP5A1Z	0.0042 (12.1)	0.0044 (12)	0.016*	-0.74	-0.74
Mean Z-linked	0.0016 (6.0)	0.0016 (4)	0.274	-1.55	-0.32

**Indicates significant values (P < 0.05).*

Isolation-migration

I used a two-population isolation-with-migration model to estimate the demographic history of divergence between chestnut teal and grey teal treating auDNA and zDNA in separate analyses (Fig. 2.2, Table 2.2). In zDNA, the coalescent estimate of divergence time between the grey and chestnut teal peaked at 0.124 (t_z= 0.124, 95% HPD=0.06-0.28). Z-linked estimates of effective population size in grey teal was larger than chestnut teal with no overlap in the posterior distributions (HPD; $\theta_{GT\text{-}Z}$ = 4.91, 95% HPD = 2.55-11.69; $\theta_{CT\text{-}Z}$ = 0.17, 95% HPD= 0.05-0.69). Similarly, ancestral zDNA effective population size was also significantly smaller than the effective population size of grey teal, but larger than that of the chestnut teal (θ_A = 0.31, 95% HPD= 0.03-0.91). There was evidence of significant gene flow (forward in time) from the grey teal into chestnut teal in zDNA, and zero gene flow was not included in the confidence interval ($m_{GT \rightarrow CT}$ = 7.92; HPD 95% = 1.52-42). However, the model was consistent with zero gene flow in the opposite direction ($m_{CT \rightarrow GT}$ = 0.02, HPD 95% = 0-1.52).

In auDNA, the coalescent isolation-migration model revealed a much more recent divergence time between the species than did the sex-linked loci (t = 0.01, 95% HPD=0.002-0.029; Fig. 2.2, Table 2.2). Consistent with the zDNA, the effective population size of the grey teal population was larger ($\theta_{GT\text{-}au}$ = 3.13, 95% HPD=1.38-22.21) than that of the chestnut teal ($\theta_{CT\text{-}au}$ = 1.51, 95% HPD=0.33-20.51) and the

ancestral population ($\theta_{A\text{-}au}$ = 1.48, 95% HPD=1.03-1.93), but the confidence intervals were overlapping among all estimates (although $\theta_{GT\text{-}au}$ and $\theta_{A\text{-}au}$ only slightly overlapped; Fig.2.2, Table 2.1). The migration rates were more consistent with zero gene flow in either direction, but the confidence limits were wide and did not differ between the two directions ($m_{GT \rightarrow CT}$ = 0.99, HPD 95% = 0-53; $m_{CT \rightarrow GT}$ = 0.21; HPD 95% = 0-50). These posterior distributions contrasted markedly with those from zDNA, which were much narrower and consistent with very low levels of gene flow.

Table 2.2: Sex-linked and autosomal demographic parameters estimated using isolation-with-migration model in IMa2. Effective number of migrants (forward in coalescence)

	Autosomal	**Z-chromosome**
$\boldsymbol{\theta_{GT}}$	3.1 (1.4-22)	4.9 (2.6-12)
$\boldsymbol{\theta_{CT}}$	1.5 (0.3-21)	0.17 (0.05-0.69)
$\boldsymbol{\theta_{A}}$	1.5 (1.0-1.9)	0.31 (0.03-0.91)
$\boldsymbol{2Nm_{GT \rightarrow CT}}$	0.37 (0-281.5)	0.03 (0-5.5)
$\boldsymbol{2Nm_{CT \rightarrow GT}}$	0.37 (0-283)	1.7 (0.62-3.3)
$\boldsymbol{t}$	0.01 (0.002-0.029)	0.12 (0.06-0.28)

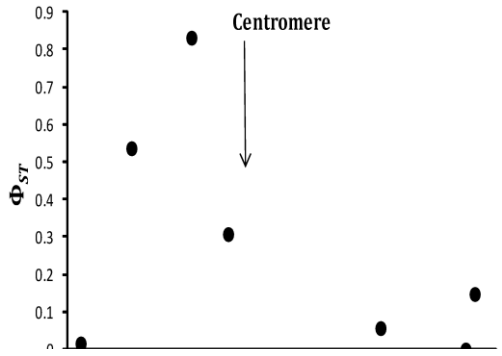

Figure 2.1 Spatial distribution of Φ_{ST} between the grey teal and chestnut teal over the Z-chromosome.

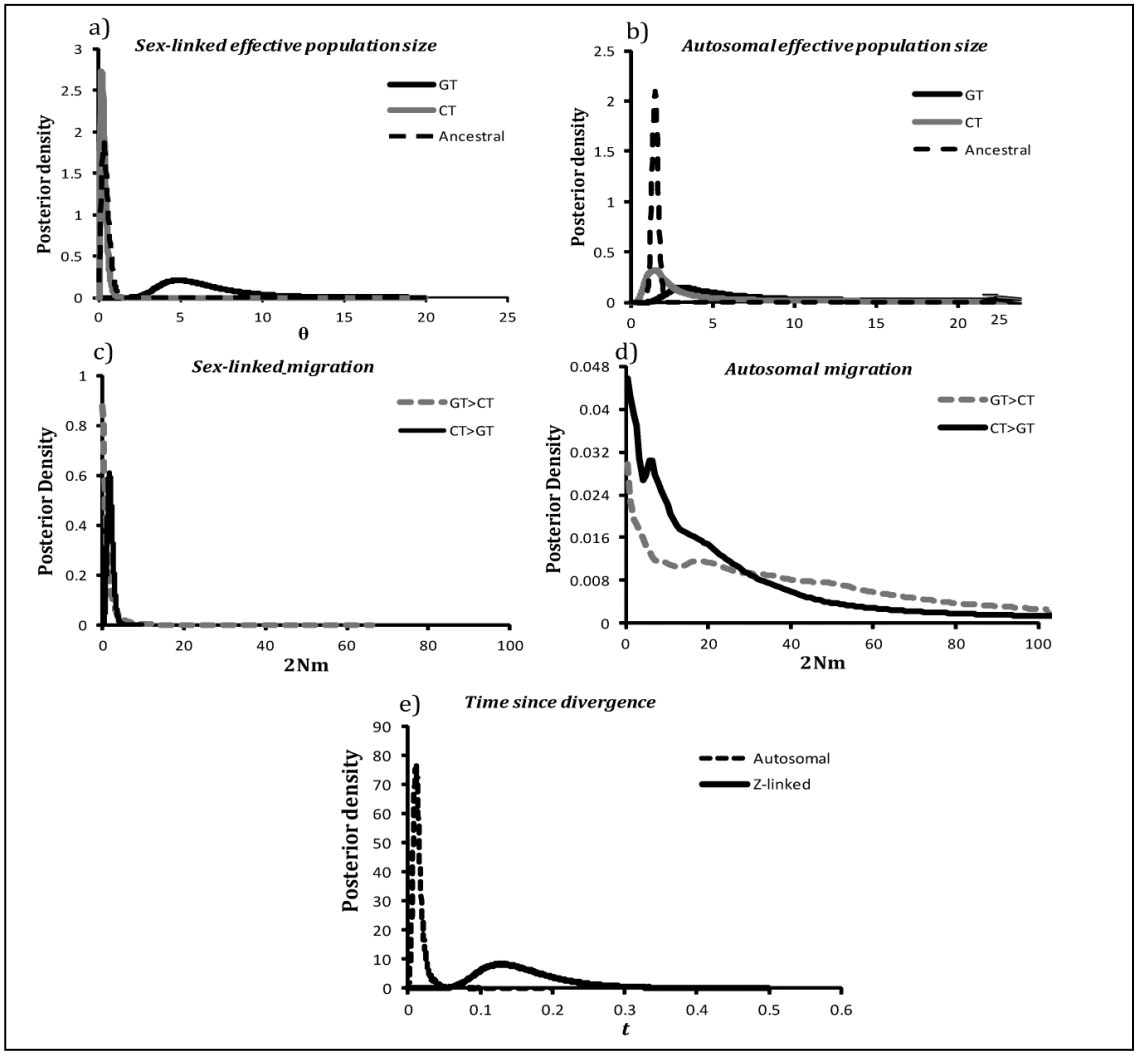

Figure 2.2 (a,b) Posterior distributions of sex-linked and autosomal effective population size; 1(c,d) Posterior distributions of migration rates in sex-linked and autosomal DNA as estimated in the IMa (scaled to mutation rate); (1e) Posterior distribution of time since divergence between the grey and chestnut teal for zDNA and auDNA.

Simulations

The goodness-of-fit test revealed a poor fit between the empirical data and the data simulated under neutrality and the demographic history estimated with the isolation-with-migration model. The mean empirical value of Φ_{ST} in auDNA was much smaller than predicted from the Z-chromosome history (Fig. 2.3a). Likewise, the empirical mean of Φ_{ST} for zDNA was much larger than the expectations obtained from the auDNA demographic history. Similarly, simulations under the isolation-migration model over-predicted Tajima's *D* in zDNA but under-predicted Tajima's *D* in auDNA in grey teal, but not chestnut teal (Fig. 2.3b). Additionally, the observed mean auDNA nucleotide diversity in both grey teal and chestnut teal deviated from expectations based upon the zDNA demographic history and selective neutrality, but this deviation was not as strong for zDNA simulated under the auDNA history (Fig. 2.3).

Locus-specific tests of Φ_{ST}, Tajima's *D* and π revealed several outlier loci in both auDNA and zDNA (Fig.2. 4 and 2.5). The isolation-migration model of neutral auDNA demographic history under-predicted the level of differentiation in zDNA for four of the loci (ADAMTS6, BRIX, BRM and MUSK). Three of these loci were located on the shorter p-arm of the Z-chromosome. In addition, three of these loci and ALDOB (also on the p- arm) had a significantly more negative Tajima's *D* in the grey teal but not in the chestnut teal. Similarly, the neutral model of auDNA

demographic history over- predicted the nucleotide diversity in all z-loci except ATP5A1Z in both species; however the empirical values fell within the 95% confidence intervals. On the contrary, several outlier auDNA loci (ODC1, SAA1, ANXA11, LDHB, NCL, LCAT, GREL, and CPD) had higher nucleotide diversity than predicted by the zDNA demographic history. Most of these auDNA loci also had a significantly higher-than-expected Tajima's *D* for grey teal (GRIN1, NCL, ODC1, and SAA1). LDHB also had a high *D* in chestnut teal. Finally, the observed levels of Φ_{ST} were over-predicted under neutrality in all autosomal loci, although the values were within the 95% confidence intervals.

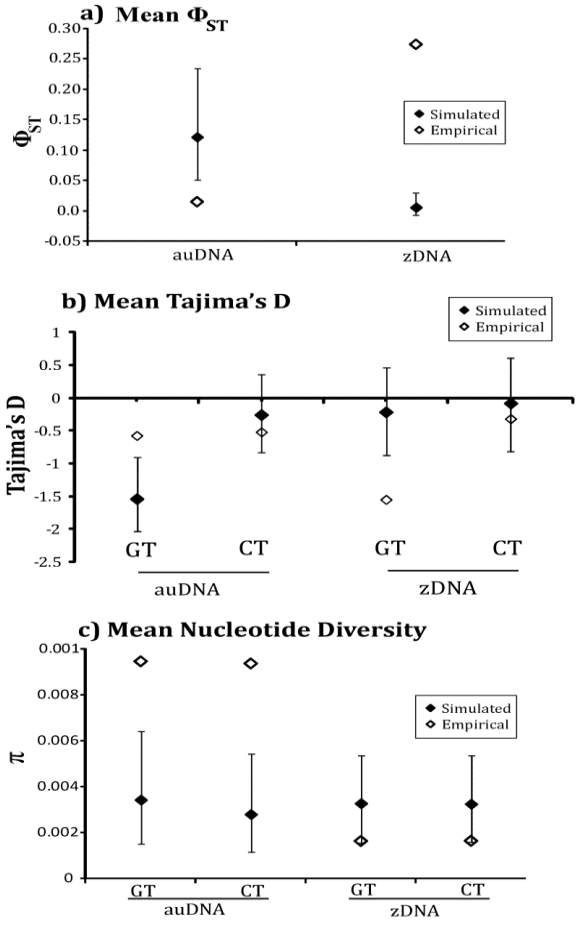

Figure 2.3 A comparison of mean Φ_{ST} (3a), Tajima's D (3b) and nucleotide diversity (3c) over the simulated autosomal and sex-linked data with that observed in the empirical autosomal and sex-linked data between grey and chestnut teal. Filled markers represent the average Φ_{ST} estimated from simulations under neutrality while the bars indicate the 95% confidence interval. Open markers represent the mean observed values.

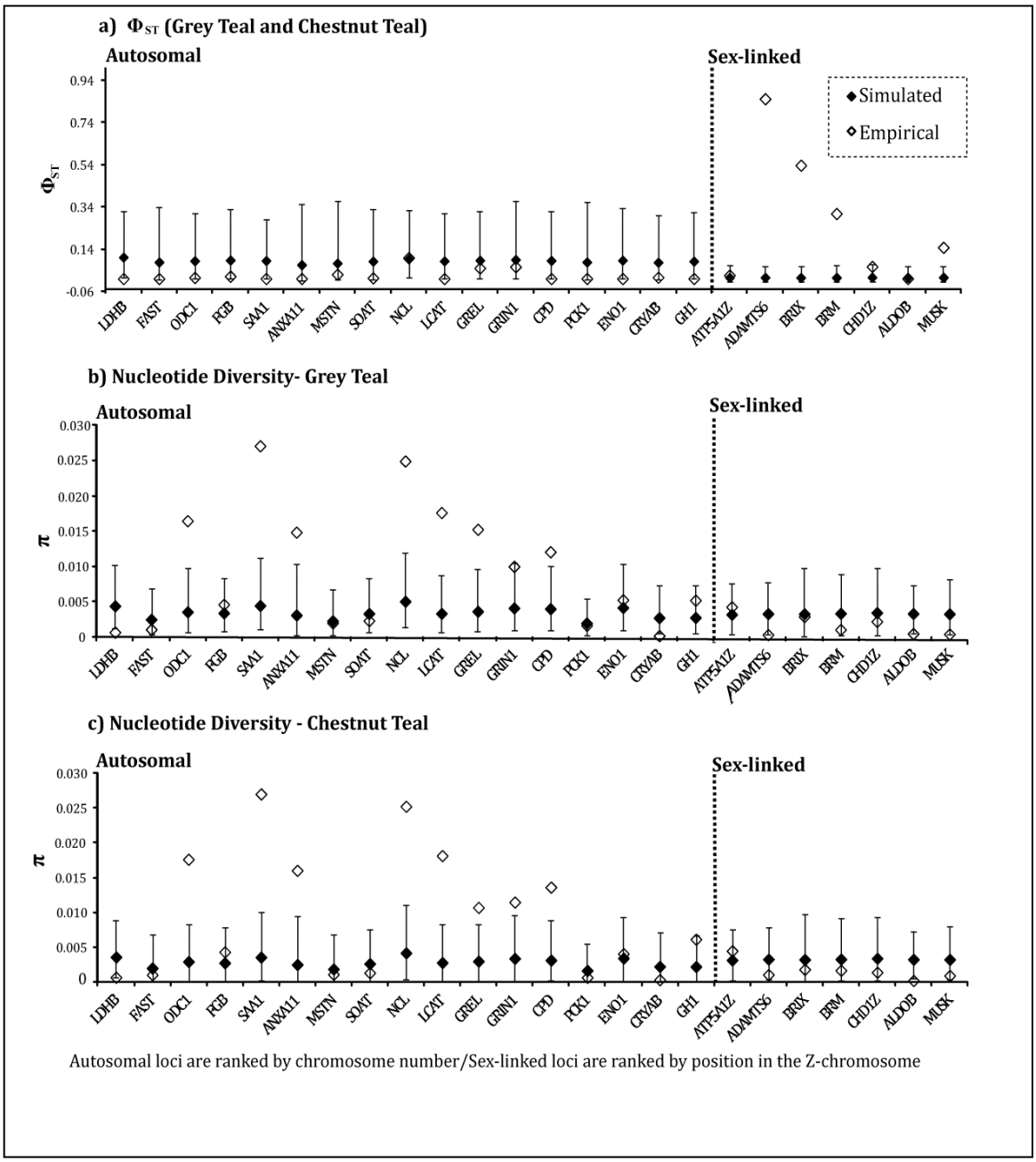

Figure 2.4 a) Locus-specific goodness of fit tests for Φ_{ST} between grey teal and chestnut teal in auDNA and zDNA. Locus-specific goodness of fit test of nucleotide diversity in the nineteen autosomal and seven Z-linked loci of (b) grey teal and (c) chestnut teal. Open markers and filled markers symbolize the empirical and expected value for the respective locus. Vertical bars represent the 95% confidence limits under the neutral expectations.

