2014

Phylogenetics, Population Genetics, and Evolution of the Mallard Complex

Philip Lavretsky
Wright State University

PHYLOGENETICS, POPULATION GENETICS, AND EVOLUTION OF THE MALLARD COMPLEX

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

By

Philip Lavretsky

B.S., University Of California-Davis, 2008

2014

WRIGHT STATE UNIVERSITY

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

APRIL 4, 2014

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Philip Lavretsky ENTITLED **Phylogenetics, population genetics, and evolution of the mallard complex** 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

_____

Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

Committee on Final Examination

_____

John O. Stireman III, Ph.D.

_____

Volker Bahn, Ph.D.

_____

Christopher Barton, Ph.D.

_____

Lisle H. Gibbs, Ph.D

ABSTRACT

Lavretsky, Philip Ph.D., Enviornmental Sciences Ph.D. Program, Wright State University, 2014. Phylogenetics, population genetics, and evolution of the mallard complex

Speciation is primarily regarded as an ancestral split that results in two distinct taxonomic units, and proceeds in stages along a continuum from initiation (i.e., population divergence) to completion (i.e., reproductively isolated species). Establishing how and why populations diverge, including the primary mechanisms influencing these events is a major objective for evolutionary scientists. Focusing on incipient forms, researchers attempt to disentangle the antagonistic nature of selection, genetic drift, and gene flow in the speciation process.

In chapter 1, I investigate the phylogenetic relationships of 14 closely related taxa within the mallard complex (*Anas* spp.) that underwent a radiation within the past one million years. Using mitochondrial DNA (mtDNA) and 20 nuclear loci for one to five individuals per taxon, I further examine how recombination and hybridization affect species tree reconstructions. In general, relationships within major clades were robust to treatment of recombination (i.e., ignoring or filtering) and inclusion or exclusion of hybridizing taxa, but branch lengths and posterior support were sensitive to both treatments. Of the 14 taxa, the most confounded relationships were those within the New World (NW) group comprising the sexually dichromatic mallard (*Anas platyrhynchos*)

and the monochromatic American black duck (*A. rubripes*; "black duck"), mottled duck (*A. fulvigula*), and Mexican duck (*A.* [*p.*] *diazi*). Finally, I address discordance between nuclear, morphometric, and mitochondrial trees, particularly with regard to the placement of the Hawaiian duck (*A. wyvilliana*), Philippine duck (*A. luzonica*), and two spot-billed ducks (*A. zonorhyncha* and *A. poecilorhyncha*) and discuss how alternative modes of speciation (i.e., hybrid speciation) may lead to variance in these relationships.

In Chapter 2, I attempt to disentangle the evolutionary relationships of the New World (NW) group using mtDNA and 17 nuclear loci for a larger per taxon sample size (24-25 individuals per taxon). In general, whereas both Florida and Gulf Coast mottled ducks were differentiated from one another and from the other taxa (mean Φ_{ST} = 0.024 – 0.064), mallards, American black ducks, and Mexican duck were not significantly differentiated among nuclear markers (mean Φ_{ST} < 0.020). Using coalescent methods to estimate rates of gene flow between mallards and each of the monochromatic taxa generally supported hybridization, but I could not reject complete isolation for any pairwise comparison. Furthermore, species tree reconstructions revealed that phylogenetic relationships were sensitive to stochastic sampling of individuals likely due to incomplete lineage sorting or hybridization. I conclude that members of the NW Mallard group appear to be adaptive incipient morphs, and that future work should focus on genomic regions under selection to better understand the stage and process of speciation in this group.

In Chapter 3, I use restriction site associated DNA (RAD) sequencing methods to generate a pseudorandom sampling of 3,563 autosomal and 172 sex-linked (Z chromosome) markers scattered across the genome to more rigorously test the mechanism of speciation between Mexican ducks (N = 105 individuals from six Mexican states and two US states) and mallards (N = 17). Specifically, I aim to determine the stage of speciation and whether speciation has been driven by few loci with large effects versus many loci with small effects, plumage associated differentiation, or genetic drift. Marker comparisons between mallards and Mexican ducks revealed strong discordance among autosomal (Φ_{ST} = 0.014), sex-linked (mean Φ_{ST} = 0.091), and mtDNA (Φ_{ST} = 0.12) markers. In general, divergence at autosomal loci followed a stepping stone model, with a gradual transition in genotypic frequencies from North to South. In contrast, Z-linked markers followed an island model of divergence, with a sharp transition in genotypic frequencies at the geographic boundary between mallards and Mexican ducks. In contrast, both autosomal (mean Φ_{ST} = 0.012) and Z-linked markers (mean Φ_{ST} = 0.018) were tightly correlated among Mexican duck sampling groups. These results suggest that, whereas genetic drift is likely influencing structure among Mexican duck populations and between Mexican ducks and mallards at autosomal loci, selection is likely influencing Z-chromosome structure between Mexican ducks and mallards. The latter finding is consistent with the evolution of post-mating isolation between Mexican ducks and mallards. Finally, I report that contemporary hybridization with mallards is likely limited to the northern edge of the Mexican duck's range, and that those from inland Mexico

appear to be “pure” and follow an isolation-by-distance model of divergence. In conclusion, these results suggest that mallards and Mexican ducks are at the earliest stages of parapatric divergence with the Z chromosome at a later stage – relative to autosomal chromosomes – of divergence, which is being driven by selection on few loci with large effects.

In Chapter 4, I test another mechanism of speciation – whether the Hawaiian duck evolved via hybrid speciation. Following from the results of Chapter 1, where I presented compelling evidence of mitochondrial-nuclear-morphological discord in the phylogenetic placement of this species, I sequenced a larger sample size of Hawaiian ducks (N = 15 individuals) and its putative parental species, the Laysan duck (*A. laysanensis*; N = 21 individuals) and mallard (N = 25 individuals). I demonstrated that the Hawaiian duck’s genome was a mosaic of mallard (59%) and Laysan duck (41%) polymorphisms. Moreover, gene flow estimates revealed significant non-zero gene flow from the Laysan duck into the Hawaiian duck under a mtDNA-like topology (Hawaiian sister to mallard) or from the mallard into the Hawaiian-Laysan duck ancestor under a nuDNA-like topology (Hawaiian sister to Laysan). Thus, regardless of the tree topology used, gene flow from the non-sister species is necessary to explain extant genetic diversity in Hawaiian ducks, further supporting a genomic mosaic. This work is one of few well-supported cases for hybrid speciation in homoploid systems, and highlights the potential for such events on island systems where the hybrid descendants can become geographically isolated from the parental species.

In Chapters 1 and 4, I found no nuclear variation in Laysan ducks, which is a critically endangered species. Consequently, in Chapter 5, I developed a PCR-based protocol to examine diversity within the Major Histocompatibility Complex (MHC) I gene in Laysan ducks. Particular attention has been given to MHC genes due to their direct correlation to an individual's immunity. The haplotype-specific primers allowed for direct genotyping after gel electrophoresis based on the presence/absence of their respective amplicons. Using the developed techniques, a total of eight unique haplotypes were isolated and assayed across 21 Laysan duck individuals from Laysan Island (N = 10) and Midway Atoll (N = 11). The protocol provides a simple, cost-effective method for isolating haplotypes and monitoring existing MHC variation in Laysan ducks that can be implemented in admixture schemes within captive breeding programs to maximize heterogeneity prior to reintroduction.

In conclusion, divergence and speciation within the mallard complex has been driven by a number of mechanisms, including allopatric divergence, parapatric divergence, and hybrid speciation. These results demonstrate the value of multi-taxa, multi-marker comparisons in resolving complex evolutionary relationships. Furthermore, each chapter builds on previous chapters, illustrating the utility of addressing speciation from macroevolutionary scales (e.g., phylogenetics), which generate testable hypotheses, to progressively more microevolutionary scales for testing those hypotheses. Given their incipient stage and evolutionary heterogeneity, the mallard complex is an excellent

system for studying the effects of various evolutionary mechanisms and demographies in the speciation process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I want to express my sincere appreciation to my advisor Dr. Jeffrey L. Peters, whose guidance, sense of humor, and continuous correspondence was instrumental in my success. Special thanks to all my committee members; John O. Stireman, Volker Bahn, Christopher Barton, and Lisle Gibbs for their encouragement and support. Moreover, I am grateful for the various collaborators and funding opportunities that made this research possible. Finally, I thank all my friends and family for their support throughout this journey, and particularly for my loving wife, Lauren Lavretsky for listening to stories about ducks, and to my collecting partner, Teal, for fetching many of the samples.

CHAPTER I. PHYLOGENETICS OF A RECENT RADIATION IN THE MALLARDS AND ALLIES (AVES: *ANAS*): INFERENCES FROM A GENOMIC TRANSECT AND THE MULTISPECIES COALESCENT

Abstract – Reconstructing species trees by incorporating information from many independent gene trees reduces the confounding influence of stochastic lineage sorting. Such analyses are particularly important for taxa that share polymorphisms due to incomplete lineage sorting or introgressive hybridization. I investigated phylogenetic relationships among 14 closely related taxa from the mallard (*Anas* spp.) complex using the multispecies coalescent and 20 nuclear loci sampled from a genomic transect. I also examined how treating recombining loci and hybridizing species influences results by partitioning the data using various protocols. In general, topologies were similar among the various species trees, with major clades consistently composed of the same taxa. However, relationships among these clades and among taxa within clades changed among partitioned data sets. Posterior support generally decreased when filtering for recombination, whereas excluding mallards (*Anas platyrhynchos*) increased posterior support for taxa known to hybridize with them. Furthermore, branch lengths decreased substantially for recombination-filtered data. Finally, concordance between nuclear and morphometric topologies conflicted with those in the mitochondrial tree, particularly with regard to the placement of the Hawaiian duck (*A. wyvilliana*), Philippine duck (*A. luzonica*), and two spot-billed ducks (*A. zonorhyncha* and *A. poecilorhyncha*). These results demonstrate the importance of maximizing sequence length and taxon sampling when inferring taxonomic relationships that are confounded by extensive allele sharing.

INTRODUCTION

Reconstructing phylogenetic relationships for recently diverged taxa can be confounded by allele sharing resulting from a recent shared ancestry (i.e., incomplete lineage sorting; Pamilo and Nei, 1988) or introgressive hybridization (Avise, 2000; Grant and Grant, 1992; Price and Bouvier, 2002). These factors result in taxa having heterogeneous genomes and discordant evolutionary histories among loci (Carstens and Knowles, 2007). Consequently, any single gene tree is unlikely to reflect the species tree (Degnan and Rosenberg, 2006). Advances in computational methods that incorporate information across numerous gene trees (Drummond and Rambaut, 2007; Kubatko *et al.*, 2009; Liu, 2008) offer researchers the tools for reconstructing species trees derived from multi-locus, genome wide datasets (Carstens and Knowles, 2007; Jacobsen and Omland, 2011b; Knowles, 2009). Although the ability of such programs to resolve relationships that are complicated by allele sharing has been tested with simulated data sets (Chung and Ané, 2011; Lanier and Knowles, 2012; Leaché and Rannala, 2011), few empirical investigations into the sensitivity of species tree reconstructions to recombination and hybridization have been conducted. The objectives of this study are to reconstruct phylogenetic relationships of 14 closely related taxa within the mallard complex (*Anas platyrhynchos* and allies) utilizing multi-locus coalescent methods, while examining the sensitivity of results to various approaches for handling recombination and hybridizing species.

Many phylogenetic and population genetic methods require making an assumption of no intralocus recombination. Doing so, however, often requires that DNA sequences are truncated, potentially resulting in a loss of information and decreased

phylogenetic resolution. Although the effects of recombination should be considered (Edwards, 2009; Rieseberg *et al.*, 2000), simulations by Lanier and Knowles (2012) show that recombination may have little or no effect on phylogenetic inferences, and instead concluded that the negative effects introduced by ignoring recombination were offset by increasing sampling effort of loci and/or individuals. Topological comparisons between empirical datasets can be used to examine the influence of filtering for recombination, especially when comparing results to simulated data. In this study I compare trees that are reconstructed with entire gene reads (i.e., "ignoring recombination") or with datasets where loci have been truncated to be consistent with no recombination (i.e., "recombination-filtered"). Based on simulated datasets (Lanier and Knowles, 2012) I expect few changes in the relationships among taxa but a decrease in the posterior support, particularly for the deepest nodes, when data is filtered for recombination.

Discordance among loci resulting from hybridization has been an important issue in avian phylogenetics (Degnan and Rosenberg, 2009; Jacobsen and Omland, 2011b; Weckstein *et al.*, 2001). The high proportion of shared polymorphisms among species has been attributed to dispersal ability (Greenwood, 1980), chromosomal stasis (Ellegren, 2010), and relatively low levels of reinforcement (Grant and Grant, 1997) in birds. Among avian orders, waterfowl (Anseriformes) experience among the highest rates of hybridization (Johnsgard, 1960; Lijtmaer *et al.*, 2003; Livezey, 1986), with 30-40% of species being capable of interbreeding (Grant and Grant, 1992) and about 20% producing viable hybrids (Scherer and Hilsberg, 1982). The mallard complex radiated around the world in the last million years (Johnson and Sorenson, 1999; Palmer, 1976). Secondary contact between species pairs has resulted in relatively high rates of introgressive

hybridization, especially between the geographically widespread mallard and the other species (Rhymer and Simberloff, 1996). Given these confounding influences, the mallard complex is an excellent study system to examine the sensitivity of phylogenetic inferences to methods of filtering data for recombination and hybridization.

Study System

There are 11-13 extant species and three or four subspecies (depending on taxonomic authority) recognized within the mallard complex (Appendix Table A1.1); these species are distributed across several major continents and islands (Johnsgard, 1978). On the basis of these distributions, Palmer (1976) proposed an "out of Africa hypothesis" which suggests an African origin, followed by a northward and eastward radiation through Eurasia, with a step-wise progression through the South Pacific, and perhaps a single colonization of North America. An African origin is also supported by mitochondrial (mt) DNA sequences (Johnson and Sorenson, 1999).

Although species within the mallard complex were likely allopatric or parapatric historically, the mallard has responded to anthropogenic influences (e.g., releases from game farms and altered landscapes) and can now be found in sympatry with most of the other species. This secondary contact has resulted in widespread hybridization with American black duck (*A. rubripes*; Avise *et al.*, 1990), mottled duck (*A. fulvigula*; McCracken *et al.*, 2001; Williams *et al.*, 2005a), Chinese spot-billed duck (*A. zonorhyncha*; Kulikova *et al.*, 2004), New Zealand (NZ) grey duck (*A. superciliosa superciliosa*; Rhymer *et al.*, 1994), Hawaiian duck (*A. wyvilliana*; Griffin and Browne, 1990), and yellow-billed duck (*A. undulata*; Pers. Obs.). As hybridization events typically produce 100% viable offspring (Avise *et al.*, 1990; but see Kirby *et al.*, 2004), the

taxonomy of this complex is uncertain (e.g., the Mexican Duck, *Anas* [*platyrhynchos*] *diazi*; AOU 1983, 2010-B, Gill at al. IOC World Bird List). Because hybridization events usually involve mallards, introgressed mallard alleles shared among the other species might confound phylogenetic inferences. To examine the influence of introgression on tree topologies, I reconstructed phylogenies with and without mallards. If introgression does not introduce biases, I predict comparable posterior support between sets of trees (ignoring recombination vs. filtering-recombination) as incomplete lineage sorting would have a similar influence regardless of data treatment. Alternatively, if recently introgressed mallard alleles have a strong influence on topologies or posterior support then I predict an increase in the posterior support for the relationships of the taxa that are influenced by these recently introgressed mallard derivatives.

Relationships within the mallard complex have been reconstructed with both morphometric data (Livezey, 1991) and mtDNA (Johnson and Sorenson, 1999; McCracken *et al.*, 2001), but the topologies of these trees differed in several ways. In particular, morphometrics supported a Pacific/southeast Asian clade that included the Pacific black duck (*A. superciliosa*), the Philippine duck (*A. luzonica*), the Indian spot-billed duck (*A. poeciliorhyncha*), and the Chinese spot-billed duck, whereas mtDNA placed the latter three species in a clade consisting of Old World (OW) mallards to the exclusion of Pacific black ducks. Chinese spot-billed ducks and mallards have polyphyletic mtDNA haplotypes (Kulikova *et al.*, 2004). Likewise, morphometrics suggested a sister relationship between the Hawaiian duck and the Laysan duck (*A. laysanensis*), but mtDNA supports a polyphyletic relationship among mallard and Hawaiian duck haplotypes that are not closely related to Laysan duck haplotypes (Fowler

et al., 2009; Johnson and Sorenson, 1999). Notably, differentiation in allozymes is more consistent with morphometrics (Browne *et al.*, 1993). Furthermore, neither of these data sets provided strong support for phylogenetic relationships among the North American monochromatic mallard-like ducks (mottled duck, American black duck, and Mexican duck), which have polyphyletic mtDNA (Avise *et al.*, 1990; McCracken *et al.*, 2001). Finally, mtDNA supports a prominent divergence between Eurasian and North American mallards (Johnson and Sorenson, 1999; Kulikova *et al.*, 2005), but at least qualitatively, there are no morphological differences between these populations. Given these conflicts between morphometric and mtDNA data, an analysis of independent characters is needed to understand the phylogenetic relationships of this recently radiated group.

Avian researchers have generally focused on mtDNA. Maternally inherited and having no recombination (Giles *et al.*, 1980; Watanabe *et al.*, 1985), mtDNA has a more rapid sorting rate and shorter coalescent intervals relative to biparentally-inherited, recombining nuclear DNA (nuDNA). This makes it particularly useful for recently diverged populations (Moore, 1995; Zink and Barrowclough, 2008). However, being maternally inherited and potentially under strong selection, its appropriateness for phylogenetics and phylogeography has been questioned (Bazin *et al.*, 2006; Edwards and Bensch, 2009; Edwards *et al.*, 2005; Hurst and Jiggins, 2005; Jacobsen and Omland, 2011b). Moreover, any single locus is sensitive to stochastic genealogical variability, which can mislead species-level phylogenies (Jacobsen and Omland, 2011b; Kubatko and Degnan, 2007; Maddison, 1997). Nevertheless, multi-locus comparisons—including between and within mitochondrial and nuclear genes—can provide insights into phenomena (e.g., historical introgression, mtDNA capture, sex-biased dispersal) that

would otherwise be lost in any single-locus analysis (Jacobsen *et al.*, 2010; Jacobsen and Omland, 2011a; Peters *et al.*, 2012a; Peters *et al.*, 2005). Thus, I also compare phylogenetic inferences among trees derived from morphometric data, mtDNA, and nuDNA.

MATERIALS AND METHODS

I sampled one to five individuals per species, subspecies or population for a total of 64 individuals from 16 operational taxonomic units (Appendix Table A1.1). Mallards from the New World (NW) and Old World (OW) and mottled ducks form the western gulf coast (WGC) and Florida (FL) were partitioned into subpopulations that were previously delineated with mtDNA (Avise *et al.*, 1990; Johnson and Sorenson, 1999; McCracken *et al.*, 2001) or nuDNA (Williams *et al.*, 2002; Williams *et al.*, 2005a; Williams *et al.*, 2005b).

Genomic DNA was isolated from each sample using a Qiagen DNAeasy blood and tissue kit (Qiagen, Valencia, CA) following manufacturer's protocol. I used previously optimized primers to amplify and sequence 19 nuclear introns (Table 1.1; Peters *et al.*, 2012b) and 640 bp of the mtDNA control region (Sorenson *et al.*, 1999; Sorenson and Fleischer, 1996). Additionally, I sequenced melanocortin 1 receptor (MC1R) because of its association with plumage characteristics in other birds (Mundy, 2005). Two sets of primers were designed to target 782 bp of exon sequence from the MC1R gene [primers MC1RR (5'ATGATGAGGATGAGGAAGAGG 3')/ MC1RFi (5' GTGGACCGCTACATCACCRT 3') and MC1RRi (5' TAGAGCACCAGCATGAGGA

3')/ MC1RF (5' CAGTGAGGGCAACCAGAG 3')]. These primers were designed from sequences downloaded from GenBank (accession numbers EU924091-EU924107 (Anas platyrhynchos); FJ605434-FJ605453 (Cairina moschata); Xia *et al.*, unpubl. data).

The polymerase chain reaction (PCR) was used to amplify each locus using 1.5 µL of template DNA (≥10 ng/µl), 2x GoTaq Green Master Mix (Promega), and 1.0 nM of each primer, in a total volume of 15 µL. PCR was conducted using an Eppendorf Mastercycler (epgradient) under the following conditions: DNA denaturation at 94°C for 7 minutes, followed by 45 cycles of denaturation at 94°C for 20 s, primer annealing at 58°C (at 52°C for mtDNA) for 20 s, and extension at 72°C for 1 minute, and a final DNA extension at 72°C for 7 minutes. Amplification was verified using gel electrophoresis with a 1.5% agarose gel, and PCR products were cleaned with AMPure XP beads following the Agencourt protocol (Beckman Coulter Co.). Sequencing was done using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) following manufacturer protocols using a 1/8 reaction. Final products were sent to the DNA Analysis Facility at Yale University for automated sequencing on an ABI 3730. Sequences were aligned and edited using Sequencher v. 4.8 (Gene Codes, Inc). All sequences have been submitted to GenBank (accession numbers: KF607919-KF609252; AY506871, AY506947, AY506948, AY506964, AY928831, AY928841-3, AY928846, Kulikova *et al.* 2004, 2005).

Table 1.1. Characteristics of 20 nuclear loci sequenced in this study: locus name, chromosomal location, and the total length, number of polymorphic sites, and number of parsimony-informative sites of non-filtered and filtered (in parentheses) datasets.

Locus	Location[1]	Non-Filtered (Filtered) Length[2]	Number of Polymorphic Sites	Number of Parsimony-Informative Sites
Chromo-helicase-DNA binding protein gene 1, intron 19	Z	326 (326)	10 (10)	3 (3)
Lactate dehydrogenase 1, intron 4	1	520 (520)	7 (7)	3 (3)
S-acyl fatty acid synthase thioesterase, intron 2	2	303 (303)	10 (10)	4 (4)
Ornithine decarboxylase, intron 7	3	302 (151)	37 (20)	24 (13)
Fibrinogen beta chain, intron 7	4	437 (246)	27 (15)	17 (8)
Serum amyloid A, intron 2	5	322 (133)	37 (12)	20 (5)
Annexin A11, intron 2	6	440 (382)	39 (34)	26 (23)
Myostatin, intron 2	7	281 (168)	26 (16)	11 (7)
Soat1-prov protein, intron 10	8	332 (332)	13 (13)	7 (7)
Nucleolin, intron 12	9	359 (98)	49 (16)	40 (14)
Melanocortin 1 receptor	11	782 (782)	28 (25)	10 (9)
Preproghrelin, intron 3	12	305 (290)	18 (17)	9 (8)
Glutamate receptor, ionotropic, N-methyl D aspartate I, intron 13	17	300 (85)	22 (1)	14 (0)
Sex determining region Y-box 9, intron 2	18	402 (120)	60 (12)	46 (11)
Carboxypeptidase D, intron 9	19	332 (127)	43 (18)	34 (13)
Phosphenolpyruvate carboxykinase, intron 9	20	333 (333)	12 (12)	10 (10)
Alpha enolase 1, intron 8	21	294 (179)	19 (11)	14 (8)
Alpha-B crystallin, intron 1	24	323 (323)	8 (8)	2 (2)
Growth hormone 1, intron 3	27	380 (379)	22 (21)	16 (12)
Lecithin-cholesterol acyltransferase, intron 3	Unk	323 (154)	36 (15)	22 (10)
Total		7396 (5431)	523 (293)	332 (170)
Percent change between filtered and non-filtered datasets		-27.0%	-44.0%	-49.0%

[1] **Location**: chromosomal location based on chicken genome (Hillier et al., 2004)

[2] **Length**: base-pairs

Nuclear Coalescent Phylogeny and ΦST Estimates

The gametic phases of nuclear alleles were determined by resolving sequences with the program PHASE v. 2.1.1 (Stephens and Donnelly, 2003), which derives the most likely state of each allele algorithmically. Additionally, indels were resolved using methods described in Peters *et al.* (2007) that determined gametic phases based on base-pair peak shifts within the chromatograms. Sequences resolved with this method were included as known alleles in the PHASE analyses. Mallard sequences were all resolved with >95% confidence from a larger data set that included extensive allele-specific priming (Peters *et al.* 2014), and these alleles were also treated as knowns.

Filtering for recombination was based on truncating loci into putatively non-recombinant fragments containing the highest number of polymorphic positions using the program IMgc (Woerner *et al.*, 2007). I iteratively adjusted chromosomal weighting so that a maximum of 5% of sequences were removed and so that both alleles from all taxa represented by a single individual were retained. Once thresholds were achieved sequences were manually truncated with the program Sequencher v. 4.8 (Gene Codes, Inc) to retain sites containing >2 states that would have been automatically removed by IMgc.

I used *Beast v.1.6.1 (Drummond and Rambaut, 2007; Heled and Drummond, 2012), which employs Markov chain Monte Carlo (MCMC) to estimate the posterior distribution of the species tree given the results from each gene tree, to reconstruct a multi-species Yule tree (Coalescent Yule-process). Analyses included (1) a non-filtered dataset for all taxa, (2) a recombination-filtered dataset for all taxa, (3) a non-filtered

dataset for all taxa excluding mallards, and (4) a recombination-filtered dataset for all taxa excluding mallards.

All loci were independently analyzed for substitution and clock models prior to species tree reconstruction. Substitution models were tested in MEGA v. 5.1 (Tamura *et al.*, 2011) and ranked based on Bayesian Information Criterion (BIC) scores that identified the Hasegawa-Kishino-Yano (HKY) model (Hasegawa *et al.*, 1985) as the most appropriate model for all datasets. Although additional parameters were not required for the truncated fragments, 12 loci within the full datasets required a gamma distribution across sites, with five of these having some proportion of invariable sites. Gene trees for each locus were analyzed with a strict clock (null model) and a Bayesian uncorrelated log-normal relaxed clock (alternative model) in *Beast v.1.6.1 and compared using Bayes Factors (BF) in Tracer v1.5 (Rambaut and Drummond, 2009). A log BF < 3 (or log BF > -3) (Li and Drummond, 2012) provided support for the null hypothesis of a strict clock for 13 loci in datasets ignoring recombination and 17 loci in recombination-filtered datasets. Species trees were then reconstructed with the appropriate substitution model and molecular clock defined for each locus (Appendix Table A1.2). A piecewise linear and constant root population size model with UPGMA starting trees (Sneath and Sokal, 1973) were used for each analysis. Sampling occurred every 2,000 iterations with runs continuing until effective sample sizes (ESS) across parameters were ≥ 100. Burn-in was set to 10% of the total number of sampled trees, and final species trees were constructed using TreeAnnotator and viewed in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree).

To examine overall levels of shared genetic variation, average pairwise ΦST was calculated for the 20 nuclear loci in Arlequin 3.5 (Excoffier and Lischer, 2010).

Mitochondrial Phylogeny

Two separate analyses were conducted using mtDNA, including a Bayesian derived individual tree reconstructed using MrBayes v. 3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) and a species tree reconstructed in *Beast v. 1.6.1 (Drummond and Rambaut, 2007; Heled and Drummond, 2012). An HKY substitution model with a gamma distribution across sites and a invariable sites model was determined as the best model based on BIC scores obtained in MEGA v. 5.1 (Tamura *et al.*, 2011). Molecular clocks were tested with similar methods as above by reconstructing species trees in *Beast v. 1.6.1 (Drummond and Rambaut, 2007; Heled and Drummond, 2012) and using Bayes Factors to compare them. A strict molecular clock was accepted, suggesting that rate variation across taxa is negligible and sequences are evolving in a clock-like fashion, which corroborates previous mtDNA studies (Weir and Schluter, 2008). The *Beast species tree obtained during the molecular clock analysis was used for direct comparison with the nuclear derived tree. A Bayesian tree illustrating relationships among all haplotypes was reconstructed in MrBayes using the same substitution and molecular clock models. The tree search comprised two concurrent runs, 3 million MCMC generations with sampling occurring every 2000 generations, and persisted until the average standard-deviation between runs was ≤ 0.01. The first 25% of trees were discarded as burn-in and the final tree was summarized and viewed in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree).

RESULTS

Nuclear Species Tree

In general, topologies across the four nuDNA coalescent trees (Fig. 1.1) were similar and included a basal African lineage consisting of the African black duck (*A. sparsa*), the yellow-billed duck (*A. undulata*), and the Meller's duck (*A. melleri*), an Australasian clade composed of the Philippine duck, NZ grey duck, and Pacific black duck, a Hawaiian clade with the Hawaiian duck and Laysan duck, and a New World clade with the NW mallard, Mexican duck, American black duck, FL mottled duck, and WGC mottled duck. In addition, analyses ignoring recombination yielded a South Pacific super clade with the Hawaiian and Australasian clades being sister, and also included the Chinese and Indian spot-billed ducks as sister lineages. Relationships within the NW clade and the OW mallard were poorly supported; however, the highest posterior support within the NW clade was obtained with the exclusion of mallards and ignoring recombination.

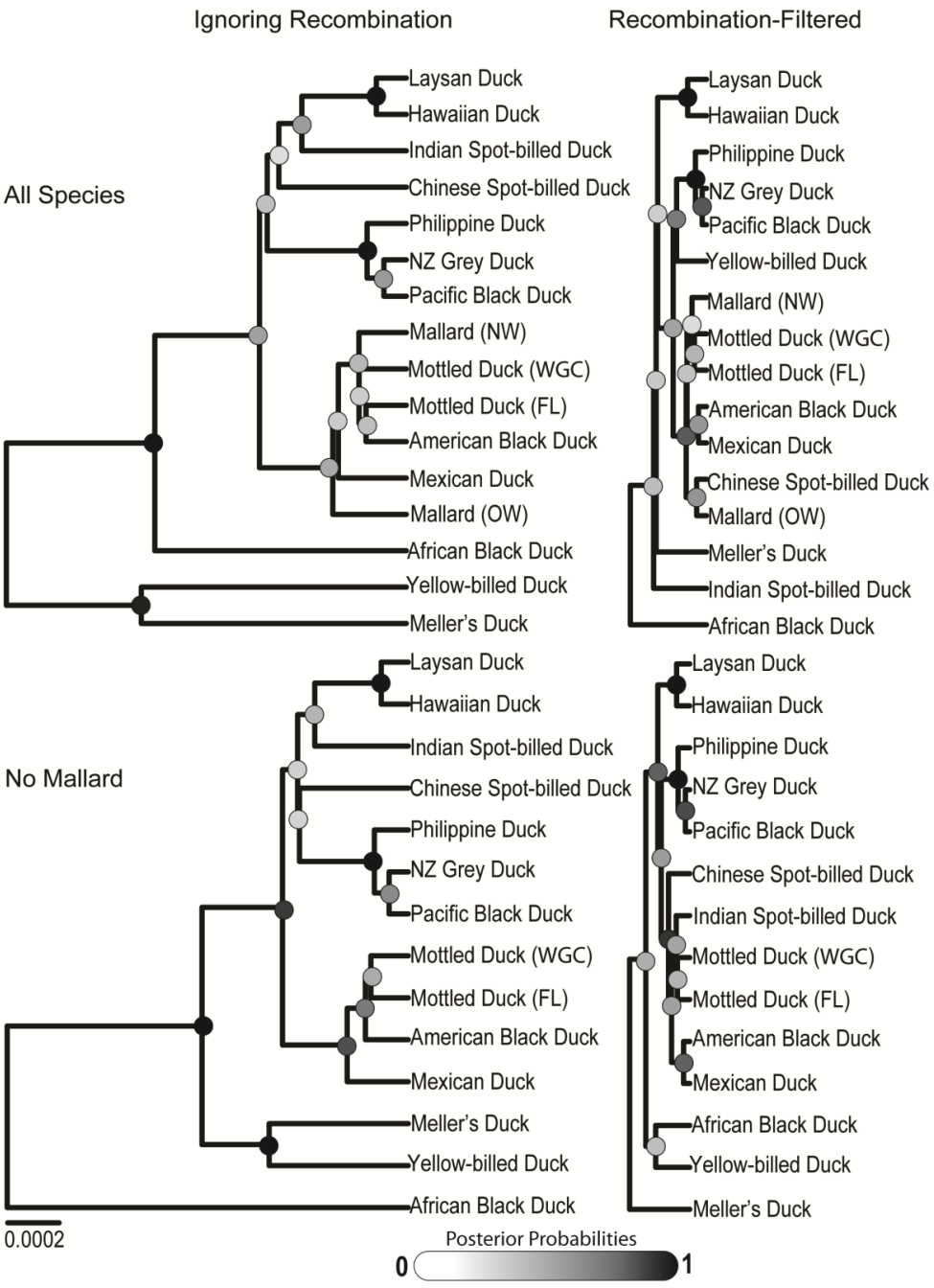


Figure 1.1. Nuclear multispecies coalescent trees reconstructed from 20 nuclear loci and 16 species/subspecies/populations of ducks. Analyses were conducted for the full dataset (ignoring recombination), recombination-filtered datasets, and including or excluding mallards, which hybridizes extensively with the other species.

Table 1.2. Average pairwise Φ_{ST} estimates for 20 nuclear loci across taxa of the mallard complex (abbreviations are defined in Appendix Table A1.1). $\Phi_{ST} \leq 0.05$ indicates a high proportion of shared polymorphisms and are shown in bold text.

	AFBD	YBDU	MELL	HAWD	LADU	PHDU	GRDU	PBDU	MALLow	SPBDch	SPBDin	MALLnw	MEDU	ABDU	MODUwgc
YBDU	0.342	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MELL	0.545	0.320	–	–	–	–	–	–	–	–	–	–	–	–	–
HAWD	0.210	0.276	0.336	–	–	–	–	–	–	–	–	–	–	–	–
LADU	0.889	0.633	0.919	0.387	–	–	–	–	–	–	–	–	–	–	–
PHDU	0.467	0.327	0.474	0.247	0.884	–	–	–	–	–	–	–	–	–	–
GRDU	0.215	0.273	0.251	0.145	0.495	0.088	–	–	–	–	–	–	–	–	–
PBDU	0.270	0.306	0.293	0.175	0.542	0.187	**0.015**	–	–	–	–	–	–	–	–
MALLow	0.167	0.220	0.169	0.096	0.460	0.090	0.065	0.089	–	–	–	–	–	–	–
SPBDch	0.176	0.206	0.151	0.112	0.484	0.151	0.083	0.089	**0.000**	–	–	–	–	–	–
SPBDin	0.414	0.285	0.467	0.172	0.878	0.382	0.151	0.187	**0.038**	0.069	–	–	–	–	–
MALLnw	0.161	0.179	0.148	0.094	0.463	0.109	0.085	0.121	**-0.001**	**0.025**	**0.025**	–	–	–	–
MEDU	0.168	0.199	0.173	0.095	0.480	0.114	0.078	0.106	**0.004**	**0.017**	0.070	**0.023**	–	–	–
ABDU	0.165	0.194	0.144	0.101	0.466	0.091	0.083	0.106	**-0.010**	**0.015**	0.052	**-0.011**	**0.012**	–	–
MODUwgc	0.136	0.168	0.146	0.104	0.474	0.120	0.094	0.116	**0.018**	**0.026**	0.056	**0.004**	**0.023**	**0.013**	–
MODUfl	0.126	0.196	0.188	0.107	0.489	0.153	0.123	0.147	**0.046**	0.063	0.056	**0.026**	0.055	**0.024**	**0.027**

Φ_{ST} estimates (Table 1.2) followed phylogenetic relationships. On average, ≥ 40% of the variability was explained by differences among taxa within the African group and between the African, South Pacific, and Australasian groups, whereas 17% of the genetic variability was explained when comparing African taxa to NW and OW taxa. Whereas pairwise Φ_{ST} estimates were on average 14% among the Hawaiian duck, Philippine duck, and the NZ grey/Pacific black duck, an average of 57% of differences were explained between these taxa and the Laysan duck. Finally, the lowest levels of differences were observed among NW taxa, OW mallards, and both spot-billed duck species ($\Phi_{ST} \leq 2\%$). Notably, the two mallard populations were indistinguishable from each other and the American black duck ($\Phi_{ST} < 0$).

Ignoring vs. Filtering for Recombination

After filtering for recombination, the total number of nucleotides, polymorphic sites, and informative sites decreased by 27%, 44%, and 49%, respectively (Table 1.1). All major groups were present with the filtered dataset, however, posterior support across nodes substantially decreased with the exception of those within the Hawaiian and Australasian clades. Although poorly supported across analyses, both spot-billed ducks (Indian and Chinese) were grouped within the Hawaiian and/or Australasian clades when ignoring recombination, but placed within the NW clade and elsewhere when analyzed with the recombination-filtered dataset. Neither dataset conclusively resolved relationships of NW taxa. Interestingly, on average, branch lengths substantially decreased when filtering for recombination (Fig. 1.1), and strongly corresponded to treatment (i.e., ignoring versus filtering for recombination) and not the inclusion/exclusion of mallards (Fig. 1.2).

Including vs. Excluding Mallards

Excluding mallards had no effect on overall relationships and little effect on posterior support of basal lineages. However, posterior support among the remaining NW taxa increased slightly when mallards were excluded (Fig. 1.1).

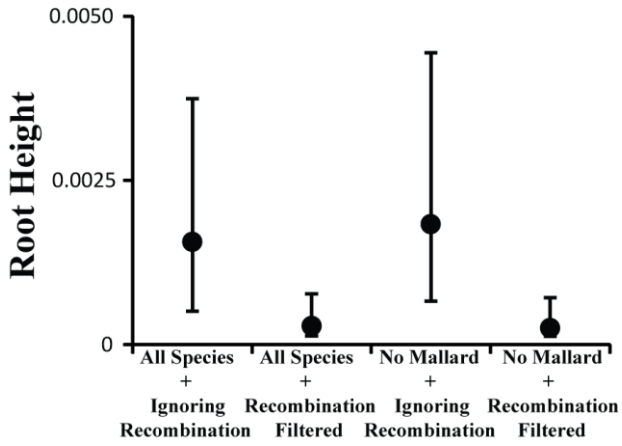


Figure 1.2. Mean and 95% highest posterior density for the root height of nuDNA species trees obtained from the four methods of data treatment (see Fig. 1.1).

Mitochondrial Derived Topologies

The mtDNA gene tree derived with Bayesian methods provided estimates of relationships among individuals, whereas the coalescent methods inferred species relationships. Although the Bayesian methods revealed a large polytomy and failed to resolve relationships among clades (Fig. 1.3), memberships within groups were generally well supported and consistent with previous studies. However, NW species, OW mallards, and both spot-billed ducks were polyphyletically intermixed, with some NW mallards grouping within the OW clade and some Chinese spot-billed ducks grouping within the NW clade (Kulikova *et al.*, 2005; Kulikova *et al.*, 2004). Chinese spot-billed duck haplotypes within the NW clade were consistent with a monophyletic subclade (Kulikova *et al.*, 2004). Within the NW clade, the placement of Hawaiian ducks was consistent with a monophyletic clade (see also Fowler *et al.*, 2009) that was sister to three of the five FL mottled ducks. In addition, a well-supported subclade consisted exclusively of Mexican ducks and WGC mottled

ducks (Fig. 1.3). FL and WGC mottled duck haplotypes were fairly divergent and consistent with previous studies (McCracken *et al.*, 2001). The Philippine duck grouped within the OW clade that consisted of OW mallards, Chinese spot-billed ducks, and Indian spot-billed ducks (Fig. 1.3). Otherwise, mtDNA haplotypes for the remaining species clustered into monophyletic clades. Two divergent groups were recovered for Pacific/New Zealand grey duck, one of which was exclusive to New Zealand (Rhymer *et al.*, 2004). Relationships within the coalescent derived species tree provided similar relationships with the exception being that the Philippine duck was recovered as sister to the yellow-billed duck (Fig. 1.4).

Morphometrics vs. mtDNA vs. nuDNA

Relationships provided by the three trees based on different character sets varied extensively, especially with regards to the placement of the Philippine duck, both Chinese and Indian spot-billed ducks, and the Hawaiian duck. The discrepancy, however, mostly lies with mtDNA, whereas topologies were nearly identical between morphometric data and nuDNA. Specifically, the sister relationship of the Philippine duck to the Pacific black duck/NZ grey duck lineage and the sister relationship between the Hawaiian duck and Laysan duck (Fig. 1.1) were consistently well supported by nuDNA and morphology, but not mtDNA (Fig. 1.4). Additionally, while the relationships of the two spot-billed ducks were poorly supported with nuDNA, they were found to be closer to the Hawaiian and Australasian clades with datasets where recombination was ignored, which once again corresponded to relationships derived from morphometric data. However, the topology of the nuDNA trees obtained from truncating sequences showed some evidence

of the spot-billed ducks grouping with the NW/OW mallards, which was consistent with the mtDNA topology.

