TESTING THE ROLE OF BAJA CALIFORNIA GENERATING BIODIVERSITY: A TEST CASE CHARACTERIZING THE POPULATION GENETIC STRUCTURE OF CASSIN'S AUKLET (*PTYCHORAMPHUS ALEUTICUS*)

By

SARAH JANE WALLACE

A thesis submitted to the Department of Biology

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

(September, 2012)

Abstract

Characterizing the population genetic structure of a species can provide insight into isolating effects of local adaptation and genetic drift, and homogenizing effects of gene flow. The relative interplay between gene flow and genetic drift can indicate whether multiple conservation units are needed to preserve biodiversity. Baja California, Mexico is a biodiversity hotspot where many individuals are genetically differentiated from conspecific individuals breeding elsewhere. Cassin's auklet (*Ptychoramphus aleuticus*), a zooplanktivorous seabird, breeds in colonies along the Pacific coast of North America. One subspecies, *P. a. australe*, is described from Baja California and another, *P. a. aleuticus*, is described from the rest of its range to the Aleutian Islands, Alaska. Many of the colonies throughout its range have been declining in numbers likely as a result of decreases in food availability in cold-water upwellings along the coast. The mitochondrial control region and eleven nuclear microsatellites were analysed using programs based on coalescent and Bayesian theory to determine how the southern subspecies compares genetically to the main subspecies.

First, population genetic structure was characterized throughout Cassin's auklet's breeding range using pairwise genetic differentiation indices, hierarchical analysis of variance, statistical parsimony and Bayesian clustering methods. The two subspecies were genetically differentiated but individuals breeding in the Channel Islands were more genetically similar to *P. a. australe* individuals. Population genetic differentiation was not evident within the two genetic groups.

Second, gene flow between the two genetic groups was estimated using coalescent and Bayesian methods. Significant gene flow was estimated from the northern group (Aleutian

Islands to Southeast Farallon Islands) into the southern group (Channel Islands to San Benito Island) but not from the southern group into the northern group since the time of divergence, possibly as a result of the non-breeding distribution. Results show that the two genetic populations diverged recently and the populations have experienced a recent population change in size. Restrictions in home range and foraging range may cause population genetic differentiation, resulting in two distinct management units. Genetic differentiation of the southern population provides support for Baja California and southern California being a biodiversity hotspot.

Acknowledgements

This thesis would not have been completed if it was not for the support of many people. First of all, I would like to whole-heartedly thank my supervisor Vicki Friesen for her guidance for this project and for life during my time as a graduate student. She encouraged me to become involved with many things in the department and I would not have had this great experience or find my passion without her support. Thank you to Stephen Lougheed and Ryan Danby for serving on my committee and providing helpful feedback. Thank you to Tim Birt for help in the lab with all the "tricky" samples and to Zhengxin Sun for genotyping and troubleshooting of microsatellites. A big thank you to Jamie Morris-Pocock, Scott Taylor and Gaby Ibarguchi for friendship, guidance and teaching me everything I know in the lab. Thank you to Amanda Xuereb for her help understanding these "complex" concepts and for her emotional support through every step of the game. Thank you to Friesen lab members Kim, Petra, Erin, Melissa, Scott, Will and Rebecca, for your friendship, support in the lab and listening to all of my practice talks. Thank you to my volunteers Kirsten, Jason and Erin for their help finishing the seemingly endless PCRs.

Second of all, range-wide population genetics projects on seabirds would never be possible without all of the sample collectors and funding agencies. I would first like to thank Tony Gaston for teaching me everything I know about field work. My field work on Reef Island, Haida Gwaii was partially funded by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to Vicki Friesen and funding from Environment Canada through Tony Gaston. A big thank you to Shaye Wolf for collecting all the samples from Mexico and the initial idea for the project. I would like to thank the following people for collecting or organizing the collection of samples: J. Williams, B. Drummond and J. Warzybok from the Alaska Maritime

National Wildlife Refuge, M. Drever under a Canadian Climate Change Action Fund Grant to D. Bertram, M. Hipfner, R. Bradley from PRBO Conservation Science and Farallon National Wildlife Refuge, L. Harvey and K. Carter with Channel Islands National Park, S. Birks, K. Winker, H. Walsh and B. Henry. Samples were also loaned from the University of Washington Burke Museum, American Museum of Natural History and University of Alaska. I am grateful to NSERC for awarding me a CGS-M scholarship and to Queen's University for a Graduate Award. Thank you to the Society for the Study of Evolution, the 5th North American Ornithological Conference and Queen's University for funding travel to meetings.

Finally, thank you to all my friends for being there for me through thick and thin and for keeping me sane (especially with coffee dates and mini-dance parties). A big thank you to my partner Korey for his endless support throughout this adventure and for always making me laugh. The largest thank you of all goes to my family for supporting me and always being there for me no matter what. I would not have made it to the end of this journey (still with a smile) without you all!

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABC: approximate Bayesian computation

AMOVA: analysis of molecular variance

ESS: effective sample size

ESUs: evolutionary significant units

Hipt: high point

HPD: highest posterior density

IUPAC: International Union of Pure and Applied Chemistry

MCMC: Markov chain Monte Carlo

MUs: management units

mya: million years ago

PCR: polymerase chain reaction

Chapter 1: Introduction and Literature Review

General introduction

Determining the population genetic variation of a species and understanding how populations diverge is an integral part of characterizing biodiversity. Gene flow can connect fragmented populations genetically and help maintain population genetic diversity by counteracting genetic drift. Conversely, if barriers to gene flow exist, populations have the potential for local adaptation and divergence. However, divergence can also occur with gene flow and quantifying the relative interplay between local adaptation, genetic drift and gene flow can provide insight into how species diversify. In addition, describing historical and contemporary gene flow between populations can aid in implementing effective conservation efforts.

Seabirds provide interesting systems for studying microevolutionary processes such as genetic drift and gene flow because population genetic structure may be influenced by both physical and non-physical barriers to gene flow (reviewed in Friesen *et al*. 2007). Most seabirds are capable of flying long distances allowing for possible population genetic admixture (e.g. wandering albatross, *Diomedea exulans*, Milot *et al*. 2008). However, seabirds often exhibit high natal philopatry which could isolate colonies and result in population genetic structure (reviewed in Friesen *et al.* 2007).

In this introduction, the hypothesis that the southern subspecies of Cassin's auklets (*Ptychoramphus aleuticus*) is genetically differentiated from the northern subspecies is reviewed, testing the importance of Baja California as a biodiversity hotspot.

Baja California as a biodiversity hotspot

Baja California, Mexico, is a biodiversity hotspot, where individuals of many species are genetically differentiated from conspecific or congeneric individuals breeding elsewhere. Baja California has been ranked as having the second highest percentage of endemic species of Indo-Pacific reef fish (Allen 2008) and is well known for its high percentage of endemic species of vascular plants (Riemann and Ezcurra 2005). Guadalupe Island in particular is recognized for its wide array of contemporary and historical endemic species across many taxonomic levels including marine invertebrates (Hubbs 1960), littoral marine plants (Neushul *et al.* 1967), seed plants (Thorne *et al*. 1969), other flora (Thorne *et al*. 1969, Garcillán *et al*. 2008), deep-sea benthic fish (Hubbs 1960) and birds (Barton *et al*. 2004 and references therein) including seabirds such as Xantus's murrelet (*Synthliboramphus hypoleucus*, Birt *et al*. 2012) and Leach's storm-petrel (*Oceandroma leucorhoa*, Friesen *et al*. unpubl.). Guadalupe Island is an oceanic volcanic island that is situated approximately 260 km west of the Baja California Peninsula, isolating individuals from conspecifics on the mainland (Neushul *et al.* 1967, Riemann and Ezcurra 2005, Garcillán *et al*. 2008).

Changes in oceanographic and climatic conditions from Baja California, Mexico to Point Conception, California, may also contribute to the high diversity of taxa in this area (Neushul *et al*. 1967, Dawson *et al*. 2001). The California Current flows south along the coast of California until reaching Point Conception where the current flows east, creating different oceanographic conditions above and below this point (Neushul *et al.* 1967). In southern and central California, a southern counter-current flows northward creating local upwellings (Dawson *et al.* 2001). Phylogeographic breaks occur at varying latitudes in this area, suggesting that specific barriers to

gene flow differ between species depending on their ecology and history, resulting in a transition zone in southern California (Dawson *et al*. 2001).

The study species: Cassin's auklet

Cassin's auklet (Charadriiformes: Alcidae), a burrow-nesting seabird, breeds along the Pacific coast of North America, with colonies on islands from Baja California, Mexico to the Aleutian Islands, Alaska (Gaston and Jones 1998, Ainley *et al*. 2011; Figure 1). Currently, one subspecies, *P. a. australe*, has been described from Baja California, and another, *P. a. aleuticus*, from the remainder of the range (Gaston and Jones 1998, Ainley *et al.* 2011). Although the subspecies are similar morphologically, the northern subspecies has a slightly longer body and larger body mass with a clinal decrease in body mass from north to south (Gaston and Jones 1998, Ainley *et al*. 2011). Natal philopatry appears to be prevalent in Cassin's auklets but experience at a breeding site was not correlated with reproductive success (Pyle 2001, Pyle *et al.* 2001). Cassin's auklets appear to be socially monogamous but most social behaviour occurs in their burrows (Gaston and Jones 1998). Observations of pairs of Cassin's auklets leaving the Triangle Island and Frederick Island breeding sites suggested pairing occurs at the nesting colony but copulation may occur at sea like *Aethia* species (Gaston and Jones 1998). Mate switching appears to be rare (Pyle *et al.* 2001, Ainley *et al*. 2011) but is more common when a breeding season is skipped (Pyle *et al*. 2001). The timing of breeding varies along the Cassin's auklet breeding range with laying of one egg in November in Baja California, March to May in central California, late March to April in British Columbia and early May in the Aleutian Islands (Gaston and Jones 1998). Cassin's auklets are zooplanktivorous, feeding on mainly copepods and euphausiids near the continental shelf break (Gaston and Jones 1998, Bertram *et al*. 2001, 2009, Abraham and Sydeman 2004).

Reproductive success depends on the availability of appropriate nesting habitat as well as proximity to and availability of prey food resources (e.g. Lee *et al*. 2007, Hipfner 2008, Wolf *et al*. 2009). Cassin's auklets are tightly associated with cold-water upwellings because reproductive success mainly depends on the availability of euphausiids (Bertram *et al*. 2001, Hyrenback and Veit 2003). The highest abundance of euphausiids in the North Pacific occurs in cold water upwellings where nutrient-rich water is brought to the surface as a result of winds hitting the coast (Huyer 1983, Legaard and Thomas 2006). The strength, duration and timing of upwellings differ along the Pacific coast. In the north, upwellings occur along the Aleutian Islands arc (Stephensen and Irons 2003) and a down-welling regime occurs frequently in the Gulf of Alaska (Day 2006). The North Pacific Current flows east from the central Pacific and bifurcates in mid-British Columbia (Huyer 1983, Legaard and Thomas 2006). North of this bifurcation, the upwellings are seasonal and fairly weak compared to the Central California Current that flows from Vancouver Island to central California and is seasonal but strong (Huyer 1983, Legaard and Thomas 2006). The southern part of the California Current from southern California to Baja California produces moderate upwellings all year round (Huyer 1983, Legaard and Thomas 2006). These predictable food resources in central California are likely the largest support for a year round resident population of Cassin's auklets and possibly support migrants from farther north (Adams *et al*. 2004).