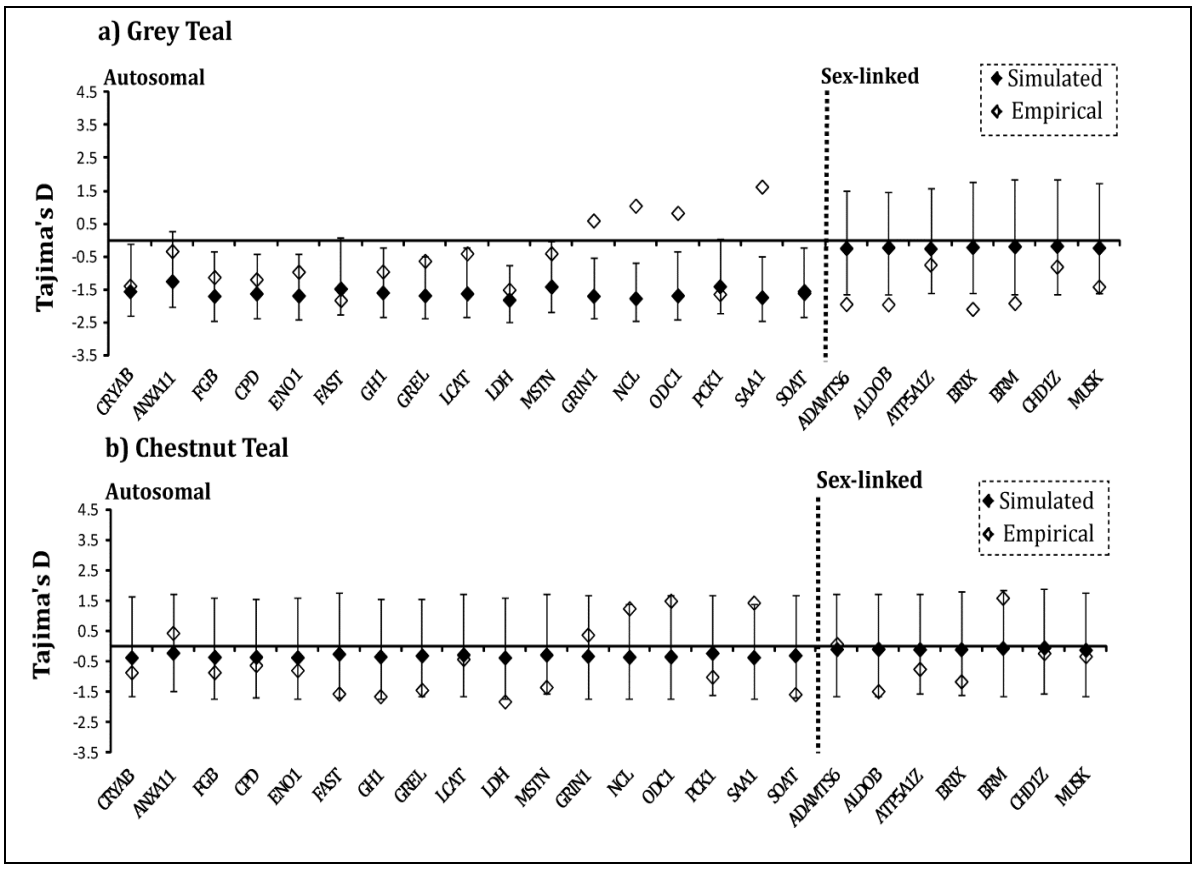

Figure 2.5 Species specific locus-wise goodness-of-fit test of Tajima's *D* in auDNA and zDNA of (a) grey teal and (b) chestnut teal. Open markers and filled markers indicate the empirical and simulated value of the locus. Vertical bars represent the 95% confidence limits under neutral expectations.

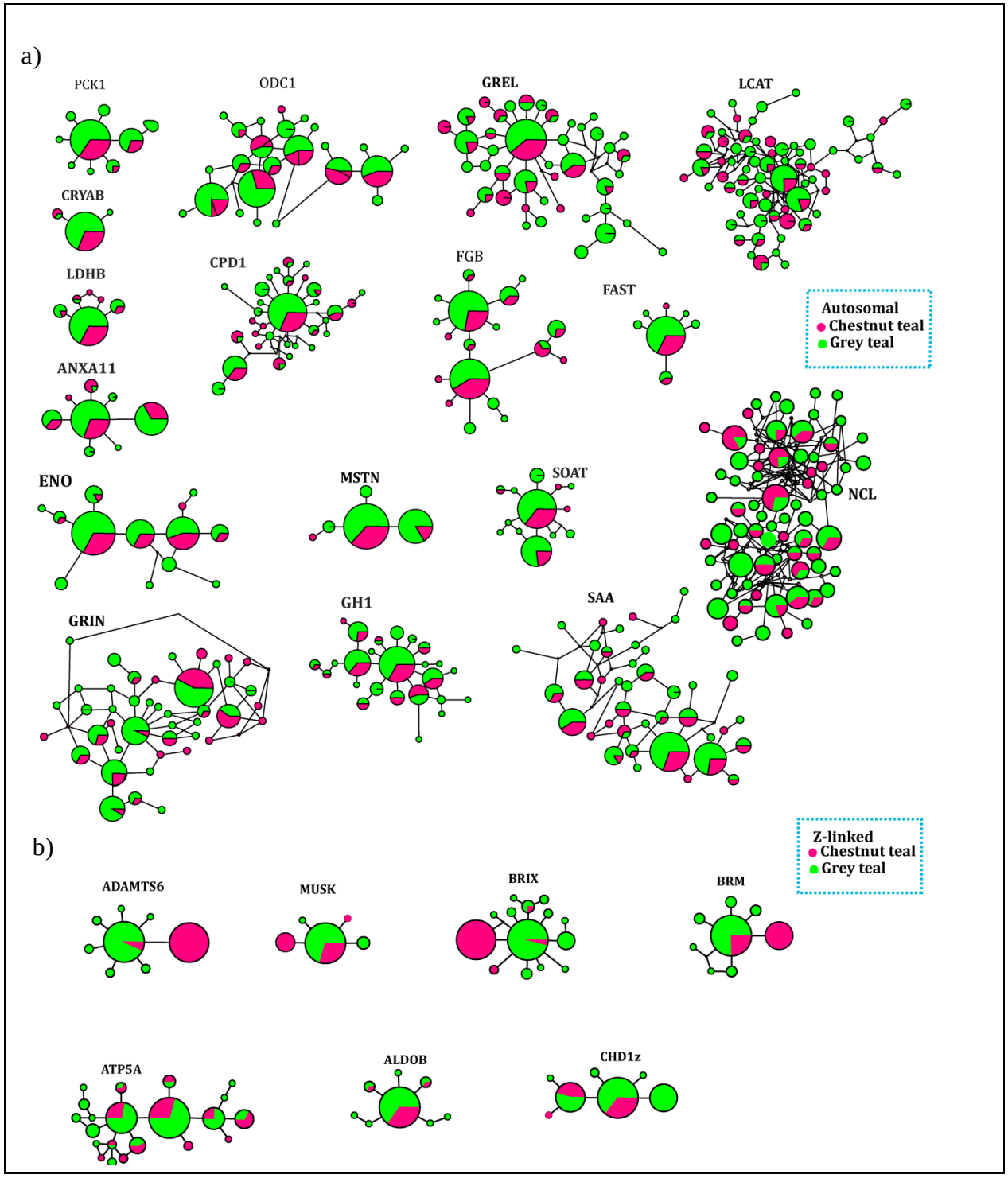

Figure 2.6 Haplotype networks between the grey teal and chestnut teal for (a) seventeen autosomal loci and (b) seven sex-linked loci.

DISCUSSION

I sequenced two different classes of nuclear DNA to test for a more rapid and larger Z-effect in two morphologically, behaviorally, and ecologically differentiated taxa during the early stages of divergence. This comparative analysis between the Australian grey teal and chestnut teal revealed nearly no differences in auDNA, which is consistent with a lack of divergence in mtDNA (Joseph *et al.*, 2009; Chapter 1), despite strong differentiation in zDNA. This contrast suggests an important role of the Z-chromosome during the early stages of speciation. In addition to the disproportionately larger divergence, we found reduced variation on the Z-chromosome, further supporting the large Z-effect hypothesis. Simulations under neutral demographic histories demonstrate that differences in substitution rates, N_e, and migration rates are insufficient to explaining the disparity in divergence and diversity between auDNA and zDNA. Finally, we also found evidence of an "island of differentiation" (Carneiro et al., 2010) near the central part of the p-arm of the Z-chromosome where differentiation was elevated relative to the remainder of the chromosome (Fig. 2.3).

Large Z-effect

The large Z-effect hypothesis proposes accelerated divergence of zDNA and an overrepresentation of Z-genes in the pool of rapidly diverging genes. The hypothesis has its foundation in the observations of male-biased mutation rates (Hurst and Ellegren, 1998), reduced recombination on the Z-chromosome, a higher influence of

genetic drift owing to a smaller N_e, and linkage of male ornamental traits (e.g., *Ficedula* flycatchers; Saether *et al*., 2007) and hybrid fitness to zDNA (Backstrom *et al*., 2010). The divergence between the Australian teals that primarily differ in male plumage is consistent with the proposed hypothesis of a large Z-effect. In support of a large Z-effect for teals, the mean value of Φ_{ST} for the zDNA was underpredicted by the neutral demographic model inferred from auDNA; thus, zDNA was strongly differentiated relative to auDNA. This larger association of highly differentiated regions to zDNA could suggest a greater role of the Z-chromosome in adaptive evolution (Ellegren, 2009; Charlesworth *et al*., 1987). Although the genes responsible for the expression of highly differentiated traits in teals are not known specifically, male plumage characters and female preference genes are Z-linked in hybridizing flycatcher species and appear to be under selection (Reeve and Pfennig, 2002; Sætre *et al*., 2003; Sæther *et al*., 2007; Backström *et al*., 2010).

In addition to high differentiation, the most divergent regions also had significantly smaller intraspecific diversity, an excess of rare polymorphisms, and high linkage disequilibrium in the grey teal, and these features are consistent with a prominent role of selection (Ellegren *et al*., 2012). Moreover, a standard neutral model predicts the neutral ratio of Z:A variation to be 0.75, assuming equal numbers of reproducing males and females, random mating, constant population size, no gene flow, and no difference in mutation rate between zDNA and auDNA. However, in birds, a lower-than-expected ratio of 0.32–0.42 and 0.24 has been observed in *Ficedula* flycatchers and the chicken, respectively, which has been attributed to the z-loci being linked to genes that underwent a selective sweep (Saetre *et al*., 2003;

Sundstrom *et al*., 2004). In this study, the Z:A ratio was 0.17 in both chestnut teal and grey teal (Table 2.2), suggesting an even stronger deviation from the neutral model.

Alternatively, variability in mutation rates between different classes of DNA and the effectiveness of genetic drift could explain the accelerated divergence of zDNA. Male-biased mutation rates can cause Z- linked loci to accumulate substitutions faster than autosomal loci, because zDNA spends 2/3 of its time in the male germline, and this rapid evolution can be especially prominent when new mutations are favorable (Kirkpatrick and Hall, 2004). On the other hand, the greater influence of drift in smaller populations could cause even weakly deleterious mutations to become fixed more rapidly on the Z-chromosome (Laporte and Charlesworth, 2002). However, we accounted for the differences in mutation rates and drift between zDNA and auDNA in our simulations, suggesting that these factors fail to explain the deviations without some role of selection.

This pattern of elevated divergence in zDNA relative to auDNA has been linked to the signatures of selection in other taxa with a deeper overall divergence. A genomic scan of divergence between two flycatcher species (i.e pied flycatcher, *Ficedula hypoleuca* and the collared flycatcher, *F. albicollis*) showed much greater divergence on Z-chromosome along with several "islands of differentiation" on auDNA that were characterized by reduced diversity, elevated linkage disequilibrium and skewed allele frequency spectra (Ellegren *et al.*, 2012). Similarly, a high density scan of 72 Z-linked loci between the species revealed similar signatures that were consistent with directional selection in Z-linked loci, which contained candidate

regions associated with plumage coloration (Saetre *et al*., 2003; Borge *et al*., 2005; Saethre *et al*., 2007; Backstrom *et al*., 2010). Likewise, two Z-linked genes with outlier F_{ST}'s between the house sparrow (*Passer domesticus*) and Spanish sparrow (*P. hispaniolensis*) were inferred to be under positive, divergent or diversifying selection (Elgvin *et al*., 2011). Overall, zDNA showed elevated interspecific divergence and reduced intraspecific variation relative to autosomes. These studies support the large Z-effect in the later stages of speciation, where mtDNA and auDNA are more differentiated than what we observed (Sætre *et al*. 2003, Borge *et al*. 2005, Carling and Brumfield, 2008, Carling *et al.*, 2010; Storchova *et al*., 2010). Thus, these teal species are perhaps one of the most recently diverged species pairs exhibiting evidence of a large Z-effect.

Differential introgression and divergent selection

Signatures of elevated divergence and lower diversity of zDNA could result from differential introgression and/or divergent selection. Haldane's Rule predicts reduced fitness of the heterogametic sex (females in birds), which causes more restricted introgression for zDNA and mtDNA than auDNA (Carling and Brumfield, 2008; Storchova *et al*., 2010). A reduction in Z-introgression is expected, because the Z-chromosome contains recessive alleles that reduce hybrid fitness and has greater cumulative effects in the hemizygous state (Haldane, 1922; Turelli and Orr, 1995). The Z-chromosome might also contain genes important in pre-zygotic isolation (Carling and Brumfield, 2009). However, the estimates of migration between the teal taxa do not conform to the expectations of Haldane's rule. Coalescent analyses were

consistent with no gene flow in auDNA (although the confidence intervals were large), but we rejected a hypothesis of no gene flow in zDNA from the grey teal into chestnut teal (Fig.2.1). Furthermore, owing to its maternal inheritance, mtDNA should be less likely to introgress (Scribner *et al*., 2001; Cianchi *et al*., 2003; Carling and Brumfield, 2009), but grey teal and chestnut teal were not differentiated at this locus (Joseph *et al.*, 2009). Thus, Haldane's Rule seems to be an unlikely explanation for the differences in zDNA and auDNA.