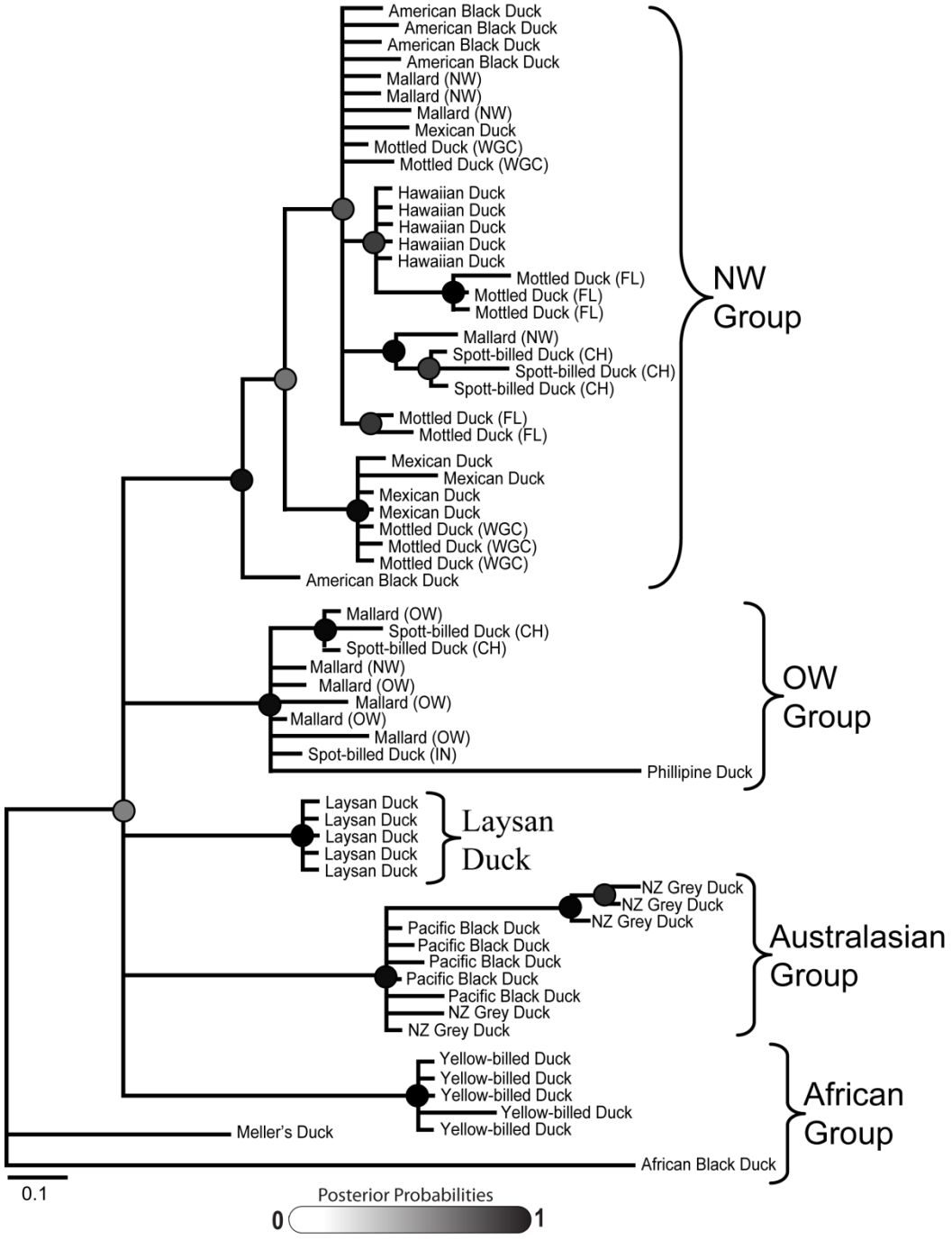


Figure 1.3. Mitochondrial gene tree reconstructed in MrBayes using 690 base pairs of the control region for 64 individuals.

DISCUSSION

Applying coalescent methods to a 20-locus dataset provided a fairly well resolved phylogeny for taxa within the mallard complex. Topologies across all protocols for handling recombination and hybridizing species were similar with major groups strongly supported. However, nodal support declined when filtering the data for recombination. Support for the more divergent lineages especially decreased, corroborating findings from simulated datasets showing that ignoring recombination decreased nodal support for deeper relationships within recently radiated taxa that have unsorted loci (Lanier and Knowles, 2012). The success of resolving relationships between taxa that are only statistically distinguishable based on allele frequencies is dependent on the presence of sufficient data (Knowles *et al.*, 2012; Maddison and Knowles, 2006). Specifically, as loci are truncated and informative variation is removed (e.g., -49% of parsimony-informative positions in this study; Table 1.1), the power for resolving relationships decreases. In contrast, ignoring recombination maximizes the number of nucleotides and individuals per taxon, which presumably enhances the phylogenetic signal obtained from statistically diagnostic markers. In general, while biases may be present when ignoring recombination, phylogenetic reconstructions of recently radiated taxa appear to be robust to violating the assumption of no recombination (Lanier and Knowles, 2012). Given the overall similarity in topologies from the recombination-filtered and the full datasets, my results are consistent with this generalization.

Unlike the tree topology, branch lengths were strongly affected when filtering the dataset for recombination (Figs. 1.1 and 1.2). Although taxonomic relationships are generally corroborated, discrepancies in branch lengths between the two datasets suggest

that estimating divergence times and the rate of diversification will be sensitive to how the data are treated (Fig. 1.2). However, it is not clear which of these methods give more realistic branch lengths. On the one hand, ignoring recombination might inflate branch lengths, because more mutations will be inferred when recombination creates new alleles. On the other hand, filtering for recombination can result in the exclusion of the most variable portions of the locus and the most variable sequences in the data set. This bias is supported by simulated data showing that filtering for recombination underestimates effective population sizes (Woerner *et al.*, 2007). Thus, analyzing recombination-filtered datasets likely biases branch lengths downwards, which would lead to underestimating divergence times. The true branch length is likely intermediate between these two extremes.

Relationships within the mallard complex

Topologies corresponded to those predicted by the ‘‘Out of Africa’’ hypothesis (Palmer, 1976), including basal African lineages (see also Johnson and Sorenson, 1999). However, whereas the ‘‘Out of Africa’’ hypothesis is based on a step-wise progression through the South Pacific after colonization of the OW (Palmer, 1976), phylogenetic (Fig. 1.1) and Φ_{ST} estimates (Table 1.2) suggest an almost simultaneous split between the Australasian clade, the Hawaiian clade, and OW mallards/NW taxa. Consequently, results from nuDNA are inconclusive regarding the step-wise progression proposed by Palmer (1976).

Few differences exist within and between NW taxa and OW mallards ($\Phi_{ST} \leq 2\%$), demonstrating the extent to which the genome is shared among them (e.g., Kraus *et al.*, 2012; Kulikova *et al.*, 2004). Moreover, of the two spot-billed ducks, pairwise UST

estimates were lowest when comparing the NW and OW mallards to the Chinese spot-billed duck (Table 1.2). In fact, the Chinese spot-billed duck appears to be indistinguishable from OW mallards ($\Phi_{ST} \leq 0\%$), which corroborates previous research from mtDNA suggesting high levels of hybridization between these two taxa (Kulikova *et al.*, 2004). In general, the inability to resolve relationships within these groups can be attributed to a recent ancestry (i.e., Upper Pleistocene; Heusmann, 1974) and ongoing introgressive hybridization between each species and the mallard. A larger sample size of individuals will be needed to increase the signal from allele-frequency differences, which can strengthen phylogenetic inferences for recently diverged taxa (Knowles, 2009; Knowles and Maddison, 2002; Lanier and Knowles, 2012). However, methods that incorporate introgression might be necessary to fully resolve these phylogenetic relationships.

The Introgressive Effect

I predicted that relationships among NW and OW taxa would be most influenced by the presence/absence of mallards because of the high incidence of hybridization between mallards and other taxa (Avise *et al.*, 1990; Kulikova *et al.*, 2004). Specifically, if relationships are significantly confounded by introgressed alleles then posterior support should increase when the introgressing species (i.e., mallard) is removed. Conversely, if relationships are predominantly affected by incomplete lineage sorting (ILS) then relationships and posterior support should remain fairly similar between tree pairs. In general, relationships remained similar and poorly supported within the NW/OW group across runs; however, the posterior support for NW taxa doubled when mallards were excluded and recombination was ignored. Similar patterns were not observed between

trees derived from recombination-filtered datasets, where the overall decreased resolution across topologies is likely attributable to the number of parsimony-informative sites that were lost (Table 1.1). Thus, results between datasets ignoring recombination suggest that the genomes of the various taxa are not swamped by mallard alleles, and while they do not have taxon-specific markers, they are likely distinguishable through frequency differences. However, when the mallard was included, shared alleles due to mallard introgression reduces the signal of these diagnostic markers. In general, these results demonstrate that the inclusion of the introgressing taxa does in fact influence the support of those taxa it is interacting with and that high rates of introgression may be having an important influence on inferences of phylogenetic relationships among the NW taxa.

Whereas Φ_{ST} estimates suggest that the two spot-billed ducks are more similar to NW taxa and OW mallards than to others, only trees reconstructed with no mallards and recombination-filtered datasets place them within the OW/NW group, and the tree reconstructed with mallards has the Chinese spot-billed duck as sister to the OW mallard (Fig. 1.1). Otherwise, trees reconstructed with datasets ignoring recombination placed them within the Australasian clade, which closely resembled the tree derived from morphometric data (Fig.1.4A; Livezey, 1991). Furthermore, unlike posterior support of NW taxa that increased when excluding mallards, those of the two spot-billed ducks remained low and unchanged across analyses. Consequently, relationships of the two spot-billed ducks seem to be more influenced by how the data are processed rather than the inclusion/exclusion of mallards, despite extensive hybridization between the Chinese spot-billed duck and the mallard (Kulikova *et al.*, 2004). In general, the Pacific relationship is likely due to retention of ancestral states that are similar to those within the

South Pacific super clade but missing in NW taxa, while extensive sharing of polymorphisms with NW taxa maintains low posterior support for these relationships (Fig. 1.3). Moreover, the loss of the spot-billed duck-from the South Pacific super clade in trees derived from recombination-filtered datasets is likely due to the loss of the ancestral states during the filtering process, which then draws the spot-billed ducks closer with OW/NW taxa.

Marker Comparison and Speciation within the Mallard Complex

I found strong discrepancies between morphometric, mitochondrial, and nuclear based phylogenies (Fig. 1.4). Generally, however, where the nuDNA and mtDNA topologies conflicted, the nuDNA was corroborated by morphometrics (Livezey, 1991). For example, mtDNA places the Hawaiian duck within the NW clade, whereas both morphology and nuDNA place it as the sister-taxon of the Laysan duck. The close affinity between the Hawaiian duck and Laysan duck is also supported by allozyme studies (Browne *et al.*, 1993). Similarly, whereas both morphology and nuDNA place the Philippine duck sister to the Pacific black duck and NZ grey duck, mtDNA suggests that it is part of the OW clade. These sister relationships received high posterior support in all four nuDNA trees. However, the placement of the Chinese and Indian spot-billed ducks is more ambiguous in the nuDNA trees and varied with the manner of treating data. When ignoring recombination, these species grouped with the South Pacific superclade; the inclusion of spot-bills and Australasian ducks within the same clade to the exclusion of Northern Hemisphere mallards is consistent with morphometrics but conflicts with mtDNA. In contrast, when filtering for recombination, spot-bills had a tendency to group

with the Northern Hemisphere ducks, which is more consistent with mtDNA relationships.

Such mito-nuclear conflict can result from a number of processes including stochastic lineage sorting and hybridization. It seems unlikely that this discord results from stochastic lineage sorting given the deep mtDNA branch lengths among the major clades, the shallow mtDNA divergence among species within clades, and the strong nodal support for the nuDNA topology. However, mitochondrial capture (Brelsford *et al.*, 2011) or hybrid speciation (Jacobsen and Omland, 2011a; Mallet, 2007) could explain this discord. First, considering the relationships presented with mtDNA and nuDNA, generations of introgressive events between female mallards and male heterospecifics can cause mtDNA to introgress and become fixed within the invaded species, resulting in mitochondrial capture. The strong support for the 20-locus nuDNA topology suggests close genomic affinities between Hawaiian and Laysan ducks and between Philippine and Pacific black ducks, which is consistent with introgression of mallard mtDNA into a genomically divergent species.

Alternatively, hybrid speciation theory dictates that a novel species evolves from historical hybridization events between two parental taxa (Seehausen, 2004). Such a mechanism for speciation within the mallard group has been suggested for the extinct Mariana mallard (*A. oustaleti*), which is believed to be descended from hybridization between the mallard and Pacific black duck (Reichel and Lemke, 1994). Thus, it is possible that the Hawaiian duck arose from hybridization between a NW mallard-like duck and Laysan duck and that the Philippine duck, and perhaps the spot-billed ducks, arose from hybridization between an OW mallard-like duck and the Pacific black duck

Additional data are required to test quantitatively these alternative hypotheses, which will need larger sample sizes (see Jacobsen and Omland, 2011a) and additional analyses (e.g., program STEM-hy; Kubatko, 2009). Finally, the mito-nuclear discordance for the Philippine duck and spot-billed ducks should be interpreted cautiously because I had only a single captive Philippine duck, which could complicate inferred relationships as ducks are well known for hybridizing in captivity (Johnsgard, 1960), and the phylogenetic placement of the spot-billed ducks received low posterior support.

CONCLUSIONS

The data presented in this study represent the most comprehensive phylogeny, both in terms of sample sizes and genomic coverage, for the mallard clade. This study illustrates the effectiveness of multi-locus data and coalescent methods in resolving phylogenetic relationships among taxa with extensive sharing of polymorphisms. Generally, posterior support across relationships, and more importantly branch lengths were reduced when filtering for recombination. Regardless, clade membership of taxa was generally supported by consistency across analyses and relatively strong posterior support for some nodes. Finally, the discordance in the placement of the Hawaiian duck, Philippine duck, as well as Indian and Chinese spot-billed ducks demonstrates how comparing trees based on different character sets can reveal phenomena that would otherwise be lost with a single tree. Testing the causes of this discordance can be important in reconstructing and understanding evolutionary history and speciation.

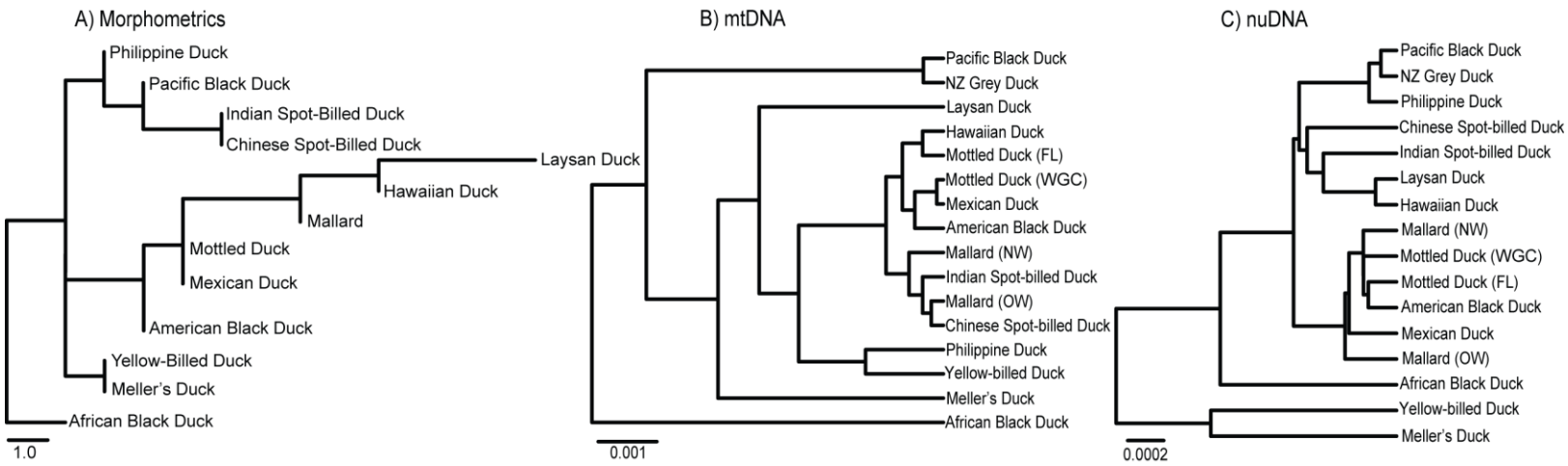


Figure 1.4. Phylogenetic relationships of 16 species/subspecies/populations from the mallard complex derived from A) morphometric data (Reconstructed with data from Livezey, 1991), B) mitochondrial DNA (mtDNA) control region, and C) 20 nuclear loci (nuDNA; ignoring recombination). Both species mtDNA and nuDNA species trees were reconstructed in *Beast v.1.6.1 (Drummond and Rambaut, 2007; Heled and Drummond, 2012) (see Materials and Methods).

ACKNOWLEDGMENTS

I am grateful to the following people and institutions for their contributions of samples to this study: Andy Engilis, Department of Wildlife, Fish and Conservation Biology, University of California, Davis; Leo Joseph, the Australian National Wildlife Collection; Graeme Cumming, the Percy FitzPatick Institute, University of Cape Town; Kevin G. McCracken with the Institute of Arctic Biology and Department of Biology and Wildlife at the University of Alaska Fairbanks; Yuri N. Zhuravlev and Irina Kulikova, Institute of Biology and Soil Sciences, Far Eastern Branch of the Russian Academy of Sciences; John Dyer and David Klee, New Zealand Fish and Game; Sylvan Heights Waterfowl Park in Scotland Neck, North Carolina, and the Burke Museum, University of Washington. This research was funded by Ducks Unlimited Richard H. G. Bonnycastle Fellowship in Wetland and Waterfowl Biology, the National Science Foundation (DEB-0926162), and the College of Sciences and Mathematics at Wright State University. The collection of yellow-billed ducks was supported by USAID via subcontract to the Wildlife Conservation Society's GAINS programme to G. Cumming. I also thank L. Joseph, G. Cumming, and K. McCracken for insightful comments on earlier drafts.

REFERENCES

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, Massachusetts.

Avise, J.C., Ankney, D.C., Nelson, W.S., 1990. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution 44, 1109–1119.

Bazin, E., Glémin, S., Galtier, N., 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312, 570–572.

Brelsford, A., MilÁ, B., Irwin, D.E., 2011. Hybrid origin of Audubon's warbler. Molecular Ecology 20, 2380–2389.

Browne, R.A., Griffin, C.R., Chang, P.R., Hubley, M., Martin, A.E., 1993. Genetic divergence among populations of the Hawaiian duck, Laysan duck, and mallard. The Auk 110, 49–56.

Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Systematic Biology 56, 400–411.

Chung, Y., Ané, C., 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Systematic Biology 60, 261–275.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genetics 2, e68.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24, 332–340.

Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.

Edwards, S., Bensch, S., 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. Molecular Ecology 18, 2930–2933.

Edwards, S.V., Kingan, S.B., Calkins, J.D., Balakrishnan, C.N., Jennings, W.B., Swanson, W.J., Sorenson, M.D., 2005. Speciation in birds: genes, geography, and sexual selection. Proceedings of the National academy of Sciences of the United States of America 102, 6550–6557.

Ellegren, H., 2010. Evolutionary stasis: the stable chromosomes of birds. Trends in Ecology & Evolution 25, 283–291.

Excoffier, L., Lischer, H.E.L., 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.

Fowler, A., Eadie, J., Engilis, A., 2009. Identification of endangered Hawaiian ducks (*Anas wyvilliana*), introduced North American mallards (*A. platyrhynchos*) and their hybrids using multilocus genotypes. Conservation Genetics 10, 1747–1758.

Giles, R.E., Blanc, H., Cann, H.M., Wallace, D.C., 1980. Maternal inheritance of human mitochondrial DNA. Proceedings of the National academy of Sciences of the United States of America 77, 6715– 6719.

Grant, P.R., Grant, B.R., 1992. Hybridization of bird species. Science 256, 193–197.

Grant, P.R., Grant, B.R., 1997. Hybridization, sexual imprinting, and mate choice. The American Naturalist 149, 1–28.

Greenwood, P.J., 1980. Mating systems, philopatry and dispersal in birds and mammals. Animal Behavior 28, 1140–1162.

Griffin, C.R., Browne, R., 1990. Genetic variation and hybridization in Hawaiian duck and mallards in Hawaii. U.S. Fish and Wildlife Service, Hawaiian and Pacific Islands National Wildlife Refuge Complex.

Hasegawa, M., Kishino, H., Yano, T.-a., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22, 160–174.

Heled, J., Drummond, A.J., 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic Biology 61, 138–149.

Heusmann, H.W., 1974. Mallard-black duck relationships in the northeast. Wildlife Society Bulletin 2, 171–177.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A.M., Delany, M.E., Dodgson, J.B., 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Hurst, G.D.D., Jiggins, F.M., 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.

Jacobsen, F., Omland, K.E., 2011a. Increasing evidence of the role of gene flow in animal evolution: hybrid speciation in the yellow-rumped warbler complex. Molecular Ecology 20, 2236–2239.

Jacobsen, F., Omland, K.E., 2011b. Species tree inference in a recent radiation of orioles (*Genus Icterus*): multiple markers and methods reveal cytonuclear discordance in the northern oriole group. Molecular Phylogenetics and Evolution 61, 460–469.

Jacobsen, F., Friedman, N.R., Omland, K.E., 2010. Congruence between nuclear and mitochondrial DNA: combination of multiple nuclear introns resolves a well supported phylogeny of New World orioles (*Icterus*). Molecular Phylogenetics and Evolution 56, 419–427.

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

Johnsgard, P.A., 1978. Ducks, Geese, and Swans. Univ. Nebraska Press, Lincoln.

Johnson, K.P., Sorenson, M.D., 1999. Phylogeny and biogeography of dabbling ducks (*Genus: Anas*): a comparison of molecular and morphological evidence. The Auk 116, 792–805.

Kirby, R.E., Sargeant, G.A., Shutler, D., 2004. Haldane's rule and American black duck x mallard hybridization. Canadian Journal of Zoology 82, 1827–1831.

Knowles, L.L., 2009. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. Systematic Biology 58, 463–467.

Knowles, L.L., Maddison, W.P., 2002. Statistical phylogeography. Molecular Ecology 11, 2623–2635.

Knowles, L.L., Lanier, H.C., Klimov, P.B., He, Q., 2012. Full modeling versus summarizing gene-tree uncertainty: method choice and species-tree accuracy. Molecular Phylogenetics and Evolution 65, 501–509.

Kraus, R.H., Kerstens, H.H., van Hooft, P., Megens, H.-J., Elmberg, J., Tsvey, A., Sartakov, D., Soloviev, S.A., Crooijmans, R.P., Groenen, M.A., 2012. Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks. BMC Evolutionary Biology 12, 45.

Kubatko, L.S., 2009. Identifying hybridization events in the presence of coalescence via model selection. Systematic Biology 58, 478–488.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56, 17–24.

Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971–973.

Kulikova, I.V., Zhuravlev, Y.N., McCracken, K.G., 2004. Asymmetric hybridization and sex-biased gene flow between eastern spot-billed ducks (*Anas zonorhyncha*) and mallards (*A. platyrhynchos*) in the Russian Far East. The Auk 121, 930–949.

Kulikova, I.V., Drovetski, S.V., Gibson, D.D., Harrigan, R.J., Rohwer, S., Sorenson, M.D., Winker, K., Zhuravlev, Y.N., McCracken, K.G., 2005. Phylogeography of the mallard (*Anas platyrhynchos*): hybridization, dispersal, and lineage sorting contribute to complex geographic structure. The Auk 122, 949–965.

Lanier, H.C., Knowles, L.L., 2012. Is recombination a problem for species-tree analyses? Systematic Biology.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Systematic Biology 60, 126–137.

Li, W.L.S., Drummond, A.J., 2012. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. Molecular Biology and Evolution 29, 751–761.

Lijtmaer, D.A., Mahler, B., Tubaro, P.L., Dunn, P., 2003. Hybridization and postzygotic isolation patterns in pigeons and doves. Evolution 57, 1411–1418.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543.

Livezey, B.C., 1986. A phylogenetic analysis of recent Anseriform Genera using morphological characters. The Auk 103, 737–754.

Livezey, B.C., 1991. A phylogenetic analysis and classification of recent dabbling ducks (Tribe *Anatini*) based on comparative morphology. The Auk 108, 471–507.

Maddison, W.P., 1997. Gene trees in species trees. Systematic Biology 46, 523–536.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology 55, 21–30.

Mallet, J., 2007. Hybrid speciation. Nature 446, 279–283.

McCracken, K.G., Johnson, W.P., Sheldon, F.H., 2001. Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). Conservation Genetics 2, 87–102.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation: mitochondrial gene trees versus nuclear-gene trees. Evolution 49, 718–726.

Mundy, N.I., 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proceedings of the Royal Society B: Biological Sciences 272, 1633–1640.

Palmer, R.S., 1976. Handbook of North American Birds. New Haven, Connecticut.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Molecular Biology and Evolution 5, 568–583.

Peters, J.L., McCracken, K.G., Zhuravlev, Y.N., Lu, Y., Wilson, R.E., Johnson, K.P., Omland, K.E., 2005. Phylogenetics of wigeons and allies (*Anatidae: Anas*): the importance of sampling multiple loci and multiple individuals. Molecular Phylogenetics and Evolution 35, 209–224.

Peters, J.L., Zhuravlev, Y., Fefelov, I., Logie, A., 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., Bolender, K.A., Pearce, J.M., 2012a. Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. Molecular Ecology 21, 3562–3575.

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

Peters, J. L., Winker, K., Millam, K., Lavretsky, P., Kulikova, I., Wilson, R., Zhuravlev, Y., Mccracken, K. G., 2014. Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus *Anas*). Molecular Ecology.

Price, T.D., Bouvier, M.M., 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56, 2083–2089.

Rambaut, A., Drummond, A.J., 2009. Tracer v1.5., Available from http://beast.bio.ed.ac.uk/Tracer.

Reichel, J.D., Lemke, T.O., 1994. Ecology and extinction of the Mariana mallard. The Journal of Wildlife Management 58, 199–205.

Rhymer, J.M., Simberloff, D., 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics 27, 83–109.

Rhymer, J.M., Williams, M.J., Braun, M.J., 1994. Mitochondrial analysis of gene flow between New Zealand mallards (*Anas platyrhynchos*) and grey ducks (*A. superciliosa*). The Auk 111, 970–978.

Rhymer, J.M., Williams, M.J., Kingsford, R.T., 2004. Implications of phylogeography and population genetics for subspecies taxonomy of grey (Pacific black) duck *Anas superciliosa* and its conservation in New Zealand. Pacific Conservation Biology 10, 57–66.

Rieseberg, L.H., Baird, S.J.E., Gardner, K.A., 2000. Hybridization, introgression, and linkage evolution. Plant Molecular Biology 42, 205–224.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Scherer, V.S., Hilsberg, T., 1982. Hybridisierung und verwandtschaftsgrade innerhalb der Anatidae – eine systematische un evolutionstheoretische betrachtung. Journal für Ornitologie 123, 357–380.

Seehausen, O., 2004. Hybridization and adaptive radiation. Trends in Ecology & Evolution 19, 198–207.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy. San Francisco.

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

Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12, 105–114.

Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. The American Journal of Human Genetics 73, 1162–1169.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using Maximum Likelihood, evolutionary distance, and Maximum Parsimony methods. Molecular Biology and Evolution 28, 2731-2739.

Watanabe, T., Mizutani, M., Wakana, S., Tomita, T., 1985. Demonstration of the maternal inheritance of avian mitochondrial DNA in chicken-quail hybrids. Journal of Experimental Zoology 236, 245–247.

Weckstein, J.D., Zink, R.M., Blackwell-Rago, R.C., Nelson, D.A., 2001. Anomalous Variation in mitochondrial genomes of white-crowned (*Zonotrichia leucophrys*) and golden-crowned (*Z. atricapilla*) sparrows: pseudogenes, hybridization, or incomplete lineage sorting?. The Auk 118, 231–236.

Weir, J.T., Schluter, D., 2008. Calibrating the avian molecular clock. Molecular Ecology 17, 2321–2328.

Williams, C.L., Brust, R.C., Rhodes Jr., O.E., 2002. Microsatellite polymorphism and genetic structure of Florida mottled duck populations. The Condor 104, 424–431.

Williams, C.L., Brust, R.C., Fendley, T.T., Tiller, G.R., Rhodes Jr., O.E., 2005a. A comparison of hybridization between mottled ducks (*Anas fulvigula*) and mallards (*A. platyrhynchos*) in Florida and South Carolina using microsatellite DNA analysis. Conservation Genetics 6, 445–453.

Williams, C.L., Fedynich, A.M., Pence, D.B., Rhodes Jr., O.E., 2005b. Evaluation of allozyme and microsatellite variation in Texas and Florida mottled ducks. The Condor 107, 155–161.

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

Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17, 2107–2121.

CHAPTER II. RAPID RADIATION AND HYBRIDIZATION CONTRIBUTE TO WEAK DIFFERENTIATION AND HINDER PHYLOGENETIC INFERENCES IN THE NEW WORLD MALLARD COMPLEX (*ANAS* SPP.)

Abstract – Of the thirteen taxa composing the Mallard complex, four occur in North America; the sexually monochromatic American Black Duck (*A. rubripes*), Mexican Duck (*A.* [*platyrhynchos*] *diazi*), and Mottled Duck (*A. fulvigula*), and the dichromatic Mallard (*Anas platyrhynchos*). Although, morphologically distinct, inferring the evolutionary relationships of this group is confounded by extensive genic sharing due to incomplete lineage sorting and ongoing hybridization. The objective of this study was to examine the underlying cause (i.e., incomplete lineage sorting vs. contemporary gene flow) of phylogenetic uncertainty. Whereas most taxa were fairly structured at mitochondrial DNA, a "star-burst" pattern of divergence consistent with a rapid radiation was recovered with 17 nuclear introns. Furthermore, nuclear-based divergence estimates and tests of population structure recovered Florida and West Gulf Coast mottled ducks as well differentiated and genetically diagnosable from each other and the remaining taxa, whereas mallards, American black ducks, and Mexican ducks were indistinguishable. In general, neither population structure analyses nor coalescent-based gene flow estimates conclusively identified the presence of hybrids or significant gene flow, suggesting that genetic similarity within the group is largely influenced by incomplete lineage sorting. However, I also cannot reject potentially high levels of gene flow. Furthermore, inconsistent relationships among species trees indicated that phylogenetic results were sensitive to which individuals were included. Taxa within the New World group are phenotypically distinguishable, yet genetically similar and without apparent reproductive

isolation that is consistent with early stages of (incomplete) speciation. Future work should focus on genomic regions under selection to better understand the stage of speciation among the various incipient forms.

INTRODUCTION

Selection can cause rapid phenotypic divergence in the absence of genomic differentiation between populations and species (Orr and Smith 1998; Funk and Omland 2003). Consequently, with the exception of key trait-specific genes, genomic divergence between incipient morphs – particularly for neutral markers – is affected by the time since isolation, population size, and the rate of introgressive hybridization (Grant and Grant 1997a). Such genetic similarities, stochastic lineage sorting, and differing gene histories (i.e. selected vs. neutral) can result in inconclusive and/or discordant phylogenetic relationships among different morphological and genetic markers (Omland 1997a; Carstens and Knowles 2007; Zink and Barrowclough 2008; Humphries and Winker 2011). Reconstructing phylogenetic history, however, can be achieved by maximizing the number of samples and genomic coverage and using Bayesian methods that incorporate genealogical differences across markers (Drummond and Rambaut 2007; Liu 2008; Kubatko et al. 2009). Establishing phylogenetic relationships can help us better understand the cause of phenotypic and genetic discordance, particularly when such discord can lead to incorrect taxonomic designations (Cicero and Koo 2012). Phenotypic-genetic discords are typically associated with rapid and/or recent adaptive radiations where morphological traits are under strong selection (e.g. trait-based assortative mating and/or niche-based selectivity; Greenberg et al. 1998), whereas the remaining genome is largely influenced by neutral processes (Humphries and Winker 2011). For instance, whereas adaptive traits can cause rapid diversification in avian lineages through premating isolation (Mayr and Ashlock 199; Grant and Grant 1997a;

Price 2008), retained genetic similarities are often attributed to strong dispersal ability (Greenwood 1980), chromosomal stasis (Ellegren 2010), and relatively low levels of reinforcement (Grant and Grant 1997b). Regardless, given suitable genomic coverage, coalescent based approaches appear capable of resolving such complex histories (Chung and Ané 2011; Leaché and Rannala 2011; Lanier and Knowles 2012).

Study System

Phylogenetic relationships within the Mallard complex (*Anas platyrhynchos* and allies) have proven difficult to resolve, owing to a recent radiation, widespread interspecific hybridization, and substantial phenotypic-mitochondrial-nuclear discordance (Livezey 1991; Johnson and Sorenson 1999; Lavretsky et al. 2014). Of the 14 taxa, the most confounded relationships are those within the New World (NW) group comprising the sexually dichromatic mallard and the monochromatic American black duck (*A. rubripes*; "Black Duck"), mottled duck (*A. fulvigula*), and Mexican duck (*A. [platyrhynchos] diazi*). Mitochondrial (mt) DNA haplotypes are polyphyletic among these taxa, suggesting a recent radiation (Avise et al. 1990; Johnson and Sorenson 1999; McCracken et al. 2001; Lavretsky et al. 2014), and ongoing hybridization between Mallards and each of the monochromatic species complicate phylogenetic inferences (Heusmann 1974; Hubbard 1977; Avise et al. 1990; Dwyer and Baldassarre 1993; Merendino et al. 1993; McCracken et al. 2001; Perez-Arteaga et al. 2002; Pérez-Arteaga and Gaston 2004; Williams et al. 2005a). Lavretsky et al. (2014), for example, demonstrated that the posterior support for the NW monochromatic taxa doubled when mallards were excluded, suggesting a confounding effect of contemporary introgression. In the absence of fixed nucleotide differences in mtDNA and nuclear (nu) DNA, allelic

frequency differences are paramount to phylogenetic reconstructions. Although coalescent methods account for incomplete lineage sorting, contemporary hybridization can bias tree reconstructions (McDade 1990, 1997; Heled et al. 2013). Consequently, resolution depends on sampling breadth of individuals and loci, and specifically on the number of individuals with mixed ancestries included in the analysis (e.g., the number of F2, F3, etc hybrid individuals present in datasets). Being phenotypically diagnosable (Palmer 1976; Livezey 1991) but genetically similar (Lavretsky et al. 2014), the NW group is an excellent system for studying phenotypic-genetic discordance that is typically associated with recent radiations (Freeland and Boag 1999; Degnan and Rosenberg 2009; Campagna et al. 2012).

The monochromatic black duck, mottled duck, and Mexican duck are endemic to North America (Johnsgard 1978). The black duck is distributed east of the Mississippi River and has migratory cycles typical of other North American waterfowl, whereas mottled ducks and Mexican ducks have more restricted distributions and are sedentary. Mottled ducks are endemic to two disjoint regions, with the first extending along the Texas-Louisiana coastline (West Gulf Coast (WGC)) and the second in Florida (FL) (Stutzenbaker 1988); these allopatric populations are genetically differentiated (McCracken et al. 2001; Williams et al. 2005a). Mexican duck distributions extend throughout central Mexico and into parts of southern Nevada, New Mexico, and Texas (Hubbard 1977; Perez-Arteaga et al. 2002). In contrast, the dichromatic mallard has a Holarctic distribution that extends across North America, Europe, and Asia, with strong mitochondrial differences between Eurasia and North America (Avise et al. 1990; Kulikova et al. 2005), but little to no nuclear differentiation among populations across

this range (Kraus et al. 2013). Once found primarily west of the Mississippi River, environmental degradation (Livezey 1991; Green 1996; Johnson and Sorenson 1999; Mank et al. 2004) and release programs (Heusmann 1974; Soutiere 1986; Hepp et al. 1988) caused an expansion of the mallard's range across North America leading to increased interspecific competition and hybridization with the monochromatic endemics. Growing interactions with mallards have negatively influenced black duck populations since the 1950s (Ankney et al. 1987; Avise et al. 1990; Dwyer and Baldassarre 1993; Merendino et al. 1993; Rhymer 2006), leading to concerns over the possibility of extinction by introgressive hybridization (Rhymer and Simberloff 1996). Moreover, the taxonomy of both Mexican ducks (Hubbard 1977; Perez-Arteaga et al. 2002) and mottled ducks (Bielefeld et al. 2010) have gone through various revisions, and continue to be debated. Given the impact of taxonomic decisions on conservation (Stutzenbaker 1988; Chesser et al. 2011), information on evolutionary relationships and population structure, including estimates of gene flow and molecular differentiation, are required.

Given the weak support for phylogenetic relationships within the NW mallard group (Lavretsky et al. 2014), the objective of this study was to examine the underlying cause (i.e., incomplete lineage sorting vs. gene flow) of phylogenetic uncertainty. I do this using a five-fold larger sample size, and (1) compare genetic differentiation among taxa in mtDNA and 17 nuclear loci, (2) estimate rates of gene flow and time since divergence between the dichromatic mallard and each of the monochromatic species, and (3) infer phylogenetic relationships while examining the influence of stochastic sampling (random subsampling of individuals) on species tree reconstructions. In this study, I treat incomplete lineage sorting as the null hypothesis. Alternatively, if gene flow is playing a

dominant role, then I expect some individuals to be assignable to taxon-specific populations whereas others will appear to have admixed genomes (i.e., hybrids) and evidence of non-zero gene flow. Finally, contemporary genetic similarities may be the result of recent historical introgression, and even perhaps repeated events due to glacial cycles (Waltari et al. 2007). However, I acknowledge that distinguishing such a scenario from incomplete lineage sorting may not be possible with the current molecular dataset.

MATERIALS AND METHODS

I added 98 individuals (19–20 individuals per taxon) to the sample of Lavretsky *et al.* (2014) for a total of 123 individuals comprising five recognized species or populations (Appendix Table A2.1; taxonomic designations based on AOU). In general, sampling spanned each taxon's range; however, black duck samples were restricted to the most northeastern part of their range where mallards are sparse and therefore may be less influenced by recent introgression (Fig. 2.1). Moreover, black ducks were collected from the USFWS waterfowl parts collection survey, and therefore, likely include individuals migrating from more northern breeding locations where mallards are absent or rare (Johnsgard 1978). Additionally, samples for mottled duck populations were obtained at the Hunter Parts survey, whereas Mexican ducks were salvaged from hunters in Mexico. In order to limit the influence of hybrids on analyses, individuals were chosen based on established "pure" wing characteristics (Carney 1992); however, I acknowledge that plumage characteristics are ineffective past the F2 stage as hybrids become phenotypically indistinguishable from parental types after multiple backcrosses (Avise *et*

al. 1990; Kirby *et al.* 2000). Finally, mallard sequences were obtained from Peters *et al.* (2014a; 2014b).

Genomic DNA was isolated from each sample using a Qiagen DNeasy blood and tissue kit (Qiagen, Valencia, CA) following manufacturer's protocol. Previously optimized primers were used to amplify and sequence 17 nuclear introns (Appendix Table A2.2; Peters *et al.* 2012) and 640 bp of the mtDNA control region (Sorenson and Fleischer 1996; Sorenson *et al.* 1999). PCR and DNA sequencing protocols are described in detail in Lavretsky *et al.* (2014). Final products were sent to the DNA Analysis Facility at Yale University for automated sequencing on an ABI 3730 (Applied Biosystems). Sequences were aligned and edited using Sequencher v. 4.8 (Gene Codes, Inc). Sequences were archived in Genbank (*accession numbers pending*).

Figure 2.1. Geographic distribution of sampling for the mallard (✹), black duck (■), Mexican duck (●), and (Florida (△) and West Gulf Coast (▲)) mottled ducks.

Gametic phases of nuclear alleles were either algorithmically determined with the program PHASE v. 2.1.1 (Stephens and Donnelly 2003) or by applying methods described in Peters *et al.* (2007) for heterozygous sequences containing indels; in the latter case, I compared the ambiguous 3'-end with the unambiguous 5'-end of forward and reverse sequences to resolve the composition and placement of gaps and the linkage of polymorphisms to those gaps. Sequences resolved with the latter method were included as known alleles in PHASE. Additionally, mallard sequences were all resolved with >95% confidence from a larger dataset that included extensive allele-specific priming (Peters *et al.* 2014b) and were also treated as known alleles in PHASE runs. PHASE was run for 1000 iterations after a burn-in of 1000 steps and a thinning interval of 100. Of the 2,091 sequences (123 individuals × 17 loci), the gametic phases for 1,857 sequences (88.8%) were resolved with greater than 90% posterior probability. Therefore, I chose the phase reconstructions that received the highest posterior probabilities for each individual per locus for further analyses.

Relationships among Individuals

A mtDNA haplotype network was constructed using the median-joining algorithm in the program Network v. 4.5.1.0 (Bandelt *et al.* 1999). In addition, unphased nuclear data were concatenated for a total of 5659 aligned base pairs and a consensus nuclear network was calculated using NeighberNet with equal angle parameters and averaging ambiguous states as implemented in SplitsTree4 (Huson and Bryant 2006). Finally, pairwise Φ_{ST} estimates for each locus were calculated in Arelquin v. 3.5 (Excoffier and

Lischer 2010). I tested for a correlation between Φ_{ST} values estimated from mtDNA and nuDNA using a Mantel test in the program *ZT* (Bonnet and Van de Peer 2002).

For each nuclear locus, alleles were coded as 1 to *n*, where *n* is the total number of alleles observed for a given locus, and entered into *Structure* v. 2.3.4 (Pritchard *et al.* 2000), which uses Bayesian clustering methods to determine the number of genetic populations and to assign individuals to those populations. I tested $K = 1–10$ populations using ten replicates of each value of *K* and 500,000 MCMC steps following a burn-in of 100,000 steps. The optimum *K* was determined by calculating ΔK in the program *Structure Harvester* (Earl and VonHoldt 2012). Final *Structure* outputs were based on the optimal clustering alignment across all ten replicates for each optimum *K* using a FullSearch algorithm as implemented in the program CLUMPP (Jakobsson and Rosenberg 2007). The nuclear data were insufficient for assigning individuals to populations in *Structure* employing an Admixture model, which likely resulted from extensive allelic sharing among these taxa. Therefore, I used the No Admixture model and independence among allele frequencies to test for subtle population structuring that may be present (Pritchard *et al.* 2000). I also partitioned the data into the two major subgroups detected by *Structure* (see Results) to test for finer structure that might be masked when analyzing the full data set.

Estimates of Gene Flow and Divergence Time

Rates of gene flow and time since divergence were estimated from the combined mtDNA and nuDNA datasets using isolation with migration (IM) models (Hey and Nielsen 2004, 2007). IM assigns posterior probability density estimates for population sizes, divergence time, and migration rates from non-recombinant sequence fragments

using Bayesian Markov chain Monte Carlo (MCMC) algorithms (Nielsen and Wakeley 2001). To meet the assumption of no intralocus recombination, all nuDNA were filtered for recombination using the program IMgc (Woerner *et al.* 2007) with weight given to maximize fragment length while maintaining the largest proportion of each population per dataset. IM analyses were run for a minimum of 10,000,000 generations following a burn-in of 1,000,000 generations; effective sample sizes (EES) were > 50 for all parameters (Hey and Nielsen 2004, 2007).

Time since divergence from mallards was simultaneously estimated with gene flow rates (i.e., IM; Hey and Nielsen 2004, 2007). Years since divergence (T) was derived as $T = t/\mu$, where t is the time since divergence parameter scaled to the geometric mean of per-locus mutation rates (μ) estimated in IM. I used an average nuclear mutation rate of 1.2×10^{-9} substitutions/site/year (Peters *et al.* 2008) and an average mitochondrial mutation rate of 4.8×10^{-8} substitutions/site/year (Peters *et al.* 2005). Multiplying these rates by the per-locus fragment lengths (Appendix Table A2.2) resulted in a geometric mean of 3.2×10^{-6} substitutions/locus/year.

