Population genetics and phylogeography of the pygmy nuthatch in Southern California

Thomas Alan Benson

Follow this and additional works at: https://scholarworks.lib.csusb.edu/etd-project

Part of the Population Biology Commons

Recommended Citation

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
POPULATION GENETICS AND PHYLOGEOGRAPHY OF THE
PYGMY NUTHATCH IN SOUTHERN CALIFORNIA

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
in
Biology

by
Thomas Alan Benson
June 2006
POPULATION GENETICS AND PHYLOGEOGRAPHY OF THE
PYGMY NUTHATCH IN SOUTHERN CALIFORNIA

A Thesis
Presented to the
Faculty of
California State University,
San Bernardino

by
Thomas Alan Benson
June 2006

Approved by:

Anthony Metcalf, Ph.D., Chair, Biology

Richard Fehn, Ph.D.

James Ferrari, Ph.D.
ABSTRACT

I examined the population genetic structure, phylogeography, and subspecies structure of pygmy nuthatch (Sitta pygmaea) populations in southern California using mitochondrial DNA sequence data. Population genetic analyses indicate that pygmy nuthatch populations fragmented in the disjunct mountain ranges of southern California exhibit low but significant levels of genetic differentiation. These results correspond to phylogeographic analyses that further suggest that the isolation of pygmy nuthatch populations occurred recently, most likely at the beginning of the current interglacial period (~10,000-20,000 years ago). The mitochondrial DNA data also support a subspecific break between S. p. melanotis and S. p. leuconucha farther north than proposed subspecific breaks that are based upon morphological characters. These results are similar to patterns found in other bird populations exhibiting similar distributions in California and western North America.
ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of various individuals and organizations. I would like to take this opportunity to thank those people.

I thank my advisor, Tony Metcalf, who provided constant support in all aspects and through all phases of this work, and whom I also consider to be a good friend. I also thank Jim Ferrari and Rich Fehn for their careful critique of this thesis, by which I have produced a better final product.

I thank the staff at the San Bernardino County Museum, Gerald Braden, Mike Rathbun, and Todd Hoggan, for assistance in collecting specimens. I also particularly thank Gene Cardiff, not only for assistance in collecting and preparing specimens, but also for teaching me most of what I know about birds.

I thank the staff at the Louisiana State University Museum of Natural Science, Steve Cardiff, Donna Dittman, Robb Brumfeld, and Chris Witt, for facilitating the tissue grant of both pygmy and brown-headed nuthatch specimens for this research.
I thank my lab mates, Ivan Phillipsen and George Walker, for providing technical support and answers to questions they had already figured out on their own.

I thank my fellow graduate students and friends, particularly Ron Spier, Ivan Phillipsen, and Heidy Contreras, for their moral support and providing much needed breaks from research, coursework, and teaching.

I thank my family and friends, MaryAnn and Susie Benson, Sean, Jean, Blake, Ryan and Josh Guarino, Bobb Elliott, and Christina Chandler, for their love, support, and understanding and for providing an easy escape from the stress of graduate school.

And finally I thank the following organizations for their financial support without which this research could not have been conducted: Los Angeles Audubon Society Ralph W. Schreiber Student Research Award; San Bernardino Valley Audubon Society Scientific Research Grant; Southern California Academy of Sciences Student Research Award; and multiple California State University San Bernardino (CSUSB) Associated Students Inc., Student Research Grants. This research was also supported by a National Science Foundation MRI grant (DBI#0321209) to A. Metcalf and J. Ferrari and a CSUSB faculty grant to A. Metcalf.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Genetic Differentiation</td>
<td>3</td>
</tr>
<tr>
<td>Mitochondrial DNA in Population Studies</td>
<td>7</td>
</tr>
<tr>
<td>Sky Island Biogeography</td>
<td>15</td>
</tr>
<tr>
<td>Pygmy Nuthatch as a Model Organism</td>
<td>18</td>
</tr>
<tr>
<td>Population Genetics and Phylogeography of Pygmy Nuthatches</td>
<td>22</td>
</tr>
<tr>
<td>CHAPTER TWO: MATERIALS AND METHODS</td>
<td>27</td>
</tr>
<tr>
<td>Data Collection</td>
<td>27</td>
</tr>
<tr>
<td>Sequence Analysis</td>
<td>35</td>
</tr>
<tr>
<td>Population Genetics</td>
<td>37</td>
</tr>
<tr>
<td>Phylogeography</td>
<td>38</td>
</tr>
<tr>
<td>Subspecies Limits</td>
<td>42</td>
</tr>
<tr>
<td>CHAPTER THREE: RESULTS</td>
<td>45</td>
</tr>
<tr>
<td>Sequence Analysis</td>
<td>45</td>
</tr>
<tr>
<td>Population Genetics</td>
<td>50</td>
</tr>
<tr>
<td>Phylogeography</td>
<td>51</td>
</tr>
<tr>
<td>Subspecies Limits</td>
<td>60</td>
</tr>
<tr>
<td>Chapter/Appendix</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>CHAPTER FOUR: DISCUSSION</td>
<td>62</td>
</tr>
<tr>
<td>Sequence Analysis</td>
<td>63</td>
</tr>
<tr>
<td>Population Genetics</td>
<td>63</td>
</tr>
<tr>
<td>Phylogeography</td>
<td>65</td>
</tr>
<tr>
<td>Subspecies Limits</td>
<td>68</td>
</tr>
<tr>
<td>Conservation Implications</td>
<td>73</td>
</tr>
<tr>
<td>Conclusions</td>
<td>74</td>
</tr>
<tr>
<td>APPENDIX A: HAPLOTYPE LISTS</td>
<td>76</td>
</tr>
<tr>
<td>APPENDIX B: DISTANCE MATRICES</td>
<td>98</td>
</tr>
<tr>
<td>APPENDIX C: INPUT FILES</td>
<td>100</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>113</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. *Sitta pygmaea* Specimen Information........... 28

Table 2. Variable Sites within 3045-Base Pair *Sitta pygmaea* Haplotypes................................. 46

Table 3. Gene Statistics for 3045-Base Pair Haplotypes......................................................... 47

Table 4. Variable Sites within 1639-Base Pair *Sitta pygmaea* Haplotypes................................. 48

Table 5. Gene Statistics for 1639-Base Pair Haplotypes......................................................... 49

Table 6. Estimates of Intrapopulation Variability........ 50

Table 7. $\Phi_{ST}$ Values for Analysis of Molecular Variance Comparisons.................................. 51

Table 8. Distribution of *Sitta pygmaea* Haplotypes and Phylogenetic Clades............................ 56
LIST OF FIGURES

Figure 1. *Sitta pygmaea* Range Map................................. 19

Figure 2. *Sitta pygmaea* Southern California Range Map................................. 21

Figure 3. Primer Map.......................................................... 32

Figure 4. *Sitta pygmaea* Subspecific Break Map............. 43

Figure 5. Bootstrapped Consensus Parsimony Tree......... 53

Figure 6. Bootstrapped Consensus Neighbor-Joining Tree........................................ 54

Figure 7. Consensus Bayesian Tree................................. 55

Figure 8. Minimum Spanning Haplotype Network............ 58

Figure 9. Nested Cladogram.............................................. 59
CHAPTER ONE

INTRODUCTION

The distribution of genetic variation in geographically structured populations is of fundamental interest to evolutionary and population biologists. For thousands of years biologists have described the variation present in nature and classified organisms into discrete units. In the 1900s, following the development of the Linnaean system of classification, Darwin’s theory of evolution, and the biological species concept, biologists began classifying organisms according to their evolutionary relationships. Many questions remained, however, regarding the definition of a species, and whether distinct geographic units were simply local races or reproductively isolated species. Often these questions could not be answered by observation of phenotypic characteristics. The ability to analyze neutral genetic markers (sequences of DNA not influenced by selection) has allowed biologists to measure the evolutionary relationships among populations directly. Observed levels of genetic differentiation among populations can provide a direct link to their evolutionary history. Low genetic differentiation among populations
suggests high levels of gene flow in the present or recent past, indicating connectivity between seemingly disjunct populations. Alternatively, high genetic differentiation among populations suggests a lack of recent gene flow, which may lead to complete isolation of populations and eventually speciation events.

Genetic differentiation is frequently observed in species exhibiting geographic structure (Avise 2000). Populations of such species may be separated by oceans, continents, mountain ranges, valleys, or numerous other geographic barriers. The extent to which these barriers are obstacles to gene flow, along with factors such as effective population size and time since divergence, influences the distribution of genetic variation and the degree of genetic differentiation among populations.

I propose to study the genetic differentiation among southern California populations of the pygmy nuthatch (*Sitta pygmaea*), a small songbird exhibiting a high level of geographic structure and with limited dispersal capabilities (Norris 1958, Kingery and Ghalambor 2001). Populations of pygmy nuthatches occupying disjunct mountain ranges that have been isolated for tens of thousands of years would be expected to show levels of genetic
differentiation consistent with recent divergence. I will examine the genetic structure within and among pygmy nuthatch populations in seven disjunct mountain ranges in southern California. These results will provide insight into the evolutionary ecology and population biology of this species, and allow inferences into the historical biogeography of pygmy nuthatches in southern California. Additionally, the study area encompasses the range of two subspecies of pygmy nuthatch, S. p. melanotis and S. p. leuconucha, the boundary of which is disputed when based upon analysis of morphological characters (Norris 1958, Grinnell and Miller 1944). This study may provide genetic evidence to corroborate one of these morphological subspecies designations.

Genetic Differentiation

Assessment of variation is the basis of any genetic study in population biology. Genetic differentiation provides information about species evolution and population history. Such data can be used to infer the degree of connectivity among populations via migration, evolutionary relationships of populations within species, and can shape decisions regarding biological conservation issues.
Beginning in the 1970s, an effort was made to correlate observed genetic differentiation with evolutionary, ecological, and biogeographic factors of the species being studied (Avise 1998). This effort, termed intraspecific phylogeography (Avise et al. 1987), compares phylogenetic information from a species to that species’ present and historical geographic distribution. Simply stated, the process begins with the identification of genetic markers (e.g., haplotypes of DNA sequences), which are obtained and analyzed for a number of individuals within the population. Gene trees (phylogenetic trees showing the relationship of haplotypes within the population) are produced from the analysis, and these phylogenetic trees are then compared with biogeographic information (i.e., morphological subspecies, geographic subpopulations) to estimate historical gene flow and evolutionary forces and relationships within the species.

The distribution of genetic variation within a species can range from completely unstructured to highly structured (Avise 2000). A genetically unstructured population generally has little geographic subdivision and relatively high gene flow across the range of the population. Highly vagile organisms often form genetically unstructured
populations. For example, several studies have found limited intraspecific genetic differentiation in marine fish species: the combination of widespread distribution, contiguous marine habitat, and high dispersal capability promotes genetic homogeneity (Graves et al. 1984, Gyllensten 1985, Ward et al. 1994).

In contrast, a genetically structured population is often geographically subdivided such that the population consists of several subpopulations among which migration (i.e., gene flow) is restricted to some extent. Many organisms display some degree of genetic population structure, an effect commonly created by the discontinuous distribution of suitable habitat (Hartl and Clark 1997). Subpopulations are separated by biogeographic barriers that can take the form of topographic features, distance, or other abiotic or biotic factors, but vary depending on the resource requirements and evolutionary ecology of the organism being considered. Populations of freshwater fish species occupying different drainages, for example, tend to exhibit high levels of genetic differentiation as a result of terrestrial and marine habitats presenting barriers to migration (Bermingham and Avise 1986, Gyllensten 1985, Ward et al. 1994). Geographic barriers, however, are not
necessarily isolating mechanisms. Although subpopulations are discontinuous, gene flow may still be maintained by individuals migrating or dispersing across these barriers. It is the level of genetic differentiation among subpopulations that is of primary interest to population biologists conducting phylogeographic studies.

Wright (1943) developed a model to describe the effects of migration and genetic drift on the distribution of genetic variation in discontinuous subpopulations. His “island model” proposes that genetic differentiation will occur among discontinuous subpopulations, or continuous subpopulations with short-range dispersal, when the effect of migration is limited (assuming effective population size and selection are sufficiently small). Migration, the movement of an individual from one subpopulation to another, is a homogenizing force that maintains genetic similarity among subpopulations. Drift, or random change in gene frequency, is an opposing evolutionary force that acts to differentiate subpopulations. Migration and drift interact to produce an equilibrium level of genetic variation among subpopulations. When even a few migrants are exchanged among subpopulations per generation, the influence of gene flow is strong enough that the effect of
Drift is negated regardless of population size (Wright 1931). In the absence of significant gene flow, however, genetic drift becomes the primary force of evolution acting on subpopulations. This, coupled with the random accumulation of mutations over time, results in genetic differentiation among subpopulations.

This model can be applied to any species that exhibits an island distribution, including those restricted to sky islands. Sky islands consist of similar habitat types distributed on mountains or mountain ranges at high elevation. Because of the geographic distribution and elevational changes often associated with mountain ranges, these habitat patches are often discontinuous. Sky islands are analogous to oceanic islands, except that instead of mountains rising from the ocean floor above sea level, mountains rise above a specific minimum elevation at which certain habitats exist; intervening lower elevations represent oceans of unsuitable habitat that present potential barriers to migration and gene flow.