Cassin's auklets appear to adjust the timing of their breeding to coincide with the highest abundance of euphausiids at the highest peak of upwellings (Abraham and Sydeman 2004, Hipfner 2008, Bertram *et al.* 2009, Wolf *et al*. 2009). However, these cold upwellings are affected by changes in the sea surface temperature (e.g. Huyer 1983, Hyrenbach and Veit 2003). As the surface temperature increases, the cold water upwellings are reduced, decreasing the

amount of primary productivity (Huyer 1983, Hyrenbach and Veit 2003). Since Cassin's auklets feed at a lower trophic level than other predators, they are more directly linked to ocean processes and thus may be more sensitive to changes in upwellings (Abraham and Sydeman 2004, Wolf *et al.* 2009). Abraham and Sydeman (2004) found that in the 1983, 1992 and 1998 breeding seasons during the warm phase of the El Niño Southern Oscillation, the hatch dates of Cassin's auklets at Southeast Farallon Island were later in the season, likely as a result in the shift of prey availability due to a reduced upwelling. Similarly, at Triangle Island, British Columbia, when sea surface temperatures were warmer, the timing and duration of the peak of highest prey availability was mismatched with the timing of Cassin's auklets breeding, causing a decrease in nestling growth rates and a lower probability of surviving to fledging age (Bertram *et al.* 2001, Hipfner 2008). To test this directly, Hipfner *et al*. (2010) removed the egg of the first breeding attempt from a sample of birds breeding on Triangle Island. The pairs that relaid (experimentally delayed timing of breeding) had a reduced fledging success rate than the control pairs that laid earlier (Hipfner *et al*. 2010).

Although numbers of Cassin's auklet have been stable at approximately 3.5 million individuals (Ainley *et al.* 2011), previous studies showed declines at some colonies throughout the breeding range. At the largest Cassin's auklet colony in California at Southeast Farallon Island, burrow occupancy declined by approximately 6.1% per year between 1971 and 2002 (Lee *et al*. 2007). Moreover, in the spring of 2005 on Southeast Farallon Island, Cassin's auklets abandoned the island completely and the reproductive success decreased to zero as a result of an atmospheric blocking event where the jet stream was farther south than usual, disrupting the California Current System (Sydeman *et al.* 2006). Although declines in the population have been noted before, in 35 years of monitoring, a decline of this magnitude has never been seen

(Sydeman *et al*. 2006). In addition to the recent decline at Southeast Farallon Island, Wolf *et al.* (2010) used a mid-level emissions model for climate change and Cassin's auklet demographic population model to predict that the Southeast Farallon Island Cassin's auklet population growth rate (λ) could decline by 11-45% by 2099.

In addition to ocean-wide climate processes differentially affecting the currents along the coast, the local oceanographic environment also varies from colony to colony (Wolf *et al*. 2009). Whereas the colonies in the California Current System were strongly impacted by oceanographic climatic conditions, impacts were less pronounced in the northern colonies (Gaston *et al*. 2009). Cassin's auklets' burrow occupancy declined by 2.3% per year at the largest Cassin's auklet colony, on Triangle Island, British Columbia, between 1984 and 1999 but the number of Cassin's auklets appeared more stable between 1999-2004 (Gaston *et al*. 2009). In addition, Rodway and Lemon (2011) noted a decline of 32% in burrow occupancy on Rankine Island in Haida Gwaii, British Columbia between 1984 and 2000, but some recovery seems to have occurred between 2000 and 2010. Similar to Southeast Farallon Island, the reduced upwellings in 2005 resulted in the lowest Cassin's auklet occupancy recorded on Triangle Island (Sydeman *et al.* 2006), but breeding success was stable in the northern British Columbia colonies (e.g. Frederick Island) where Cassin's auklets depend more on downwellings for prey (Gaston *et al*. 2009). Currently, the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) is reviewing whether Cassin's auklet should be listed in Canada.

Along the Aleutian Islands the current distribution of Cassin's auklets is much more fragmented than it was previously, despite the availability of suitable nesting habitat (Bailey 1993, Stephensen and Irons 2003, Byrd *et al.* 2005). Islands that currently have an abundant population of Cassin's auklets (e.g. Buldir Island) never had foxes (red foxes, *Vulpes vulpes* and

Arctic foxes, *Alopex lagopus*) or Norway rats (*Rattus norvegicus*), suggesting that introduced predators extirpating colonies have created this fragmented distribution (Byrd *et al*. 2005). This led the United States Fish & Wildlife Service (2006) to designate Cassin's auklets as having high conservation concern in Alaska. Furthermore, breeding colonies of Cassin's auklets are absent from the Gulf of Alaska (Gaston and Jones 1998) despite seemingly suitable habitat (Stephensen and Irons 2003). Stephensen and Irons (2003) suggested the lack of least and crested auklets (*Aethia pusilla* and *Aethia cristatella*, respectively) in the Gulf of Alaska may be due to the limited appropriate nesting habitat due to vegetation growing quickly over the rock crevices, blocking entrances to nesting sites. Many of the Cassin's auklet colonies on islands in Mexico have also been extirpated by predators, although some islands have been recolonized (Wolf *et al.* 2006). The various magnitudes of decline of Cassin's auklet colonies demonstrate the importance of characterizing population genetic diversity to determine the appropriate number of conservation units.

Describing population genetic variation for conservation

Defining conservation units can lead to more effective allocation of limited funds for populations that are in need of management. Evolutionary significant units (ESUs) are populations that are both geographically and genetically distinct (reviewed in Moritz 1994, Crandall *et al.* 2000, Funk *et al.* 2012). The amount of genetic differentiation suggested to consider populations as separate ESUs varies among studies (Moritz 1994). In general, reciprocal monophyly of mitochondrial alleles suggests historical isolation of populations leading to population distinctiveness (Moritz 1994). More recently, restricted gene flow between lineages has been used to diagnose ESUs (Fraser and Bernatchez 2001). However, both neutral and adaptive variation should be assessed to delineate ESUs (Crandall *et al.* 2000, Fraser and

Bernatchez 2001, Funk *et al.* 2012). In contrast, contemporary population structure based on neutral variation can be used to infer management units (MUs), which are demographically (Moritz 1994) but not genetically isolated (Faubet *et al*. 2007, Funk *et al.* 2012).

Determining population genetic variation and structure

To obtain independent estimates of genetic variation within and between populations of Cassin's auklets, the mitochondrial control region and 11 nuclear microsatellites were analysed in the present study. The mitochondrial control region is a non-coding, hypervariable region of the mitochondrial genome that has a high mutation rate (Baker and Marshall 1997). Mitochondrial DNA is haploid and maternally inherited in vertebrates, thus has one-fourth the effective population size of nuclear DNA, making it more sensitive to changes in population size and gene flow than nuclear DNA (Moore 1995). Microsatellites are short tandem repeats of two to six nucleotides in the nuclear genome (McDonald and Potts 1997). Due to slippage of DNA polymerase during replication and errors in crossing-over, different numbers of repeats arise producing distinct alleles (McDonald and Potts 1997). As a result of typically high mutation rate of microsatellites, recent changes in gene flow and population size are detectable, but homoplasy occurs more frequently than in mitochondrial DNA (Estoup *et al.* 2002).

Genetic variation can be measured within and among populations to characterize population genetic structure and gene flow within a species. Wright's (1951, 1965) F-statistics and Cockerham's (1969) θ-statistics are based on the relative allele frequencies at the individual, sub-population and total population levels. F_{ST}, the standardized index of genetic variance, is used to characterize population genetic differentiation by estimating the variance in allele frequencies between sub-populations within the total population (i.e. between populations within

a species; Holsinger and Weir 2009). By using an analysis of molecular variance (AMOVA) to estimate hierarchical F-statistics, such as F_{CT} (the genetic variance between population groups) and F_{SC} (the genetic variance within population groups), populations can be partitioned into groups to maximize F_{CT} and minimize F_{SC}.

Determining population genetic structure from F_{ST} estimates introduces bias because the populations have to be defined *a priori* (Pritchard *et al*. 2000). In contrast, the program STRUCTURE (Pritchard *et al*. 2000) uses a Bayesian maximum likelihood assignment method using a Markov Chain Monte Carlo (MCMC) to estimate the number of genetic clusters that best describes the data by assigning individuals to clusters based on their individual genotype and not sampling location (Pritchard *et al*. 2000). The genetic clusters are formed by minimizing deviations from Hardy-Weinberg equilibrium and linkage equilibrium (Pritchard *et al.* 2000).

Estimating gene flow

Estimating past migration rates from contemporary population genetic variation can be difficult due to the generally complex histories of populations (Hey and Nielsen 2004). In Whitlock and McCauley's (1999) review of Wright's (1931) population island model, they discuss how estimating the number of migrants (Nm; where N is the effective population size and m is the migration rate) from F_{ST} values makes many assumptions that are not usually true in wild populations (e.g. equal exchange of genetic material between populations, large and constant population sizes, migration-drift equilibrium). Hey and Nielsen (2004) developed an isolation-with-migration model based on coalescent theory that accounts for the possibility of gene exchange during divergence. It considers a situation in which one ancestral population diverges into two populations at a certain time, following which symmetrical or asymmetrical

migration may have occurred (Hey and Nielsen 2004). However, the isolation-with-migration-analytic model implemented in IMa (Hey and Nielsen 2007) has the following assumptions: the molecular markers used are selectively neutral, recombination rate is zero, other populations are not exchanging genes with the two focal populations, and population structure is not evident within the two focal populations (Hey and Nielsen 2004).

The isolation-with-migration model is useful to estimate average gene flow among populations since the time of divergence (Hey and Nielsen 2004). However, the timing of gene flow also can provide insight into historical versus contemporary evolutionary processes. Specifically, the amount of contemporary gene flow can indicate demographically isolated populations, useful for determining management units (Faubet *et al.* 2007). The program BAYESASS uses a Bayesian method to estimate the amount of gene flow that has occurred among populations within the last few generations (Wilson and Rannala 2003). Some key assumptions are relaxed in the model implemented in BAYESASS, where deviations from Hardy-Weinberg equilibrium are tolerated but linkage equilibrium is assumed (Wilson and Rannala 2003, Faubet *et al.* 2007).

Objectives and hypotheses

To determine whether the two subspecies of Cassin`s auklets are genetically differentiated, this thesis had three objectives: 1) summarize the population genetic variation of sampling sites from throughout the breeding range, 2) characterize the population genetic structure throughout the range, and 3) test whether any gene flow has occurred between distinct genetic groups since the time of divergence. The gaps in the Cassin's auklet breeding distribution, the documented declines of Cassin's auklet throughout the range, and previous

research on species with similar ranges suggest that the two subspecies will be genetically differentiated, and that Cassin's auklets will likely exhibit strong population genetic structure throughout the range with some gene flow occurring. This study will add to the knowledge of whether the Baja California subspecies is genetically differentiated from the northern subspecies as well as providing information about appropriate conservation units for management decisions.