Species Tree Reconstructions

*BEAST v. 1.7.4 (Drummond *et al.* 2012) which uses Markov chain Monte Carlo (MCMC) to estimate the posterior distribution of the species tree given the results from each gene tree (Heled and Drummond 2012), was used to reconstruct multi-species trees (Coalescent Yule-process) using the nuclear data. Given that ignoring recombination provided stronger support for phylogenetic relationships but did not appear to bias topologies within the mallard complex (Lavretsky *et al.* 2014; see also Lanier and Knowles 2012), full sequences were used in all phylogenetic analyses. *BEAST ran

slowly and failed to converge when analyzing the full nuclear data set of 123 individuals (~246 alleles/locus) sequenced for 17 loci. Therefore, to effectively run *BEAST, a total of ten individuals per taxa were randomly chosen without replacement for two separate analyses – this was repeated five times for a total of ten species trees. By doing so, I was able to examine the sensitivity of phylogenetic reconstructions to stochastic sampling, as similar and well supported relationships between replicates would strengthen conclusions. Each locus was tested for the most appropriate substitution and clock models. Base-pair substitution models and rate parameters (i.e., gamma distribution, invariable sites) were tested in MEGA v. 5.1 (Tamura *et al.* 2011) and ranked based on Bayesian Information Criterion (BIC). Molecular clocks were tested for each locus by reconstructing gene trees in *BEAST v.1.7.1 with a strict clock (null model) or a Bayesian uncorrelated log-normal relaxed clock (alternative model). Bayes Factors (BF) calculated in Tracer v. 1.5 (Rambaut and Drummond 2009) were used to distinguish between models (i.e., a log BF < 3 or BF > -3 provided support for the null hypothesis of a strict clock; Li and Drummond 2012). Species trees were then reconstructed with appropriate substitution and molecular clock models (Appendix Table A2.3). A piecewise linear and constant root population size model with UPGMA starting trees (Sneath and Sokal 1973) was used for each analysis, which consisted of 500,000,000 MCMC iterations with sampling every 5000 steps for a total of 100,000 trees and a burn-in of 10%. All runs were analyzed in Tracer v. 1.5 (Rambaut and Drummond 2009) to confirm that effective sample sizes (ESS) were ≥ 100 for all parameters (Rambaut and Drummond 2009). A "consensus" species tree was reconstructed by summarizing the entire posterior tree set derived from all ten species trees. Finally, posterior tree sets were visualized with

the DensiTree program (Bouckaert 2010), and subsequently superimposed over their respective species tree.

RESULTS

Genetic Differentiation and Population Structure

Two lineages corresponding to previously described A and B haplogroups (Avise *et al.* 1990; Johnson and Sorenson 1999; Kulikova *et al.* 2005) were observed in the mtDNA haplotype network (Fig. 2.2A). Seven mallards, three American black ducks, one Mexican duck, and one WGC mottled duck had group A haplotypes, whereas all others had B haplotypes. Two notable B group haplotypes included one consisting of twelve Mexican Ducks, three WGC Mottled Ducks, and two Mallards, and another with72% of all FL Mottled Ducks (also see McCracken et al. 2001). All other haplotypes were polyphyletic within the B haplogroup, and some haplotypes were shared between mallards and black ducks and between WGC mottled ducks and Mexican ducks; the FL mottled duck was the only taxon that did not share

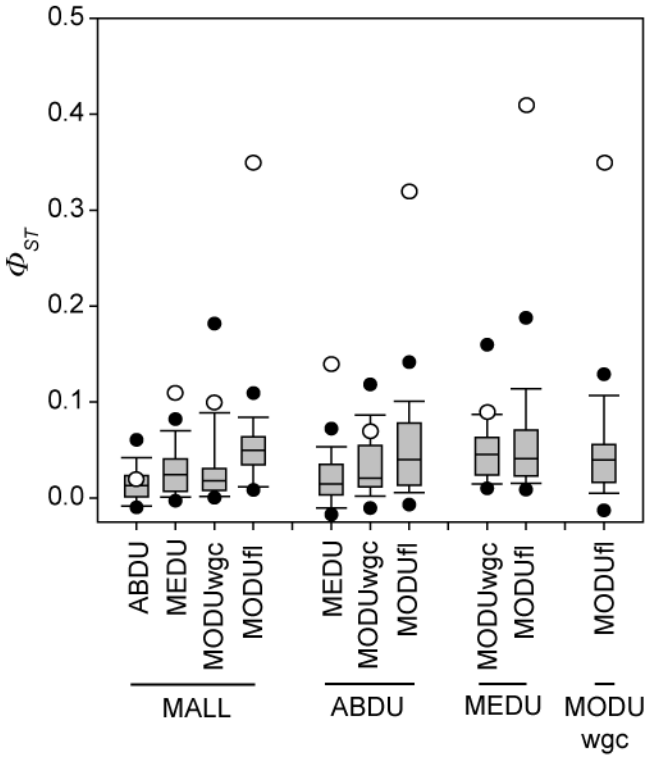


Figure 2.3. Box-plots of Φ_{ST} estimates per pair-wise comparisons. Open circles correspond to mtDNA based Φ_{ST} estimates. [ABDU = American black duck; MALL = mallard; MEDU = Mexican duck; MODUwgc = West Gulf Coast mottled duck; MODUfl = Florida mottled duck]

haplotypes with any other taxon. Φ_{ST} values corresponded to network patterns; significant Φ_{ST} values were observed for all pairwise comparisons except between mallards and black ducks (Φ_{ST} = 0.023; Fig. 2.3; Appendix Table A2.4). Mexican ducks and WGC mottled ducks were similarly differentiated from each other and from the remaining taxa (Φ_{ST} = 0.07–0.14), whereas Florida mottled ducks were the most differentiated overall ($\Phi_{ST} \geq$ 0.32).

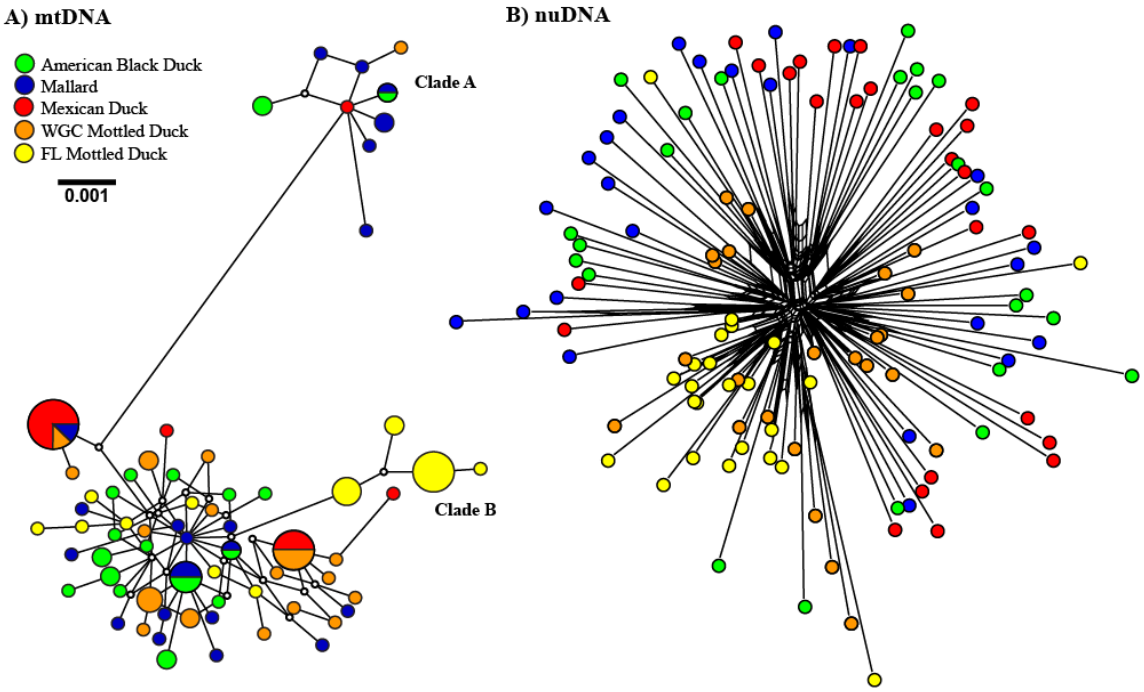


Figure 2.2. (A) Mitochondrial DNA median-joining network—size of circles corresponds to total number of individuals (range 1- 16) with that haplotype and branch lengths indicate the number of mutations separating haplotypes. (B) Neighbor-net nuclear network obtained from 17 nuclear loci.

Similar mtDNA-like structuring was not observed in nuDNA. Specifically, the NeighborNet appeared 'star-like,' demonstrating that taxa were broadly polyphyletic and indicating that many polymorphisms were shared among taxa (Fig. 2.2B). However, the two mottled duck populations tended to cluster together, suggesting some differentiation in allelic frequencies. These interpretations were further supported by pairwise Φ_{ST} values that indicated extensive genomic sharing and similar allele frequencies across taxa (Fig. 2.3; Appendix Table A2.4). Only 1–2% of the variation was explained by differences among mallards, black ducks, and Mexican ducks; however, 2.5 – 6.5% of the total genetic variation was explained by differences between FL and WGC mottled ducks (mean Φ_{ST} = 0.042) and between each mottled duck population and the other three species (mean Φ_{ST} = 0.024 – 0.064). On average, Φ_{ST} values for mtDNA were about 5 times larger than values for nuDNA, but mtDNA and nuDNA differentiation was significantly correlated among the 10 pairwise comparisons (Mantel test, r = 0.842, P = 0.017). *Structure* analyses corroborated Φ_{ST} estimates. First, the best-supported number of populations was K = 2 when analyzing all five populations together. Under this model, 19 of the 24 black ducks, 24 of the 25 mallards, and all Mexican ducks were assigned to population one, whereas all mottled ducks were assigned to population two (Fig. 2.4A). Sub-clade analyses did not provide additional resolution among the mallards, black ducks, and Mexican ducks; although K = 2 was the best-supported model, only a single black duck was assigned to the second population (Fig. 2.4B). However, sub-clade analyses revealed that most FL and WGC mottled ducks were assigned to separate

populations, although five WGC mottled ducks clustered with FL mottled ducks (K = 2 was the best-supported model; Fig. 2.4C).

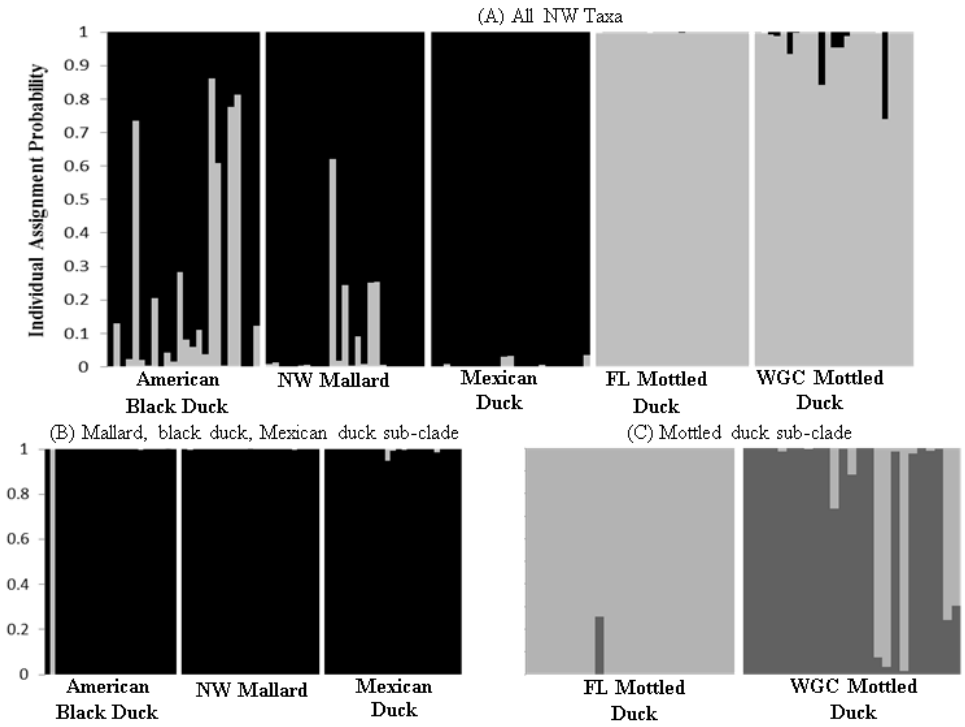


Figure 2.4. Structure assignment probabilities for (A) New World taxa and sub-clade analyses of (B) mallards, black ducks, and Mexican ducks, as well as (C) Florida (FL) and west gulf coast (WGC) mottled ducks.

Gene Flow and Divergence Estimates

Migration estimates suggested nearly equal bi-directional gene flow between mallards and each of the monochromatic taxa, and although consistent with low to moderate levels of gene flow, the estimates were also consistent with no gene flow (Fig. 2.5). Specifically, the lowest bin was contained within the 95% highest posterior distributions for all estimates of gene flow rates. The posterior distributions for gene flow

were flat between black ducks and mallards, and from mallards into Mexican ducks; thus for these species, the data are consistent with both no gene flow and high rates of gene flow (Fig. 2.5).

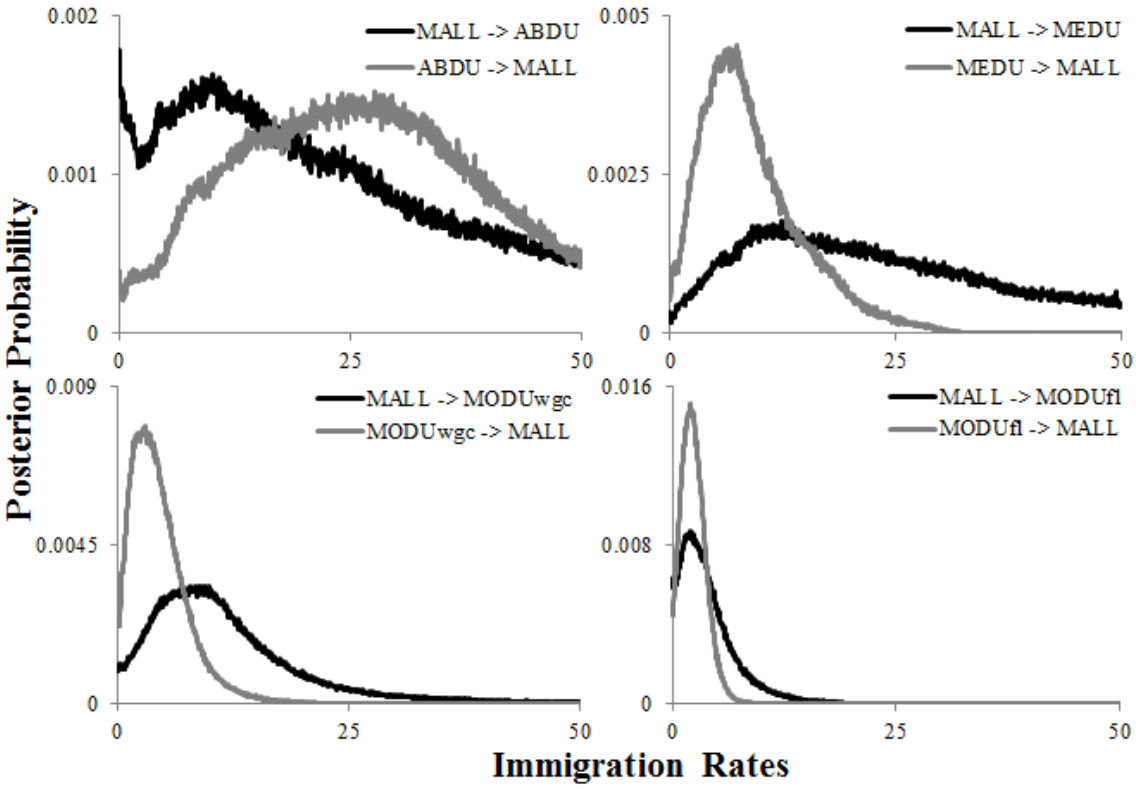


Figure 2.5. Immigration rates estimated in IM (Hey and Nielsen 2004, 2007) for each monochromatic-mallard pair-wise comparison. [ABDU = American black duck; MALL = mallard; MEDU = Mexican duck; MODUwgc = West Gulf Coast mottled duck; MODUfl = Florida mottled duck; → = gene flow into]

Time since divergence from mallards suggested that FL mottled ducks have been diverging for the longest time (390,000 years; 95% CI = 230,000–600,000 years), followed by Mexican ducks (325,000 years; 95% CI = 190,000-600,000 years), WGC

mottled ducks (245,000 years; 95% CI = 150,000–600,000 years), and black ducks (180,000 years; 95% CI = 100,000–400,000 years) (Fig. 2.6).While these divergence times appeared to be staggard, the confidence intervals were broadly overlapping among all pairwise comparisons.

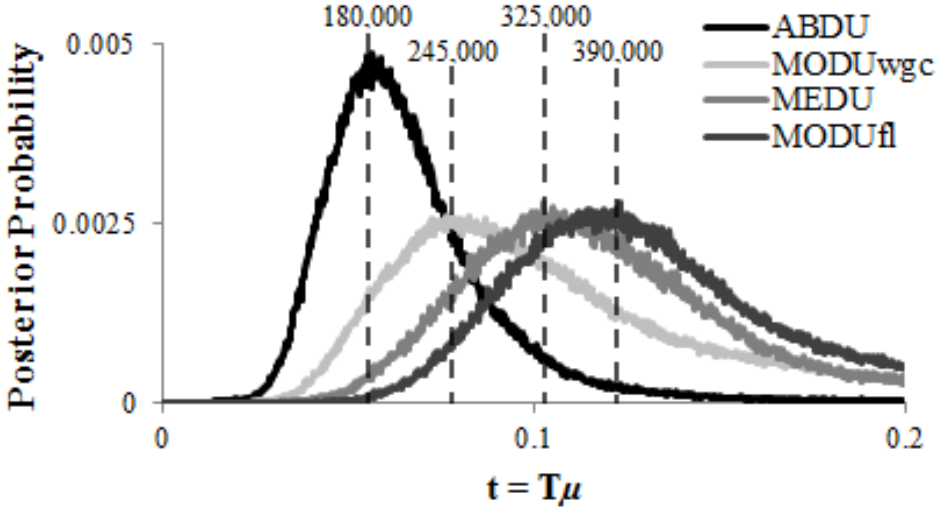


Figure 2.6. Pair-wise time since divergence from mallards estimated in IM (Hey and Nielsen 2004, 2007). Peak posterior supported time is depicted by a dashed line with Time = t/μ in years before present provided for each monochromatic-mallard comparison order from youngest to oldest. [ABDU = American black duck; MODUwgc = West Gulf Coast mottled duck; MEDU = Mexican duck; MODUfl = Florida mottled duck]

Phylogenetic Relationships

Phylogenetic analyses using the multispecies coalescent most frequently supported the two mottled duck populations as sister groups (8 of 10 trees) and grouped mallards, black ducks, and Mexican ducks as a monophyletic group (7 of 10 trees; Fig. 2.7) . These two groups were also most frequently supported when examining the entire

posterior set of trees across runs and the resulting consensus tree. However, the inferred sister relationships varied considerably among the individual species trees. For example, mallards were recovered as being sister to black ducks in 7 trees and sister to Mexican ducks in 3 trees, and each of these relationships received high posterior support in at least one analysis. Likewise, among the separate analyses, both phylogenetic placements of the WGC mottled duck as sister to the FL mottled duck or as part of the mallard-Black-Mexican group received strong posterior support, and one tree had high posterior support for the Mexican duck being the most basal lineage. Integrating results from all ten trees into a consensus tree, all phylogenetic relationships received low posterior support, suggesting that tree topologies were sensitive to which samples were included in the analysis.

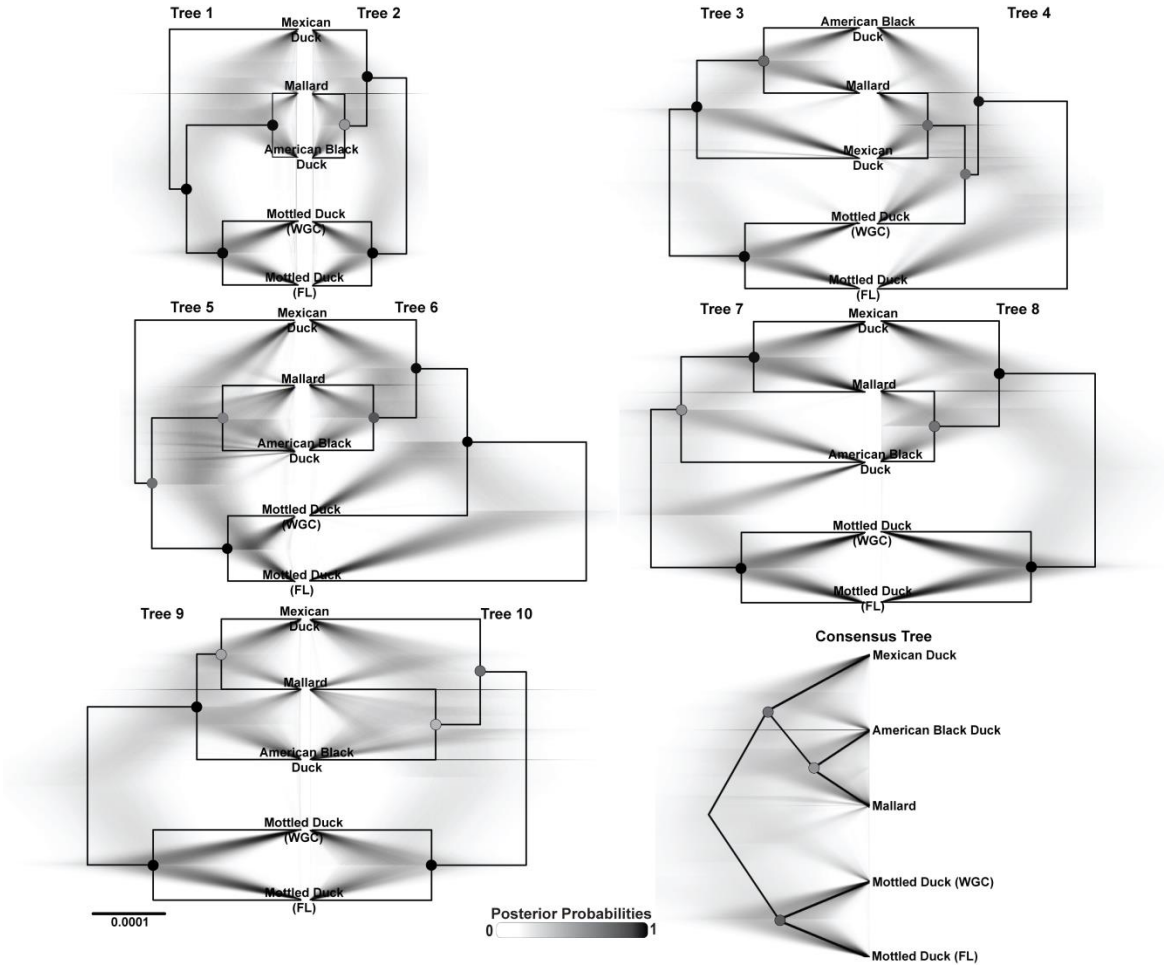


Figure 2.7. Ten species trees obtained from randomly partitioning the data for 17 nuclear introns into two sets of 10 individuals per taxon. The entire posterior sets of trees are superimposed for each analysis demonstrating the uncertainty in phylogenetic reconstructions. The consensus tree was obtained by combining the results of all ten replicates of species tree reconstructions.

DISCUSSION

Whereas the majority of pair-wise comparisons among species were significantly structured at mtDNA, the group was weakly differentiated across nuclear markers (Fig. 2.3; Appendix Table A2.4). Differences in sorting rates are likely sufficient to explain much of the variance between these marker types; Φ_{ST} values for mtDNA were five times larger than, but significantly correlated with, values for nuDNA, which is consistent with expectations based on mtDNA having ¼ the effective population size of nuDNA (Zink and Barrowclough 2008). This weak differentiation is likely due to a recent and rapid radiation, coupled with gene flow between the mallard and each of the monochromatic species, which hinders our ability to confidently reconstruct phylogenetic relationships.

Although there were few frequency differences (Fig. 2.3) within the nuDNA dataset, subtle population structure was recovered. Specifically, the *Structure* results, Φ_{ST} values, and coalescent trees all supported the two mottled duck populations as being most differentiated from the other taxa (Fig. 2.4A) and from each other (Fig. 2.7; Appendix Table A2.4; Fig. A2.4C). Significant differentiation between these populations is also corroborated by mtDNA, allozymes, and microsatellites (McCracken *et al.* 2001; Williams *et al.* 2005b). Elevated levels of differentiation in the mottled duck populations as compared to the other taxa, might be attributable to their relatively smaller population sizes and sedentary behavior (Stutzenbaker 1988; Ballard *et al.* 2001; Bielefeld *et al.* 2010). In addition, the distributions of mottled ducks coincide with possible glacial refugia (Waltari *et al.* 2007), which is consistent with these populations diverging in allopatry since the last glaciation. Such demographic and temporal attributes would result

in higher molecular sorting rates in these populations as compared to those with larger population sizes (i.e. black ducks or mallards) (Kimura and Ohta 1978) and suggests that neutral genetic drift might explain the population divergence. Interestingly, however, if demographic pressures are the primary cause of marker sorting, then why does the Mexican duck (also sedentary with a small population size) not show similar trends?

Phylogenetic relationships among mallards, black ducks, and Mexican ducks remain inconclusive despite examining 18 independent loci. However, whereas mallards and black ducks were not significantly structured at either mtDNA or nuDNA, mallards and Mexican ducks were significantly differentiated in mtDNA (Fig. 2.2A-B; Fig. 2.3; Appendix Table A2.4). One possible explanation for the apparent mito-nuclear discordance is that the sorting rate of nuDNA is too slow to track their recent divergence (McCracken and Sorenson 2005; Zink and Barrowclough 2008), which is consistent with the observed correlation and five-fold difference between mtDNA and nuDNA Φ_{ST} values. Alternatively, the discord could be a result of a hybridization bias where male mallards pair with female Mexican ducks and hybrids backcross into the Mexican duck population. Although mallard abundance has steadily declined by approximately 4.2% per year in Mexico (Pérez-Arteaga and Gaston 2004), past hybridization might have introduced mallard alleles into the population (Scott and Reynolds 1984). Furthermore, the greatest opportunities for contemporary hybridization likely occur in the southwestern part of the US where Mexican duck populations continue to regularly interact with mallards, and introgressed alleles have the potential to percolate into southern Mexican duck populations. However, our estimates of gene flow rates were consistent with complete isolation, although the posterior distribution of gene flow from mallards into

Mexican ducks was relatively flat and was also consistent with high levels of gene flow. More comprehensive sampling of Mexican ducks across their range is needed to better test hypotheses regarding the nuclear similarity between these species.

Stochastic Sampling and Hybridization

Inconsistent phylogenetic reconstructions based on 17 nuclear loci for mallards, black ducks, Mexican ducks, and the two populations of mottled ducks demonstrate the difficulties in resolving evolutionary relationships of recently radiated and currently hybridizing taxa. Despite a substantial increase in sample sizes relative to Lavretsky *et al.* (2014), relationships remained inconsistent across replicated species trees. The most common species tree was concordant with Φ_{ST} estimates and *Structure* results, supporting two primary lineages: a lineage consisting of mallards, black ducks, and Mexican ducks, and one consisting of FL and WGC mottled ducks (*Consensus Tree*, Fig. 2.7). However, only the FL mottled duck was independent of the mallard in all sets of species trees (Fig. 2.7), and the regular occurrence of various other relationships demonstrates that reconstructing these phylogenetic relationships was sensitive to stochastic sampling. Although I suspect that the inconsistencies among trees partially resulted from the inclusion of introgressed alleles, IM analyses were unable to conclusively demonstrate gene flow between mallards and each of the monochromatic species (Fig. 2.5). Furthermore, the posterior distributions of times since divergence were broadly overlapping among all pairwise comparisons when using models that incorporated gene flow (IM; Fig. 2.6), emphasizing the difficulties in reconstructing the history of divergence and phylogenetic relationships within this group. Comparing the results of the isolation-migration models with those from the multispecies coalescent suggests that

incomplete lineage sorting due to a rapid radiation might be contributing to phylogenetic uncertainties more so than hybridization. However, using a 6-fold larger sample size (but 1/3 the number of loci), Peters *et al.* (2014a) found significant evidence of gene flow from mallards into WGC mottled ducks, suggesting that gene flow could be playing a role in the inconsistent placement of WGC mottled ducks among phylogenetic trees. Regardless of the cause of inconsistencies among replicated species trees, the strong posterior support observed in some replicates provides a false confidence for relationships within this group. Interpreting the well-supported trees as evolutionarily likely or correct could have significant implications if applied to taxonomy, conservation, etc. (DeSalle *et al.* 2005; Oyler-McCance *et al.* 2010).

Future work will benefit from distinguishing between the effects of incomplete lineage sorting and hybridization within datasets. Although increasing sample sizes might offer higher resolution, knowledge on the frequency and geography of ongoing hybridization can further minimize the influence of contemporary introgression by excluding individuals from such areas *a priori*. For example, increased geographic sampling across the Mexican duck's range with subsequent genomic assays and comparisons between Mexican ducks and mallards could establish parental genotypes and help identify individuals with a hybrid ancestry. This would allow a direct assessment of the influence of hybridization on species tree reconstructions for this group.

Phenotypic-Genetic Discord

Dichromatism is presumed to be under sexual selection in populations where species recognition and the partner's quality must be accurately assessed amidst other

species and in short time periods (Johnsgard 1968). However, once selection is relaxed dichromatism can quickly be lost (Wiens 2001), as in numerous island taxa (Webster 1980). Such a scenario has been suggested for black ducks; Heusmann (1974) hypothesized that selection favored darker plumage that would be less conspicuous among the dark timber of Northeastern North America. Moreover, although "pure" Mexican ducks are distinguishable from mallards, their monochromatic plumage is similar to female mallards (Huey 1961; Hubbard 1977) and likely the ancestral state of the entire mallard clade (Johnsgard 1961; Omland 1997b; Johnson 1999). Alternatively, while the presence of "vestigial" mallard characters that have been described in black ducks and Mexican ducks were considered to be due to recent hybridization (Hubbard 1977; Livezey 1991), these may also be remnants of a recent dichromatic ancestor within the NW taxa (Omland 1997b).

Nuclear data revealed that mallards, black ducks, and Mexican ducks are three morphologically differentiated populations that are genetically indistinguishable (Ankney et al 1986; Hepp *et al.* 1988), much like the sexually dichromatic chestnut teal (*Anas castanea*) and monochromatic grey teal (*A. gracilis*) in Australia (Dhami *et al.* 2013). The plumage-genetic discrepancy can be explained by either (1) neutral alleles moving freely between populations coupled with selection inhibiting or preventing alleles at other loci from introgressing or (2) recent divergence among taxa with rapid phenotypic divergence that is not tracked by neutral variation (Winker 2009). Under the first scenario, neutral markers might provide false signals of divergence due to hybridization swamping the evolutionary signal (Palmer 1976; Johnson *et al.* 1999; McCracken *et al.* 2001; Kulikova *et al.* 2004), whereas under the second scenario the time since divergence

has been insufficient for drift to have had a major influence on neutral allele frequencies (Avise *et al.* 1990; Omland 1997b). Furthermore, this group might be best represented by nearly simultaneous divergence and a hard polytomy (Hoelzer and Melnick 1994), rather than a simple bifurcating tree, as has been suggested for other groups of ducks that have undergone a rapid radiation (Bulgarella *et al.* 2010). The identification of diagnostic markers that might be under selection will be instrumental in understanding the evolutionary histories of these taxa.

Considering Marker Variance in Taxonomy

Species recognition in avian lineages has been the subject of extensive debate due to the high variance of pre-zygotic and post-zygotic isolating mechanisms among genera (Grant and Grant 1992, 1997a). Without observable isolating mechanisms, taxonomic status is often based on morphometric data, niche partitioning, genetic relatedness among individuals, and the phylogenetic species concept (Mayr 1963, 1982). Among the NW taxa, extensive genic and phenotypic sharing has led to several taxonomic revisions and currently, three of the NW groups are considered species, one pair of subspecies (mallard & Mexican duck), and two subpopulations (FL & WGC mottled ducks; Table 2.1). However, our results largely disagree with these designations. Particularly, the two mottled duck subpopulations are nearly as divergent from each other as they are from the other taxa, and they might constitute different taxonomic units (e.g., subspecies; Callaghan 2005; Bielefeld *et al.* 2010). In contrast, mallard, black duck, and Mexican duck genetic relationships are shallow despite strong morphological differences. The discordance between morphological and genetic traits is suggestive of an adaptive radiation (Freeland and Boag 1999; Degnan and Rosenberg 2009; Campagna *et al.* 2012)

where selective or intrinsic factors influence morphological traits while the remaining genome is largely unaffected (Palmer 1976; Humphries and Winker 2011). A recent radiation is also supported by the 'star-burst' nuclear tree (Aleixandre *et al.* 2013) (Fig. 2.2B) and the overlapping estimates of time since divergence from mallards (Fig. 2.6).

In such instances of a rapid radiation accompanied by phenotypic-genetic discordance, a few genes might be responsible for maintaining species integrity (specifically, maintenance of those characters that lead us to recognize different species or subspecies), whereas shared polymorphisms are retained throughout the majority of the genome and/or can freely introgress between species. Under such a scenario, each taxon examined in this study could be considered a different species under the genic species concept (Wu 2001). Alternatively, numerous species develop reproductive barriers only after secondary contact when genetic incompatibilities are built up and lead to species barrier reinforcement (Short 1969; Grant and Grant 1992). Although these species might be genetically cryptic (Grant and Grant 1997a), until speciation genes are uncovered, the weak or non-existent genetic differentiation suggests that the NW taxa may be incipient morphs. In general, selection on genomic regions responsible for species integrity needs to be stronger than gene flow rates in order to resist amalgamation (Slatkin 1987; Charlesworth *et al.* 1997; Wu 2001). Higher genomic coverage is necessary (i.e., through next-generation sequencing) to successfully uncover and test for the presence/absence of speciation genes, and resolving the evolutionary relationships of the NW mallards may require thousands of loci (e.g., African rift-lake cichlids; Keller *et al.* 2012). Nevertheless, speciation is a dynamic process and studies of recently radiated taxa will need to consider the adaptive advantages of populations that are at present

unexplained by molecular divergence, but yet maintain species integrity (Price *et al.* 2003).

ACKNOWLEDGMENTS

I am grateful to Ken Richkus with U.S. Fish and Wildlife Services and the North American Flyway wingbees, Eduardo Carrera and Ducks Unlimited-Mexico, Ruben Del Castillo, Todd Scott and Wingshooters Lodge-Mexico, Kevin G. McCracken with the Institute of Arctic Biology and Department of Biology and Wildlife at the University of Alaska Fairbanks, Blanca E. Hernández-Baños with the Departamento de Biología Evolutiva, Universidad Nacional Autónoma de México, Distrito Federal, México, and Andrew Engilis, Jr. and the University of California, Davis Museum of Wildlife and Fish Biology for the contributions of samples to this study. This research was funded by Ducks Unlimited Richard H. G. Bonnycastle Fellowship in Wetland and Waterfowl Biology, American Museum of Natural History Chapman Grant, the National Science Foundation (DEB-0926162), Ohio Waterfowl Association Graduate fellowship, the U.C. Davis Museum of Wildlife and Fish Biology, and a Research Initiation Grant from the Research Council at Wright State University. I also thank Andrew Engilis, Jr. at the UC-Davis Museum of Wildlife and Fish Biology for comments on earlier drafts.

REFERENCES

Aleixandre, P., Montoya, J.H., Milá, B., 2013. Speciation on oceanic islands: rapid adaptive divergence vs. cryptic speciation in a Guadalupe Island songbird (Aves: *Junco*). PLoS ONE 8, e63242.

Ankney, C. D., Dennis, D. G., Wishard, L. N., Seeb. J. E., 1986. Low genic variation between black ducks and mallards. The Auk 103, 701-709.

Ankney, D.C., Dennis, D.G., Bailey, R.C., 1987. Increasing mallards, decreasing American black ducks: coincidence or cause and effect? The Journal of Wildlife Management 51, 523-9.

Avise, J.C., Ankney, D.C., Nelson, W.S., 1990. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution 44, 1109-19.

Ballard, B.M., Merendino, M.T., Terry, R.H., Tacha, T.C., 2001. Estimating abundance of breeding mottled ducks in Texas. Wildlife Society Bulletin 29, 1186-92.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16, 37-48.

Bielefeld, R.R., Brasher, M.G., Moorman, T.E., Gray, P.N., 2010. Mottled duck (*Anas fulvigula*). In: The Birds of North America Online (ed. by Poole A), Ithaca: Cornell Lab of Ornithology.

Bonnet, E., Van De Peer, Y., 2002. zt: a software tool for simple and partial Mantel tests. Journal of Statistical Software 7, 1-12.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372-3.

Bulgarella, M., Sorenson, M.D., Peters, J.L., Wilson, R.E., Mccrakcen. K.G., 2010. Phylogenetic relationships of *Amazonetta, Speculanas, Lophonetta,* and

Tachyeres: four morphologically divergent duck genera endemic to South America. Journal of Avian Biology 4, 186-199.

Callaghan, D., 2005. Mottled duck *Anas fulvigula*. In: ducks, Geese and Swans (ed. by Kear J.), pp. 517–20. Oxford University Press, New York, NY.

Campagna, L., Benites, P., Lougheed, S.C., Lijtmae,r D.A., Di Giacomo, A.S., Eaton, M.D., Tubaro, P.L., 2012. Rapid phenotypic evolution during incipient speciation in a continental avian radiation. Proceedings of the Royal Society B: Biological Sciences 279, 1847-56.

Carney, S. M., 1992. Species, age and sex identification of ducks using wing plumage. US Department of the Interior, US Fish and Wildlife Service Washington, DC Jamestown, ND.

Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. Systematic biology 56, 400-11.

Charlesworth, B., Nordborg, M., Charlesworth. D., 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical research 70, 155-174.

Chesser, R.T., Banks, R.C., Barker, F.K., Cicero, C., Jon, L.D., Kratter, A.W., Irby, J.L., Rasmussen, P.C., Remsen, J.V., Rising, J.D., Stotz, D.F., Winker, K., 2011. Fifty-second supplement to the american ornithologists' union check-list of north american birds. The Auk 128, 600-13.

Chung, Y., Ané, C., 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Systematic biology 60, 261-75.

Cicero, C., Koo, M.S., 2012. The role of niche divergence and phenotypic adaptation in promoting lineage diversification in the sage sparrow (*Artemisiospiza belli*, Aves: *Emberizidae*). Biological Journal of the Linnean Society 107, 332-54.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24, 332-40.

DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 1905-16.

Dhami, K.K., Joseph, L., Roshier, D.A., Heinsohn, R., Peters, J.L., 2013. Multilocus phylogeography of Australian teals (*Anas* spp.): a case study of the relationship between vagility and genetic structure. Journal of Avian Biology 44, 169-78.

Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology 7, 214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969-73.

Dwyer, C.P., Baldassarre, G.A., 1993. Survival and nest success of sympatric female mallards, *Anas platyrhynchos*, and American black ducks, *A. rubripes*, breeding in a forested environment. Canadian Field-Naturalist 107, 213-6.

Earl, D.A., VonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4, 359-61.

Ellegren, H., 2010. Evolutionary stasis: the stable chromosomes of birds. Trends in Ecology & Evolution 25, 283-91.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564-7.

Freeland, J.R., Boag, P.T., 1999. The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. Evolution, 1553-63.

Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 34, 397-423.

Grant, P.R., Grant, B.R., 1992. Hybridization of bird species. Science 256, 193-7.

Grant, P.R., Grant, B.R., 1997a. Genetics and the origin of bird species. Proceedings of the National academy of Sciences of the United States of America 94, 7768-75.

Grant, P.R., Grant, B.R., 1997b. Hybridization, sexual imprinting, and mate choice. The American Naturalist 149, 1-28.

Green, A.J., 1996. Analyses of globally threatened anatidae in relation to threats, distribution, migration patterns, and habitat use. Conservation Biology 10, 1435-45.

Greenberg, R., Cordero, P.J., Droege, S., Fleischer, R.C., 1998. Morphological adaptation with no mitochondrial DNA differentiation in the coastal plain swamp sparrow. The Auk, 706-12.

Greenwood, P.J., 1980. Mating systems, philopatry and dispersal in birds and mammals. Animal Behaviour 28, 1140-62.

Heled, J., Drummond, A.J., 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic biology 61, 138-49.

Heled, J., Bryant, D., Drummond. A. J., 2013. Simulating gene trees under the multispecies coalescent and time-dependent migration. BMC evolutionary biology 13, 1-10.

Hepp, G.R., Novak, J.M., Scribner, K.T., Stangel, P.W., 1988. Genetic distance and hybridization of black ducks and mallards: a morph of a different color? The Auk 105, 804-7.

Heusmann, H.W., 1974. mallard-black duck relationships in the northeast. Wildlife Society Bulletin 2, 171-7.

Hey, J., Nielsen. R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167, 747-760.

Hey, J., Nielsen. R., 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National academy of Sciences of the United States of America 104, 2785-2790.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C. P., Bork, P., Burt, D.W., Groenen, M.A.M., Delany, M.E., Dodgson. J.B., 2004. Sequence

and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695-716.

Hoelzer, G. A., Melnick. D.J., 1994. Patterns of speciation and limits to phylogenetic resolution. Trends in Ecology & Evolution 9, 104-107.

Hubbard, J.P., 1977. The biological and taxonomic status of the Mexican duck. In: New Mexico Department of Game and Fish Bulletin, pp. 1-56.

Huey, W.S., 1961. Comparison of female mallard with female New Mexican duck. The Auk 78, 428-31.

Humphries, E.M., Winker, K., 2011. Discord reigns among nuclear, mitochondrial and phenotypic estimates of divergence in nine lineages of trans-Beringian birds. Molecular Ecology 20, 573-83.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254-67.

Jakobsson, M., Rosenberg. N. A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801-1806.

Johnsgard, P.A., 1961. Evolutionary relationships among the North American mallards. The Auk 78, 3-43.

Johnsgard, P.A., 1968. Waterfowl: their biology and natural history. University of Nebraska Press Lincoln, Nebraska.

Johnsgard, P.A., 1978. ducks, geese, and swans of the world. In: Lincoln, Univ. Nebraska Press.

Johnson, K. P., 1999. The evolution of bill coloration and plumage dimorphism supports the transference hypothesis in dabbling ducks. Behavioral Ecology 10, 63-67.