The most prominent difference between auDNA and zDNA from coalescent analyses was in estimates of time since divergence rather than differences in rates of introgression. The divergence in zDNA was about an order of magnitude deeper than auDNA. Because estimates of *t* from IMa2 are scaled to the mutation rate (i.e., $t = T\mu$), this result could be explained if mutation rates for zDNA are an order-of-magnitude faster compared to auDNA. Although several studies support a higher mutation rate for zDNA, which has been attributed to male-biased mutation (Ellegren and Fridolfsson, 1997; Kahn and Quinn, 1999; Axelsson *et al*., 2004), the observed <2-fold difference is unlikely to account for our results. In ducks specifically, mutation rate estimates for CHD1Z were similar to estimates for auDNA (Peters *et al*. 2008, 2012c). In contrast, analyses of auDNA and mtDNA yielded concordant estimates of divergence times after applying a correction for differences in mutation rates (Chapter 1). Finally, a higher mutation rate for zDNA should also cause zDNA to have higher genetic diversity than auDNA, but I found significantly less diversity in zDNA. Thus, a faster mutation rate is insufficient to explain the differences in estimates of *t* between auDNA and zDNA. Overall, the data suggest more rapid

divergence of zDNA than auDNA, which is consistent with strong divergent selection

Z-effect and speciation

Exclusively Z linked divergence between the teals with almost negligible differentiation on the mtDNA and auDNA is suggestive of sex-linked speciation between the grey and chestnut teal. Evidence suggests that prezygotic and postzygotic traits of reproductive isolation, including sexual dichromatism, mate preferences and hybrid sterility, are disproportionately controlled by Z-linked genes (Noor *et al*., 2001; Servedio and Seatre., 2003; Saether *et al*., 2007; Carling and Brumfield, 2009). Chestnut teal are strongly sexually dichromatic with brightly colored males, whereas grey teal are monochromatic and have dull plumage. Given the results of previous research and the contrast between zDNA and auDNA in this study, I hypothesize that this morphological trait is linked to the Z-chromosome and that sexual selection is a primary driver of divergent selection in this pair. However, failure to exclude the complete genetic isolation between the grey teal and chestnut teal in the coalescent based analyses suggests that these taxa have not achieved complete reproductive isolation, although gene flow is probably rare. This emphasizes the possible importance of pre-zygotic traits between the teal taxa as the major mechanism contributing to isolation and preventing complete genetic homogenization. This case study is consistent with the proposed hypothesis of linkage of sexual dimorphism to sex-chromosomes (Mank *et al*., 2009), suggesting sexual selection might be an important player in the early stages of divergence. Although evidence for genes encoding for ecological traits is lacking, it is possible

that zDNA could also have contributed to ecological speciation between the grey and chestnut teal. Given the strong divergence in morphology, behavior, and ecology coupled with weak overall divergence in DNA, this pair of taxa is an excellent model system for further testing the role of selection in driving the speciation process.

Z-Island of genetic differentiation

We propose the central region on the *p*-arm of the Z-chromosome as an "island of differentiation" between the teal species. Locus-specific comparisons on the zDNA detected four outliers (ADAMTS6, MUSK, BRIX and BRM) with their observed Φ_{ST} much greater than expected under neutrality (Fig. 2.3). Three of these were located within the same region (Fig. 2.2). Significantly negative Tajima's *D* in these outlier loci, evidence of linkage disequilibrium, and reduced diversity with some gene flow suggest a strong role of divergent selection in this island. However, we cannot rule out the possibility of several other islands of differentiation spread throughout auDNA. A larger dataset from a larger sample size of loci, a genomic scan in particular, could be beneficial in detecting the overall heterogeneity in differentiation in auDNA and zDNA.

Conclusion

I conclude that selection on the Z-chromosome was important during the early stages of speciation in these morphologically, behaviorally, and ecologically divergent teal. Thus, zDNA can differentiate between these taxa when the fast evolving mtDNA cannot. The role of the Z-chromosome in the later stages of speciation is evident from several other studies but this is the first study to document such a large effect

during the earliest stages of divergence. Comparisons of divergence and gene flow suggest that divergent selection rather than differential introgression has a major role in the generation and maintenance of genetic differentiation on the Z-chromosome.

REFERENCES

Axelsson E., N. Smith, H. Sundstrom, S. Berlin, and H. Ellegren. 2004. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. Mol. Biol. Evol. 21:1538-1547.

Backstroem N., and U. Vaeli. 2011. Sex- and species-biased gene flow in a spotted eagle hybrid zone. Bmc Evolutionary Biology 11:100.

Backstrom N., J. Lindell, Y. Zhang, E. Palkopoulou, A. Qvarnstrom, G. Saetre, and H. Ellegren. 2010. A High-Density Scan of the Z Chromosome in *Ficedula* Flycatchers Reveals Candidate Loci for Diversifying Selection. Evolution 64:3461-3475.

Bandelt H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37-48.

Barrett R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322:255-257.

Borge T., M. T. Webster, G. Andersson, and G. P. Saetre. 2005. Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. 171:1861-1873.

Carling M. D., and R. T. Brumfield. 2008. Haldane's rule in an avian system- using cline theory and divergence population genetics to test for differential introgression of

mitochondrial, autosomal, and sex-linked loci across the Passerina bunting hybrid zone. Evolution 62:2600-2615.

Carling M. D., and R. T. Brumfield. 2009. Speciation in Passerina buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. Mol. Ecol. 18:834-847.

Carling M. D., I. J. Lovette, and R. T. Brumfield. 2010. Historical Divergence and Gene Flow: Coalescent Analyses of Mitochondrial, Autosomal and Sex-Linked Loci in Passerina Buntings. Evolution 64:1762-1772.

Carneiro M., J. A. Blanco-Aguiar, R. Villafuerte, N. Ferrand, and M. W. Nachman. 2010. Speciation in the European Rabbit (*Oryctolagus Cuniculus*): Islands of Differentiation on the X Chromosome and Autosomes. Evolution 64:3443-3460.

Charlesworth B., J. Coyne, and N. Barton. 1987. The Relative Rates of Evolution of Sex-Chromosomes and Autosomes. Am. Nat. 130:113-146.

Cianchi R., A. Ungaro, M. Marini, and L. Bullini. 2003. Differential patterns of hybridization and introgression between the swallowtails *Papilio machaon* and *P. hospiton* from Sardinia and Corsica islands (Lepidoptera, Papilionidae). 12:1461-1471.

Crochet P. A., J. J. Z. Chen, J. M. Pons, J. D. Lebreton, P. D. N. Hebert, and F. Bonhomme. 2003. Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: Sex-biased interspecific gene flow? Evolution 57:2865-2878.

Elgvin T. O., J. S. Hermansen, A. Fijarczyk, T. Bonnet, T. Borge, S. A. Saether, K. L. Voje, and G. Saetre. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? Mol. Ecol. 20:3823-3837.

Ellegren H., and A. Fridolfsson. 1997. Male-driven evolution of DNA sequences in birds. Nat. Genet. 17:182-184.

Ellegren H. 2009. Genomic evidence for a large-Z effect. Proceedings of the Royal Society B-Biological Sciences 276:361-366.

Ellegren H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W. Wolf. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature advance online publication.

Emelianov I., F. Marec, and J. Mallet. 2004. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proceedings of the Royal Society B-Biological Sciences 271:97-105.

Excoffier L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.

Flot J. -. 2010. SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. Molecular Ecology Resources 10:162-166.

Frith H. J. 1967. Waterfowl in Australia. Angus & Robertson, Sydney.

Frith H. J.1963. Movements and mortality rates of the Black Duck and Grey teal in South-eastern Australia. CSIRO Wildl. Res. 8:119-131.

Frith H. J. 1962. Movements of the grey teal, *Anas gibberifrons* Muller (Anatidae). CSIRO Wildl. Res. 7:50-70.

Frith H. J. 1959. The ecology of wild ducks in inland New South Wales. II. Movements. CSIRO Wildl. Res. 4:108-130.

Grant P. R., and B. R. Grant. 2006. Evolution of character displacement in Darwin's finches. Science 313:224-226.

Hagen R., and J. Scriber. 1989. Sex-Linked Diapause, Color, and Allozyme Loci in *Papilio-Glaucus* - Linkage Analysis and Significance in a Hybrid Zone. J. Hered. 80:179-185.

Haldane J. 1922. Sex ratio and unisexual sterility in hybrid animals. Journal of Genetics 12:101-109.

Helbig A., M. Salomon, S. Bensch, and I. Seibold. 2001. Male-biased gene flow across an avian hybrid zone: evidence from mitochondrial and microsatellite DNA. J. Evol. Biol. 14:277-287.

Hendry A. 2004. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. Evol. Ecol. Res. 6:1219-1236.

Hey J. 2010. Isolation with Migration Models for More Than Two Populations. Mol. Biol. Evol. 27:905-920.

Hogner S., S. A. Sather, T. Borge, T. Bruvik, A. Johnsen, and G. Satre. 2012. Increased divergence but reduced variation on the Z chromosome relative to autosomes in Ficedula flycatchers: differential introgression or the faster-Z effect? Ecology and evolution 2: 379-396.

Hudson R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337-338.

Hurst L., and H. Ellegren. 1998. Sex biases in the mutation rate. Trends in Genetics 14:446-452.

Itoh Y., K. Kampf, C. N. Balakrishnan, and A. P. Arnold. 2011. Karyotypic polymorphism of the zebra finch Z chromosome. Chromosoma 120:255-264.

Jacobsen F., N. R. Friedman, and K. E. Omland. 2010. Congruence between nuclear and mitochondrial DNA: Combination of multiple nuclear introns resolves a well-supported phylogeny of New World orioles (*Icterus*). Mol. Phylogenet. Evol. 56.

Joseph L., G. J. Adcock, C. Linde, K. E. Omland, R. Heinsohn, R. T. Chesser, and D. Roshier. 2009. A tangled tale of two teal: population history of the grey (*Anas gracilis*) and chestnut teal (*A.castanea*) of Australia. 40:430-439.

Kahn N., and T. Quinn. 1999. Male-driven evolution among eoaves? A test of the replicative division hypothesis in a heterogametic female (ZW) system. J. Mol. Evol. 49:750-759.

Kirby R. E., G. A. Sargeant, and D. Shutler. 2004. Haldane's rule and American black duck x mallard hybridization. 82:1827-1831.

Kirkpatrick M., and D. Hall. 2004. Male-biased mutation, sex linkage, and the rate of adaptive evolution. Evolution 58:437-440.

Kraus R. H. S., A. Zeddeman, P. van Hooft, D. Sartakov, S. A. Soloviev, R. C. Ydenberg, and H. H. T. Prins. 2011. Evolution and connectivity in the world-wide migration system of the mallard: Inferences from mitochondrial DNA. Bmc Genetics 12:99.

Laporte V., and B. Charlesworth. 2002. Effective population size and population subdivision in demographically structured populations. Genetics 162:501-519.

Mank J. E. 2009. Sex Chromosomes and the Evolution of Sexual Dimorphism: Lessons from the Genome. Am. Nat. 173.

Mank J. E., E. Axelsson, and H. Ellegren. 2007. Fast-X on the Z: Rapid evolution of sex-linked genes in birds. Genome Res. 17:618-624.

McCracken K. G., M. Bulgarella, K. P. Johnson, M. K. Kuhner, J. Trucco, T. H. Valqui, R. E. Wilson, and J. L. Peters. 2009a. Gene Flow in the Face of Countervailing Selection: Adaptation to High-Altitude Hypoxia in the beta A Hemoglobin Subunit of Yellow-Billed Pintails in the Andes. Mol. Biol. Evol. 26:815-827.

McCracken K. G., C. P. Barger, M. Bulgarella, K. P. Johnson, M. K. Kuhner, A. V. Moore, J. L. Peters, J. Trucco, T. H. Valqui, K. Winker, and R. E. Wilson. 2009b. Signatures of High-Altitude Adaptation in the Major Hemoglobin of Five Species of Andean Dabbling Ducks. Am. Nat. 174:631-650.

Miyata T., H. Hayashida, K. Kuma, K. Mitsuyasu, and T. Yasunaga. 1987. Male-Driven Molecular Evolution - a Model and Nucleotide-Sequence Analysis. Cold Spring Harb. Symp. Quant. Biol. 52:863-867.

Noor M., K. Grams, L. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. U. S. A. 98:12084-12088.

Norman F. I., and R. S. Brown. 1988. Aspects of the Distribution and Abundance of Chestnut Teal in Southeastern Australia. Emu 88.

Nosil P., and J. L. Feder. 2012. Widespread yet heterogeneous genomic divergence. Mol. Ecol. 21:2829-2832.

Nosil P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375-402.

Orr H., and J. Coyne. 1989. The Genetics of Postzygotic Isolation in the Drosophila-*Virilis* Group. Genetics 121:527-537.

Peichel C., K. Nereng, K. Ohgi, B. Cole, P. Colosimo, C. Buerkle, D. Schluter, and D. Kingsley. 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414:901-905.

Peters J. L., Y. N. Zhuravlev, I. Fefelov, E. M. Humphries, and K. E. Omland. 2008. Multilocus phylogeography of a Holarctic duck: Colonization of North America from Eurasia by gadwall (*Anas strepera*). Evolution 62:1469-1483.

Peters J. L., Y. Zhuravlev, I. Fefelov, A. Logie, and K. E. Omland. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution 61:1992-2006.

Peters J. L., K. A. Bolender, and J. M. Pearce. 2012a. Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. Mol. Ecol. 21:3562-3575.

Peters J. L., K. G. McCracken, C. L. Pruett, S. Rohwer, S. V. Drovetski, Y. N. Zhuravlev, I. Kulikova, D. D. Gibson, and K. Winker. 2012b. A parapatric propensity for breeding precludes the completion of speciation in common teal (*Anas crecca*, sensu lato). Mol. Ecol. 21: 4563-4577.

Peters J. L., T. E. Roberts , K. Winker , and K. G. McCracken . 2012c. Heterogeneity in genetic diversity among non-coding loci fails to fit neutral coalescent models of population history. PLoS ONE 7:e31972.

Presgraves D. C. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution 56:1168-1183.

Reeve H., and D. Pfennig. 2003. Genetic biases for showy males: Are some genetic systems especially conducive to sexual selection? Proc. Natl. Acad. Sci. U. S. A. 100:1089-1094.

Roshier D., M. Asmus, and M. Klaassen. 2008. What drives long-distance movements in the nomadic Grey Teal *Anas gracilis* in Australia? Ibis 150:474-484.

Roshier D. A., N. I. Klomp, and M. Asmus. 2006. Movements of a nomadic waterfowl, Grey Teal *Anas gracilis*, across inland Australia - results from satellite telemetry spanning fifteen months. 94:461-475.

Rozas J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Saether S. A., G. Saetre, T. Borge, C. Wiley, N. Svedin, G. Andersson, T. Veen, J. Haavie, M. R. Servedio, S. Bures, M. Kral, M. B. Hjernquist, L. Gustafsson, J. Traeff, and A. Qvarnstroem. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science 318:95-97.

Saetre G., T. Borge, J. Lindell, T. Moum, C. Primmer, B. Sheldon, J. Haavie, A. Johnsen, and H. Ellegren. 2001. Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. Mol. Ecol. 10:737-749.

Saetre G. P., T. Borge, K. Lindroos, J. Haavie, B. C. Sheldon, C. Primmer, and A. C. Syvanen. 2003. Sex chromosome evolution and speciation in Ficedula flycatchers. Proceedings of the Royal Society B-Biological Sciences 270.

Saint-Laurent R., M. Legault, and L. Bernatchez. 2003. Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (Osmerus mordax Mitchill). Mol. Ecol. 12:315-330.

Sattler G. D., and M. J. Braun. 2000. Morphometric variation as an indicator of genetic interactions between Black-capped and Carolina chickadees at a contact zone in the Appalachian mountains. Auk 117:427-444.

Schluter D. 2009a. Evidence for Ecological Speciation and Its Alternative. Science 323:737-741.