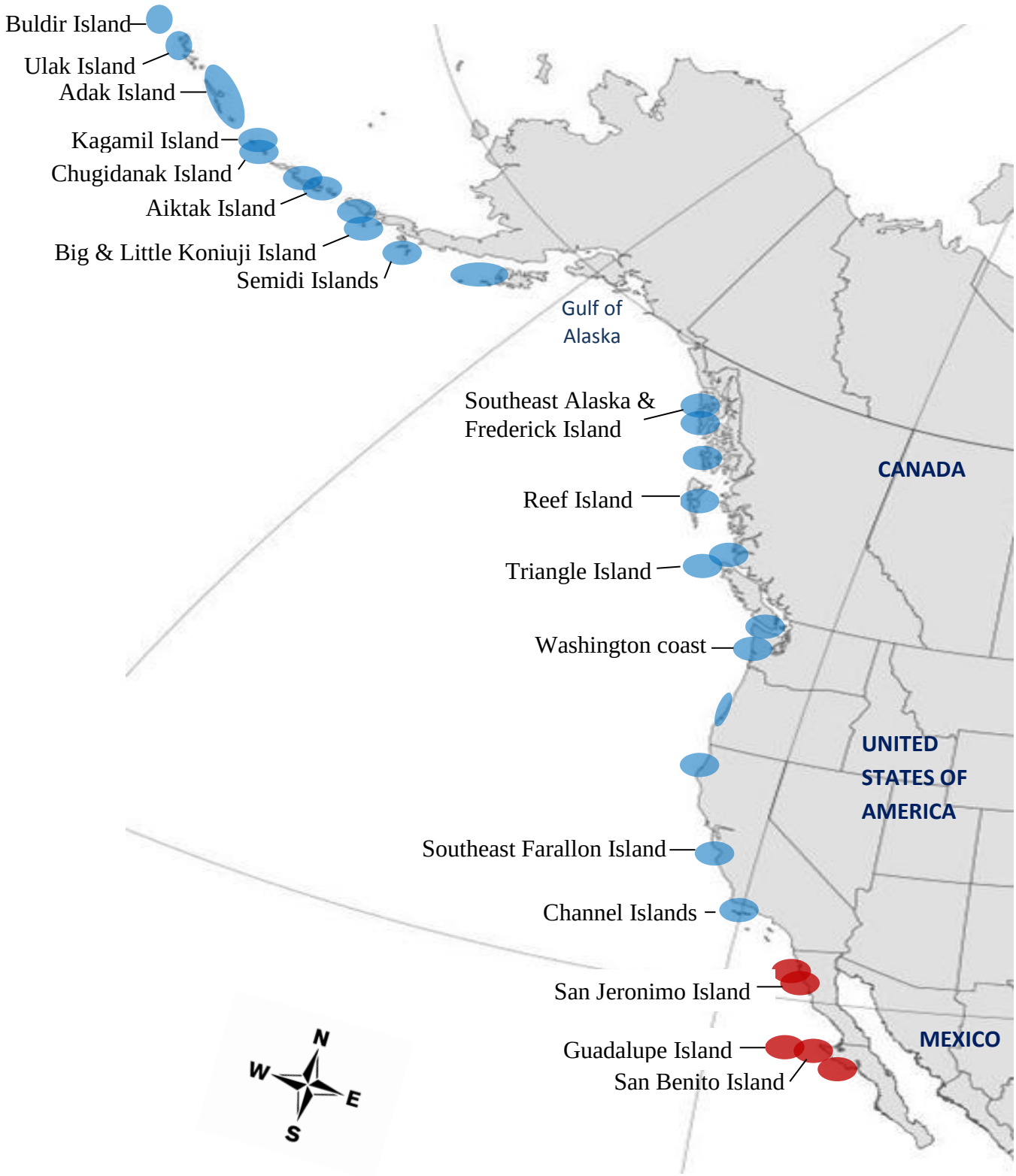


Figure 1.The breeding distribution of Cassin's auklets and sampling locations (labelled) for this study. Blue shaded areas represent *P. a. aleuticus* breeding distribution and red shaded areas represent *P. a. australe* breeding distribution. The number of samples collected from each location are listed in Table 1.

Chapter 2: Materials and Methods

Sample collection and DNA extraction

Blood or solid tissue samples were collected over a 22 year span from 250 breeding Cassin's auklet adults or chicks at eight sampling sites along the Aleutian Islands, in addition to nine sampling sites from the rest of the breeding range (Figure 1, Table 1). Most samples were blood that were collected by pricking the individuals' ulnar vein where it crosses the joint between the humerus and radius/ulna and drying a few drops of blood on filter paper or preserving the blood in ethanol. For some individuals from the Aleutian Islands and Washington, solid tissue samples (e.g. liver, heart) were collected from carcasses found in the breeding season or shortly after (to increase confidence that the individuals were breeding in that area) and were archived in museums (Table 1). One family group (two adults and a chick) was sampled from Reef Island, British Columbia to confirm correct scoring of microsatellites (Dewoody *et al*. 2006). All samples are archived at -80^{o}C at Queen's University, Kingston, Ontario. DNA extractions were performed using the standard proteinase-K phenol/chloroform method (Sambrook *et al.* 1989). Samples that did not amplify well were re-extracted using the ChargeSwitch Forensic DNA Purification kit (Invitrogen, Carlsbad, CA).

Laboratory methods: *mitochondrial control region*

Published sequences of regions flanking the control region from Cassin's auklet mitochondrial DNA (NADH dehydrogenase subunit 6, Moum *et al.* 1994; 12S ribosomal RNA gene, Baker *et al.* 2007) and mitochondrial control region sequences of closely related species (family: Alcidae) were used to develop polymerase chain reaction (PCR) primers for Cassin's auklet control region (ca_cr11L and ca_cr960H, Table 2, Appendix A1). Internal primers

(ca_cr600H and ca_cr652L, Table 2, Appendix A1) were designed in the conserved domain II of the control region for better amplification of the more variable domains I and III. For some individuals, a new primer (ca_cr652Lb) was modified from the internal primer ca_cr652L to obtain clearer sequence at the 5' end of domain III (Table 2). PCRs were performed in 15 µL reaction volumes with 1x Multiplex Mix (Qiagen, Mississauga, Ontario), 0.4 mM each of the heavy and light strand primers and approximately 50-100 ng of template. The reactions were amplified in a TGradient (Biometra, Montreal, PQ) thermocycler. Due to the HotStarTaq in the Multiplex Mix, the amplifications began with an initial denaturation step of 15 minutes at 94^{o}C (Qiagen), followed by 35 cycles of denaturation at 94^{o}C for 30s, annealing at 50^{o}C for 30s and extension at 72^{o}C for 45s. A final extension of three minutes at 72^{o}C was applied. The PCR products were visualized on 2% agarose gels and sequenced in one direction using the 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA) at the Genome Quebec Innovation Centre (Montreal, QC). The heavy strand was sequenced for the 5' end and the light strand was sequenced for the 3' end based on clearer sequences. Chromatograms were checked manually and ambiguous sites were labelled with the appropriate International Union of Pure and Applied Chemistry IUPAC code. Sequences were aligned by using CLUSTALW (Larkin *et al.* 2007) as implemented in BIOEDIT (Version 7.0.9.0, Hall 1999).

Laboratory methods: *microsatellites*

Primers for five microsatellite loci designed for whiskered auklet (*Aethia pygmaea*, Dawson *et al.* 2005) were tested in ten Cassin`s auklet individuals. Two loci amplified well, were polymorphic, and thus were amplified in all of the individuals (Apy02, Apy10, Table 3). For microsatellite primer development in Cassin's auklet, 1 µg of DNA was sheared to create an Illumina paired-end shotgun library on a sample preparation Covaris S220 System (Applied

Biosystems, Foster City, CA) using the standard protocol in the Illumina TruSeq DNA Library Kit and a multiplex identifier adaptor index (Savannah River Ecology Laboratory). Paired-end reads of 100 base pairs (bp) were sequenced on the HiSeq automated sequencer (Illumina, California, USA). Reads were analyzed with the program PAL_FINDER_v0.02.03 (Castoe *et al.* 2012) to select sequences that included microsatellite repeats ranging from dinucleotide to hexanucleotide repeats. The program PRIMER3 (version 2.0.0, Rozen and Skaletsky 2000) was used to select loci with flanking primer sites that only occurred once in the library. Loci that were between eight and fifteen repeats long, were tetra- or pentanucleotides, and had a "GC" clamp at the end of the primers were amplified in a subset of individuals (details below). Forty-three pairs of primers were tested on three individuals, of which 26 pairs produced clear amplifications and were polymorphic (Table 3). For the polymorphic loci, sixteen individuals from throughout the range were amplified to determine variation in lengths of fragments. Loci were chosen for further analysis if the size of the fragment was less than 400 bp, alleles appeared to differ in length in a stepwise fashion, and genotype frequencies did not deviate significantly from Hardy-Weinberg equilibrium (see below). Several loci were grouped into either triplex (Triplex 1: Pal11, Pal18, Pal26) or duplex reactions (Duplex 1: Pal17, Pal25, Duplex 2: Pal21, Pal38, Duplex 3: Pal37, Pal42). The other loci were amplified in a single reaction.

For PCR, M13 forward sequence was added to the 5' end of the forward primers (M13F: 5'-CACGACGTTGTAAAACGA-3'). PCR reactions were performed in 5 μL reactions with 1x Qiagen Multiplex Mix (containing 3mM Mg^{2+}), 1 μM M13 forward primer labelled with D4 dye (Sigma-Aldrich, Oakville, ON), 0.15 mM of each forward and reverse primer, and approximately 20 ng of DNA. Most loci were amplified using a touchdown protocol with an initial denaturation of 15 minutes at 95^{o}C, followed by 17 cycles of denaturation at 95^{o}C for 45s, annealing for 45s

at a temperature from 60-52^{o}C decreasing by 0.5^{o}C each cycle, and an extension of 30s at 72^{o}C. Another 21 cycles followed, with denaturation at 94^{o}C for 45s, annealing at 52^{o}C for 45s and an extension of 72^{o}C for 30s. A final extension of 10 minutes at 72^{o}C was applied. For loci that did not amplify well with a touchdown protocol (Pal21, Pal38) a protocol of initial duration of 95^{o}C for 15 minutes, followed by 35 cycles of 95^{o}C for 45s, 52^{o}C for 1 minute and 72^{o}C for 30s was performed. A final extension of five minutes was performed at 72^{o}C. The PCR products were genotyped for length variation using a Beckman Coulter CEQ 8000 capillary automated sequencer at Queen's University Department of Biology Core Genotyping Facility, and alleles were scored visually using the Beckman Coulter CEQ 8000 genetic analysis system (version 9.0). The loci were tested for null alleles and scoring errors in MICRO-CHECKER (version 2.2.3, Oosterhout *et al.* 2004).

Population genetic analysis of the control region

ARLEQUIN was used to estimate standard diversity indices including haplotypic diversity (*h*, Nei 1987) and nucleotide diversity (π, Tajima 1983). The standard deviations were reported after the estimates. Many analyses of population genetic structure and gene flow assume that the molecular markers are selectively neutral. Therefore, to test for deviations from selective neutrality in the mitochondrial control region, Fu's F test (Fu 1997) and Tajima's D test (Tajima 1989) were performed in the program ARLEQUIN (version 3.5.1.2, Excoffier and Lischer 2010) assuming an infinite site mutation model. Chakraborty's (Chakraborty 1990) and Ewens-Watterson's (Ewens 1972, Watterson 1978) tests of deviation from selective neutrality were also estimated in ARLEQUIN. A maximum likelihood program based on coalescent theory, FLUCTUATE (version 1.4, Kuhner *et al.* 1998) was used to estimate whether populations were growing. Thirty short chains of 1000 steps and four long chains of 100,000 steps were run,

sampling every 20 steps. Watterson's estimate of θ was used as a prior and the population size was allowed to vary. A second run was conducted with all parameters the same except that the population growth rate (g) was constrained to zero. Significance was determined by comparing G, estimated by two times the difference between the ln likelihood of the maximum value of g and the ln likelihood at g=0, to the critical value of χ^2 at one degree freedom (M. Kuhner, *pers. comm.* in Veit *et al.* 2005).