Johnson, K.P., McKinney, F., Sorenson, M.D., 1999. Phylogenetic constraint on male parental care in the dabbling ducks. Transactions of the Royal Society of London. Series B: Biological Sciences 266, 759-63.

Johnson, K.P., Sorenson, M.D., 1999. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): A comparison of molecular and morphological evidence. The Auk 116, 792-805.

Keller, I., Wagner, C., Greuter, L., Mwaiko, S., Selz, O., Sivasundar, A., Wittwer, S., Seehausen, O., 2012. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular Ecology 22, 2848–2863.

Kirby, R.E., Reed, A., Dupuis, P., Obrecht Iii, H.H., Quist. W.J., 2000. Description and identification of American black duck, mallard, and hybrid wing plumage. US Geological Survey, Biological Resources Division, Biological Science Report USGS/BRD/BSR-2000-0002, 1-26.

Kimura, M., Ohta, T., 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National academy of Sciences of the United States of America of the United States of America 75, 2868-72.

Kraus, R.H., Hooft, P., Megens, H.J., Tsvey, A., Fokin, S.Y., Ydenberg, R.C., Prins, H.H., 2013. Global lack of flyway structure in a cosmopolitan bird revealed by a

genome wide survey of single nucleotide polymorphisms. Molecular Ecology 22, 41-55.

Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971-3.

Kulikova, I.V., Zhuravlev, Y.N., McCracken, K.G., 2004. Asymmetric hybridization and sex-biased gene flow between eastern spot-billed ducks (*Anas zonorhyncha*) and mallards (*A. platyrhynchos*) in the Russian Far East. The Auk 121, 930-49.

Kulikova, I. V., Drovetski, S.V., Gibson, D.D., Harrigan, R.J., Rohwer, S., Sorenson, M.D., Winker, K., Zhuravlev, Y.N., Mccracken. K.G., 2005. Phylogeography of the mallard (*Anas platyrhynchos*): hybridization, dispersal, and lineage sorting contribute to complex geographic structure. The Auk 122, 949-965.

Lanier, H.C., Knowles, L.L., 2012. Is recombination a problem for species-tree analyses? Systematic biology 61, 691-701.

Lavretsky, P., McCracken, K.G., Peters J.L., 2014. Phylogenetics of a recent radiation in the mallards and allies (genus *Anas*): inferences from a genomic transect and the multispecies coalescent. Molecular Phylogenetics and Evolution 70, 402-411.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Systematic biology 60, 126-37.

Li, W.L.S., Drummond, A.J., 2012. Model averaging and bayes factor calculation of relaxed molecular clocks in bayesian phylogenetics. Molecular Biology and Evolution 29, 751-61.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542-3.

Livezey, B.C., 1991. A phylogenetic analysis and classification of recent dabbling ducks (Tribe *Anatini*) based on comparative morphology. The Auk 108, 471-507.

Mank, J.E., Carlson, J.E., Brittingham, M.C., 2004. A century of hybridization: decreasing genetic distance between American black ducks and mallards. Conservation Genetics 5, 395-403.

Mayr, E., 1963. Animal Species And Evolution. Belknap Press of Harvard University Press, Cambridge.

Mayr, E., 1982. Of what use are subspecies? The Auk 99, 593-5.

Mayr, E., Ashlock P.D., 1991. Principles of Systematic Zoology. McGraw-Hill New York.

McCracken, K.G., Johnson, W.P., Sheldon, F.H., 2001. Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). Conservation Genetics 2, 87-102.

McCracken, K.G., Sorenson, M.D., 2005. Is homoplasy or lineage sorting the source of incongruent mtdna and nuclear gene trees in the stiff-tailed ducks (*Nomonyx Oxyura*)? Systematic biology 54, 35-55.

McDade, L., 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution, 1685-700.

McDade, L.A., 1997. Hybrids and Phylogenetic Systematics III. Comparison with Distance Methods. Systematic Botany 22, 669-83.

Merendino, M.T., Ankney, C.D., Dennis, D.G., 1993. Increasing mallards, decreasing American black ducks: more evidence for cause and effect. The Journal of Wildlife Management 57, 199-208.

Omland, K.E., 1997a. Correlated rates of molecular and morphological evolution. Evolution, 1381-93.

Omland, K.E., 1997b. Examining two standard assumptions of ancestral reconstructions: Repeated loss of dichromatism in dabbling ducks (*Anatini*). Evolution 51, 1636-46.

Orr, M.R., Smith, T.B., 1998. Ecology and speciation. Trends in Ecology & Evolution 13, 502-6.

Oyler-McCance, S.J., John, J.S., Quinn, T.W., 2010. Rapid evolution in lekking grouse: implications for taxonomic definitions. Ornithological Monographs 67, 114-22.

Palmer, R.S., 1976. Handbook of North American Birds, New Haven, Connecticut.

Pérez-Arteaga, A., Gaston, K., 2004. Wildfowl population trends in Mexico, 1961-2000: a basis for conservation planning. Biological Conservation 115, 343-55.

Perez-Arteaga, A., Gaston, K.J., Kershaw, M., 2002. Population trends and priority conservation sites for Mexican duck *Anas diazi*. Bird Conservation International 12, 35-52.

Peters, J.L., Sonsthagen, S.A., Lavretsky, P., Rezsutek, M., Johnson, W.P., Mccracken. K. G., 2014a. Interspecific hybridization contributes to high genetic diversity and apparent effective population size in an endemic population of mottled ducks. Conservation Genetics 15, 509-520.

Peters, J.L., Winker, K., Millam, K., Lavretsky, P., Kulikova, I.,Wilson, R., Zhuravlev, Y., Mccracken. K. G., 2014b. Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus *Anas*). Molecular Ecology.

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

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

Peters, J.L., Zhuravlev, Y., Fefelov, I., Logie, A., 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.

Price, T.D., 2008. Speciation in Birds. Roberts and Company, Greenwood Village, CO.

Price, T.D., Qvarnström, A., Irwin, D.E., 2003. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences 270, 1433-40.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–59.

Rambaut, A., Drummond, A.J., 2009. Tracer v. 1.5., Available from http://beast.bio.ed.ac.uk/Tracer.

Rhymer, J.M., 2006. Extinction by hybridization and introgression in anatine ducks. Acta Zoologica Sinica 52, 583-6.

Rhymer, J.M., Simberloff, D., 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics 27, 83-109.

Scott, N.J., Jr., Reynolds, R.P., 1984. Phenotypic variation of the Mexican duck (*Anas platyrhynchos diazi*) in Mexico. The Condor 86, 266-74.

Short, L. L. 1969. Taxonomic aspects of avian hybridization. The Auk 86, 84-105.

Slatkin, M. 1987. Gene flow and the geographic structure of natural. Science 236, 787-792.

Sneath, P.H., Sokal, R.R., 1973. Numerical Taxonomy. The principles and practice of numerical classification. W. H. Freeman, San Francisco, CA.

Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12, 105-14.

Sorenson, M.D., Fleischer, R.C., 1996. Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proceedings of the National academy of Sciences of the United States of America 93, 15239-43.

Soutiere, E.C., 1986. Hand-reared mallard releases on 3 private farms in Maryland. In: Proceedings of the Annual Conferance Southeast Association, pp. 438-45. Fish and Wildlife Agencies.

Stephens, M., Donnelly P., 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. The American Journal of Human Genetics 73, 1162-9.

Stutzenbaker, C.D., 1988. The mottled duck, its life history, ecology and management. (ed. by Department Texas Parks & Wildlife Department), Austin, Texas.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution. 28.10, 2731-2739.

Waltari, E., Hijmans, R.J., Peterson, A.T., Nyári, Á.S., Perkins, S.L., Guralnick, R.P., 2007. Locating pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PLoS ONE 2, e563.

Webster, W.M., 1980.. Iowa State University Press, Ames, Iowa.

Wiens, J.J., 2001. Widespread The Island Waterfowl loss of sexually selected traits: how the peacock lost its spots. Trends in Ecology & Evolution 16, 517-23.

Williams, C.L., Brust, R.C., Fendley, T.T., Tiller, G.R., Rhodes, O.E., 2005a. A comparison of hybridization between mottled ducks (*Anas fulvigula*) and mallards (*A. platyrhynchos*) in Florida and South Carolina using microsatellite DNA analysis. Conservation Genetics 6 445-53.

Williams, C.L., Fedynich, A.M., Pence, D.B., Rhodes, O.E., 2005b. Evaluation of allozyme and microsatellite variation in Texas and Florida mottled ducks. The Condor 107, 155-61.

Winker, K. 2009. Reuniting phenotype and genotype in biodiversity research. BioScience 59, 657-665.

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

Wu, C. I. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology 14, 851-865.

Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17, 2107-21.

CHAPTER III. SPECIATION GENOMICS AND A ROLE FOR THE SEX CHROMOSOME IN THE MALLARD AND MEXICAN DUCK

Abstract – Speciation is a continuous and dynamic process. Distinguishing between evolutionary forces influencing the speciation process can be effectively achieved by studying organisms at early stages of divergence. I conducted genomic scans across the mitochondrial DNA (mtDNA) control region, 3563 autosomal loci, and 172 Z-sex chromosome loci in the mallard (*Anas platyrhynchos*; $N = 17$) and Mexican duck (*A.* [*p.*] *diazi*; $N = 105$ individuals from six Mexican and two US states) to determine their evolutionary relationship. Between mallards and Mexican ducks, divergence estimates varied across autosomal ($\Phi_{ST} = 0.014$), Z-linked ($\Phi_{ST} = 0.091$), and mtDNA ($\Phi_{ST} = 0.12$) markers. Whereas population structure between mallards and Mexican ducks at autosomal markers was consistent with a stepping-stone model of divergence, Z-linked loci followed a two-island model of divergence with few loci under positive selection having large effects. In contrast, divergence at autosomal (mean $\Phi_{ST} = 0.012$) and Z-linked markers (mean $\Phi_{ST} = 0.018$) were tightly correlated among Mexican duck sampling groups. I conclude that speciation between mallards and Mexican ducks is likely proceeding via selection on a few sex-linked markers, whereas divergence at the remaining genome, as well as among Mexican duck sampling groups, is largely driven by genetic drift. I highlight how analyzing and comparing different marker-types can reveal the differential roles of selection and genetic drift across recently diverged genomes.

INTRODUCTION

Speciation proceeds in stages along a continuum from initiation to completion (Hohenlohe et al., 2010; Nosil et al., 2009a; Nosil et al., 2009b). Establishing how and why populations diverge, including the primary mechanisms (e.g., selection, genetic drift, and gene flow) influencing these events, is a key objective for evolutionary biologists (Mayr, 1982; Seehausen, 2004; Wolf et al., 2010). Moreover, determining whether divergence is driven by a few genes with large effects or many genes with small effects is of particular interest (Orr, 2001; Templeton, 1981; Wu, 2001). In general, simulated and empirical data suggest that the number and effect of genes driving divergence at early (incipient) stages corresponds to the extent of isolation (or gene-flow) (Feder et al., 2012; Seehausen et al., 2014). For instance, allopatric populations likely accumulate genome-wide divergence driven by many loci having smaller effects and via genetic drift, whereas those experiencing gene flow (i.e., parapatric, sympatric) are more likely to speciate at a few highly selected on genes with large pleiotropic effects (Andrew and Rieseberg, 2013; Feder et al., 2012; Martin et al., 2013; Nosil and Feder, 2013; Rockman, 2012; Stölting et al., 2013; Yeaman and Whitlock, 2011). Disentangling between the antagonistic forces of selection, genetic drift, and gene flow can be achieved by studying how genomes are shaped early in the speciation process (Coyne and Orr, 2004; Dobzhansky, 1940; Schluter, 2009).

Advances in next-generation sequencing (NGS) techniques are transforming the field of speciation genomics by providing cost-effective methods to attain genomic insight across non-model organisms (Ellegren, 2008; Stapley et al., 2010). Studies are

revealing how divergent patterns are localized across genomes, determining the mechanisms driving patterns across taxonomic lineages, and using the overall genomic signal to determine stages of speciation (Nosil and Feder, 2013; Seehausen et al., 2014). In particular, genomic surveys can detect regions associated with speciation (i.e., speciation genes) and the mechanistic cause(s) of genetic heterogeneity (Nosil and Feder, 2013; Nosil and Schluter, 2011; Seehausen et al., 2014). Currently, results across studies have varied, suggesting that evolutionary and demographic factors that are specific to each study play an integral role and need to be carefully considered (reviewed in Seehausen et al., 2014).

Given the possible heterogeneous nature of any single genome, comparisons across marker-types (i.e., autosomal, sex-linked, mtDNA) are important in understanding the cause of any discordance among genetic, as well as phenotypic markers that can arise during radiations (Edwards et al., 2005; Pryke and Griffith, 2009; Winker, 2009). Interestingly, results across various genera have identified divergent properties frequently arising on sex chromosomes (e.g., birds (Minvielle et al., 2000; Pryke, 2010; Sæther et al., 2007), insects (Martin et al., 2013; Phadnis and Orr, 2009), mammals (Sutter et al., 2013; Tucker et al., 1992)), and particularly, at the incipient stage (Frank, 1991; Haldane, 1948; Phadnis and Orr, 2009; Reeve and Pfennig, 2003). To date, important isolating mechanisms, such as male sterility, sexually selected male plumage traits, assortative mating, and post-mating isolation have been linked to sex chromosomes (Abbott et al., 2013; Carling and Brumfield, 2009; Minvielle et al., 2000; Pease and Hahn, 2013; Phadnis and Orr, 2009; Pryke, 2010; Sæther et al., 2007; Stölting et al., 2013; Turelli and Moyle, 2007). In this study, I explore the genomes of two incipient and parapatric taxa in

an attempt to differentiate among evolutionary mechanisms and to determine the presence, number, and location of potential diverging elements.

Study System

The dichromatic mallard (*Anas platyrhynchos*) and its close monochromatic relative, the Mexican duck (*A.* [*p.*] *diazi*) are a part of the recently radiated mallard complex (Johnsgard, 1978; Palmer, 1976). Incomplete lineage sorting and widespread interspecific hybridization have made resolving relationships within this group difficult, particularly with respect to the four species/subspecies from the New World (Avise et al., 1990; Johnson and Sorenson, 1999; Lavretsky et al., 2014; McCracken et al., 2001). Indeed, phenotypic, mitochondrial, and nuclear markers support different species tree topologies, revealing mito-nuclear-phenotypic discord (Johnson and Sorenson, 1999; Lavretsky et al., *in press;* Lavretsky et al., 2014; Livezey, 1991).

Whereas mallards have a Holarctic distribution and are migratory, Mexican ducks are non-migratory and endemic to North America from southwestern US (i.e., Arizona, New Mexico, and Texas) and extending southward into the central highlands of Mexico (Aldrich and Baer, 1970; Bellrose, 1976; Stutzenbaker, 1988). Mexican ducks have gone through several taxonomic revisions due to the observed variance in mallard-like plumage expression across sampled populations (AOU, 1983; Conover, 1922; Hubbard, 1977; Huber, 1920; Ridgway, 1886). A recent proposition for taxonomic reevaluation to establish them as a single monotypic *A. diazi* species was considered based on mtDNA results (McCracken et al., 2001), but remained unchanged (Chesser et al., 2011; AOU petition 2010-B-6) due to insufficient knowledge about hybridization levels between Mexican ducks and mallards (Scott and Reynolds, 1984). Moreover, the five-fold

difference in divergence estimates that was recently reported between mitochondrial DNA (mtDNA; Φ_{ST} = 0.11) and nuclear DNA (nuDNA; mean Φ_{ST} across 17 introns = 0.020) (Lavretsky et al., *in press*) suggests that multi-marker comparisons are necessary.

Historically, mallards naturally wintered in the Mexican duck's most northern range (Palmer, 1976), and recently have become a regular and limited breeder in southwestern US as a result of introductions into urban settings and on shooting preserves for sport hunting. Consequently, opportunity for hybridization between mallards and Mexican ducks has been and continues to be a potentially important force. Although, pre-mating barriers (i.e., segregated courting groups and timing of pair formation) between Mexican ducks and mallards have been suggested (Aldrich and Baer, 1970; Palmer, 1976), establishing the frequency of hybridization across the Mexican duck's range, as well as the connectivity between Mexican duck populations that could facilitate mallard alleles to percolate across their range is required to understand the possible implication(s) of introgression on speciation and management of these two taxa (Aldrich and Baer, 1970; Pérez-Arteaga and Gaston, 2004; Perez-Arteaga et al., 2002).

The objective of this study is to determine the heterogeneous nature of the mallard and Mexican duck genomes to infer mechanisms of divergence, including historical and contemporary levels of introgression. Specifically, I address the following aims.

1. Conduct a genomic scan to determine the number and distribution of divergent loci between Mexican ducks and mallards. Given that the two species are phenotypically diagnosable yet genetically indistinguishable (Lavretsky et al., *in press*), and they likely have experienced a parapatric history, I predict that they are diverging at a few key genes/genic regions. Specifically, I will identify loci in two separate analyses, including

those that contribute towards population structure and ones identified as under selection. First, the number of markers needed to recover population structure will discern whether divergence is due to a few loci with large effects or many loci with smaller effects. Next, if analyses testing for selection identify the same loci, then this would support that population structure is driven by markers under selection. In general, between Mexican ducks and mallards, I expect few loci having high loadings (i.e., effect) and under selection. In contrast, among Mexican duck sampling groups, I expect few (if any) loci having large effects, but rather lots of loci with smaller effects and none of which to be under selection.

2. For a finer examination of population structure, I will separately analyze autosomal, Z-linked, and mtDNA markers. If genomic scans (Objective 1) reveal marker-specific variance in genomic divergence, then I expect to recover population structure following these differences. For example, if speciation is largely driven by selection on the sex chromosome (Carling and Brumfield, 2009; Minvielle et al., 2000; Pease and Hahn, 2013; Pryke, 2010; Sæther et al., 2007; Trier et al., 2014; Turelli and Moyle, 2007), then I expect Z-linked population structure to follow an island model of divergence in which genetic variation largely differentiates between mallards and Mexican ducks. Conversely, if genetic drift and/or introgression (Lavretsky et al., *in press*) is primarily influencing molecular variance, then I expect comparable population structure regardless of marker type.

3. Finally, being the most comprehensive genetic analysis of Mexican ducks, I will use divergent patterns across Mexican duck sampling locations to establish whether Mexican ducks are a single continuous population (or isolated by distance), and

determine the effect of hybridization with mallards across their range. Given their current geographic association with mallards, I expect hybrids to be most frequent in the northern portion of the Mexican duck's range.

MATERIALS AND METHODS

Sampling and DNA Extraction

I sampled 105 Mexican ducks from six Mexican (N = 92) and two US states (N = 13) and 17 mallards throughout North America (Fig. 3.1; Appendix Table A3.1). Genomic DNA was extracted using a DNeasy Blood & Tissue kit (Qiagen, Valencia, CA). Before preparing libraries, all extractions were quantified using a nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc.) to ensure a minimum concentration of 0.02 µg/µL; samples failing this quality check were re-extracted.

ddRADseq Library Preparation

Sample preparation for RAD sequencing followed protocols outlined in DaCosta and Sorenson (*in review*). In brief, ~1 ug of genomic DNA was double digested using 10 U of restriction enzymes SbfI and EcoRI. Adapters containing sequences compatible for Illumina sequencing and barcodes for de-multiplexing reads were ligated to the sticky ends generated by the restriction enzymes. The adapter-ligated DNA fragments were then size-selected using gel electrophoresis (2% low-melt agarose) and a MinElute gel extraction kit (Qiagen, Valencia, CA). Fragments of 300-450 bp were selected, but fragments as small as ~40 bp are reliably captured using this method (see DaCosta and Sorenson, *in review*). Size-selected fragments were then amplified using a polymerase

chain reaction (PCR) with Phusion high-fidelity DNA polymerase (Thermo Scientific, Pittsburgh, PA), and the amplified products were cleaned using magnetic AMPure XP beads (Beckman Coulter, Inc., Indianapolis, IN). The concentration of purified PCR products was estimated with quantitative PCR using an Illumina library quantification kit (KAPA Biosystems, Wilmington, MA), and samples with compatible barcode combinations were pooled in equimolar concentrations. Multiplexed libraries were sequenced on an Illumina HiSeq 2000 at the University of California-Berkley Vincent J. Coates Genomics Sequencing laboratory.

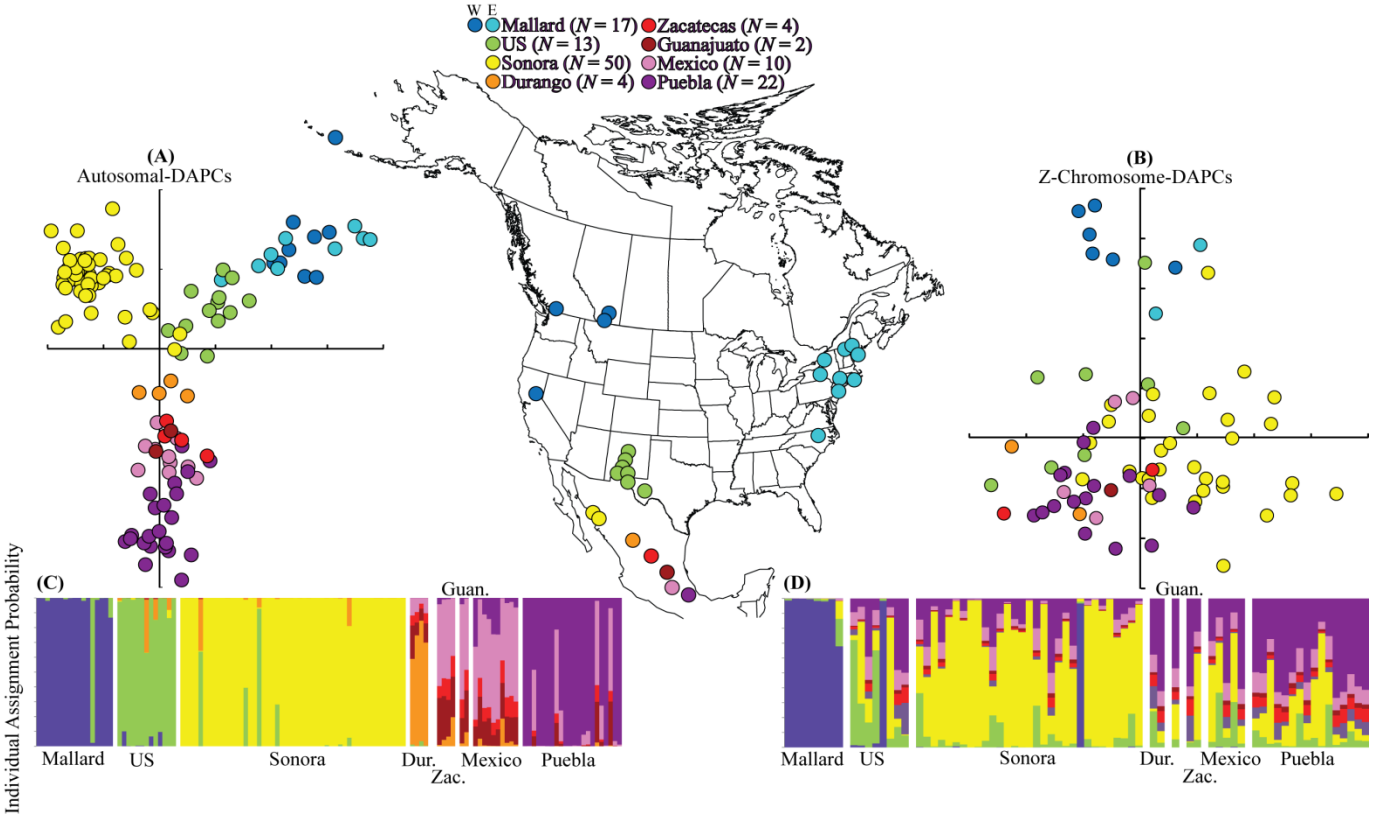


Figure 3.1. Sampling locations of Mexican ducks and mallards and results of $DAPC_S$ obtained from 3,695 RADseq markers. The map displays sample locations color-coded by sample group (Appendix Table A3.1; *N* = number of samples). Discriminant functions 1 (x-axis) and 2 (y-axis) from $DAPC_S$ are plotted for (A) 3,523 autosomal (*N* = 105 Mexican ducks and 17 mallards) and (B) 172 Z loci (*N* = 64 Mexican ducks and 8 mallards; males only, because adegenet does not accommodate heterogamy). Population assignment posterior probabilities are based on (C) autosomal and (D) Z loci. Colors correspond to those shown in the sampling map.

Bioinformatics of ddRADseq Data

Raw Illumina reads were processed using a pipeline described by DaCosta & Sorenson (*in review*). Custom Python scripts used in the pipeline are available at http://github.com/BU-RAD-seq/ ddRAD-seq-Pipeline. Reads were assigned to individual samples based on barcode sequences. For each sample, low quality reads were filtered and identical reads were collapsed (maintaining a read count and the highest quality score at each position). Filtering was achieved using the UCLUST function in USEARCH v. 5 (Edgar, 2010), with reads that were >10% divergent and an average Phred score < 20 being removed from the data set. Condensed and filtered reads from all samples were concatenated and clustered with an –id setting of 0.85 in UCLUST. The highest quality read from each cluster was mapped to the mallard reference genome (accession numbers SS263068950 - SS263191362; Huang et al., 2013; Kraus et al., 2011) using BLASTN v. 2 (Altschul et al., 1990), and clusters with similar BLAST hits were combined. The reads within each cluster (i.e. putative loci) were aligned using MUSCLE v. 3 (Edgar, 2004), and samples within each aligned cluster were genotyped using the Python script *RADGenotypes.py*. Alignments with end gaps due to indels and/or a polymorphism in one of the restriction sites were either automatically trimmed or flagged for manual editing during genotyping. Alignments with ≥2 polymorphisms in the first or last five base-pairs were also flagged for manual inspection. Polymorphisms were scored using read depths for major and minor alleles and a population-aware algorithm (i.e. more skewed major:minor allele depths were allowed if the minor allele was present in other samples). Individual genotypes fall into four general categories: "missing" (no data), "good" (unambiguously genotyped), "low depth" (recovered data, but could not reliably score as

homozygous or heterozygous because of low depth), and "flagged" (recovered heterozygous genotype, but with counts of major and other alleles below acceptable thresholds). Loci with <10% missing genotypes and ≤6 flagged genotypes were retained for downstream analyses. Unlike other protocols (e.g., Stacks; Catchen et al., 2013), the developed pipeline retains loci containing indels and high variability by flagging them for manual editing. By including these flagged loci, I increased the total number of retained markers by ~15%, while reducing any bias resulting from discarding loci with indels or high variability.

Although most loci generated a BLAST hit on the mallard reference genome, the current build of this genome (v1.0) contains 78,487 contigs that are not yet assigned to chromosomes. I therefore categorized ddRADseq loci as either autosomal or Z-linked based on two criteria. First, all loci were BLASTed to the reference chicken genome (*Gallus gallus*; accession numbers PRJNA10807-08, PRJNA13342, PRJNA202483). BLAST results against mallard and chicken genomes were used to discover mallard contigs that are likely part of the Z chromosome, and all loci with BLAST hits to these mallard contigs were categorized as Z-linked. Cross-validation of sex versus autosomal chromosome assignment was based on depth and homozygosity across markers (Appendix Fig. A3.1). Because females are heterogametic for the Z chromosome, Z-linked markers from females will have no heterozygosity and about one half the depth of males. I therefore also used sex-specific depth and heterozygosity information to cross-reference loci assigned to the Z chromosome based on mallard and chicken BLAST results. While birds are known to exhibit strong genomic synteny (Backström et al., 2008;

Grant and Grant, 1997; Nanda et al., 2008), cross-validation of marker assignment will only be possible once the mallard genome is assembled into chromosomes.

Mitochondrial DNA

Previously optimized primers (L78 and H774) were used to isolate 653 bp of the mtDNA control region across Mexican duck samples (Sorenson et al., 1999; Sorenson and Fleischer, 1996) and were amplified with PCR using 1.5 µL of template DNA (10 ng/ µL), 2x GoTaq Green Master Mix (Promega), and 1.0 nM of each primer, in a total volume of 15 µL. PCR conditions, amplification verification, and subsequent sample prep for sequencing followed methods described in Lavretsky et al. (2014). Final products were sent to the Yale University DNA Analysis Facility for automated sequencing on an ABI 3730. Sequences were aligned and edited using Sequencher v. 4.8 (Gene Codes, Inc). All sequences have been submitted to GenBank (*accession numbers pending*). For mallards, 25 previously published sequences (*accession numbers* KF608514-518 (Lavretsky et al., 2014); KF857589, KF857591, KF857593, KF857596, KF857598, KF857599, KF857600-603, KF857606-607, KF857627, KF857635-636, KF857641-642, KF857644, KF857646, KF857649 (Peters et al., 2014)) were used as representative of the species (Lavretsky et al., *in press*). Finally, DNA sequences were converted to the Roehl format in DnaSP v. 5 (Librado and Rozas, 2009) and used to reconstruct a median-joining haplotype network (Bandelt et al., 1999) as implemented in Network Publisher (Fluxus Technology).

General Population Genetics and Outlier Locus Analyses

Pair-wise population Φ_{ST} estimates for all marker types (i.e. mtDNA, autosomal ddRADseq, and Z-linked ddRADseq) were calculated with Arlequin 3.1 (Excoffier and

Lischer, 2010) between Mexican ducks and mallards, as well as among eight Mexican duck sampling groups (Appendix Table A3.1). Estimates of nucleotide diversity (π) were obtained using DnaSP v. 5 (Librado and Rozas, 2009).

Outliers across ddRADseq markers were independently detected using two different procedures. Both analyses were conducted for Mexican ducks (all sampling locations combined) versus mallards and among Mexican duck sampling groups. First, I used probabilistic models employed in the discriminant analysis of principal components ($DAPC_S$) as implemented in the package Adegent v. 1.3.5 (Jombart, 2008; Jombart et al., 2010) in R (R Development Core Team 2013) to identify autosomal or Z-linked markers that had a significant contribution to population structure (i.e., outlier analysis). Specifically, locus contributions (i.e., loadings) were assessed after retaining an optimum number of principle components (PCs) and eigenvalues per analysis (see below for specifics). For Mexican ducks versus mallards, all individuals were assigned as either "Mexican duck" or "mallard" *a priori*, whereas individuals were assigned to their respective sampling group (Appendix Table A3.1) *a priori* for the within Mexican duck analysis. Unfortunately, $DAPC_S$ analyses are sensitive to missing data and so only males (the sex with two copies of the Z chromosome) were included in the genomic surveys to ensure direct comparison of autosomal and Z-linked marker contribution to population structure. All loci with substantial contributions (Loading $\geq$ 0.002) were tested for linkage disequilibrium in Arlequin v. 3.1 (Excoffier and Lischer, 2010) to ensure that population structure was unbiased.

Second, I used *BayeScan* v. 2.1 (Foll and Gaggiotti, 2008) to identify outlier loci that are likely under selection between Mexican ducks versus mallards and among

Mexican duck sampling groups – individuals were assigned to "populations" *a priori* matching those in $DAPC_S$ analyses. *BayeScan* uses a reversible-jump MCMC to assign posterior probability support to each locus by comparing F_{ST} distribution models with and without selection, and simultaneously distinguishes between positive ($\alpha > 0$) and balancing/purifying ($\alpha < 0$) selection. Using simulated data, Pérez-Figueroa et al. (2010) demonstrated that *BayeScan* was efficient in detecting outlier loci with relatively low rates of false positives (< 1%), particularly when analyzing closely related taxa, with average genomic divergence (F_{ST}) of $\leq 2.5\%$. The analyses were run with default settings that included 20 pilot runs, each a length of 5,000 steps, followed by 50,000 burn-in and 5,000 sampling steps with a thinning interval of 10. The prior odd for the neutral model was set at 10. Posterior distributions for all parameters were analyzed for efficient mixing using the provided *plot* R function. Finally, outliers were identified using a false discovery rate (FDR) of 0.01 with the *plot_bayescan* R function. For direct comparison to $DAPC_S$ results, only males were used and both autosomal and Z-linked loci were analyzed together, which also ensured that outlier identification was against the overall genomic background. However, analyses were repeated with both males and females to test for outlier correspondence and the sensitivity of *BayeScan* to "missing" Z-linked data in females.

Population Structure Within and Between Mexican Ducks and Mallards

Given the amount of data generated by NGS methods, Bayesian clustering algorithms (e.g., STRUCTURE; Pritchard et al., 2000) appear to be ineffective at resolving large datasets, particularly in systems under non-island models (e.g., stepping stone model; Nei, 1972) (Jombart et al., 2010). I therefore once again used the

multivariate $DAPC_S$ analysis to assess genomic structure among Mexican ducks and between Mexican ducks and mallards. Rather than plotting between two variables that describe the most variance (i.e., principal component analyses), $DAPC_S$ achieves an optimum number of PCs when discriminating among individuals by simultaneously maximizing between-group variation while minimizing within-group variation across PCs. Consequently, $DAPC_S$ analyses are not sensitive to underlying family structure, and related individuals do not need to be removed (Jombart et al., 2010). To minimize the bias of over-fitting the model through the inclusion of too many PCs, the proportion of successful assignments (i.e., assignment proportions from observed discriminations / random discriminations) was maximized and corrected for the number of retained PCs using the *optim.a.score* function. Moreover, retention of discriminant functions (i.e., eigenvalues) was based on the minimum number of eigenvalues that effectively captured the genetic structure within the data (Jombart et al., 2010). All individuals were assigned to their respective *a priori* sample group, or "populations" described in Appendix Table A3.1, as this allowed us to examine how individuals among the various groups were genetically related.

Isolation-By-Distance

Correlations between genetic and geographic distance was tested using a simple Mantel's test as implemented in the zt program (Bonnet and Van de Peer, 2002). Specifically, I tested for correlations between geographic distance and each set of pair-wise Φ_{ST} estimates derived from mtDNA, Z-linked, or autosomal markers. Distance was calculated using the average latitude and longitude among individuals for each sampling

group (see "populations" in Appendix Table A3.1). Analyses were run for 100,000 iterations.

RESULTS

Nuclear Divergence and Outlier Loci

After quality-filtering, ddRADseq recovered 3695 variable loci, with 3523 assigned to autosomal and 172 to the Z-sex chromosome. Between mallards and Mexican ducks, an average Φ_{ST} of 5.2% was recovered across ddRADseq loci (Fig. 3.2), however, Z-linked (Φ_{ST} = 0.091) markers were 6.5 times more differentiated then autosomal loci (Φ_{ST} = 0.014) (Fig. 3.2). While, the overall distribution of Φ_{ST} estimates revealed an exponential decrease in the number of highly divergent loci for both marker types, the Z chromosome possessed an extended tail of divergent loci (Fig. 3.3). In contrast, Φ_{ST} estimates from autosomal (mean Φ_{ST} = 0.012; ± 0.006 StDev) and Z-linked loci (mean Φ_{ST} = 0.018; ± 0.021 StDev) were similar across Mexican duck pair-wise comparisons (Fig. 3.2). Finally, nucleotide diversity was similar between mallards and Mexican ducks and among sampling locations. However, autosomal loci had ~2-3

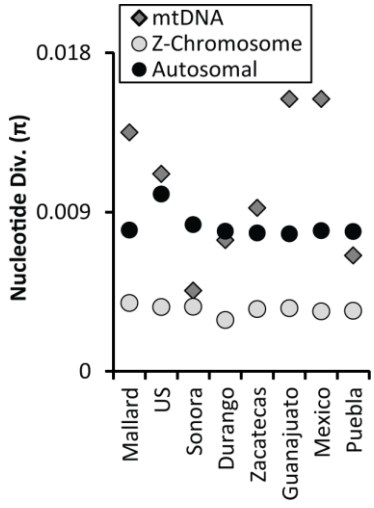


Figure 3.4. Nucleotide diversity for the mitochondrial (mtDNA) control region, 172 Z-chromosome loci, and 3523 autosomal loci for mallards and seven Mexican duck sampling locations.

times more nucleotide diversity than Z-linked markers (Fig. 3.4).

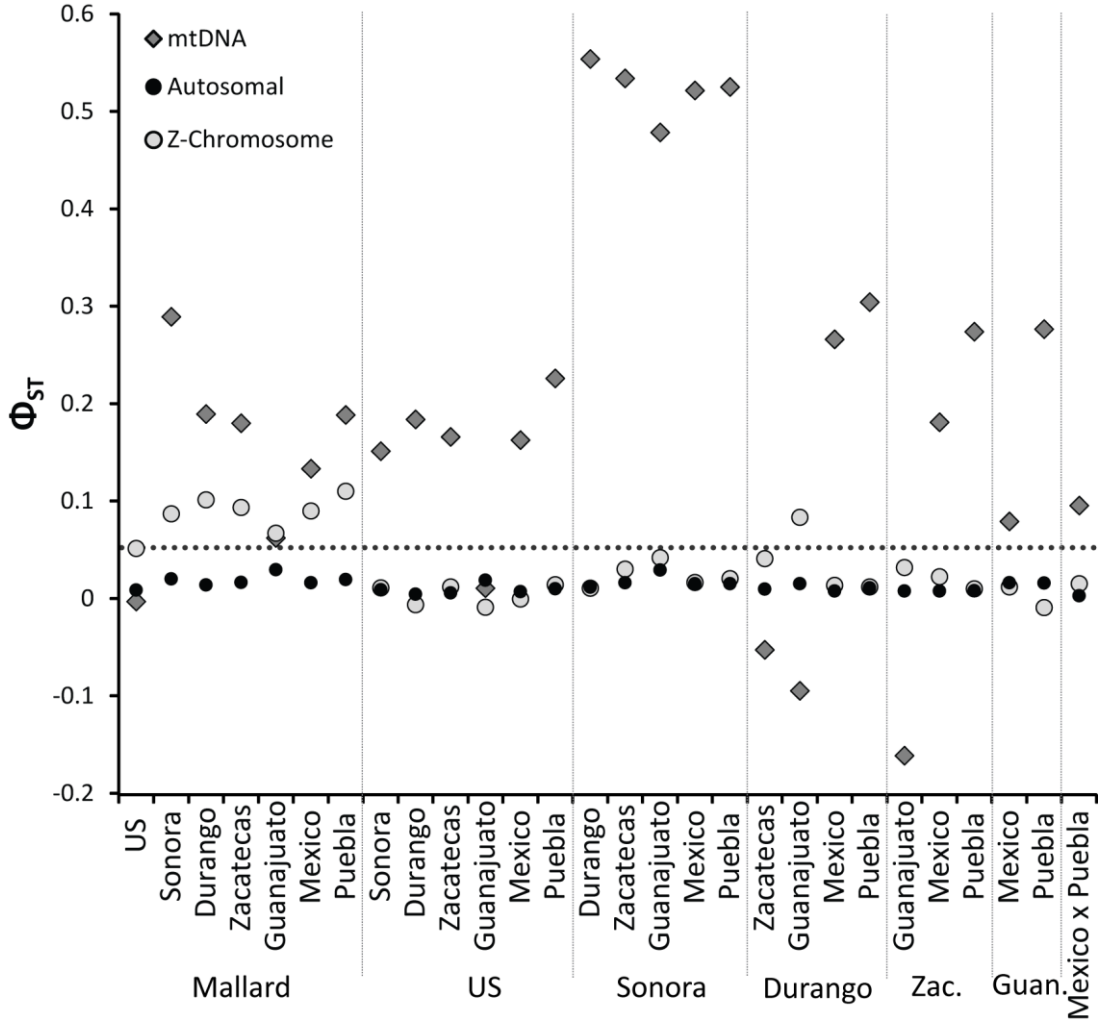


Figure 3.2. Average pair-wise Φ_{ST} estimates for the mtDNA control region, 3,523 autosomal loci, and 172 Z loci for mallards and seven Mexican duck sample groups (see Fig. 3.1). The dotted line denotes the average Φ_{ST} (0.052) between mallards and Mexican ducks across all 3,695 RADseq loci.

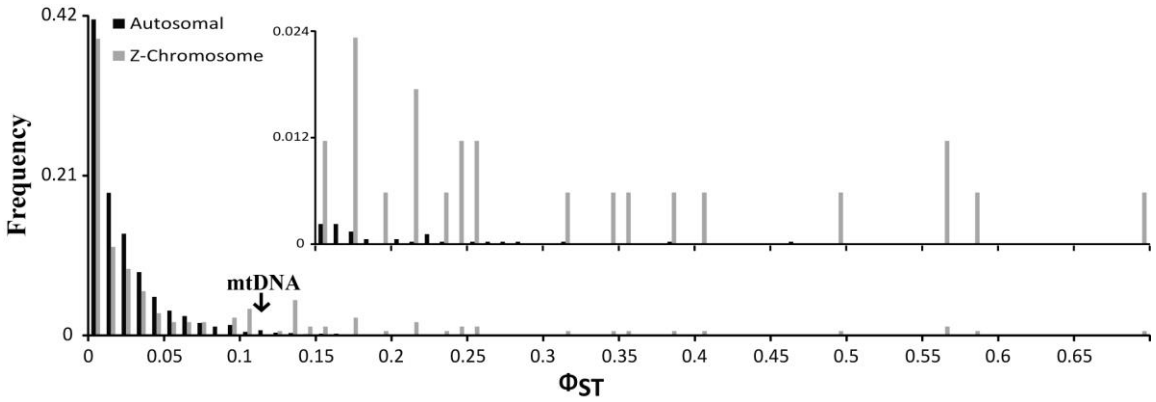


Figure 3.3. Frequency distribution of Φ_{ST} estimates between mallards and Mexican ducks across 3523 autosomal loci and 172 Z-linked loci. Average mtDNA Φ_{ST} estimate between the two taxa (Φ_{ST} = 0.12) is indicated by the arrow. Inset includes frequency distribution of Φ_{ST} estimates from 0.15 - 0.70.

