PATTERNS OF GENETIC VARIATION AMONG VARIOUS BERINGIAN BIRDS

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Elizabeth Marella Humphries, B.S., B.A.

Fairbanks, Alaska

December 2008

Abstract

Every animal has two different sets of genes, the mitochondrial genome and the nuclear genome. They are inherited independently of each other and can have dissimilar evolutionary histories. For this thesis, I examined the correlation of the mitochondrial (using direct sequencing of mitochondrial DNA) and nuclear (using a fragment-based technique, AFLPs) evolutionary histories in Beringian birds. The first study looked at relationships among a within-Beringia radiation of birds, the *Aethia* auklets. I was able to infer a resolved phylogeny using mitochondrial DNA, but found a polytomy using nuclear DNA. The lack of correlation between the genomes is either the result of incomplete lineage sorting or reflects the fact that AFLPs may not be an appropriate technique to survey nuclear diversity at this level of divergence. In the second study, I examined the correlation between taxonomic status, mitochondrial divergence, and nuclear divergence among nine lineages of trans-Beringian birds. I found no relationship between any of the three variables, suggesting that trans-Beringian birds have complex and discordant genetic and phenotypic histories.

Table of Contents

List of Figures

List of Tables

Page

List of Other Materials

List of Appendices

Page

Acknowledgments

This research has been supported by the National Science Foundation (DEB-9981915), the National Geographic Society, the Alaska Maritime National Wildlife Refuge, and a University of Alaska Fairbanks EPSCoR fellowship. This research was done under the supervision of K. Winker. For auklet specimens, I thank Hector Douglas and the collectors for the University of Alaska Museum, including K. Bell, V. Byrd, R. Dickerman, D. Johnson, E. Miller, D. Rocque, L. Slater, A. Will, and S. Wright. For trans-Beringian specimens, I thank Sievert Rohwer and Sharon Birks of the Burke Museum for a tissue loan of Russian birds and the collectors for the University of Alaska Museum. D. Sikes, C. L. Pruett, T. Roberts, J. L. Peters, M. J. Lelevier, H. C. Lanier, C. Topp, and N. Takebayashi provided valuable feedback and help.

GENERAL INTRODUCTION

Every animal has two genomes. The mitochondrial genome is found inside the mitochondrion, an organelle that is passed from mother to offspring. In birds this genome is approximately 18,000 base pairs long, divided into 37 genes (Mindell et al. 1999). The mitochondrial genome (mtDNA) is circular, and all the genes are inherited as a single unit; recombination between different genes is not expected to occur (Avise 2004). The nuclear genome is much larger and contains most of the genes necessary for development and life. The nuclear genes are divided among many chromosomes (two copies of each unique chromosome), and each chromosome is inherited separately from the other (Snustad and Simmons 2000). Additionally, each parent contributes half of the chromosomes inherited by the offspring.

Because they are inherited differently, the mitochondrial and nuclear genomes have different evolutionary histories. The nuclear genome has four times the effective population size of mtDNA (because mtDNA is inherited only from the mother); the time required until fixation of mtDNA mutations in a population is one-fourth the time required until fixation of a nuclear mutation (Zink and Barrowclough 2008). Additionally, birds can exhibit a bias in dispersal between the sexes. If a species has male-biased dispersal, gene flow for mtDNA is very different than gene flow for the nuclear genome. Gene flow between the two genomes will be more similar in a species with female-biased dispersal.

In the following chapters, I sampled the mitochondrial genome via direct sequencing, and the nuclear genome was sampled using the Amplified Fragment Length Polymorphism (AFLP) method. The AFLP technique is a fragment-based method that allows us to sample a large number of anonymous, presumably unlinked loci from across the entire nuclear genome (Vos et al. 1995). It is particularly useful when examining very shallow or recent divergences in which direct sequencing of the nuclear genome has been uninformative (African cichlids, Allender et al. 2003; snow and McKay's buntings, Maley 2006; *Laupala* crickets, Mendelson et al. 2004). The following two studies cover a range of divergence depths, making the AFLP method the most appropriate for examining genetic variation across the entire genome.

Both of these studies examine genetic differentiation in Beringian birds. Beringia is a region between eastern Alaska and eastern Russia. During the Pleistocene (2 million years ago to 10,000 years ago), Alaska and Russia were periodically connected by the Bering Land Bridge, a land mass that stretched over 1,000 miles north to south (Hopkins 1959, Hopkins et al. 1965). The first chapter examines the phylogenetics of the auklets (family Alcidae, genera *Ptychoramphus* and *Aethia*), a group of seabirds that live within the part of Beringia currently covered by the Bering Sea. In the second chapter, genetic variation is examined in relation to the effects of the Pleistocene glaciations among three different lineages (orders Anseriformes, Charadriiformes, and Passeriformes) of birds that are found on either side of the Bering Strait.

CHAPTER 1:

Working through polytomies: Auklets revisited[1]

1.1 ABSTRACT

Polytomies, or phylogenetic "bushes," are the result of a series of internodes occurring in a short period of evolutionary time (which can result in data that do not contain enough information), or data that have too much homoplasy to resolve a bifurcating branching pattern. In this study we used the *Aethia* auklet polytomy to explore the effectiveness of different methods for resolving polytomies: mitochondrial DNA gene choice, number of individuals per species sampled, model of molecular evolution, and AFLP loci. We recovered a fully-resolved phylogeny using NADH dehydrogenase subunit 2 (ND2) sequence data under two different Bayesian models. We were able to corroborate this tree under one model with an expanded mtDNA dataset. Effectiveness of additional intraspecific sampling varied with node, and fully 20% of the subsampled datasets failed to return a congruent phylogeny when we sampled only one or two individuals per species. We did not recover a resolved phylogeny using AFLP data. Conflict in the AFLP dataset showed that nearly all possible relationships were supported at low levels of confidence, suggesting that either AFLPs are not useful at the genetic depth of the *Aethia* auklet radiation (7 – 9% divergent in the mtDNA ND2 gene), perhaps resulting in too much homoplasy, or that the *Aethia* auklets have experienced incomplete lineage sorting.

[1]Humphries, E.M. and K. Winker. 2008. Working through polytomies:Auklets revisited. Prepared for submission to Molecular Phylogenetics and Evolution.

1.2 Introduction

The goal of most phylogenetic studies is to infer a completely resolved, bifurcating phylogeny. The presence of polytomies, or phylogenetic "bushes," in the final phylogeny is often seen as a failure (Rokas and Carroll, 2006). These polytomies may be artifacts of the inference process— "soft"—or may be biologically real—"hard" (Maddison, 1989). Theoretically, soft polytomies can be resolved if one corrects the problems in the inference process, although this becomes more difficult if the amount of time since the radiation (length of the terminal branches) is long compared to the amount of time between speciation events (length of the internode; Rokas and Carroll, 2006; Whitfield and Lockhart, 2008). Hard polytomies, by definition, cannot be resolved into bifurcating relationships.

Different methods have been proposed for determining whether a polytomy is soft or hard, and many of these are tree-based. For example, Slowinski (2001) advocated a method that involved topology testing, which Poe and Chubb (2004) expanded by suggesting that researchers should determine whether multiple independent gene trees are congruent, as opposed to simply testing whether a resolved topology better fits the data than a polytomy. McCracken and Sorenson (2005) also used a multiple-locus approach to resolving polytomies and included both parametric and nonparametric bootstrapping. All of these methods used sequence data. DNA sequencing is a useful phylogenetic tool, as evidenced by its popularity in systematic studies; and it has many advantages, including

repeatability and well-characterized models of evolution (Avise, 2004). However, analyses using sequence data can be hampered by choice of gene and by limited genomic representation. Practically, sequence-based studies can only examine a very small fraction of the total genome and, thus, of the total phylogenetic information that the genome contains. Fragment-based methods, such as amplified fragment length polymorphisms (AFLPs; Vos et al., 1995), may be able to compensate for this particular weakness of sequence-based methods. There are several benefits to the AFLP method: 1) it involves a sampling of the entire genome, resulting in a broad assessment of genomic phylogenetic signal; 2) it samples many (presumably) unlinked loci; and 3) many of the loci sampled are polymorphic (often more than 50%, depending on how closely taxa are related), which may result in greater phylogenetic signal (e.g., Ribiero et al., 2002; Schneider et al., 2002). Koopman (2005) suggested that AFLPs have more phylogenetic signal at lower divergence levels than ITS sequence data in fungi, although this relationship was not clearly seen in plants or bacteria. Spooner et al. (2005) suggested that AFLPs were more useful than sequence data in resolving relationships among wild tomatoes. AFLPs have become widely used in plants but are relatively uncommon in vertebrate studies (Bensch and Åkesson, 2005).

Increasing the number of individuals per species sampled may also increase phylogenetic resolution. Many studies have examined whether choice of different ingroup taxa sampled is more important for inferring resolved phylogenies than increasing the number of loci sampled; most suggest that a larger number of loci results in greater phylogenetic resolution (Rosenberg and Kumar, 2003; Rokas et al., 2005; but

see also Hedtke et al., 2006). Jackman et al. (1999) suggested analyzing subsampled datasets with fewer taxa to resolve polytomies. However, examining only one individual per species potentially overlooks phylogenetically informative intraspecific variation and could effectively mean ignoring sampling error and incorporating sampling bias (based on which individual is chosen) into the dataset (Ives et al., 2007).

One basic assumption in resolving soft polytomies is that repeating analyses with larger and larger datasets will result in a greater resolution until the polytomy disappears (Maddison, 1989; DeSalle et al., 1994). However, because of systematic bias (such as long branches rich with homoplasy), phylogenetic analysis of large datasets can produce fully resolved and well supported but erroneous topologies (Rokas and Carroll, 2006). Thus, increasing the size of a dataset may not be as informative as an approach that can examine phylogenetic signal in the data independent of a single topology (Charleston and Page, 1999). Network methods, such as spectral analysis (Hendy and Penny, 1993), NeighborNet (Bryant and Moulton, 2004) and consensus networks (Holland and Moulton, 2003; Holland et al., 2005; Holland et al., 2006) allow visualization of all possible phylogenetic relationships in the data, not just the majority or plurality relationship (Kennedy et al., 2005). In particular, consensus networks show all bifurcations that appear in a given input of trees (Kennedy et al., 2005), permitting one to examine more than just the consensus tree in likelihood or Bayesian phylogenetic analyses.

In this study, we examined the utility of increased sampling (genomic and individual) and improved evolutionary models and analytical methods to re-examine the

polytomy described by Walsh et al. (1999) in *Aethia* auklets (Aves: Alcidae). Auklets are small seabirds that breed on rocky shorelines and spend the rest of the year at sea (Byrd and Williams, 1993; Jones, 1993a, b; Manuwal and Thoresen, 1993; Jones et al., 2001). Auklets belong to three genera: 1. *Cerorhinca*, a monotypic genus consisting of *C. monocerata* (rhinoceros auklet, a widely acknowledged distant relative to the group of interest here), 2. *Ptychoramphus*, another monotypic genus (*P. aleuticus*, Cassin's auklet), and 3. *Aethia*, a genus with four members: *A. pusilla* (least auklet), *A. psittacula* (parakeet auklet), *A. cristatella* (crested auklet), and *A. pygmaea* (whiskered auklet). The earliest molecular and morphological phylogenies of the Alcidae placed *Ptychoramphus* as sister to a polytomy formed by all the members of *Aethia* (Strauch, 1985; Moum et al., 1994; Friesen et al., 1996). However, despite repeated study, researchers have been unable to resolve the relationships within *Aethia* (Moum et al., 1994; Friesen et al., 1996; Walsh et al., 1999; Pereira and Baker, 2008). Walsh et al. (1999) speculated that the *Aethia* auklets diverged ~2.6 million years ago and that their speciation may have been the result of late Pliocene/early Pleistocene glaciations. They used a power analysis approach and determined that if the *Aethia* auklets had speciated over 100,000 years (the range of the shortest interglacial period in the late Pliocene/early Pleistocene), enough DNA (~3,000 base pairs) had been sequenced in their study to resolve the *Aethia* polytomy if it was soft. Additionally, they discovered that the number of base pairs required to resolve a polytomy increases exponentially (Walsh et al., 1999). Thus, they estimated that ~22,000 base pairs of equivalently informative mtDNA would be required to resolve the polytomy if the range of radiation decreased by a power of 10 (i.e., to about

10,000 years). Because the avian mtDNA genome only has ~17,000 base pairs (Mindell et al., 1999), this polytomy was considered essentially unresolvable by using mtDNA sequence data (assuming the radiation occurred over 10,000 years or less). Pereira and Baker (2008) combined sequences from five mitochondrial and one nuclear gene; they suggested that the unresolved *Aethia* relationships are the result of incomplete lineage sorting. Thus, the *Aethia* polytomy would remain no matter how much data were added.

The presence of an apparently hard polytomy in the auklets, coupled with a radiation presumed to have occurred over less than 100,000 years, make *Aethia* an ideal genus for testing various methods for resolving polytomies. We employed three strategies to crack the *Aethia* polytomy: increasing genomic data, sampling multiple individuals within each species, and running more sophisticated analyses. First, we tested Walsh et al.'s (1999) conclusion that the auklet polytomy is "hard" using a different mtDNA gene (NADH dehydrogenase 2, ND2), amplified fragment length polymorphisms (AFLPs), and a Bayesian mixture model of nucleotide substitution. We also combined our data with the sequence data from Pereira and Baker (2008). We then examined how phylogenetic signal and support in the data changed as we increased the number of individuals per species. Finally, we used a coalescent method to estimate the duration of the *Aethia* radiation.