To test for a significant correlation between genetic and geographic distance throughout the breeding range, a Mantel's test was applied in ARLEQUIN to a matrix of pairwise Slatkin's linearized ϕ_{ST} values (Slatkin 1995) and a matrix of great-circle distances based on latitude and longitude of sampling sites (calculated from http://www.movabletype.co.uk/scripts/latlong.html). Ten thousand permutations were performed and significance was determined at an alpha level of 0.05.

Pairwise ϕ_{ST} estimates were calculated between sampling sites using ARLEQUIN. Individuals were grouped based on their sampling location, except sampling sites with fewer than nine individuals which were grouped by region (e.g. Aleutian Islands; southeast Alaska and Frederick Island, Table 1). Ten thousand permutations were performed using pairwise distances at a significance level of 0.05. A Benjamini-Yekutieli correction was applied to reduce the chance of a type I error but to retain enough power to avoid type II errors (Narum 2006). Hierarchical F-statistics were calculated with an AMOVA in ARLEQUIN, with sampling sites grouped by (1) subspecies, (2) significant pairwise ϕ_{ST} estimates, or (3) genetic breaks hypothesized from geographic separation (e.g. Gulf of Alaska) to determine what grouping of sampling sites maximized the variation between groups (highest ϕ_{CT} value) and minimized the

variation within groups (low ϕ_{SC} value). Ten thousand permutations were performed using a significance level of 0.05.

Finally, the relationships between control region haplotypes were summarized in a haplotype network estimated using statistical parsimony (Templeton *et al.* 1992) in TCS (version 1.21, Clement *et al.* 2000).

Population genetic analysis of microsatellites

ARLEQUIN was used to test for deviations from Hardy-Weinberg equilibrium and linkage equilibrium within each sampling site. Allelic richness, number of alleles and gene diversity were estimated using FSTAT (version 2.9.3.2, Goudet 1995). The program BOTTLENECK (version 1.2.02, Cornuet and Luikart 1996) was used to test for evidence of a recent population expansion or bottleneck (i.e. within the last few generations) for distinct genetic groups. As the mutation model for microsatellites is unknown, the two-phase mutation model with 70% stepwise mutations and 30% infinite alleles was run for 1000 iterations. To estimate population genetic structure, pairwise F_{ST}s were estimated, and AMOVA and Mantel`s tests were conducted on F_{ST} in ARLEQUIN using the infinite-alleles model as above.

In addition, STRUCTURE (version 2.3.2, Pritchard *et al.* 2000) was used to determine the number of genetic clusters (K value) that most accurately described the genotypic data by assigning individuals to genetic clusters based on their genotype. An "admixture" model was applied allowing allele frequencies to be correlated between populations that allows for the possibility of populations experiencing different amounts of drift (Falush *et al.* 2003). The model was run for 1.0×10^6 generations. The number of genetic clusters (K) ranged from a minimum of 1 to a maximum of 10 (the number of sampling sites) and each was run for 10 iterations. Since

structure appeared to be weak, a "no admixture" model was applied with the prior "locprior", which uses sampling sites (1-10) as prior information to help cluster individuals (Pritchard *et al*. 2000, Hubisz *et al*. 2009). A burnin of 1.0×10^5 generations was applied and the number of genetic clusters (highest ΔK) was determined using Evanno's method (Evanno *et al.* 2005). For the highest ΔK value, the "Greedy" option was used in CLUMPP (version 1.1.2, Jakobsson and Rosenberg 2007) to summarize the 10 iterations from STRUCTURE, and DISTRUCT (version 1.1, Rosenberg 2004) was used to display the results graphically. STRUCTURE is only able to detect the highest level of genetic structure (Evanno *et al*. 2005), thus subsets of the data were run separately to assess any sub-structuring within the major genetic clusters.

Estimation of gene flow

Gene flow was estimated from the mitochondrial control region dataset using an isolation-with-migration model as implemented in IMa (Hey and Nielsen 2007). Initial runs included wide uninformative priors on parameters (ancestral population size, θ_A; two population sizes, θ_1, θ_2; maximum time for population splitting, t), and migration parameters (m_1 and m_2) were set to zero. Priors were optimized by adjusting the values based on the posterior probabilities from preliminary runs reaching convergence by the distribution reaching zero. When the population size parameters and divergence time parameters were optimized, migration parameters were added and optimized. Single chains were run without heating parameters, recording results every 0.5 hour. A burnin of 1.0 hour was applied to ensure stationarity of the runs. The analysis was run for at least 10,000,000 steps, and effective sample size (ESS) values were monitored. Each analysis was run three times starting at a different random number seed to ensure consistent results (Hey and Nielsen 2007).

To estimate the population migration rate, the number of migrants per generation (2Nm) was calculated by multiplying the high point (Hipt) value of the population size by one half of the Hipt of the migration parameter (Hey and Nielsen 2007). The time since divergence was calculated by $T=t/\mu$, where T is the time since divergence in years and μ is the neutral mutation rate. A mutation rate for the control region in Cassin's auklets has not been estimated, thus a range of values was applied. The average divergence rate for the avian mitochondrial genome has been estimated at 2% per million years (Shields and Wilson 1987, Weir and Schluter 2008) and was used as the minimum sequence divergence rate for Cassin's auklet control region. A maximum sequence divergence rate was set as 20% per million years as the control region mutation rate is often 10 times the average mitochondrial sequence divergence rate (as was done in Kidd and Friesen 1998 for guillemots, *Cepphus* spp.). Estimates using a sequence divergence rate of 3.3% were also conducted as this was the rate of sequence divergence calculated for divergence between razorbill (*Alca torda*) and common guillemot (*Uria aalge*) control region and cytochrome *b* (Moum and Árnason 2001). The divergence rates were converted to a mutation rate per year per locus (M. Miller, J. Hey, *pers. comm*.). The sizes of the ancestral population and two descendant populations were estimated by $N=\theta/4\mu g$ where N is the effective female population size and g is the generation time of Cassin's auklet. The generation time was estimated by $\alpha+(s/(1-s))$ (Lande *et al.* 2003) where α is the age at first reproduction (3.4 years for Cassin's auklet, Pyle 2001) and s is the annual survival rate (0.77 in female Cassin's auklet at the Farallon Islands, Lee *et al.* 2007). The 90% confidence intervals were calculated using the lower and upper bound of the estimated 90% highest posterior density (HPD90Lo and HPD90Hi, respectively).

Finally, BAYESASS (version 1.3, Wilson and Rannala 2003) was used to estimate contemporary migration rates between genetic populations diagnosed from STRUCTURE using the from the microsatellite genotypic data. Three million MCMC iterations were run with a burnin of 999,999 runs, sampling every 2000 runs. Trace files were monitored for convergence and non-migration rates were monitored to ensure sufficient power in the data to estimate migration rates. Standard deviations are reported after the estimates.

Table 1. Sampling information of Cassin's auklet and the number of individuals (N).

Location	Sampling site	Code	Latitude	Longitude	Sampling year	Sample type	Sample collector	N
P. a. aleuticus								
Aleutian Islands, USA	Buldir I.	AL	52°36'N	175°02'E	1989-2011	Soft tissue, blood	J.Warzybok/H.Walsh	15[a]
	Ulak I.		51°51'N	178°47'W	1997,2011	Blood	H.Walsh/J.Williams/ B.Drummond	5
	Adak I.		51°41'N	176°18'W	1991,2002	Soft tissue	UAM	2[b]
	Chugidanak I.		52°28'N	169°20'W	1990	Soft tissue	AMNH	2[c]
	Kagamil I.		52°59'N	169°42'W	1993	Soft tissue	AMNH	1[c]
	Aiktak I.		54°57'N	164°58'W	1990	Soft tissue	AMNH	2[c]
	Big & Little Koniuji Is.		55°57'N	159°59'W	1990,2001	Soft tissue	AMNH/UAM	3[d]
	Semidi Is.		56°52'N	156°01'W	2000	Soft tissue	UAM	2[b]
Southeast Alaska, USA	Lowrie I.	SA	54°51'N	133°32'W	1996	Soft tissue	UAM	5[b]
British Columbia, Canada	Frederick I.	FR	53°55'N	133°09'W	2000	Blood	M.Drever/D.Bertram	30
	Reef I.	RE	52°51'N	131°31'W	2011	Blood	S.Wallace/A.Gaston	20
	Triangle I.	TR	50°51'N	129°05'W	2002	Blood	M.Hipfner	30
Washington, USA	Washington coast	WA	46°02'N	123°55'W	1997-2000	Soft tissue	UWBM	9[e]
California, USA	Southeast Farallon I.	SF	37°41'N	123°00'W	2011	Blood	R.Bradley	30
	Channel Is. (Prince I. and Santa Rosa I.)	CH	34°15'N	120°20'W	2011	Blood	L.Harvey	23

P. a. australe								
Baja California,	San Jeronimo I.	SJ	29°51'N	115°48'W	2003	Blood	S.Wolf	11
Mexico	Guadalupe I.	GU	29°13'N	118°23'W	2003	Blood	S.Wolf	30
	San Benito I.	SB	28°18'N	115°34'W	2003	Blood	S.Wolf	30
Total:								250

[a] 8 samples from American Museum of Natural History (AMNH; DOT1179, DOT1205-1211), 1 from University of Alaska Museum (UAM; UAMX1578)
[b] samples from UAM (UAMX2982, RWD24221, UAMX2265, UAMX1251, KSW4732, UAMX4561, UAMX4562, UAMX4565, JJW1195, UAMX4564)
[c] samples from AMNH (DOT1364, DOT1380, DOT1407, DOT1575, DOT1576, DOT1827)
[d] 2 samples from Big Koniuji Island from AMNH (DOT1494, DOT1550) and 1 sample from Little Koniuji Island from UAM (UAMX2242)
[e] samples from University of Washington Burke Museum (UWBM: 86712, 85562, 85603, 86749, 86752, 86753, 86754, 86755, 90293)

Table 2. Primers designed for the mitochondrial control region in Cassin's auklet. Locations of these primers are displayed in Appendix 1 except ca_ND6L is located at the 3' end of ND6 and ca_12sH is located at the 5' end of 12SrRNA. A site labelled "K" represents a G and a T.

Primer	Sequence
ca_ND6L	5'-GAAGGATTAGACGCAACTGCC-3'
ca_cr11L	5'-ATGCACGAGCGTGTCTCG-3'
ca_cr600H	5'-GGTCGGATGTCGCGATCACG-3'
ca_cr652L	5'-CCCCTCCAGKGCACCGAGG-3'
ca_cr652Lb	5'-CCCCTCCAGTGCACCGAGG-3'
ca_cr960H	5'-ATAGTTTGTGTCGTGTTGGTCG-3'
ca_12sH	5'-GTAAGGTTAGGACTAAGTC-3'

Table 3. Primers designed for microsatellite loci in Cassin's auklet except for the loci indicated with $^{\Psi}$ (Dawson *et al.* 2005). The subscript number beside the repeat motif indicates the number of repeats. A touch-down thermocycler protocol was used if a range of annealing temperatures was given. See text for details.