Mitochondrial DNA in Population Studies

Phylogenetic studies at the intraspecific level attempt to reveal genetic relationships among populations
of a single species or species complex, usually with regard to subspecific designations, morphological characteristics, biogeographic structuring, or a combination of the above. The primary approach used to reveal these relationships is now the direct analysis of DNA. DNA studies have high resolution in determining genetic differentiation, and in some cases show variation and differentiation not revealed by previous methods such as allozyme electrophoresis (Metcalf et al. 2001, Burton and Lee 1994).

Mitochondrial DNA (mtDNA) has proven particularly useful at the intraspecific level. The advantage of using mitochondrial DNA in intraspecific phylogeographic studies stems from the combination of its haploid chromosome number, maternal inheritance, lack of significant recombination, and relatively fast rate of evolution. MtDNA is a circular genome found in the mitochondria of all animal cells, and it is passed to offspring only by females (but see below). These two qualities yield an effective mitochondrial population size that can be one-fourth the effective population size of the nuclear genome (Birky et al. 1983). Decreased effective population size increases the rate of coalescence (the point at which genealogies converge on a common ancestor), which means that a gene
tree based on mtDNA has a better likelihood of reflecting the true species (or population) tree than does a gene tree based on nuclear DNA (Moore 1995). Additionally, because mtDNA rarely recombines, it tends to generate cleaner gene trees than nuclear DNA in which recombination can obscure branching structure (Moore 1995). Another advantage of mtDNA is its relatively quick rate of evolution, which is 5-10 times faster than in single copy nuclear genes (Brown et al. 1979). This makes mtDNA useful for inferring phylogenies of recently diverged taxa or intraspecific population structure. These qualities give mtDNA a distinct advantage over nuclear DNA in inferring the phylogenies within or among closely related genera, species, or subspecies.

MtDNA does have potential limitations that can lead to incorrect phylogenetic inferences if undetected or not properly addressed. Homoplasy (similarity of sequences in different organisms due to convergent evolution rather than shared common ancestry) due to saturation at substitution sites makes mtDNA less effective for analyzing taxa that diverged more than approximately 10 million years ago (Moore and DeFilippis 1997). Nuclear homologues of mtDNA (Numts) are short sequences of the mitochondrial genome
that have been transposed into the nuclear genome, making it possible to amplify and analyze nuclear rather than mtDNA. This is potentially significant because nuclear homologues evolve more slowly than mtDNA (Sorenson and Quinn 1998). Amplification of Numts can be greatly reduced by using purified mtDNA during PCR reactions, or by extracting whole genomic DNA from tissues rich in mtDNA; careful analysis of resultant sequences for anomalies can also be helpful in the detection of Numts (Sorenson and Quinn 1998). Heteroplasmy (the coexistence of multiple mtDNA genotypes within an individual) in the mitochondrial genome can arise via two pathways: paternal leakage or mutation. Paternal leakage involves the transmission of mtDNA from males to their offspring. While this phenomenon is mostly detected in interspecific crosses, it has been recorded intraspecifically in a few species including fruit flies (Kondo et al. 1990), mussels (Hoeh et al. 1991, Zouros et al. 1992), and birds (Kvist 2003). Heteroplasmy can also arise through mutation via replication error. Recombination has been demonstrated in some animal mtDNA (reviewed by Rokas et al. 2003); studies in birds, however, have failed to detect evidence of mtDNA recombination (Berlin and Ellegren 2001, Berlin et al. 2004). Both
heteroplasmy and recombination are believed to be rare and have little impact on phylogeographic studies (Avise et al. 1987, Berlin et al. 2004). Sex biased dispersal may also introduce error into mtDNA-based studies. Because mtDNA is maternally inherited, mtDNA may fail to detect gene flow at the nuclear level when females exhibit philopatry (limited dispersal from the birth site) and males are the dispersing sex. This would reduce the effectiveness of mtDNA over nuclear DNA in inferring species (or population) trees from gene trees (Hoelzer 1997). Thorough research of the life history of the organism and selection of the appropriate molecular marker in accordance with its evolutionary ecology can avoid this type of error. Overall, mtDNA is the most effective genetic marker for inferring phylogenies among recently diverged groups.

Two mitochondrial markers have been widely used in phylogeographic studies: cytochrome b (cyt b; a protein-coding gene) and the control region (a non-coding region). Cyt b has been the most commonly used mitochondrial marker in avian sequence-based studies (Moore 1997). This gene is particularly attractive for phylogenetic studies because most nucleotide substitutions are synonymous at this level. This allows sequences from individuals in different groups
to be easily aligned and compared using amino acid sequences, and allows for informative nucleotide sequence variation at third-position codon sites (Moore 1997). For example, Kircham et al. (2000) used cyt b data to show subdivision of cave swallow (Petrochelidon fulva) populations in North America into two distinct clades, and that a population in Ecuador is best considered a separate species (P. rufocollaris), consistent with recent classification changes based on morphology. Prychitko and Moore (1999) employed cyt b to analyze the relationships among 8 genera of woodpeckers (subfamily Picinae), demonstrating its usefulness at higher taxonomic levels. The value of cyt b in phylogenetic studies has been well established over the last decade.

The control region is a rapidly evolving portion of the mitochondrial genome that has been employed in a number of phylogenetic studies. For example, analysis of control region sequences have been used to demonstrate restricted gene flow between sedentary and migratory populations of prairie warblers (Dendroica discolor) in the eastern United States (Buerkle 1999), Pleistocene origins of disjunct populations of azure-winged magpies (Cyanopica cyanus) in Europe and Asia (Fok et al. 2002), and bottlenecking and
recent expansion of red knots (*Calidris canutus*) worldwide (Baker et al. 1994). The utility of the control region in these studies arises from relatively low selective constraints (Baker and Marshall 1997). The control region is divided into Domains I, II, and III based on degree of variability and base composition. Domain II and a short sequence (CSB-I) in Domain III are highly conserved across a wide range of taxa, while Domain I and the remainder of Domain III show large amounts of variation. Approximately two-thirds of the nucleotide sites within the control region are potentially phylogenetically informative, compared to approximately one-third in cytochrome *b*. Studies of human mtDNA have revealed substitution rates in the control region ranging from 2.8 to 5 times faster than the rest of the mitochondrial genome (Cann et al. 1984, Aquadro and Greenberg 1983), and comparisons of red knot mtDNA have shown control region substitution rates approximately 2.6 times faster than cyt *b* in that species (Baker and Marshall 1997). As a result, the control region may reveal differentiation among populations or recently diverged taxa that is undetected by cyt *b* or other mitochondrial or nuclear markers.
Numerous phylogenetic studies have sequenced hypervariable regions of mtDNA to analyze genetic relationships among populations (Avise 2000). For example, Idaghdour et al. (2004) provided evidence of genetic differentiation in the houbara bustard (*Chlamydotis undulata*). Populations in the Middle East (C. u. *macqueenii*) were found to be significantly differentiated from populations in North Africa (C. u. *fuertaventurae* and C. u. *undulata*). The authors were also able to infer that the population on the Canary Islands (C. u. *fuertaventurae*) had been colonized twice, but has been relatively isolated since the second colonization about 20,000 years ago.

Other mtDNA sequence-based studies have shown a significant lack of differentiation among populations. Pearce et al. (2002) discovered little geographic subdivision (attributed to high gene flow) among ancient murrelet (*Synthia bororamphus antiquus*) populations sampled across the northern Pacific Ocean from China to British Columbia. Kvist et al. (2001) found no significant genetic differentiation among willow tit subspecies (*Parus montanus*) across northern Eurasia. The same study, however, also revealed significant genetic differentiation between the northern subspecies and the more southerly, conspecific songar tit (*P. m. songarus* and
*P. m. affinis*). These studies bear out the utility of mtDNA in phylogeographic studies, and attest to the range in degree of structure observed in vertebrate populations.

**Sky Island Biogeography**

Sky islands provide an opportunity to study the effects of genetic differentiation on recently diverged taxa. The coniferous forests and alpine habitats of western North America were much more widespread and contiguous during the last ice age than they are today (Pielou 1991). These habitats underwent several cycles of expansion (downward shift) and contraction (upward shift) during the cooler glacial and warmer interglacial periods in the Pleistocene Epoch approximately 1,000,000 to 10,000 years ago (Dorf 1976, Pielou 1991). Habitat contractions resulted in the creation of sky islands of coniferous forest. The current level of genetic differentiation among sky island populations therefore depends on the degree of isolation during interglacial periods and the extent of population admixture during glacial periods. Sky island populations in western North America potentially may have been geographically isolated for anywhere between
approximately 1,000,000 and 10,000 years (Dorf 1976, Pielou 1991).

Few studies have been conducted on the intraspecific phylogeography of sky island fauna. Lamb et al. (1997) examined the distribution of genetic variation in tassel-eared squirrels (Sciurus aberti) on sky islands in the American southwest, revealing distinct eastern and western assemblages, with further genetic subdivision observed in the eastern group. Based on the pattern of differentiation, the authors concluded that climatological events during the Pleistocene may have been important in the divergence of east-west assemblages, while habitat expansion and dispersal were responsible for observed variation within assemblages.

Masta (2000) found that populations of jumping spiders (Habronattus pugillis) inhabiting smaller sky islands in southeastern Arizona were monophyletic, while populations on larger islands were not, a result attributed to incomplete lineage sorting. The author concluded that topography (i.e., elevation) played an important role in the differentiation of populations, and that multiple vicariance events were responsible for the observed patterns of genetic variation.
DeChaine and Martin (2004) examined population structure in an alpine butterfly (*Parnassius smintheus*) in the Rocky Mountains. Analysis suggested distinct northern and southern groups separated by a large dispersal barrier, while population structure within these groups was attributed to cyclic habitat fragmentation and expansion caused by climate change during the last 400,000 years.

Barrowclough et al. (2004) found strong geographic structure corresponding to subspecies groups in a study of blue grouse (*Dendragapus obscurus*) in coniferous forests of western North America. Gene flow was extensive, however, among populations within the *D. o. fuliginosus* subspecies group, and within northern and southern clades of the *D. o. obscurus* subspecies group. The authors suggested that differences between subspecies groups were due to separation during the late Pleistocene, while extensive continuous habitat during glacial epochs was responsible for gene flow within groups.

These studies suggest that western North American sky island populations exhibit varying degrees of differentiation consistent with Pleistocene climate change, and that these populations are recently diverged, having become genetically differentiated only within the last

Pygmy Nuthatch as a Model Organism

The pygmy nuthatch (Sitta pygmaea) is an ideal species for examining genetic variation among recently diverged sky island taxa. It exhibits well-documented morphological variation among subspecies, has a widespread range showing geographic structure and habitat specificity, and displays extremely low vagility in light of its powers of flight (Bent 1948, Norris 1958). These qualities present the foundation for potentially high levels of genetic differentiation among discontinuous sky island populations of pygmy nuthatches.

Pygmy nuthatches are small songbirds (family Sittidae) that inhabit coniferous forests throughout western North America. They range from the Pacific coast east to the Rocky Mountains, and from British Columbia south into central Mexico (Figure 1). There are currently six recognized subspecies: S. p. pygmaea occurring along the Pacific coast in central California; S. p. melanotis inhabiting mountains from British Columbia, Canada south to
southern California (Pacific interior) and northern Mexico (western North American interior); *S. p. leuchonucha* occupying mountains from southern California to Baja California Norte, Mexico; and *S. p. elii, S. p. flavinucha*, and *S. p. brunnescens*, which occur in the mountains of central and northern Mexico (Kingery and Ghalambor 2001).

Pygmy nuthatches are restricted to coniferous forests (Small 1994), and as such, their distribution varies from continuous to fragmented throughout their range. In southern California the current patchy distribution of coniferous forests results in large distances between populations of pygmy nuthatches (Figure 2). *S. p. leuconucha* inhabits ponderosa pine forest, generally above 4000 feet elevation, in seven separate populations located in Riverside County in the San Jacinto and Santa Rosa Mountains, and in San Diego County in the Laguna, Cuyamaca, Volcan, and Hot Springs Mountains and on Palomar Mountain. Although most authorities ascribe the Riverside County birds to this subspecies (Grinnell and Miller 1944, Bent 1948, Small 1994), some have described them as belonging to, or intergrading with, *S. p. melanotis* (Norris 1958, Kingery and Ghalambor 2001). *S. p. melanotis* shares similar habitat requirements and ranges from San Bernardino County.
through northern California. It occurs in disjunct populations in San Bernardino County in the San Bernardino Mountains, Los Angeles County in the San Gabriel Mountains, Ventura County on Frazier Mountain and Mount Pinos, and Kern County on Cerro Noroeste, the Tehachapi Mountains, and the Piute Mountains. The distribution of this subspecies becomes continuous in northern Kern County at the southern
edge of the Sierra Nevada. *S. p. pygmaea* is relatively continuously distributed in San Luis Obispo and Monterey Counties in the Santa Lucia Range and along the coastal fog belt where it prefers long-needled pines (Kingery & Ghalambor 2001).