1.3 Methods

Mitochondrial DNA (mtDNA) sequencing

DNA was extracted from muscle tissue from 9 *Ptychoramphus*, 10 *A. psittacula*, 10 *A. pusilla*, 10 *A. pygmaea*, and 10 *A. cristatella* individuals (Table 1) using a DNeasy Tissue

Kit (Qiagen, Valencia, CA, USA). We amplified the mtDNA NADH dehydrogenase subunit 2 (ND2) gene using primers L5215 (Hackett, 1996) and H6313 (Sorenson et al., 1999). Amplification was done following standard PCR protocols and using an annealing temperature of 48˚C. Cycle-sequencing was done using ABI Big-Dye Terminator mix and an annealing temperature of 50˚C. Products were sequenced on an ABI 3130 automated sequencer (Applied Biosystems, Inc., USA). We aligned sequences without gaps using Sequencher (Genecodes Corporation, Inc., USA).

Amplified fragment length polymorphism (AFLP) analyses

We chose 5 individuals from each species for AFLP analysis (Table 1). We followed a modified ABI (Applied Biosystems Inc., USA) plant-mapping (large genome) protocol to generate loci (Parchman et al., 2006). Restriction of extracted DNA using enzymes MseI and EcoRI was done concurrently with ligation of MseI and EcoRI adaptors. Preselective amplification was done using ABI's preselective primers. We used 11 primer pair combinations for selective amplification (Appendix 1). Genotyping was done on an ABI 3100 automated sequencer (Applied Biosystems Inc., USA). We scored the chromatograms using GeneMapper ver. 3.7 (Applied Biosystems, Inc., USA). Only loci that could be determined unambiguously as present or absent in each individual were scored.

Phylogenetic analyses

mtDNA – We used MrModelTest (ver. 2.2) by Nylander (2004) in combination with PAUP* (Swofford, 2002) to choose the most appropriate model (GTR + I + Γ) for phylogenetic analysis. In addition, we also ran analyses under the GTR + site-specific

(SS) model, in which the data were partitioned by codon site. Under this model there is no gamma distribution or invariant sites assumption; instead, rates of mutation vary among partitions. (The GTR+SS model was 200 log-likelihood units better than the GTR + I + Γ model.) Phylogenetic analyses were done using MrBayes (ver. 3.1.1, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Analyses were run using 4 chains and flat Dirichlet priors. Both analyses were run for 8 million generations with a burnin of 3 million generations and a tree sampling interval of 10,000, for a total of 501 trees sampled. The length of the burnin period for this run and all subsequent MrBayes runs was chosen based on when the differences between the –ln of the likehoods for the cold chain and one randomly-chosen heated chain were consistently less than 0.1

We also used the sequence data of Pereira and Baker (2008) for the five auklet species (Appendix 2). We downloaded a total of 43 sequences, a combination of five mitochondrial genes, one nuclear gene (recombination activating gene 1 (RAG-1)) and 18 individuals: 5 individuals each for cytochrome b, NADH dehydrogenase subunit 2 (ND2), 16S and 12S, and 18 individuals for cytochrome *c* oxidase subunit I (COI). We only included sequence data with an associated vouchered specimen. We concatenated all mitochondrial data from the same individual; there were only five individuals (one *Ptychoramphus* and one of each of the 4 *Aethia* species) with sequences for all five mitochondrial genes. We then aligned the remaining downloaded mitochondrial sequences (all for COI) with the COI genes in the concatenated dataset and coded the remaining genes as missing data. We repeated this for our ND2 dataset. In the end, we had an incomplete data matrix containing 68 individuals and 4744 bp of mtDNA. We

also created a second mitochondrial matrix by combining our GenBank dataset with only three individuals per species from our ND2 dataset. (We chose to include three individuals because on average that is the number of additional sequences we downloaded from GenBank for COI.) We chose to concatenate the data because the genes were all from the mitochondrial genome and thus share a similar evolutionary history; we also chose to include individuals for which we did not have all genes sequenced because even data sets with missing data can infer accurate phylogenies (Wiens et al., 2005) We analyzed the RAG-1 dataset separately because genes from the nuclear genome do not necessarily share the same history as the mitochondrial genome. We partitioned both mitochondrial matrices by gene and used MrModelTest (ver. 2.2; Nylander, 2004) in conjunction with PAUP* (Swofford, 2002) to determine the most appropriate model for each gene. We also used this method to determine the best model for the RAG-1 dataset. In all cases, the GTR+I+Γ model was found to be the most appropriate model. We used the same priors as before, but ran analyses for the combined mitochondrial datasets for 12 million generations, sampling trees every 10,000 generations. We used a burnin of 5 million generations and sampled a total of 701 trees. We then repeated analysis of both mitochondrial matrices using the GTR+SS model and the same priors and settings. The RAG-1 dataset was analyzed using both models and run for 8 million generations, with a burnin of 3 million generations and trees sampled every 10,000 generations.

To determine the effect of increased sample size on polytomy resolution, we randomly subsampled the ND2 dataset without replacement and created phylogenies

using 1 - 9 individuals per species. We replicated this five times. These phylogenies were created using the GTR + SS models with the same settings as the full ND2 dataset and were reconstructed to answer two questions: did the analysis of the subsampled dataset result in the same phylogeny as the full dataset; and what was the posterior probability of the internal nodes?

AFLPs – One of the limitations of the AFLP method for phylogenetic studies is the lack of a satisfactory model of AFLP evolution. Most researchers use parsimony for AFLP analysis. Because parsimony has been shown to fail in some cases (Felsenstein, 2004), or to falsely resolve a real polytomy (Slowinski, 2001), using this method for phylogenetic analysis could result in incorrect inference. We inferred phylogenies using both parsimony and Bayesian methods. Parsimony analysis was performed in PAUP* (Swofford, 2002). The Bayesian analysis was done in MrBayes (ver. 3.1.1, Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) using the restriction site (binary) model. These analyses were run using 4 chains for 8 million generations with a burnin of 3 million generations and a tree sampling interval of 10,000. To examine the conflict in the dataset, we used the bipartition probabilities file from MrBayes in SplitsTree4 (ver 4.6; Huson and Bryant, 2006). We only examined bipartitions that were present in greater than 5% of trees.

All trees from mtDNA sequence data and AFLP data were rooted with *Ptychoramphus*, a monotypic genus known to be basal to the *Aethia* polytomy (Strauch, 1985; Friesen et al., 1996).

Estimation of the length of the Aethia *radiation*

We used the program BEAST (Drummond and Rambaut, 2007) to estimate the time to most recent common ancestor (TMRCA) at each node and the duration of the *Aethia* radiation. TMRCA estimates were made in standard time units (STU), which are the estimated number of substitutions per 1041 base pairs (the length of the ND2 gene) along a branch. We downloaded ND2 sequences for *Fratercula arctica*, *Cerorhinca monocerata*, *Cepphus columba*, *Uria lomvia*, *Alca torda*, *Alle alle*, *Brachyramphus brevirostris*, and *Synthliboramphus antiquus* from GenBank (accession numbers DQ385092, EF373230, EF373229, EF373273, EF373220, EF373221, EF373227, and EF373269, respectively), and subsampled our ND2 dataset, randomly choosing three different individuals per *Aethia* species. TMRCA was estimated using an HKY + I + Γ model under a strict molecular clock with a mean substitution rate of 1.0. (We were unable to get reasonable posterior distributions for all the parameters when using a more complex model.) We used a Yule process tree prior and estimated our starting tree using UPGMA. We ran BEAST for 14 million generations, sampling every 10,000 generations and using a burnin of 10%.

1.4 Results

Aethia *phylogeny*

1041 base pairs of mtDNA were generated for 49 individuals (Table 1). We generated a single consensus tree (using the 50% majority rule) for each Bayesian run using only our ND2 dataset. The genus *Aethia* was recovered as monophyletic with a posterior probability of 1.0 under both the GTR + I + Γ and the GTR + SS models. Additionally, each species was monophyletic, as expected, and trees from both models of molecular

evolution showed the same topology: *A. cristatella* and *A. pygmaea* were recovered as sister taxa, *A. psittacula* was sister to the *A. cristatella–A. pygmaea* clade, and *A. pusilla* was recovered as basal (Fig. 1).

We also generated 898 AFLP loci for a total of 25 individuals using 11 different primer pair combinations (average 81.6 loci per primer pair with a standard deviation of 15.4 loci). Loci were 75-300 bp in size. This dataset proved to have little strength to resolve relationships in this group. Parsimony failed to resolve the polytomy, and Bayesian analysis produced a consensus tree that recovered *A. cristatella* and *A. pusilla* as sister with a posterior probability of 0.55, with *A. pygmaea* as sister to the *A. cristatella–A. pusilla* clade (Fig. 2). *A. psittacula* was recovered as basal. All species were recovered as monophyletic with a posterior probability of 1.0. When we examined the conflict in the dataset (i.e., other possible bipartitions recovered during Bayesian analysis), we discovered that most possible sister relationships among the auklets were present in approximately the same percentage (25%) of trees from the posterior distribution (Fig. 3). The *A. pygmaea–A. cristatella* pairing and the *A. psittacula–A. pusilla* pairing deviated from this, appearing in less than 5% of the trees, and 55% of trees showed the *A. cristatella–A. pusilla* relationship.

Using our complete expanded data matrix (i.e., the matrix that included all of our ND2 sequences), we recovered two different topologies. Under the GTR + I + Γ model, Node A had a posterior probability of 0.86, while Node B was not recovered (thus, the topology showed a three-species polytomy); these values increased to 1.0 (Node A) and 0.97 (Node B) under the GTR + SS model. When we included only 3 of our ND2

sequences per species (to control for the possibility that our ND2 dataset was driving the phylogenetic inference), Node A had a posterior probability of 0.99 (under the GTR + I + Γ model) and Node B had a posterior probability of 0.63 (under the same model). When the truncated data matrix was analyzed using the GTR + SS model, the posterior probabilities were 1.0 (Node A) and 0.98 (Node B). This pattern holds when only one ND2 sequence was used.

The tree generated from the RAG-1 dataset recovered two other topologies (trees not shown); under the under GTR + I + Γ model, *A. psittacula*, *A. cristatella*, and *A. pygmaea* formed a three species polytomy with a posterior probability of 0.68. However, under GTR + SS model, *A. psittacula* and *A. cristatella* were recovered as sister, with a posterior probability of 0.82. This clade was sister to *A. pygmaea*; the posterior probability of this node was 0.78.

Contribution of increased taxon sampling

Under the better molecular model (GTR + SS) on our ND2 dataset, the posterior probability of Node A was unaffected by sample size (Fig. 4). For Node B (the more weakly supported node), trees with a sample size of 3 or more showed more robust support than trees with a sample size of only 1 or 2. Additionally, when we sampled only one or two individuals per species, one replicate of the five failed to return a topology congruent to the topology returned by the full ND2 dataset (not shown).

Estimation of the length of the Aethia *radiation*

The most recent common ancestor (MRCA) of the auklets (genera *Ptychoramphus* and *Aethia*) was estimated to have occurred 64.79 standard time units

(STU) ago, with a 95% highest posterior density (HPD) of 52.61–77.69 STU (Fig. 5). The genus *Aethia* shared a MRCA 51.10 (40.18–61.48) STU ago. Within *Aethia*, the TMRCA for Node A was 45.52 (35.44–56.25) STU. Node B on the MrBayes ND2 tree (Fig. 1) did not exist in the BEAST tree; instead, BEAST placed *A. cristatella* and *A. psittacula* as sister species, with *A. pygmaea* as basal. The TMRCA for *A. cristatella* and *A. psittacula* was 38.87 (31.40–47.60) STU. The *Aethia* radiation lasted for a total of 12.23 (8.79–13.88) STU.

1.5 Discussion

Aethia *phylogeny*

We were able to infer a fully bifurcating, resolved phylogeny using our ND2 dataset (Fig. 1).Under some models, the GenBank data matrices corroborated this topology, although the full matrix analyzed using a GTR + I + Γ model returned a unresolved topology. Morphological evidence matches partially with the topology recovered from our ND2 dataset. *A. cristatella* and *A. pygmaea* were recovered as sisters in our mtDNA tree. Morphologically, these two species were assumed to be sister because they share several traits (facial plumes, a citrus odor, and vocalizations) that are not found in any of the other true auklet species (Jones, 1993a). However, morphology also suggested that *A. pusilla* is more closely related to the presumed *A. cristatella–A. pygmaea* clade (Jones, 1993b). Our mitochondrial data suggest instead that *A. psittacula* is closer to the *A. cristatella–A. pygmaea* clade. If the true species relationship matches the one inferred by the mitochondrial genome, then the morphological evidence suggesting a closer relationship between *A. pusilla* and *A. pygmaea* (size, mostly, but also

bill color) is the result of one of three events: a loss of phenotypic characters on the *A. psittacula* branch; a loss of phenotypic characters in the ancestor of the *A. psittacula*-*A. cristatella*-*A. pygmaea* clade, followed by a reversal in the ancestor for the *A. cristatella*-*A. pygmaea* clade; or convergence between *A. pusilla* and the *A. cristatella*-*A. pygmaea* clade.