Primer	Sequence	Repeat motif	Annealing temperature (^{o}C)
Apy02$^{\Psi}$	F:5'-GGAAATTCTTTCCTGAAGTCTCC-3' R:5'-AAAATCTGAGACTATCACCTTTATCAT-3'	$(TATC)_{12}$	60-52^{o}C
Apy10$^{\Psi}$	F:5'-GCAGTTCTTGTTTAGCATGGC-3' R:5'-AGTGCAAGACCGAACCTCAG-3'	$(TATC)_{13}(T)_4$	60-52^{o}C
Pal11	F:5'-AAACTGAAGAACTACACTCTGAAGGG-3' R:5'-CAACCGCTTCACTTAGCTTGG-3'	$(ATGG)_{13}$	60-52^{o}C
Pal17	F:5'-AGGAGGCAGGTGCAAGTCC-3' R:5'-GGTTTGGTGCAACAACTCGG-3'	$(ATGG)_{12}$	60-52^{o}C
Pal18	F:5'-CAGAACTAAAGCATAACTATGAACAGC-3' R:5'-GCTGTCATTAGCCCGAACC-3'	$(AATAG)_{13}$	60-52^{o}C
Pal21	F:5'-GGGAAAGCGAAAGACAAAGG-3' R:5'-ACTGTGAGCTCCACGGGC-3'	$(ATGG)_{13}$	52^{o}C
Pal25	F:5'-TTGCTCTCCTGTGGACACC-3' R:5'-CCCAACCTCTCAACAAAGG-3'	$(AATAG)_{11}$	60-52^{o}C
Pal26	F:5'-GCTGTTTCCGACCTGATGG-3' R:5'-GCTGGCTTATGCTTACCTGC-3'	$(ATGG)_{11}$	60-52^{o}C
Pal37	F:5'-TGTCTAAGCACTCCATTACCATCG-3' R:5'-GTACATCGCACACAGCTCTCC-3'	$(ATGG)_{10}$	60-52^{o}C
Pal38	F:5'-GCAGAGGATTACAGGCAGAGG-3' R:5'-GAGTGAGGCAAGGAGCGG-3'	$(ATGG)_{10}$	52^{o}C
Pal39	F:5'-CTTGTAGCCGCCGTCTAAGC-3' R:5'-CCAGTCTCTGAACCTGGACCC-3'	$(ATGG)_{10}$	60-52^{o}C
Pal42	F:5'-GGATGAGTAGCTCAGTGGATGG-3' R:5'-CGTCACTGATGGAGACTGGG-3'	$(ATGG)_{6}$	60-52^{o}C

Chapter 3: Results

Variability and assumptions of the mitochondrial control region

The full mitochondrial control region (from $tRNA^{glu}$ to 12SrRNA) was sequenced for 10 individuals using the primers ca_ND6L and ca_12sH (Table 2, Appendix A1). Variation was not found in $tRNA^{glu}$ and after six base pairs of the 5' end of domain I a poly-C region caused ambiguous sequences. Thus subsequent individuals were amplified using the internal primer ca_CR11L rather than ca_ND6L. The control region was amplified in two fragments divided by a poly-T region that resulted in unclear sequences. A repeated sequence (CAACA) at the 3'end of domain III resulted in uncertainty in the length of the fragment, thus only 190 bp of domain III were sequenced for all individuals. The two fragments were concatenated, resulting in 706 bp sequence (237 bp of domain I, 275 bp of domain II and 190 bp of domain III).

Several lines of evidence demonstrated that true mitochondrial DNA was amplified: 1) $tRNA^{glu}$ folded into the correct cloverleaf-like structure (Desjardins and Morais 1990), 2) regions with high similarity to previously known conserved sequences, F box, D box, C box, CSB-1 and bird similarity box, were found (Desjardins and Morais 1990, Moum and Bakke 2001, Ruokonen and Kvist 2002; Appendix A1), and 3) the position of variable sites was as expected (38 in domain I and 10 in domain III; Appendix A2). Only individuals with successful amplification of both fragments were included in further analyses. For 227 individuals, 88 haplotypes with 49 variable sites were found, 44 of which were transitions, three were transversions and two had multiple states.

Haplotype diversity was similar among sampling sites ranging from 0.78 ± 0.07 in Triangle Island to 0.97 ± 0.02 in the Aleutian Islands (Table 4). Nucleotide diversity also was

similar among sampling sites ranging from 0.004 ± 0.003 in Southeast Alaska/Frederick Island to 0.008 ± 0.004 in the Aleutian Islands and Guadalupe Island (Table 4). Although the control region is non-coding, Fu's F_S test estimated large negative numbers that were significantly different from zero for the northern and southern populations (Table 4). Significant values with Fu's F_S test can indicate an excess of rare haplotypes that may result from selection on the control region or other mitochondrial gene, or the occurrence of a recent population expansion (Fu 1997, de Jong *et al*. 2011). Estimates of Tajima's D were mostly negative (except Southeast Alaska/Frederick Island and Reef Island) suggesting a population expansion or selection (de Jong *et al*. 2011) but the values did not deviate significantly from zero (Table 4). Deviations from selective neutrality were not evident from the Ewens-Watterson's tests and only the northern population showed significance from the Chakraborty's tests (Table 4). The results from FLUCTUATE suggest a significant recent population growth in both the northern population (G=6.94 p=0.01) and southern population (G=8.04 p=0.005).

Population genetic analysis of the control region

Throughout the Cassin's auklet breeding range, genetic distance based on the control region increased with increasing geographic distance (Mantel's test, R=0.44, p=0.01; Figure 2). Pairwise ϕ_{ST} estimates were not significant between sampling sites within *P. a. aleuticus* or between sampling sites within *P. a. australe* after using a Benjamini-Yekutieli correction (Table 5). In contrast, most pairwise ϕ_{ST} estimates between *P. a. australe* and *P. a. aleuticus* sampling sites were significantly different from zero (Table 5). One exception was that the Channel Islands sampling site was significantly differentiated from all other *P. a. aleuticus* sampling sites and was not significantly genetically differentiated from any *P. a. australe* sampling sites after a Benjamini-Yekutieli correction (Table 5). Global ϕ_{ST} estimates were similar if sampling sites

were grouped by subspecies (ϕ_{ST}= 0.42 p<0.001) or if Channel Islands was grouped with *P. a. australe* (ϕ_{ST}= 0.43 p<0.001) but more of the variation was captured between groups rather than within groups for the latter grouping (Table 6). When the Aleutian Islands were partitioned separately from the rest of the northern subspecies, the global ϕ_{ST} and ϕ_{CT} values were low and not significantly different from zero (Table 6).

Although the two subspecies were not reciprocally monophyletic in the haplotype network, there was evidence of phylogeographic structure (Figure 3). Haplotypes mainly grouped by subspecies in the network with a few haplotypes that were sampled in one subspecies range grouping with haplotypes from the other subspecies (Figure 3). Most haplotypes occurred in only one or two individuals in a single sampling site; however, 22 haplotypes were shared between two or more sampling sites (Figure 3, Appendix A3). Haplotype A was shared between all of the sampling sites except for Channel Islands and Guadalupe Island (Appendix A3). Sampling sites within the *P. a. aleuticus* range did not share haplotypes with sampling sites within the *P. a. australe* range with two exceptions: 1) one haplotype was shared between one individual in Southeast Farallon Island and five individuals in Guadalupe Island and 2) three haplotypes were shared with similar frequencies between Channel Islands and sampling sites farther south (Appendix A3). Several haplotypes were shared between all sampling sites within each subspecies (Appendix A3).

Variability and assumptions for the nuclear microsatellites

All microsatellite loci were highly variable, with the number of alleles per locus within populations ranging from 4-13 with an average of 4.60-11.20 alleles per locus globally (Appendix A4). Gene diversity also was similar among sites, ranging from an average of 0.77 in

Triangle Island to 0.83 in San Jeronimo Island (Appendix A4). In addition, allelic richness was similar among sampling sites, ranging from an average of 4.65 in Southeast Alaska/Frederick Island and Reef Island to 5.07 in Channel Islands (Appendix A4). Results from MICRO-CHECKER suggests that there may be alleles present in three of the eleven loci in the northern population (Pal11, Pal37, Pal39) and one locus in the southern population (Pal11) as indicated by the excess of homozygosity of the allele size classes.

Some loci in some populations did show evidence of deviation from linkage equilibrium; however, significant deviations from linkage equilibrium were not consistent throughout populations or across loci (Appendix A5). Linked markers can create bias in the results, but Bayesian assignment methods are usually robust to low levels of linkage disequilibrium (Falush *et al.* 2003). Many of the loci deviated from linkage equilibrium in the Aleutian Islands population, but this may be a result of missing data or the Wahlund effect (Sinnock 1975; Appendix A5). Significant deviations from Hardy-Weinberg equilibrium were not evident for most loci in most populations (Appendix A4). However, four instances occurred where one locus showed significant deviations from Hardy-Weinberg equilibrium in one population, and two loci showed significant deviations in three populations (Appendix A4). In most cases, observed heterozygosity was lower than expected heterozygosity except for one locus in the Aleutian Islands and two loci in Guadalupe Island where observed heterozygosity was higher than expected under Hardy-Weinberg equilibrium (Appendix A4). The significant deviations from Hardy-Weinberg equilibrium were not consistent across populations or across loci, thus all loci were included in subsequent analyses. The more conservative tests implemented in BOTTLENECK detected heterozygosity excess in the southern population (p_{Sign}=0.03,

$p_{Wilcoxin}$=0.01) but failed to detect heterozygosity excess in the northern population (p_{Sign}=0.51, $p_{Wilcoxin}$=0.24).

Population genetic structure of microsatellites

Genetic distance estimated from microsatellites was not correlated with geographic distance globally (Mantel's test, R=0.25, p=0.09, Figure 2) but the F_{ST} values were very low throughout the range. Significant pairwise F_{ST} values occurred between sampling sites in the northern versus southern parts of the range (Table 7). Significant pairwise F_{ST} estimates within *P. a. aleuticus* were not evident except that individuals breeding at Channel Islands were significantly genetically differentiated from some sampling sites farther north (Table 7). Cassin's auklets from Guadalupe Island were significantly differentiated from those from all other sampling sites except San Jeronimo Island (Table 7). Population genetic variation was best characterized when Channel Islands was grouped with *P. a. australe* as more of the genetic variation was described between groups (F_{CT}=0.018, p=0.0049) compared to within groups (F_{SC}=0.0041, p=0.051, Table 6).

Using the program STRUCTURE, the K value that best described the genotypic data was K=2. For the "admixture" analysis, the highest ΔK value as estimated by Evanno's method was 27.10 at K=2 compared to ΔK values of 3.75 and 18.56 for K=3 and K=4, respectively. Although Evanno's method cannot evaluate a ΔK value for K=1, the likelihood of K=2 (-9207.72) was higher than the likelihood of K=1 (-9200.61). The "admixture" analysis displayed weak structure and all individuals assigning to the same cluster did not group together when compared to sampling location (Figure 4a). Similarly, for the no-admixture analysis with "locprior", the highest ΔK value was 54.08 at K=2 compared to ΔK values of 1.64 and 4.53 for K=3 and K=4,

respectively. The likelihood of K=1 (-9200.70) was lower than the likelihood of K=2 (-9034.86) as well. However, the "no-admixture" analysis with "locprior" displayed strong population genetic structure with individuals from the Channel Islands having a high probability of assigning to the same cluster as *P. a. australe* individuals (Figure 4c). In addition, K=1 had the highest likelihood when the two genetic clusters were analysed separately in STRUCTURE suggesting no sub-structuring within the two genetic clusters.