For Mexican duck versus mallard $DAPC_S$ analysis, I retained 7 PCs and only one eigenvalue – which was the maximum given that only two populations were compared – that explained 12.5% of the variance (Appendix Fig. A3.2). Only one autosomal and four Z-linked markers, none of which were in linkage disequilibrium, had significant contribution to population structure between Mexican ducks and mallards (Fig. 3.5). *BayeScan* analysis of males identified two autosomal and four Z-linked markers likely under positive/diversifying selection, and an additional autosomal locus likely influenced by purifying selection (Fig. 3.5). Importantly, all markers with the highest $DAPC_S$ loadings between the two taxa were identified by *BayeScan* to be under positive/diversifying selection. Interestingly, *BayeScan* results including/excluding females were nearly identical (Appendix Table A3.2), suggesting that *BayeScan* is not sensitive to missing data like $DAPC_S$ analyses (Appendix Fig. A3.3). However, while recovering outlier autosomal markers including/excluding Z-linked markers were

reproducible, doing so with Z-linked markers only were not (Appendix Table A3.2; Fig. A3.3). Specifically, *BayeScan* results reveal that between Mexican ducks and mallards, the Z chromosome is significantly diverged and distinguishing Z-linked loci under positive/balancing selection against the elevated Z-chromosome background is difficult (Appendix Fig. A3.3). This result demonstrates the importance of comparing individual loci against the overall genomic background. Finally, running a BLAST search in GenBank, I recovered two of the four putatively selected Z-linked genes to be annotated and functional for a Zinc-finger domain (accession number KB743159) and a Kinase involved in riboflavin biosynthesis (accession number KB742655).

For the Mexican duck $DAPC_S$ analysis, I retained 7 PCs and two eigenvalues that explained 13.6% of the variance (Appendix Fig. A3.2). Although, three autosomal markers, which were not in linkage disequilibrium, had significant contributions to population structure, markers in general had small and similar contributions across the genome (Fig. 3.5). *BayeScan* recovered eight autosomal markers, all of which were likely under positive/balancing selection, including one of the two identified in $DAPC_S$ analyses (Fig. 3.5). When analyzing the Z-chromosome only, and regardless of excluding/including females, I did not identify any markers to be under selection (Appendix Fig. A3.3), which corresponds to $DAPC_S$ results (Fig. 3.5). Once again, excluding/including females did not change results (Appendix Table A3.2; Fig. A3.3).

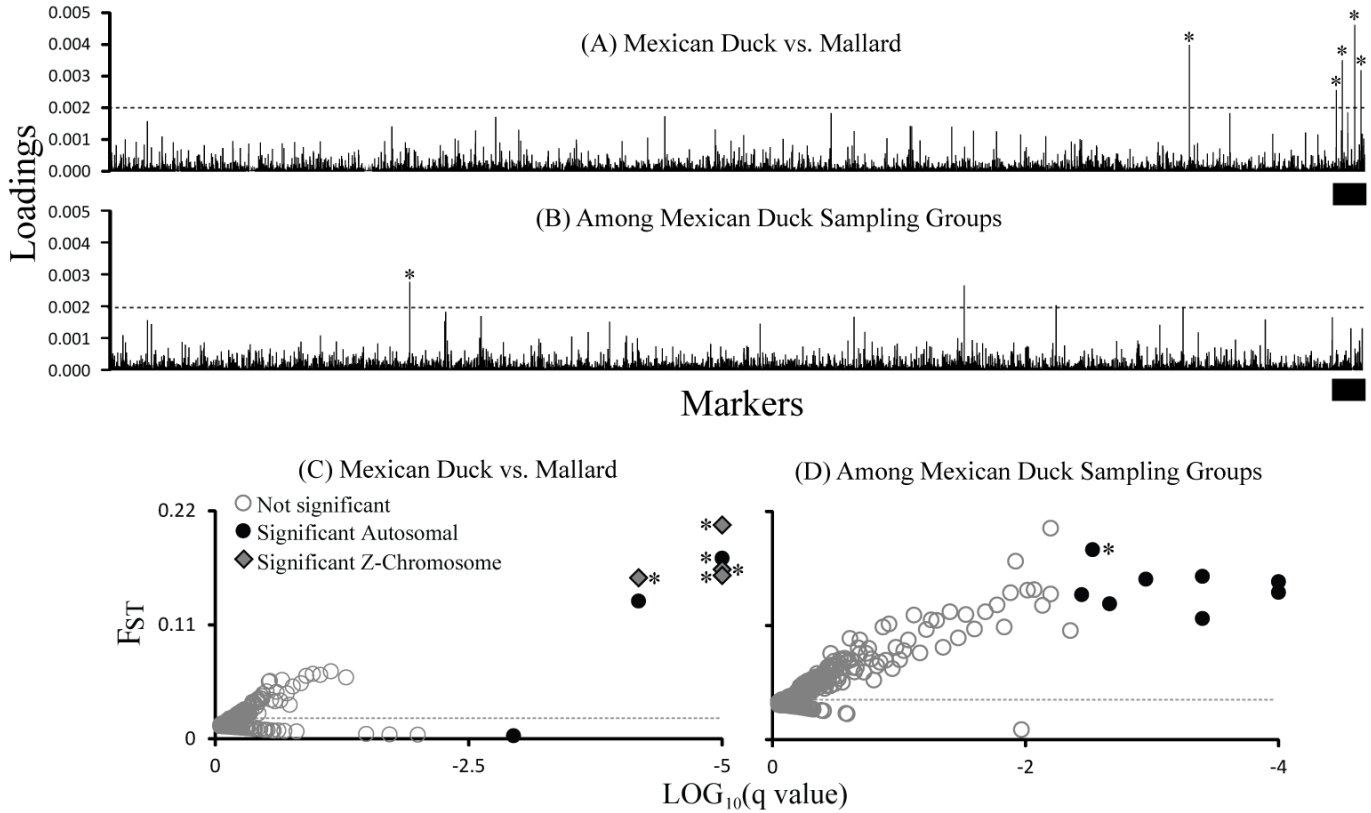


Figure 3.5. Contribution (i.e., Loadings) of 3523 autosomal and 172 Z-chromosome (above thick black line) loci to population structure (A) between mallards and Mexican ducks and (B) among Mexican duck sampling groups – bars extending above the dotted line denote a significant contribution (Loading ≥ 0.002). *BayeScan* outlier results are provided (C) between mallards and Mexican ducks, and (D) within Mexican duck sampling groups – the dotted line denotes loci under diversifying (above) or purifying (below) selection. Asterisks correspond to the same markers identified between the paired analyses.

Nuclear DNA Population Structure

For the autosomal $DAPC_S$ analysis that included males and females, and samples identified by "population" for both mallards and Mexican ducks (Appendix Table A3.1), I retained 11 PCs and two eigenvalues that explained 13.1% of the variance (Appendix Fig. A3.4). Plotting the two retained discriminant functions uncovered structure that followed a stepping-stone model of divergence (Fig. 3.1A). Moreover, I found a significant correlation between genetic divergence and geographic distance (Mantel's test; $r = 0.50$; $p < 0.05$), supporting isolation-by-distance. In contrast to the isolation-by-distance observed among Mexican duck groups, there was no indication of substructure among western and eastern mallards (Fig. 3.1; see also Kraus et al., 2013).

Once again, because $DAPC_S$ is sensitive to missing data, only males (the sex with two copies of the Z chromosome) were analyzed for Z-linked population structure. I am confident that excluding females did not bias overall population structure (note that outlier analyses including/excluding females do not substantially change results; Appendix Table A3.2; Fig. A3.3). The optimum number of PCs was one; however, to run analyses I retained two PCs – note optimization scores did not differ between the retention of one or two PCs – that explained 11.3% of the variance (Appendix Fig. A3.4). In contrast to structure recovered with autosomal markers, plotting the two retained discriminant functions primarily differentiated mallards from Mexican ducks (Fig. 3.1B). Furthermore, a non-significant Mantel's test between Z-linked marker divergence and geographical distance ($r = 0.65$; $p > 0.05$) suggests that genetic structure does not follow an isolation-by-distance model. Moreover, there was no indication of population substructure among eastern and western mallards. Consequently, population structure at

Z-linked markers is consistent with a two-island model of divergence (Mexican duck versus mallard); although there was some evidence of the Sonoran sampling group being slightly differentiated from interior Mexican duck sampling groups (Fig. 3.1B; 3.1D). Furthermore, a single Mexican duck from US and Sonoran sampling groups clustered with mallards (Fig. 3.1B).

Although corresponding to $DAPC_S$ results, individual posterior support revealed only three US Mexican ducks with admixture proportions that included "mallard" for autosomal markers (Fig. 3.1C). Individuals from Sonora, Durango, and Puebla were largely assigned to separate populations, whereas individuals from Guanajuato, Zacatecas, and Mexico were assigned to the same population with similar probabilities. Interestingly, one mallard from the east coast was assigned to the US Mexican duck population. In addition, two individuals from Sonora were assigned with US Mexican ducks, and four individuals from Puebla were assigned to the Guanajuato-Zacatecas-Mexico population. For Z-linked loci, all mallards were identified as a single population that also included one Mexican duck from each US and Sonora (Fig. 3.1D); only the US individual was also identified to include some "mallard" with autosomal markers (Fig. 3.1C). All remaining Mexican duck individuals were similarly admixed, although there was some evidence that US, Sonora, and interior locations comprised weakly differentiated populations (Fig. 3.1D).

Mitochondrial DNA Divergence within Mexican Ducks and Between Mallards

Of the three marker types, mtDNA was most differentiated between Mexican ducks and mallards (Φ_{ST} = 0.12). Within the haplotype network, two mtDNA haplogroups (A and B) that are characteristic of the mallard complex were recovered

(Ankney et al., 1986; Johnson and Sorenson, 1999; Kulikova et al., 2004; Lavretsky et al., 2014) (Fig. 3.6). Seven mallards and two US Mexican ducks were within haplogroup A. Although the majority of mallards and Mexican ducks were within haplogroup B, two mallards and Mexican ducks from Sonora (N = 38) and the US (N = 4) shared a divergent haplotype within haplogroup B (Fig. 3.6). In general, Mexican ducks possessed either unique haplotypes or shared haplotypes with individuals from the nearest sampled state (see Fig. 3.1); however, testing for an association between mtDNA divergence and geographic distance was not significant (Mantel's test; r = 0.095; $p \geq 0.05$), suggesting that mtDNA does not follow an isolation-by-distance model of divergence. With the exception of US Mexican ducks that were indistinguishable (Φ_{ST} = -0.0069) from mallards, Φ_{ST} recovered structure between mallards and each of the Mexican duck sampling groups (mean Φ_{ST} = 0.14 ± 0.095 StDev). Finally, Φ_{ST} values among Mexican duck sampling locations were as elevated (mean Φ_{ST} = 0.22; ± 0.22 StDev) relative to divergence between the two taxa (Fig. 3.2). I note that while the Sonoran sample group was most differentiated (mean Φ_{ST} = 0.52; ± 0.028 StDev), nucleotide diversity was the lowest relative to the remaining groups (Fig. 3.4).

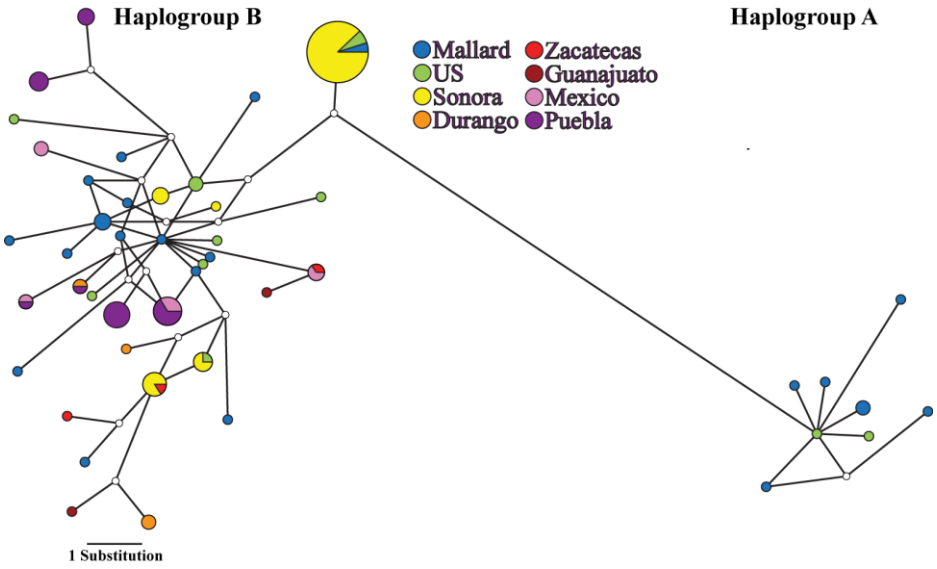


Figure 3.6. Mitochondrial DNA median-joining network – size of circles corresponds to total number of individuals with that haplotype and branch lengths indicate the number of mutations separating haplotypes. Population color codes correspond to Fig. 3.1.

DISCUSSION

Genomic Scans Identify Divergent (Speciation) Regions

Genomic scans revealed that Mexican ducks and mallards are at the earliest stages of divergence, with speciation likely driven by selection on the Z chromosome. In general, the distribution of Φ_{ST} values for both ddRADseq marker types (Fig. 3.3) fit expectations from simulations for "adjacent" populations with gene flow in which the number and extent of markers with the highest Φ_{ST} estimates is due to selection counteracting the effects of gene flow (Feder et al., 2012; Nosil et al., 2012). More specifically, relative to autosomal markers, the elevated divergence (Fig. 3.2; Appendix

Fig. A3.3) and broad frequency distribution of Φ_{ST} for Z-linked markers (Fig. 3.3) suggest that regions within the Z chromosome are likely under selection. Divergence between Mexican ducks and mallards can be attributed to a few genes with large effects on the Z chromosome (Fig. 3.5), all of which were also identified with *BayeScan* as being under positive/balancing selection when compared to the genomic background (Fig. 3.5). In contrast to the Z-chromosome, autosomal markers generally have uniformly low loadings (with the exception of one locus), suggesting that many loci with small effects contribute to autosomal differentiation between mallards and Mexican ducks; similar loadings were identified at both autosomal and Z-linked markers among Mexican duck sampling groups (Fig. 3.5). This "uniformity" in the frequency distributions of Φ_{ST} estimates and marker loadings is consistent with genetic drift primarily driving autosomal divergence between Mexican ducks and mallards, as well as both ddRADseq marker types within Mexican ducks (Feder et al., 2012; Nosil et al., 2012; Wu, 2001). In general, *BayeScan* analyses reveal that divergence is elevated at the Z chromosome as compared to autosomes (Appendix Fig. A3.3). Among Mexican duck sampling groups, I attribute the subtle Z chromosome divergence (Appendix Fig. A3.3) to the three-fourths effective population size causing faster sorting rates of Z-linked loci (i.e., genetic drift). In contrast, the substantially higher Z chromosome divergence between Mexican ducks and mallards is unlikely to be attributable to genetic drift alone, and is further support for selection on the Z chromosome playing an integral role in the speciation of these two taxa.

Population structure among mallards and Mexican duck populations revealed a stepping-stone model of divergence for autosomal DNA, whereas Z-linked markers

follow a two-island model of divergence (Fig. 3.1A & 1C). Specifically, for autosomal markers, the isolation-by-distance effect (Fig. 3.1A) and the uniformity of marker loadings (Fig. 3.5) are consistent with gradual changes in allelic frequencies from north to south. In contrast, Z-linked markers primarily differentiate Mexican ducks from mallards (Fig. 3.1C), suggesting a sharp transition in allelic frequencies that coincides with the geographic transition between the two species. Finally, as compared to the autosomal markers, the lower nucleotide diversity at the Z chromosome across mallards and Mexican ducks (Fig. 3.4) also suggests that selection may be maintaining lower, more taxon-specific diversity at the Z chromosome (Liu and Burke, 2006). Overall results from the two markers are consistent with expectations under a parapatric/sympatric existence, prolonged effects of gene flow, and a recent ancestry (Feder et al., 2012; Nosil et al., 2012).

Speciation Driven By the Sex Chromosome

I provide compelling evidence that speciation genes are likely present on the Z sex chromosome. Epistatic interactions between sex chromosomes and the remaining genomes of mallards and Mexican ducks would suggest the evolution of post-zygotic isolation that is consistent with Haldane's rule (i.e., at the incipient stage, the absence or decreased representation of the heterogametic sex suggests post-zygptic isolation; Haldane, 1922). A breeding experiment between mallards and American black ducks (*A. rubripes*) – another New World monochromatic taxon that is part of a phylogenetic polytomy with Mexican ducks and mallards (Lavretsky et al., *in press;* McCracken et al., 2001) – found a disproportionate number of viable male relative to female F_1 hybrids, suggesting these species fit Haldane's rule (Kirby et al., 2004). Given this cross-breeding

observation, as well as our genomic results, I predict that species integrity and evolution within the New World complex is likely proceeding via sex chromosomes. These results build upon mounting evidence that link important evolutionary mechanisms to sex chromosomes, and that sex chromosomes are likely hotspots for harboring speciation genes that maintain taxonomic integrity at the incipient stage (Andrew and Rieseberg, 2013; Feder et al., 2012; Martin et al., 2013; Nosil and Feder, 2013; Rockman, 2012; Stölting et al., 2013; Yeaman and Whitlock, 2011). Among these findings, divergence at sex chromosomes have resolved evolutionary relationships in which taxa are weakly, if at all, structured at nuclear markers but are morphologically identifiable (Axelsson et al., 2004; Kunte et al., 2011; Pryke, 2010; Reeve and Pfennig, 2003). To date, Mexican ducks and mallards have been indistinguishable at nuclear markers (Lavretsky et al., *in press*), while individuals are diagnosable via phenotypic characters (Hubbard, 1977; Scott and Reynolds, 1984). Thus, the phenotypic variance between Mexican ducks and mallards may be Z-linked. Putatively selected Z-linked genes include one coding for a zinc finger that is involved in facial development, fibroblast migration, skeletal system morphogenesis, hair follicle development, and one coding for riboflavin kinase, which was also noted to be involved in maintenance of morphological features. This preliminary assessment is consistent with the Z chromosome likely coding for plumage/morphological characteristics (Minvielle et al., 2000; Sæther et al., 2007), which may be under the influence of sexual selection (Johnsgard, 1994; Promislow et al., 1994). I note that the identified putatively selected on loci may not directly be under selection, but rather "hitchhiking" with genes under selection (Feder et al., 2012). Future work will benefit from full sex chromosome sequencing for a finer examination of possible

"islands" of divergence (Coyne and Orr, 2004; Nosil and Schluter, 2011; Price, 2008) and for associating the genetic variance with morphological characters.

Phylogeography of Mexican Ducks

This study is the most comprehensive molecular assessment of Mexican ducks to date with samples spanning nearly the entirety of the taxon's range. In general, I found evidence for five differentiated subpopulations (i.e., US, Sonora, Durango, Zacatecas/Guanajuato/Mexico, and Puebla) with divergence following a stepping-stone model that corresponds with geography (Fig. 3.1A): each sampling group most closely resembles its geographic neighbor for all marker types (Fig. 3.1 & 3.3). Unlike the loadings between Mexican ducks and mallards, there is no indication of any set of loci overwhelmingly contributing to the recovered structure among Mexican duck sampling groups (Fig. 3.5), which is consistent with the effects of genetic drift.

Among the sampling groups, contemporary hybridization seems most problematic for US and Sonoran localities (Fig. 3.3). However, both autosomal and Z-linked markers recovered relatively few putative hybrids, suggesting that hybridization may not be as prevalent as once thought, or has more recently subsided (Hubbard, 1977; 2004; Perez-Arteaga et al., 2002). Nevertheless, if northern Mexican ducks continue to regularly interact with mallards, introgressed mallard alleles could potentially percolate into southern Mexican duck subpopulations (Lavretsky et al., *in press*). Gene flow from mallards into Mexican ducks could explain the similar estimates of nucleotide diversity (Fig. 3.4) that are inconsistent with census sizes (N = 55,500 Mexican ducks and 10 million mallards; Delany and Scott, 2006; Perez-Arteaga et al., 2002; Peters et al., 2014). However, I did not find any case of "inland" individuals that shared mtDNA haplotypes

(Fig. 3.6) or probability of assignment to the "mallard" population (Fig. 3.1C & D). Although hybridization (contemporary or ancestral) cannot be discounted, I hypothesize that recent divergence and retention of the ancestral gene pool, which can have results that are similar to those expected under gene flow (Noor and Bennett, 2009; Seehausen et al., 2014), may be the cause of the phenotypic-genetic discordance in which Mexican ducks expressing "mallard" characteristics are not genetically identified as hybrids. Ongoing efforts to reevaluate Mexican duck plumage variability, and particularly among males, are finding that morphological variance is geographically and/or age associated (Engilis unpub. data), suggesting that the residual "mallard-like" characters are more consistent with recent ancestry rather than hybridization.

Finally, historically found on inland Mexico marshes, the advancement of irrigation channels and drainage ditches in last few decades may have functioned as corridors that facilitated Mexican duck expansion into coastal habitats, especially into western Sonora during the past 20 years (Perez-Arteaga et al., 2002; Scott and Reynolds, 1984). Such a founder event is supported by mtDNA, for which Sonora was the most differentiated (Fig. 3.2; $\Phi_{ST} = 0.52 \pm 0.028$ StDev) and had the lowest nucleotide diversity relative to all other locations (Fig. 3.4). These results can be attributed to the prevalent mtDNA haplotype found in 78% of the Sonoran samples (Fig. 3.5), that also included 31% of haplotypes from the US, but none from interior Mexico. However, a similar founder effect was not evident in autosomal or Z-linked markers (Fig. 3.4), which might be attributable to overall larger effective population sizes of these markers; regardless, allelic frequency differences at autosomal and Z-linked markers support Sonora as a genetically unique subpopulation. The source for the Sonoran population is

likely a combination of individuals from neighboring US and Mexican states. Unfortunately, I was unable to sample Chihuahua, Mexico, which is a strong candidate for the source population given the geographic proximity. Increasing samples from US states, and attaining samples from Chihuahua, and other coastal states (e.g., Sinaloa, Mexico) will be important to understand the expansion into the new coastal niche.

On Taxonomy and the Selection Criterion

High variance in mechanisms involved in the speciation process, as well as the extent of isolation across avian lineages has resulted in extensive taxonomic debates (Grant and Grant 1992, 1997a). In particular, determining evolutionary relationships within rapid radiations in which genomes are largely free to move between species and/or are retained due to recent ancestry can be especially difficult (Grant and Grant, 1997; Lavretsky et al., 2014) unless genes maintaining species integrity (i.e., speciation genes) are found (Rundell and Price, 2009; Wu, 2001; Wu and Ting, 2004). Although identifying these selected-upon genes help, taxonomist must foremost consider the strength of selection/isolation that is conferred by these putative speciation genes (Charlesworth et al., 1997; Slatkin, 1987; Wu, 2001). As a result, I propose a species concept for incipient forms based on a selection criterion in which species assignment is based on the probability that species boundaries are retained or strengthened regardless of the extent of gene flow. For example, if divergence between mallards and Mexican ducks is driven via the sex chromosome, then determining the probability of species maintenance based on the strength of selection on this region(s) will be necessary for resolving their taxonomy.

To date, Mexican ducks have gone through several taxonomic revisions based on the variance in the presence of "mallard-like" plumage displayed across their range (AOU, 1983; Conover, 1922; Hubbard, 1977; Huber, 1920; Ridgway, 1886). Taxonomic reevaluation was recently proposed (Chesser et al., 2011; AOU petition 2010-B-6) based on mtDNA divergence estimates (McCracken et al., 2001). However, among the three markers, mtDNA had the highest Φ_{ST} estimates across pair-wise comparisons; the exceptions were US birds versus mallards and Guanajuato (Fig. 3.2) – although Guanajuato was likely hampered due to a small sample size (N = 2). Mitochondrial structure (Fig. 3.6), particularly among Mexican duck sampling groups, can be attributed to several factors, including (1) genetic drift acting on a maternally inherited marker with one quarter the effective population size of nuclear DNA (Zink and Barrowclough, 2008), (2) strong female philopatry that is characteristic of ducks (Doums et al., 2002; Peters et al., 2012), and (3) overall sedentary lifestyle of Mexican ducks that further limits genetic exchange. More importantly, the significant variance across markers (Fig. 3.2) demonstrates the importance of multi-marker comparisons, including identifying primary evolutionary influences, for proper decision making. Whereas our genomic assessment between Mexican ducks and mallards suggests post-zygotic isolation linked to the Z chromosome, which is consistent with incipient species designations, the clinal variance in autosomal markers and overall absence of taxon-specific markers within this dataset is more consistent with allo- or morpho-species designations (Lavretsky et al., *in press*). Once again, future taxonomic decisions will benefit from determining the effectiveness of the proposed isolation mechanism by identifying the prevalence of male versus female F_1, F_2, etc., hybrid individuals through either captive breeding experiments (i.e., see

American black duck x mallard experiments; Kirby et al., 2004) or observational studies in the wild; however, the latter requires the development of a morphological key to distinguish among mallards, Mexican ducks, and their hybrids.

CONCLUSIONS

Genomes, like populations, are dynamic and are continuously shaped by multiple evolutionary forces (Nosil et al., 2009b). Early stages of speciation likely proceed via changing selective pressures on genic regions and/or genetic drift between populations. I present compelling evidence that speciation of two incipient duck forms is being driven by selection on a few key sex-linked genes with large effects, whereas the remaining genome is largely affected by genetic drift. With advances in next-generation sequence methods, the field of speciation genomics is only beginning to open the mechanistic "black box" of speciation (Seehausen et al., 2014). Each taxonomic comparison continues to shed light into the behavior of genomes during speciation and subsequently revealing the process(es) that have resulted in the evolution of species complexes, as well as how these evolutionary mechanisms contribute to overall biodiversity.

ACKNOWLEDGMENTS

I am grateful to Eduardo Carrera and Ducks Unlimited-Mexico, Ruben Del Castillo, Todd Scott and Wingshooters Lodge-Mexico, Patricio Gaudiano, Kevin G. McCracken, Blanca E. Hernández-Baños, and Andrew Engilis, Jr. for the contributions of samples to

this study. I thank Jeffrey M. DaCosta and Michael D. Sorenson for their help with bioinformatics. I also thank Evangeline Shank for her contribution in sample preparation. This research was funded by Ducks Unlimited Richard H. G. Bonnycastle Fellowship in Wetland and Waterfowl Biology, American Museum of Natural History Chapman Grant, the National Science Foundation (DEB-0926162), the U.C. Davis Museum of Wildlife and Fish Biology, and a Research Initiation Grant from the Research Council at Wright State University. I thank Caitlin P. Wells for her reviews of previous drafts.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J., Baird, S., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C., Buggs, R., 2013. Hybridization and speciation. Journal of Evolutionary Biology 26, 229-246.

Aldrich, J.W., Baer, K.P., 1970. Status and Speciation in the Mexican Duck (*Anas diazi*). The Wilson Bulletin 82, 63-73.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Andrew, R.L., Rieseberg, L.H., 2013. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution 67, 2468-2482.

Ankney, C.D., Dennis, D.G., Wishard, L.N., Seeb, J.E., 1986. Low genic variation between black ducks and mallards. The Auk 103, 701-709.

AOU, 1983. Check-list of North American Birds, Washington, DC.

Avise, J.C., Ankney, D.C., Nelson, W.S., 1990. Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. Evolution 44, 1109-1119.

Axelsson, E., Smith, N.G.C., Sundström, H., Berlin, S., Ellegren, H., 2004. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. Molecular Biology and Evolution 21, 1538-1547.

Backström, N., Karaiskou, N., Leder, E.H., Gustafsson, L., Primmer, C.R., Qvarnström, A., Ellegren, H., 2008. A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. Genetics 179, 1479-1495.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16, 37-48.

Bellrose, F.C., 1976. Ducks, geese and swans of North America. Stackpole Books, Harrisburg, Pennsylvania.

Bonnet, E., Van de Peer, Y., 2002. zt: a software tool for simple and partial Mantel tests. Journal of Statistical software 7, 1-12.

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

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Molecular ecology 22, 3124-3140.

Charlesworth, B., Nordborg, M., Charlesworth, D., 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical research 70, 155-174.

Chesser, R.T., Banks, R.C., Barker, F.K., Cicero, C., Jon, L.D., Kratter, A.W., Irby, J.L., Rasmussen, P.C., Remsen, J.V., Rising, J.D., Stotz, D.F., Winker, K., 2011. Fifty-Second Supplement to the American Ornithologists' Union check-list of North American Birds. The Auk 128, 600-613.

Conover, B.H., 1922. *Anas diazi novimexicana* and *Gris amiiericantain* Nebraska. The Auk 39, 412.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland, MA.

Delany, S., Scott, D., 2006. Waterbird population estimates. Wetlands International Global Series, the Netherlands.

Dobzhansky, T., 1940. Speciation as a stage in evolutionary divergence. The American Naturalist 74, 312-321.

Doums, C., Cabrera, H., Peeters, C., 2002. Population genetic structure and male-biased dispersal in the queenless ant Diacamma cyaneiventre. Molecular Ecology 11, 2251-2264.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32, 1792-1797.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460-2461.

Edwards, S.V., Kingan, S.B., Calkins, J.D., Balakrishnan, C.N., Jennings, W.B., Swanson, W.J., Sorenson, M.D., 2005. Speciation in birds: genes, geography, and

sexual selection. Proceedings of the National academy of Sciences of the United States of America of the United States of America 102, 6550-6557.

Ellegren, H., 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology 17, 1629-1631.

Excoffier, L., Lischer, H.E.L., 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.

Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends in Genetics 28, 342-350.

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

Frank, S.A., 1991. Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. Evolution, 262-267.

Grant, P.R., Grant, B.R., 1997. Genetics and the origin of bird species. Proceedings of the National Academy of Sciences of the United States of America 94, 7768-7775.

Haldane, J., 1948. The theory of a cline. Journal of Genetics 48, 277-284.

Haldane, J.B.S., 1922. Sex ratio and unisexual sterility in hybrid animals. Journal of genetics 12, 101-109.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genetics 6, e1000862.

Huang, Y., Li, Y., Burt, D.W., Chen, H., Zhang, Y., Qian, W., Kim, H., Gan, S., Zhao, Y., Li, J., 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nature genetics.

Hubbard, J.P., 1977. The biological and taxonomic status of the Mexican Duck. New Mexico Department of Game and Fish Bulletin, pp. 1-56.

Huber, W., 1920. Description of a new North American duck. The Auk 37, 273-274.

Johnsgard, P.A., 1978. Ducks, geese and swans of the World. Lincoln, Univ. Nebraska Press.

Johnsgard, P.A., 1994. Arena birds: sexual selection and behavior. Smithsonian Institution Press Washington, DC.

Johnson, K.P., Sorenson, M.D., 1999. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): A comparison of molecular and morphological evidence. The Auk 116, 792-805.

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

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

Kirby, R.E., Sargeant, G.A., Shutler, D., 2004. Haldane's rule and American black duck × mallard hybridization. Canadian Journal of Zoology 82, 1827-1831.

Kraus, R., Kerstens, H., Van Hooft, P., Crooijmans, R., Van Der Poel, J., Elmberg, J., Vignal, A., Huang, Y., Li, N., Prins, H., Groenen, M., 2011. Genome wide SNP

discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). BMC Genomics 12, 150.

Kraus, R.H., Hooft, P., Megens, H.J., Tsvey, A., Fokin, S.Y., Ydenberg, R.C., Prins, H.H., 2013. Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. Molecular ecology 22, 41-55.

Kulikova, I.V., Zhuravlev, Y.N., McCracken, K.G., 2004. Asymmetric hybridization and sex-biased gene flow between eastern spot-billed ducks (*Anas zonorhyncha*) and mallards (*A. platyrhynchos*) in the Russian Far East. The Auk 121, 930-949.

Kunte, K., Shea, C., Aardema, M.L., Scriber, J.M., Juenger, T.E., Gilbert, L.E., Kronforst, M.R., 2011. Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. PLoS Genetics 7, e1002274.

Lavretsky, P., Hernández Baños, B.E., Peters, J.L., , *in press*. Weak differentiation across eighteen molecular markers from four incipient taxa of the new world mallard complex (*anas* spp.). The Auk.

Lavretsky, P., McCracken, K.G., Peters, J.L., 2014. Phylogenetics of a recent radiation in the mallards and allies (Aves: *Anas*): Inferences from a genomic transect and the multispecies coalescent. Molecular phylogenetics and evolution 70, 402–411.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452.

Liu, A., Burke, J.M., 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173, 321-330.

Livezey, B.C., 1991. A phylogenetic analysis and classification of recent dabbling ducks (Tribe *Anatini*) based on comparative morphology. The Auk 108, 471-507.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., Jiggins, C.D., 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome research 23, 1817-1828.

Mayr, E., 1982. Of What Use Are Subspecies? The Auk 99, 593-595.

McCracken, K.G., Johnson, W.P., Sheldon, F.H., 2001. Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). Conservation Genetics 2, 87-102.

Minvielle, F., Ito, S., Inoue-Murayama, M., Mizutani, M., Wakasugi, N., 2000. Brief communication. Genetic analyses of plumage color mutations on the Z chromosome of Japanese quail. Journal of Heredity 91, 499-501.

Nanda, I., Schlegelmilch, K., Haaf, T., Schartl, M., Schmid, M., 2008. Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. Cytogenetic and genome research 122, 150-156.

Nei, M., 1972. Genetic distance between populations. American naturalist, 283-292.

Noor, M.A., Bennett, S.M., 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103, 439-444.

Nosil, P., Feder, J.L., 2013. Genome evolution and speciation: toward quantitative descriptions of pattern and process. Evolution 67, 2461-2467.

Nosil, P., Funk, D.J., Ortiz-Barrientos, D., 2009a. Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18, 375-402.

Nosil, P., Gompert, Z., Farkas, T.E., Comeault, A.A., Feder, J.L., Buerkle, C.A., Parchman, T.L., 2012. Genomic consequences of multiple speciation processes in a stick insect. Proceedings of the Royal Society B: Biological Sciences 279, 5058-5065.

Nosil, P., Harmon, L.J., Seehausen, O., 2009b. Ecological explanations for (incomplete) speciation. Trends in Ecology & Evolution 24, 145-156.

Nosil, P., Schluter, D., 2011. The genes underlying the process of speciation. Trends in Ecology & Evolution 26, 160-167.

Orr, H.A., 2001. The genetics of species differences. Trends in Ecology & Evolution 16, 343-350.

Palmer, R.S., 1976. Handbook of North American Birds, New Haven, Connecticut.

Pease, J.B., Hahn, M.W., 2013. More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. Evolution 67, 2376-2384.

Pérez-Arteaga, A., Gaston, K., 2004. Wildfowl population trends in Mexico, 1961-2000: a basis for conservation planning. Biological Conservation. Biological Conservation 115, 343-355.

Perez-Arteaga, A., Gaston, K.J., Kershaw, M., 2002. Population trends and priority conservation sites for Mexican Duck *Anas diazi*. Bird Conservation International 12, 35-52.

Pérez-Figueroa, A., García-Pereira, M., Saura, M., Rolán-Alvarez, E., Caballero, A., 2010. Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology 23, 2267-2276.

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

Peters, J.L., Sonsthagen, S.A., Lavretsky, P., Rezsutek, M., Johnson, W.P., McCracken, K.G., 2014. Interspecific hybridization contributes to high genetic diversity and apparent effective population size in an endemic population of mottled ducks (*Anas fulvigula maculosa*). Conservation Genetics 15, 509-520.

Phadnis, N., Orr, H.A., 2009. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science 323, 376-379.

Price, T.D., 2008. Speciation in Birds. Roberts and Company, Greenwood Village, CO.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945-959.

Promislow, D., Montgomerie, R., Martin, T.E., 1994. Sexual selection and survival in North American waterfowl. Evolution, 2045-2050.

Pryke, S.R., 2010. Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. Evolution 64, 1301-1310.

Pryke, S.R., Griffith, S.C., 2009. Postzygotic genetic incompatibility between sympatric color morphs. Evolution 63, 793-798.

Reeve, H.K., Pfennig, D.W., 2003. Genetic biases for showy males: Are some genetic systems especially conducive to sexual selection? Proceedings of the National Academy of Sciences of the United States of America 100, 1089-1094.

Ridgway, R., 1886. Preliminary descriptions of some new species of birds from southern Mexico, in the collection of the Mexican Geographical and Exploring Commission. The Auk 3, 331-333.

Rockman, M.V., 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66, 1-17.

Rundell, R.J., Price, T.D., 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends in Ecology & Evolution 24, 394-399.

Sæther, S.A., Sætre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M.R., Bureš, S., 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science 318, 95-97.

Schluter, D., 2009. Evidence for ecological speciation and its alternative. Science 323, 737-741.

Scott, N.J., Jr., Reynolds, R.P., 1984. Phenotypic variation of the Mexican duck (*Anas platyrhynchos diazi*) in Mexico. The Condor 86, 266-274.

Seehausen, O., 2004. Hybridization and adaptive radiation. Trends in Ecology & Evolution 19, 198-207.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.-P., Bank, C., Brännström, Å., 2014. Genomics and the origin of species. Nature Reviews Genetics 15, 176-192.

Slatkin, M., 1987. Gene flow and the geographic structure of natural. Science 236, 787-792.

Sorenson, M.D., Ast, J.C., Dimcheff, D.E., Yuri, T., Mindell, D.P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12, 105-114.

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

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., Slate, J., 2010. Adaptation genomics: the next generation. Trends in Ecology & Evolution 25, 705-712.

Stölting, K.N., Nipper, R., Lindtke, D., Caseys, C., Waeber, S., Castiglione, S., Lexer, C., 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Molecular Ecology 22, 842-855.

Stutzenbaker, C.D., 1988. The mottled duck, its life history, ecology and management. In: Department, T.P.a.W. (Ed.), Austin, Texas.

Sutter, A., Beysard, M., Heckel, G., 2013. Sex-specific clines support incipient speciation in a common European mammal. Heredity 110, 398-404.

Templeton, A.R., 1981. Mechanisms of Speciation--A Population Genetic Approach. Annual Review of Ecology and Systematics 12, 23-48.

Trier, C.N., Hermansen, J.S., Sætre, G.-P., Bailey, R.I., 2014. Evidence for Mito-Nuclear and Sex-Linked Reproductive Barriers between the Hybrid Italian Sparrow and Its Parent Species. PLoS genetics 10, e1004075.

Tucker, P.K., Sage, R.D., Warner, J., Wilson, A.C., Eicher, E.M., 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution 46, 1146-1163.

Turelli, M., Moyle, L.C., 2007. Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. Genetics 176, 1059-1088.

Winker, K., 2009. Reuniting phenotype and genotype in biodiversity research. BioScience 59, 657-665.

Wolf, J.B.W., Lindell, J., Backström, N., 2010. Speciation genetics: current status and evolving approaches. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 1717-1733.

Wu, C.-I., 2001. The genic view of the process of speciation. Journal of Evolutionary Biology 14, 851-865.

Wu, C.-I., Ting, C.-T., 2004. Genes and speciation. Nature Reviews Genetics 5, 114-122.

Yeaman, S., Whitlock, M.C., 2011. The genetic architecture of adaptation under migration–selection balance. Evolution 65, 1897-1911.

Zink, R.M., Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17, 2107-2121.

CHAPTER IV. ISLANDS AS VENUES FOR HOMOPLOID HYBRID SPECIATION: A CASE FOR THE ENDANGERED HAWAIIAN DUCK

Abstract – Speciation is regarded primarily as a bifurcation from an ancestral species resulting in two distinct taxonomic units. Although hybrid speciation is known to occur, such events have been regarded as rare in homoploid systems. Here, I provide several lines of evidence supporting the hypothesis that the endangered Hawaiian duck (*Anas wyvilliana*) is descended from ancient hybridization between the mallard (*A. platyrhynchos*) and Laysan duck (*A. laysanensis*). I discuss how island systems might act as arenas for interspecific hybridization leading to speciation as a result of rapid isolation between hybrids and parental species. Hybrid speciation may be a more common mechanism than previously thought for generating biodiversity, especially during rapid radiations.

INTRODUCTION

Hybrid speciation (Mallet, 2007) is emerging as an important mechanism for species formation (Brelsford et al., 2011; Dowling and Secor, 1997; Jacobsen and Omland, 2011; Mallet, 2007; MavÁRez and Linares, 2008; Schwarz et al., 2005; Seehausen, 2004) and the generation of biodiversity (Baack and Rieseberg, 2007; Mallet, 2007). Traditionally regarded as being more important in polyploid systems through allopolyploid formation (Husband, 2000; Ramsey and Schemske, 2002; Soltis et al., 2004; Wood et al., 2009), hybrid speciation was considered unlikely in homoploid systems (Mallet, 2007; MavÁRez and Linares, 2008) due to the high chance of continued gene flow between hybrids and parental species (Mallet, 2005, 2007). However, molecular data have revealed compelling cases of hybrid speciation across a variety of homoploid taxonomic groups (e.g. plants (Rentsch and Leebens-Mack, 2012; Soltis et al., 2004; Wood et al., 2009); fish (DeMarais et al., 1992; Keller et al., 2012; Nolte et al., 2005); insects (Consortium 2012; Kunte et al. 2011; Schwarz et al. 2005); birds (Brelsford et al., 2011; Hermansen et al., 2011)). Several hypotheses for homoploid hybrid speciation have been formulated, including hybrid trait speciation (Arnold et al., 1999; Keller et al., 2012;

Kunte et al., 2011; Salazar et al., 2010) and adaptive niche availability (Gompert et al., 2006; Kunte et al., 2011; Nolte et al., 2005; Rieseberg, 2006; Seehausen, 2004). Here I explore a third possibility: island systems in which closely related taxa interact by happenstance producing offspring that are isolated and speciate in allopatry (Jacobsen and Omland, 2011; Mallet, 2007).

Although homoploid hybrid speciation has been suggested in a number of systems, systematically testing and ruling out alternative evolutionary possibilities remains difficult (Jacobsen and Omland, 2011). Major critera recently outlined by Jacobsen and Omland (2011) in support of hybrid speciation include the existence of (1) three identifiable taxa (i.e. hybrid speciation effectively increased biodiversity (Schwarz et al., 2005) in which (2) a complex evolutionary history within the putative hybrid species is supported by a mito-nuclear discord that is (3) further supported by a genomic mosaic consisting of parental alleles from both species within the putative hybrid species. Focusing on the evolution of the Hawaiian duck (*Anas wyvilliana*), I present compelling support for a hybrid origin and discuss how the Hawaiian Islands, and islands in general, might provide the appropriate venue for hybrid species formation.

The Hawaiian duck is one of fourteen incipient taxa within the mallard complex (Lavretsky et al., 2014). Whereas morphological (Livezey, 1991) and nuclear (Lavretsky et al., 2014) data suggest a sister relationship with the Laysan duck (*A. laysanensis*), mitochondrial (mt) DNA supports a close affinity to the mallard (*A. platyrhynchos*) (Fowler et al., 2009). Specifically, Hawaiian ducks possess mtDNA haplotypes that are nested within, and probably derived from, a clade consisting of mallard and other New World mallard-like ducks (Fowler et al., 2009; Lavretsky et al., 2014). Given the mito-

nuclear discord among the three identifiable taxa, I tested for the presence of a genomic mosaic within Hawaiian ducks. In addition, I used coalescent methods to test for gene flow under different evolutionary scenarios to test *a priori* predictions under a hypothesis of hybrid speciation. If the Hawaiian duck is a derivative of hybrid speciation, then regardless of the pre-defined topology, gene flow from the non-sister taxon will be required to explain the observed genetic diversity within Hawaiian ducks. Specifically, I predict non-zero gene flow from the basal lineage into the Hawaiian duck or its ancestor. Alternatively, if results are driven by common ancestry and stochastic lineage sorting, then no gene flow will be necessary to explain the genetic variability observed in Hawaiian ducks.