A phylogeny built using just mtDNA is not necessarily representative of the true species phylogeny (Ballard and Whitlock, 2004). The relationships supported by a single locus may be the actual species relationships, but they can also be the result of stochastic lineage sorting or a selective sweep (Ballard and Whitlock, 2004). Additionally, mtDNA is maternally inherited and thus can cross species boundaries when females of one species interbreed with males of another. If widespread hybridization occurred soon after diversification, the "wrong species" mtDNA could have easily become fixed; in this case, the evolutionary history of mtDNA would not be the evolutionary history of the species (Ballard and Whitlock 2004). Ideally, we could test whether the relationships inferred by mtDNA are correct using an independent dataset, such a nuclear DNA. However, both phylogenies inferred using two nuclear data sets (the RAG-1 dataset from Pereira and Baker (2008), downloaded from GenBank, and our AFLP dataset) had two completely different topologies, neither of which matched the mtDNA topology. In the case of Pereira and Baker's (2008) RAG-1 data, *A. psittacula* and *A. cristatella* were found to be sister taxa; analysis of our AFLP data set returned an unresolved tree (Fig. 2). The lack of concordance among the various gene trees is not necessarily unexpected. The mitochondrial and nuclear genomes are inherited independently and have different rates

of sorting (Moore, 1995). Lineage sorting can cause different loci to infer different relationships, especially when the lengths of internal branches are short, as we see in all resolved *Aethia* phylogenies. This is the explanation favored by Pereira and Baker (2008). However, gene trees are not species trees, and the probability of a 5-taxa gene tree being discordant with the species tree is high, especially when the internode lengths are short. It is possible that none of the three topologies supported by the three genetic datasets is the species tree.

Why would we get a resolved tree with mtDNA and nuclear sequence data (under one model) but still see a polytomy with an AFLP sampling of the whole genome? One possible explanation is that we have not generated enough AFLP loci for phylogenetic analysis. However, 70% of our 898 loci were polymorphic. In other studies (*Pinus pinaster*, Ribeiro et al., 2002; *Solanum* L. section *Lycopersicon*, Spooner et al., 2005), ~1000 or fewer loci with a smaller percentage of polymorphisms have provided enough signal to resolve shallow phylogenetic relationships. Another possible explanation for the polytomous topology recovered in our Bayesian consensus tree of the AFLP data is the stochastic process of lineage sorting. Based on our mtDNA coalescent analyses, the speciation events occurred so closely together that our estimated 95% HPDs overlapped (Fig. 5). The stochastic process of DNA lineage sorting during short internode time intervals is notorious for causing different genes to have different evolutionary histories (Avise, 2004). With a method such as AFLPs, which amplifies stretches of DNA over many different genes, conflicting phylogenetic signals across the genome could result in a consensus topology that either showed a polytomy or very low nodal support.

Another possible explanation is that the *Aethia* auklets are too divergent for the AFLP technique to be useful in resolving the phylogeny. As species become more divergent, lack of homology of generated fragments becomes a serious issue (see Lerceteau and Szmidt, 1999; Robinson and Harris, 1999). Bremer (1991) suggested that treating non-homologous loci as homologous artificially increases homoplasy and thus obscures any phylogenetic signal in RFLP data. AFLPs may be subject to similar problems. Lerceteau and Szmidt (1999) discovered that AFLPs failed when attempting to reconstruct the deeper relationships within the genus *Pinus*, although the method was successful in recovering shallower relationships. Our AFLP data showed that the Bayesian consensus tree (without collapsed branches) recovered a relationship not found in any mtDNA or morphological tree (*A. cristatella* and *A. pusilla* as sister taxa). With the exception of the bipartitions containing *A. cristatella*, all other possible pairwise relationships were supported in approximately 25% of the trees of the posterior distribution; this same result would occur if we randomly assigned AFLP loci to each species (assuming we only sampled one individual per species). If the lack of resolution in our AFLP data was caused by lineage sorting issues, we suspect that the support for the various bipartitions would be less evenly distributed. However, two lines of evidence suggest that AFLP technique is not at fault: 1. all individuals were correctly assigned to their respective species, and each species formed a monophyletic clade, and 2. the lengths of the branches leading to each species clade were not the same. We suspect that the branches would be the same length if loci had truly been randomly assigned.

Contribution of increased sample size toward resolution of the polytomy

Fully 20% ($N = 5$) of phylogenetic reconstructions using ND2 and one or two individuals recovered a tree that was incongruous with the tree recovered using the full dataset (Fig. 4, Node B). Using ND2, it appears that for divergences at this depth adding individuals past three per taxon only increased the intraspecific variation. Maddison and Knowles (2006) found this same relationship with a simulated dataset. No matter how many loci they used to build their trees, the accuracy of species tree inference plateaued when three or more individuals were sampled. Felsenstein (2006) also noticed this relationship when calculating phylogeny-based likelihood parameters. Our results provide additional, empirical evidence.

Estimation of times to most recent common ancestor (TMRCA) between species

The range of the *Aethia* radiation can be calculated by subtracting the estimate of TMRCA for the most recent split (in this case, the *A. pygmaea-A.psittacula* divergence) from the estimate of TMRCA for the entire genus. We can convert STU into years if we multiply the estimates by a substitution rate. Previous estimates suggested that the auklet radiation occurred over a very short period (less than 100,000 years; Walsh et al., 1999), but our estimates suggest that it took 800,000 years at the very least (assuming the overall mtDNA genome clock of 2.7% per million years from Walsh et al. 1999). If we use a standard mtDNA clock of 2% per million years, the radiation would have taken place over 1.18 million years, and doubling the fastest mutation rate (see Ho et al., 2007) only decreases the estimated range by half. In fact, we would have to use an unlikely mutation rate of 122.3% to confine our coalescence-estimated range of this radiation to 100,000 years.

However, the reality of the molecular clock has been questioned, and in birds the standard clock has been calibrated in only a few lineages, none of which are alcids (Lovette, 2004). Pereira and Baker (2008) used fossil calibration points to estimate the ages of the most recent common ancestor of the *Aethia* auklets, a method that does not require a clock estimate. Their estimated length of the radiation (calculated using the means of their posterior distributions) is still around 1 million years.

A tree made using TMRCA estimates produced a topology different from that obtained using MrBayes. However, because our HPD intervals overlapped, any topology created using only TMRCA estimates is suspect. Although hard or near-hard polytomies have been shown to mislead Bayesian inference (Lewis et al., 2005; Steel and Matsen, 2007), the topology we recovered using Bayesian methods was also recovered using both parsimony and maximum likelihood (not shown). We feel confident that we are not seeing a Bayesian "star paradox" (Steel and Matsen, 2007) with our ND2 data.

Conclusion

Polytomies can exist both at the gene/locus level and at the species level. In the *Aethia* auklets, the mtDNA phylogeny of Walsh et al. (1999) showed a gene-level polytomy, although the true *Aethia* species tree may be polytomous. We were able to confidently resolve the mtDNA polytomy when using a different gene from the same locus (ND2), a larger concatenated dataset from data on GenBank, better analytical techniques using improved models of molecular evolution, and by sampling more than a single individual per species. We were unsuccessful in reconstructing a bifurcating species-level phylogeny of the *Aethia* auklets using AFLPs. The technique may not be

useful for species that are 7-9% divergent in mtDNA or that have short internode distances at that depth. The polytomous AFLP consensus tree may also reflect a history of incomplete lineage sorting among the *Aethia* auklet species.

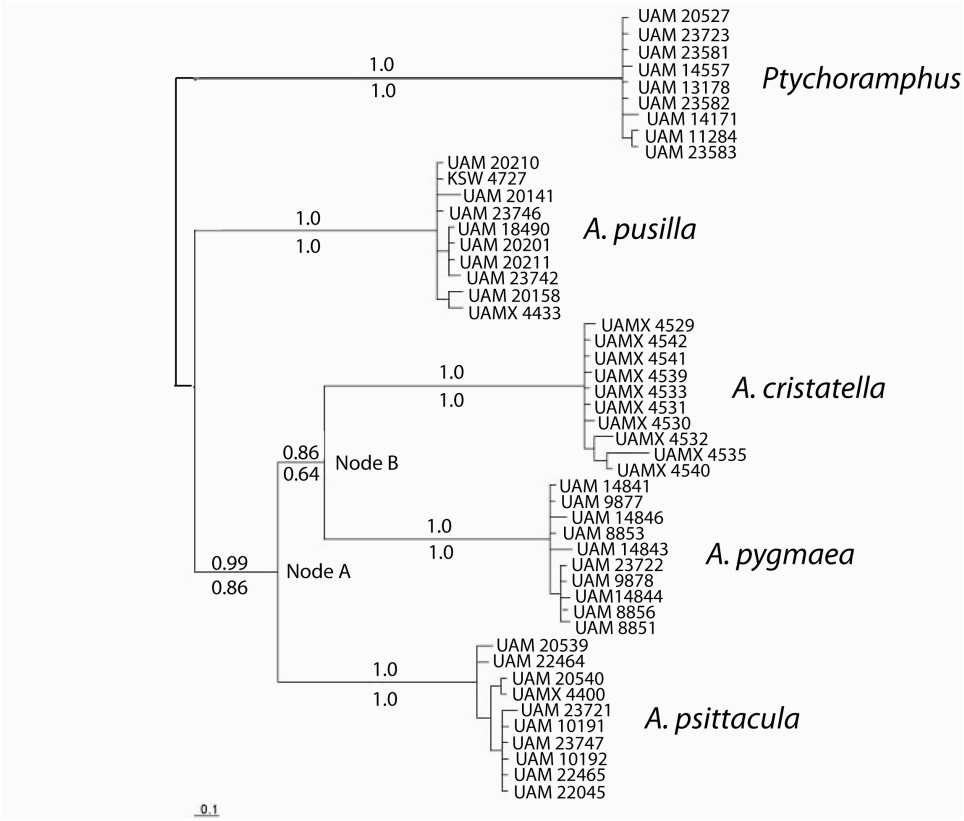


Figure 1.1. Bayesian mtDNA phylogeny (ND2 data). The same topology was recovered under both models of evolution. The number on top of the nodal branches is the posterior probability under the GTR + I + Γ model of evolution; the number below the branches is the posterior probability under the GTR + SS model. The two nodes tested for sample size effects are labeled A and B.

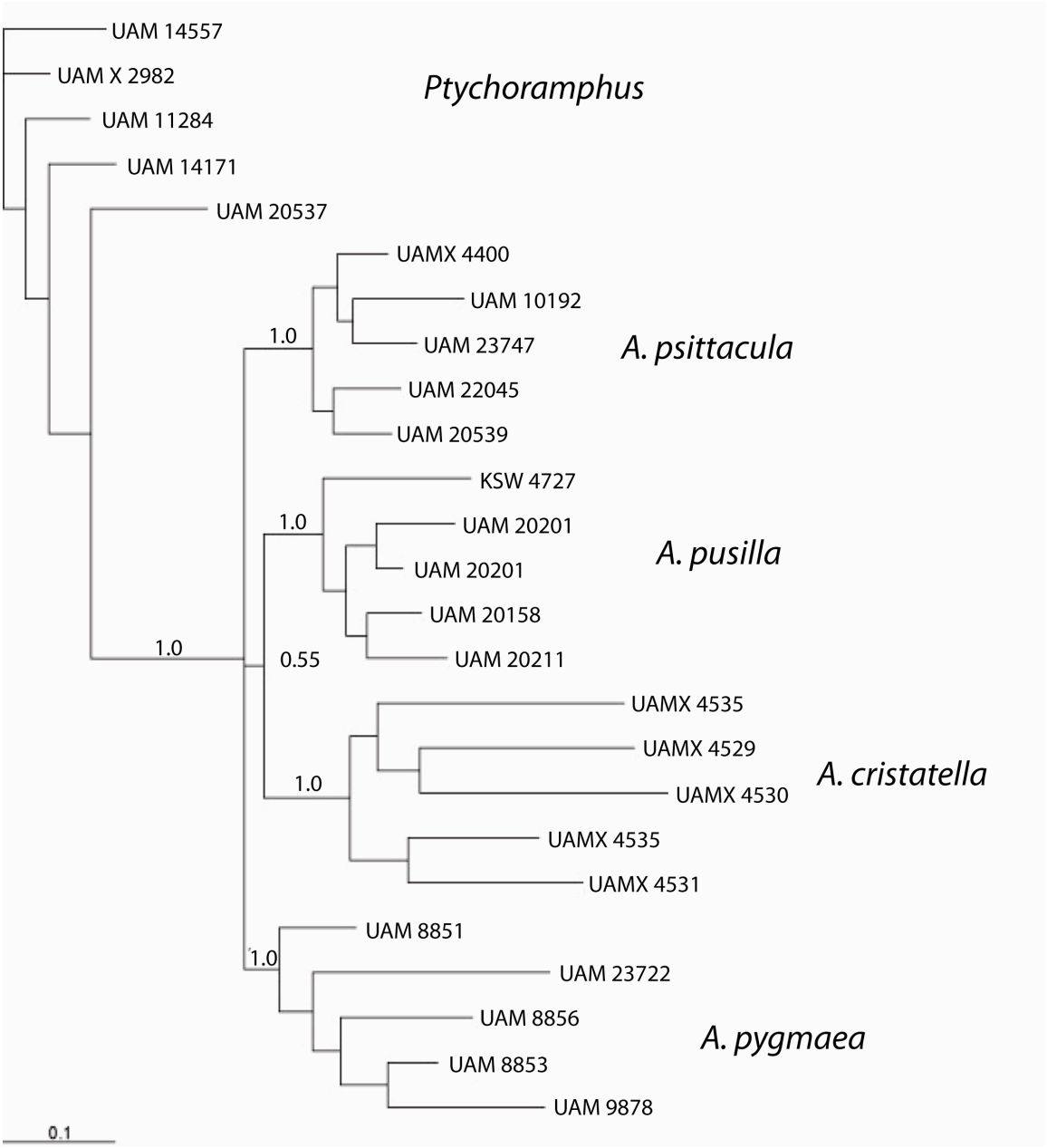


Figure 1.2. Bayesian AFLP phylogeny. Phylogeny reconstructed from 898 AFLP loci using the restriction site (binary) model in MrBayes ver. 3.1.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003).