In addition to being geographically structured, pygmy nuthatches are relatively sedentary. They are a resident species that does not migrate, and shows limited dispersal. Norris (1958) reported mean natal dispersal distances of 286.5 meters, and a maximum recorded dispersal of 533 meters. Therefore, natal dispersal would not promote migration between disjunct mountain ranges. However, pygmy nuthatches are occasionally observed outside their habitat, with a few sightings recorded less than annually in the desert or coastal lowlands (Garrett and Dunn 1981). Thus pygmy nuthatches are capable of rare long distance flights, and could potentially span the distance between disjunct mountain ranges.

Population Genetics and Phylogeography of Pygmy Nuthatches

The primary focus of this study will be to characterize the genetic variation present within and among
populations of two subspecies of pygmy nuthatch occurring in southern California (Figure 2). From these data several other questions relating to the population biology, evolutionary ecology, and biogeography of this species will be addressed. Questions concerning connectivity and gene flow arise about many species with discontinuous distributions. Evolutionary and island models predict that when migration is restricted (and selection is negligible), random drift will act to differentiate populations inhabiting separate islands. This study will examine the degree of genetic differentiation among sky island populations of pygmy nuthatch and assess the hypothesis that sky islands are genetically isolated and that migration (gene flow) is restricted among them.

Up to seven subspecies of pygmy nuthatch have been described based on plumage coloration and skeletal measurements (Norris 1958). Two of these subspecies occur in the study area, with the subspecific designation of one population in dispute. This study will also evaluate whether differences in mitochondrial DNA sequences substantiate these morphological subspecific designations. Analysis of neutral genetic markers could corroborate
morphological subspecies, and shed light on the genetic relationship of disputed populations.

Genetic differentiation can also be used to estimate divergence times of populations and subspecies and infer the historical biogeography of populations. These estimates can be compared with the geological record to determine if paleoclimatological events were important in genetic divergence. Recent studies have shown that Pleistocene climate has played a significant role in both initiating divergence events in North American passerines as well as maintaining previously formed divergence events (Avise and Walker 1998, Johnson and Cicero 2004). The divergence times of pygmy nuthatch populations may be approximated by applying calibrated molecular clocks derived from fossil data from other passerines to observed sequence differentiation (Lovette 2004), potentially allowing the evaluation of the role of Pleistocene climate change on the evolution of pygmy nuthatches. While there is significant variation in molecular evolutionary rates among species, among both genetic markers and species, most rate calibrations of molecular clocks in birds have clustered around 2% per million years (Lovette 2004). A study by Fleischer et al. (1998) on partial cyt b sequences
of Hawaiian honeycreepers revealed a substitution rate of 1.6 - 1.9% per million years. For molecular clock estimates to be accurate, however, the populations being compared must be reciprocally monophyletic (Cicero and Johnson 2004).

The results of this project will be an important contribution to the knowledge of the biology of pygmy nuthatches, as well to the understanding of the evolution and ecology of recently diverged taxa, especially sky island species. Little is known about the evolutionary ecology and the population genetics of pygmy nuthatches. This investigation will provide valuable information to fill in the gaps in our knowledge of this species, with particular regard to the evolutionary relationships and history of this species within and among populations and mountain ranges throughout southern California. Significant genetic differentiation, if observed, would indicate that disjunct populations are isolated and gene flow is restricted among populations. Alternatively, lack of genetic differentiation could be attributed to recent or current migration among populations, providing direction for further research into the mechanisms by which a purportedly sedentary species maintains gene flow.
On a broader scale this project will improve the understanding of the population genetics and phylogeography of sky island species. The results of this research will provide insight into the evolution and behavior of sedentary species occurring on sky islands, especially birds.
CHAPTER TWO

MATERIALS AND METHODS

Data Collection

Two of the three subspecies of pygmy nuthatch occurring in southern California, *S. p. melanotis* and *S. p. leuconucha*, are included within the scope of this study. Eleven sampling locations throughout southern California were selected based on representation of subspecies, availability of specimens, and feasibility of collection. Sampling locations were also chosen to create a hierarchical geographic structure, allowing for the analysis of genetic variation within mountain ranges, among mountain ranges, and between subspecies. Seven specimens were sampled from each location, except Thomas Mountain and Piute Mountain where 6 specimens each were sampled. Eight locations represent the widespread *S. p. melanotis* and three locations the southern *S. p. leuconucha*. Three brown-headed nuthatch (*Sitta pusilla*) samples were included for outgroup analysis. Table 1 summarizes subspecies and collection information for all specimens used in this study.
<table>
<thead>
<tr>
<th>Museum Accession Number</th>
<th>GenBank Accession Number</th>
<th>3045-bp haplotype</th>
<th>1639-bp haplotype</th>
<th>Subspecies</th>
<th>Collection Location</th>
<th>Mountain Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBCM 57726</td>
<td>DQ153996</td>
<td>A1</td>
<td>A</td>
<td>melanotis</td>
<td>Piute Mountain</td>
<td>Piute</td>
</tr>
<tr>
<td>SBCM 57727</td>
<td>DQ194292</td>
<td>A1</td>
<td>A</td>
<td>melanotis</td>
<td>Piute Mountain</td>
<td>Piute</td>
</tr>
<tr>
<td>SBCM 57728</td>
<td>DQ153998</td>
<td>B1</td>
<td>B</td>
<td>melanotis</td>
<td>Piute Mountain</td>
<td>Piute</td>
</tr>
<tr>
<td>SBCM 57729</td>
<td>DQ153985</td>
<td>C</td>
<td>C</td>
<td>melanotis</td>
<td>Piute Mountain</td>
<td>Piute</td>
</tr>
<tr>
<td>SBCM 57730</td>
<td>DQ194297</td>
<td>B1</td>
<td>B</td>
<td>melanotis</td>
<td>Piute Mountain</td>
<td>Piute</td>
</tr>
<tr>
<td>SBCM 57731</td>
<td>DQ194298</td>
<td>B1</td>
<td>B</td>
<td>melanotis</td>
<td>Piute Mountain</td>
<td>Piute</td>
</tr>
<tr>
<td>SBCM 57737</td>
<td>DQ155593</td>
<td>D</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>SBCM 57738</td>
<td>DQ155594</td>
<td>E</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>SBCM 57739</td>
<td>DQ155591</td>
<td>A</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>LSU B-41563</td>
<td>DQ194306</td>
<td>A</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>LSU B-41564</td>
<td>DQ194307</td>
<td>A</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>LSU B-41566</td>
<td>DQ194328</td>
<td>C</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>LSU B-41567</td>
<td>DQ194329</td>
<td>C</td>
<td>melanotis</td>
<td>Cerro Noroeste</td>
<td>Cerro Noroeste</td>
<td></td>
</tr>
<tr>
<td>SBCM 57734</td>
<td>DQ194308</td>
<td>D</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>SBCM 57735</td>
<td>DQ194309</td>
<td>D</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>SBCM 57736</td>
<td>DQ194310</td>
<td>D</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>LSU B-41557</td>
<td>DQ194330</td>
<td>C</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>LSU B-41558</td>
<td>DQ194331</td>
<td>C</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>LSU B-41560</td>
<td>DQ194332</td>
<td>C</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>LSU B-41561</td>
<td>DQ194314</td>
<td>E</td>
<td>melanotis</td>
<td>Frazier Mountain</td>
<td>Frazier Mountain</td>
<td></td>
</tr>
<tr>
<td>SBCM 57633</td>
<td>DQ154000</td>
<td>E1</td>
<td>E</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
</tr>
<tr>
<td>SBCM 57634</td>
<td>DQ153986</td>
<td>F</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 57635</td>
<td>DQ194312</td>
<td>E</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 57636</td>
<td>DQ194325</td>
<td>C</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 57637</td>
<td>DQ194311</td>
<td>D</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 57638</td>
<td>DQ194313</td>
<td>E</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 57639</td>
<td>DQ194303</td>
<td>A</td>
<td>melanotis</td>
<td>Winston Peak</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 56877</td>
<td>DQ153987</td>
<td>G</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 56878</td>
<td>DQ194334</td>
<td>G</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 56898</td>
<td>DQ153997</td>
<td>A2</td>
<td>A</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
</tr>
<tr>
<td>SBCM 56899</td>
<td>DQ194293</td>
<td>A1</td>
<td>A</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
</tr>
<tr>
<td>SBCM 56900</td>
<td>DQ194294</td>
<td>A1</td>
<td>A</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
</tr>
<tr>
<td>SBCM 56901</td>
<td>DQ194295</td>
<td>A1</td>
<td>A</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
</tr>
<tr>
<td>Museum Accession Number</td>
<td>GenBank Accession Number</td>
<td>3045-bp haplotype</td>
<td>1639-bp haplotype</td>
<td>Subspecies</td>
<td>Collection Location</td>
<td>Mountain Range</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>---------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>SBCM 56902</td>
<td>DQ155595</td>
<td>H</td>
<td>melanotis</td>
<td>Blue Ridge</td>
<td>San Gabriel</td>
<td></td>
</tr>
<tr>
<td>SBCM 57596</td>
<td>DQ154001</td>
<td>H1</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57597</td>
<td>DQ194315</td>
<td>H</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57598</td>
<td>DQ194316</td>
<td>H</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57599</td>
<td>DQ154004</td>
<td>I1</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57640</td>
<td>DQ194326</td>
<td>C</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57641</td>
<td>DQ194327</td>
<td>C</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57642</td>
<td>DQ155592</td>
<td>B</td>
<td>melanotis</td>
<td>Delamar Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57591</td>
<td>DQ154005</td>
<td>J1</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57592</td>
<td>DQ153988</td>
<td>K</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57593</td>
<td>DQ154006</td>
<td>L1</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57594</td>
<td>DQ155597</td>
<td>L</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57630</td>
<td>DQ194301</td>
<td>J1</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57631</td>
<td>DQ153989</td>
<td>M</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 57632</td>
<td>DQ194324</td>
<td>L</td>
<td>melanotis</td>
<td>Onyx Peak</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56616</td>
<td>DQ155596</td>
<td>J</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56617</td>
<td>DQ153990</td>
<td>N</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56618</td>
<td>DQ194317</td>
<td>H</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56619</td>
<td>DQ153991</td>
<td>O</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56678</td>
<td>DQ194336</td>
<td>N</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56679</td>
<td>DQ194333</td>
<td>F</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56681</td>
<td>DQ194318</td>
<td>H</td>
<td>melanotis</td>
<td>Grinnell Mountain</td>
<td>San Bernardino</td>
<td></td>
</tr>
<tr>
<td>SBCM 56813</td>
<td>DQ154335</td>
<td>K</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 56814</td>
<td>DQ153999</td>
<td>D1</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 56816</td>
<td>DQ194296</td>
<td>A1</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 56817</td>
<td>DQ194304</td>
<td>A</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57765</td>
<td>DQ154002</td>
<td>H2</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57766</td>
<td>DQ194319</td>
<td>H</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57767</td>
<td>DQ153992</td>
<td>P</td>
<td>leuconucha</td>
<td>Black Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57583</td>
<td>DQ194320</td>
<td>H</td>
<td>leuconucha</td>
<td>Thomas Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57584</td>
<td>DQ153993</td>
<td>Q</td>
<td>leuconucha</td>
<td>Thomas Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57585</td>
<td>DQ194321</td>
<td>H</td>
<td>leuconucha</td>
<td>Thomas Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57586</td>
<td>DQ194305</td>
<td>A</td>
<td>leuconucha</td>
<td>Thomas Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. *Sitta pygmaea* Specimen Information (continued).

<table>
<thead>
<tr>
<th>Museum</th>
<th>GenBank Accession Number</th>
<th>GenBank Accession Number</th>
<th>3045-bp haplotype</th>
<th>1639-bp haplotype</th>
<th>Subspecies</th>
<th>Collection Location</th>
<th>Mountain Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBCM 57587</td>
<td>DQ194980</td>
<td>D1</td>
<td>D</td>
<td>leuconucha</td>
<td>Thomas Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57588</td>
<td>DQ154003</td>
<td>H3</td>
<td>H</td>
<td>leuconucha</td>
<td>Thomas Mountain</td>
<td>San Jacinto</td>
<td></td>
</tr>
<tr>
<td>SBCM 57579</td>
<td>DQ194302</td>
<td>J1</td>
<td>J</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>SBCM 57580</td>
<td>DQ194300</td>
<td>H3</td>
<td>H</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>SBCM 57581</td>
<td>DQ194299</td>
<td>D1</td>
<td>D</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>SBCM 57626</td>
<td>DQ194322</td>
<td>H</td>
<td>H</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>SBCM 57627</td>
<td>DQ194338</td>
<td>P</td>
<td>P</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>SBCM 57628</td>
<td>DQ194337</td>
<td>N</td>
<td>N</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>SBCM 57629</td>
<td>DQ194323</td>
<td>H</td>
<td>H</td>
<td>leuconucha</td>
<td>Laguna Mountain</td>
<td>Laguna</td>
<td></td>
</tr>
<tr>
<td>LSU B-2485</td>
<td>DQ154007</td>
<td>R1</td>
<td>R</td>
<td><em>S. pusilla</em></td>
<td>Louisiana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSU B-2486</td>
<td>DQ153994</td>
<td>S</td>
<td>S</td>
<td><em>S. pusilla</em></td>
<td>Louisiana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSU B-2487</td>
<td>DQ153995</td>
<td>T</td>
<td>T</td>
<td><em>S. pusilla</em></td>
<td>Louisiana</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SBCM—San Bernardino County Museum, LSU—Louisiana State University Museum of Natural Sciences Collection of Genetic Resources; subspecies designations based on Grinnell and Miller (1944).
Tissue samples were acquired from collections at the San Bernardino County Museum and the Louisiana State University Museum of Natural Sciences. Protocols for tissue digestion and DNA extraction follow Hillis et al. (1996) and are summarized as follows. For each specimen, 50 milligrams of frozen liver tissue was macerated on a watch glass and transferred to a 1.5-milliliter centrifuge tube and digested with proteinase-K in a 55°C water bath for a minimum of 8 hours. Total genomic DNA was extracted from the resulting tissue homogenate using 2 phenol-chloroform-isoamyl (25:24:1) and 1 chloroform-isoamyl (24:1) extractions, followed by precipitation with sodium acetate and ethanol. The DNA precipitate was centrifuged and the excess solution decanted. The DNA pellet was air dried in a hood for one hour and then dissolved in distilled water. Concentrations of total genomic DNA were determined by absorbance at 280 nm using a spectrophotometer.