Schluter D., and G. L. Conte. 2009b. Genetics and ecological speciation. Proc. Natl. Acad. Sci. U. S. A. 106:9955-9962.

Schluter D., K. B. Marchinko, R. D. H. Barrett, and S. M. Rogers. 2010. Natural selection and the genetics of adaptation in threespine stickleback. Philosophical Transactions of the Royal Society B-Biological Sciences 365:2479-2486.

Scribner K. T., M. R. Petersen, R. L. Fields, S. L. Talbot, J. M. Pearce, and R. K. Chesser. 2001. Sex-biased gene flow in spectacled eiders (Anatidae): Inferences from molecular markers with contrasting modes of inheritance. 55:2105-2115.

Servedio M. R., and G. P. Saetre. 2003. Speciation as a positive feedback loop between postzygotic and prezygotic barriers to gene flow. Proceedings of the Royal Society B-Biological Sciences 270.

Stephens M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73:1162-1169.

Stephens M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978-989.

Storchova R., J. Reif, and M. W. Nachman. 2010. Female Heterogamety and Speciation: Reduced Introgression of the Z Chromosome between Two Species of Nightingales. Evolution 64:456-471.

Sundstrom H., M. Webster, and H. Ellegren. 2004. Reduced variation on the chicken Z chromosome. Genetics 167:377-385.

Tajima F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by Dna Polymorphism. Genetics 123:585-595.

Tucker P., R. Sage, J. Warner, A. Wilson, and E. Eicher. 1992. Abrupt Cline for Sex-Chromosomes in a Hybrid Zone between 2 Species of Mice. Evolution 46:1146-1163.

Turelli M., and H. Orr. 1995. The Dominance Theory of Haldanes Rule. Genetics 140:389-402.

Turner T. L., and M. W. Hahn. 2010. Genomic islands of speciation or genomic islands and speciation? Mol. Ecol. 19:848-850.

Turner T. L., and M. W. Hahn .2007. Locus- and population-specific selection and differentiation between incipient species of Anopheles gambiae. Mol. Biol. Evol. 24:2132-2138.

Vicoso B., and B. Charlesworth. 2009. Effective Population Size and the Faster-X Effect: an Extended Model. Evolution 63:2413-2426.

Woerner A. E., M. P. Cox, and M. F. Hammer. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics 23:1851-1853.

CHAPTER III. COALESCENT HISTORY OF NUCLEAR INTROGRESSION BETWEEN HOLARCTIC GADWALLS AND EURASIAN FALCATED DUCKS FAILS TO EXPLAIN HETEROGENEITY IN GENETIC DIVERSITY

Abstract. Large variances in genetic diversity and differentiation among loci have been observed for many species. This heterogeneity can arise from a number of processes, including stochastic variance in introgression of alleles, complex demographic histories, and selection influencing some loci. The gadwall (*Anas strepera*) is a species of duck distributed across North America and Eurasia that has high heterogeneity in genetic diversity that fails to fit neutral coalescent models of population history. Both selection and neutral hybridization with the falcated duck (*A. falcata*) are plausible explanations for this heterogeneity. The objective of this study was to assess the rate of introgression using a genomic transect of non-coding loci (19 introns that map to 19 different chromosomes). I found strong evidence of introgression of nuclear alleles from falcated duck into North American gadwalls, but not Eurasian gadwalls. However, simulating genetic diversity under the model of neutral population history estimated using coalescent analyses, I found that introgression was insufficient to explain the observed heterogeneity in genetic diversity for both species. Furthermore, four loci were significantly more differentiated between species than expected. These analyses suggest a prominent role of selection in the among-locus heterogeneity in non-coding DNA.

INTRODUCTION

Levels of genetic diversity and differentiation can vary extensively across the genomes of a population or species (Avise, 2000; Hammer *et al*., 2004; Borge *et al*., 2005; Hadrill *et al*., 2005; Carneiro *et al*., 2010; Nosil and Feder, 2012; Peters *et al*., 2012c; Strasberg *et al*., 2012). Some level of among-locus heterogeneity is expected given the stochastic nature of mutation and genetic drift: new mutations (or substitutions) increase diversity by the random generation of new polymorphisms, whereas drift decreases diversity through the random loss of alleles. Among-locus variation in baseline mutation rates and differences in effective population sizes (N_e) among autosomal, sex-linked, and cytoplasmic DNA, contribute further to this heterogeneity. Under an assumption of no gene flow, diversity at a neutral locus will depend on the antagonistic and stochastic interaction between mutation and drift, which can yield high heterogeneity in genetic diversity across the genome (Rosenberg and Nordborg, 2002; Hudson and Turelli, 2003; Knowles and Richards, 2005). The stochastic interaction between these two forces, both of which contribute to the accumulation of genetic differences between diverging populations or species, can also cause among-locus heterogeneity in differentiation.

Fluctuations in population sizes (i.e., demography) and gene flow leave footprints in the pattern of genetic variation across the genome. For example, population bottlenecks cause a loss of genetic variation. Conversely, rare variants in the DNA sequences arise with population expansion. These population size changes can alter the level of genetic variation disproportionately across loci (Pool and Nielsen, 2007). The temporal variation in mutation and drift resulting from

fluctuations in population size add to among-locus heterogeneity within and between populations. In addition, introgression or gene flow adds diversity by causing new alleles to enter populations (Sweigart and Willis, 2003; Johannesan *et al*., 2006; Minder and Widmer, 2008; Kaiser *et al*., 2011) and enhance the genetic similarity between populations and species. On the other hand, populations evolving in complete isolation should have fewer shared polymorphisms and greater divergence than populations experiencing gene flow. Hence, differences in introgression rates among populations add to the pattern of among-locus heterogeneity in genetic diversity and differentiation between the populations.

Selection can influence genetic diversity across the genome by favoring polymorphisms from standing genetic variation or from a pool of new mutations. Positive selection, for example, can cause selectively advantageous alleles to spread throughout a population or between populations or species, resulting in low genetic variation and low differentiation at affected loci (Maynard-Smith and Haigh, 1974). Likewise, strong negative selection eliminates deleterious polymorphisms from the linked sites within a genome in a population (Charlesworth *et al*., 1993), whereas balancing selection maintains high levels of genetic diversity at some loci (Charlesworth, 1997). Similarly, divergent selection favors different alleles in two populations residing in different environments selectively enhancing the genetic differentiation at those loci, despite higher similarity throughout the remainder of the genome (Schluter, 2001; Nosil, 2009), whereas balancing selection inhibits population differentiation (Charlesworth *et al*., 1993). This variability in selective

pressures throughout the genome differentially influences variation within and between populations, creating high among-locus heterogeneity.

Apart from the influences of individual demographic or genomic forces, the interaction between selection and introgression is also important if gene flow is prevented at some loci by selection. However, the interaction between selection and introgression is not always antagonistic. Positive selection can facilitate the introgression of alleles at some loci and homogenize the genomes between the closely related taxa (Bachtrog *et al*., 2006; Currat *et al*., 2008), but divergent selection can prevent gene flow at other loci that are important for species integrity, thereby enhancing genetic differentiation at those loci (Noor *et al*., 2001; Morjan and Rieseberg, 2004; Kulathinal *et al*., 2009; Nosil *et al*., 2009). Hence, variability in selection pressures among loci creates heterogeneous diversity and differentiation by filtering introgressed alleles at some loci (Charlesworth, 1997; Postma and Noordwijk, 2005; Storz and Kelly, 2008; McCracken, 2009). Overall, among-locus patterns of genetic diversity and differentiation can arise from various demographic and genomic forces; a multi-locus approach can help elucidate the roles of these forces in generating among-locus heterogeneity

Study taxa

Sequence data from twenty-two non-coding loci in the gadwall (*Anas strepera*) revealed large among-locus heterogeneity in genetic diversity that varied by two orders of magnitude (Peters *et al*., 2012c). The gadwall has a Holarctic distribution that extends across Eurasia (Old World, OW) and North America (New World, NW; Fig.3.1) and is genetically characterized by haplotype frequency differences between

OW and NW in mitochondrial DNA (mtDNA) and nuclear DNA (nuDNA) that likely resulted from a founder effect (colonization of NW from OW) and restricted gene flow (Peters *et al*. 2008, 2012c). However, intragenomically, the observed among-locus heterogeneity failed to fit this model of population history, but neither selection nor introgression could be rejected as plausible contributing factors (Peters *et al*., 2012c).

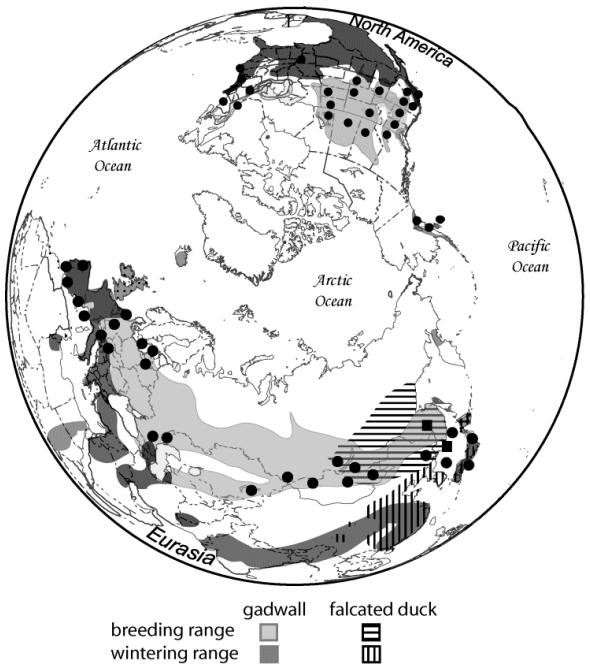

Figure 3.1 Falcated Ducks have a restricted distribution in eastern Asia (eastern Russia, Japan, northern China, and Mongolia), whereas Gadwalls have a Holarctic distribution extending across Europe, Asia, and North America. Black dots and squares represent sampling locations of gadwalls and falcated ducks with sample sizes >1 indicated. Modified from Peters *et al*. (2007). Sampling details have been provided in Peters *et al*., (2012c).

The gadwall hybridizes with its closest extant relative, the falcated duck *(Anas falcata*), in the wild (Johnsgard, 1960). Unlike the gadwall, the falcated duck has a restricted distribution in eastern Asia, where its breeding range overlaps with the gadwall (Fig. 3.1). Molecular data reveal introgression of falcated duck DNA into gadwalls. Falcated duck shared one haplotype each at mtDNA and one of the

two nuclear loci sequenced with sympatric OW gadwalls (Peters *et al*., 2007). In addition, 5.5% of North American gadwalls had mtDNA that was more similar to falcated ducks than to other gadwalls, although there was no evidence of nuclear introgression (Peters and Omland, 2007; Peters *et al*., 2007). However, one of the two nuclear markers sequenced was located on the Z-chromosome, which may be less susceptible to introgression (Borge *et al*, 2005; Storchova, 2010). Therefore, a multi-locus assessment is necessary to examine the extent of nuclear introgression between the falcated duck and each gadwall population to better examine the role of introgression in generating among-locus heterogeneity.

The main objectives of this study were i) to examine genetic diversity sampled from genomic transect of in falcated ducks and compare it to the among-locus heterogeneity observed in gadwall; ii) to estimate rates of introgression between species in sympatry and allopatry; and iii) to examine the role of introgression in among-locus heterogeneity in genetic diversity and differentiation. This multi-locus comparison of polymorphic data between falcated ducks and the gadwall populations will contribute to disentangling the roles of demographic history, introgression, and selection in generating among-locus heterogeneity in these taxa.

MATERIALS AND METHODS

I sequenced a genomic transect for 24 falcated ducks (Fig. 3.1), which included nineteen non-coding regions of nuclear DNA that map to different chromosomes in the chicken (*Gallus gallus*) genome using previously published primers (Peters *et al*.,

2012c; Table 3.1). Homologous data for 25 NW and 25 OW gadwalls were obtained from Peters *et al*. (2012c). Each locus was amplified using PCR and cleaned using AMPure XP beads following the Agencourt protocol (Beckman Coulter Co., Brea, CA). I sequenced PCR products using the BigDye v. 3.1 Terminator Cycle Sequencing Kit following manufacturer protocols (Applied Biosystems, Foster City, CA). Automated sequencing was performed on an ABI 3730 at the DNA Sequencing Facility on Science Hill, Yale University, CT. I edited the falcated duck sequences and aligned them with gadwall sequences using Sequencher 4.1 (Gene Codes, Ann Arbor, MI). I determined the gametic phases of sequences that were heterozygous at more than one nucleotide position using the software PHASE 2.1 (Stephens *et al*., 2001, Stephens and Donnelly, 2003).

Genetic Diversity and Demography

I quantified genetic variation within the populations and differentiation among populations in terms of the nucleotide diversity (π, the average number of nucleotide differences per site between pairs of randomly selected individuals from a population), pairwise Φ_{ST} between falcated duck and gadwall (the proportion of genetic diversity attributable to differences among populations), and Tajima's *D* (a measure of the relative frequency of rare polymorphisms to common polymorphisms) in Arlequin v3.11 (Excoffier and Lischer, 2010). I used linear regression to compare π between falcated ducks and gadwalls. A paired *t*-test was used to compare Φ_{ST} between falcated ducks and gadwalls with Φ_{ST} between the two gadwall populations with Φ_{ST} being paired by loci. I constructed haplotype networks using the median-joining algorithm in NETWORK ver. 4.1 (Bandelt *et al.*, 1999).

To infer aspects of the population histories of falcated ducks and OW and NW gadwalls, I applied the MCMC Bayesian approach in a three-population isolation with migration model in the coalescent program IMa2 (Hey, 2010). Demographic history was estimated under three possible scenarios of migration: a full migration model with both ancestral and ongoing gene flow, a model of recent secondary contact that assumes no gene flow between ancestral populations, and a model of ancestral migration that assumes no-ongoing migration. The estimated parameters included time since divergence (t_0 and t_1, the divergence times between OW and NW gadwalls and between falcated ducks and gadwalls, respectively, where $t = T\mu$, and T is the time since divergence in years and μ is the geometric mean of mutation rates per locus among all loci), the effective population sizes of the ancestral populations (Θ_{A0} at t_0 and Θ_{A1} at t_1, where $\Theta_A = 4N_{eA}\mu$ and N_{eA} is the ancestral effective population size), and the effective population size of each daughter population (Θ_f, Θ_{ow}, and $\Theta_{nw,}$ for falcated ducks, OW gadwalls and NW gadwalls, respectively). The full model included eight migration parameters: two parameters (bidirectional migration) between each population pair and between the falcated duck and the ancestral gadwall population (M_{ij} is migration of alleles into population i from population j forward in time, where $M_{ij} = m_i/\mu$ and m_i is the rate at which alleles enter population i from population j). The model of secondary contact only included migration between extant populations (six migration parameters), whereas the ancestral migration model only included migration between the two gadwall populations and between the falcated duck and the ancestral gadwall population (four parameters). I converted migration rates into the number of effective migrants per

generation as θM_{ij} /2. Because IMa2 assumes no recombination within loci, I chose blocks of nucleotides consistent with no recombination that contained the maximum number of variable sites for each locus in IMgc (Woerner *et al.*, 2007). I iteratively adjusted the chromosomal weighting so that a maximum of 5% of chromosomal copies were removed from the analysis. I ran IMa2 on this recombination-filtered, nineteen-locus data set for $2x10^7$ steps following a burn-in of one million steps using thirty markov chains (one hot and 29 cold chains). I replicated the analysis three times with different random number seeds to check for convergence.