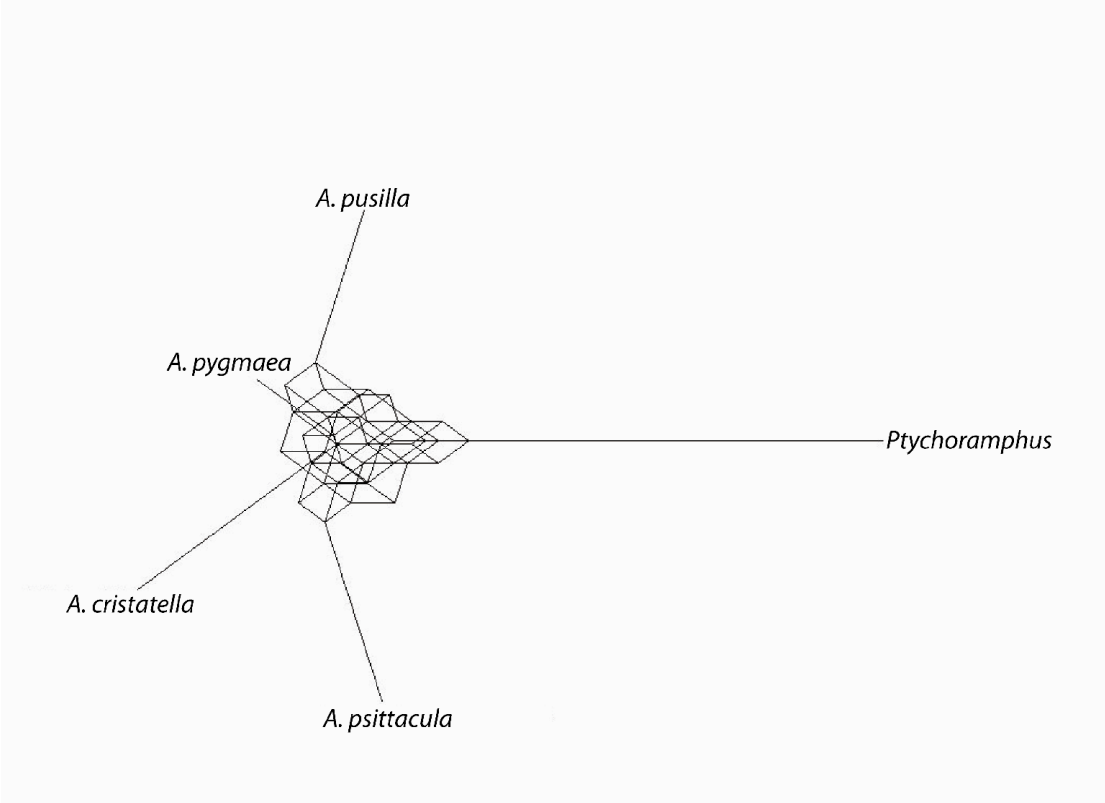


Figure 1.3. Network of possible sister relationships (AFLP data). Relationships that appear in more than 5% of trees in the posterior distribution are included. Each vertex represents another possible data bipartition. Only relationships between the species are shown.

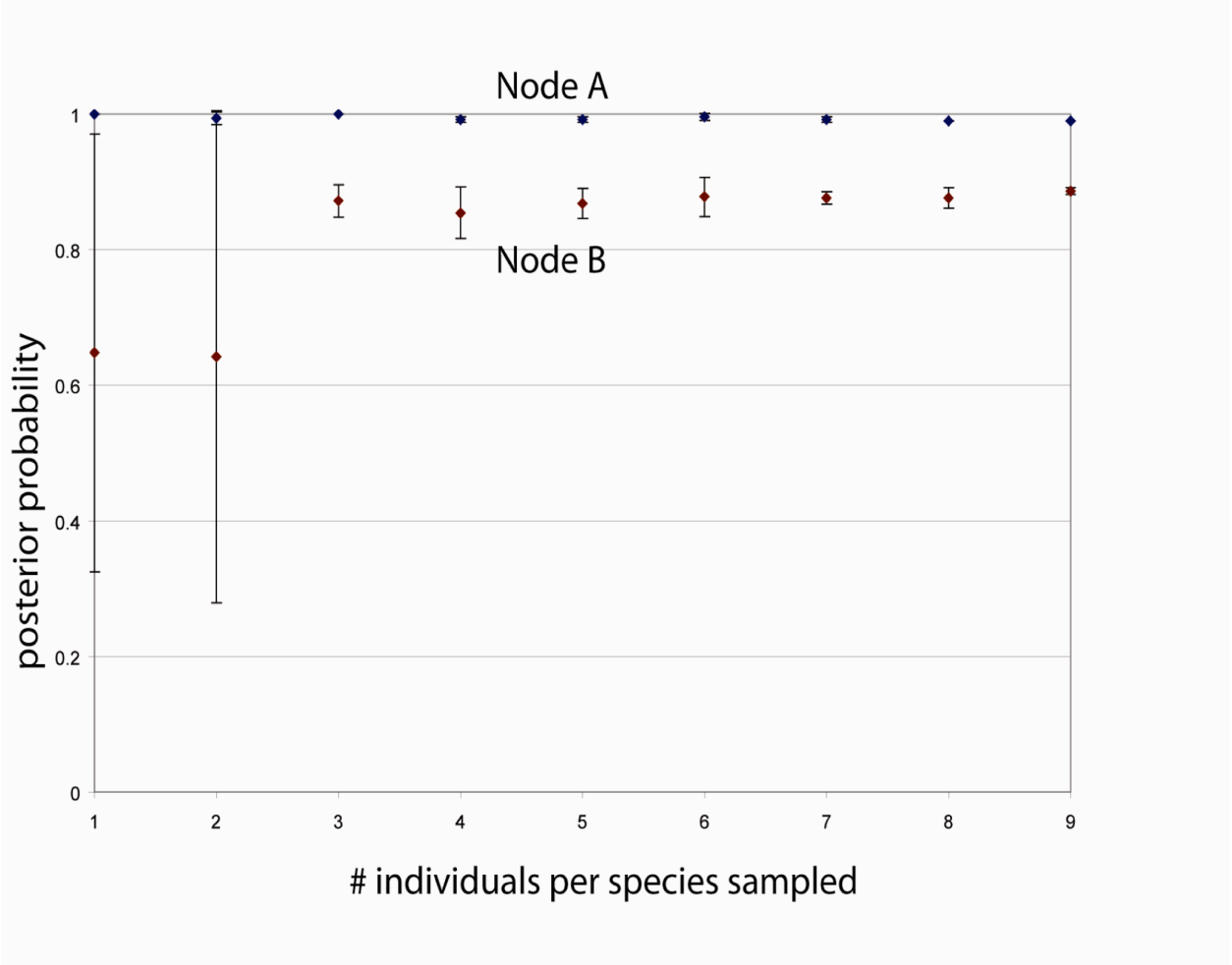


Figure 1.4. Change in posterior probabilities of nodes. Posterior probabilities were calculated under the GTR+SS model. Node A is the node that unites *A. pygmaea, A. cristatella*, and *A. psittacula*, and Node B unites *A. pygmaea* and *A. cristatella* (Fig. 1). We sampled a minimum of 1 individual per species and a maximum of 9 individuals per species.

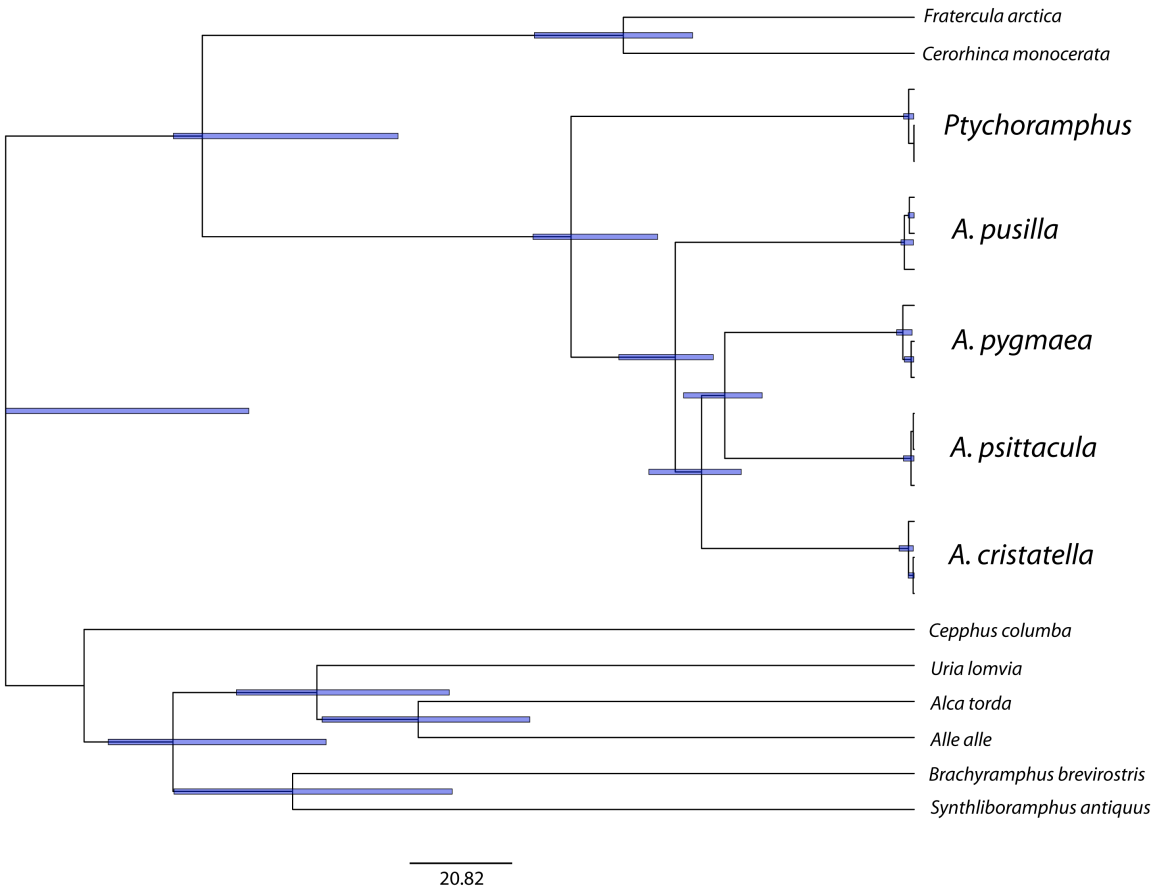


Figure 1.5. BEAST tree (ND2 data). The gray boxes show the 95% highest posterior density of estimated times to most recent common ancestors. Branches are scaled to estimated number of substitutionsper 1041 base pairs of mtDNA sequence (the length of the ND2 gene).

Table 1.1. Individuals used in this study. Collecting localities (all from Alaska), and identifiers (specimen voucher numbers) are included. All specimens (except *Aethia cristatella*) are deposited in the University of Alaska Museum. All individuals were used for sequencing analysis; starred individuals were used for AFLP analysis.

[1] Specimen only used for AFLP analysis.

Species	Voucher No.	Collection localities
Ptychoramphus aleuticus	UAM 13178, 11284*, 14171*, 14557*, 20537*, 23581, 23582, 23583, 23723, UAMX 2982*[1]	Buldir Is., Chowiet Is., Isl. of Four Mts, Little Koniuji Is., Lowrie Is. (5).
Aethia cristatella	UAMX 4529*, UAMX 4530*, UAMX 4531*, UAMX 4532*, UAMX 4533, UAMX 4535*, UAMX 4539, UAMX 4541, UAMX 4540, UAMX 4542	St. Lawrence Is. (10)
Aethia psittacula	UAM 10191, UAM 10192*, UAM 20539*, UAM 20540, UAM 22045*, UAM 22464, UAM 22465, UAM 23721, UAM 23747*, UAMX 4400*	Koniuji Is. (5), St. George Is., St. Paul Is. (4)
Aethia pusilla	UAM 18490, UAM 20141*, UAM 20158*, UAM 20201*, UAM 20210, UAM 20211*, UAM 23742, UAM 23745, UAM 23746, KSW 4727*	Birch Creek, Buldir Is., Little Diomede Is., St. George Is. (6), St. Paul Is.
Aethia pygmaea	UAM 8851*, UAM 8853*, UAM 8856*, UAM 9877, UAM 9978*, UAM 14841, UAM 14843, UAM 14844, UAM 14846, UAM 23722*	Konuiji Is., Sedanka Is. (4), Seguam Is., Umak Is. (4)

1.8 Literature Cited

Avise, J.C. 2004. Molecular markers, natural history, and evolution. 2nd ed. Sinauer Associates, Sunderland, Massachusetts.

Bensch, S. and M. Åkesson. 2005. Ten years of AFLP in ecology and evolution: why so few animals? Mol. Ecol. 14:2899-2914.

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

Bremer, B. 1991. Restriction data from chloroplast DNA for phylogenetic reconstruction: is there only one accurate way of scoring? Pl. Syst. Evol. 175:39-54.

Bryant, D. and V. Moulton. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic methods. Mol. Biol. Evol. 21:255-265.

Byrd, G.V. and J.C. Williams. 1993. Whiskered Auklet (*Aethia pygmaea*). *In* The Birds of North America, No. 76 (A. Poole and F. Gill, Eds.). Philadelphia: The Academy of Natural Sciences; Washington, D.C.: The American Ornithologists' Union.

Charleston, M.A. and R.D.M. Page. 1999. Spectral analysis – A brief introduction. Pages. 255-267 *in* Molecular systematics and plant evolution (P.M. Hollingsworth, R.M. Bateman, and R.J. Gornall, eds.). Taylor & Francis, London.

DeSalle, R., R. Absher, and G. Amato. 1994. Speciation and phylogenetic resolution. Trends Ecol. Evol. 9:297-298.

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

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Massachusetts.

Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol. Biol. Evol. 23:691-700.

Friesen, V.L., A.J. Baker, and J.F. Piatt. 1996. Phylogenetic relationships within the Alcidae (Charadriiformes: Aves) inferred from total molecular evidence. Mol. Biol. Evol. 13:359-367.