MATERIALS AND METHODS

Sample Preparation and Nuclear Marker Amplification

Genomic DNA was isolated from 21 Laysan ducks, 15 Hawaiian ducks, and 25 mallards using a Qiagen DNA extraction kit (Qiagen) according to the manufacturer's protocol.

Nineteen nuclear intronic loci, each from a different chromosomes, that were previously optimized in gadwall (*Anas strepera*) (Appendix Table A4.1; Peters et al., 2012) were used. Putatively neutral markers (i.e., introns) were used as these are expected to differ in allopatric systems as a result of stochasticity and population demography (Dobzhansky, 1940; Mayr, 1963) rather than selection, which can quickly drive favorable alleles to fixation and decrease the "hybrid" signal (Nolte and Tautz,

2010; Seehausen, 2004). Amplification by PCR was carried out with 1.5 μL of an individual's DNA combined with 1 nM of both forward and reverse primers, and 2x GoTaq Green Master Mix (Promega) for a total of a 15 μL reaction per individual per locus. PCR was conducted using an Eppendorf Mastercycler (ep*gradient*) thermocycler under the following conditions: DNA denaturation at 94°C for 7 minutes, followed by 45 cycles of DNA denaturation at 94°C for 20 s, primer annealing at 58°C for 20 s, and DNA extension at 72°C for 1 minute, and a final DNA extension at 72°C for 7 minutes. Amplification was verified using gel electrophoresis with a 1.5% agarose gel. PCR products were cleaned with AMPure XP beads, following Agencourt protocol (Beckman Coulter Co.). Sequencing was performed using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) following supplier protocols. Sequenced products were sent to the DNA Analysis Facility at Yale University for automated sequencing on an ABI 3730. Sequences were aligned and edited using Sequencher v. 4.8 (Gene Codes, Inc). All sequences were submitted to GenBank (*accession numbers pending*).

Gametic phases were resolved first for sequences with indels by methods outlined in Peters et al. (2007) and then used as "knowns" when resolving the remaining sequences with the program PHASE (Stephens and Donnelly, 2003). PHASE derives the most likely state of each allele algorithmically by comparing all known alleles. Additionally, all mallard sequences were previously resolved with >95% confidence from a larger data set that included extensive allele-specific priming (Peters et al. unpubl. data) and were also treated as "knowns." Linkage between loci was not considered as all markers are found on different chromosomes.

Identifying a Genomic Mosaic

A finer examination of overall genetic connectivity among individuals with linked nodes representing reticulate events (i.e. hybridization or recombination) was conducted in SplitsTree (Huson and Bryant, 2006). An unrooted phylogenetic network was reconstructed from 19 nuclear loci that were first concatenated for each individual with IUPAC nucleotide codes used for ambiguous sites. A neighbor net analysis with character transformations based in an uncorrected P and an equal angle for both splits and reticulate transformations were used.

A locus-by-locus AMOVA was used to determine the most informative single nucleotide polymorphism (SNP) between Laysan and mallard ducks per locus in Arlequin 3.5 (Excoffier and Lischer, 2010) (see also Appendix Table A4.1). These SNPs were then isolated in Hawaiian ducks. SNPs were subsequently imported into *Structure* v. 2.3.4 (Pritchard et al., 2000) that uses Bayesian clustering methods to estimate admixture proportions from molecular data. *Structure* was run for 500,000 iterations after a burn-in of 100,000. All loci were considered independent, and the admixture model was used to determine percent genome composition. I expect individuals with a genomic mosaic to display admixture proportions relative to the contribution of its parental taxa when analyzed with a two population model (K = 2), while displaying an independent lineage when analyzed with a three population model (K = 3), demonstrating that the hybrid species is genetically diagnosable (Gompert et al., 2006; Kunte et al., 2011).

Estimating Gene Flow and Divergence Time

IMa2 assigns posterior probability density estimates for population sizes and migration rates from non-recombinant sequence fragments for several populations (N = 2-10) using Bayesian Markov chain Monte Carlo (MCMC) algorithms (Nielsen and

Wakeley, 2001). To do so, all loci were tested for recombination using the program IMgC (Woerner et al., 2007) (Appendix Table A4.1) and then manually truncated in order to retain polymorphic sites (>2 states) that would have been automatically removed by IMgC. Weight was given to maximize fragment length, unless sample size was decreased by > 10% of each population, in which case fragment lengths were reduced to maximize sample size. Phylogenetic relationships were manually entered into IMa2 and ran until the effective sample sizes (EES) for parameters were ≥ 50. Given the discord in sister-relationships derived from mtDNA versus nuDNA markers (Lavretsky et al., 2014), gene flow estimates were derived under two alternative tree topologies that included an mtDNA-like (Hawaiian duck is sister to mallard) and nuDNA-like relationships (i.e., Hawaiian duck is sister to Laysan duck). Once again, regardless of the pre-defined topology and under a hybrid origin, I expect gene flow from the non-sister taxon into Hawaiian ducks or its ancestor.

Years since divergence (T) was derived as $T = t/\mu$, t being the time since divergence parameter in IMa2. The mutation rate (μ) to be 2.67 x 10^{-7} substitutions/locus/site/year was derived from the geometric mean number of base-pairs (222.32 bp) and previously calculated average mutation rate (μ = 1.2 x 10-9 substitutions/locus/site/year; Peters et al., 2008) (Appendix Table A4.1).

RESULTS

Identifying a Genomic Mosaic

For nuclear DNA, individual relationships revealed that Hawaiian ducks cluster at intermediate positions between mallards and Laysan ducks, and share many reticulate events with both species, which is consistent with a genomic mosaic (Fig. 4.1A). In contrast, under a classical bifurcating history, I would have expected Hawaiian ducks to cluster more closely and share more reticulations (resulting from incomplete lineage sorting) with its sister species, as observed in the mtDNA neighbor net tree (Fig. 4.1B).

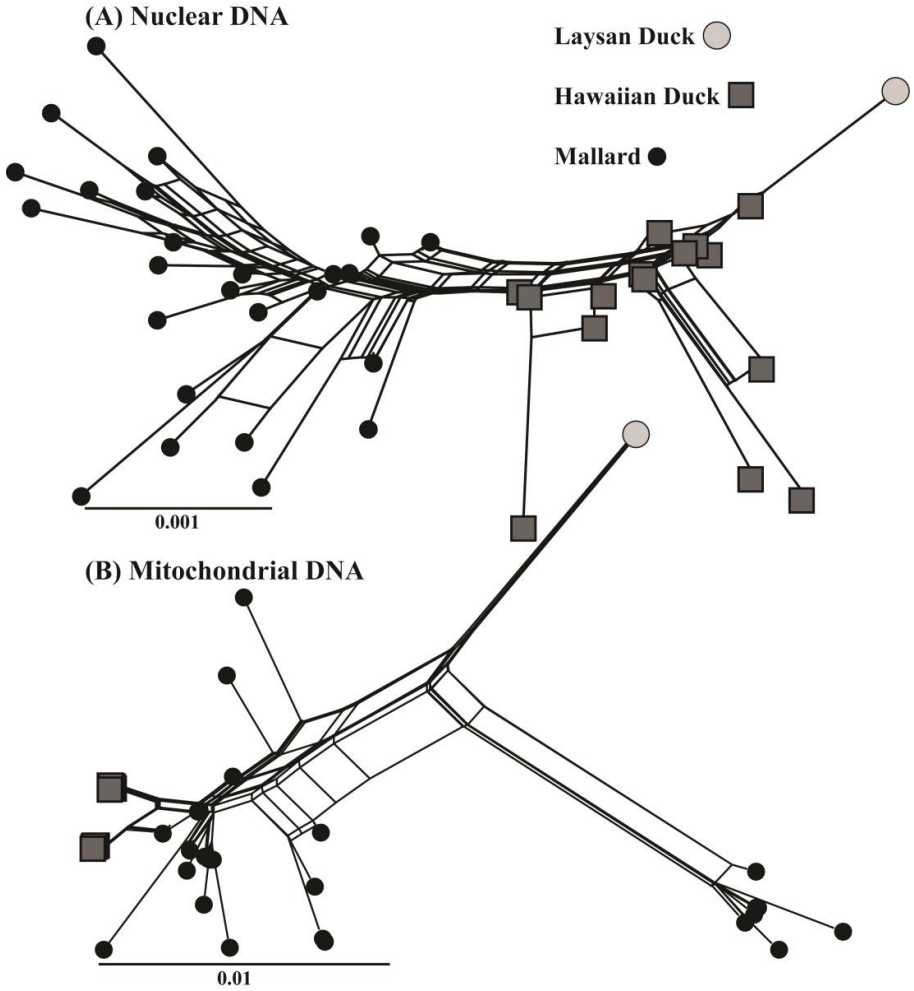


Figure 4.1. Neighbor-net trees for (A) nuclear DNA (6,682 aligned nucleotides) showing Hawaiian ducks as being intermediate between mallard and Laysan duck and (B) mitochondrial DNA control region (645 bp) showing Hawaiian duck to be deeply nested within mallard and distinct from Laysan duck.

Structure (Pritchard et al., 2000) assigned all Laysan ducks to population A with an average probability of 99% (± 0.0% SD) and all mallards to population B with an average probability of 98% (± 1.3% SD). Consistent with a genomic mosaic, all Hawaiian duck individuals were assigned to both parental groups with an average assignment of 41% (± 9.9% SD) to population A and 59% (± 9.9% SD) to population B (Fig. 4.2A). In general, the nuclear genome of the Hawaiian duck was consistent with a 50:50 mosaic. Furthermore, SNP frequencies reveal that Hawaiian ducks are fixed at three loci, two specific to Laysan ducks and one specific to mallards, whereas the remaining fourteen SNPs had intermediate frequencies as expected for a putative hybrid species (Appendix Table A4.1). Thus, Criteria 3 (i.e., genetic mosaic; Jacobsen and Omland, 2011) is supported by these data. In a three population model, Hawaiian ducks were recovered as a distinct population with an average of 95% (± 8.1% SD) probability (Fig. 4.2B), which is consistent with the Hawaiian duck being a genetically distinct cluster, and thus meeting Criteria 1: persistence of three distinct lineages (Jacobsen and Omland, 2011).

If the apparent mosaic was due to stochastic lineage sorting then I expect other species within the mallard complex to show assignment probabilities similar to the Hawaiian duck when analyzed with the same set of SNPs. However, I found no evidence of this when assigning other mallard-like species to a two- or three-population model (Appendix Fig. A4.1). In a two-population model, all other species were assigned with strong posterior support to the same population as mallards.

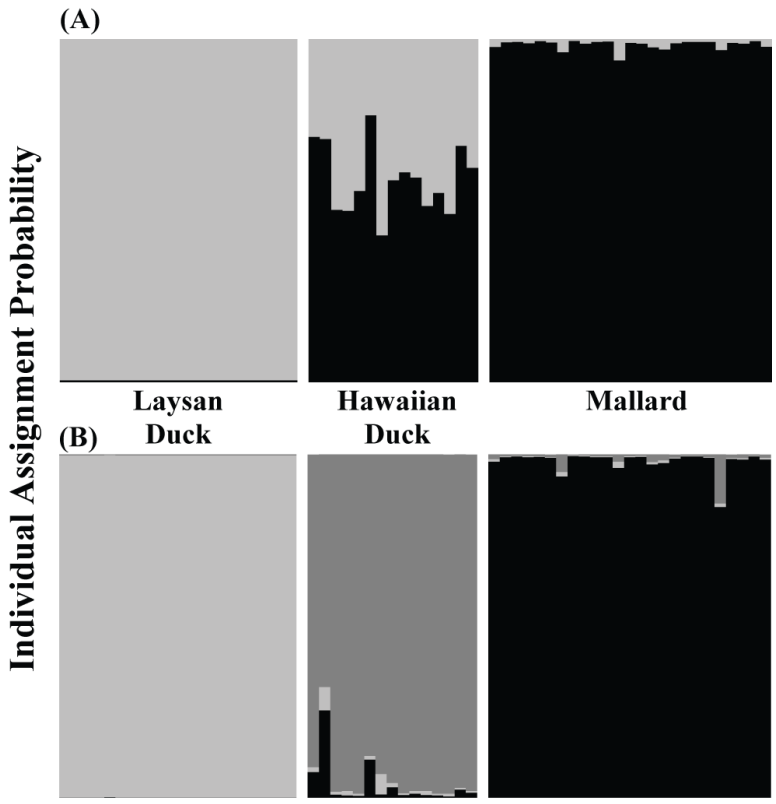


Figure 4.2A-B. Assignment probabilities obtained from 17 diagnostic SNPs ascertained by comparing 21 Laysan ducks and 25 mallards (Appendix Table A4.1) and assayed in 15 Hawaiian ducks. (A) K = 2 populations; (B) K = 3 populations.

Gene Flow

The only non-zero gene flow observed under the mtDNA-like topology (Hawaiian duck sister to mallard; Fig. 4.3A) was from Laysan ducks into the Hawaiian duck ($2Nm$ = 1.58; 95% CI 0.52-8.8), whereas under the nuDNA-like topology (Hawaiian duck sister to Laysan duck; Fig. 4.3B), non-zero gene flow from mallards into the ancestor of Hawaiian and Laysan duck was supported ($2Nm$ = 1.37; 95% CI 0.87-26.11). Thus, a simple bifurcating history was insufficient for explaining the evolution of this group. Moreover, all Hawaiian duck individuals were sampled from Kauai, which is thought to

be free from contemporary hybridization, and each individual was previously molecularly vetted as “pure” (Fowler et al., 2009). Consistent with this, the non-zero gene flow from mallards into the Hawaiian-Laysan duck ancestor supported ancient, rather than recent, hybridization, suggesting that contemporary gene flow is unlikely to explain the genomic mosaic.

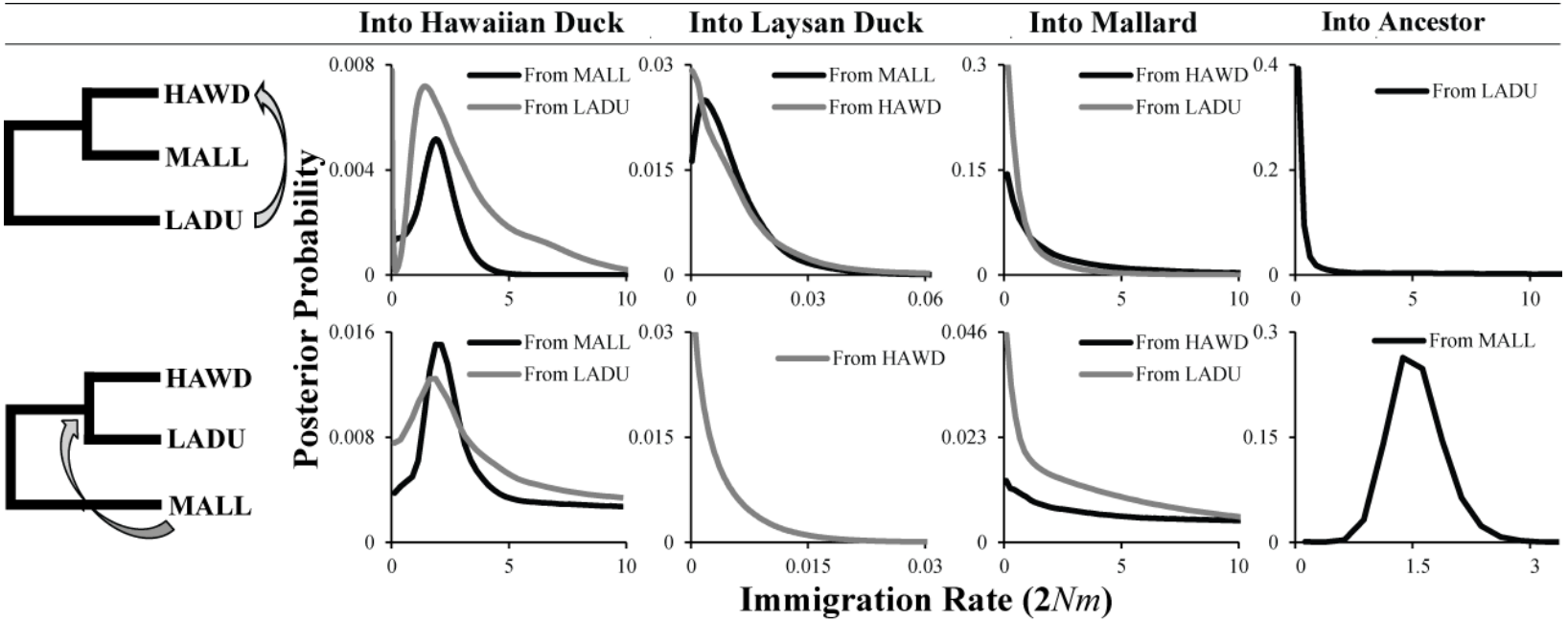


Figure 4.3. Population migration rates (2*Nm*) estimated from 19 nuclear loci and defining a (A) mtDNA-like topology and (B) nuDNA-like topology (Lavretsky et al., 2014; HAWD = Hawaiian Duck; LADU = Laysan Duck; MALL= Mallard). The 95% highest posterior distributions that did not include zero gene flow (i.e., rejected complete isolation) was from Laysan ducks into Hawaiian ducks under the mtDNA-like topology and from mallards into the Hawaiian-Laysan duck ancestor under the nuDNA-like topology. Thus, consistent with the hybrid speciation hypothesis, gene flow from the non-sister species is necessary to explain the genetic variability within Hawaiian ducks.

Divergence Time

Divergence times were reliably obtained (ESS ≥ 50) under the nuDNA-like topology only (Appendix Fig. A4.2). The inability to obtain a divergence estimate under the mtDNA-like topology is likely due to forcing nuDNA to resolve a "false" scenario in which the Hawaiian duck is sister to mallards (Fig. 4.3). In general, divergence estimates between the Laysan-Hawaiian duck ancestor and mallard was estimated to be ~650,000 years before present (YBP) (95% CI = 364,000–1,100,000 YBP), which is consistent with the Laysan duck being one of the older lineages within the mallard complex (Johnson and Sorenson, 1999; Lavretsky et al., 2014). Divergence time between Hawaiian ducks and Laysan ducks was estimated at ~3,000 YBP (95% CI = 560–240,000 YBP). The latter divergence estimate corresponds to those estimated from species tree reconstructions with mitochondrial (i.e., Hawaiian duck sister to Mallard; 23,000 YBP; 95% CI = 0 – 80,000) or nuclear (i.e., Hawaiian

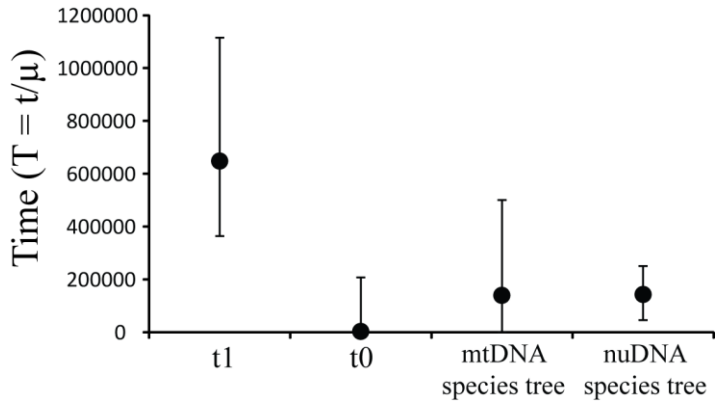


Figure 4.4. IMa2 (Nielsen and Wakeley, 2001) time since divergence (with 95% CI) estimates for derived from 19 nuclear introns under the nuDNA-like topology (see Fig. 4.3) for t1 (i.e., divergence between mallards and Hawaiian-Laysan duck ancestor) and t0 (i.e., divergence between Hawaiian and Laysan ducks). Additional divergence estimates derived from species trees (Lavretsky et al., 2014) reconstructed with mtDNA (i.e., divergence between Hawaiian duck and mallard) and nuDNA.

duck sister to Laysan duck; 60,000 YBP; 95% CI = 18,000-100,000 YBP) markers – branch lengths derived from Lavretsky et al (2014); $\mu_{MTdna} = 4.8x10^{-8}$ (Peters et al., 2005), $\mu_{NUdna} = 1.2x10^{-9}$ (Peters et al., 2008) (Fig. 4.4).

DISCUSSION

My results satisfy Jacobsen and Omland's (2011) three criteria for hybrid speciation. First, there are three extant taxa that are genetically differentiated. Second, there is significant mito-nuclear discordance (Lavretsky et al., 2014). Third, the nuclear genome appears to be a mosaic of the two parental lines. In addition, I demonstrate that gene flow from the non-sister taxon, regardless of tree topology, is required to explain the genetic diversity observed within Hawaiian ducks, and that contemporary gene flow is an unlikely explanation (Fig. 4.3). Thus, I conclude that hybrid speciation played an integral role in the evolution of the Hawaiian duck.

Morphological characteristics of Hawaiian ducks are also suggestive of a hybrid origin. Hawaiian ducks have intra-appendicular skeletal and sternal dimensions that are intermediate between Laysan ducks and continental mallards (Livezey, 1993). Interestingly, the skeletal dimensions of juvenile Hawaiian ducks are similar to adult Laysan ducks, whereas these characteristics cluster adult Hawaiian ducks and juvenile mallards (Livezey, 1993). Moreover, ongoing studies continue to note high variation in plumage characteristics within Hawaiian ducks that again appear to be intermediate between its putative parental taxa (Engilis et al., 2002), as well as corroborate a morphology-based phylogeny that placed the Hawaiian duck as intermediate between the

Laysan duck and mallard (Livezey, 1991). These phenotypic-based studies further support an admixed history.

Estimates of Divergence Time Correspond to the Fossil Record

Additional evidence for a complex evolutionary history is found within the Hawaiian fossil record which has Laysan-like duck forms dating to the mid-Pleistocene, intermediate Laysan-Hawaiian duck forms dating to the Holocene (Burney et al., 2001; Cooper et al., 1996; Olson and James, 1991), but only a few recent bones are attributable to modern Hawaiian ducks (H. James, pers. obs.). In general, divergence estimates from coalescent methods closely correspond with the fossil record, suggesting a Pleistocene divergence between Laysan ducks and mallards, and a much more recent divergence between Hawaiian ducks and its putative sister species (Fig. 4.4). Given the fossil and molecular data, I hypothesize an ancestral hybridization event near the Pleistocene-Holocene transition between the once widespread Laysan duck (Cooper et al., 1996) and mallards that arrived on the Hawaiian Islands by happenstance during migration (dispersal by "migratory drop-outs" of several species of Holarctic waterbirds continues to be documented on the Islands; Engilis Jr et al., 2004; Pyle and Pyle, 2009). This hybridization event resulted in a hybrid swarm that became isolated from its parental species and subsequently speciated (Jacobsen and Omland, 2011; Mallet, 2007).

Scenarios for Hybrid Speciation in the Hawaiian Duck

Laysan ducks only recently disappeared from the main Hawaiian Islands (800-900 YBP; Pyle, 1988). The sympatric existence between the incipient Hawaiian duck population and their Laysan parental species would have resulted in backcrossing and diminishing hybrid signal, unless the two were somehow isolated. First, given that

mallards are known to be relatively dominant when in contact with other conspecifics (Brodsky et al., 1988), the offspring of mallards may have outcompeted their smaller Laysan parent (i.e., adaptive hybrid trait advantage hypothesis; Keller et al., 2012; Salazar et al., 2010). Alternatively, hybrid individuals may have been able to take advantage of underutilized habitat (i.e., adaptive niche hypothesis; Gompert et al., 2006; Nolte et al., 2005; Rieseberg, 2006; Seehausen, 2004), which is supported by the fossil record. Specifically, recoveries of Laysan duck fossils across terrestrial and even montane sites of the main Hawaiian Islands suggest that they were more terrestrial than aquatic (Cooper et al., 1996; Moulton and Marshall, 1996; Olson and Ziegler, 1995). Conversely, Hawaiian ducks have not been recovered from terrestrial fossil sites, and are known to be strongly associated with water (e.g., perennial streams, lowland marshes, and wetlands). Consequently, an ecological shift within the hybrid swarm could have facilitated initial isolation. The unexplained extirpation of Laysan ducks from the main islands prior to Polynesian arrival suggests that Laysan Island might have acted as a refuge from a "shifting hybrid zone" (Rheindt and Edwards, 2011) that finally isolated the hybrid swarm from Laysan ducks. Moreover, assortative mating within the hybrid swarm could explain the complete lack of Laysan-like mtDNA haplotypes if female mallards were more likely to mate with male Laysan ducks; however, a single mtDNA lineage could also have become fixed as a result of selection or drift in the small population size of Hawaiian ducks and their ancestors. Examining the temporal and spatial distributions of fossil morphotypes, coupled with ancient DNA analyses (Huynen et al., 2003; Willerslev and Cooper, 2005), could provide explicit tests for these isolating mechanism hypotheses.

Islands as Venues for Hybrid Speciation

A major criticism of homoploid hybrid speciation is that continued interactions between hybrids and parental species inhibit speciation (Mallet, 2005, 2007), resulting in the creation of hybrid zones (Barton and Hewitt, 1989), the reversal of speciation (Seehausen, 2006), or extinction by hybridization (Rhymer, 2006). However, island systems might reduce interactions between hybrids and parental species by imposing strong barriers to expansion, and thus maintaining hybrid populations and enabling speciation. My results suggest that perhaps given the available niche space and the chance of becoming isolated, there is a non-trivial likelihood of hybrid speciation when incipient species come into secondary contact on islands (Nolte and Tautz, 2010), and if so, this may be an important mechanism in the evolution of island biodiversity. Although hybrid speciation as a mechanism has previously been proposed for the radiation of other island forms in the mallard complex (e.g., Mariana mallard, *Anas oustaleti*; Reichel and Lemke, 1994), my analyses are the first quantitative evaluation of this hypothesis. The Laysan-Hawaiian-mallard complex provides an intriguing model system to understand how selection, genetic drift, and the overall consequences of genomic admixture interact in the formation of new species on islands.

ACKNOWLEDGMENTS

I am grateful to the following people and institutions for their contributions of samples to this study: Andy Engilis Jr. & Irene Engilis, Museum of Wildlife and Fish Biology, University of California, Davis, and the Burke Museum, University of Washington. Research was funded by Ducks Unlimited Richard H. G. Bonnycastle Fellowship in

Wetland and Waterfowl Biology, the National Science Foundation (DEB-0926162), the College of Sciences and Mathematics at Wright State University, and the Selma Herr Fund for Ornithology at U.C. Davis. Hawaiian duck specimens were provided by the USFWS Refuges, Hawaiian Islands and Laysan duck specimens were provided by the USGS-NWHC-Hawaii Field Office. I thank Caitlin P. Wells and John M. Eadie, University of California, Davis, for their helpful comments on previous drafts.

REFERENCES

Arnold, M.L., Bulger, M.R., Burke, J.M., Hempel, A.L., Williams, J.H., 1999. Natural hybridization: how low can you go and still be important? Ecology 80, 371-381.

Baack, E.J., Rieseberg, L.H., 2007. A genomic view of introgression and hybrid speciation. Current Opinion in Genetics & Development 17, 513-518.

Barton, N.H., Hewitt, G.M., 1989. Adaption, speciation and hybrid zones. Nature 341, 497-503.

Brelsford, A., MilÁ, B., Irwin, D.E., 2011. Hybrid origin of Audubon's warbler. Molecular Ecology 20, 2380-2389.

Brodsky, L.M., Ankney, C.D., Dennis, D.G., 1988. The influence of male dominance on social interactions in black ducks and mallards. Animal Behaviour 36, 1371-1378.

Burney, D.A., James, H.F., Burney, L.P., Olson, S.L., W Kikuchi, Wagner, W.L., Burney, M., 2001. Holocene lake sediments in the Maha'ulepu caves of Kaua'i: evidence for a diverse biotic assemblage from the Hawaiian lowlands and its transformation since human arrival. Ecological Monographs 71, 615–642.

Consortium, H.G., 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487, 94-98.

Cooper, A., Rhymer, J., James, H.F., Olson, S.L., Mclntosh, C.E., Sorenson, M.D., Fleischer, R.C., 1996. Ancient DNA and island endemics. Nature 381, 484-484.

DeMarais, B.D., Dowling, T.E., Douglas, M.E., Minckley, W.L., Marsh, P.C., 1992. Origin of Gila seminuda (Teleostei: *Cyprinidae*) through introgressive hybridization: implications for evolution and conservation. Proceedings of the National Academy of Sciences of the United States of America 89, 2747-2751.

Dobzhansky, T., 1940. Speciation as a stage in evolutionary divergence. The American Naturalist 74, 312-321.

Dowling, T.E., Secor, C.L., 1997. The role of hybridization and introgression in the diversification of animals. Annual Review of Ecology and Systematics 28, 593-619.

Engilis, A., Uyehara, K.J., Giffin, J.G., 2002. Hawaiian Duck: *Anas Wyvilliana*. Birds of North America, Incorporated.

Engilis Jr, A., Pyle, R., David, R., 2004. Status and occurrence of migratory birds in the Hawaiian Islands: PART 1—Anseriformes: Anatidae (Waterfowl). Bishop Museum Occasional Paper 81.

Excoffier, L., Lischer, H.E.L., 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.

Fowler, A., Eadie, J., Engilis, A., 2009. Identification of endangered Hawaiian ducks (*Anas wyvilliana*), introduced North American mallards (*A. platyrhynchos*) and their hybrids using multilocus genotypes. Conservation Genetics 10, 1747-1758.

Gompert, Z., Fordyce, J.A., Forister, M.L., Shapiro, A.M., Nice, C.C., 2006. Homoploid hybrid speciation in an extreme habitat. Science 314, 1923-1925.

Hermansen, J.S., SÆTher, S.A., Elgvin, T.O., Borge, T., Hjelle, E., SÆTre, G.-P., 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Molecular Ecology 20, 3812-3822.

Husband, B.C., 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. Proceedings of the Royal Society of London. Series B: Biological Sciences 267, 217-223.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254-267.

Huynen, L., Millar, C.D., Scofield, R., Lambert, D.M., 2003. Nuclear DNA sequences detect species limits in ancient moa. Nature 425, 175-178.

Jacobsen, F., Omland, K.E., 2011. Increasing evidence of the role of gene flow in animal evolution: hybrid speciation in the yellow-rumped warbler complex. Molecular Ecology 20, 2236-2239.

Johnson, K.P., Sorenson, M.D., 1999. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): A comparison of molecular and morphological evidence. The Auk 116, 792-805.

Keller, I., Wagner, C.E., Greuter, L., Mwaiko, S., Selz, O.M., Sivasundar, A., Wittwer, S., Seehausen, O., 2012. Population genomic signatures of divergent adaptation,

gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular Ecology 22, 2848-2863.

Kunte, K., Shea, C., Aardema, M.L., Scriber, J.M., Juenger, T.E., Gilbert, L.E., Kronforst, M.R., 2011. Sex chromosome mosaicism and hybrid speciation among tiger swallowtail butterflies. PLoS Genetics 7, e1002274.

Lavretsky, P., McCracken, K.G., Peters, J.L., 2014. Phylogenetics of a recent radiation in the mallards and allies (Aves: *Anas*): Inferences from a genomic transect and the multispecies coalescent. Molecular Phylogenetics and Evolution 70, 402–411.

Livezey, B.C., 1991. A phylogenetic analysis and classification of recent dabbling ducks (Tribe *Anatini*) based on comparative morphology. The Auk 108, 471-507.

Livezey, B.C., 1993. Comparative morphometrics of *Anas* ducks, with particular reference to the Hawaiian Duck *Anas wyvilliana*, Laysan Duck *A. laysanensis*, and Eaton's Pintail *A. eatoni*. Wildfowl 44, 75-100.

Mallet, J., 2005. Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20, 229-237.

Mallet, J., 2007. Hybrid speciation. Nature 446, 279-283.

MavÁRez, J., Linares, M., 2008. Homoploid hybrid speciation in animals. Molecular Ecology 17, 4181-4185.

Mayr, E., 1963. Animal Species and Evolution. Belknap Press of Harvard University Press, Cambridge.

Moulton, D.W., Marshall, A.P., 1996. Laysan Duck (*Anas laysanensis*), Philadelphia, PA.

Nielsen, R., Wakeley, J., 2001. Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. Genetics 158, 885-896.

Nolte, A.W., Freyhof, J., Stemshorn, K.C., Tautz, D., 2005. An invasive lineage of sculpins, Cottus sp. (*Pisces*, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. Proceedings of the Royal Society B: Biological Sciences 272, 2379-2387.

Nolte, A.W., Tautz, D., 2010. Understanding the onset of hybrid speciation. Trends in Genetics 26, 54-58.

Olson, S.L., James, H.F., 1991. Descriptions of thirty-two new species of birds from the Hawaiian Islands, Part I. Non-Passeriformes. Ornithological Monographs 45, 1-88.

Olson, S.L., Ziegler, A.C., 1995. Remains of land birds from Lisianski Island, with observations on the terrestrial avifauna of the northwestern Hawaiian Islands. Pacific Science 49, 111-125.

Peters, J.L., McCracken, K.G., Zhuravlev, Y.N., Lu, Y., Wilson, R.E., Johnson, K.P., Omland, K.E., 2005. Phylogenetics of wigeons and allies (Anatidae: *Anas*): the importance of sampling multiple loci and multiple individuals. Molecular Phylogenetics and Evolution 35, 209-224.

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

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

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

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. . Genetics 155, 945–959.

Pyle, R., Pyle, P., 2009. The birds of the Hawaiian islands: occurrence, history, distribution, and status. BP Bishop.

Pyle, R.L., 1988. Checklist of the Birds of Hawaii, 1988.

Ramsey, J., Schemske, D.W., 2002. Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33, 589-639.

Reichel, J.D., Lemke, T.O., 1994. Ecology and extinction of the Mariana mallard. The Journal of Wildlife Management 58, 199-205.

Rentsch, J.D., Leebens-Mack, J., 2012. Homoploid hybrid origin of Yucca gloriosa: intersectional hybrid speciation in Yucca (*Agavoideae, Asparagaceae*). Ecology and Evolution 2, 2213-2222.

Rheindt, F.E., Edwards, S.V., 2011. Genetic introgression: an integral but neglected component of speciation in birds. The Auk 128, 620-632.

Rhymer, J.M., 2006. Extinction by hybridization and introgression in anatine ducks. Acta Zoologica Sinica 52 (Supplement), 583-586.

Rieseberg, L.H., 2006. Hybrid speciation in wild sunflowers. Annals of the Missouri Botanical Garden 93, 34-48.

Salazar, C., Baxter, S.W., Pardo-Diaz, C., Wu, G., Surridge, A., Linares, M., Bermingham, E., Jiggins, C.D., 2010. Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. PLoS Genetics, p. e1000930.

Schwarz, D., Matta, B.M., Shakir-Botteri, N.L., McPheron, B.A., 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. Nature 436, 546-549.

Seehausen, O., 2004. Hybridization and adaptive radiation. Trends in Ecology & Evolution 19, 198-207.

Seehausen, O., 2006. Conservation: losing biodiversity by reverse speciation. Current Biology 16, R334-R337.

Soltis, D.E., Soltis, P.S., Tate, J.A., 2004. Advances in the study of polyploidy since plant speciation. New Phytologist 161, 173-191.

Stephens, M., Donnelly, P., 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. The American Journal of Human Genetics 73, 1162-1169.

Willerslev, E., Cooper, A., 2005. Review paper. Ancient DNA. Proceedings of the Royal Society B: Biological Sciences 272, 3-16.

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

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L.H., 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences of the United States of America 106, 13875-13879.

CHAPTER V. MAJOR HISTOCOMPATIBILITY I GENE DIVERSITY IN THE CRITICALLY ENDANGERED LAYSAN DUCK (*ANAS LAYSANENSIS*)

Abstract – Quantifying the genetic composition of founder populations is important to the success of reintroduction programs, especially for bottlenecked and/or specialized species, such as island endemics. By implementing admixture schemes based on genetic variability, captive breeding programs can minimize detrimental genetic effects (e.g., bottlenecking, inbreeding depression, outbreeding depression). Particular attention has been given to genes within the major histocompatibility complex (MHC) due to their direct correlation to an individual's immunity. However, isolating and amplifying MHC haplotypes remains difficult owing to the high diversity and paralogous nature. I describe a method of MHC I haplotype isolation based on an iterative process of primer design for the endangered island endemic, the Laysan duck (*Anas laysanensis*). Ultimately, haplotype-specific primers allow for direct genotyping after gel electrophoresis based on the presence/absence of their respective amplicons. Using the developed techniques, a total of eight unique haplotypes were isolated and assayed across 21 Laysan duck individuals from Laysan Island (N = 10) and Midway Atoll (N = 11). The presence/absence of seven haplotypes were variable across individuals with three haplotypes present in 95% of individuals, three in 38% of individuals, and one in 90% of individuals. The protocols described herein provide a simple, cost-effective method for isolating haplotypes and monitoring existing MHC variation in Laysan ducks, and the general approach can be applied to other molecular markers and species with low genetic diversity.

INTRODUCTION

When introducing species to novel locations, it is important to assess the adaptive capability of those individuals prior to release (Frankham *et al.* 1986, Lande 1988, Brekke *et al.* 2011). This is particularly important for island species that naturally might have low genetic variability due to demographic constraints and/or might be specialized for certain biotic and abiotic conditions (Hedrick and Kalinowski 2000, Jamieson *et al.* 2006). For instance, the translocation of island species can cause immediate isolation, and without a genetically admixed founder population, deleterious alleles due to breeding between homozygous individuals can lead to a loss of adaptability (Keller and Waller 2002, Briskie and Mackintosh 2004). Introductions of endangered species can be further complicated by the relatively low number of remaining individuals, which are likely already genetically similar (Spielman *et al.* 2004). Captive breeding programs can typically rescue such endangered populations (Doyle *et al.* 2001, Frankham 2008); however, the need for constant augmentation can persist if maladaptive individuals are used (Vrijenhoek 1998, Doyle *et al.* 2001, Woodworth *et al.* 2002). Conservation efforts can benefit by initially quantifying available genetic variability that then can be used for admixture schemes (i.e. breeding individuals that are genetically dissimilar) to maximize genomic variability in the founding population. Moreover, monitoring these reintroductions provides a way to study the effects of genetic drift or loss of genetic variability due to stochastic processes in wild populations that may not be evident in laboratory settings (Frankham 2000, Brekke *et al.* 2011). Marker development has primarily focused on neutral or non-coding regions (e.g. microsatellites, introns) that are

largely influenced by stochastic processes (e.g. population size) rather than selective forces (Kimura, 1985). Consequently, neutral markers may not directly correspond to a population's adaptive diversity (Holderegger *et al.*, 2006). For instance, some studies have shown a correlation between neutral and non-neutral markers (Mikko and Anderson, 1995; Campos *et al.*, 2006), while others have not (Hansson and Richardson, 2005; van Oosterhout et al. 2006). Consequently, conservation initiatives, specifically during reintroductions with captive populations should include markers across the genome, including those that correspond to an individual's adaptive potential.

Coding for antigen recognition (Lundqvis *et al.* 2001), major histocompatibility complex (MHC) genes are the cornerstone of an individual's immune system (Klein 1986), and have become a focal non-neutral marker in population and conservation genetics (Sommer 2005). Populations with higher levels of MHC polymorphism often rebound faster when encountering novel diseases or stochastic environmental events (Apanius *et al.* 1997, Beacham *et al.* 2004, Miller *et al.* 2004, Neff *et al.* 2008), and MHC variability can be maintained in populations or species that are predominantly homogenic (Hansson and Richardson 2005, van Oosterhout et al. 2006), as well as selectively driven by local parasitic environments (Sommer 2005; Spurgin and Richardson, 2010). Genic duplications and positive/balancing selection have been attributed to higher MHC locus and allelic heterogeneity, respectively. Unfortunately, the high diversity and paralogy of MHC I genes has made it difficult to directly isolate loci and/or haplotypes (Moon *et al.* 2005). Typically, MHC genes are isolated through cDNA cloning and sequencing, which is labor intensive, costly, and does not always yield primers that can be used with genomic DNA (Lundqvis *et al.* 2001, Moon *et al.* 2005,

Skinner *et al.* 2009). Recently, high-throughput methods have also been applied to MHC studies (Babik *et al.* 2009, Ekblom *et al.* 2010), but again, these methods are expensive and time-consuming (see review of techniques in Babik 2010). The objective of this study was to isolate MHC I variants in the critically endangered Laysan duck (*Anas laysanensis*) through an iterative process of designing primers that specially target individual MHC I haplotypes. The process bypasses cloning and permits the use of genomic DNA as a template that is more stable than RNA, which is typically used in the cDNA cloning process. Once haplotype-specific primers are developed, my method allows the detection of MHC I variants in a presence/absence framework that is low-cost and time-efficient.

Study System

The Hawaiian Islands are a biodiversity hotspot, but anthropogenic changes have endangered numerous species (Olson and Ames 1982). Of the 113 endemic bird species once found across the Hawaiian Islands, 71 are extinct and 31 are currently federally listed (http://www.abcbirds.org/newsandreports/releases/080918.html). Laysan duck populations were decimated through the introduction of non-native fauna, and by 1912, there were approximately 12 individuals left in the wild (Dill and Bryan 1912). The entire population was confined to Laysan Island where they specialized on hyper-saline wetlands. However, whether this was a facultative or obligate adaptation is debatable as they historically occurred across the Hawaiian archipelago (Olson and James 1991, Cooper *et al.* 1996, Burney *et al.* 2001). To date, several translocations have been attempted with both failures (e.g. Pearl and Hermes Reef; Berger 1981) and successes (e.g. Midway Atoll; Reynolds and Klavitter 2006). As a result of conservation initiatives,

~1000 Laysan ducks now inhabit Laysan and Midway Atoll ("Midway") (Reynolds and Klavitter 2006, Reynolds *et al.* 2013). However, with the entire species fluctuating between 100 and 600 individuals, the Laysan duck is vulnerable to stochastic environmental events and novel pathogen introductions (e.g. 2008 botulism outbreak on Midway;Work *et al.* 2010). To decrease the probability of extinction by stochastic events, primary conservation initiatives are to establish Laysan duck populations on neighboring islands (Butchart and Hughes 2003, USFWS 2004). Determining MHC diversity in extant Laysan duck populations will benefit future reintroductions by providing a tool for maximizing genetic diversity, and hence adaptability, of founder populations prior to release.