An approximately 3100-base pair fragment of mtDNA was initially amplified and sequenced in 23 samples. This fragment includes the control region, ND6, tRNA^{Glu}, tRNA^{Pro}, tRNA^{Thr}, and a portion of cytochrome b (Figure 3). Primers
<table>
<thead>
<tr>
<th>Name</th>
<th>5′-3′ Sequence</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 L14990</td>
<td>CCATCCACATCTCACGCATGATGAAA</td>
<td>Gill 1999</td>
</tr>
<tr>
<td>2 LCYTbC</td>
<td>AGAGACCTGAAACGTAGGAG</td>
<td>This study</td>
</tr>
<tr>
<td>3 LCYTb1</td>
<td>GAGCCTGAGGTGGATTCTCTG</td>
<td>This study</td>
</tr>
<tr>
<td>4 LCYTb2</td>
<td>CCGCCTCGTCTAGTCTTAT</td>
<td>This study</td>
</tr>
<tr>
<td>5 LND6A1</td>
<td>GCCCGAATYGCCCCCMCGAGA</td>
<td>This study</td>
</tr>
<tr>
<td>6 LND6*</td>
<td>GCCCGAATAGC CCCCGTGA</td>
<td>This study</td>
</tr>
<tr>
<td>7 LGLU</td>
<td>ATTCGCCACCTGGCTCTCTCC</td>
<td>This study</td>
</tr>
<tr>
<td>8 LCR2*</td>
<td>CATTAACCCAAATGCTAC</td>
<td>This study</td>
</tr>
<tr>
<td>9 L437</td>
<td>CTCACCGAGAACCAGGCTACT</td>
<td>Tarr 1995</td>
</tr>
<tr>
<td>10 LCR1</td>
<td>TAAATAGGCCACTTCCGGC  GACTGA</td>
<td>This study</td>
</tr>
<tr>
<td>11 H15487</td>
<td>GATCCTGTTCGTGAGGAGG</td>
<td>Pasquet 1998</td>
</tr>
<tr>
<td>12 HCYTB*</td>
<td>CAGAGAGGATTGAGGAG</td>
<td>This study</td>
</tr>
<tr>
<td>13 H16065</td>
<td>GGAGTCTTCAGTCTCTGTTTACAAGAC</td>
<td>Gill 1999</td>
</tr>
<tr>
<td>14 HND6A</td>
<td>AAYCCYTCTCCYTATATGGGT</td>
<td>This study</td>
</tr>
<tr>
<td>15 HGLU*</td>
<td>GGAGAGGGCCAGGTTGGA</td>
<td>This study</td>
</tr>
<tr>
<td>16 HCR437B</td>
<td>ACCTCGAGATGGTCTGGCTCT</td>
<td>This study</td>
</tr>
<tr>
<td>17 HCR3*</td>
<td>CAGACGTTATGTTGCCGGTC</td>
<td>This study</td>
</tr>
<tr>
<td>18 HCR2</td>
<td>AAAAGTAAATACGGCGGTTCAT</td>
<td>This study</td>
</tr>
<tr>
<td>19 H1248</td>
<td>CTCCTGAGTGCTCTGCT</td>
<td>Tarr 1995</td>
</tr>
<tr>
<td>M13 forward</td>
<td>CACGACGTTGTAATAACGAC</td>
<td></td>
</tr>
<tr>
<td>M13 reverse</td>
<td>GGATAACAATTTCACACAGG</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Primer Map.
Approximate locations of primers used in PCR and sequencing. T, P, E, and F are the tRNAs for threonine, proline, glutamic acid, and phenylalanine, respectively. Figure not to scale. *Primers used for amplification of the 1639-base pair fragment.
used for PCR amplification of this fragment were L14990 (Gill 1999) and H1248 (Tarr 1995). Internal sequencing primers were designed from alignments of the mitochondrial genomes of *Corvus frugilegus*, *Vidua chalybeata*, and *Smithornis sharpei* downloaded from GenBank (accession numbers Y18522, AF090341, NC000879, respectively) and pygmy nuthatch sequence data. Figure 3 lists the PCR and sequencing primers used in this study. PCR reactions were carried out using Eppendorf MasterTaq Kit (Brinkmann Instruments, Inc.) in 50 µL volumes: 5.0 µL template DNA (approximately 0.2 µg/mL), 1.0 µL each primer (20 mM), 0.5 µL polymerase, 5.0 µL dNTPs, 5.0 µL buffer, and 10.0 µL TaqMaster. The profile for PCR reactions was: initial denaturation at 94°C for 2 minutes; 35 cycles of denaturation at 94°C for 50 seconds, annealing at 50°C for 1 minute, and extension at 72°C for 2 minutes; and final extension at 72°C for 2 minutes. PCR products were visualized with gel electrophoresis using a 0.8% agarose gel and a 1-kilobase ladder, and purified using Montage PCR Centrifugal Filter Devices (Millipore Corporation). Purified PCR products were sequenced by Laragen, Inc. (Los Angeles, CA). Double-stranded sequences were constructed
using Sequitherm EXCEL II DNA Sequencing Kit-LC (Epicentre) in 20 µL volumes: 1.0 µL purified PCR product, 1.0 µL each M13 forward and reverse primers, 7.2 µL buffer, 1.0 µL polymerase, and 8.8 µL distilled water. The profile for cycle sequencing reactions was: initial denaturation at 92°C for 2 minutes; 20 cycles of denaturation at 92°C for 30 seconds, annealing at 54°C for 30 seconds, and extension at 70°C for 1 minute. Cycle sequencing products were sequenced on a LI-COR 4300 automated sequencer. Double-stranded sequences were constructed with AlignIR version 2.0 software (LI-COR, Inc.) and aligned in ClustalX (Thompson et al. 1997).

Sequence Analysis

The goal of this study is to characterize the genetic variation and structure of southern California populations of the pygmy nuthatch, and to determine the evolutionary relationships among these populations using mitochondrial DNA sequence data. This study specifically analyzes the potential effect of isolation on the population genetic structure of pygmy nuthatches in southern California, whether the observed genetic structure corresponds to population geographic structure, and the role of historical
biogeographic processes in shaping population genetic structure. To achieve this goal this research examines genetic structure both within and among mountain ranges. This study also examines the correlation of the genetic data with putative morphological subspecies breaks.

Prior to analysis the reverse complement of the ND6 sequence was taken in both the 3045- and 1639-bp fragments using MacClade 4.01 (Maddison and Maddison 2000). This was done to maintain consistent codon positioning with cyt b. A partition homogeneity test was conducted in Paup* 4.10b (Swofford 2003) to determine if the three mitochondrial gene fragments could be combined into a single 1639-base pair sequence for analysis (Farris et al. 1995). Genetic variation is described using basic descriptive statistics including nucleotide frequency, number of haplotypes, number of polymorphic sites, number and type of substitutions, number of transitions (Ti) and transversions (Tv), Ti/Tv ratio, and uncorrected genetic distance among haplotypes. Descriptive statistics were calculated using Paup* 4.10b (Swofford 2003), Mega 2.1 (Kumar et al. 2001) and Modeltest 3.06 (Posada and Crandall 1998), or were done by hand.
Population Genetics

Population genetic analyses use F-statistics or other genetic distance measures to evaluate genetic differentiation within and among populations. Wright’s F-statistics are fixation indices that were developed to characterize genetic differentiation within and among subpopulations (Hartl and Clark 1997). The most commonly used of these, $F_{st}$, measures genetic differentiation among subpopulations that constitute the total population. F-values range from 0 to 1, with increasing values indicating greater degrees of genetic differentiation. Analogs of F-statistics, such as $N_{st}$ (Lynch and Crease 1990), $K_{st}$ (Hudson et al. 1992), and $\Phi_{ST}$ (Excoffier et al. 1992) have been developed to analyze DNA sequence data.

Analysis of molecular variance (AMOVA) and Mantel tests use molecular-based F-statistics (or their analogs) to test for significant population structure. AMOVA employs a nested analysis of variance approach to compare levels of genetic differentiation in hierarchically structured populations (Excoffier et al. 1992). Arlequin 2.0 (Schneider et al. 2000) was used to implement the AMOVA. Individuals were grouped into populations by mountain range. Population comparisons were based on a
Kimura 2-parameter distance matrix (calculated by Arlequin) and significance testing of F values and variance components was based on 16000 permutations and evaluated at p=0.05.

A Mantel test was used to examine the correlation between population pairwise geographic distances (in km) and population pairwise F_{ST} values (calculated by Arlequin). Significant results from a Mantel test indicate isolation by distance among populations. Significance was determined by running 1000 permutations and evaluating at p=0.05.

Fu’s F_s, nucleotide diversity (Π), and percentage of variable sites were calculated as measures of intrapopulation variability in Arlequin 2.0 (Schneider et al. 2000). Fu’s F_s can be used to detect population expansion, and the relative values of nucleotide diversity (Π) among populations may indicate the direction of that expansion if present (Sgariglia and Burns 2003).

Phylogeography

Evolutionary relationships among haplotypes were estimated using three methods of phylogenetic inference: genetic distance, parsimony, and maximum likelihood. Distance methods use genetic distance data to generate gene
trees. Character data are transformed into pairwise distance values (distances between pairs of haplotypes) and plotted in a matrix. Trees can be constructed from the matrix using various approaches, with two of the more common methods being neighbor-joining and minimum evolution. Neighbor-joining connects nearest neighbors within the matrix until all haplotypes are represented in the tree, while minimum evolution searches multiple trees to produce the tree(s) with the shortest overall distance. Neighbor-joining analysis was conducted in Paup* 4.0b10 (Swofford 2003). Modeltest 3.06 (Posada and Crandall 1998) selected the HKY85 (Hasegawa et al. 1985) model of evolution as the best model for this data set. Trees were constructed using an HKY85 model of evolution, branch-swapping via tree-bisection-reconnection, and by conducting 1000 bootstrap replicates.

Parsimony constructs gene trees based on the fewest number of evolutionary steps required to explain differences among haplotypes. Optimality criteria are defined for the transformation of characters, and the most parsimonious tree (or trees) is selected from all possible trees based on the defined criteria. Maximum parsimony is the most commonly used method of phylogenetic inference.
Parsimony analysis was conducted in Paup* 4.0b10 (Swofford 2003). Parsimony trees were constructed using both branch-and-bound (the method guaranteed to find the minimum-length tree) and heuristic searches. Bootstrap analysis was performed using 1000 bootstrap replicates and a full heuristic search, with the initial trees obtained via stepwise addition and branch-swapping via tree-bisection-reconnection.

Maximum likelihood selects the tree that best fits the character data based on evolutionary models chosen by the investigator. The maximum likelihood approach is attractive because it often the estimation method least affected by sampling error, tends to conform to the assumptions of its models, and frequently outperforms other methods when evaluated under multiple models of sequence evolution (Swofford et al. 1996). Maximum likelihood analyses were implemented using MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001). Four chains were run for 30 million generations sampling every 1000th generation with a burnin of 3000 sampled generations. The default settings from MrBayes 3.0b4 were used, except that lset nst = 2 and prset rater = variable.
Patterns of historical gene flow were inferred using nested clade analysis (NCA). NCA evaluates the null hypothesis of no geographic association among haplotypes by testing the significance of three values: clade distance ($D_c$), a measure of the geographic range of a clade, nested clade distance ($D_n$), a measure of the geographic distribution of a clade relative to its sister clades, and the difference of these values between interior and tip clades (I-T; Templeton 1998). The relationship among significant values within a clade can then be used to infer restricted gene flow, population fragmentation, or population expansion for haplotypes within that clade. A haplotype network was created with TCS 1.18 (Clement et al. 2000) and nested according to the rules established by Templeton et al. (1987) and Templeton and Sing (1993). NCA was implemented in GEODIS 2.2 (Posada et al. 2000) with 10000 random permutations. The inference key of Templeton (1998) was used to infer patterns of historical gene flow from the analysis.