Hackett, S.J. 1996. Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). Mol. Phylogenet. Evol. 5:368-382.

Hedtke, S.M., T.M. Townsend, and D.M. Hillis. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55:522-529.

Hendy, M.D. and D. Penny. 1993. Spectral analysis of phylogenetic data. J. Classif. 10:5-24.

Ho, S.Y.W., B. Shapiro, M.J. Phillips, A. Cooper and A.J. Drummond. 2007. Evidence for time dependency of molecular rate estimates. Syst. Biol. 56:515-522.

Holland, B.R., F. Delsuc, and V. Moulton. 2005. Visualizing conflicting evolutionary hypotheses in large collections of trees: Using consensus networks to study the origins of placentals and hexapods. Syst. Biol. 54:66-76.

Holland, B.R., L.S. Jermiin, and V. Moulton. 2006. Improved consensus network techniques for genome-scale phylogeny. Mol. Biol. Evol. 23:848-855.

Holland, B.R. and V. Moulton. 2003. Consensus networks: A method for visualizing incompatibilities in collections of trees. Pages 165-176 *in* WABI 2003. Springer-Verlag, Berlin, Heidelberg, Budapest.

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

Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networds in evolutionary studies. Mol. Biol. Evol. 23:254-267.

Ives, A.R., P.E. Midford, and T. Garland, Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst. Biol. 56:252-270.

Jackman, T.R., A. Larson, K. de Queiroz, and J.B. Losos. 1999. Phylogenetic relationships and tempo of early diversification in *Anolis* lizards. Syst. Biol. 48:254-285.

Jones, I.L. 1993a. Crested Auklet (*Aethia cristatella*). *In* The Birds of North America, No. 70 (A. Poole and F. Gill, Eds.). Philadelphia: The Academy of Natural Sciences; Washington, D.C.: The American Ornithologists' Union.

Jones, I.L. 1993b. Least Auklet (*Aethia pusilla*). *In* The Birds of North America, No. 69 (A. Poole and F. Gill, Eds.). Philadelphia: The Academy of Natural Sciences; Washington, D.C.: The American Ornithologists' Union.

Jones, I.L., N.B. Konyukhov, J.C. Williams, and G.V. Byrd. 2001. Parakeet Auklet (*Aethia psittacula*). *In* The Birds of North America, No. 594 (A. Poole and F. Gill, eds.). The Birds of North America, Inc., Philadelphia, PA.

Kennedy, M., B.R. Holland, R.D. Gray, and H.G. Spencer. 2005. Untangling long branches: Identifying conflicting phylogenetic signals using spectral analysis, Neighbor-Net, and consensus networks. Syst. Biol. 54:620-633.

Koopman, W. J. M. 2005. Phylogenetic signal in AFLP data sets. Syst. Biol. 54:197-217.

Lerceteau, E. and A.E. Szmidt. 1999. Properties of AFLP markers in inheritance and genetic diversity studies of *Pinus sylvestris* L. Heredity 82:252-260.

Lewis, P.O., M.T. Holder and K.E. Holsinger. 2005. Polytomies and Bayesian phylogenetic inference. Syst. Biol.54:241-253.

Lovette, I.J. 2004. Mitochondrial dating and mixed-support for the "2% rule" in birds. Auk 121:1-6.

Maddison, W. 1989. Reconstructing character evolution on polytomous cladograms. Cladistics 5:365-377.

Maddison, W. P. and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21-30.

Manuwal, D. A. and A. C. Thoresen. 1993. Cassin's Auklet (*Ptychoramphus aleuticus*). *In* The Birds of North America, No. 50 (A. Poole and F. Gill, Eds.). Philadelphia: The Academy of Natural Sciences; Washington, D.C.: The American Ornithologists' Union.

McCracken, K.G. and M.D. Sorenson. 2005. Is homoplasy or lineage sorting the source on incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*)? Syst. Biol. 54:35-55.

Mindell, D.P., M.D. Sorenson, D.E. Dimcheff, M. Hasegawa, J.C. Ast, T. Yuri. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48:138-152.

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

Moum, T., S. Johansen, K.E. Erikstad, and J.F. Piatt. 1994. Phylogeny and evolution of the auks (subfamily Alcinae) based on mitochondrial DNA sequences. Proc. Natl. Acad. Sci. USA 91:7912-7916.

Nylander, J.A.A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Parchman, T.L., C.W. Benkman, and S.C. Britch. 2006. Patterns of genetic variation in the adaptive radiation of New World crossbills (Aves : Loxia). Molecular Ecology 15:1873-1887.

Pereira, S.L. and A.J. Baker. 2008. DNA evidence for a Paleocene origin of the Alcidae (Aves: Charadriiformes) in the Pacific and multiple dispersals across northern oceans. Mol. Phylog. Evol. 46:430-445.

Poe, S. and A.L. Chubb. 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. Evolution 58:404-415.

Ribiero, M.M., S. Mariette, G.G. Vendramin, E. Szmidt, C. Plomion, and A. Kremer. Comparison of genetic diversity estimates within and among populations of maritime pine using chloroplast simple-sequence repeat and amplified fragment length polymorphism data. 2002. Molecular Ecology 11:869-877.

Robinson, J.P. and S.A. Harris. 1999. Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective *in* Which DNA Marker for Which Purpose? Gillet, E.M. (ed.). Final Compendium of the Research Project Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees in the European Union DGXII Biotechnology FW IV Research Programme Molecular Tools for Biodiversity.

Rokas, A., D. Krüger, and S.B. Carroll. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22:1337-1344.

Rokas, A. and S.B. Carroll. 2006. Bushes in the tree of life. PLoS Biology 4:1899-1904.

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

Rosenberg, M.S. and S. Kumar. 2003. Taxon sampling, bioinformatics, and phylogenomics. Syst. Biol. 52:119-124.

Schneider, M.V., L.W. Beukeboom, G. Driessen, L. Lapchin, C. Bernstein, and J.J.M. van Alphen. 2002. Geographical distribution and genetic relatedness of sympatrical thelytokous and arrhentotokous populations of the parasitoid *Venturia canescens* (Hymenoptera). J. Evol. Biol. 15:191-200.

Slowinksi, J.B. 2001. Molecular polytomies. Mol. Phylogenet. Evol. 19:114-120.

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

Spooner, D.M., I.E. Peralta, and S. Knapp. 2005. Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section *Lycopersicon* (Mill.) Wettst.]. Taxon 54:43-61.

Strauch, J.G. Jr. 1985. The phylogeny of the Alcidae. Auk 102:520-539.

Steel, M. and F.A. Matsen. 2007. The Bayesian 'star paradox' persists for long finite sequences. Mol. Biol. Evol. 24:1075-1079.

Swofford, D.L. 2002. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, and M. Kuiper. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.

Walsh, H.E., M.G. Kidd, T. Moum, and V.L. Friesen. 1999. Polytomies and the power of phylogenetic inference. Evolution 53:932-937.

Weins, J.J., J.W. Fetzner Jr., C.L. Parkinson, and T.W. Reeder. 2005. Hylid frog phylogeny and sampling strategies for speciose clades. Syst. Biol. 54:719-748.

Whitfield, J.B. and P.J. Lockhart. 2008. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22: 258-265.

Appendix 1.1 . Selective amplification primer pairs. Only the last three base pairs of the primers are reported. Primers are from Applied Biosystems, Inc (USA).

Primer combination	MseI primer	EcoRI primer
1	CAA	ACT
2	CTT	ACA
3	CAC	AAC
4	CAT	AGC
5	CTG	AGC
6	CTG	AAG
7	CTT	ACC
8	CAG	ACG
9	CAT	ACT
10	CTG	ACC
11	CTG	ACT

Appendix 2. GenBank accession numbers for downloaded *Ptychoramphus* and *Aethia* sequences.

Species	Museum Voucher	Cytochrome *b*	Cytochrome Oxidase subunit I	NADH dehydrogenase subunit 2	12S	16S
Ptychoramphus	1B2650	U37302	EF380330	EF373261	EF373103	EF380293
Ptychoramphus	JP2652	AY666369				
A. cristatella	JP3166	U37087	EU525290	EF373219	EF373064	EF380278
A. cristatella	JP3175		EU525291			
A. cristatella	JP3162		EU525289			
A. cristatella	JP2221		AY666339			
A. psittacula	PAAU-2340	U37296	EF380327	EF373235	EF373077	EF380290
A. psittacula	JP2366		AY666218			
A. psittacula	JP2351		AY666219			
A. psittacula	JP2340		EU525292			
A. psittacula	JP2369		EU525293			
A. pusilla	JP3140	U37104	EF380316	EF380337	EF380303	EF380279
A. pusilla	AMNH-DOT1410		DQ432699			
A. pusilla	JP3139		AY666223			
A. pygmaea	JP3395	U37286	EU525294	EF380338	EF380304	EF380280
A. pygmaea	JP3386		AY666212			
A. pygmaea	JP3387		EU525295			
A. pygmaea	JP3394		EU525296			

CHAPTER 2:

Patterns of divergence within nine lineages of trans-Beringian birds[1]

2.1 ABSTRACT

During the Pleistocene Era (2 million years ago to 10,000 years ago), Beringia was a large, ice-free land mass that experienced occasional inundations during interglacial periods. During periods of glaciation, tundra vegetation covered the exposed land mass and supported a variety of animal life; when the climate warmed and the glaciers melted, the land fauna that was connected across the intermittently exposed Beringian land was vicariantly split by water to occupy both sides of the newly-formed Bering Strait. This cycle repeated at least 9 and perhaps as many as 20 times. The biotic effects of each flooding event presumably was shared among birds on both sides of the Bering Sea. In this study we examined patterns of both mitochondrial and nuclear genetic divergence at three taxonomic levels (populations, subspecies, and species) using paired (trans-Beringean) samples from three lineages each in three avian orders (Anseriformes, Charadriiformes, and Passeriformes). Our nine lineage pairs showed evidence of 3-9 distinct divergence events. We also found a lack of concordance between mtDNA and AFLP estimates of divergence and little or no relationship between taxonomic relatedness and genetic differentiation between taxon pairs.

[1]Humphries, E.M. and K. Winker. 2008. Patterns of divergence within nine lineages of trans-Beringian birds. Prepared for submission to Molecular Ecology.

2.2 Introduction

During the Pleistocene Era (2 million years ago to 10,000 years ago), Beringia was a large, ice-free land mass that experienced occasional inundations during interglacial periods (Hopkins 1967). Although the name "Bering Land Bridge" implies a narrow strip of land connecting Asia and North America, at its narrowest Beringia stretched 1,000 miles north to south (Hopkins 1959, Hopkins et al. 1965). At the heights of glaciation, Beringia was isolated from North America and western Asia by ice sheets; as a result, it functioned as an important Pleistocene refugium (Pielou 1991). Thus, passage of species through the land bridge to North America (and vice-versa) may have been impeded during the glacial maxima and during the interglacial flooding of Beringia.

The cycle of glacial and interglacial periods during the Pleistocene caused a matching cycle of exposure and submergence of the Beringian land mass (Hewitt 1996, 2004). During the Pleistocene, the Bering-Chukchi platform, which connects the North American and Asian continents, was exposed at least nine times (Hopkins 1967) and possibly as many as 20 times (Pielou 1991). When exposed, this land mass was covered with vegetation and supported a variety of animal life (Pielou 1991). As the climate warmed and the glaciers melted, the Bering-Chukchi platform would flood and divide North America and Asia into two separate continents (Hopkins 1959, 1967). Terrestrial life that was connected across the intermittently exposed Beringian land was vicariantly split by water to occupy both sides of the newly-formed Bering Strait. Eventually, the

climate cooled down and the glaciers reformed, exposing the Beringian land mass again, allowing both plant and animal life from each continent to expand towards the other.

Each time the cycle of exposure and submergence of the Bering-Chukchi platform repeated itself, the same basic cycle (possible separation and reuniting of taxa) occurred. The biotic effects of this cycle of events should be shared among birds on either side of Beringia. Additionally, taxa that split during an early flooding event should be more genetically divergent than taxa that split during one of the more recent interglacial periods. The repetitive vicariance barrier that Beringian emersion and immersion cycles have produced among terrestrial organisms might cause multiple, coherent, and detectable 'ripple' effects in the genetic divergences of these organisms on each side of the Bering Strait.

We examined genetic divergences (mitochondrial and nuclear) within avian lineages at three levels of phenotypic divergence (populations, subspecies, and species) to encompass a wide period, using taxonomy as a categorical measure of phenotypic divergence. Patterns of genetic divergence were determined using paired samples from three lineages each in three avian orders (Anseriformes, Charadriiformes, and Passeriformes). We estimated the probable number of divergence events that occurred among these nine paired samples, and we also examined whether there was any correlation between phenotypic divergence (taxonomic status) and genetic divergence based on mtDNA. Finally, we examined the relationship between mtDNA and AFLP estimates of divergence.

2.3 Methods

Genotypic divergence between closely-related taxa is often measured using mitochondrial DNA (mtDNA) sequence data. Hundreds of studies have used mtDNA (Zink and Barrowclough 2008), although this method is not without its problems. The mitochondrial genome is inherited as a single unit, so mtDNA sequence tracks the history of a single locus, which is not necessarily the same as the evolutionary history of the entire genome (Ballard and Whitlock 2004). Thus, corroboration of the patterns occurring in mtDNA with patterns in nuclear DNA is desirable (Rubinoff and Holland 2005), but perhaps elusive.