Estimation of gene flow

To reach convergence (ESS values over 200), the maximum estimates of the priors from the control region analysis in IMa were as follows: θ_A=10, $\theta_{1(\text{northern})}$=50, $\theta_{2(\text{southern})}$=50, t=10, $m_{1(\text{southern into northern})}$=0.5 and $m_{2(\text{northern into southern})}$=0.5. The estimated number of migrants from the southern genetic group into the northern genetic group going forward in time was not zero at 0.27 (0.01-2.23) migrants per generation, whereas 0.47 (0.001-2.26) migrants per generation were estimated from the northern genetic group into the southern genetic group (Table 8, Figure 5). Using the 3.3% per million years divergence rate estimate for control region (20%/my – 2%/my in parentheses), the time of divergence was estimated at 130,483 (21,530 - 215,297) years ago (see Table 8 for HPD90 intervals, Figure 5). The effective female size for the ancestral population was estimated at 11,226 (1,850 - 18,497) individuals. The female effective size for the northern population was not different from the southern population at 76,541 (12,611 - 126,114) individuals versus 37,987 (6,259 - 62,591) individuals, respectively. The microsatellite analysis in BAYESASS resulted in non-migration rates that were not sufficient to provide robust migration rates. BAYESASS identified the potential migrants (indicated by a mismatch mitochondrial haplotype) as second generation migrants.

Table 4. Estimates of haplotype diversity (h ± standard deviation), nucleotide diversity (π ± standard deviation) and results of tests for deviations from selective neutrality (Fu's F_S,Tajima's D, Ewens-Watterson's, and Chakraborty's tests) for each sampling site. Sampling site abbreviations are the same as listed in Table 1.

Sampling site	h	π	F_S	D	E-W*	Chakraborty$^{\gamma}$
Northern population	0.90 ±	0.006 ±	-25.70	-1.15	0.10/0.04	49/23.2
	0.02	0.003	p<0.001	p=0.11	p=1.00	p<0.001
AL	0.97 ±	0.008 ±	-11.37	-0.46	0.06/0.06	22/21.74
	0.02	0.004	p<0.001	p=0.38	p=0.71	p=0.15
SA&FR	0.86 ±	0.004 ±	-0.93	0.17	0.20/0.21	7/7.02
	0.07	0.003	p=0.27	p=0.61	p=0.59	p=0.23
RE	0.82 ±	0.006 ±	-1.54	0.42	0.23/0.16	9/6.50
	0.10	0.004	p=0.19	p=0.70	p=0.99	p=0.08
TR	0.78 ±	0.005 ±	-1.68	-0.61	0.24/0.16	11/7.35
	0.07	0.003	p=0.22	p=0.31	p=0.95	p=0.04
WA	0.89 ±	0.005 ±	-1.16	-0.71	0.22/0.20	6/5.63
	0.11	0.003	p=0.19	p=0.27	p=1.00	p=0.31
SF	0.94 ±	0.006 ±	-7.85	-0.74	0.10/0.08	18/15.9
	0.03	0.004	p=0.001	p=0.25	p=0.92	p=0.11
Southern population	0.97 ±	0.007 ±	-25.66	-0.19	0.04/0.04	44/41.39
	0.01	0.004	p<0.001	p=0.50	p=0.78	p=0.08
CH	0.96 ±	0.006 ±	-5.46	-0.24	0.08/0.10	14/15.68
	0.02	0.003	p=0.01	p=0.45	p=0.08	p=0.13
SJ	0.96 ±	0.007 ±	-2.94	-0.57	0.12/0.13	9/9.26
	0.05	0.004	p=0.06	p=0.30	p=0.75	p=0.30
GU	0.96 ±	0.008 ±	-7.20	-0.15	0.08/0.07	19/18.7
	0.02	0.004	p=0.01	p=0.51	p=0.71	p=0.16
SB	0.93 ±	0.006 ±	-4.55	-0.16	0.10/0.11	15/14.92
	0.03	0.004	p=0.02	p=0.47	p=0.55	p=0.16

*First number indicates the observed F value and the second number indicates the expected F value

$^{\gamma}$First number indicates the number of observed haplotypes and the second number indicates the number of expected haplotypes under neutral evolution

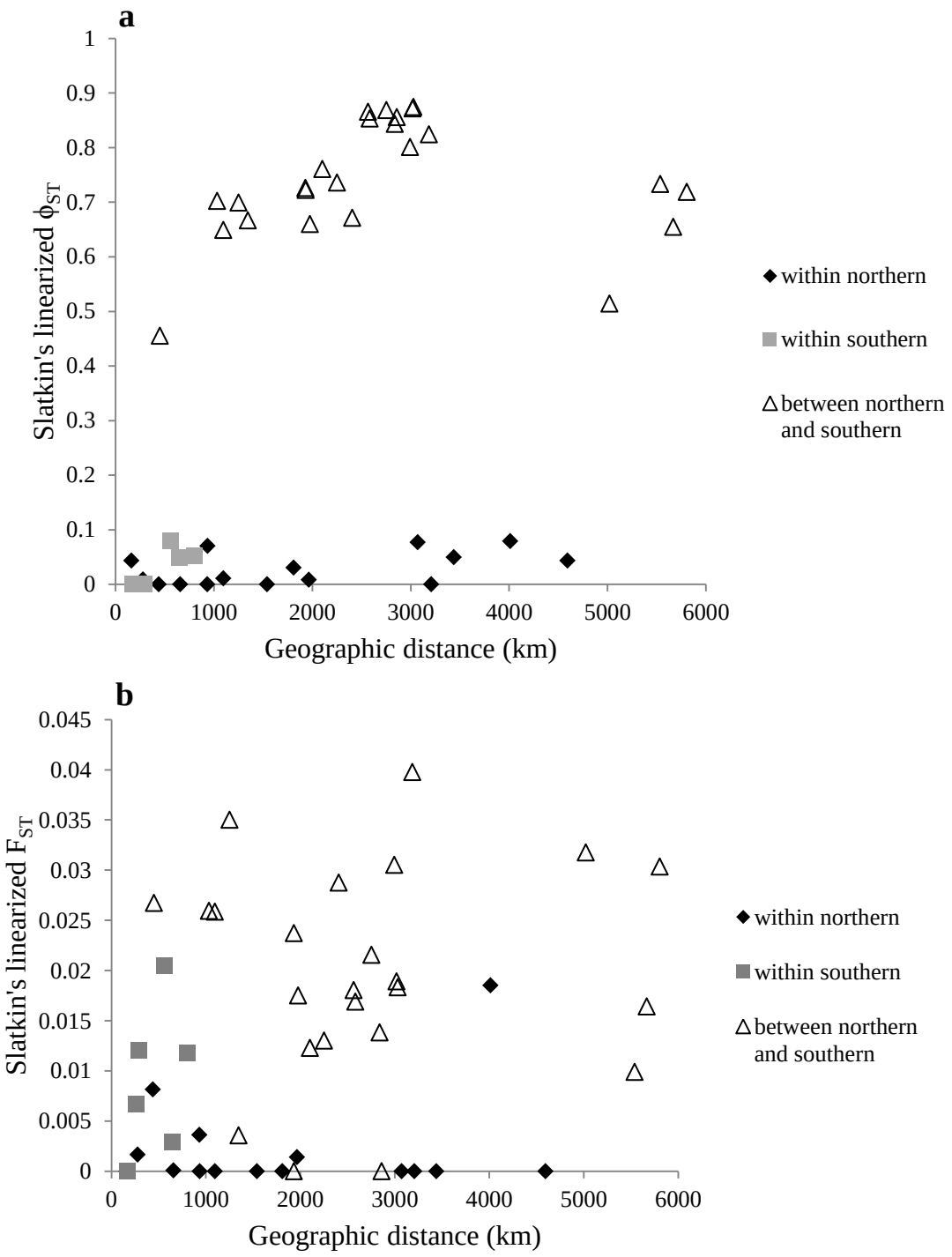


Figure 2. Genetic distance (Slatkin's linearized ϕ_{ST}/F_{ST}) versus geographic distance (km) between sampling sites of Cassin's auklet for a) control region and b) microsatellites. "Northern" and "southern" refer to the two genetic populations of Cassin's auklet as diagnosed using STRUCTURE.

Table 5. Pairwise ϕ_{ST} estimates from the mitochondrial control region. One asterisk (*) represents a significant p-value with an α value of 0.05 whereas two asterisks (**) represents significance after a Benjamini-Yekutieli correction ($p<0.011$). Sampling site abbreviations are the same as given in Table 1.

		P. a. aleuticus							*P. a. australe*	
		AL	SA&FR	RE	TR	WA	SF	CH	SJ	GU
P. a. aleuticus	AL									
	SA&FR	0.07*								
	RE	-0.03	0.04							
	TR	0.05	-0.02	0.01						
	WA	0.07	0.01	0.07	-0.01					
	SF	0.04*	0.01	0.03	-0.01	-0.00				
	CH	0.40**	0.47**	0.46**	0.46**	0.42**	0.39**			
P. a. australe	SJ	0.42**	0.44**	0.46**	0.46**	0.42**	0.41**	-0.03		
	GU	0.42**	0.45**	0.47**	0.46**	0.43**	0.41**	-0.04	0.00	
	SB	0.34**	0.40**	0.42**	0.40**	0.40**	0.31**	0.05	0.07*	0.05

Table 6. Results of the hierarchical analyses of molecular variance (AMOVA) for the control region and microsatellites. Sampling sites were grouped by subspecies, by the results of pairwise ϕ_{ST} and F_{ST} estimates (CH grouping with *P. a. australe*), or with AL in a separate group. Significance estimated based on α=0.05. Sampling site abbreviations are the same as listed in Table 1.

	control region			microsatellites		
Grouping	ϕ_{ST}	ϕ_{CT}	ϕ_{SC}	F_{ST}	F_{CT}	F_{SC}
By subspecies (AL, SA&FR, RE, TR, WA, SF and CH vs. SJ, GU and SB)	0.42 p<0.001	0.35 p=0.01	0.11 p<0.001	0.02 p<0.001	0.02 p=0.03	0.01 p<0.001
CH grouping with *P. a. australe* (AL, SA&FR, RE, TR, WA and SF vs. CH, SJ, GU and SB)	0.43 p<0.001	0.41 p=0.01	0.02 p=0.04	0.02 p<0.001	0.02 p<0.001	0.00 p=0.05
AL genetically differentiated? (AL* vs. SA&FR, RE, TR, WA and SF)	0.05 p=0.08	0.05 p=0.17	0.00 p=0.29	-0.00 p=0.69	-0.00 p=1.00	0.00 p=0.58

*in the third scenario, AL is the only sampling location in the group, thus ϕ_{SC}/ F_{SC} values are unreliable.