I also used the MCMC Bayesian method in the coalescent program LAMARC v2.1.6 (Kuhner, 2006) to jointly estimate recombination rates (r, *where* $r = C/\mu$,C is the rate of recombination per inter-site link per generation, and μ is the mutation rate per site) for each locus in falcated duck. I jointly estimated Θ (where $\Theta = 4N_e\mu$, and N_e is the effective population size) and the exponential growth rate (g, where $\Theta_t = \Theta_o exp^{-gt}$, and Θ_o is an index of the current N_e and Θ_t is an index of N_e at time t). I used the Felsenstein 84 model of substitution (ti:tv=2.5; the average ratio among loci) and ran the program for a burn-in of 2,000,000 generations, sampling every 1,000 generations for a total of 20,000 samples. To verify the consistency of the estimates, I replicated the run with a different random number seed.

Coalescent Simulations

I simulated genetic diversity and differentiation in each population to assess the role of introgression in the among-locus heterogeneity in genetic diversity under the assumptions of neutrality. For these simulations, I followed the protocol described by

Peters *et al*. (2012c), which incorporated the demographic parameters estimated from isolation-with-migration models, recombination rates from the LAMARC analyses, and evolutionary substitution rates estimated from a comparison of eight deeply divergent taxa (obtained from Peters *et al*. 2012c). I simulated genetic diversity and differentiation under each of the three migration models: full migration, secondary contact, and ancestral migration. For each parameter, I randomly sampled 1000 values from their respective posterior distributions so that uncertainty in these values was incorporated into the simulations.

Simulations were conducted under an assumption of neutral population history in the program MS (Hudson, 2002). All parameters were scaled to Θ_f, and the parameters for *CHD1z* were scaled by a factor of 0.75 to reflect the difference in effective population size resulting from linkage to the sex-chromosome Z. Polymorphism data were simulated 1,000 times for each locus (each replicate had a slightly different population history as described above) under each of the three models (19,000 simulations per model). From each simulated data set, I calculated π, Φ_{ST}, and Tajima's *D* in the program MS.output (Peters *et al*. 2012c).

Goodness-of-fit test

I performed a goodness-of-fit test to test the fit of the empirical data to the models of population history (Becquet and Przeworski, 2009; Peters *et al*., 2012c). For the population level goodness-of-fit tests, I compared empirical values of mean π and Φ_{ST} and their associated coefficients of variation (CV) with the expected values for a 19-locus dataset obtained from the simulated data sets (1,000 values per model). I also compared locus-specific values of each parameter to determine whether any loci

were consistent outliers from model expectations. I rejected the null hypothesis of no difference between expected and empirical values if the empirical values were outside the 95% CI of expected values.

RESULTS

Genetic Variation and Population structure

Sequence data from 19 non-coding loci revealed that heterogeneity in nucleotide diversity (π) for the falcated duck was similar to that observed in NW and OW gadwall populations (Table 3.2, S1). Overall, π in falcated duck (mean π = 0.0097, range =0.0002–0.0251) was similar to values observed in both OW (mean π = 0.0091, range = 0.0001–0.0231) and NW (mean π = 0.0090, range = 0.0001–0.0243) gadwalls. Indeed, nucleotide diversity among the 19 loci in falcated ducks was significantly correlated with that in OW gadwalls (R^2 = 0.88, df = 18, P = 3 x10^{-9}) and NW gadwalls (R^2 = 0.80, df = 18, P = 2x10^{-7}). Average Tajima's D was negative for each of the three populations (D_{FD} = -0.52 $\pm$ 0.97 StDev; D_{OW} = -0.44 $\pm$ 0.79 StDev; D_{NW} = -0.11 $\pm$ 0.74 StDev). The index was significantly negative for four loci in falcated ducks (*CRYAB, FAST, LDHB, GRIN1*) and OW gadwalls (*Sf3A2, ENO1, FAST, GRIN1*) and for two loci in NW gadwalls (*Sf3A2, GRIN1*) (Table 3.2).

Population pairwise comparisons indicated that the falcated duck is significantly differentiated from both gadwall populations (mean $\Phi_{ST\ (OW\text{-}FD)}$ = 0.281, range = 0.035–0.965; mean $\Phi_{ST\ (NW\text{-}FD)}$ = 0.286, range = 0.032–0.901). Differentiation was significantly lower between OW and NW gadwalls ($\Phi_{ST\ (OW\text{-}NW)}$ = 0.057, range = -0.014–0.184; t = 1.73, df = 18, P < 0.005). Falcated duck was more differentiated

from NW gadwalls than OW gadwalls in nine of the nineteen loci (Table 3.2, S1); however, these differences were not significant (t = 1.73, df = 18, P = 0.8). Haplotype networks revealed that many polymorphisms were shared between falcated ducks and both gadwall populations in most of nuclear loci investigated; only *CRYAB*, *LDHB* and *CHD1Z* were consistent with reciprocal monophyly between the species (Fig. 3.2). On the other hand, NW and OW gadwalls shared polymorphisms in all nuclear markers.

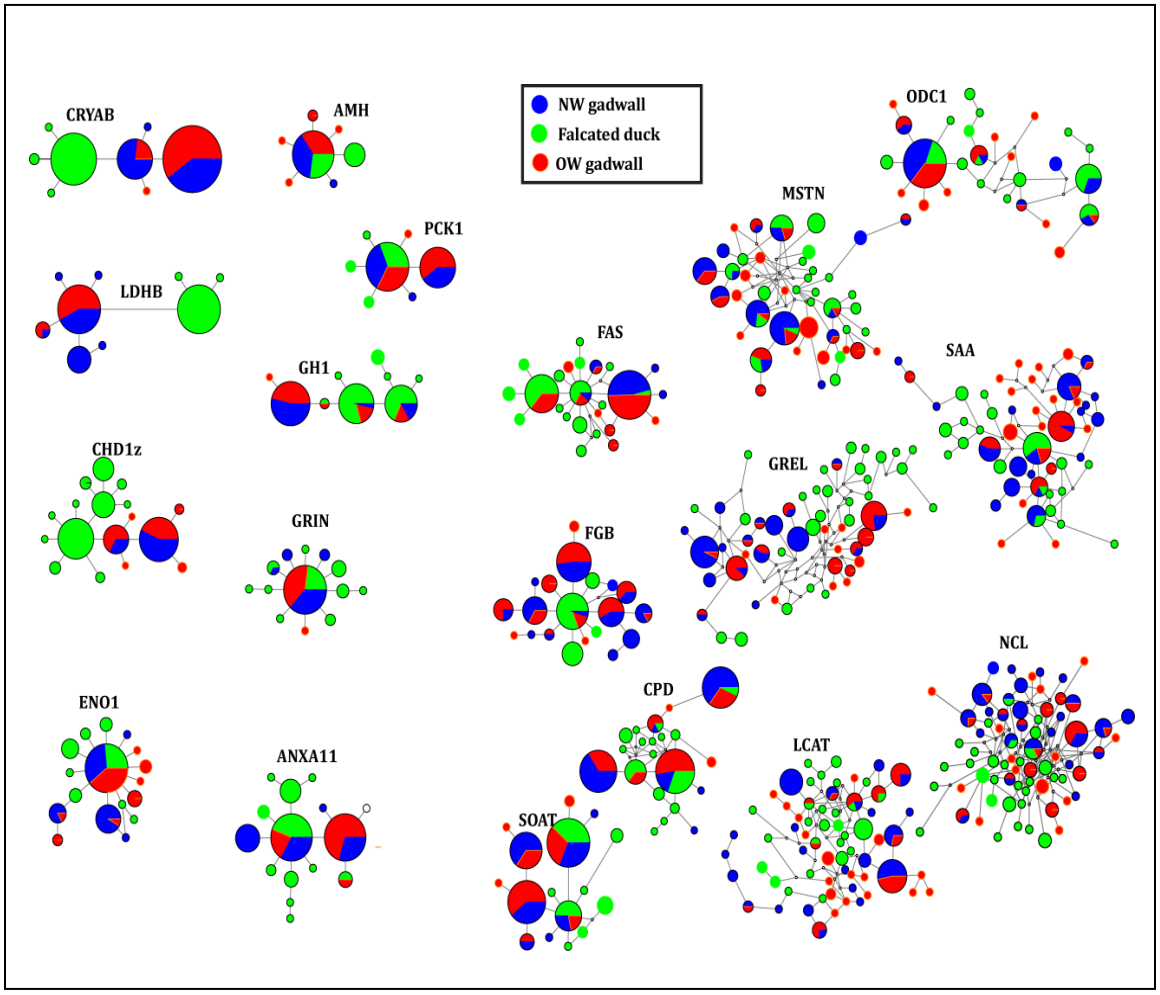

Figure 3.2 Haplotype networks for the nineteen non-coding loci. Each circle represents a different allele and the area of each is proportional to allele frequencies. Branch lengths between alleles are proportional to the number of mutations.

Table 3.1 Characteristics of the nineteen non-coding loci sequenced in the gadwalls and falcated ducks. [1] Locus abbreviations follow standards put forth by the chicken Gene nomenclature Committee (). [2] Chromosome location within the chicken genome and the zebrafinch genome, respectively. ?=unknown. *p<0.05

locus	Abbreviations[1]	Location[2]	Introns #	Length (bp)
Chromo-helicase-DNA binding protein gene 1	CHD1Z	Z/Z	19	272
Lactate dehyrogenase 1	LDHB	1/1A	3	470
S-acyl fatty acid synthase thioesterase	FAST	2/2	2	305
Ornithine decarboxylase	ODC1	3/3	5	276
Fibrinogen beta chain	FGB	4/4	7	350
Serum amyloid A	SAA	5/5	2	311
Annexin A11	ANXA11	6/6	5	191
Myostatin	MSTN	7/7	2	238
Sterol O-acyltransferase	SOAT1	8/?	12	346
Nucleolin	NCL	9/9	12	262
Lecithin-cholesterol acyltransferase	LCAT	?/11	2	200
Preproghrelin	GHRL	12/?	3	332
Glutamate receptor,ionotropic,N-methyl D aspirate I	GRIN1	17/17	11	256
Carboxypeptidase D	CPD	19/19	9	161
Phosphenolpyruvate carboxykinase	PCK1	20/20	9	169
Alpha enolase 1	ENO1	21/21	8	175
Alpha-B crystallin	CRYAB	24/24	1	276
Growth hormone 1	GH1	27/?	3	363
Splicing factor 3A subunit 2	Sf3A2	28/?	8	268

Table 3.2 Locus specific estimates of nucleotide diversity (π), genetic differentiation (Φ_{ST}) and Tajima's D in each population of falcated duck (FD), old world gadwalls (OW), and new world gadwalls (NW).

	π			Φ_{ST}			Tajimas's D		
Locus	π (FD)	π(OW)	π(NW)	$\Phi_{ST/FD-OW}$	$\Phi_{ST/FD-NW}$	$\Phi_{ST/OW-NW}$	D (FD)	D(OW)	D(NW)
GHRL	0.0251	0.0223	0.0194	0.195	0.251	0.184	0.905	0.028	0.338
LCAT	0.0242	0.0239	0.0243	0.098	0.094	0.022	-0.052	1.052	0.640
MSTN	0.0241	0.0222	0.0190	0.049	0.095	0.027	0.066	0.254	-0.375
ODC1	0.0232	0.0142	0.0129	0.135	0.109	0.009	1.147	-1.152	-0.001
NCL	0.0180	0.0204	0.0198	0.067	0.112	0.033	-0.546	-0.099	1.019
CPD	0.0148	0.0187	0.0226	0.057	0.18	0.049	-0.462	0.577	1.072
SAA	0.0142	0.0179	0.0150	0.227	0.156	0.057	0.790	-0.266	-0.004
SOAT1	0.0085	0.0072	0.0072	0.141	0.14	-0.014	-0.056	0.337	1.095
FAST	0.0063	0.0028	0.0025	0.372	0.31	0.03	-1.592*	-1.599*	-0.723
ANXA11	0.0047	0.0034	0.0051	0.389	0.194	0.152	-1.355	-0.056	-0.277
ENO1	0.0045	0.0041	0.0068	0.035	0.116	0.088	-1.386	-1.776*	-0.280
CHD1Z	0.0045	0.0022	0.0006	0.647	0.719	0.038	-1.274	-1.101	-0.238
GRIN1	0.0032	0.0001	0.0007	0.039	0.032	0.013	-1.868*	-1.102*	-1.459*
GH1	0.002	0.0022	0.0014	0.617	0.702	0.005	-0.330	-0.190	-0.469
FGB	0.0021	0.0073	0.0080	0.148	0.193	0.036	0.059	-0.333	-0.282
PCK1	0.0015	0.0026	0.0023	0.172	0.124	-0.009	-1.412	0.019	-0.224
Sf3A2	0.0012	0.0007	0.0001	0.079	0.162	0.081	0.623	-1.764*	-1.102*
CRYAB	0.0007	0.0010	0.0019	0.917	0.861	0.147	-1.764*	-0.642	0.362
LDHB	0.0002	0.0002	0.0011	0.965	0.901	0.142	-1.482*	-0.650	-1.267

Demographic History, Migration, and Divergence

The three-population model with all migration parameters in IMa2 showed a finite posterior distribution for most of the demographic parameters (Fig.3.3, Table 3.3). In this model, θ_{FD} was the largest among all θ parameters (θ_{FD} = 2.1, 95% HPD= 1.6–2.9). There was no overlap in the 95% HPD between θ_{FD} and θ_{NW}, which had the smallest population size (θ_{NW} = 0.61, 95% HPD = 0.26–0.93). On the other hand, θ_{OW} was intermediate (θ_{OW} =1.10, 95% HPD = 0.63–2.06) with 95% HPDs that

overlapped both θ_{FD} and θ_{NW}. The ancestral population of gadwalls (θ_{A0}) was smaller than either NW or OW gadwall population (θ_{A0} = 0.35, 95% HPD= 0.01–4.6), whereas the ancestral population of gadwall and falcated duck (θ_{A1} = 0.66, 95% HPD= 0.14–1.40) was similar to that of NW gadwalls but smaller than OW gadwall and falcated duck. Thus, the model suggested population expansions for all three populations following divergence. Similar values of θ were obtained from both the secondary-contact and the ancestral migration models (Table 3.4 & 3.5; Fig. 3.3). However, analyses of population size changes in falcated ducks obtained from LAMARC were consistent with a stable population size (g = -0.976, 95% CI = -8.8–37.2).

In the full migration model, the rates of introgression from OW and NW gadwall populations into the falcated duck (forward in time) did not vary considerably and peaked at the lowest value of $\leq$0.025 migrants per generation (95% HPD = 0–0.62 and 0–0.57, respectively; Table 3.3, Fig. 3.4a). Thus, the model was consistent with little to no introgression from gadwalls into falcated ducks. Similarly, the introgression rate from falcated ducks into OW gadwalls peaked near zero ($M_{FD \rightarrow OW} \leq$0.025 migrants per generation, 95% HPD = 0–0.82). However, gene flow from falcated ducks into NW gadwalls was low, but non-zero ($M_{FD \rightarrow NW}$ = 0.78 migrants per generation, 95% HPD = 0.07–1.82). Likewise, estimates of gene flow between the ancestral populations in the full model suggested asymmetrical gene flow with higher migration rates from the falcated duck into the ancestral gadwall population, although confidence intervals were large and I could not reject the possibility of no gene flow ($M_{FD \rightarrow A}$ = 1.6 migrants per generation, HPD 95% = 0–50;

$M_{A \rightarrow FD}$ = 0.025 migrants per generation, HPD 95% = 0–46). The model also supported asymmetrical gene flow between the two gadwall populations with higher gene flow into the OW population ($M_{NW \rightarrow OW}$ = 20 migrants per generation, HPD 95% = 0–32; $M_{OW \rightarrow NW}$ = 2.0 migrants per generation, HPD 95% = 0–35); however the posterior distribution was bimodal in both directions and the minor peak for $M_{NW \rightarrow OW}$ was consistent with no gene flow.