There are two different classes of methods for examining diversity and divergence in the nuclear genome – direct sequencing of nuclear genes and fragment analysis. When working with shallow divergences (especially between groups that are not reciprocally monophyletic in mtDNA), nuclear DNA presents two distinct problems: one, on average nuclear DNA mutates more slowly than mtDNA, so nuclear genes generally show less variation than mtDNA (Brown 1983); and two, nuclear DNA has a larger effective population size than mtDNA, so variation takes longer to sort among structured populations (Zink and Barrowclough 2008), making any given nuclear gene less likely than mtDNA to track the taxon tree (Moore 1995).

We used amplified fragment length polymorphisms (AFLPs; Vos et al. 1995) to determine genetic divergence in the nuclear genome. Because fragment analyses such as AFLPs compare the sizes of fragments among samples, a single nucleotide change can provide phylogenetic signal. Thus, these techniques are often able to differentiate between taxa at very shallow levels of divergence, even when mtDNA is uninformative

(e.g., African cichlids, Allender et al. 2003; snow and McKay's buntings, Maley 2006; *Laupala* crickets, Mendelson et al. 2004). Additionally, the AFLP technique generates a large number of presumably unlinked nuclear loci, thus allowing a broad assessment of the entire nuclear genome. The main assumption in using the AFLP technique is that bands that are the same length are homologous; this assumption appears to be increasingly violated as taxonomic distance increases (Bremer 1991; Mechanda et al. 2004). Thus, there may be a genetic depth beyond which AFLPs are no longer useful.

Mitochondrial DNA (mtDNA) sequencing

DNA was extracted from muscle tissue of 162 birds collected in Alaska and Russia using a DNeasy Tissue Kit (Qiagen, Valencia, CA, USA). Species from three different orders at three different taxonomic levels were sampled (see Appendix 1). We used currently accepted taxonomic designations (Gibson and Kessel 1997, Dickinson 2003) to choose "sister" pairs. We amplified the mtDNA NADH dehydrogenase subunit 2 (ND2) gene using primers L5215 (Hackett 1996) and H6313 (Sorenson et al. 1999). Amplification was performed following standard PCR protocols, using an annealing temperature of 50˚C. Cycle-sequencing was performed using ABI Big-Dye Terminator mix and an annealing temperature of 50˚C. Samples were bidirectionally sequenced using a combination of the primers used for amplification and the internal primers L5682 and H5776. Products were sequenced on an ABI 3130 automated sequencer (Applied Biosystems, Inc., USA). We aligned sequences by eye without gaps using Sequencher (Genecodes Corporation, Inc., USA).

Amplified fragment length polymorphism (AFLP) analyses

All samples (Appendix 1) were also used for AFLP analyses. We followed the same modified ABI protocols as detailed in Chapter 1. We used six MseI/EcoRI selective amplication primer pairs (Appendix 2).

Comparison of divergence estimates from mtDNA and AFLPs

We estimated net nucleotide difference (D_A, the average number of substitutions per site between populations minus the average number of substitutions per site within populations; Nei 1978) between our Alaskan and Russian samples for our mtDNA sequence data in Arlequin ver. 3.0 (Excoffier et al. 2005). We also calculated Nei's (1978) genetic distance based on Lynch and Milligan's (1994) method for each taxon comparison for our AFLP data using the program AFLP-SURV (Vekemans 2002). We chose to calculate genetic distance using a "band-based" method as opposed to an "allele-based" method to accommodate our low sample size of at most 10 individuals per taxon (Bonin et al. 2007). We then plotted D_A versus Nei's (1978) genetic distance to examine how the two estimates were correlated. We used Statistica (StatSoft Inc., Tulsa, OK) to fit a regression line and calculate the corresponding *P*-value.

We also estimated F_{ST} values (and *P*-values) for all of our comparisons using Arlequin (for mtDNA data; Excoffier et al. 2005) and AFLP-SURV (for AFLP data; Vekemans 2002).

Estimation of number of divergence events

We estimated the most probable number of vacariant events that caused divergence among our nine comparisons using the program msBayes (Hickerson et al. 2006; Hickerson et al. 2007). We ran 3 million simulations, bounding the prior distribution for

theta between 0.5 and 100. We set the upper bound for the prior distributions of τ at 40.0, migration rate at 10, and ancestral population size at 0.5. All other parameters were set to 0. We analyzed our simulations using the recommended tolerance of 0.0005.

2.4 Results

Comparison of divergence estimates from mtDNA and AFLPs

Net nucleotide differences (D_A) ranged from 0.0017 to 0.0273 for our 9 pairwise mtDNA comparisons; Nei's genetic distance was 0.0 - 0.1981 (Table 1). Under both methods, *Luscinia svecica* showed the least amount of differentiation, and the *Pica hudsonia*/*Pica pica* comparison showed the greatest amount. Level of taxonomic divergence was not necessarily an accurate predictor of the degree of differentiation of two taxa (Figure 2). Only two comparisons (*Anas americana*/*Anas penelope* and *Pica hudsonia*/*Pica pica*) had both mtDNA and AFLP estimates of population differentiation that were significantly different from 0 ($P \leq 0.05$, from F_{ST} estimates; Table 2); an additional four pairs (*Tringa incana*/*Tringa brevipes*, *Numenius phaeopus hudsonicus*/*Numenius phaeopus variegatus*, *Anas crecca carolinensis*/*Anas crecca crecca* and *Pinicola enucleator flammula*/*Pinicola enucleator kamschatkensis*) had only mtDNA estimates that were significantly different from 0. Two more comparisons (*Clangula hyemalis*/*Clangula hyemalis* and *Pluvialis squatarola*/*Pluvialis squatarola*) had only AFLP estimates that were significantly different from 0.

Linear regression showed an insignificant relationship between these two measures of genetic divergence ($P > 0.05$; Figure 2).

Estimation of number of divergence events

The posterior distribution of ψ (the number of possible vicariant events that caused diversification among the 9 lineage pairs) showed low density values from 3 to 8 divergence events and peaked at 9 divergence events (Figure 3). The 95% probability density was 2.70 to 9; we thus conclude that 3-9 separate vicariant events affected the 9 lineages.

2.5 Discussion

The mitochondrial genome has an effective population size that is one-fourth the effective population size of the nuclear genome (Avise 2004). Thus, the mitochondrial genome should show evidence of differentiation before the nuclear genome. Additionally, if we were to compare D_A values from mtDNA sequence data to D_A values from nuclear sequence data, we would expect the slope of the regression line to be exactly 0.25. However, we are comparing a genetic differentiation estimate from mtDNA sequence data with an estimate from a restriction technique. Our estimation using mtDNA is fairly standard, but estimates of differentiation using AFLPs are expected to be biased downward due to allele homoplasy (Koopman and Gort 2004). This can limit the power of the analysis (Vekemans et al. 2002). So our only expectation of the slope of any regression line will be less than 1.

Mitochondrial and nuclear divergence estimates failed to show any correlation. We were also only able to reject panmixia (or fail to reject to panmixia) in both the mitochondrial and nuclear genomes in three of our nine lineages: *Luscinia svecica* (failure to reject panmixia), *Anas americana*/*Anas penelope*, and *Pixa hudsonia*/*Pica pica*. The mitochondrial and nuclear genomes in our study show different evolutionary

histories. Because the mitochondrial genome is maternally inherited as a single unit, it has a smaller effective population size and is more likely to be affected by drift, introgression, or selective sweeps (Ballard and Whitlock 2004). The nuclear genome is biparentally inherited and contains many independently inherited chromosomes, which theoretically track the history of the species more accurately than the mitochondrial genome (Edwards et al. 2005).

If dispersal is male biased, as in most birds, we would expect mtDNA to show greater evidence of divergence than the nuclear genome. Indeed, in two lineages where there is female-biased philopatry (the *Anas crecca* lineage and the *Anas Americana*/*Penelope* lineage; Johnson 1995, Mowbray 1999), the mtDNA differentiation was higher than the nuclear genome differentiation. All birds except the Anseriformes have male-biased philopatry; the *Pica* passerine lineage has an mtDNA divergence estimate for this lineage that is similar in magnitude to the estimates for the *Anas* duck lineages, but the nuclear DNA divergence estimate was 500 times greater than the larger of the two *Anas* nuclear DNA estimate. In most of the other (non-Anseriform) lineages, only the mtDNA estimates were significantly different from panmixia. (Neither the mtDNA or the nuclear genome estimates in the *Luscinia* lineage, a population-level split, were significantly different from panmixia.) However, two lineages do not follow expectations. In the *Clangula* duck lineage, which also exhibits female-biased philopatry (Robertson and Savard 2002), the nuclear genome estimate was both higher than the mtDNA estimate and the only one of the two estimates that was significantly different from panmixia, while the *Pluvialis* lineage (which has male-biased philopatry) shows the

opposite pattern. The stochastic process of gene sorting and the confounding influence of gene flow between separate groups make determination of precise divergence levels uncertain, especially when little time has passed after separation of the groups. All individuals from our species-level comparisons could be correctly assigned to continent using mtDNA; nearly all individuals from our subspecies-level comparisons could also be correctly assigned to continent. Despite this, there are still rampant stochastic processes at work, both within populations (in mtDNA) and between populations (nuclear DNA).

Taxonomic status was a very poor predictor of both mitochondrial and nuclear genetic divergence in most instances (Hendry et al. 2000; Avise and Walker 2000). Taxonomy is based on phenotypic characters, which do not necessarily correlate to genotype. Taxonomy in birds specifically has been based on external morphology (mostly plumage), which is likely to be under both natural and sexual selection (e.g., Mayr and Ashlock 1991). Phenotype can change faster than genotype, and phenotypic changes can precede genotypic ones (Brumfield et al. 2001; Kondo et al. 2008); the reverse is also true (Omland et al. 2000; Winker and Pruett 2006). Large phenotypic changes can occur as the result of a small environmental change (West-Eberhard 2003), which can then create a reproductive barrier between two previously interbreeding populations. The resulting genotypic change would lag in magnitude behind the phenotypic change, but would not necessarily contradict the phenotypic signal. Using mtDNA, Zink et al. (1995) also found a lack of consistency between taxonomic status and genetic differentiation among 13 trans-Beringian species pairs. This pattern may be

the result of differing selection pressures on phenotype across different avian lineages, or it may reflect a lack of consistency in amount of phenotypic divergence necessary for elevation to species or subspecies status in birds.

In our study there was not a correlation between divergence estimated using mtDNA and divergence estimated using AFLPs, nor was there correlation between taxonomic status and genotypic differentiation. The lack of genetic correlation is likely the result of the different evolutionary histories of the mitochondrial and nuclear genomes. Additionally, our nine lineages were affected by at least three and possibly as many as nine unique vicariant events. These results suggest that Beringian birds have complex and discordant genetic and phenotypic histories.

a)

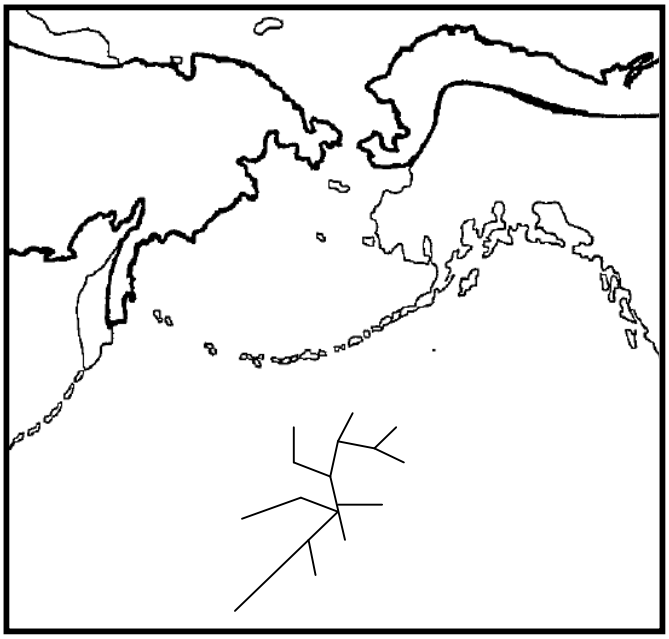
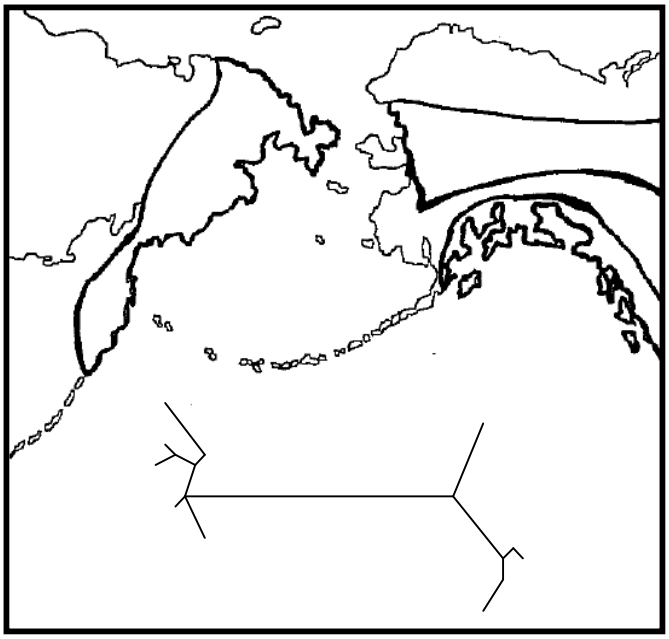


Figure 2.1. Distributions and mtDNA networks of taxa. OTUs are unlabelled. Networks were inferred using parsimony. a) *Luscinia svecica*; b) *Pinicola enucleator flammula*/*Pinicola enucleator kamschatkensis*; c) *Pica hudsonia*/*Pica pica*; d) *Pluvialis squatarola*; e) *Numenius phaeopus hudsonicus*/*Numenius phaeopus variegatus*; f) *Tringa incana*/*Tringa brevipes*; g) *Clangula hyemalis*; h) *Anas crecca carolinensis*/*Anas crecca crecca*; i) *Anas americana*/*Anas penelope*.