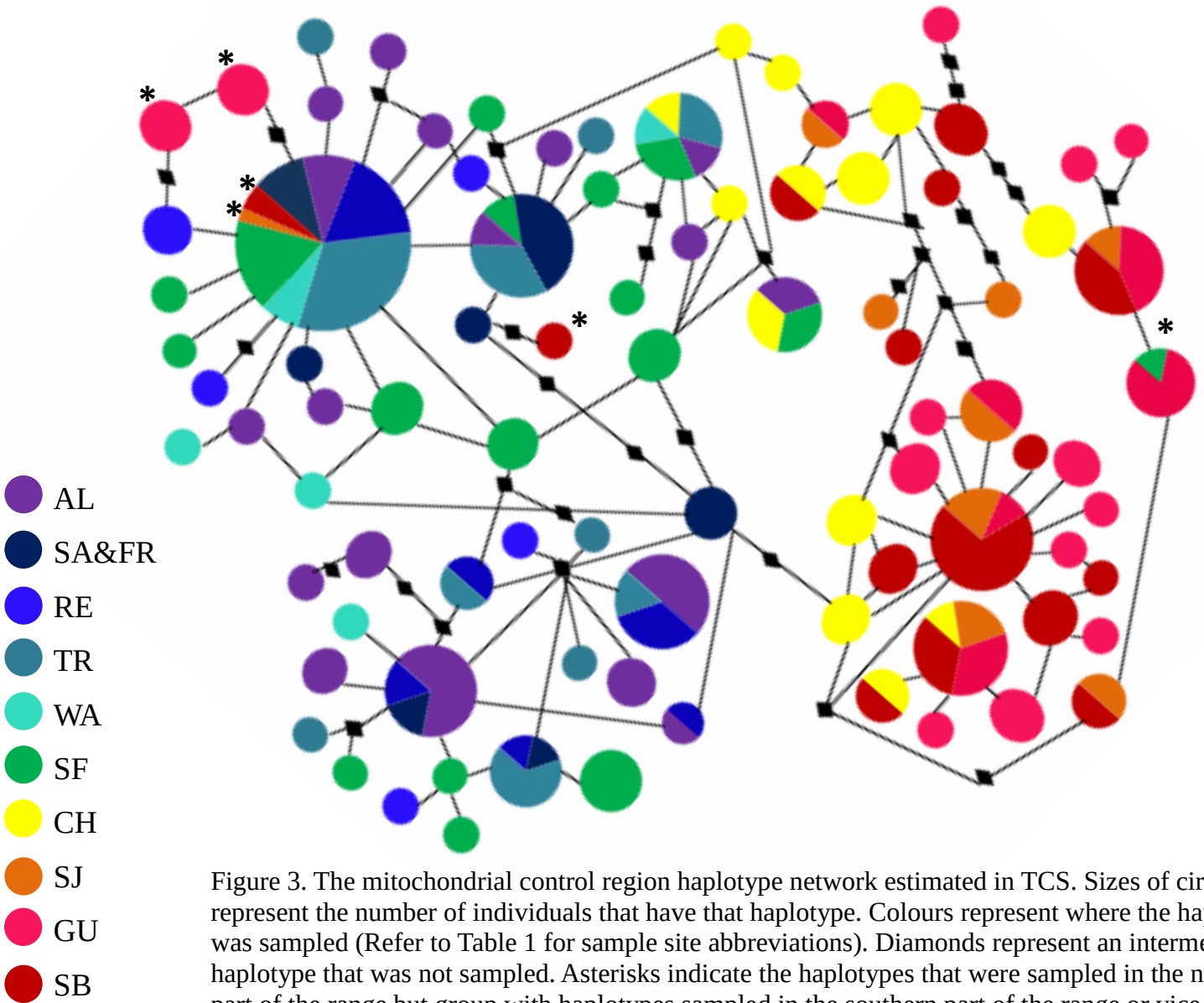


Figure 3. The mitochondrial control region haplotype network estimated in TCS. Sizes of circles represent the number of individuals that have that haplotype. Colours represent where the haplotype was sampled (Refer to Table 1 for sample site abbreviations). Diamonds represent an intermediate haplotype that was not sampled. Asterisks indicate the haplotypes that were sampled in the northern part of the range but group with haplotypes sampled in the southern part of the range or vice versa.

Table 7. Pairwise F_{ST} estimates from microsatellites. One asterisk (*) represents a significant p-value with an α value of 0.05 whereas two asterisks (**) represents significance after a Benjamini-Yekutieli correction (p<0.011). Sampling site abbreviations are the same as given in Table 1.

	P. a. aleuticus							*P. a. australe*		
	AL	SA&FR	RE	TR	WA	SF	CH	SJ	GU	SB
AL										
SA&FR	-0.00									
RE	-0.00	-0.00								
TR	-0.01	0.01	0.00							
WA	0.02	-0.01	-0.01	0.00						
SF	-0.01	0.00	-0.00	-0.00	0.00					
CH	0.03**	0.03**	0.01	0.02**	0.00	0.03**				
SJ	0.02	0.02*	-0.00	0.02*	-0.01	0.03*	0.00			
GU	0.01*	0.03**	0.01*	0.02**	0.02*	0.03**	0.02**	0.01		
SB	0.03**	0.04**	0.02*	0.02**	0.01	0.03*	0.01*	-0.00	0.01*	

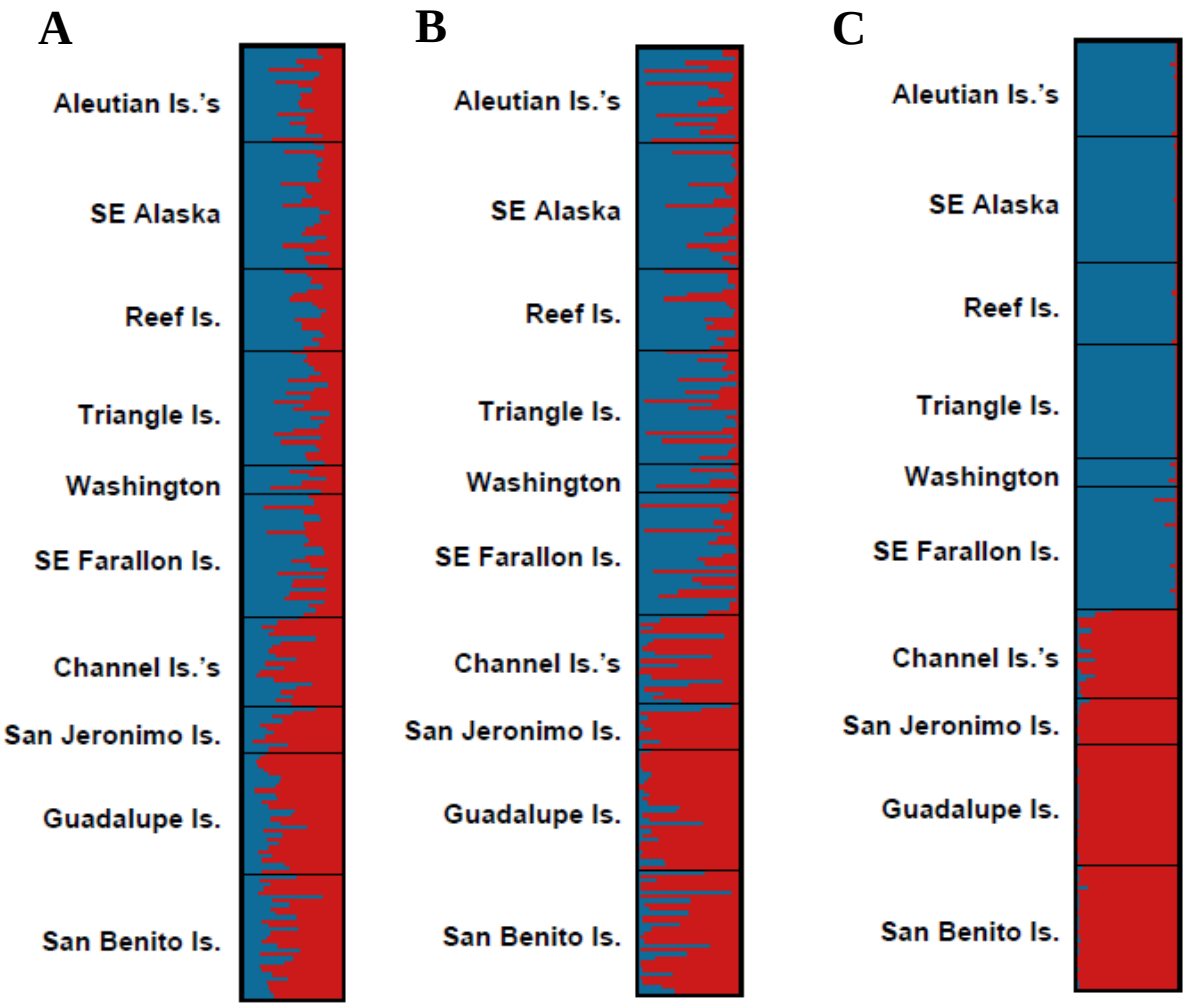


Figure 4. Bar plots from STRUCTURE showing individual "admixture" coefficients (A) and probabilities of assignment of individuals to two genetic clusters (B and C) for K=2 for A) "admixture" analysis, B) "no admixture" analysis, and C) "no admixture" analysis with "locprior".

Table 8. Time since divergence and female effective population sizes of the northern, southern and ancestral populations from IMa. Results are scaled by the sequence divergences of 2%, 3.3% and 20% per million years. The HPD90Lo and HPD90Hi are given in parentheses.

	2% per million years	3.3% per million years	20% per million years
Time since divergence in years	215,297 (141,360-569,405)	130,483 (85,673-345,094)	21,530 (14,136-56,941)
Female effective population size – northern population	126,114 (82,435-231,399)	76,541 (50,031-140,440)	12,611 (8,244-23,140)
Female effective population size – southern population	62,591 (36,736-100,311)	37,987 (22,296-60,881)	6,259 (3,674-10,031)
Female effective population size – ancestral population	18,497 (3,005-49,845)	11,226 (1,824-30,252)	1,850 (301-4,985)