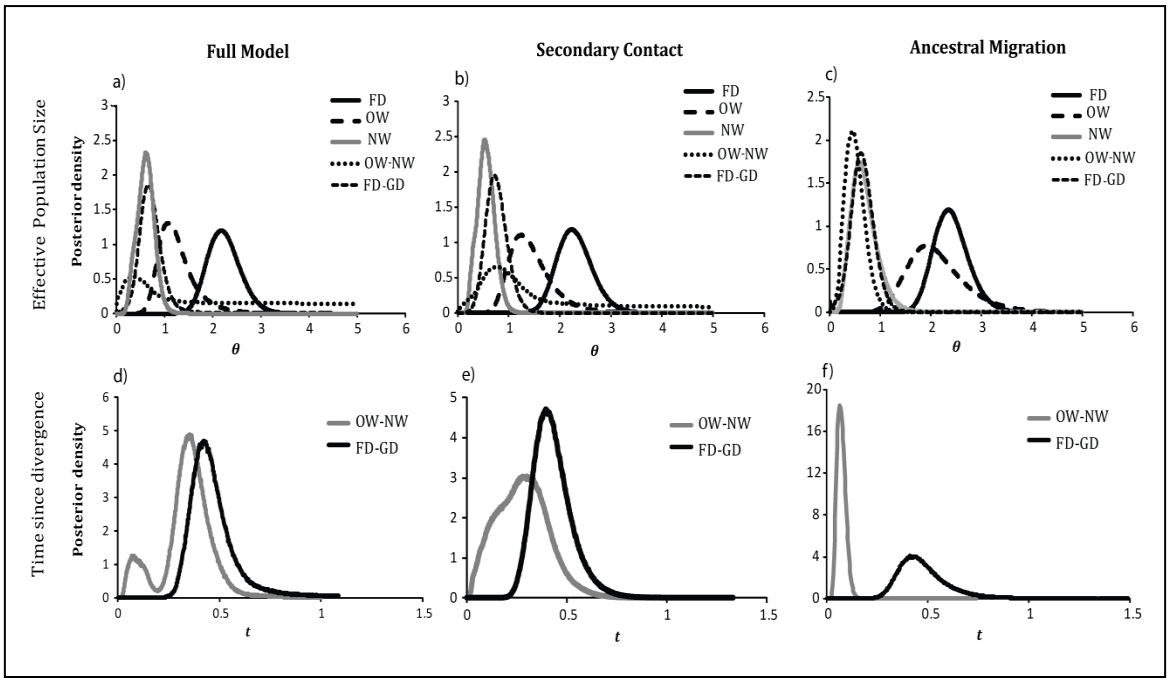

Figure 3.3 Posterior distributions of demographic parameters estimated in IMa2 (scaled to the neutral mutation rate) estimated under three migration models: full migration (a,d), secondary contact (b,e), and ancestral migration (c,f,); a,b,c) effective population sizes of the falcated duck, OW gadwall, NW gadwall and ancestral populations; def) Time since divergence between the falcated duck and gadwall and between OW and NW gadwalls.

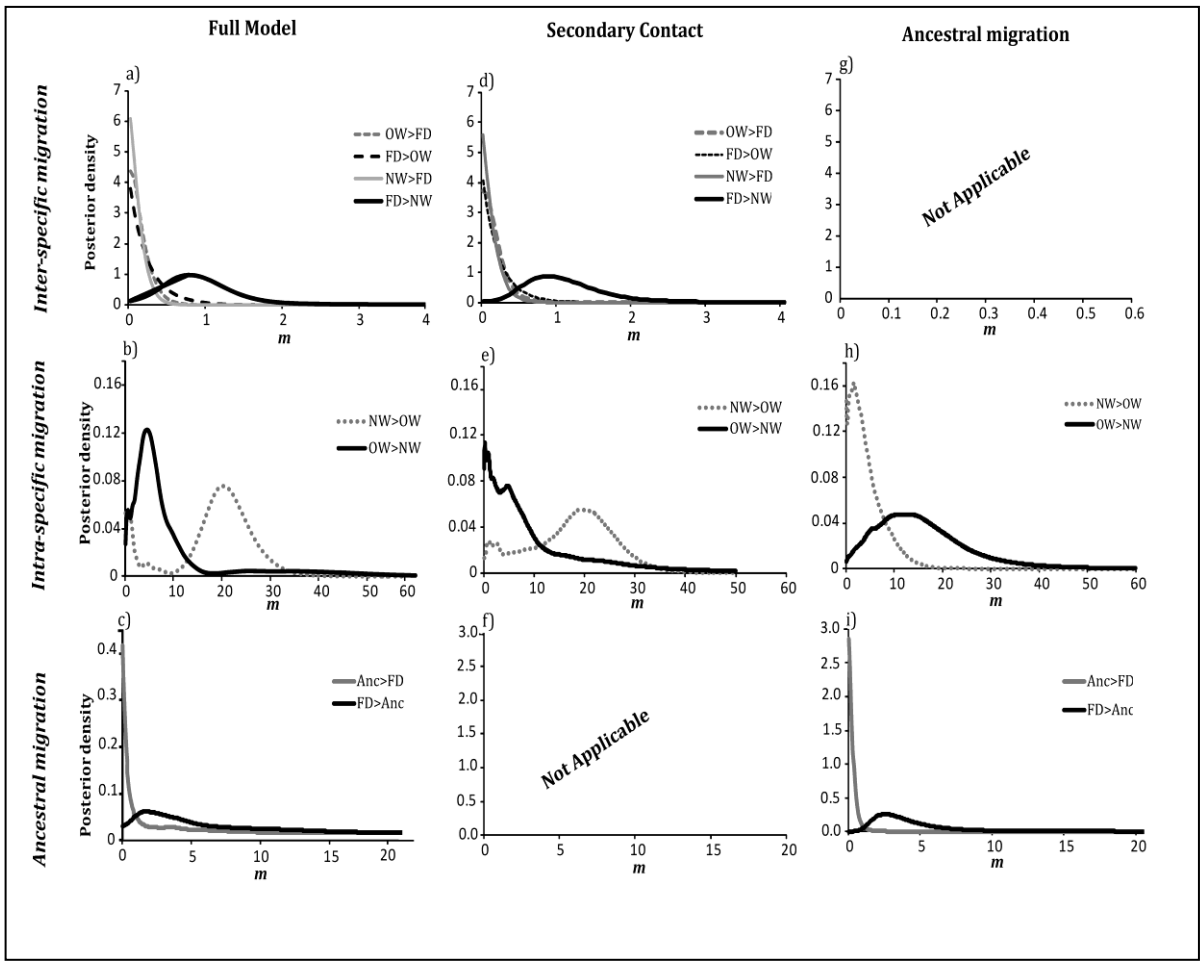

Figure 3.4 Posterior distributions of migration rates estimated in IMa2 in three migration models: full migration, secondary contact, and ancestral migration; a,d,g) interspecific migration rates between falcated duck and gadwall populations; b,e,h) migration estimates between the OW and NW gadwall populations; c,f,i) migration estimates between the falcated duck and ancestral population.

These estimates of migration rates were similar in the model of recent, secondary contact (no ancestral gene flow; Table 3.4, Fig.3. 4d, e, f). Moreover, the posterior distribution of $M_{NW \rightarrow OW}$ was bimodal, but the minor peak was much smaller than in the full model. . In contrast, all the estimates of migration rates had unimodal distributions in the ancestral migration model (Fig.3.4g, h, i, Table 3.5). There was

clear evidence of gene flow from the falcated duck into the ancestral gadwall population ($M_{FD \rightarrow A1}$ = 2.6 migrants per generation, HPD 95% = 0.69–9.2), but the posterior distribution was most consistent with zero gene flow in the opposite direction ($M_{A1 \rightarrow FD}$= 0.025 migrants per generation, HPD 95% = 0–10). Similarly, the model estimated asymmetrical gene flow between the gadwall populations with greater introgression from the OW into the NW gadwall population ($M_{NW \rightarrow OW}$ = 1.7 migrants per generation, HPD 95% = 0–11; $M_{OW \rightarrow NW}$ = 11 migrants per generation, HPD 95% = 0.09–34), which was the reverse direction compared to the full migration model and the secondary-contact model.

The estimates of time since divergence between the falcated duck and ancestral gadwall and the two gadwall populations peaked at different points in the full model (Fig. 3. 3, Table 3.3). The model supported a deep divergence between the gadwall and falcated duck (t_1= 0.42, HPD 95% = 0.25-1.6), but only a slightly more recent divergence between OW and NW gadwall (t_0= 0.35, HPD 95% = 0.03–0.55). However, the posterior distribution of divergence time between OW and NW gadwalls was bimodal with a minor peak that was substantially more recent. In the secondary-contact model, t_1 was similar to the full model, but t_0 showed a broad posterior distribution that encompassed both peaks from the full model (Fig. 3.2b; Table 3.4). In the ancestral migration model, t_0 was more recent and consistent with the minor peak in the full model (t_0= 0.06, HPD 95% = 0.02–0.11), whereas t_1 was similar to the previous two models. Unlike the previous models, there was no overlap in the two divergence time estimates (Fig. 3.3c, Table 3.5), and t_0 was similar to the estimate obtained from the two-population model examined in Peters *et al*. (2012c).

Table 3.3 Demographic parameters estimated from the three population isolation-with-migration model in the IMa2

***t*, Time since divergence**			**θ = Effective population size $\theta=4Ne\mu$**			
Falcated duck & Gadwalls	OW & NW gadwalls	Falcated duck, θ_{FD}	OW gadwall θ_{OW}	NW gadwall θ_{NW}	Ancestral (OW-NW) θ_{A1}	Ancestral (Falcated-gadwall) θ_{A2}
0.42 (0.25-1.59)	0.35 (0.03-0.55)	2.18 (1.60-2.91)	1.10 (0.62-2.06)	0.61 (0.26-0.93)	0.35 (0.01-4.65)	0.66 (0.14-1.40)

Introgression, M (forward in coalescence)							
$\mathbf{M_{OW \rightarrow FD}}$	$M_{FD \rightarrow OW}$	$M_{NW \rightarrow FD}$	$M_{FD \rightarrow NW}$	$M_{NW \rightarrow OW}$	$M_{OW \rightarrow NW}$	$M_{A1 \rightarrow FD}$	$M_{FD \rightarrow A1}$
0.02 (0-0.62)	0.02 (0-0.82)	0.02 (0-0.57)	0.77 (0.07-1.82)	20.02 (0-32.52)	1.97 (0-35.73)	0.02 (0-46.42)	1.67 (0-50)

Table 3.4: Demographic parameters estimated from the three -population model with recent migration in the IMa2.

***t*, Time since divergence**			**θ = Effective population size $\theta=4Ne\mu$**			
Falcated duck & Gadwalls	OW & NW gadwalls	Falcated duck, θ_{FD}	OW gadwall θ_{OW}	NW gadwall θ_{NW}	Ancestral (OW-NW) θ_{A1}	Ancestral (Falcated-gadwall) θ_{A2}
0.39 (0.25-0.57)	0.11 (0.03-0.45)	2.22 (1.64-2.98)	1.24 (0.71-2.32)	0.52 (0.23-0.87)	0.75 (0.06-4.47)	0.71 (0.36-1.18)

Introgression, M (forward in coalescence)							
$\mathbf{M_{OW \rightarrow FD}}$	$M_{FD \rightarrow OW}$	$M_{NW \rightarrow FD}$	$M_{FD \rightarrow NW}$	$M_{NW \rightarrow OW}$	$M_{OW \rightarrow NW}$	$M_{FD \rightarrow A1}$	$M_{A1 \rightarrow FD}$
0.02 (0-0.67)	0.02 (0-0.77)	0.02 (0-0.62)	0.87 (0.22-2.32)	20 (0-30)	0.17 (0-32)	-	-

Table 3.5 Demographic parameters estimated from the three -population model with ancestral migration in the IMa2.

***t*, Time since divergence**			**θ = Effective population size θ=4Neμ**			
Falcated duck & Gadwalls	OW & NW gadwalls	Falcated duck θ_{FD}	OW gadwall θ_{OW}	NW gadwall θ_{NW}	Ancestral (OW-NW) θ_{A1}	Ancestral (Falcated-gadwall) θ_{A2}
0.409 (0.26-0.73)	0.06 (0.02-0.11)	2.35 (1.75-3.10)	1.92 (1.11-3.33)	0.61 (0.24-1.19)	0.45 (0.15-0.89)	0.61 (0.20-1.11)
Introgression, M (back in coalescence)						
$\mathbf{M_{FD \rightarrow OW}}$	$M_{OW \rightarrow FD}$	$M_{FD \rightarrow NW}$	$M_{NW \rightarrow FD}$	$M_{NW \rightarrow OW}$	$M_{OW \rightarrow NW}$	$M_{A1 \rightarrow FD}$	$M_{FD \rightarrow A1}$
-	-	-	-	1.71 (0-11.07)	11.13 (0.09-34.47)	0.03 (0-9.99)	2.55 (0.69-9.21)

Simulated Models of Population History

To test the role of introgression in among-locus heterogeneity, I simulated DNA sequences using the parameters estimated from the three models of demographic history and selective neutrality. Simulations under all the three models under-predicted mean π within each population and mean Φ_{ST} between the falcated duck and each gadwall population (Fig. 3.5). Furthermore, empirical values of π and Φ_{ST} were consistently within the 95% confidence intervals of the simulated values under the full migration model only. Mean π was significantly higher than expected for all three populations under the secondary-contact and ancestral-migration models, and Φ_{ST} was significantly greater than expected under secondary contact. In contrast, there was higher-than-expected heterogeneity (coefficients of variation) in the data

for all parameters, except Φ_{ST} between OW and NW gadwalls, for all three migration models (Fig. 3.5c, d).

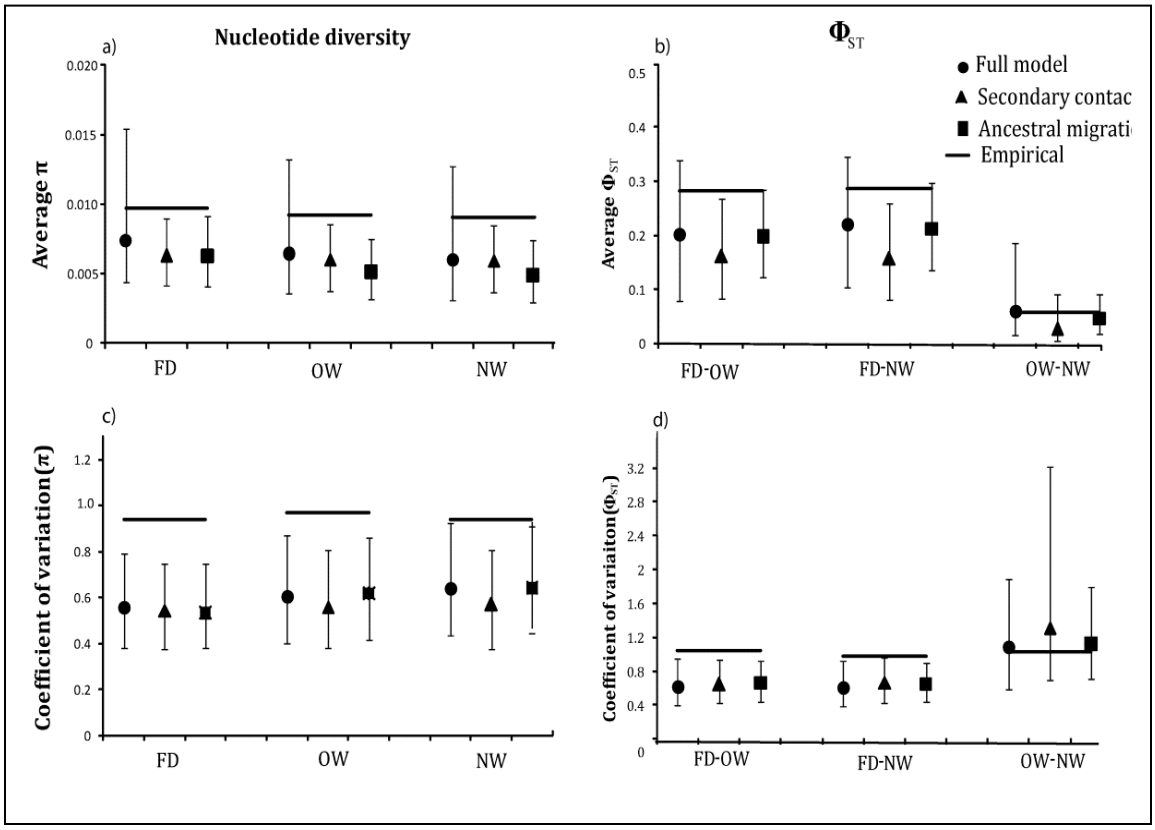

Figure 3.5 Empirical and simulated values of mean a) nucleotide diversity for the nineteen locus data for each population b) Φ_{ST} between each population pair for three migration models. Empirical and simulated values of coefficients of variation for c) nucleotide diversity in falcated ducks, OW gadwall, and NW gadwall, and d) Φ_{ST} between each population pair under three migration models. Black circles, triangles and squares represent the simulated values for the full migration model, secondary contact, and ancestral migration model, respectively; the horizontal bars show the empirical values.