A substitution rate of 1.6-1.9% per million years (Fleischer et al. 1998) was used to estimate the divergence times of haplotypes between pygmy and brown-headed nuthatches and among pygmy nuthatch populations. Because
pygmy nuthatch populations are not reciprocally monophyletic; these estimates do not correspond to the divergence times of geographic populations. These estimates do, however, provide an approximation of the maximum divergence time among haplotypes as well as a threshold value for divergence of populations.

Subspecies Limits

The mitochondrial DNA sequence data was used to examine the correlation of genetic structure with morphological subspecies described by both Grinnell and Miller (1944) and Norris (1958). Grinnell and Miller (1944) supported a subspecific break between *S. p. melanotis* and *S. p. leuconucha* at the Banning Pass between the San Bernardino Mountains and the San Jacinto Mountains based on wing length, bill size, and plumage coloration (Figure 4). Norris (1958) supported a break between these two subspecies immediately south of the San Jacinto and Santa Rosa Mountains based primarily on the ratio of bill width to bill depth, but also on wing, tail, bill, and tarsus lengths. Individuals were partitioned by subspecies and an AMOVA was conducted across the subspecific break supported by each author using Arlequin 2.0 (Schneider et
Figure 4. *Sitta pygmaea* Subspecific Break Map. Southern California distribution and hypothesized subspecific breaks between *S. p. melanotis* and *S. p. leuconucha* and their source: 1-immediately south of the San Jacinto and Santa Rosa Mountains (Norris 1958), 2-Banning Pass (Grinnell and Miller 1944), 3-Cajon Pass (this study), 4-immediately north of the San Gabriel Mountains (this study). Map adapted from Grinnell and Miller (Grinnell, J., and A.H. Miller. 1944. The Distribution of the Birds of California. Cooper Ornithological Club. Pacific Coast Avifauna: Number 27. pp. 318-321.).

Observation of the distribution of haplotypes suggested that a stronger genetic subspecific break may occur further north than proposed by Grinnell and Miller (1944) or Norris (1958). Therefore, an AMOVA was also
conducted for each of two additional subspecific breaks: at the Cajon Pass between the San Bernardino and San Gabriel Mountains, and immediately north of the San Gabriel Mountains.
CHAPTER THREE

RESULTS

Sequence Analysis

Analysis of the 3045-base pair genetic marker from 23 pygmy nuthatch specimens in five populations yielded 11 haplotypes (Table 2). Base frequencies for the total sequence were $A=0.304$, $T=0.252$, $C=0.318$, and $G=0.126$. There were 22 polymorphic sites: 8 in cytochrome $b$, 4 in ND6, and 10 in the control region. All coding region substitutions were synonymous and, with the exception of one 1st-position substitution in cyt $b$, occurred at 3rd positions. The average uncorrected pairwise distance among all haplotypes was approximately 0.0020, substantially lower than values reported for other bird species. The overall transition-transversion ratio was 22.0, with 8, 4, and 9 transitions occurring in cyt $b$, ND6, and the control region respectively, and 1 transversion occurring in the control region. Base frequency and substitution data for the total sequence and for individual gene regions are summarized in Table 3.
Table 2. Variable Sites within 3045-Base Pair *Sitta pygmaea* Haplotypes.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>n</th>
<th>Cyt b</th>
<th>ND6</th>
<th>Control Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>6</td>
<td>G</td>
<td>C</td>
<td>A A T T T G A C C C</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>. . . . . . T</td>
</tr>
<tr>
<td>B1</td>
<td>3</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>D1</td>
<td>3</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>E1</td>
<td>1</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>H1</td>
<td>1</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>H2</td>
<td>1</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>H3</td>
<td>2</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>I1</td>
<td>1</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>J1</td>
<td>3</td>
<td>G</td>
<td>C</td>
<td>A A T T G A C C C</td>
</tr>
<tr>
<td>L1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>. . . . . . T</td>
</tr>
</tbody>
</table>

*S. pygmaea* haplotype frequency and variable site data for 3045-bp sequences. Numbers above columns indicate nucleotide position relative to 3045-bp sequence; dots indicate identity with haplotype A1.
Table 3. Gene Statistics for 3045-Base Pair Haplotypes.

<table>
<thead>
<tr>
<th>Base frequencies</th>
<th>Overall (3045 bp)</th>
<th>Cyt b (partial; 1045 bp)</th>
<th>ND6 (partial; 519 bp)</th>
<th>Control Region (1248 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.304</td>
<td>0.273</td>
<td>0.397</td>
<td>0.286</td>
</tr>
<tr>
<td>T</td>
<td>0.252</td>
<td>0.251</td>
<td>0.122</td>
<td>0.304</td>
</tr>
<tr>
<td>C</td>
<td>0.318</td>
<td>0.340</td>
<td>0.394</td>
<td>0.276</td>
</tr>
<tr>
<td>G</td>
<td>0.126</td>
<td>0.136</td>
<td>0.087</td>
<td>0.134</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Substitutions</th>
<th>1st pos</th>
<th>2nd pos</th>
<th>3rd pos</th>
<th>non-coding</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>8</td>
<td>4</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>10.0</td>
<td>1.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Base frequencies are presented for the light strand of the overall sequence and the individual gene regions. Substitution data are provided for the light strand of cyt b and the control region, and the heavy strand of ND6.

The partition homogeneity test found no significant differences among the three gene fragments, therefore they were combined into a single sequence for analysis.

Analysis of the 1639-base pair concatenated sequence from 75 pygmy nuthatch specimens in seven populations yielded 17 haplotypes (Table 4), and results were similar overall to the 3045-base pair sequence. Base frequencies for the concatenated sequence were A=0.306, T=0.228, C=0.337, and G=0.129. There were 26 polymorphic sites: 12 in cytochrome b and 7 each in ND6 and the control region. In cyt b seven 3rd and two 1st-position coding region
Table 4. Variable Sites within 1639-Base Pair *Sitta pygmaea* Haplotypes.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Cyt b</th>
<th>ND6</th>
<th>Control Region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>13</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>G</td>
<td>2</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>H</td>
<td>14</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>J</td>
<td>4</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>K</td>
<td>2</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>L</td>
<td>3</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>O</td>
<td>1</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>P</td>
<td>2</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>Q</td>
<td>1</td>
<td>G</td>
<td>A</td>
</tr>
</tbody>
</table>

S. *pygmaea* haplotype frequency and variable site data for 1639-bp sequences. Numbers above columns indicate nucleotide position relative to 1639-bp sequence; dots indicate identity with haplotype A.
substitutions were synonymous, while one 2nd and two 1st-position substitutions were non-synonymous. All ND6 substitutions occurred at the 3rd position and were synonymous. The overall transition-transversion ratio was 26.02, with 12, 7, and 6 transitions occurring in cyt b, ND6, and the control region respectively, and 1 transversion occurring in the control region. Base frequency and substitution data for the concatenated sequence and for individual gene regions are summarized in Table 5. The average uncorrected pairwise (p) distance among all haplotypes was approximately 0.0035, with a

Table 5. Gene Statistics for 1639-Base Pair Haplotypes.

<table>
<thead>
<tr>
<th></th>
<th>Overall (1639 bp)</th>
<th>Cyt b (621 bp)</th>
<th>ND6 (504 bp)</th>
<th>Control Region (514 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base frequencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.306</td>
<td>0.276</td>
<td>0.401</td>
<td>0.248</td>
</tr>
<tr>
<td>T</td>
<td>0.228</td>
<td>0.249</td>
<td>0.125</td>
<td>0.301</td>
</tr>
<tr>
<td>C</td>
<td>0.337</td>
<td>0.343</td>
<td>0.393</td>
<td>0.277</td>
</tr>
<tr>
<td>G</td>
<td>0.129</td>
<td>0.131</td>
<td>0.081</td>
<td>0.174</td>
</tr>
<tr>
<td>Substitutions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st pos</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd pos</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd pos</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>non-coding</td>
<td>7</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>12</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Ti</td>
<td>25</td>
<td>12</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Tv</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Ti/Tv</td>
<td>26.0</td>
<td></td>
<td></td>
<td>7.0</td>
</tr>
</tbody>
</table>

Base frequencies are presented for the light strand of the overall sequence and the individual gene regions. Substitution data are provided for the light strand of cyt b and the control region, and the heavy strand of ND6.
maximum p-distance of 0.0049. The average p-distance between \( S. \) pygmaea and \( S. \) pusilla haplotypes was 0.0702.

**Population Genetics**

Population genetic analyses were used to infer the levels of genetic differentiation within and among pygmy nuthatch populations in southern California. The number of haplotypes, percentage of variable sites, nucleotide diversity (\( \Pi \)), and Fu’s \( F_s \) were calculated as measures of population variability and are summarized in Table 6. No values for Fu’s \( F_s \) were significant. Results from the overall AMOVA analysis indicated significant, although relatively low levels of genetic differentiation, reporting \( \Phi_{ST}=0.139 \) (\( p<0.05 \)) with 13.9% of the total variation

<table>
<thead>
<tr>
<th>Mountain range</th>
<th>n</th>
<th>Number of haplotypes</th>
<th>% variable sites</th>
<th>Nucleotide diversity (( \Pi ))</th>
<th>Fu’s ( F_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piute Mountain</td>
<td>6</td>
<td>3</td>
<td>0.18%</td>
<td>0.0009</td>
<td>0.38</td>
</tr>
<tr>
<td>Cerro Noroeste</td>
<td>7</td>
<td>4</td>
<td>0.61%</td>
<td>0.0022</td>
<td>1.04</td>
</tr>
<tr>
<td>Frazier Mountain</td>
<td>7</td>
<td>3</td>
<td>0.55%</td>
<td>0.0028</td>
<td>3.27</td>
</tr>
<tr>
<td>San Gabriel Mountains</td>
<td>14</td>
<td>7</td>
<td>0.92%</td>
<td>0.0027</td>
<td>0.20</td>
</tr>
<tr>
<td>San Bernardino Mountains</td>
<td>21</td>
<td>11</td>
<td>1.16%</td>
<td>0.0031</td>
<td>-1.48</td>
</tr>
<tr>
<td>San Jacinto Mountains</td>
<td>13</td>
<td>6</td>
<td>0.67%</td>
<td>0.0023</td>
<td>0.54</td>
</tr>
<tr>
<td>Laguna Mountain</td>
<td>7</td>
<td>5</td>
<td>0.79%</td>
<td>0.0026</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

Statistics are based on the 1639-base pair haplotype data set. No values of Fu’s \( F_s \) were significant.
occurring among mountain ranges (Table 7). Differences within populations accounted for 84.1% of the variance. A Mantel test found a significant correlation (0.462, p=0.007) between population pairwise $F_{ST}$ values and geographic distances (Appendix B), suggesting isolation by distance among pygmy nuthatch collecting locations.

### Table 7. $F_{ST}$ Values for Analysis of Molecular Variance Comparisons.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$F_{ST}$ Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among all mountain ranges</td>
<td>0.139</td>
<td>0.0005</td>
</tr>
<tr>
<td>Between Laguna Mountains and all individuals to the north</td>
<td>0.102</td>
<td>0.0517</td>
</tr>
<tr>
<td>Between individuals north and south of Banning Pass</td>
<td>0.140</td>
<td>0.0010</td>
</tr>
<tr>
<td>Between individuals north and south of Cajon Pass</td>
<td>0.205</td>
<td>0.0001</td>
</tr>
<tr>
<td>Between individuals north and south of San Gabriel Mountains</td>
<td>0.093</td>
<td>0.0104</td>
</tr>
</tbody>
</table>

$F_{ST}$ values and their significance for overall AMOVA and AMOVAs testing subspecies breaks at various locations. Significant $F_{ST}$s are shown in bold.

### Phylogeography

Phylogenetic relationships were evaluated using parsimony, distance, and maximum likelihood-based methods. Consensus trees generated by all three methods were congruent (Figures 5-7), and although each showed well-
supported clades, these clades showed no strong geographic structure. Parsimony analysis using a branch and bound search produced three equally likely minimum length trees differing only in the placement of haplotype N (distributed in the San Bernardino and Laguna Mountains; Table 8). The bootstrap consensus tree (Figure 5), which used a full heuristic search and was congruent with the branch-and-bound analysis, provided strong support for two separate clades containing haplotypes F, I and O (87%; distributed in the San Gabriel and San Bernardino Mountains), and D, H, J, K, and Q (91%; distributed throughout all populations). The consensus tree also weakly supported a clade containing haplotypes C, G, and P (55%; distributed throughout all populations except Piute Mountain). These three clades and the remaining 6 haplotypes formed a basal polytomy with respect to the outgroup.