MATERIALS AND METHODS

Sample

The University of California, Davis Museum of Wildlife and Fish Biology (MWFB) provided tissue samples from 21 specimens (Laysan = 10, Midway = 11) archived at the MWFB. The Midway population was established in 2004-2005 with a total of 43 individuals from Laysan Island and has since grown to ~100 individuals (Reynolds *et al.* 2008). Specimens were provided to the MWFB from the US Geological Survey, National Wildlife Health Center - Honolulu Field Station and US Fish and Wildlife Service. Tissue samples (breast and leg muscle) were sampled at the lab of the MWFB, where they are archived along with round skins of adult birds as voucher

specimens. Genomic DNA was isolated from each tissue sample using a Qiagen DNA extraction kit (Qiagen) according to the manufacturer's protocol.

MHC Markers

Believed to be one of the most polymorphic regions in the vertebrate genome (Lundqvis *et al.* 2001), I targeted exon 2 of MHC I that codes for the peptide binding region (PBR) (Promerová *et al.*, 2009). First, the exon 2 region was amplified using published primers D26E2R1/D26E2F1 (1263 bp; Moon *et al.* 2005) and degenerate primers E2R/E2F (~355 bp) and E2R2/E2F2 (~238 bp) that were designed across conserved regions of published MHC I mallard (*Anas platyrhynchos*) sequences obtained from GenBank (Appendix Table A5.1). PCR amplification was performed using 1.5 μL template DNA (≥10 ng/μl), 2x GoTaq Green Master Mix (Promega), and 1 nM of each primer, in a total volume of 15 μL. PCR was conducted using an Eppendorf Mastercycler (ep*gradient*) thermocycler under the following conditions: DNA denaturation at 94°C for 7 minutes, followed by 45 cycles of DNA denaturation at 94°C for 20 s, primer annealing at primer-specific temperatures (Appendix Table A5.2) for 20 s, and DNA extension at 72°C for 1 minute, and a final DNA extension at 72°C for 7 minutes. Amplification was verified using gel electrophoresis with a 1.5% agarose gel and the presence of a band corresponding to product lengths (Appendix Table A5.2). PCR products were cleaned with AMPure XP beads, following Agencourt protocol (Beckman Coulter Co.). Sequencing was done using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) following manufacturer protocols. Automated Sanger sequencing was conducted at the DNA Analysis Facility at Yale University on an ABI 3730. Sequences were aligned and edited using Sequencher v. 4.8 (Gene Codes, Inc).

Multiple haplotypes were obtained using primer sets E2R/E2F and E2R2/E2F2. Using these sequences and following the concept of allele-specific priming (Bottema and Sommer 1993), I designed new reverse primers targeting polymorphic sites which resulted in amplification of variants with those nucleotide(s) (Appendix Table A5.2); these included the combination of E2R2 with MHC1aF, MHC1bF, and MHC1cF, designated as primer sets 1a, 1b, and 1c, respectively. Primer sets 1a and 1b targeted a maximum of two variants and primer set 1c targeted a single MHC I variant within individuals (Appendix Table A5.2). Subsequently, primers targeting each haplotype individually were designed (see Results*;* Appendix Table A5.2). Primer specificity was maximized by increasing the number of nucleotide mismatches between haplotypes providing a presence/absence framework for examining genetic variation (Table 5.1) – an amplicon was obtained only when the primer matched a haplotype variant present within the individual's genome. Developing such primer pairs is an effective method for delineating alleles/haplotypes in loci that have undergone duplication events (Lavretsky *et al.* 2012). Moreover, the primers were designed to target amplicons varying in length to permit the pooling of PCR products of a single individual prior to gel electrophoresis.

PCR conditions were similar to those described above, but in some cases, annealing temperatures varied (Appendix Table A5.2) and a total volume of 10 μL that included 1μL of template DNA (≥10 ng/μl), 2x GoTaq Green Master Mix (Promega), and 1 nM of each primer was used. Although, non-targeted amplicons were present for certain primer pairs, these products were easily distinguishable on an agarose gel from the desired ones (Fig. 5.1). Primer specificity was increased by using a touch down method (TD-PCR) (Korbie and Mattick, 2008) that eliminated the secondary products

(Appendix Table A5.2 & A5.3). TD-PCR uses an initial annealing temperature above the primer-specific temperature (I used +5^{o}C) and progressively transitions to lower temperatures in successive cycles and was the optimum condition for amplification of haplotypes 1,3,4,6, and 7 (Appendix Table A5.3).

Table 5.1. Presence (shaded) or absence (open) of MHC I exon 2 haplotypes per population for each Laysan duck individual.

	Laysan										Midway										
Haplotyp	WFB8621	WFB8639	WFB8622	WFB8708	WFB8643	WFB8709	WFB1052	WFB1052	WFB1053	WFB1140	WFB1050	WFB1050	WFB1050	WFB1050	WFB1050	WFB1050	WFB1051	WFB9381	WFB8638	WFB8710	WFB8711
1	■	■			■	■		■	■										■	■	
2	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■
3	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
4	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
5	■	■			■	■		■	■										■	■	
6	■	■			■	■		■	■										■	■	
7	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
8	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

The PCR conditions were DNA denaturation at 94°C for 7 minutes, followed by 5 successive cycles of denaturation at 94°C for 15 s, primer annealing at 71-66°C (for haplotypes 1 and 6) or 65-60°C (for haplotypes 3, 4, and 7) for 15 s decreasing by 1°C in each successive cycle, and DNA extension at 72°C for 45 s. This was then followed by

30 cycles of denaturation at 94°C for 15 s, primer annealing at 66°C (for haplotypes 1 and 6) or 60°C (for haplotypes 3, 4, and 7) for 15 s, and DNA extension at 72°C for 45 s, after which a final DNA extension at 72°C for 7 minutes occurred. Subsequent product verification was based on presence/absence of products on a 1.5% agarose gel. PCR products from a subset of individuals were cleaned and sequenced using the above protocols to verify that primers were targeting desired haplotypes.

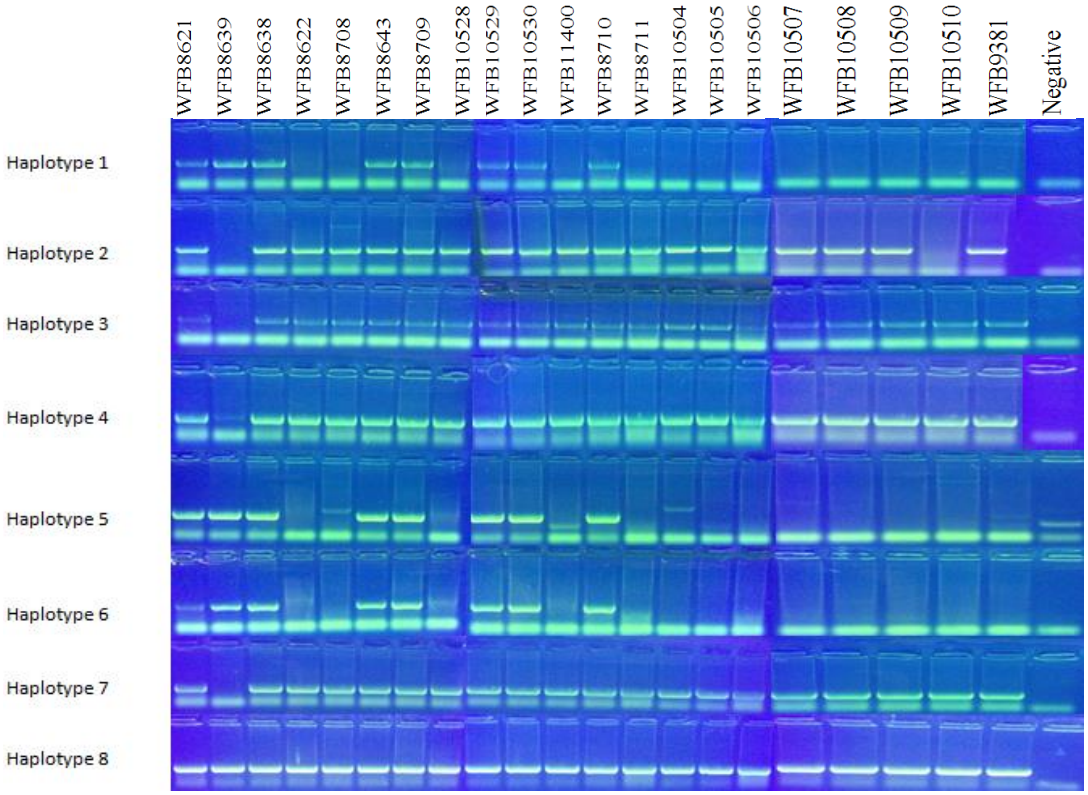


Figure 5.1. Gel electrophoresis of 8 MHC I exon 2 haplotypes for 21 Laysan ducks.

Phylogenetic Reconstruction

A phylogenetic MHC I exon 2 gene tree with unconstrained branch lengths was constructed using mallard and Laysan duck sequences (Appendix Table A5.2) in

MrBayes (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003) and viewed in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree). The data were separated by codon position and evaluated using a General Time Reversible (GTR) model and gamma-distributed rates across sites. Two separate analyses were run for 2 million generations, with sampling every 200 iterations until the standard deviation between sampling events was < 0.01. The first 25% of the samples were discarded as burn in.

RESULTS

The primers D26E2R1/D26E2F1 from Moon *et al.* (2005) yielded a single haplotype (HAP_7; Table 5.3) across Laysan duck individuals. Degenerate primer set E2R/E2F revealed highly polymorphic sequences (>2 variants) for 14 individuals and non-polymorphic sequences for 7 individuals (HAP_1). Primer set E2R2/E2F2 produced sequences with >2 variants at multiple base positions for 9 individuals and non-polymorphic sequences for 12 individuals (HAP_3; Appendix Table A5.2).

A total of 197 bp, 238 bp, and 197 bp of the peptide binding region were amplified with primer sets 1a, 1b, and 1c, respectively. Primer sets 1a and 1b yielded a mix of heterozygous and homozygous individuals. Specifically, for primer set 1a, a total of 13 individuals were homozygous for HAP_2 (or 3) (HAP_2/3 = primers do not amplify the region containing polymorphisms distinguishing the two haplotypes) whereas 8 individuals were heterozygous for HAP_5 and HAP_6. For primer set 1b, 1 individual was homozygous for HAP_1, 5 individuals were homozygous for HAP_2/3, 8 individuals were heterozygous for HAP_1 and HAP_2, and 7 individuals were heterozygous for

HAP_3 and HAP_4. No individuals were heterozygous for any combination of HAP_1 and HAP_3. The combination of products from 1a and 1b yielded the same polymorphic positions as observed for individuals that contained polymorphic sites for primer set E2R/E2F or E2R2/E2F2 (see above). Primer set 1c was haplotype specific, with 20 individuals' sequences yielding HAP_2/3 and 1 individual having a novel haplotype, HAP_8 (Appendix Table A5.2). Finally, the haplotype obtained with D26E2R1/D26E2F1 (HAP_7) was not amplified using primer sets 1a, 1b, or 1c. Phylogenetic results demonstrate that many of the haplotypes are alleles of different loci (Fig. 5.2).

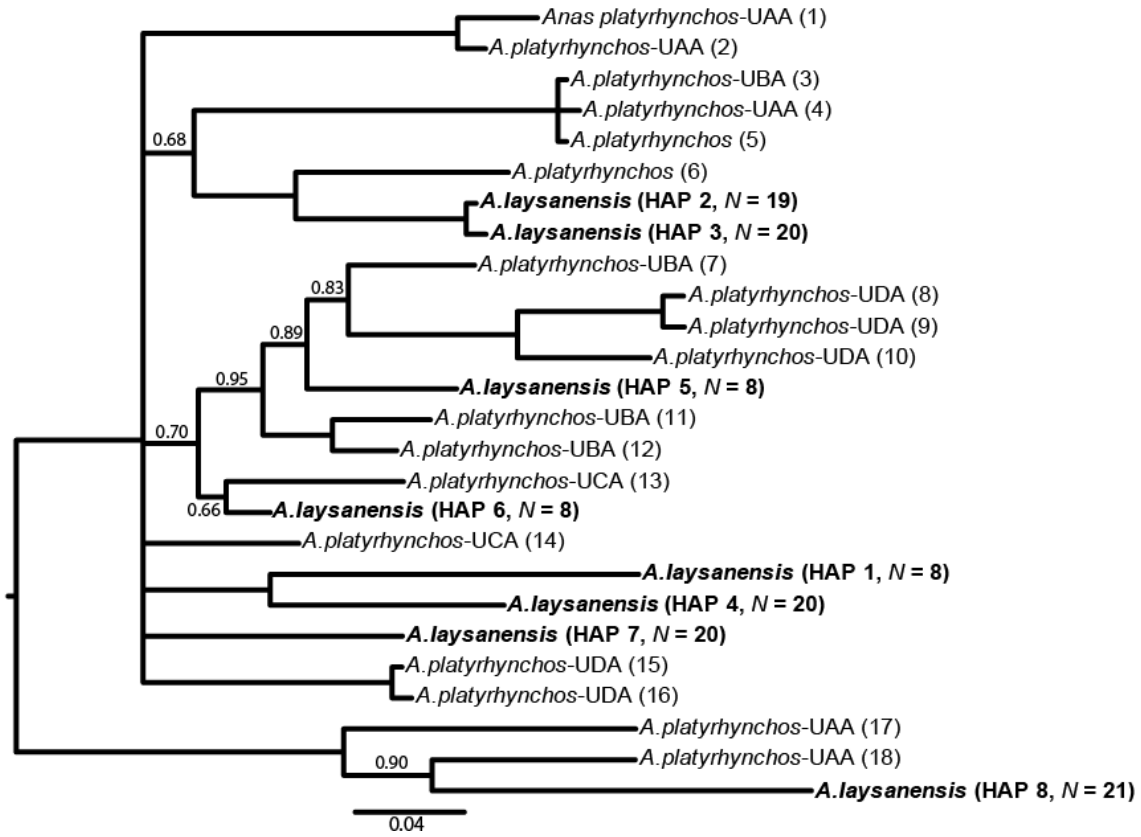
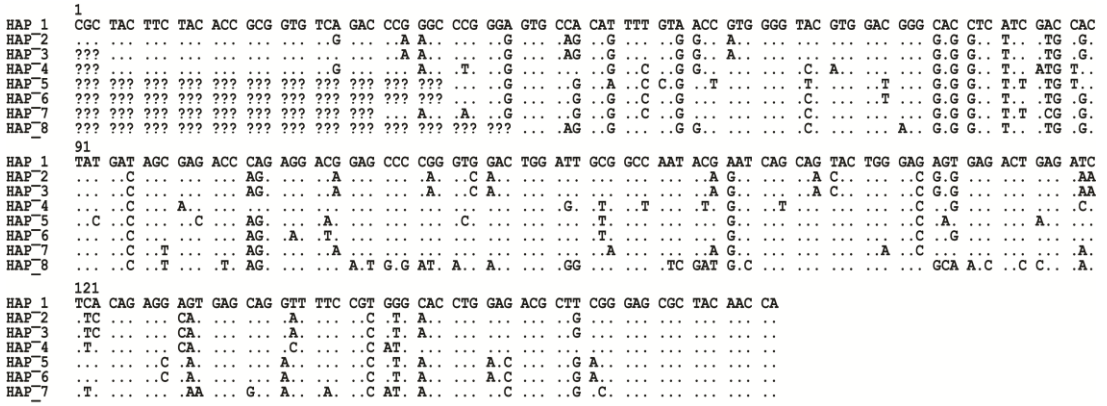


Figure 5.2. A Bayesian tree for MHC I variants in Laysan ducks and mallards. Previously identified MHC I loci are provided for mallard sequences (Appendix Table A5.1) following a dash. *N* indicates the total number of Laysan duck individuals observed with each respective haplotype.

(A)



(B)

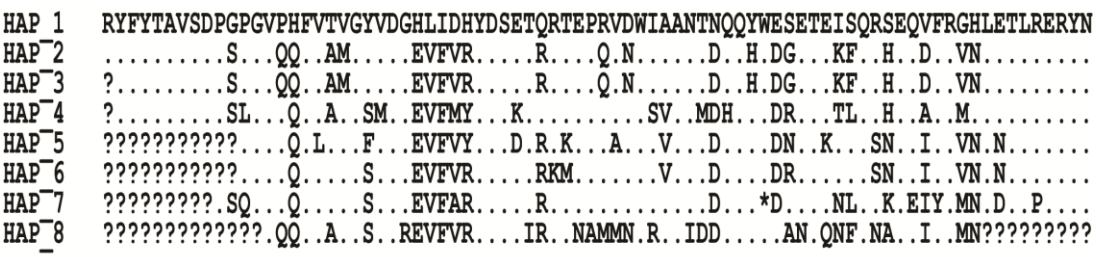


Figure 5.3. (A) Sequence and (B) amino acid alignments for 8 MHCI Laysan duck haplotypes isolated in this study. Dots indicate identity with the top alignment. A bolded star indicates a stop-codon in the protein alignment.

Presence/Absence Analysis

Using newly developed primers that targeted each haplotype individually (Appendix Table A5.2), I documented genetic variation for the presence/absence of all but one haplotype. HAP_8 was present in all individuals (Table 5.1; Fig. 5.1) despite initially being found only in one individual. HAP_2 was present in 19 individuals, whereas HAP_3 and HAP_4 were in 20 individuals. Finally, HAP_1, HAP_5 and HAP_6 were present in 8 individuals, and appear to be in high linkage disequilibrium.

All haplotypes were protein coding with no stop codons except HAP_7 that had a single stop codon, suggesting that it is a pseudogene (Fig. 5.3A-B), which are known to

occur in passerine MHC class I and II genes (Westerdahl *et al.* 1999, Edwards *et al.* 2000, Reusch *et al.* 2004). Laysan MHC haplotypes were phylogenetically intermixed with mallard MHC I haplotypes, although the pairs HAP_2/HAP_3 and HAP_1/HAP_4 were sister lineages, respectively (Fig. 5.2). Nevertheless, the MHC gene tree reveals that isolated Laysan duck haplotypes likely span across all major loci previously determined to comprise MHC I in ducks (Xia *et al.* 2004, Mesa *et al.* 2004, Moon *et al.* 2005). Specifically, sister relationships between *A. platyrhynchos* 18 and Laysan HAP_8, *A. platyrhynchos* 11 and Laysan HAP_5, *A. platyrhynchos* 13 and Laysan HAP_6 suggest that these haplotypes comprise alleles of the UAA, UBA, and UCA loci, respectively.

DISCUSSION

MHC Haplotype Identification

Phylogenetic analyses delineated that at least three different loci were amplified using my degenerate and haplotype-specific primers (Fig. 5.2) and that much of the sequence variation is likely across loci rather than within loci. However, sister relationships between HAP_1 and HAP_4, as well as HAP_2 and HAP_3 suggest that these might be alleles of the same locus (Fig. 5.2). Consequently, 20 individuals might have been heterozygous for HAP_2 and HAP_3, whereas one individual was homozygous for HAP_3. Likewise, assuming HAP_1 and HAP_4 are the same locus would suggest that one individual was homozygous for HAP_1, 13 individuals were homozygous for HAP_4, and seven were heterozygous for HAP_1 and HAP_4 (Table 5.1; Fig. 5.1). However, assigning haplotypes to MHC loci is exceedingly difficult,

because highly divergent alleles can be found at a single locus whereas more similar alleles can correspond to different loci (Moon *et al.* 2005). Therefore, sequence variation at any single locus cannot be conclusively demonstrated without further analyses. Similarly, the absence of haplotypes in individuals–as the case for HAP_5 and HAP_6 that are likely alleles of different loci (Fig. 5.2) – suggests that additional alleles of those loci are present but are not being amplified with my methods. Although, future work can capture these "missing" haplotypes through additional primer pair iterations or cloning, the presence/absence of these haplotypes still represents MHC I variability. As a result, while the methods are not specific enough to identify alleles of loci, the designed presence/absence framework still readily provides measures of MHC I diversity.

Conservation Implications

Whether using captive bred or wild individuals, it is important to quantify and maintain existing genomic variation of the potential founder population (Frankham 2008). Using homogenic or inbred individuals can increase the chances of disease susceptibility and the fixation of maladaptive traits (Soulé and Wilcox 1980). Without the need to sequence once primer sets are established, methods described here can be applied with low cost. Specifically, designed primers allow for genomic amplification without the need for RNA extraction that is sensitive to rapid degradation (Bustin 2002). Although, described methods do not provide the same amount of information that can be obtained from cDNA conversion and cloning (Lundqvis *et al.* 2001, Moon *et al.* 2005, Skinner *et al.* 2009), they are time and cost effective in readily obtaining genotyping assays, which was the primary objective of this study. I acknowledge that these methods would not be entirely suitable for taxa with high genetic diversity as the number of primer pairs would

significantly increase, and thus these methods are likely to be more effective for species of conservation concern that have experienced an extensive loss of genetic diversity.

Presently, this is the first study to isolate and report on MHC I diversity in the Laysan duck. Although I cannot conclusively provide the total number of loci being amplified, I show that MHC I variation was retained within Laysan duck individuals. A total of four genotypes were described, with 2 shared between Laysan and Midway Islands, as well as one specific to each island (Table 5.1; Fig. 5.1). Consequently, the reintroduction of 43 individuals onto Midway Atoll from Laysan Island appears to have captured MHC I exon 2 variation; however, additional individuals need to be assayed to determine whether the Midway population contains all extant variants from Laysan. Nevertheless, these results have important implications to conservation initiatives for this species, especially with respect to future reintroduction efforts (Reynolds *et al.* 2013). Protocols described here for assaying MHC I variation can be used by breeding programs to establish admixture schemes that in theory can increase the viability of future reintroductions. For example, based on the presence/absence of haplotypes 1, 5, and 6 (Table 5.1; Fig. 5.1), overall heterozygosity would increase in the offspring of individuals WFB8622 and WFB8643. Including their progeny in founder populations will increase the probability that both variants will be maintained in future generations. Although, any further loss of MHC variability can be detrimental to the species overall adaptability and future survival (Hughes 1991), I caution against basing reintroductions on a single gene due to possible negative effects of outbreeding depression or loss of diversity at other loci (Amos and Balmford 2001, Neff 2004). As a result, future work should include examining additional molecular markers (e.g. introns, microsatellites) and increasing

sample sizes to provide a more conclusive measure of overall genetic variability between Laysan duck individuals. Finally, I suggest populations be evaluated for genetic variability for several generations after the initial reintroduction and augmented with additional individuals to maintain or increase variability as described in the recovery plan for the Laysan duck (USFWS 2004). The protocol that I described can be used as a tool for these efforts.

ACKNOWLEDGEMENTS

I thank Andy Engilis, Jr., Peter L. Gibert, Mana Hattori, and Irene Engilis from the University of California-Davis Museum of Wildlife and Fish Biology for assistance and specimen preparation. I thank Robert Rameyer of the USGS National Wildlife Health Center – Honolulu Field Office for donation of specimens. Funds were provided by Ducks Unlimited Richard H. G. Bonnycastle Fellowship in Wetland and Waterfowl Biology.

REFERENCES

Amos, W., Balmford, A., 2001. When does conservation genetics matter? Heredity 87, 257-265.

Apanius, V., Penn, D., Slev, P. R., Ruff, L. R., Potts, W. K., 1997. The nature of selection on the major histocompatibility complex. Critical Reviews in Immunology 17, 79-224.

Babik, W., Taberlet, P., Ejsmond, M. J. , Radwan, J., 2009. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Molecular Ecology Resources 9, 713-719.

Babik, W., 2010. Methods for MHC genotyping in non-model vertebrates. Molecular Ecology Resources 10, 237-251.

Beacham, T. D., Lapointe, M., Candy, J. R., Miller, K. M., Withler, R. E., 2004. DNA in action, Rapid application of DNA variation to sockeye salmon fisheries management. Conservation Genetics 5, 411-416.

Berger, A., 1981. Hawaiian Birdlife. University of Hawaii Press, Honolulu, HI.

Bottema, C. D. K., Sommer, S. S., 1993. PCR amplification of specific alleles, rapid detection of known mutations and polymorphisms. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 288, 93-102.

Brekke, P., Bennett, P. M., Santure, A. W. , Ewen, J. G., 2011. High genetic diversity in the remnant island population of hihi and the genetic consequences of re-introduction. Molecular Ecology 20, 29-45.

Briskie, J. V. , Mackintosh, M., 2004. Hatching failure increases with severity of population bottlenecks in birds. Proceedings of the National academy of Sciences of the United States of America 101, 558–561.

Burney, D. A., James, H. F., Burney, L. P., Olson, S. L., W Kikuchi, Wagner, W. L., Burney, M., 2001. Holocene lake sediments in the Maha'ulepu caves of Kaua'i, evidence for a diverse biotic assemblage from the Hawaiian lowlands and its transformation since human arrival. Ecological Monographs 71, 615–642.

Bustin, S., 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR), trends and problems. Journal of Molecular Endocrinology 29, 23-39.

Butchart, S., Hughes, B., 2003. 2004 Red list changes for waterfowl. Threatened Waterfowl Res. Group Newsletter 14, 82–84.

Campos, J. L., Posada, D., Mora´n, P., 2006. Genetic variation at MHC, mitochondrial and microsatellite loci in isolated populations of Brown trout (*Salmo trutta*). Conservation Genetics 7, 515-530.

Cooper, A., Rhymer, J., James, H. F., Olson, S. L., Mclntosh, C. E., Sorenson, M. D., Fleischer, R. C., 1996. Ancient DNA and island endemics. Nature 381, 484-484.

Dill, H. R., Bryan, W. A., 1912. Report of an expedition to Laysan Island in 1911. US Department of Agriculture Biological Survey Bulletin 42.

Doyle, R. M., Perez-Enriquez, R., Takagi, M., Taniguchi, N., 2001. Selective recovery of founder genetic diversity in aquacultural broodstocks and captive, endangered fish populations. Genetica 111, 291-304.

Edwards, S. V., Nusser, J., Gasper, J., 2000. Characterization and evolution of major histocompatibility complex (MHC) genes in non-model organisms, with examples from birds. Molecular Methods in Ecology 168-207.

Ekblom, R., Balakrishnan, C. N., Burke, T., Slate, J., 2010. Digital gene expression analysis of the zebra finch genome. BMC Genomics 11, 219.

Frankham, R., Hemmer, H., Ryder, O. A., Cothran, E. G., Soulé, M. E., Murray, N. D., Snyder, M., 1986. Selection in captive populations. Zoo Biology 5, 127-138.

Frankham, R., 2000. Quantitative Methods. Pp. in Conservation Biology ed by Springer, New York.

Frankham, R., 2008. Genetic adaptation to captivity in species conservation programs. Molecular Ecology 17, 325-333.

Hansson, B., Richardson, D. S., 2005. Genetic variation in two endangered *Acrocephalus* species compared to a widespread congener, estimates based on functional and random loci. Animal Conservation 8, 83-90.

Hedrick, P. W., Kalinowski, S. T., 2000. Inbreeding depression in conservation biology. Annual Review of Ecology, Evolution, and Systematics 31, 139–162.

Holderegger, R., Kamm, U., Gugerli, F., 2006. Adaptive vs. neutral genetic diversity, implications for landscape genetics. Landscape Ecology 21, 797-807.

Huelsenbeck, J. P., and Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Hughes, A. L., 1991. MHC polymorphism and the design of captive breeding programs. Conservation Biology 5, 249–251.

Jamieson, I. G., Wallis, G. P., Briskie, J. V., 2006. Inbreeding and endangered species management, is New Zealand out of step with the rest of the world. Conservation Biology 20, 38.

Keller, L. F., Waller, D. M., 2002. Inbreeding effects in wild populations. Trends in Ecology & Evolution 17, 230–241.

Kimura, M., 1985. The neutral theory of molecular evolution, Cambridge University Press.

Korbie, D. J., Mattick, J. S., 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nature Protocols 3, 1452-1456.

Klein, J., 1986. Natural History of the Major Histocompatibility Complex. Wiley New York.

Lande, R., 1988. Genetics and demography in biological conservation. Science 241, 1455-1460.

Lavretsky, P., Truong, T., McGowin, A., Balazs, G., Peters, J., 2012. New primers reveal the presence of a duplicate histone H3 in the marine turtle leech *Ozobranchus branchiatus*. Conservation Genetics Resources 1-4.

Lundqvis, M. L., Middleton, D. L., Hazard, S., Warr, G. W., 2001. The immunoglobulin heavy chain locus of the duck, genomic organization and expression of D, J, and C region genes. The Journal of Biological Chemistry 276, 46729–46736.

Mesa, C. M., Thulien, K. J., Moon, D. A., Veniamin, S. M., Magor, K. E., 2004. The dominant MHC class I gene is adjacent to the polymorphic TAP2 gene in the duck, *Anas platyrhynchos*. Immunogenetics 56, 192-203.

Mikko, S., Anderson, L., 1995. Low major histocompatibility complex class II diversity in European and North American moose. Proceedings of the National Academy of Sciences USA 92, 4259-4263.

Miller, K. M., Winton, J. R., Schulze, A. D., Purcell, M. K., Ming, T. J., 2004. Major histocompatibility complex loci are associated with susceptibility of Atlantic salmon to infectious hematopoietic necrosis virus. Environmental Biology of Fishes 69, 307-316.

Moon, D. A., Veniamin, S. M., Parks-Dely, J. A., Magor, K. E., 2005. The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes1. The Journal of Immunology 175, 6702-6712.

Neff, B. D., 2004. Stabilizing selection on genomic divergence in a fish population. Proceedings of the National academy of Sciences of the United States of America USA 101, 2381-2385.

Neff, B. D., Garner, S. R., Heath, J. W., Heath, D. D., 2008. The MHC and non-random mating in a captive population of Chinook salmon. Heredity 101, 175–185.

Olson, S. L., Ames, H. F., 1982. Fossil birds from the Hawaiian Islands, evidence for wholesale extinction by man before western contact. Science 217, 633–635.

Olson, S. L., James, H. F., 1991. Descriptions of thirty-two new species of birds from the Hawaiian Islands, Part I. Non-Passeriformes. Ornithological Monographs 45, 1-88.

van Oosterhout, C., Joyce, D. A., Cummings, S. M., Blais, J., Barson, N. J., Ramnarine, I. W., Mohammed, R. S., Persad, N., Cable, J., 2006. Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). Evolution 60, 2562-2574.

Promerová, M., Albrecht, T., Bryja, J., 2009. Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). Immunogenetics 61, 451-461.

Reusch, T. B. H., Schaschl, H., Wegner, K. M., 2004. Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. Immunogenetics 56, 427-437.

Reynolds, M., Klavitter, J., 2006. Translocation of wild Laysan duck *Anas laysanensis* to establish a population at Midway Atoll National Wildlife Refuge, United States and US Pacific Possession. Conservation Evidence 3, 6-8.

Reynolds, M. H., Seavy, N. E., Vekasy, M. S., Klavitter, J. L., Laniawe, L. P., 2008. Translocation and early post-release demography of endangered Laysan teal. Animal Conservation 11, 160-168.

Reynolds, M. H., Weiser, E., Jamieson, I., and Hatfield. J. S., 2013. Demographic variation, reintroduction, and persistence of an island duck (*Anas laysanensis*). The Journal of Wildlife Management 77, 1094-1103.

Ronquist, F., and Huelsenbeck, J. P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Skinner, B., Robertson, L., Tempest, H., Langley, E., Ioannou, D., Fowler, K., Crooijmans, R., Hall, A., Griffin, D., Volker, M., 2009. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. BMC Genomics 10, 357.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology 2, 16.

Soulé, M. E., Wilcox, B. A., 1980. Conservation biology, an evolutionary-ecological perspective. Sinauer Associates, Sunderland, Mass.

Spielman, D., Brook, B. W., Frankham, R., 2004. Most species are not driven to extinction before genetic factors impact them. Proceedings of the National Academy of Sciences of the United States of America 101, 15261–15264.

Spurgin, L. G., Richardson, D. S., 2010. How pathogens drive genetic diversity, MHC, mechanisms and misunderstandings. Proceedings of the Royal Society B, Biological Sciences 277, 979-988.

USFWS. 2004. Draft revised recovery plan for the Laysan duck (*Anas laysanensis*) in U. F. a. W. Service, editor, Portland, OR.

Vrijenhoek, R. C., 1998. Conservation genetics of freshwater fish. Journal of Fish Biology 53, 394-412.

Westerdahl, H., Wittzell, H., Von Schantz, T., 1999. Polymorphism and transcription of Mhc class I genes in a passerine bird, the great reed warbler. Immunogenetics 49, 158-170.

Woodworth, L. M., Montgomery, M. E., Briscoe, D. A., Frankham, R., 2002. Rapid genetic deterioration in captive populations, Causes and conservation implications. Conservation Genetics 3, 277-288.

Work, T. M., Klavitter, J. L., Reynolds, M. H., Blehert, D., 2010. Avian Botulism, A case study in translocated endangered Laysan ducks (*Anas laysanensis*) on Midway Atoll. Journal of Wildlife Diseases 46, 499-506.

Xia, C., Lin, C. Y., Xu, G. X., Hu, T. J., Yang, T. Y., 2004. cDNA cloning and genomic structure of the duck (*Anas platyrhynchos*) MHC class I gene. Immunogenetics 56, 304-309.

APPENDIX

Table A1.1. Species, sub-species, and populations of the mallard complex included in analyses with their respective sample sizes.

Species	Sample Size
Mallard (*Anas platyrhynchos*; MALL)	
Eurasia (OW)	5
North America (NW)	5
American Black Duck (*A. rubripes*; ABDU)	5
Mottled Duck (MODU)	
Florida, FL (*A. f. fulvigula*)	5
Western Gulf Coast, WGC (*A. f. maculosa*)	5
Mexican Duck (*A. p. diazi*; MEDU)	5
Hawaiian Duck (*A .wyvilliana*; HAWD)	5
Laysan Duck (*A. laysanensis*; LADU)	5
Chinese Spot-Billed Duck (*A. zonorhyncha*; $SPBD_{CH}$)	5
Indian Spot-billed Duck (*A. poecilorhyncha*; $SPBD_{IN}$)	1
Philippine Duck (*A. luzonica*; PHDU)	1
African Black Duck (*A. sparsa*; AFBD)	1
Yellow-Billed duck (*A. undulata*; YBDU)	5
Meller's Duck (*A. melleri*; MELL)	1
Pacific Black Duck (*A. superciliosa rogersii*; PBDU)	5
New Zealand Grey Duck (*A. s. superciliosa*; GRDU)	5

Table A1.2. Locus-specific substitution models with associated parameters and the molecular clock (strict vs. Bayesian uncorrelated log-normal relaxed clock) used during species tree reconstruction for datasets ignoring verse filtering for recombination.

	Ignoring Recombination			Recombination-Filtered		
Locus	**MODELS**	**STRICT**	**LOGNORM**	**MODELS**	**STRICT**	**LOGNORM**
Chromo-helicase-DNA binding protein gene 1, intron 19	HKY	X		HKY	X	
Lactate dehydrogenase 1, intron 4	HKY	X		HKY	X	
S-acyl fatty acid synthase thioesterase, intron 2	HKY	X		HKY	X	
Ornithine decarboxylase, intron 7	HKY+ Gamma		X	HKY	X	
Fibrinogen beta chain, intron 7	HKY+ Gamma	X		HKY	X	
Serum amyloid A, intron 2	HKY+ Gamma+ Invariable Sites		X	HKY	X	
Annexin A11, intron 2	HKY+ Gamma	X		HKY		X
Myostatin, intron 2	HKY+ Gamma		X	HKY	X	
Soat1-prov protein, intron 10	HKY	X		HKY		X
Nucleolin, intron 12	HKY+ Gamma+ Invariable Sites		X	HKY	X	
Melanocortin 1 receptor	HKY	X		HKY	X	
Preproghrelin, intron 3	HKY		X	HKY	X	
Glutamate receptor, ionotropic, N-methyl D aspartate I, intron 13	HKY+ Gamma	X		HKY	X	
Sex determining region Y-box 9, intron 2	HKY+ Gamma+ Invariable Sites	X		HKY	X	
Carboxypeptidase D, intron 9	HKY+ Gamma+ Invariable Sites		X	HKY	X	
Phosphenolpyruvate carboxykinase, intron 9	HKY	X		HKY	X	
Alpha enolase 1, intron 8	HKY+ Gamma	X		HKY	X	
Alpha-B crystallin, intron 1	HKY		X	HKY		X
Growth hormone 1, intron 3	HKY+ Gamma	X		HKY	X	
Lecithin-cholesterol acyltransferase, intron 3	HKY+ Gamma+ Invariable Sites	X		HKY	X	

Table A2.1. Sample sizes of each operational taxonomic unit used in this study.

Species	Sample Size
North America (NW) Mallard (*Anas platyrhynchos*)	25
American Black Duck (*A. rubripes*)	24
Mottled Duck (*A. fulvigula*)	
Florida (FL)	24
Western gulf coast (WGC)	25
Mexican Duck (*A.* [*p.*] *diazi*)	25

Table A2.2. Characteristics of 17 nuclear loci sequenced in this study, including locus name, chromosomal location, and the total and recombination filtered lengths (in base pairs).

Locus Name	Chromosomal Location[1]	Total Length / recombination filtered
CHD1Z-b	Z	327 / 327
LDH1-4	1	460 / 460
ODC1-7	3	300 / 131
FGB-7	4	439 / 244
SAA-2	5	306 / 144
ANXA11-2	6	441 / 225
MSTN-2	7	281 / 139
SOAT1-10	8	327 / 327
NCL-12	9	359 / 137
GHRL-3	12	305 / 271
GRIN1-13	17	274 / 177
CPD-9	19	315 / 108
PCK1-9	20	307 / 307
ENO1-8	21	295 / 147
GH1-3	27	373 / 311
Sf3A2	28	227 / 171
LCAT-3	Unk	323 / 133

[1] Location: chromosomal location based on chicken genome (Hillier et al., 2004)

Table A2.3. Locus-specific substitution models with associated parameters (Gamma = G; Invariable sites = I) & the molecular clock (strict (S) vs. Bayesian uncorrelated log-normal relaxed clock (R)) used during species tree reconstruction. Stars indicate identity with the models used with Tree 1 sample set.

Locus Name	TREE 1	TREE 2	TREE 3	TREE 4	TREE 5	TREE 6	TREE 7	TREE 8	TREE 9	TREE 10
CHD1Z-b	HKY & S	*	*	*	*	*	*	*	*	*
LDH1-4	HKY & S	*	*	*	*	*	*	*	*	*
ODC1-7	HKY + G & S	*	*	*	*	*	*	*	*	*
FGB-7	HKY + I & S	*	*	*	*	*	*	*	*	*
SAA-2	HKY + G & S	HKY + G + I & S	HKY + G & S	HKY + G + I & S	HKY + G & S	HKY + G + I & S	HKY + G & S	HKY + G	HKY + G	HKY + G + I
ANXA11-2	HKY + G & R	*	*	*	*	*	*	*	*	*
MSTN-2	HKY+G & R	*	*	*	*	*	*	*	*	*
SOAT1-10	HKY & S	*	*	*	*	*	*	*	*	*
NCL-12	HKY + G + I & S	*	*	*	*	*	*	*	*	*
GHRL-3	HKY & R	*	*	*	*	*	*	*	*	*
GRIN1-13	HKY+G & S	*	*	*	*	*	*	*	*	*
CPD-9	HKY + G + I & R	*	*	*	*	*	*	*	*	*
PCK1-9	HKY & S	*	*	*	*	*	*	*	*	*
ENO1-8	HKY + I & S	*	*	*	*	*	*	*	*	*
GH1-3	HKY & S	*	*	*	*	*	*	*	*	*
Sf3A2	HKY + G & S	*	*	*	*	*	*	*	*	*
LCAT-3	GTR + I & S	HKY + I & S	GTR + G + I & S	HKY + I & S	GTR + I & S	HKY + I & S	HKY + G & S	GTR + G + I & S	GTR + G + I & S	HKY + G + I & S

Table A2.4. Pair-wise Φ_{ST} estimates averaged across 17 nuclear (below diagonal) and the mtDNA control region (above diagonal).

	American Black Duck	Mallard	Mexican Duck	Mottled Duck (WGC)	Mottled Duck (FL)
American Black Duck	–	0.023*	0.14	0.069	0.31
Mallard	0.011	–	0.11	0.10	0.34
Mexican Duck	0.017	0.020	–	0.087	0.40
Mottled Duck (WGC)	0.031	0.024	0.042	–	0.34
Mottled Duck (FL)	0.059	0.055	0.064	0.042	–

* = not statistically significant ($p > 0.05$)

Table A3.1. Sample information on "population," sex (M = male; F = female), age (A = Adult; I = Immature), location, and collection date.