b)



c)

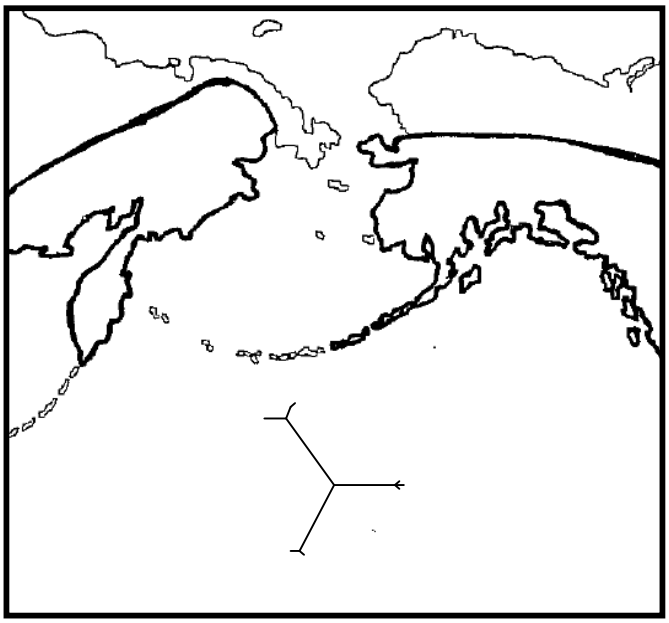
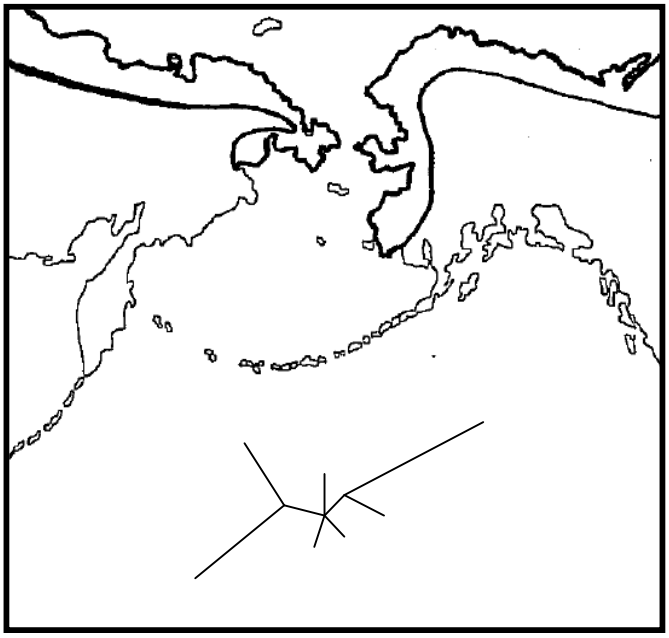


Figure 2.1 continued…

d)



e)

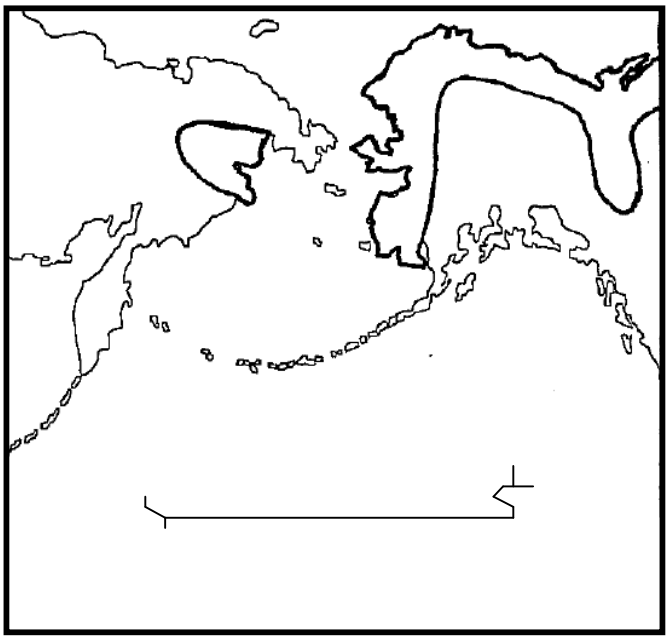
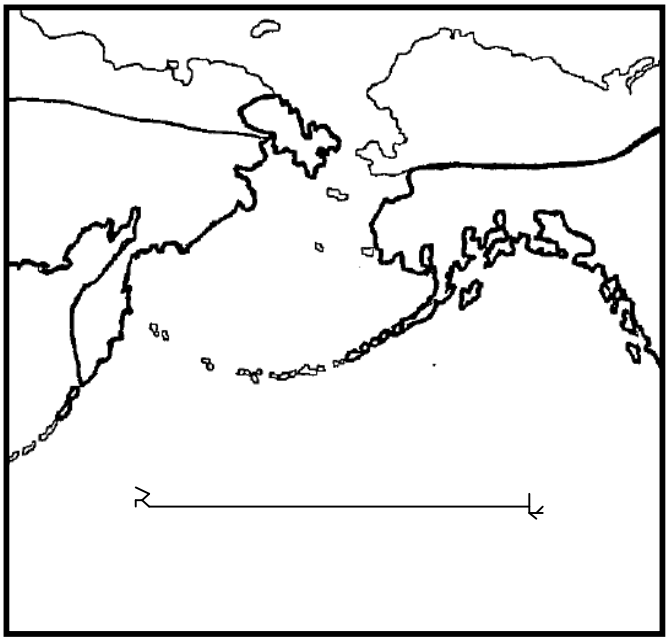


Figure 2.1 continued…

f)



g)

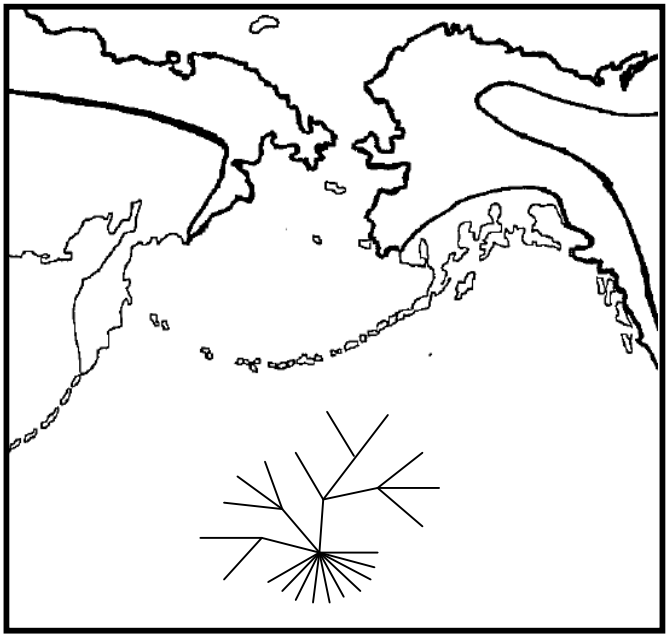
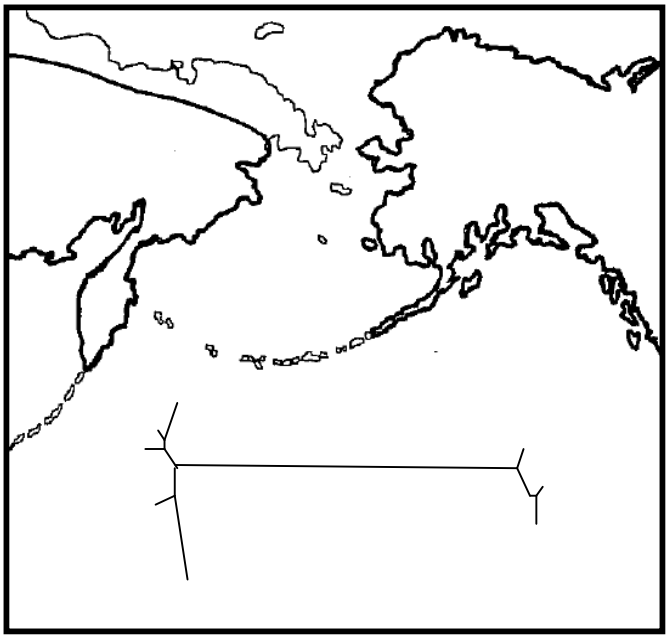


Figure 2.1 continued…

h)



i)

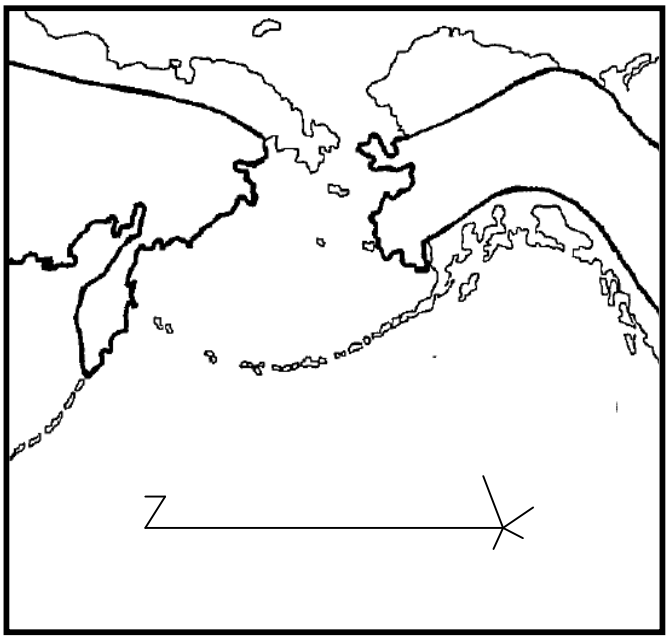


Figure 2.1 continued…

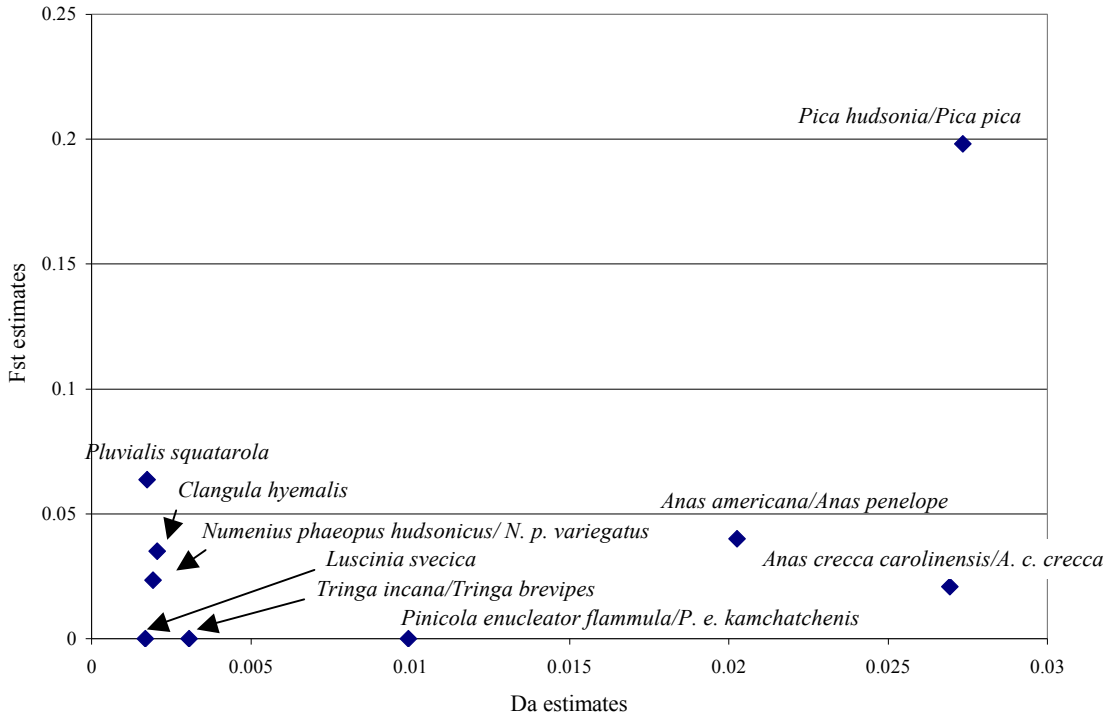


Figure 2.2. D_A plotted against corresponding Nei's genetic distance values. D_A estimates were calculated using mtDNA sequence data, whereas Nei's genetic distance values were calculated using AFLP data.

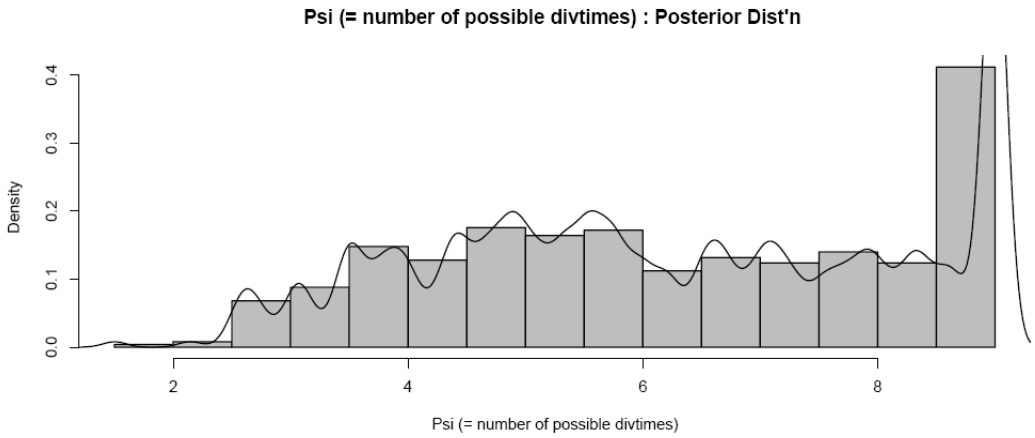


Figure 2.3. Posterior distribution of number of vicariant events. Figure from msBayes output.