The bootstrap consensus of the neighbor-joining analysis produced a similar tree to the bootstrapped consensus parsimony tree, supporting clades F-I-O and D-H-J-K-Q at 90% each, and clade C-G-P at 70% (Figure 6). The neighbor-joining analysis also provided weak to moderate
Figure 5. Bootstrapped Consensus Parsimony Tree. The consensus parsimony tree shows strong to moderate bootstrap support for clades C-G-P, D-H-J-K-Q, and F-I-O. There is no strong geographic structure among these clades (Table 8).
Figure 6. Bootstrapped Consensus Neighbor-Joining Tree. Like the parsimony analysis, the consensus neighbor-joining tree shows strong bootstrap support for clades C-G-P, D-H-J-K-Q, and F-I-O. There is no strong geographic structure among these clades (Table 8).
Figure 7. Consensus Bayesian Tree.
The Bayesian tree shows strong support for clades D-H-J-K-Q, C-G-P, A-B-E-L-M-CGP, and N-DHJKQ-ABELMCGP. The Bayesian analysis supports more structure than the parsimony and neighbor-joining analyses, however, lacks the support for clade F-I-O shown by these two analyses. Again there is no strong geographic structure among these clades (Table 8).
Table 8. Distribution of *Sitta pygmaea* Haplotypes and Phylogenetic Clades.

| Subspecies          | Mountain Range | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---------------------|----------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *S. p. melanotis*   | Piute          | 2 | 3 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                     | Cerro Noroeste | 3 | 2 | 1 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                     | Frazier        | 3 | 3 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                     | San Gabriel    | 5 | 1 | 1 | 3 | 1 | 2 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |
|                     | San Bernardino | 1 | 2 | 1 | 5 | 1 | 3 | 1 | 3 | 1 | 2 | 1 |   |   |   |   |   |   |   |   |
| *S. p. leuconucha*  | San Jacinto    | 3 | 2 | 5 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                     | Laguna         | 1 | 3 | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                     |                | 13| 4 | 9 | 8 | 5| 2 | 2 | 14| 1 | 4 | 2 | 3 | 1 | 3 | 1 | 2 | 1 |   |

<table>
<thead>
<tr>
<th>Represented subspecies</th>
<th>ml</th>
<th>m</th>
<th>m</th>
<th>ml</th>
<th>m</th>
<th>m</th>
<th>m</th>
<th>ml</th>
<th>ml</th>
<th>ml</th>
<th>ml</th>
<th>m</th>
<th>m</th>
<th>ml</th>
<th>m</th>
<th>l</th>
<th>l</th>
</tr>
</thead>
</table>

Upper portion of table shows the distribution of *Sitta pygmaea* haplotypes with number of individuals of each haplotype occurring in each population. Represented subspecies indicates the subspecies represented by each haplotype: m=melanotis, l=leuconucha, and ml=melanotis and leuconucha. Example: haplotype A contains 2 individuals sampled from Piute Mountain and is represented by both *S. p. melanotis* and *S. p. leuconucha*. Lower portion of table shows distribution of *Sitta pygmaea* phylogenetic clades with number of individuals of each clade occurring in each population. Example: 5 individuals within clade A-B-E-L-M were sampled from Piute Mountain. Subspecies designations are based on Grinnell and Miller (1944).
support for groupings within these three clades. Haplotypes F and I (distributed in the San Gabriel and San Bernardino Mountains) were grouped together with 76% bootstrap support, and H, J, K and Q (distributed in the San Gabriel, San Bernardino, San Jacinto and Laguna Mountains) were grouped with 74% bootstrap support. As in the parsimony consensus tree, the three main clades were arranged in a basal polytomy with the remaining 6 haplotypes with respect to the outgroup.

Bayesian analysis produced a consensus tree similar overall to parsimony and distance-based methods, but with more resolution of the interior nodes (Figure 7). Haplotypes F, I, and O and a clade containing the remaining 14 haplotypes (83% posterior probability) formed a basal polytomy with respect to the outgroup. Within the clade containing the 14 haplotypes, haplotype N stood alone, the D-H-J-K-Q clade was supported at 100%, and a second clade including haplotypes A, B, E, L, M, C, G, and P (distributed throughout all populations) was supported at 84%, with haplotypes C, G, and P comprising an additional group (91% posterior probability) within this clade.

Phylogeographic relationships were also analyzed using nested clade analysis. TCS produced a single haplotype
network with no ambiguities, inferring 11 unsampled haplotypes (Figure 8). The haplotype network was grouped into seven 1-step clades, four 2-step clades, and one 3-step clade (the total cladogram; Figure 9). Nested clade analysis yielded 4 clades with significant values: 1-5, 1-7, 2-3, and 3-1. The pattern of gene flow in Clade 1-5 was determined to be inconclusive (inference key 1-2-11-17no). Within Clade 1-7, $D_c$ and $D_n$ values for haplotype D were significantly large, $D_c$ and $D_n$ values for haplotype H were

![Figure 8. Minimum Spanning Haplotype Network. Each line represents a single mutational step; circle size is relative to number of individuals with each haplotype; empty circles represent unsampled haplotypes. See Table 8 for geographic distribution of haplotypes.](image)
Figure 9. Nested Cladogram.
Nested cladogram showing the successive grouping of sister clades (haplotypes). Nested clade analysis uses these clades in conjunction with geographic information to infer evolutionary history of the clades.

significantly small, and the $D_n$ value for I-T was significantly small, leading to an inference of contiguous range expansion (inference key 1-2-11-12no). Within Clade 2-3, $D_n$ values for Clade 1-6 and I-T were significantly large, leading to an inference of restricted gene flow with isolation by distance (inference key 1-2-11-17-4no). Clade 3-1 also suggested restricted gene flow with isolation by distance (inference key 1-2-3-4no) based on a significantly small $D_c$ value for Clade 2-4.
Based on a substitution rate of 1.6-1.9% (Fleischer et al. 1998), and a mean uncorrected pairwise genetic distance of 7.02% between *S. pygmaea* and *S. pusilla* haplotypes, pygmy nuthatches diverged from brown-headed nuthatches approximately 4.4-3.7 million years ago. The maximum uncorrected pairwise genetic distance among pygmy nuthatch haplotypes (several pairings) is 0.49%. Using the same substitution rate these haplotypes diverged approximately 250,000 to 300,000 years ago.

Subspecies Limits

The results of the AMOVA for the hypothesized subspecific breaks are summarized in Table 7. The AMOVA analysis provided no significant genetic support for Norris' hypothesis of a subspecific break immediately south of the Santa Rosa Mountains (Figure 4). The AMOVA analysis did provide significant genetic support for Grinnell and Miller's hypothesis of a subspecific break at the Banning Pass. The $\Phi_{ST}=0.140$ indicates that 14% of the genetic differentiation among pygmy nuthatches can be accounted by differences between individuals north and south of the Banning Pass. This is similar to the value among mountain ranges irrespective of subspecies. I tested two additional
hypotheses based on my observation of the haplotype distribution data (Table 8). Both AMOVA analyses yielded significant results. An $\Phi_{ST}=0.205$ was reported for the AMOVA across the Cajon Pass, indicating that over 20% of the genetic differentiation among pygmy nuthatches can be accounted by differences between individuals north and south of Cajon Pass. This was the highest $\Phi_{ST}$ reported among the four AMOVAs conducted at hypothesized subspecific breaks. A significant $\Phi_{ST}=0.093$ was reported for the subspecific break immediately north of the San Gabriel Mountains.
Overall the analyses reveal that pygmy nuthatch populations in southern California are currently stable with restricted gene flow among populations. There is significant genetic differentiation of haplotypes among populations and a significant association between linear distance and population subdivision. There is, however, no strong geographic structure of haplotypes among populations. Considered together, this suggests that: 1) historical gene flow was extensive, 2) populations were isolated relatively recently, and 3) the current distribution of pygmy nuthatches is the result of vicariance rather than dispersal. Analyses of the subspecific break between *S. p. melanotis* and *S. p. leuconucha* indicate that the strongest genetic difference occurs at the Cajon Pass, which is further north than the hypothesized subspecific breaks supported by morphological characters.
Sequence Analysis

Analysis of both the 3045- and 1639-base pair sequences revealed low levels of genetic variation within southern California populations of the pygmy nuthatch. Overall genetic diversity was low, as was the degree of genetic differentiation among populations and subspecies (Table 7). Nucleotide composition and substitution data of both the 3045 and 1639-base pair sequences (Tables 3 and 5) are generally consistent with values reported for other birds at the intraspecific level (Baker and Marshall 1997, Moore and DeFilippis 1997). Specifically the low G content is characteristic of D-loop in birds (Baker and Marshall 1997), and the high proportion of 3\textsuperscript{rd} position, synonymous substitutions within coding regions is consistent with recently diverged birds (Moore and DeFilippis 1997). The average pairwise genetic distance among haplotypes (0.0035) is also similar to values found in other passerines (Buerkle 1999, Sgariglia and Burns 2003, Burns and Barhoum 2006).

Population Genetics

Results of the population genetic analyses revealed a low but significant level of genetic differentiation among
populations (Table 7). Specifically, the AMOVA indicated that 13.9% of the genetic variance is due to differences among rather than within mountain ranges. There is also a significant isolation by distance effect among collecting locations. These results suggest that there has been extensive historical gene flow, but that this gene flow has been restricted in the recent past. The lack of lineage sorting from the phylogenetic analyses also suggest that gene flow was only recently restricted.

These results are consistent with the historical distribution of the pygmy nuthatch’s coniferous forest habitat. Studies in the sky islands of Arizona (Masta 2000) and the Rocky Mountains (DeChaine and Martin 2004) have shown that habitats moved southward and to lower elevations during glacial periods. These results suggest the coniferous forests in California would have been more contiguous and widely distributed at lower elevations during the last glacial period approximately 100,000 to 10,000 years ago, providing greater opportunity for migration among pygmy nuthatch populations at that time. During the present interglacial period, coniferous forests have contracted and elevated, becoming isolated. This study suggests that this, combined with the sedentary
nature of pygmy nuthatches, has resulted in the recent restriction of gene flow among mountain ranges.

Estimates of nucleotide diversity (H) were similar to values reported in other intraspecific studies of passerines, while the number of haplotypes [per number of individuals] per population was generally lower (Table 6; Buerkle 1999, Sgariglia and Burns 2003). Values of Fu’s Fs for all populations were not significant (Table 6), indicating that population sizes are currently stable and that no populations have undergone a recent range expansion. These results further support the recent isolation of pygmy nuthatch populations and the role of vicariance in the separation of populations. Pygmy nuthatches have remained relatively sedentary during periods of interglacial isolation while gene flow during glacial periods was facilitated by proximity.

Phylogeography

Parsimony, neighbor-joining, and maximum likelihood analyses revealed strong to moderate support for four distinct haplotype clades: F-I-O, D-H-J-K-Q, A-B-E-L-M-C-G-P, and C-G-P by itself (Figures 5-7). Of these, only the F-I-O clade shows any association with geography.
Haplotypes F, I, and O are restricted to the San Gabriel and San Bernardino Mountains within the Transverse Ranges of southern California (Table 8). The remaining clades in the phylogenetic analysis are widespread in their distribution and occur in all or all but one of the sampled mountain ranges (Table 8).

Nested clade analysis revealed the overall pattern of historical gene flow in southern California pygmy nuthatch populations to be influenced by both past contiguous range expansion and restriction of gene flow with isolation by distance. Clade 1-7 (D-H-J-K-Q), which is represented in all but the Piute Mountains, was inferred to have arisen through contiguous range expansion. The majority of individuals represented in this clade are distributed in the San Bernardino and San Jacinto Mountains (Table 8), with fewer individuals represented to the north and south. This suggests that these may have been concentrated in a glacial refugium in the vicinity of the San Bernardino and San Jacinto Mountains and then expanded northward as glaciers retreated during interglacial episodes. This pattern of haplotype distribution, however, may be an artifact of the sampling design with the San Bernardino and San Jacinto Mountains being two of the most heavily sampled
ranges. Clades 2-3 (A-B-E-L-M-C-G-P; represented in all populations except Laguna Mountain) and 3-1 (entire cladogram) were inferred to be the result of restricted gene flow with isolation by distance. With the exception of a few widespread haplotypes (A, D, and H), all of the haplotypes are geographically restricted either from the San Bernardino Mountains to the south (haplotypes I-Q) or from the San Bernardino Mountains to the north (haplotypes B-C and E-G). This suggests that pygmy nuthatches have experienced significant but local gene flow during past glacial episodes as habitats became connected. The Mantel test also showed a significant isolation by distance effect among pygmy nuthatch collecting locations.

The maximum divergence time between southern California pygmy nuthatch haplotypes based on a 1.6-1.9% substitution rate (Fleischer et al. 1998) is approximately 250,000 to 300,000 years ago. While geographic populations are not reciprocally monophyletic, this estimate provides a maximum threshold value for the restriction of gene flow among populations, providing further evidence for the recent isolation of pygmy nuthatch populations. Thus southern California populations have become isolated on
disjunct mountain ranges well within the last 300,000 years.

Due to the sedentary nature of pygmy nuthatches it is likely that even though their habitat was widespread during glacial epochs, dispersal was still limited and migration occurred only among adjacent populations. Current populations on isolated mountain ranges exchanged genes with populations on adjacent isolated mountain ranges during glacial periods when their habitat was connected at lower elevations.