Locus-specific goodness-of-fit tests revealed that 13 of the 19 loci had either significantly greater (*GHRL, MSTN, LCAT, ODC1 NCL, SAA* and *CPD*) or lower diversity (*CRYAB, Sf3A2, FGB, LDH1, GRIN, CHD1z*) than expected for at least one

population (Fig. 3.6). Nucleotide diversity for four loci (*GHRL*, *LCAT*, *LDHB* and *MSTN*) consistently differed from expectations under the neutral models in both taxa and all three models (Fig. 3.6). Similarly, locus-specific tests for Φ_{ST} between falcated duck and gadwall revealed four loci (*CRYAB*, *CHD1z*, *GH1* and *LDHB*) in which the empirical levels of differentiation deviated significantly from the simulated values (Fig. 3.7). At all four loci, the empirical values of Φ_{ST} were greater than expected for both population pairs. However, empirical values of Φ_{ST} between the two gadwall populations were within the simulated values for all 19 loci.

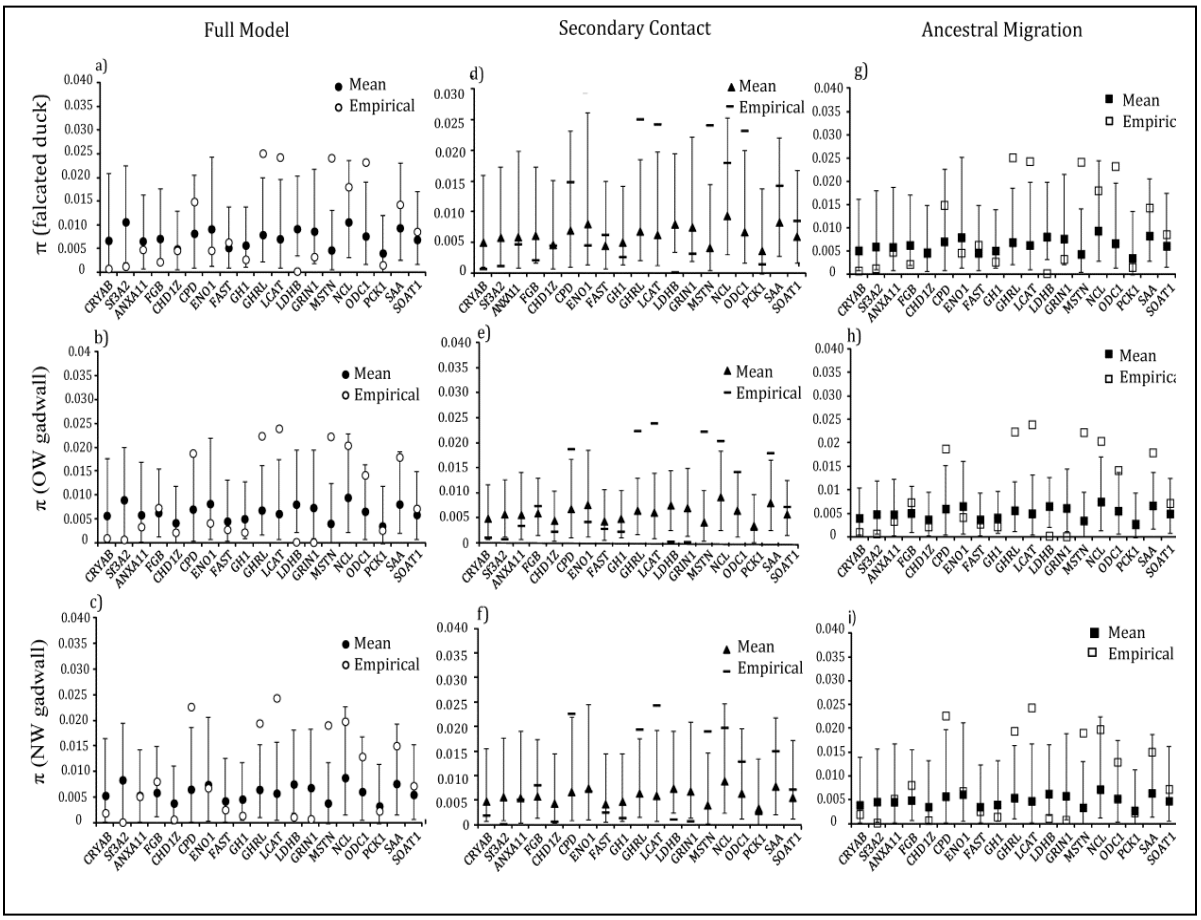

Figure 3.6 Locus-specific goodness of fit tests for nucleotide diversity in the falcated duck(FD), OW gadwall, and NW gadwall under three migration models: full migration model (a,b,c), secondary contact (d,e,f), and ancestral migration (g,h,i). Open symbols mark the empirical data; filled symbols mark the expected values (and the 95% confidence interval) under each model.

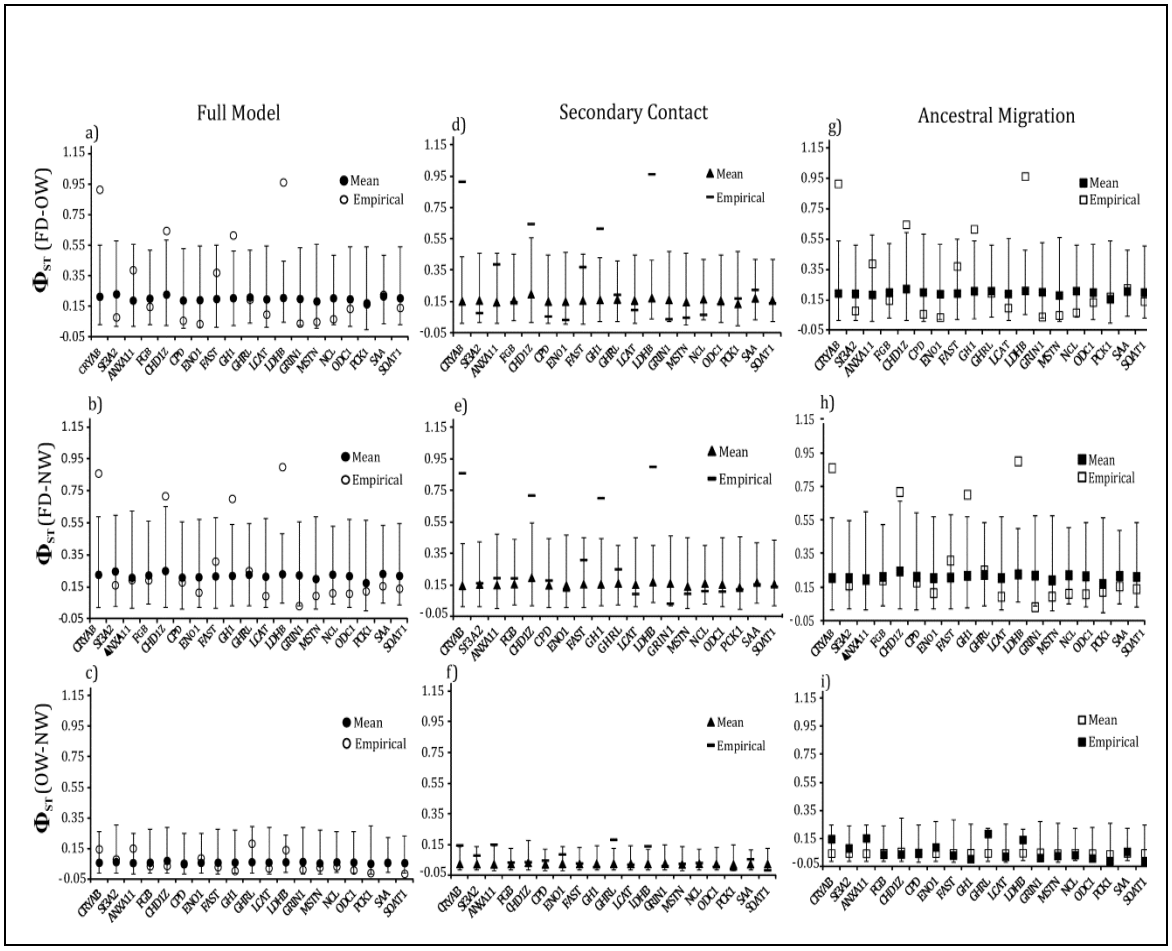

Figure 3.7 Locus-specific goodness-of-fit tests for mean Φ_{ST} between each population pair under the full migration model (a,b,c), secondary contact (d,e,f), and ancestral migration (g,h,i). Open symbols mark the empirical data; filled symbols mark the expected values (and the 95% confidence interval) under each model.

DISCUSSION

Among-locus heterogeneity in genetic diversity can be a function of a complex demography, selection or a combination of these. Peters *et al*. (2012c) found that either introgression or selection could explain the observed heterogeneity in gadwall. However, their estimates of interspecific gene flow were obtained from only three loci, the mtDNA control region, CHD1Z, and LDHB (Peters *et al*., 2007). Using this more comprehensive data set for falcated ducks, I was able to reject introgression as causing the observed among-locus heterogeneity. Furthermore, my results revealed lower-than-expected nucleotide diversity for LDHB and higher-than-expected differentiation at both LDHB and CHD1Z, which combined with regular selective sweeps in mtDNA (Ballard and Whitlock, 2003; Hurst and Jiggins, 2005; Galtier *et al*., 2009), likely misled the results from Peters *et al*. (2007, 2012c). In addition, I found that nucleotide diversity in the falcated duck also varied over 100-fold among the 19 loci, yet I found no evidence of DNA introgression from gadwall into falcated duck. Furthermore, differentiation between the falcated duck and the gadwall varied more than 20-fold among the sequenced loci. In both cases, heterogeneity was greater than expected under the inferred neutral models. Various lines of evidence suggest the influence of selection, rather than hybridization, as a better explanation for the among-locus heterogeneity observed in both taxa.

Deviations from the Models

Coalescent analyses of nuDNA supported introgression either from falcated duck into allopatric NW gadwall or from falcated ducks into the ancestral gadwall population but not into sympatric OW gadwalls. Previous mtDNA analysis also revealed this direction of gene flow and evidence for ancient introgression into NW gadwalls, although ongoing gene flow could not be rejected in sympatry (Peters *et al*., 2007). In this sense, analyses of mtDNA and nuDNA provide concordant results suggesting that NW gadwalls harbor a significant proportion of falcated duck DNA within their gene pool. Whatever the true scenario might be (ancient or ongoing gene flow), introgression fails to account for among-locus heterogeneity in genetic diversity in the taxa for several reasons.

First, the goodness-of-fit tests revealed a poor fit between empirical data and the neutral models of demographic history under all three migration scenarios. Specifically, the empirical coefficients of variation for diversity and interspecific differentiation failed to fit within the expected simulated values under all three models (Fig. 3.5). Similarly, several loci had values of genetic diversity and differentiation that deviated significantly from the expected values under all three neutral models (Fig. 3.6, 3.7). Secondly, the stochasticity of mutation and drift is unlikely to explain this high heterogeneity as the tested models incorporated the variance in these evolutionary forces. Locus-specific mutation rates estimated from independent data were also included in the simulations, and the uncertainty in coalescent estimates of population-level parameters was also incorporated. Given the

amount of noise included in the simulated models, the deviations from the expected patterns are particularly striking.

Thirdly, genetic variation in nuDNA is expected to reflect the species abundance and distributions (Frankham, 1996; Bazin *et al*., 2006; Mccusker and Bentzen, 2010), and therefore, I expected the smaller population of falcated duck (estimated census size of 90,000 individuals; Cao *et al*., 2008) to have lower diversity than the more abundant Holarctic gadwall (>3,000,000 individuals; Delany and Scott, 2006). However, despite the thirty-times smaller population size of falcated ducks relative to gadwalls, the two species had similar genetic diversity, and the falcated duck had the largest effective population size in all three models. A large historical population size in the falcated duck followed by a population decline could be one possible explanation for the observed deviation. However, analyses of population size changes were consistent with falcated ducks having a stable population size, and there was no evidence of a major population decline. Alternatively, a much smaller ancestral population size for gadwalls followed by a population expansion could explain this deviation, which is supported by coalescent analyses (Peters *et al*. 2008, 2012c). Regardless, this deviation from expectations questions the relative roles of genetic drift, introgression, and selection, all three of which might have contributed to the level of diversity observed in falcated ducks. The large effective population size estimated in the falcated duck relative to the gadwall, the lack of evidence of gene flow into falcated duck from any of the gadwall populations, and a lack of evidence suggesting a major population decline in falcated ducks, suggest that selection might be playing an important role.

Locus-specific deviations

The available evidence suggests that selection is a strong candidate for the cause of at least some of the among-locus heterogeneity in these taxa. I propose two markers (*CRYAB* and *LDHB*) as candidate loci under strong positive selection for their exceptional patterns. Both loci have lower genetic diversity than predicted for both gadwall populations and falcated ducks. *CRYAB* and *LDHB* are also more strongly differentiated between falcated ducks and gadwalls than predicted and among the most differentiated loci between the two gadwall populations. These loci were also more differentiated, relative to other loci, between populations of other species of ducks (*CRYAB* and *LDHB* in green-wing teal *Anas crecca* and CRYAB in common merganser *Mergus merganser*; Peters *et al*., 2012 a, b)].

LDHB, the locus with the lowest diversity and the highest divergence, appeared consistently as an outlier in sixteen of eighteen sets of simulated data under all three models of population history. All three models over-predicted the diversity for this locus in all three populations and under-predicted the divergence between falcated ducks and gadwalls. *LDHB* was the only locus among the nineteen loci that never conformed to neutral expectations. Also, the near star-like pattern of the network topology and a significantly negative Tajima's *D* are consistent with positive selection that may have increased the levels of genetic differentiation among populations (Tajima, 1989; Galtier *et al*., 2000). Functionally, the gene is expressed both in the heart of ducks and in the eyes as lens structural proteins (ε- crystalline; Hendriks, 1988) and there is evidence of adaptive evolutionary changes occurring in the sequence of *LDHB* (Crawford and Powers, 1989; Kraft *et al*., 1994). For

example, the presence or absence of repressor elements in the regulatory sequence of *LDHB* is responsible for the adaptive difference in *LDHB* transcription between northern and southern populations of *Fundulus. heteroclitu* (Schulter, 2000). The high divergence of the locus and its association with adaptive evolution in other species supports the proposed hypothesis of non-neutrality at this locus.