Table 2.1. D_A and Nei's genetic distance values. The Alaskan taxon is listed first; the Russian taxon is listed second.

Taxononic comparison	D_A	Nei's genetic distance
Luscinia svecica/ *Luscinia svecica*	0.00169	0.0000
Pinicola enucleator flammula/ *Pinicola enucleator kamchatchensis*	0.00994	0.0000
Pica hudsonia/ *Pica pica*	0.02735	0.1981
Pluvialis squatarola/ *Pluvialis squatarola*	0.00174	0.0638
Numenius phaeopus hudsonicus/ *Numenius phaeopus variegatus*	0.00193	0.0235
Tringa incana/ *Tringa brevipes*	0.00306	0.000
Clangula hyemalis/ *Clangula hyemalis*	0.00206	0.0350
Anas crecca carolinensis/ *Anas crecca crecca*	0.02694	0.0209
Anas americana/ *Anas penelope*	0.02025	0.0401

Table 2.2. F_{ST} for all comparisons. Statistics with $P \leq 0.05$ are in bold.

Taxononic comparison	F_{ST} (mtDNA)	F_{ST} (AFLP)
Luscinia svecica/ *Luscinia svecica*	0.1287	-0.0641
Pinicola enucleator flammula/ *Pinicola enucleator kamchatchensis*	**0.8055**	-0.0921
Pica hudsonia/ *Pica pica*	**0.6989**	**0.3902**
Pluvialis squatarola/ *Pluvialis squatarola*	-0.0526	**0.0882**
Numenius phaeopus hudsonicus/ *Numenius phaeopus variegatus*	**0.9547**	0.0224
Tringa incana/ *Tringa brevipes*	**0.8437**	-0.0480
Clangula hyemalis/ *Clangula hyemalis*	-0.0393	**0.0530**
Anas crecca carolinensis/ *Anas crecca crecca*	**0.7450**	0.0315
Anas americana/ *Anas penelope*	**0.9772**	**0.0661**

Table 2.3. Individuals used in this study. UAM and Burke Museum voucher numbers included. The side of the Bering Strait each taxon represents is indicated beneath the species name.

Species	Voucher No.
Anas crecca carolinensis (Alaska)	UAM 11920, UAM 14961, UAM 11251, UAM 14034, UAM 11340, UAM 11339, UAM 11338
Anas crecca crecca (Russia)	UWBM 56971, UWBM 71261, UWBM 71264, UWBM 56569, UWBM 71263, UWBM 44476, UWBM 43947, UWBM 71265, UAM 9255, UAM 14100
Anas americana (Alaska)	UAM 11908, UAM 11909, UAM 11916, UAM 11919, UAM 11922, UAM 11923, UAM 11924, UAM 11927, UAM 11928, UAM 11929
Anas penelope (Russia)	UAM 8758, UAM 9359, UAM 10008, UAM 11803, UAM 11804, UAM 14529, UAM 17756, UAM 14595, UAM 15180, UAM 14538
Clangula hyemalis (Alaska)	UAM 9395, UAM 11468, UAM 11602, UAM 13154, UAMX 3357, REW 620, REW 583, REW 532, REW 618, REW 619, REW 531, REW 584
Clangula hyemalis (Russia)	UWBM 43893, UWBM 43894, UWBM 43895, UWBM 43913, UWBM 43916, UWBM 43917, UWBM 43918, UWBM 43919, UWBM 4370
Luscinia svecica (Alaska)	UAM 8584, UAM 10809, UAM 8622, UAM 8944, UAM 8945, UAM 8946, UAM 13718, UAM 17727, UAM 15419

Table 2.3 continued…

Luscinia svecica (Russia)	UWBM 44629, UWBM 44630, UWBM 44233, UWBM 44242, UWBM 44243, UWBM 44246, UWBM 44360, UWBM 44361, UWBM 44362, UWBM 44363
Numenius phaeopus hudsonicus (Alaska)	UAM 11044, UAM 13349, UAM 13423, UAM 13925, UAM 9328, UAM 9260, UAM 11508, UAM 11760, UAM 17980, JMM 384
Numenius phaeopus variegatus (Russia)	UAM 9426, UAM 10049, UAM 11351, UAM 14231, UAM 14230, UAM 14225, UAM 14229, UAM 14227, UAM 14223, DDG 1920
Pica hudsonia (Alaska)	UAM 10139, UAM 10140, UAM 10141, UAM 10142, UAM 12453, UAM 13048, UAM 13052, UAM 13053, UAM 13238, UAM 14665
Pica pica (Russia)	UWBM 44584, UWBM 44585, UWBM 47197, UWBM 72084, UWBM 72091, UWBM 74569, UWBM 74697, UWBM 74872, UWBM 44581
Pinicola enucleator flammula (Alaska)	UAM 11285, UAM 11287, UAM 11286, UAM 10157, UAM 8797, UAM 10158, UAM 8794, UAM 7362, UAM 8563, UAM 11831
Pinicola enucleator kamschatkensis (Russia)	UWBM 44628, UWBM 47313, UWBM 47314, UWBM 47315, UWBM 47316, UWBM 51627, UWBM 51628, UWBM 51642, UWBM 51643
Pluvialis squatarola (Alaska)	UAM 14237, UAM 14241, UAM 14238, UAM 14239, UAM 14240, UAM 13347, UAM 13389, UAM 13488, UAM 13390

Table 2.3 continued…

Pluvialis squatarola (Russia)	UWBM 51608, UWBM 43931, UWBM 43963, UWBM 43964, UWBM 44500
Tringa brevipes (Russia)	UAM 7534, UAM 7535, UAM 8521, UAM 8805, UAM 9398, UAM 9939, UAM 9400, UAM 9402, UAM 9404, UAM 10112
Tringa incana (Alaska)	UAM 8420, UAM 10101, UAM 10135, UAM 10176, UAM 10496, UAM 13434, UAM 21813, UAM 11759, UAM 13569, UAM 15181

2.8 Literature Cited

Allender, CJ, Seehausen O, Knight ME, Turner GF, Maclean N (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. *Proceedings of the National Academy of Science*, **100**, 14074-14079.

Avise, J, Walker, J (2000) Abandon all species concepts? A response. *Conservation Genetics* **1**, 77-80.

Avise, J (2004) Molecular markers, natural history and evolution. Sinauer Associates, Inc. Sunderland, Massachusetts.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729-744.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists. *Molecular Ecology*, **16**, 3737-3758.

Bremer B (1991) Restriction data from chloroplast DNA for phylogenetic reconstruction: is there only one accurate way of scoring? *Plant Systematics and Evolution*, **175**, 39-54.

Brown WM (1983) Evolution of animal mitochondrial DNA. In: *Evolution of Genes and Proteins* (eds Nei M, Koehn RK), pp. 62-88. Sinauer Associates, Sunderland, Massachusetts.

Brumfield, RT, Jernigan, RW, McDonald, DB, Braun, MJ (2001) Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. *Evolution*, **55**, 2070-2087.

Dickinson, E, ed (2003) The Howard and Moore complete checklist of the birds of the world. 3rd Edition. Christopher Helm, London.

Edwards, S, Jennings, WB, Shedlock, A (2005) Phylogenetics of modern birds in the era of genomics. *Proceedings of the Royal Society B-Biological Sciences* **272**, 979-992.

Excoffier, L, Laval, G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Gibson, D. Kessel, B (1997) Inventory of the species and subspecies of Alaska birds. *Western Birds* **28**, 45-95.

Hackett SJ (1996) Molecular phylogenetics and biogeography of tanagers in the genus Ramphocelus (Aves). *Molecular Phylogenetics and Evolution*, **5**, 368-382.

Hendry A., Vamosi S, Latham S, Heilbuth J, Day T (2000) Questioning species realities. *Conservation Genetics* **1**, 67-76.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hewitt GM (2004) Biodiversity: a climate for colonization. *Heredity*, **92**, 1-2.

Hickerson MJ, Stahl E, HA Lessios (2006) Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, **60**, 2435-2453.

Hickerson MJ, Stahl E, N Takebayashi (2007) msBayes: Pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics*, **8**, 268.

Hopkins DM (1959) Cenozoic history of the Bering Land Bridge. *Science*, **129**, 1519-1528.

Hopkins DM, ed. (1967) The Cenozoic history of Beringia-a synthesis. In The Bering land bridge. Stanford: Stanford University Press.

Hopkins DM, MacNeil FS, Merklin RL, Petrov OM (1965) Quaternary correlations across Bering Strait. *Science*, **147**, 1107-1114.

Johnson K (1995) Green-winged Teal (Anas crecca), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/193

Koopman WJM, Gort G (2004) Significance tests and weighted values for AFLP similarities, based on Arabidopsis *in silico* AFLP fragment length distributions. *Genetics*, **167**, 1915-1928.

Kondo B, Peters JL, Rosensteel BB, Omland KE (2008) Coalescent analyses of multiple loci support a new route to speciation in birds. *Evolution*, **62**, 1182-1191.

Lynch M, Milligan BG (1994) Analysis of population genetic-structure with RAPD markers. *Molecular Ecology*, **3**, 91-99.

Maley J (2006) A morphological and genetic investigation of the highest-latitude endemic passerine: McKay's Bunting. Masters Thesis. University of Alaska Fairbanks.

Mayr E, Ashlock P (1991) Principles of Systematic Zoology, 2nd edition. McGraw-Hill, New York.

Mendelson TC, Siegel AM, Shaw KL (2004) Testing geographical pathways of speciation in a recent island radiation. *Molecular Ecology*, **12**, 3787-3796.

Mechanda SM, Baum BR, Johnson DA, Arnason JT (2004) Sequence assessment of comigrating AFLP™ bands in *Echinacea* – implications for comparative biological studies. *Genome*, **47**, 15-25.

Mowbray T (1999) American Wigeon (Anas americana), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/401

Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49,** 718-726.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **76**, 379-390.

Omland KE, Tarr CL, Boarman WI, Marzluff JM, Fleischer RC (2000) Cryptic genetic variation and paraphyly in ravens. *Proceedings of the Royal Society of London B*, **267**, 2475-2482.

Pielou EC (1991) After the ice age: The return of life to glaciated North America. University of Chicago Press, Chicago.

Robertson, GJ, Savard, J-PL (2002) Long-tailed Duck (*Clangula hyemalis*), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/651 doi:10.2173/bna.651

Rubinoff D, Holland BS (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, **54**, 952-961.

Sorenson MD, Ast JD, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution*, **12**, 105-114.

Vekemans X (2002) AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139-151.

Vos P, Hagers R, Bleeker M *et al.* (1995) AFLP:new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407-4414.

West-Eberhard MJ (2003) Developmental plasticity and evolution. Oxford University Press, New York.

Winker K, Pruett CL (2006) Seasonal migration, speciation and morphological convergence in the avian genus *Catharus* (Turdidae). *Auk*, **123**, 1052-1068.

Zink RM, Rohwer S, Andreev A, Dittmann D (1995) Trans-Beringia comparisons of mitochondrial DNA differentiation in birds. *The Condor* **97**, 639-649.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107-2121.

Appendix 2.1. Selective amplification primer pairs. Only the last three base pairs of the primers are reported. Primers are from Applied Biosystems, Inc. (USA).

Primer combination	MseI primer	EcoRI primer
1	AGG	CTA
2	AAG	CAG
3	ACT	CTT
4	ACA	CAA
5	AAC	CTA
6	ACC	CTC

GENERAL CONCLUSIONS

In both studies, the mitochondrial and nuclear genomes provided discordant genetic signals about the relationships among and between the birds studied. In the first study, we were able to infer a fully-resolved and bifurcating phylogeny of the *Aethia* auklet species using our mtDNA data set. However, the phylogeny inferred from the nuclear data set failed to corroborate the relationships found using the mtDNA data set; in fact, the nuclear data set did not provide any clear phylogenetic signal about the relationships between any of the auklet species. In the second study, the mtDNA and AFLP data sets showed different magnitudes of genetic differentiation between diverging lineages among nine taxa at three taxonomic levels. The mitochondrial and nuclear genomes of Beringian birds appear to have very different evolutionary histories. Much of the discordance discovered in these studies may be due to the different natural histories of the two different genomes.

Literature Cited

Allender, C.J., O. Seehausen, M.E. Knight, G.F. Turner, and N. Maclean. 2003. Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. Proc. Nat. Acad. Sci. 100:14074-14079.

Avise, J.C. 2004. Molecular markers, natural history, and evolution, 2nd ed. Sinauer Associates, Sunderland, Massachusetts.

Hopkins, D.M. 1959. Cenozoic history of the Bering Land Bridge. Science. 129:1519-1528.

Hopkins, D.M., F.S. MacNeil, R.L. Merklin, and O.M. Petrov. 1965. Quaternary correlations across Bering Strait. Science. 147:1107-1114.

Maley, J. 2006. A morphological and genetic investigation of the highest-latitude endemic passerine: McKay's Bunting. Masters Thesis. University of Alaska Fairbanks.

Mendelson, T.C., A.M. Siegel, and K.L. Shaw. 2004. Testing geographical pathways of speciation in a recent island radiation. Molecular Ecology. 12:3787-3796.

Mindell, D.P., M.D. Sorenson, D.E. Dimcheff, M. Hasegawa, J.C. Ast, and T. Yuri. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48:138-152.

Snustad, D.P. and M.J. Simmons. Principles of genetics, 2nd ed. John Wiley and Sons, Inc., New York, New York.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, and M. Kuiper. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.

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