ID	Species	Population	Sex	Age	Country	State	City/landmark	Longitude	Latitude	Date Collected
UAMX1739	*Anas platyrhynchos*	Mallard	M	Unk	US	AK	Unk	52.9025	-172.909	Unk
KGM1412	*Anas platyrhynchos*	Mallard	M	Unk	Canada	Unk	Unk	49.9	-113.1	Unk
KGM1414	*Anas platyrhynchos*	Mallard	F	Unk	Canada	Unk	Unk	49.2	-113.3	Unk
KGM1429	*Anas platyrhynchos*	Mallard	M	Unk	Canada	Unk	Unk	49.2	-122.2	Unk
CAMall12	*Anas platyrhynchos*	Mallard	M	Unk	US	California	Colusa	39.3299	-121.914	2004
CAMall05	*Anas platyrhynchos*	Mallard	M	Unk	US	California	Colusa	39.3299	-121.914	2004
CAMall01	*Anas platyrhynchos*	Mallard	M	Unk	US	California	Colusa	39.3299	-121.914	2004
CAMall17	*Anas platyrhynchos*	Mallard	F	Unk	US	California	Colusa	39.3299	-121.914	2004
PL701	*Anas platyrhynchos*	Mallard	M	A	US	NC	Orange	35.9206	-79.0839	10/8/2009
PL824	*Anas platyrhynchos*	Mallard	F	I	US	NY	Yates	42.6839	-76.9572	12/5/2009
PL832	*Anas platyrhynchos*	Mallard	M	I	US	NY	Jefferson	44.3358	-75.9147	12/1/2009
PL844	*Anas platyrhynchos*	Mallard	F	A	US	NY	Dutchess	41.8528	-73.9222	11/26/2009
PL852	*Anas platyrhynchos*	Mallard	F	A	US	VT	Orleans	44.9442	-72.2044	10/10/2009
PL861	*Anas platyrhynchos*	Mallard	F	I	US	CT	Middlesex	41.3517	-72.4161	10/16/2009
PL923	*Anas platyrhynchos*	Mallard	F	A	US	ME	Cumberland	43.9194	-70.4667	11/9/2009
PL944	*Anas platyrhynchos*	Mallard	F	A	US	NJ	Middlesex	40.345	-74.4806	10/13/2009
PL962	*Anas platyrhynchos*	Mallard	F	I	US	ME	Franklin	44.9664	-70.7737	10/6/2009
JAT1000	*Anas [p] diazi*	US	M	Unk	US	Texas	Jeff Davis	30.5377	-103.801	5/25/2011
PL508	*Anas [p] diazi*	US	F	A	US	New Mexico	Valencia	34.72	-106.8	12/5/2009
PL513	*Anas [p] diazi*	US	M	A	US	New Mexico	Valencia	34.72	-106.8	11/12/2009
PL532	*Anas [p] diazi*	US	M	A	US	New Mexico	Dona Ana	32.3122	-106.778	11/13/2009
PL538	*Anas [p] diazi*	US	F	I	US	New Mexico	Socorro	34.02	-106.93	1/9/2010
PL680	*Anas [p] diazi*	US	F	I	US	New Mexico	Socorro	34.02	-106.93	10/30/2009
KGM927	*Anas [p] diazi*	US	M	Unk	US	New Mexico	Dona Ana	32.3122	-106.778	2003
KGM933	*Anas [p] diazi*	US	F	Unk	US	Texas	El Paso	31.7903	-106.423	2003

KGM946	*Anas [p] diazi*	US	M	Unk	US	New Mexico	Sierra Co.	33.2333	-107.317	2003
KGM954	*Anas [p] diazi*	US	M	Unk	US	New Mexico	Dona Ana	32.3122	-106.778	2003
KGM965	*Anas [p] diazi*	US	M	Unk	US	New Mexico	Luna County	32.2611	-107.756	2003
KGM968	*Anas [p] diazi*	US	F	Unk	US	Texas	El Paso	31.7903	-106.423	2003
KGM969	*Anas [p] diazi*	US	M	Unk	US	New Mexico	Sierra Co.	33.2333	-107.317	2003
PL1000	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.3494	-109.988	2/4/2012
PL1001	*Anas [p] diazi*	Sonora	F	I	Mexico	Sonora	Ciudad Obregon	27.3494	-109.988	2/4/2012
PL1002	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.3494	-109.988	2/4/2012
PL1003	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.3494	-109.988	2/4/2012
PL1004	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.3494	-109.988	2/4/2012
PL1005	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.3494	-109.988	2/4/2012
PL1006	*Anas [p] diazi*	Sonora	F	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1007	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1008	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1009	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1010	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1011	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1012	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1013	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1014	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1015	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1016	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1017	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1018	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1019	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1020	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1021	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1022	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1023	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012

PL1024	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1025	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1026	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1027	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1028	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1029	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1030	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1031	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1032	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1033	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1034	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1035	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1036	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1037	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1038	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	27.4459	-110.15	2/4/2012
PL1039	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1040	*Anas [p] diazi*	Sonora	F	I	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1041	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1042	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1043	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1044	*Anas [p] diazi*	Sonora	F	A	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1045	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1046	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1047	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1048	*Anas [p] diazi*	Sonora	M	I	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL1049	*Anas [p] diazi*	Sonora	M	A	Mexico	Sonora	Ciudad Obregon	26.9979	-109.897	2/4/2012
PL2033	*Anas [p] diazi*	Durango	M	Unk	Mexico	Durango	Nuevo Ideal	24.8875	-105.073	1/12/2013
PL2034	*Anas [p] diazi*	Durango	F	Unk	Mexico	Durango	Nuevo Ideal	24.8875	-105.073	1/12/2013
PL2035	*Anas [p] diazi*	Durango	M	Unk	Mexico	Durango	Nuevo Ideal	24.8875	-105.073	1/12/2013

PL2036	*Anas [p] diazi*	Durango	F	Unk	Mexico	Durango	Nuevo Ideal	24.8875	-105.073	1/12/2013
MMMEDU3	*Anas [p] diazi*	Zacatecas	F	Unk	Mexico	Zacatacas	Unk	23.2928	-102.701	Unk
MMMEDU6	*Anas [p] diazi*	Zacatecas	M	Unk	Mexico	Zacatacas	Unk	23.2928	-102.701	Unk
MMMEDU7	*Anas [p] diazi*	Zacatecas	F	Unk	Mexico	Zacatacas	Unk	23.2928	-102.701	Unk
MMMEDU9	*Anas [p] diazi*	Zacatecas	M	Unk	Mexico	Zacatacas	Unk	23.2928	-102.701	Unk
PL2037	*Anas [p] diazi*	Guanajuato	F	Unk	Mexico	Guanajuato	Presa la morilla	21.4958	-100.659	2/9/2013
PL2038	*Anas [p] diazi*	Guanajuato	M	Unk	Mexico	Guanajuato	Presa la morilla	21.4958	-100.659	2/9/2013
PL2023	*Anas [p] diazi*	Mexico	F	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	2/19/2013
PL2024	*Anas [p] diazi*	Mexico	F	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	2/19/2013
PL2025	*Anas [p] diazi*	Mexico	F	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	2/19/2013
PL2026	*Anas [p] diazi*	Mexico	M	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	2/19/2013
PL2027	*Anas [p] diazi*	Mexico	M	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	2/19/2013
PL2028	*Anas [p] diazi*	Mexico	F	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	11/11/2013
PL2029	*Anas [p] diazi*	Mexico	F	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	11/11/2013
PL2030	*Anas [p] diazi*	Mexico	M	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	11/11/2013
PL2031	*Anas [p] diazi*	Mexico	M	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	11/11/2013
PL2032	*Anas [p] diazi*	Mexico	M	Unk	Mexico	Mexico	Logo Boximo	19.6039	-99.7033	11/11/2013
PL2001	*Anas [p] diazi*	Puebla	F	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2002	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2003	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2004	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2005	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2006	*Anas [p] diazi*	Puebla	F	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2007	*Anas [p] diazi*	Puebla	F	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2008	*Anas [p] diazi*	Puebla	F	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2009	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2010	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2011	*Anas [p] diazi*	Puebla	F	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2012	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013

PL2013	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2014	*Anas [p] diazi*	Puebla	F	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2015	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2016	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2017	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2018	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2019	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2020	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2021	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013
PL2022	*Anas [p] diazi*	Puebla	M	Unk	Mexico	Puebla	San Jose Chiapa	19.2596	-97.6659	2/18/2013

Table A3.2. Markers identified across *BayeScan* analyses as likely under positive/balancing or purifying selection – these correspond to *BayeScan* results presented in Fig. 3.5 and Appendix Fig. A3.2.

	Males Only		Males and Females	
	Mexican Duck versus Mallard	Among Mexican Duck Sampling Groups	Mexican Duck versus Mallard	Among Mexican Duck Sampling Groups
Autosomal + Z- Chromosome	Z135679*, Z214842*, Z502513*, Z803027*; Aut517447, Aut844307*, Aut961719	Aut165057, Aut175689*, Aut184913, Aut322960, Aut452129, Aut882275, Aut1031756, Aut1112617	Z135679, Z214842, Z318715, Z502513, Z803027; Aut805541, Aut844307, Aut943308, Aut961719	Aut175689, Aut368811, Aut720330, Aut844307, Aut961719
Autosomal	Aut517447, Aut844307, Aut961719	Aut1031756, Aut1112617, Aut165057, Aut175689, Aut184913, Aut322960, Aut452129, Aut882275	Aut805541, Aut844307, Aut961719	Aut100383, Aut175689, Aut368811, Aut630790, Aut720330, Aut844307, Aut1031756
Z-Chromosome	Z421108	NA	Z17127, Z417097, Z421108, Z841971,	NA

*: Markers also identified in $DAPC_S$ outlier analyses (Fig. 3.5).

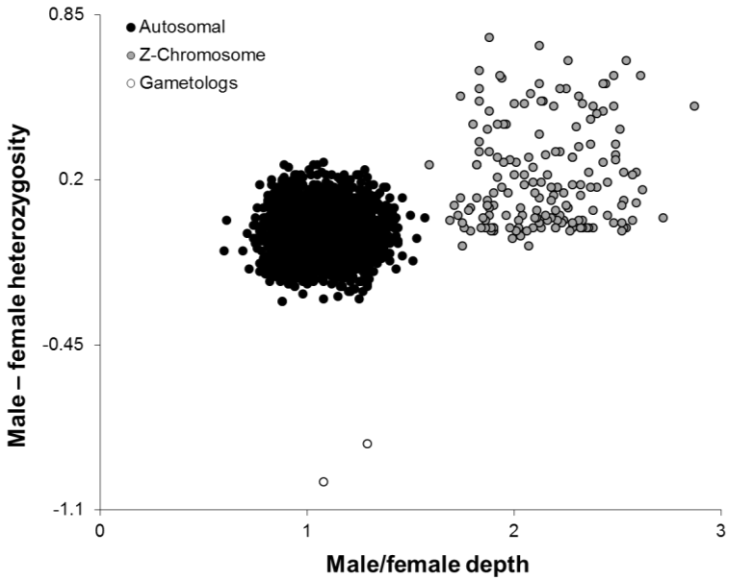


Figure A3.1. Differentiating between autosomal and Z loci based on male and female sequencing depth and heterozygosity. Colors indicate our chromosomal assignments based on this information. Two markers were found to be gametologs, which were excluded from analyses.

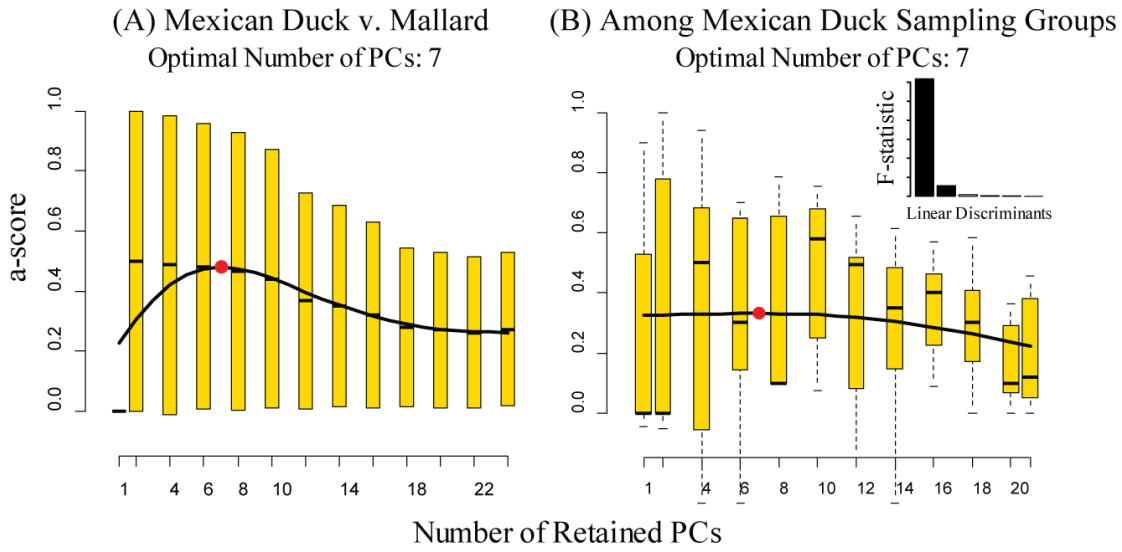


Figure A3.2. Optimum number of retained PCs (identified by red dot) and Eigenvalue for $DAPC_S$ analyses 3695 RAD sequencing markers (A) between Mexican ducks and mallards, and (B) among Mexican duck sampling groups. Only one eigenvalue was retained for analysis A due to it being a two population comparison.

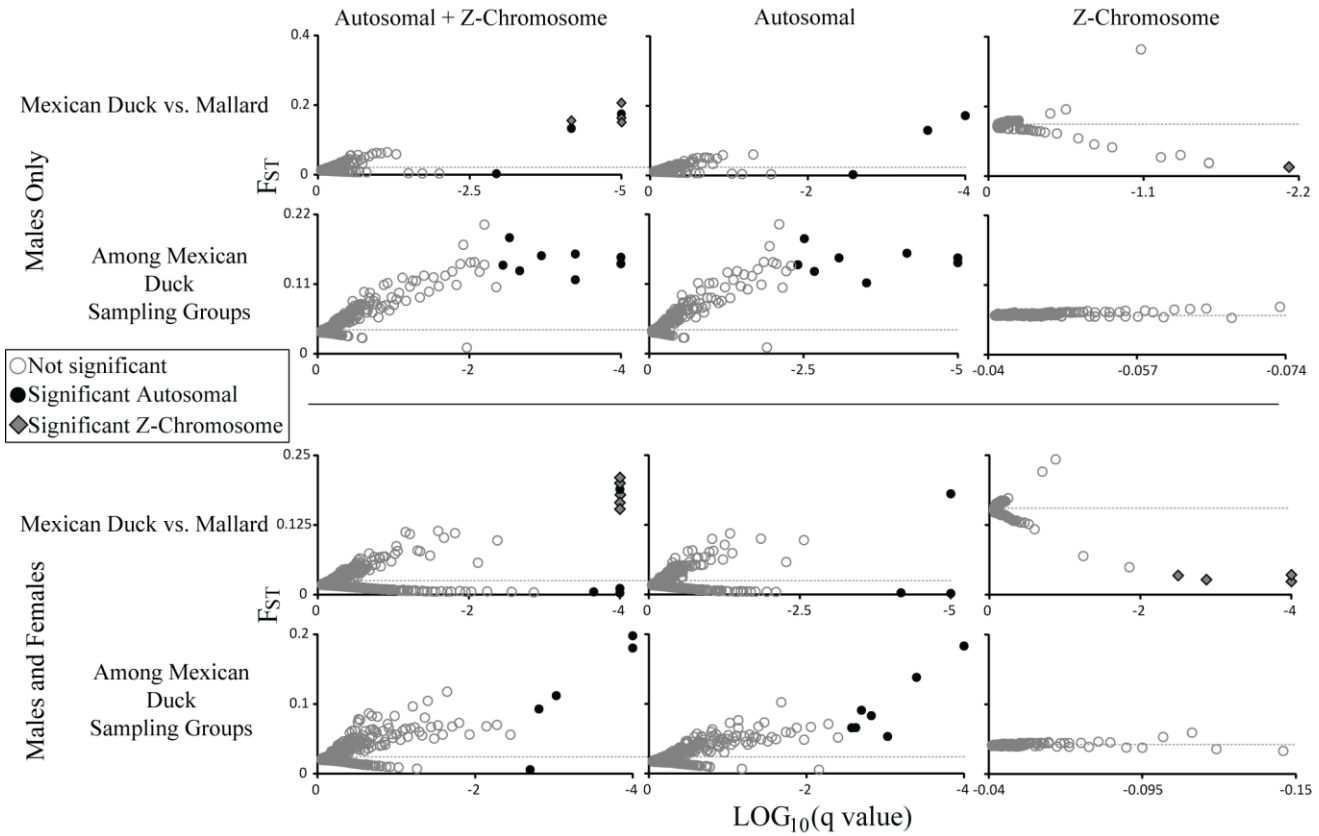


Figure A3.3. Comparison of *BayeScan* outlier results between Mexican ducks and mallards and among Mexican duck sampling groups across 3523 autosomal and 172 Z-chromosome ddRADseq markers – the dotted line denotes loci under diversifying (above) or purifying (below) selection. In an effort to test the sensitivity of *BayeScan* analyses to "missing" data of the heterogametic sex, results include analyses with and without females. Markers identified as likely under selection are listed in Appendix Table A3.2.

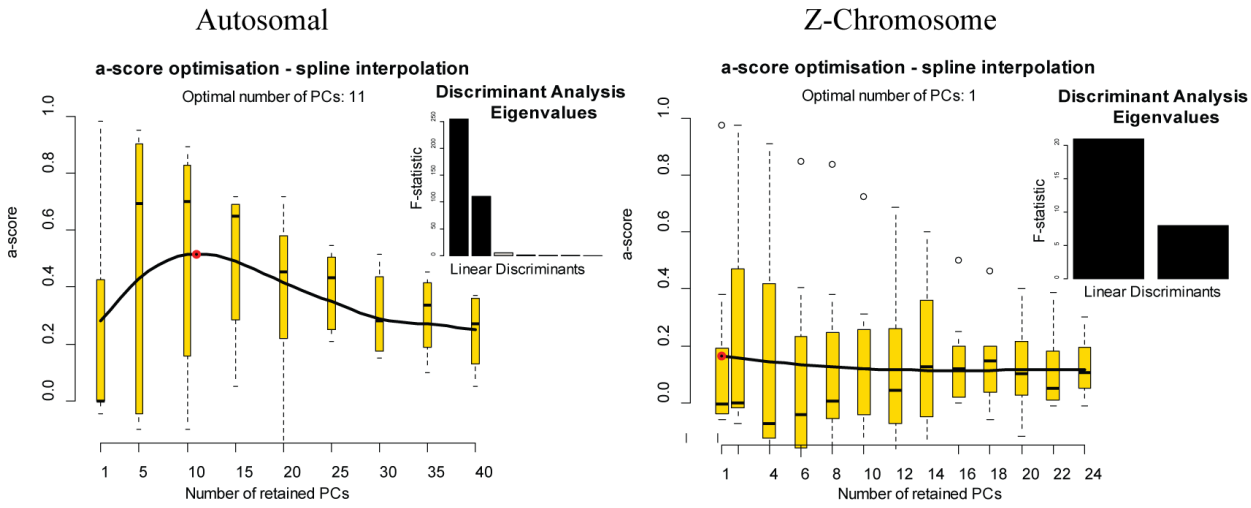


Figure A3.4. Optimum number of retained PCs (identified by red dot) and Eigenvalue for $DAPC_S$ population structure analyses with autosomal or Z-chromosome markers (Fig. 3.1).

Table A4.1. Characteristics of 19 nuclear loci. Contrasting sequence lengths are provided when data were either non-filtered or filtered for recombination. Columns 4-6 provide SNP frequencies for 16 loci used in population structure analyses (**.** indicates that the marker did not contain an informative SNP at a frequency of > 0.05).

Locus	**Location**[1]	**Non-Filtered (Filtered) Length**[2]	Laysan	Hawaiian Duck	Mallard
Chromo-helicase-DNA binding protein gene 1, intron b	**Z**	**306 (306)**	1.00:0.00	1.00:0.00	0.88:0.12
Lactate dehydrogenase 1, intron 4	**1**	**530 (470)**	1.00:0.00	0.00:1.00	0.00:1.00
S-acyl fatty acid synthase thioesterase, intron 2	**2**	**294 (294)**	**.**	**.**	**.**
Ornithine decarboxylase, intron 7	**3**	**341 (242)**	1.00:0.00	0.50:0.50	0.12:0.88
Fibrinogen beta chain, intron 7	**4**	**439 (255)**	1.00:0.00	0.80:0.20	0.56:0.44
Serum amyloid A, intron 2	**5**	**320 (145)**	1.00:0.00	0.53:0.47	0.12:0.88
Annexin A11, intron 2	**6**	**444 (384)**	1.00:0.00	0.07:0.93	0.18:0.82
Myostatin, intron 2	**7**	**280 (175)**	1.00:0.00	0.70:0.30	0.00:1.00
Soat1-prov protein, intron 10	**8**	**336 (283)**	1.00:0.00	0.90:0.10	0.96:0.04
Nucleolin, intron 12	**9**	**359 (90)**	1.00:0.00	0.73:0.27	0.22:0.78
Preproghrelin, intron 3	**12**	**320 (309)**	1.00:0.00	0.87:0.13	0.84:0.16
Glutamate receptor, ionotropic, N-methyl D aspartate I, intron 13	**17**	**312 (198)**	1.00:0.00	0.87:0.13	0.84:0.16
Sex determining region Y-box 9, intron 2	**18**	**402 (90)**	1.00:0.00	0.63:0.37	0.02:0.98
Carboxypeptidase D, intron 9	**19**	**323 (141)**	1.00:0.00	0.57:0.43	0.00:1.00

Phosphenolpyruvate carboxykinase, intron 9	**20**	**333 (272)**	1.00:0.00	1.00:0.00	0.86:0.14
Alpha enolase 1, intron 8	**21**	**306 (174)**	1.00:0.00	0.27:0.73	0.22:0.78
Alpha-B crystallin, intron 1	**24**	**334 (334)**	**.**	**.**	**.**
Growth hormone 1, intron 3	**27**	**380 (362)**	1.00:0.00	0.87:0.13	0.56:0.44
Lecithin-cholesterol acyltransferase, intron 3	**Unk**	**323 (136)**	1.00:0.00	0.92:0.08	0.40:0.60

[1] **Location**: chromosomal location based on chicken genome [50]

[2] **Length**: base-pairs

[3] Number of sites for non-filtered datasets above and those filtered for recombination is below in parentheses

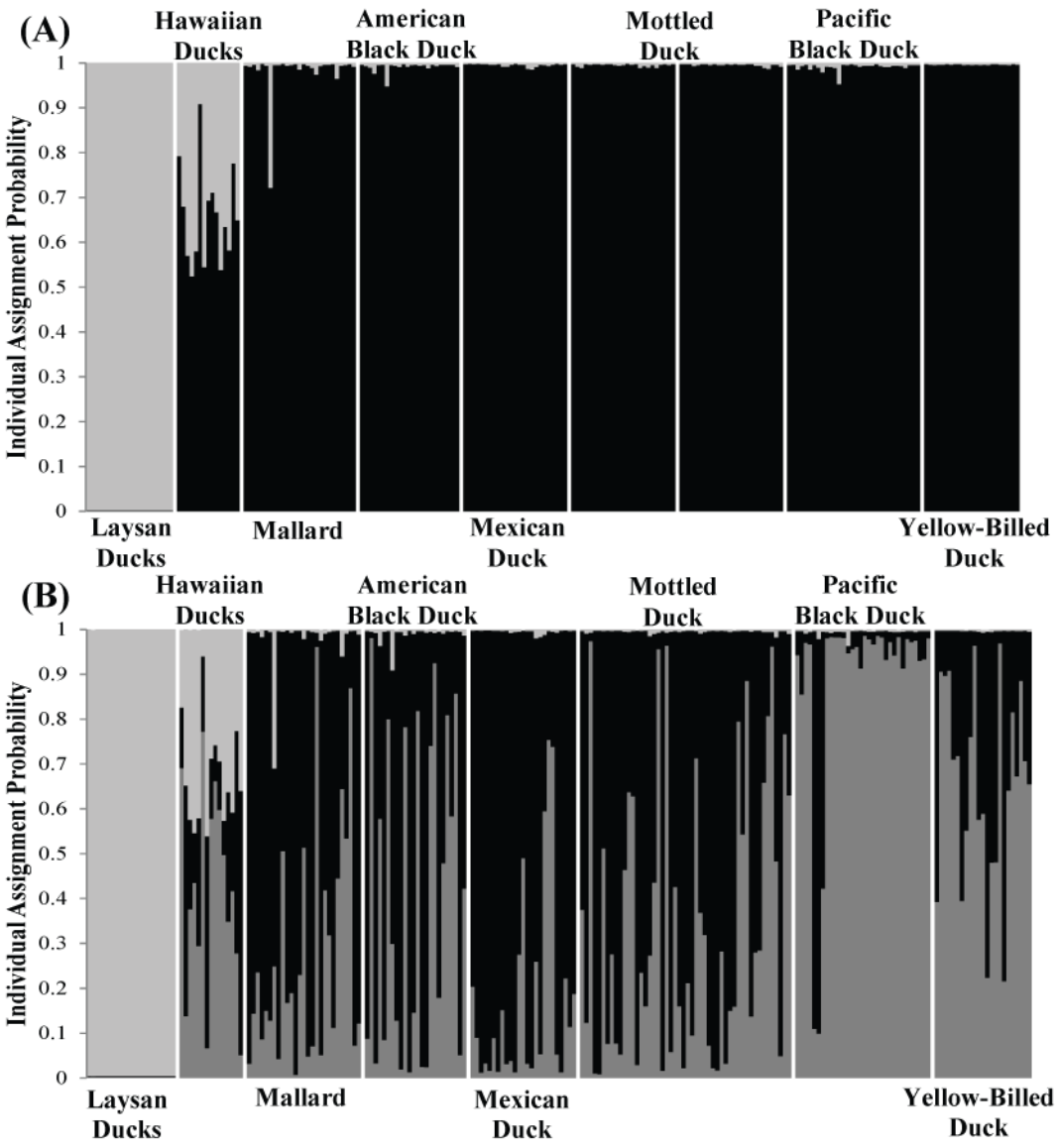


Figure A4.1A-B. *Structure* (Pritchard et al., 2000) results for 21 Laysan ducks, 15 Hawaiian ducks, 25 mallards, 24 American black ducks, 49 mottled ducks, 25 Mexican ducks, 32 Pacific black ducks, 23 yellow-billed ducks using 16 SNPs that were diagnostic between Laysan ducks and mallards (see text). (A) K = 2 and (B) K= 3.

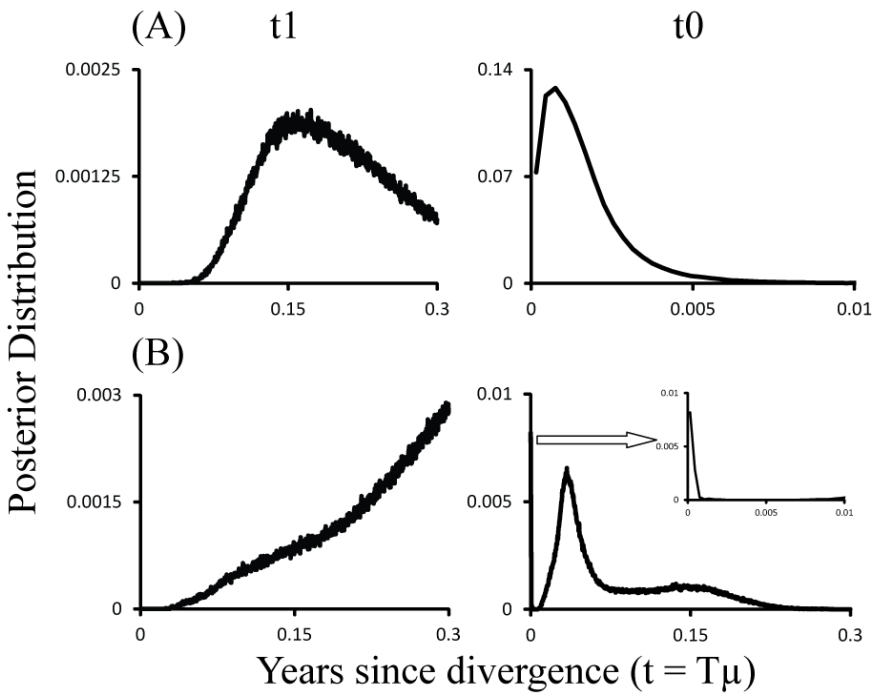


Figure A4.2A-B. IMa2 (Nielsen and Wakeley, 2001) posterior distribution of t1 (i.e., basal lineage divergence) and t0 (i.e., divergence within the sister relationship) divergence estimates under the (A) nuDNA-like or (B) mtDNA-like topology (see Fig. 4.3). Note the exponential increase for t1 and tri-modul t0 distributions – inset provides a visual of the first peak – that did not allow for resolution under the mtDNA-like toplogy.

Table A5.1. Mallard sequences of MHC I exon 2 obtained from GenBank. Number refers to the haplotype designation within parentheses given in Fig. 5.2.

Species	Number	GenBank Accession Number
Anas platyrhynchos	1	AB115242[1]
A. platyrhynchos	2	AY294417[2]
A. platyrhynchos	3	AB115243[1]
A. platyrhynchos	4	AF393511[4]
A. platyrhynchos	5	AY841883[3]
A. platyrhynchos	6	AY841882[3]
A. platyrhynchos	7	AB115244[1]
A. platyrhynchos	8	AB119993[1]
A. platyrhynchos	9	AY841881[3]
A. platyrhynchos	10	AY294418[2]
A. platyrhynchos	11	AY841884[3]
A. platyrhynchos	12	AY885227[2]
A. platyrhynchos	13	AB115245[1]
A. platyrhynchos	14	AY294419[2]
A. platyrhynchos	15	AB115240[1]
A. platyrhynchos	16	AB115246[1]
A. platyrhynchos	17	AB115241[1]
A. platyrhynchos	18	AY294416[2]
A. laysanensis	HAP 1	KF612477
A. laysanensis	HAP 2	KF612478
A. laysanensis	HAP 3	KF612479
A. laysanensis	HAP 4	KF612480
A. laysanensis	HAP 5	KF612481
A. laysanensis	HAP 6	KF612482
A. laysanensis	HAP 7	KF612483
A. laysanensis	HAP 8	*

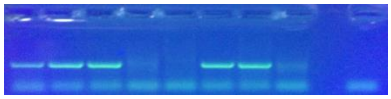
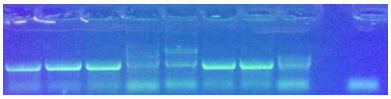
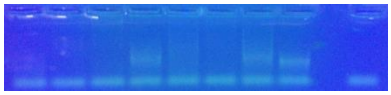
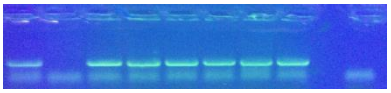
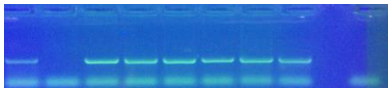
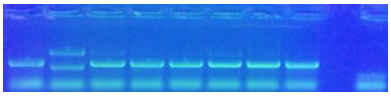
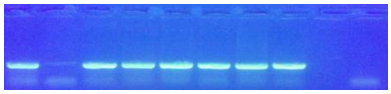
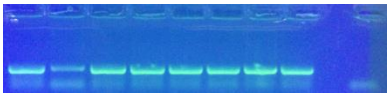
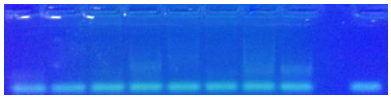
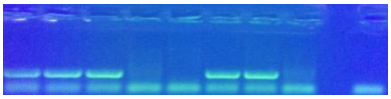
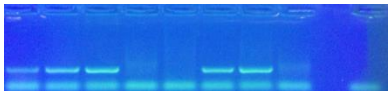
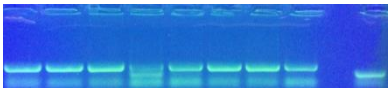
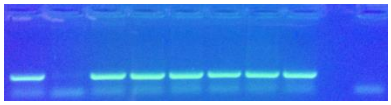
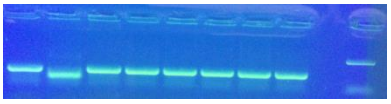
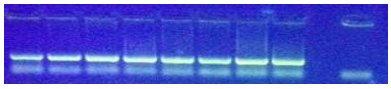
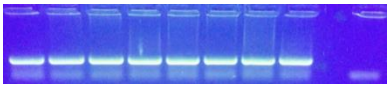
[1], (Xia *et al.* 2004); **[2],** (Mesa *et al.* 2004); **[3],** (Moon *et al.* 2005); **[4]**, (Chan *et al.* unpublished); *, Product length too small (< 200 bp) for GenBank submission

Table A5.2. Primers and primer pairs designed to amplify the antigen binding site of the Major Histocompatibility Complex I exon 2 gene in Laysan ducks with respective annealing temperatures and product sizes (base pairs) per primer pair. Optimized method used for haplotype-specific primer pairs in brackets.

Primer Pairs	Primer	Sequence (5' - 3')	Annealing Temperature (^{o}C)	Product Size (bp)	Haplotypes Amplified
E2R\E2F	E2R	GAGCCCCACTCMMTKCGCTAYTTC	65	~355	1 or 2+3+7 or
	E2F	CAGTAGCRTGSGGGMAGG			2+3+4+7
E2R2\E2F2	E2R2	TACTTCTACACCGCGGTGTC	62	238	3 or 1+5+6+7
	E2F2	TGCTCTGGTTGTAGCGCT			or 2+3
E2R2\MHC1aF (1a)	MHC1aF	TGGTTGTAGCGCTCCCTC	62	197	2/3[1] or 5+6
MHC1bF\E2F2 (1b)	MHC1bF	TAGCGCTCCCGMAGCGTC	64	238	1 or 2/3 or
					1+2 or 3+4
MHC1cF\E2F2 (1c)	MHC1cF	TGGTTGTAGCGCTCCCGC	60	204	2/3 or 8
MHC_hap1F\ MHC_hap1F	MHC_hap1R	CGGGAGTGCCACATTTTGTAA	66	191	1
	MHC_hap1F	AAGCGTCTCCAGGTGCCC	[TD 71-65]		
MHC_hap2R\MHC_hap2F	MHC_hap2R	TTCTACACCGCGGTGTCG	60	133	2
	MHC_hap2F	GCGGAAATCCTGCTCATG	[PCR]		
MHC_hap3R\MHC_hap3F	MHC_hap3R	TACTTCTACACCGCGGTGTCA	62	163	3
	MHC_hap3F	CGTCCCAGTGTTGCTGATCT	[TD 65-60]		
MHC_hap4R\MHC_hap4F	MHC_hap4R	ATGTACTATGACAGCAAGACCCAGAG	66	132	4
	MHC_hap4F	CAGGTGCATGCGGAAAGC	[TD-65-60]		
MHC_hap5R\MHC_hap5F	MHC_hap5R	GGATGGGGAGGTCTTTGTGT	66	125	5
	MHC_hap5F	CATTGCTCTGTGAGATCTTAGTCTCAT	[PCR]		
MHC_hap6R\MHC_hap6F	MHC_hap6R	ATGGGGAGGTCTTCGTGC	64	120	6
	MHC_hap6F	TGCTCTGTGAGATCTCAGTCTCC	[TD 71-65]		
MHC_hap7R\MHC_hap7F	MHC_hap7R	CGGGTGGACTGGATTGCA	62	100	7
	MHC_hap7F	TGTCCAGGTTCATGCGGT	[TD-65-60]		
MHC_hap8R\MHC_hap8F	MHC_hap8R	GGACGAATGCGATGATGA	58	79	8
	MHC_hap8F	TTCTGAAAGTTCTGGGTGTTTG	[PCR]		

[1]Sequences do not extend into polymorphisms distinguishing haplotypes 2 and 3

Table A5.3. Gel electrophoresis results during PCR optimization of haplotype specific primers with 8 Laysan duck individuals (Appendix Table A5.2). The optimum PCR conditions are noted.

Haplotype (optimum condition)	TD PCR W/ 5 CYCLES THEN 30 CYCLES	STANDARD PCR W/ 45 CYCLES
HAP 1 (TD-PCR 71-65)	 TD-PCR 71-65	 PCR 66
HAP 2 (PCR 60)	 TD-PCR 65-60	 PCR 60
HAP 3 (TD-PCR 65-60)	 TD-PCR 65-60	 PCR 62
HAP 4 (TD-65-60)	 TD-PCR 65-60	 PCR 66
HAP 5 (PCR 66)	 TD-PCR 71-65	 PCR 66
HAP 6 (TD-PCR 71-65)	 TD-PCR 71-65	 PCR 66
HAP 7 (TD-PCR 65-60)	 TD-PCR 65-60	 PCR 62
HAP 8 (PCR-58)	 TD-PCR 63-58	 PCR-58

GLOSSARY

Adaptive niche availability: A hypothesis in which the unique genetic combination within homoploid hybrid individuals results in a phenotype suited for an environment/habitat that is distinct from their parental taxa. These hybrid individuals are isolated and subsequently speciate via homoploid hybrid speciation.

Allopatric speciation: The evolution of a reproductively isolated population due to geographical/vicariant isolation.

Base-pair substitution models (e.g., gamma distribution, invariable sites): Various models that simulate different rates of changes in the mutation or changes involving replacement or substitution of a single nucleotide base with another.

Bayes factor: A test statistic that uses likelihood factors to estimate the probability between alternative models/hypotheses where K (Bayes Factor) = $Pr(D|M_1) / Pr(D|M_2)$ [D = data; | = "given"; M = model].

Bayesian Markov chain Monte Carlo (MCMC) algorithms: A class of algorithms for sampling from probability distributions that uses prior probabilities and likelihood functions to compute posterior probabilities.

Bottlenecking: A significant decrease in population size that results in the reduction of genetic variation within the population.

Clinal variation: Traits that occur as a gradient (frequency differences) across a population and is geographically associated.

Coalescence: The merging of two genetic lineages in a common ancestor. Specifically, it is tracing the genealogy of all alleles within a population to a common ancestor (i.e., most recent common ancestor).

Discriminant analysis of principle components: A multivariate method designed to identify and describe clusters of genetically related individuals from genetic data.

Divergence with gene flow: Populations that are diverging (or speciating) even though interbreeding continues to move genes between them.

Epistatic interactions: A phenotype that is specific to the particular combination and interaction of two or more genes.

Evolutionary mechanisms: Divergence leading to speciation depends on the interactions of various mechanisms (e.g., gene flow, selection, and genetic drift), all of which differentially influence genomes and the subsequent outcome(s).

Extinction by introgressive hybridization: The introduction of genes from one population or species into another through hybridization (or gene flow) that results in the

loss of genetically identifiable individuals, and thus the extinction of a "pure" population or species.

Gene tree: A representation of the evolutionary history of a single gene.

Genetic drift: The change in allelic frequency of a gene within a population due to random or stochastic processes.

Genetically cryptic: Taxa that are phenotypically diagnosable but genetically indistinguishable.

Genomic heterogeneity: The variation of the influence by evolutionary mechanisms across a single genome.

Genomic Mosaic: A genome that is composed of alleles derived from two different parental taxa.

Genomic Scan: The comparison of genome-wide patterns of diversity within and between populations using thousands of genetic markers.

Haplotype network: A representation of the relationships of alleles/haplotypes of a single gene based on the number of base-pair differences across samples.

Homoploid hybrid speciation: Speciation of a population derived by an ancestral hybridization event between two homoploid (organisms that have the same number of chromosomes) organisms.

Hybrid trait speciation: A hypothesis in which the unique genetic combination within a homoploid hybrid individual results in a phenotype (e.g., change in mate preference, song, mating time) that instantaneously limits gene flow with their parental taxa, and results in homoploid hybrid speciation.

Hybrid zone: A geographic area or contact zone where two populations or species produce "hybrid" individuals.

IM & IMa2 (Isolation-with-migration): Programs that use Markov chain Monte Carlo algorithms to simulate gene genealogies in order to estimate population parameters that include population size, time since divergence, and rates of gene flow.

Inbreeding depression: The reduction in the fitness of a population due to excessive breeding between related individuals.

Incipient taxa: Groups of individuals that have recently diverged and are at the early stages of speciation.

Incomplete lineage/stochastic sorting: A situation in which the ancestry across genes varies within single taxa, and in particular when some alleles share a more common ancestry with the alleles of another species than within the same species.

Introgressive hybridization: The movement of genes from one taxa into another due to hybridization (or gene flow).

Islands of Divergence: Large sections within a genome that show significantly higher divergence as compared to the remaining genome, and thus are likely under selection and important in the speciation process.

Isolation-by-distance: A consequence of limited dispersal across space resulting in pairs of populations which are genetically closer to one another than to populations farther away. Such a phenomenon can be explained by the stochastic change in allelic frequencies across space via genetic drift.

Linkage disequilibrium: The statistical association of the alleles at two loci within the gametes of a population.

Molecular clock rate (e.g., strict clock, Bayesian uncorrelated log-normal relaxed): The rate in which a gene or sequence changes. For example, a strict molecular clock would suggest that the rate of change across a DNA sequence changes at a constant rate.

Outbreeding depression: Offspring from crosses between individuals from different populations have a lower fitness then progeny from crosses between individuals from the same population.

Outlier loci: A gene/locus showing divergence that is statistically different from overall genomic levels. These loci are typically associated with regions that are important in divergence and speciation.

Parapatric speciation: The evolution of a reproductively isolated population that is spatially adjacent to another, closely related taxa, and which may not have any spatial barriers to gene flow.

Polytomy: A section of a phylogeny in which the evolutionary relationships cannot be fully resolved and is represented by a node with >2 descending branches.

Population Structure: Nonrandom geographic clustering of alleles.

Posterior sets of trees: A posterior distribution of tree topologies from a single species tree analysis.

Post-zygotic isolation: Effects of barriers that act after fertilization in which negative epistatic interactions between the two genomes confers isolation, including hybrid sterility, hybrid zone, F2 inviability, decreased fecundity.

Restriction site associated DNA (RAD) markers: A type of next-generation sequencing method that generates a pseudorandom set of markers across a genome. Specifically, DNA makers that are flanked by specific restriction sites are excised using the appropriate restriction enzyme(s) and subsequently sequenced.

Reversal of speciation: The loss of a unique species due to excessive hybridization/gene flow (also see extinction by introgressive hybridization).

Speciation: The evolutionary process by which new biological organisms arise.

Speciation Continuum: Variance in the strength of reproductive isolation across different groups of individuals.

Speciation genes: Genes that are likely under strong selection, which limits the effects of gene flow between two taxa, and are thus important for divergence and the speciation process.

Speciation genomics: Using next-generation technology to study speciation and the processes underlying divergence.

Species tree (e.g., *Beast): Bayesian methods that incorporate genealogical differences across markers to reconstruct overall evolutionary relationships. In general, species tree

reconstructions incorporate the genealogical variance that can exist across a set of markers.

Sympatric speciation: The evolution of a reproductively isolated population that is spatially overlapping of another, closely related taxa with no spatial barriers to gene flow.

Touch Down(TD)-PCR: A method of polymerase chain reaction in which the annealing temperature progressively changes from higher (specific) to lower (less specific) temperatures. This ensures that the sequence of interest is initially amplified, and continues to be in subsequent steps. The exponential increase of product produced by PCR ensures that the initial sequence product will outcompete nonspecific sequences that the primers may bind at lower temperatures.

Φ_{ST}: A measure of population differentiation computed as the difference in nucleotide diversity (π; or pair-wise differences across a nuclear sequence) between two randomly chosen individuals from two different populations then from the same population [$\Phi_{ST} = \pi_{between} - \pi within / \pi_{between}$]. Φ_{ST} of 0 indicates no differences between sampled populations, where a value of 1 indicates two completely divergent populations (one allele per population).