CRYAB, also a low diversity locus with high differentiation, also failed to fit with the expected values under neutrality in several tests. A significant excess of rare polymorphisms and a star like pattern in the haplotype network topology support the possibility of positive selection/selective sweep at this locus. The evolutionary trajectory of this gene, which codes for eye-lens crystallins, varies between mammalian and avian taxa. In contrast to the high conservation of the gene among mammals, only a few blocks of the gene are conserved in birds. For example, the duck *CRYAB* homologues have lost the heat shock response seen in mammalian homologues (Wistow and Grahm, 1995). This partial conservation of gene elements in ducks and the variability in heat-shock response suggest taxon-specific patterns of expression. It is intriguing that both CRYAB and LDHB are expressed in eye-lens crystallins and both deviate from neutral expectations.

Genetic hitchhiking can influence nucleotide diversity of non-coding loci and potentially maintain the high diversity of non-coding regions (several times that of neutral loci), when these regions are in linkage disequilibrium with a coding region under balancing selection (Smith and Haigh, 1974; Orengo and Aguade, 2004). Alternatively, hitchhiking can cause reduced variation in non-coding DNA linked to loci subjected to selective sweeps (Kaplan *et al*., 1989). Thus, elevated or reduced

diversity in non-coding regions might not necessarily be due to selection acting directly on components of the introns, but rather a result of strong linkage to coding regions that are the targets of selection. Hitchhiking depends upon the recombination rate and the distance from the target of selection (Maynard-Smith and Haigh, 1974). However, hitchhiking could be prominent in the non-coding loci with lower recombination rates. The major outlier loci, CRYAB and LDHB, in falcated ducks and gadwalls were both consistent with no intra-locus recombination (Peters *et al*., 2012c). Therefore, selection on the coding regions of these loci coupled with hitchhiking could explain the inferred non-neutrality that was detected.

Kraus *et al*. (2011) suggested that hybridization among more divergent species of *Anas* ducks likely explained the high number of polymorphisms shared among species. Our models do not account for the possibility of gene flow with these additional species, and it is possible that this confounding variable could explain much of the heterogeneity that I observed. In particular, the high diversity loci might reflect broad introgression. However, broad-scale hybridization cannot fully account for the low diversity found at some loci without the combined effects of selection preventing the introgression of alleles at those loci. Thus, complete neutrality is unlikely even under this more complex population history.

Differential Introgression, divergent selection, and demography

Heterogeneity in genetic divergence across the genome of divergent taxa is expected under divergent selection. The counteraction between introgression and divergent selection prevents complete homogenization of genomes when divergent selection

restricts gene flow at some loci. Therefore, these loci can have higher genetic differentiation than neutral loci (Saint-Laurent *et al.*, 2003; Emilianov *et al.*, 2004; McCracken *et al*. 2009; Nosil *et al*., 2009; McCracken and Wilson, 2011). In accordance with the predictions of differential introgression caused by divergent selection, this study detected several outlier loci that exhibited higher genetic differentiation than expected under neutrality (Fig. 3.7). The same outlier loci were observed under all the three models of introgression. Despite evidence of introgression from falcated duck into the NW gadwall population and strong support for zero gene flow between falcated duck and the OW gadwall population, the same four loci were detected as outliers in both comparisons. On the other hand, Φ_{ST} of these loci was consistent with expectations under neutrality between OW and NW gadwalls, suggesting inter-specific selective pressures. In particular, the sex-linked locus CHD1Z was consistently an outlier and empirical data from numerous taxa suggest that the Z-chromosome is often less likely to introgress than autosomal chromosomes (Carling and Brumfield, 2008, 2009; Storchova *et al*., 2010)

I found evidence of higher nuDNA introgression in allopatry, which was consistent with patterns observed in mtDNA (Peters et al., 2007). Infrequent sightings of male falcated ducks in NorthAmerica (http://www.fws.gov/sacramentovalleyrefuges/wo_sightings.html) raises the speculation of ongoing gene flow between falcated duck and NW gadwalls in accordance with Hubb's Principle or Desperation hypothesis, which predicts hybridization when one species is rare in sympatry. Absence of conspecifics and

restricted mate choice in North America could cause these rare Asian visitors to hybridize with the more abundant gadwalls, as has been demonstrated in other species of ducks (McCracken and Wilson, 2011). Alternatively, the introgression of falcated duck genes into NW gadwalls could be explained by ancient introgression (Peters *et al*., 2007). Genetic evidence suggests that the gadwall colonized North America from Eurasia during the late Pleistocene (Peters *et al*., 2008). If a falcated duck or a hybrid was among the original founders, then it could have had a large genetic contribution to the extant gene pool. The observation that some mtDNA haplotypes in NW gadwall were similar to, but not shared with, falcated duck haplotypes is consistent with this scenario (Peters *et al*., 2007). Furthermore, our model that allowed only ancient introgression converged better than the other migration models, suggesting that it might be a more appropriate model. Unfortunately, distinguishing between ancient gene flow and secondary contact can be difficult with genetic data (Becquet and Perzeworski, 2010; Strasberg and Rieseberg, 2010) preventing conclusive tests of these hypotheses.

Conclusion

I conclude that gene flow between falcated ducks and gadwalls fails to explain the heterogeneity in genetic diversity and differentiation under various models of demographic history. Simulating models of introgression under neutrality failed to explain the high empirical diversity for some loci (*GHRL*, *LCAT* and *MSTN*) and lower empirical diversity observed for other loci (*LDHB* and *GRIN1*). Inter-specifically, *CRYAB* and *LDHB* were strong outliers with exceptionally greater Φ_{ST},

and these two loci are among the most structured intraspecifically. I suggest *CRYAB* and *LDHB* as strong candidate loci under positive selection, perhaps resulting from low recombination and high linkage disequilibrium with polymorphisms in coding regions. Selection might also have had a major effect on the other loci, thereby contributing to the strong among-locus heterogeneity in genetic diversity and differentiation.

REFERENCES

Avise J. C. 2000. Phylogeography- The History and Formation of Species. London.

Bachtrog D., K. Thornton, A. Clark, and P. Andolfatto. 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. Evolution 60:292-302.

Ballard J., and M. Whitlock. 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13:729-744.

Bandelt H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37-48.

Bazin E., S. Glémin, and N. Galtier. 2006. Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. Science 312:570-572.

Becquet C., and M. Przeworski. 2009. Learning about Modes of Speciation by Computational Approaches. Evolution 63.

Borge T., M. T. Webster, G. Andersson, and G. P. Saetre. 2005. Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. Genetics 171:1861-1873.

Cao L., M. Barter, and G. Lei. 2008. New Anatidae population estimates for eastern China: Implications for current flyway estimates. Biol. Conserv. 141:2301-2309.

Carling M. D., and R. T. Brumfield. 2009. Speciation in Passerina buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. Mol. Ecol. 18:834-847.

Carling M. D., and R. T. Brumfield. 2008. Haldane's Rule in an Avian System: using Cline Theory and Divergence Population Genetics to Test for Differential Introgression of Mitochondrial, Autosomal, and Sex-Linked Loci Across the Passerina Bunting Hybrid Zone. Evolution 62:2600-2615.

Crawford D. L., and D. A. Powers. 1989. Molecular-Basis of Evolutionary Adaptation at the Lactate Dehydrogenase-B Locus in the Fish Fundulus-Heteroclitus. Proc. Natl. Acad. Sci. U. S. A. 86.

Currat M., M. Ruedi, R. J. Petit, and L. Excoffier. 2008. The hidden side of invasions: Massive introgression by local genes. Evolution 62:1908-1920.

Delany, S. and Scott, D. (2006) *Waterbird Population Estimates*. Fourth Edition. Wageningen: Wetlands International.

Excoffier L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.

Frankham R. 1996. Relationship of genetic variation to population size in wildlife. Conserv. Biol. 10.

Galtier N., F. Depaulis, and N. Barton. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics 155:981-987.

Galtier N., B. Nabholz, S. Glemin, and G. D. D. Hurst. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. 18:4541-4550.

Haddrill P., K. Thornton, B. Charlesworth, and P. Andolfatto. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. 15:790-799.

Hammer M., D. Garrigan, E. Wood, J. Wilder, Z. Mobasher, A. Bigham, J. Krenz, and M. Nachman. 2004. Heterogeneous patterns of variation among multiple human x-linked loci: The possible role of diversity-reducing selection in non-Africans. Genetics 167:1841-1853.

Hendriks W., J. W. M. Mulders, M. A. Bibby, C. Slingsby, H. Bloemendal, and W. W. Dejong. 1988. Duck Lens Epsilon-Crystallin and Lactate Dehydrogenase-B4 are Identical - a Single-Copy Gene-Product with 2 Distinct Functions. Proc. Natl. Acad. Sci. U. S. A. 85.

Hey J. 2010. Isolation with Migration Models for More Than Two Populations. Mol. Biol. Evol. 27:905-920.

Hudson R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337-338.

Hurst G., and F. Jiggins. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited

symbionts. Proceedings of the Royal Society B-Biological Sciences 272:1525-1534.

Johannesen J., B. Johannesen, E. M. Griebeler, I. Baran, M. R. Tunc, A. Kiefer, and M. Veith. 2006. Distortion of symmetrical introgression in a hybrid zone: evidence for locus-specific selection and uni-directional range expansion. J. Evol. Biol. 19:705-716.

Johnsgard, P. A. 1960. Hybridization in the *Anatidae* and its taxonomic implications. Condor 62:25–33.

Kaiser V. B., R. Bergero, and D. Charlesworth. 2011. A new plant sex-linked gene with high sequence diversity and possible introgression of the X copy. Heredity 106:339-347.

Kaplan N. L., R. R. Hudson, and C. H. Langley. 1989. The Hitchhiking Effect Revisited. Genetics 123.

Knowles L., and C. Richards. 2005. Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. Mol. Ecol. 14:4023-4032.

Kraft H. J., C. E. M. Voorter, L. Wintjes, N. H. Lubsen, and J. G. G. Schoenmakers. 1994. The Developmental Expression of Taxon-Specific Crystallins in the Duck Lens. Exp. Eye Res. 58.

Kraus R. H. S., A. Zeddeman, P. van Hooft, D. Sartakov, S. A. Soloviev, R. C. Ydenberg, and H. H. T. Prins. 2011. Evolution and connectivity in the world-

wide migration system of the mallard: Inferences from mitochondrial DNA. Bmc Genetics 12:99.

Kuhner M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768-770.

Kulathinal R. J., L. S. Stevison, and M. A. F. Noor. 2009. The Genomics of Speciation in Drosophila: Diversity, Divergence, and Introgression Estimated Using Low-Coverage Genome Sequencing. Plos Genetics 5:e1000550.

Maynard Smith, J., and J. Haigh, 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23**:** 23–35.

McCracken K. G., M. Bulgarella, K. P. Johnson, M. K. Kuhner, J. Trucco, T. H. Valqui, R. E. Wilson, and J. L. Peters. 2009. Gene Flow in the Face of Countervailing Selection: Adaptation to High-Altitude Hypoxia in the beta A Hemoglobin Subunit of Yellow-Billed Pintails in the Andes. Mol. Biol. Evol. 26:815-827.

McCracken K. G., and R. E. Wilson. 2011. Gene Flow and Hybridization between Numerically Imbalanced Populations of Two Duck Species in the Falkland Islands. Plos One 6:e23173.

McCusker M. R., and P. Bentzen. 2010. Positive relationships between genetic diversity and abundance in fishes. Mol. Ecol. 19.

Minder A. M., and A. Widmer. 2008. A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. Mol. Ecol. 17:1552-1563.

Morjan C. L., and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. 13:1341-1356.

Nosil P., and J. L. Feder. 2012. Widespread yet heterogeneous genomic divergence. Mol. Ecol. 21:2829-2832.

Nosil P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375-402.

Orengo D. J., and M. Aguade. 2004. Detecting the footprint of positive selection in a European population of *Drosophila melanogaster*: Multilocus pattem of variation and distance to coding regions. Genetics 167.

Peters J. L., and K. E. Omland. 2007. Population structure and mitochondrial polyphyly in North American Gadwalls (*Anas strepera*). Auk 124:444-462.

Peters J. L., Y. N. Zhuravlev, I. Fefelov, E. M. Humphries, and K. E. Omland. 2008. Multilocus phylogeography of a Holarctic duck: Colonization of North America from Eurasia by gadwall (*Anas strepera*). Evolution 62:1469-1483.

Peters J. L., Y. Zhuravlev, I. Fefelov, A. Logie, and K. E. Omland. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution 61:1992-2006.

Peters J. L., K. A. Bolender, and J. M. Pearce. 2012a. Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. Mol. Ecol. 21:3562-3575.

Peters J. L., K. G. Mccracken, C. L. Pruett, S. Rohwer, S. V. Drovetski, Y. N. Zhuravlev, I. Kulikova, D. D. Gibson, and K. Winker. 2012b. A parapatric propensity for breeding precludes the completion of speciation in common teal (*Anas crecca*, sensu lato). Mol. Ecol. 21:4563-4577.

Peters J. L., T. E. Roberts, K. Winker, and K. G. McCracken. 2012c. Heterogeneity in genetic diversity among non-coding loci fails to fit neutral coalescent models of population history. PLoS One 7:e31972.

Pool J. E., and R. Nielsen. 2007. Population size changes reshape genomic patterns of diversity. Evolution 61:3001-3006.

Postma E., and A. J. van Noordwijk. 2005. Gene flow maintains a large genetic difference in clutch size at a small spatial scale. Nature 433:65-68.

Saint-Laurent R., M. Legault, and L. Bernatchez. 2003. Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax Mitchill*). Mol. Ecol. 12:315-330.

Schluter D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372-380.

Schulte P. M., H. C. Glemet, A. A. Fiebig, and D. A. Powers. 2000. Adaptive variation in lactate dehydrogenase-B gene expression: Role of a stress-responsive regulatory element. Proc. Natl. Acad. Sci. U. S. A. 97.

Smith J. M., and J. Haigh. 1974. Hitch-Hiking Effect of a Favorable Gene. Genet. Res. 23.

Stephens M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73:1162-1169.

Stephens M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978-989.

Storchova R., J. Reif, and M. W. Nachman. 2010. Female Heterogamety and Speciation: Reduced Introgression of the Z Chromosome between Two Species of Nightingales. Evolution 64:456-471.

Storz J. F., and J. K. Kelly. 2008. Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: Insights from deer mouse globin genes. Genetics 180:367-379.

Strasburg J. L., and L. H. Rieseberg. 2010. How Robust Are "Isolation with Migration" Analyses to Violations of the IM Model? A Simulation Study. Mol. Biol. Evol. 27.

Strasburg J. L., N. A. Sherman, K. M. Wright, L. C. Moyle, J. H. Willis, and L. H. Rieseberg. 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? Philos. Trans. R. Soc. B-Biol. Sci. 367:364-373.

Sweigart A., and J. Willis. 2003. Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. Evolution 57:2490-2506.

Wistow G., and C. Graham. 1995. The duck gene for alpha-B-crystallin shows evolutionary conservation of discrete promoter elements but lacks heat and osmotic shock response. Biochim. Biophys. Acta 1263:105-113.

Woerner A. E., M. P. Cox, and M. F. Hammer. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics 23:1851-1853.