Subspecies Limits

Pygmy nuthatch populations in southern California exhibit clinal variation in morphology: mean size increases and plumage coloration becomes paler from north to south (Kingery and Ghalambor 2001). This clinal variation encompasses two described subspecies, but exactly where this subspecific break occurs is uncertain (Figure 4; Grinnell and Miller 1944, Norris 1958). Analysis of the distribution of sequence variation using AMOVA supports a significant break within the Transverse Ranges at the Cajon Pass between the San Gabriel and San Bernardino Mountains (Table 7). Twenty percent of the genetic differentiation
can be accounted by individuals north and south of the Cajon Pass. By comparison, the mtDNA data accounted for only 14% of the genetic variance across Banning Pass between the San Bernardino Mountains (Transverse Ranges) and the San Jacinto Mountains (Peninsular Ranges), the subspecific break supported by Grinnell and Miller (1944) based on morphological characters. This is similar to the level of variation (13.9%) explained by the overall AMOVA analyzing differences among all mountains ranges. Our mtDNA data revealed no statistically significant genetic differentiation between individuals in the Laguna Mountains and those to the north, providing no genetic support for Norris’ (1958) proposed subspecific break.

Recent studies on comparative phylogeography of California fauna provide support for a subspecific break at the Cajon Pass, at least in an evolutionary perspective. A comprehensive review of phylogeographic studies of California fauna found that most species exhibit a distinct genetic break near (north of) the Transverse Ranges (Figure 4), although this pattern was not as apparent in birds (Calsbeek et al. 2003). More recently studies by Sgariglia and Burns (2003), Burns and Barhoum (2006), and Alexander and Burns (unpublished) have provided additional
phylogeographic information on California's avifauna supporting this genetic break near the Transverse Ranges.

In a study of the phylogeography of the wrentit (Chamaea fasciata), Burns and Barhoum (2006) found a strong genetic break among populations at the Cajon Pass, an area with no strong morphological variation and which did not conform to a described morphological subspecies break. This is nearly identical to the pattern found in this pygmy nuthatch study, except wrentit populations south of the Cajon Pass form a monophyletic grouping whereas pygmy nuthatch populations do not. The reason for the lack of monophyly observed in pygmy nuthatches may be differences in habitat during Pleistocene climate changes. The pygmy nuthatch is a resident of coniferous forest while the wrentit inhabits coastal sage scrub and chaparral. Both habitats probably experienced similar down slope and southward shifts during glacial periods, however, because coastal sage scrub and chaparral occur at lower elevation, they probably were more isolated at these times restricting gene flow among wrentit populations (Burns and Barhoum 2006). Alternatively, coniferous forests were probably more widespread and contiguous during glacial periods.
facilitating opportunities for gene flow among adjacent pygmy nuthatch populations.

A study on the white-headed woodpecker (*Melanerpes albolarvatus*), with which the pygmy nuthatch is co-distributed in Pacific North America, also found a nearly monophyletic grouping of populations from Mount Pinos south through the Peninsular Ranges (Alexander and Burns unpublished). In this case, however, the authors postulated that the southern populations were colonized by a southward expansion from the north, and that monophyly was the result of a founder event or bottleneck as opposed to vicariance caused by the occurrence of a southern glacial refugium for wrentits (Alexander and Burns unpublished, Barhoum and Burns 2006). Thus while similar patterns of genetic differentiation are observed among many co-distributed taxa, the proposed evolutionary processes by which these patterns occur are different for different species.

This research supports three significant genetic breaks of varying degree at the northern edge of the San Gabriel Mountains, at the Cajon Pass between the San Gabriel and San Bernardino Mountains, and at the Banning Pass at the northern edge of the San Jacinto Mountains.
The strongest genetic break occurs at the Cajon pass, in contrast to the strongest morphological break observed by Grinnell and Miller (1944) occurring at the Banning Pass. That the subspecific breaks supported by mitochondrial DNA and morphological data do not correspond is not unusual or unprecedented. Genetically-based morphological traits may represent adaptations to local conditions, and would be subject to local environmental selective pressures that would not affect neutral mtDNA markers (Burns and Barhoum 2006). In a review of avian mitochondrial DNA studies, Zink (2004) found that 97% of continental North American bird subspecies lack a corresponding significant population genetic structure.

While a subspecific break at the Cajon Pass is supported by our mtDNA data and evidence from other species, caution is warranted. Much debate has occurred over the concept of subspecies and the criteria for diagnosing them (Patten and Unitt 2002). Recently Rising (2005) has argued that subspecies should be distinctive, geographically isolated populations. While southern California populations meet the criteria of being geographically isolated, these populations are more clinal and less distinctive. This study encompasses only a small
portion of the range of *S. p. melanotis*, which is widely distributed throughout western North America, and does not include samples from all of the populations representing *S. p. leuconucha*, particularly those in northern Baja California Norte, Mexico, and several populations located between the Laguna and San Jacinto Mountains in southern California. Therefore one must also consider that all of these populations may comprise a single subspecies. Even Grinnell and Miller (1944) recognized the cline in morphological variation and considered the placement of the subspecific break between *S. p. melanotis* and *S. p. leuconucha* "arbitrary".

**Conservation Implications**

The population and phylogenetic data generated from this research also has important application to the management and conservation of land in southern California. Long term habitat loss and fragmentation have made habitat conservation and connectivity a priority for land management in southern California. The South Coast Missing Linkages Project has recently identified the pygmy nuthatch as an important indicator for identifying and preserving dispersal corridors among large habitat preserves on public
and private lands (South Coast Wildlands 2005). The results of this research, however, suggest that pygmy nuthatches are a poor choice as indicators of corridor usage. The current restriction of gene flow among mountain ranges suggests that migration and dispersal among mountain ranges is a rare event. Due to their sedentary nature and habitat affinity, pygmy nuthatches would be better suited as indicators of habitat quality for coniferous forests.

This research provides baseline genetic data which could be used to monitor changes in genetic composition of pygmy nuthatch populations over time, from which inferences about the health and quality of their coniferous forest habitat could be made. For example, if at some time in the future a population of pygmy nuthatches exhibited reduced genetic diversity relative to this baseline data, this could indicate a reduction in population size corresponding to a decline in the quality or availability of suitable habitat.

Conclusions

Pygmy nuthatch populations in southern California exhibit relatively low levels of genetic variation and differentiation. During glacial periods greater than about 10,000 years ago, coniferous forests were more widespread
and contiguous (Pielou 1991, Dorf 1976), allowing for extensive gene flow among populations. The current wide distribution of haplotypes and lack of lineage sorting in the phylogenetic analysis is the result of this historical gene flow. Within the last 10,000 years during the current interglacial period, coniferous forests have become isolated in disjunct mountain ranges. As a result pygmy nuthatch populations have undergone a recent restriction of gene flow, and show significant (although low) levels of genetic differentiation among mountain ranges. This genetic differentiation is greatest at the Cajon Pass, which may also represent a subspecific break between S. p. melanotis and S. p. leuconucha. If so, this subspecific break would occur further north than previously thought, and would not conform to proposed morphologically-based subspecific breaks.
APPENDIX A

HAPLOTYPE LISTS
<table>
<thead>
<tr>
<th>A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTTTGATCACTCCTAGGTATCTGTTAGTCACCCAAATCTGTAACAGGCCTGCTACTAGCAACACACTACA CGCAAGACCTCTGCTTGGCCTAGTCTGCTTAGTCACCCAAATCGTAACAGGCCTGCTACTAGCAACACACTACA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTTTGATCACTCCTAGGTATCTGTTAGTCACCCAAATCTGTAACAGGCCTGCTACTAGCAACACACTACA CGCAAGACCTCTGCTTGGCCTAGTCTGCTTAGTCACCCAAATCGTAACAGGCCTGCTACTAGCAACACACTACA</td>
</tr>
</tbody>
</table>
1639-BASE PAIR HAPLOTYPE LIST

A

GTAGGAGTTATTTCTCCTACTAACCCTCAATAGCAACTGCTTTCGTAGGATACGTTCTACCATGAGGACAAAT
ATCCTCTCGAGGGCTACAGCTCATCAACCTATTTCTCCGCAATCCCTCCTATCAGCGGCAAACACATAGTAG
ATATGGCCCTGATGACCTTGCATGCAAACCATCCTATTTCTCCCTTGGGTTACCGGCGACTTCTTACTT
CCATTCGTCATCGCAGGCATTACATTAGTTCACCTCACCTTCCTACACGAAACAGGCTCAAACAACCCCCT
AGGAATCCCTCAGACTGTCAGAACAAATTTCTCAACTCCCCACCTACCCACATCAAAACCAATGATCTGGTCTC
CTACATGCTTATTCTCCTAGTCACTCTAGCACTATTCTCCCCAAACCTACTAGGAGACCCAGAAAACTTC
ACACCAGCTAACCCCCTAGTCACACCCCCACACATCAAACCCGAATGATATTTCCTATTCGCATACGCTAT
TCTCCGATCCATCCCAAACAAACTAGGAGGAGTACTAGCCCTTGCCGCCTCCGTCCTAGTCTTATTCCTTA
TGCCCTTACTACACACATCCAAACAACGCTCAAACAACGCTCAAACAACGCTCAAACAACGCTCAAACAACGCT
TTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
TGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
CCAGAAGCCGCAAGCGCTAGCTGATCTCCCTGTGCATACCATCCGGGAGTGCTGTGCTTTGCTTCCTACAGTGAC

B

GTAGGAGTTATTTCTCCTACTAACCCTCAATAGCAACTGCTTTCGTAGGATACGTTCTACCATGAGGACAAAT
ATCCTCTCGAGGGCTACAGCTCATCAACCTATTTCTCCGCAATCCCTCCTATCAGCGGCAAACACATAGTAG
ATATGGCCCTGATGACCTTGCATGCAAACCATCCTATTTCTCCCTTGGGTTACCGGCGACTTCTTACTT
CCATTCGTCATCGCAGGCATTACATTAGTTCACCTCACCTTCCTACACGAAACAGGCTCAAACAACCCCCT
AGGAATCCCTCAGACTGTCAGAACAAATTTCTCAACTCCCCACCTACCCACATCAAAACCAATGATCTGGTCTC
CTACATGCTTATTCTCCTAGTCACTCTAGCACTATTCTCCCCAAACCTACTAGGAGACCCAGAAAACTTC
ACACCAGCTAACCCCCTAGTCACACCCCCACACATCAAACCCGAATGATATTTCCTATTCGCATACGCTAT
TCTCCGATCCATCCCAAACAAACTAGGAGGAGTACTAGCCCTTGCCGCCTCCGTCCTAGTCTTATTCCTTA
TGCCCTTACTACACACATCCAAACAACGCTCAAACAACGCTCAAACAACGCTCAAACAACGCTCAAACAACGCT
TTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
TGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG
CCAGAAGCCGCAAGCGCTAGCTGATCTCCCTGTGCATACCATCCGGGAGTGCTGTGCTTTGCTTCCTACAGTGAC
APPENDIX B

DISTANCE MATRICES
### DISTANCE MATRICES

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Piute Mountain</td>
<td>0.0457</td>
<td>0.2209</td>
<td>0.1413</td>
<td>0.1367</td>
<td>0.1229</td>
<td>*0.2270</td>
<td>*0.1777</td>
<td>0.0917</td>
<td>*0.1882</td>
<td>*0.2023</td>
<td></td>
</tr>
<tr>
<td>2 Cerro Noroeste</td>
<td>102.1</td>
<td>0.0427</td>
<td>0.0000</td>
<td>0.0225</td>
<td>0.1185</td>
<td>*0.1905</td>
<td>*0.1429</td>
<td>0.0000</td>
<td>0.1104</td>
<td>*0.1493</td>
<td></td>
</tr>
<tr>
<td>3 Frazier Mountain</td>
<td>92.0</td>
<td>23.4</td>
<td>0.0375</td>
<td>*0.3095</td>
<td>0.1318</td>
<td>*0.2381</td>
<td>*0.1905</td>
<td>*0.1377</td>
<td>0.1864</td>
<td>*0.1630</td>
<td></td>
</tr>
<tr>
<td>4 Winston Peak</td>
<td>128.2</td>
<td>129.4</td>
<td>106.2</td>
<td>*0.1704</td>
<td>0.1312</td>
<td>*0.1667</td>
<td>0.1007</td>
<td>0.0616</td>
<td>0.1290</td>
<td>0.1250</td>
<td></td>
</tr>
<tr>
<td>5 Blue Ridge</td>
<td>138.2</td>
<td>151.6</td>
<td>128.3</td>
<td>23.8</td>
<td>*0.2138</td>
<td>*0.2619</td>
<td>*0.1809</td>
<td>0.0128</td>
<td>0.1227</td>
<td>0.1884</td>
<td></td>
</tr>
<tr>
<td>6 Delamar Mountain</td>
<td>185.2</td>
<td>218.7</td>
<td>195.2</td>
<td>92.4</td>
<td>68.7</td>
<td>*0.1905</td>
<td>0.0233</td>
<td>0.0233</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>7 Onyx Peak</td>
<td>206.0</td>
<td>240.5</td>
<td>217.2</td>
<td>113.5</td>
<td>89.9</td>
<td>22.0</td>
<td>*0.1064</td>
<td>*0.1250</td>
<td>*0.1951</td>
<td>0.1312</td>
<td></td>
</tr>
<tr>
<td>8 Grinnell Mountain</td>
<td>206.5</td>
<td>237.1</td>
<td>213.8</td>
<td>109.0</td>
<td>85.6</td>
<td>21.0</td>
<td>9.2</td>
<td>0.0148</td>
<td>0.0040</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>9 Black Mountain</td>
<td>234.1</td>
<td>253.1</td>
<td>230.1</td>
<td>124.0</td>
<td>103.1</td>
<td>53.5</td>
<td>42.0</td>
<td>34.9</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>10 Thomas Mountain</td>
<td>256.2</td>
<td>270.2</td>
<td>247.6</td>
<td>142.1</td>
<td>122.9</td>
<td>77.8</td>
<td>65.3</td>
<td>58.9</td>
<td>24.4</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>11 Laguna Mountain</td>
<td>339.5</td>
<td>338.5</td>
<td>317.3</td>
<td>217.0</td>
<td>202.0</td>
<td>165.6</td>
<td>152.3</td>
<td>146.6</td>
<td>112.2</td>
<td>87.9</td>
<td></td>
</tr>
</tbody>
</table>

Values below the diagonal are pairwise geographic distances in kilometers for *Sitta pygmaea* collecting locations. Values above the diagonal are pairwise $F_{ST}$ values for *Sitta pygmaea* collecting locations. * indicates significance at $p=0.05$. Negative values were changed to zero.
APPENDIX C

INPUT FILES
ARLEQUIN INPUT FILE

[Profile]

Title="Sitta pygmaea AMONG MOUNTAIN RANGES AMOVA"
NbSamples= 7
DataType= DNA
GenotypicData= 0
LocusSeparator= NONE
GameticPhase= 0
RecessiveData= 0
RecessiveAllele= null
MissingData= '?'

[Data]

[[HaplotypeDefinition]]
HaplListName="Sitta pygmaea Haplotype List"
HaplList={
[Insert 1639-base pair haplotype list without haplotypes R, S and T here]
}

[[Samples]]

SampleName="PIUTE"
SampleSize=6
SampleData= {
  A  2
  B  3
  C  1
}

SampleName="NOROESTE"
SampleSize=7
SampleData= {
  A  3
  C  2
  D  1
  E  1
}

SampleName="FRAZIER"
SampleSize=7
SampleData= {
  C  3
  D  3
  E  1
}

SampleName="SAN GABRIEL"
SampleSize=14
SampleData= {

101
SampleName="SAN BERNARDINO"
SampleSize=21
SampleData= }
  B 1
  C 2
  D 1
  E 3
  F 1
  G 2
  H 1
}

SampleName="SAN JACINTO"
SampleSize=13
SampleData= }
  A 3
  D 2
  H 5
  K 1
  P 1
  Q 1
}

SampleName="LAGUNA"
SampleSize=7
SampleData= }
  D 1
  H 3
  J 1
  N 1
  P 1
}

[[Structure]]

StructureName = "AMONG MOUNTAIN RANGES"
NbGroups=1
Group={  "PIUTE"
[Profile]

Title="Sitta pygmaea subspecies AMOVA"
NbSamples= 2
DataType= DNA
GenotypicData= 0
LocusSeparator= NONE
GameticPhase= 0
RecessiveData= 0
RecessiveAllele= null
MissingData= '?'

[Data]

[[HaplotypeDefinition]]
HaplListName="Sitta pygmaea Haplotype List"
HaplList={
[Insert 1639-base pair haplotype list without haplotypes R, S and T here]
}

[[Samples]]

SampleName="melanotis"
SampleSize=68
SampleData= {
A  13
B  4
C  9
D  7
E  5
F  2
G  2
H 11
I  1
J  3
K  2
L  3
M  1
N  2
O  1
P  1
Q  1
}

103
SampleName="leuconucha"
SampleSize=7
SampleData= {
    D 1
    H 3
    J 1
    N 1
    P 1
}

[[Structure]]

StructureName = "between subspecies san jacinto=melanotis"
NbGroups=1
Group={
    "melanotis"
    "leuconucha"
}

[Profile]

Title="Sitta pygmaea subspecies AMOVA"
NbSamples= 2
DataType= DNA
GenotypicData= 0
LocusSeparator= NONE
GameticPhase= 0
RecessiveData= 0
RecessiveAllele= null
MissingData= '?'

[Data]

[[HaplotypeDefinition]]

HaplListName="Sitta pygmaea Haplotype List"
HaplList={
    [Insert 1639-base pair haplotype list without haplotypes R, S and T here]
}

[[Samples]]

SampleName="melanotis"
SampleSize=55
SampleData= {
    A 10
    B 4
    C 9
    D 5
    E 5
    F 2
}

104
SampleName="leuconucha"
SampleSize=20
SampleData={
  A 3
  D 3
  H 8
  J 1
  K 1
  N 1
  P 2
  Q 1
}

[[Structure]]

  StructureName = "between subspecies san jacinto=leuconucha"
  NbGroups=1
  Group={
    "melanotis"
    "leuconucha"
  }

[Profile]

  Title="Sitta pygmaea subspecies AMOVA"
  NbSamples= 2
  DataType= DNA
  GenotypicData= 0
  LocusSeparator= NONE
  GameticPhase= 0
  RecessiveData= 0
  RecessiveAllele= null
  MissingData= '?'

[Data]

[[HaplotypeDefinition]]
  HaplListName="Sitta pygmaea Haplotype List"
  HaplList=
  [Insert 1639-base pair haplotype list without haplotypes R, S and T here]
[[Samples]]

SampleName="melanotis"
SampleSize=34
SampleData= { A 10 B 3 C 7 D 5 E 5 F 1 G 2 H 1 }

SampleName="leuconucha"
SampleSize=41
SampleData = { A 3 B 1 C 2 D 3 F 1 H 13 I 1 J 4 K 2 L 3 M 1 N 3 O 1 P 2 Q 1 }

[[Structure]]

StructureName = "between subspecies san bernardino=leuconucha"
NbGroups=1
Group={
  "melanotis"
  "leuconucha"
}

[Profile]

Title="Sitta pygmaea AMOVA"
NbSamples=11
DataType= DNA
GenotypicData= 0
LocusSeparator= NONE
GameticPhase= 0
RecessiveData= 0
Recessive Allele= null
MissingData= '?'

[[Data]]

[[HaplotypeDefinition]]
  HaplListName="Sitta pygmaea Haplotype List"
  HaplList=
  [Insert 1639-base pair haplotype list without haplotypes R, S and T here]
}

[[Samples]]

  SampleName="PIUTE"
  SampleSize=6
  SampleData= { 
    A 2
    B 3
    C 1
  }

  SampleName="NOROESTE"
  SampleSize=7
  SampleData= { 
    A 3
    C 2
    D 1
    E 1
  }

  SampleName="FRAZIER"
  SampleSize=7
  SampleData= { 
    C 3
    D 3
    E 1
  }

  SampleName="WINSTON"
  SampleSize=7
  SampleData= { 
    A 1
    C 1
    D 1
    E 3
    F 1
  }

  SampleName="BLUE RIDGE"
  SampleSize=7
SampleData= { 
    A  4  
    G  2  
    H  1  
}

SampleName="DELAMAR"
SampleSize=7
SampleData= { 
    B  1  
    C  2  
    H  3  
    I  1  
}

SampleName="ONYX"
SampleSize=7
SampleData= { 
    J  2  
    K  1  
    L  3  
    M  1  
}

SampleName="GRINNELL"
SampleSize=7
SampleData= { 
    F  1  
    H  2  
    J  1  
    N  2  
    O  1  
}

SampleName="BLACK"
SampleSize=7
SampleData= { 
    A  2  
    D  1  
    H  2  
    K  1  
    P  1  
}

SampleName="THOMAS"
SampleSize=6
SampleData= { 
    A  1  
    D  1  
    H  3  
    Q  1  
}

SampleName="LAGUNA"
```plaintext
SampleSize=7
SampleData= {
  D 1
  H 3
  J 1
  N 1
  P 1
}

[[Mantel]]
MatrixSize=11
MatrixNumber=2
Ymatrix=Fst
YmatrixLabels= {
  "PIUTE"
  "NOROESTE"
  "FRAZIER"
  "WINSTON"
  "BLUE RIDGE"
  "DELAMAR"
  "ONYX"
  "GRINNELL"
  "BLACK"
  "THOMAS"
  "LAGUNA"
}
DistMatMantel= {
  0.00 102.1 0.00 92.0 23.4 0.00 128.2 129.4 106.2 0.00 138.2 151.6 128.3 23.8 0.00 185.2 218.7 195.2 92.4 68.7 0.00 206.0 240.5 217.2 113.5 89.9 22.0 0.00 206.5 237.1 213.8 109.0 85.6 21.0 9.2 0.00 234.1 253.1 230.1 124.0 103.1 53.5 42.0 34.9 0.00 256.2 270.2 247.6 142.1 122.9 77.8 65.3 58.9 24.4 0.00 339.5 338.5 317.3 217.0 202.0 165.6 152.3 146.6 112.2 87.9 0.00
}
UsedYMatrixLabels= {
  "PIUTE"
  "NOROESTE"
  "FRAZIER"
  "WINSTON"
  "BLUE RIDGE"
  "DELAMAR"
  "ONYX"
  "GRINNELL"
  "BLACK"
  "THOMAS"
  "LAGUNA"
}
```
MRBAYES INPUT FILE

#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=20 NCHAR=1639;
FORMAT DATATYPE=DNA MISSING=? GAP=-;

MATRIX
[Insert 1639-base pair haplotype list here]
;
END;

begin mrbayes;
log start filename = pynuhap.2005e.log;
outgroup T;
charset molecules = 1-1639;
charset Cytb = 1-621;
charset ND6 = 622-1125;
charset Dloop = 1126-1639;
charset 1st_pos = 1-1125\3;
charset 2nd_pos = 2-1125\3;
charset 3rd_pos = 3-1125\3;
charset non_coding = 1126-1639;
partition by_codon = 4: 1st_pos, 2nd_pos, 3rd_pos, non_coding;
partition genes = 3: Cytb, ND6, Dloop;
partition all = 1: molecules;
set partition = by_codon;
databreaks 621 1125;
lset nst=2 code=vertmt;
prset ratepr=variable;
set autoclose=yes;
mcmcp ngen=30000000 printfreq=10000 samplefreq=1000
nchains=4 savebrlens=yes filename=pynuhap.2005e.bay;
mcmc;
plot filename=pynuhap.2005e.bay.p;
sumt filename=pynuhap.2005e.bay.t burnin=3000 contype=halfcompat;
log stop;
end;
GEODIS INPUT FILE

Sitta pygmaea

11
1 PIUTE
6 35 36 21 N 118 23 37 W
2 NOROESTE
7 34 49 41 N 119 13 25 W
3 FRAZIER
7 34 46 10 N 118 58 39 W
4 WINSTON
7 34 21 10 N 117 56 06 W
5 BLUE
7 34 20 52 N 117 40 33 W
6 DELAMAR
7 34 17 05 N 116 55 58 W
7 ONYX
7 34 12 19 N 116 42 53 W
8 GRINNELL
7 34 08 30 N 116 46 41 W
9 BLACK
7 33 49 42 N 116 45 04 W
10 THOMAS
6 33 37 28 N 116 41 11 W
11 LAGUNA
7 32 51 08 N 116 26 06 W
6

Clade 1.1
3
F I O
1 0 1
3
4 6 8
1 0 1
0 1 0
0 0 1

Clade 1.5
3
C G P
0 1 1
8
1 2 3 4 5 6 9 11
1 2 3 4 0 2 0 0
0 0 0 0 2 0 0 0
0 0 0 0 0 0 1 1

Clade 1.6
3
B L M
0 1 1
3
1 6 7
3 1 0
0 0 3
0 0 1
Clade 1.7
5
D H J K Q
1 0 1 1 1
10
2 3 4 5 6 7 8 9 10 11
1 3 1 0 0 0 0 1 1 1
0 0 0 1 3 0 2 2 3 3
0 0 0 0 0 2 1 0 0 1
0 0 0 0 0 1 0 1 0 0
0 0 0 0 0 0 0 0 1 0
Clade 2.3
4
1.3 1.4 1.5 1.6
1 1 1 0
10
1 2 3 4 5 6 7 9 10 11
0 1 1 3 0 0 0 0 0 0
2 3 0 1 4 0 0 2 1 0
1 2 3 4 2 2 0 1 0 1
3 0 0 0 0 1 4 0 0 0
Clade 3.1
4
2.1 2.2 2.3 2.4
1 1 1 0
11
1 2 3 4 5 6 7 8 9 10 11
0 0 0 1 0 1 0 2 0 0 0
0 0 0 0 0 0 0 2 0 0 1
6 6 4 8 6 3 4 0 3 1 1
0 1 3 1 1 3 3 3 4 5 5
REFERENCES

Intraspecific phylogeography and adaptive divergence in the white-headed woodpecker.


genes in mitochondria and chloroplasts, and some results. Genetics. 103:513-527.


the Hawaiian Islands to estimate molecular evolutionary rates. *Molecular Ecology*. 7:533-545.


118


nested-cleade analysis of mitochondrial-DNA variation.

*The Auk.* 120:346-361.