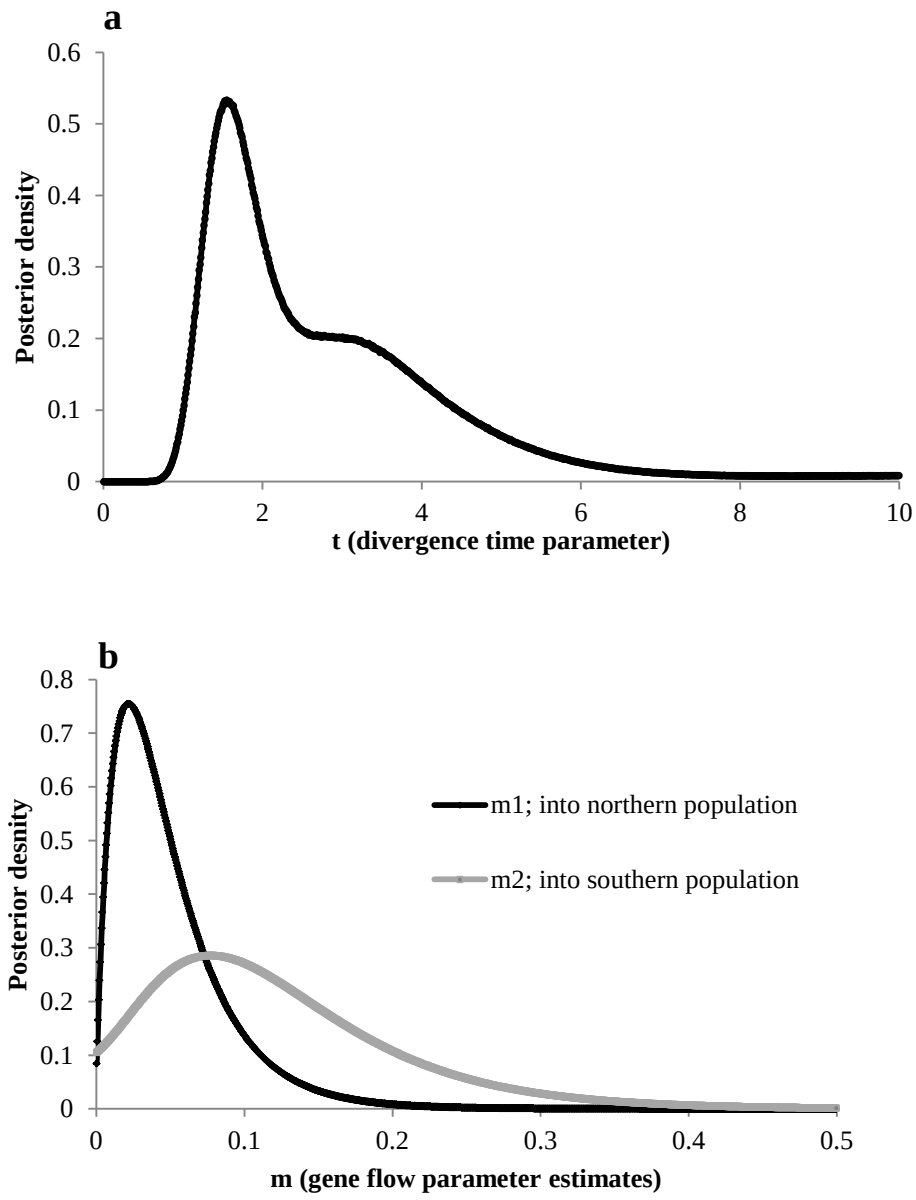


Figure 5. Posterior density distributions from IMa for Cassin's auklet a) divergence time (t) between the northern and southern genetic populations and b) gene flow estimate (m) between the two populations.

Chapter 4: Discussion

Both the control region and microsatellite analyses indicate that Cassin's auklets have significant population genetic structure, characterized mainly by two genetic clusters. These genetic groups roughly map to the original two subspecies ranges, except that individuals breeding at Channel Islands originally were grouped with *P. a. aleuticus* but are more similar to *P. a. australe* in their control region sequences and microsatellite allele frequencies. Population genetic differentiation was not evident within either the northern or southern genetic population, suggesting that the isolation by distance effect is likely due to the strong differentiation between the northern and southern genetic groups. Despite the strong population genetic differentiation between the two genetic clusters, evidence shows that gene flow has occurred both northward and southward directions since the time of divergence. Evidence and possible mechanisms for this divergence are detailed below.

Population genetic structure throughout the range

Results of the control region analyses provided strong support for the existence of two genetic populations (Aleutian Islands to Southeast Farallon Island and Channel Islands to San Benito Island). Pairwise ϕ_{ST} estimates were highest and significantly different from zero (0.31-0.47) between sampling sites in the northern and southern genetic populations. More of the genetic variation was characterized among groups rather than within groups when Channel Islands was grouped with *P. a. australe* in a hierarchical AMOVA. In addition, haplotypes from the two genetic populations tended to group together in the haplotype network. However, the two groups were not reciprocally monophyletic and the few haplotypes that were sampled in one

subspecies' range but grouped with the other subspecies may be a result of either retained ancestral variation or secondary contact.

The microsatellite analyses also supported two genetic populations within the Cassin's auklet breeding range, but the division between the populations was not as clear. Pairwise F_{ST} estimates were low throughout the range but were significantly different from zero for pairwise estimates between the northern and southern genetic populations including Channel Islands. Significant genetic variation was not found between sampling sites within groups when Channel Islands was grouped with sampling sites farther south, and sampling sites farther north grouped together in the AMOVA analysis. Moreover, the STRUCTURE analysis supported two genetic clusters but individuals from the same sampling site were not always assigned to the same cluster. The "admixture" model used for the unbiased Bayesian assignment method in STRUCTURE is likely the most appropriate model for wild populations as it allows mixed ancestry of individuals and does not require distinct populations (Pritchard *et al*. 2000). The analysis with "admixture" resulted in K=2 having the highest likelihood of representing the genotypic data but did not define any distinct geographic clusters for Cassin's auklet. STRUCTURE has difficulty assigning individuals to the correct cluster if population genetic structure is weak or if genetic variation is continuous (i.e. isolation by distance; Waples and Gaggiotti 2006, Hubisz *et al.* 2009). The analysis with "no-admixture" can usually detect weaker structure as it assigns a probability to each individual of assigning to one cluster or another (Pritchard *et al*. 2000). The "no-admixture" analysis found two genetic clusters as having the highest likelihood but individuals from the same sampling site did not always have a high probability of assignment to the same cluster. When a "locprior" was used in the analysis, individuals assigned with high probability to the same cluster as other indivduals from the same

sampling site. Some researchers consider that results using "locprior" are unreliable because sampling effects could bias the results (J. Morris-Pocock *pers. comm*.) but other researchers suggest the model with "locprior" can provide useful information when weak structure is present (Hubisz *et al*. 2009). The sampling information that was used in "locprior" for this analysis was the sampling site (1-10), not the subspecies, so any spatial grouping bias would involve sampling sites not subspecies. Furthermore, other analyses in this study also supported the existence of two genetic clusters.

Results from IMa imply that on-going gene flow has occurred between the two genetic populations since the time of divergence suggesting that the pattern of variation seen in the haplotype network is a result of some contemporary gene flow. A tendency was evident for more migrants to enter the southern population (Channel Islands to San Benito Island) from the northern population (Aleutian Islands to Southeast Farallon Island) rather than the reverse. Thus, divergence of the two genetic populations appears to have occurred with asymmetrical gene flow, but confidence intervals overlap in the IMa estimates.

The IMa results suggest that despite gene flow, the two genetic populations diverged between 21,530 - 215,297 years ago. The large range is a result of the range of mutation rates used to calculate the time estimates. These estimates are consistent with previous estimates of divergence within the Alcidae. Friesen *et al.* (1996) estimated that the family Alcidae diverged from other charadriiform birds around 11.8-7.4 mya using cytochrome *b* and allozyme loci.

Population genetic structure at a local scale

Several results suggest that neither the northern or southern genetic populations are in migration/drift or mutation/drift equilibrium: 1) results from Fu's F_S test suggested that an excess

number of rare haplotypes exist in both populations, which may be a consequence of a deviation from neutral evolution or a recent population expansion, 2) a non-zero growth parameter had a higher likelihood in both populations in FLUCTUATE, 3) the weak correlation between genetic and geographic distance suggested that gene flow homogenized local sampling sites but genetic drift had more of an influence on distant sites (Hutchinson and Templeton 1999), and 4) heterozygosity excess was detected by BOTTLENECK in the southern population.

Although evidence suggests that the two genetic populations of Cassin's auklet are strongly differentiated, evidence does not support further genetic division of the two populations. Population genetic differentiation was not detected within either of the two genetic populations of Cassin's auklets: 1) pairwise ϕ_{ST} and F_{ST} estimates were very low and most did not significantly differ from zero, 2) haplotypes sampled within the same sampling site did not group together in the haplotype network, and 3) STRUCTURE did not detect further sub-structuring within the two genetic clusters.

The control region and microsatellite analyses did not provide evidence for Cassin's auklet individuals breeding at the Aleutian Islands being genetically differentiated from individuals breeding farther south. Pairwise ϕ_{ST} estimates between the Aleutian Islands and other sampling sites in the northern part of the range were generally low and not statistically significant suggesting that the gap in the breeding distribution in the Gulf of Alaska did not result in population genetic structure. In addition, significant genetic differentiation was not found by the AMOVA when the Aleutian Islands were grouped separately from sampling sites from Southeast Alaska/Frederick Island to Southeast Farallon Island. These results are unusual compared to other species that breed in the Aleutian Islands and Gulf of Alaska. For example, marbled murrelets (*Brachyramphus marmoratus*) exhibited population genetic differentiation

between the Aleutian Islands and other areas (Congdon *et al.* 2000). More specifically, marbled murrelets were genetically differentiated between the western Aleutian Islands, central Aleutian Islands and mainland Alaska (Friesen *et al*. 2005). Pigeon guillemots (*Cepphus columba*) also exhibited high population genetic structure throughout their range on the Pacific coast of North America, with population genetic variation based on mitochondrial and nuclear DNA apportioned into three groups: Alaska, British Columbia and California/Oregon (Friesen and Piatt 2003).

Within the southern genetic population, individuals breeding at Guadalupe Island were weakly differentiated at microsatellites from individuals elsewhere. Pairwise F_{ST} estimates were low but significantly different from zero between Guadalupe Island and other sampling sites except for San Jeronimo Island. These results differ from those for Guadalupe Island Xantus's murrelet (*S. h. scrippsi*) which is genetically distinct from individuals elsewhere (*S. h. hypoleucus*).

Mechanisms of divergence

Ongoing gene flow between the two genetic groups of Cassin's auklet may be related to Cassin's auklets' non-breeding distribution. In a review of phylogeographic studies of seabirds, Friesen *et al*. (2007) found an important correlate of population genetic structure was the non-breeding distribution: e.g., seabird species that have more than one non-breeding area tend to exhibit phylogeographic structure. However, if Cassin's auklets disperse to a common area during the non-breeding season, interactions could facilitate population admixture. For example, Burg and Croxall (2001) did not find population genetic structure in the grey-headed albatross (*Thalassarche chrysostoma*), an oceanic species that does not have population specific non-

breeding sites. Auklets have an extended molt period allowing for long-distance dispersal from colonies after breeding in search of food resources (Adams *et al.* 2004). Renner *et al*. (2008) reported a decrease in the number of planktivorous seabirds on the Aleutian Islands during winter months compared to summer months. Copepods and some euphausiid species go into diapause during the winter which results in less food available during the non-breeding season (Renner *et al*. 2008). In addition, storms in the Gulf of Alaska during the winter months cause downwellings along the Aleutian arch (Day 2006 and references therein). Although migration routes of Cassin's auklets breeding at the northern colonies have not been tracked, inferences indicate that northern Cassin's auklets travel south during the non-breeding season (Briggs *et al*. 1987, Ainley *et al*. 2011). Population numbers of Cassin's auklets have been documented to rise in California after the breeding season and the additions likely originated from colonies farther north where copepods are depleted (Briggs *et al*. 1987). In contrast, Cassin's auklet individuals breeding in central California do not disperse far during the non-breeding season (Ainley and Boekelheide 1990). Radio-marked Cassin's auklets breeding at the Channel Islands dispersed northward approximately 600 km to a region north of the Farallon Islands and Point Conception after the breeding season likely due to the predictable upwellings in this area (Adams *et al.* 2004). As a result, population admixture could occur between the northern colonies and the southern colonies in the over-wintering area in California (e.g. Friesen *et al*. 2007). Cassin's auklets in southern California and Baja California are resident throughout the year (Adams *et al*. 2004) and breed more asynchronously from November through to March which could cause these individuals to become isolated from individuals further north.

Population genetic structure in Cassin's auklets also may be a function of home range and foraging range. Populations foraging in different areas could result in restricted gene flow among

populations (Friesen *et al.* 2007). For example, black-browed albatrosses (*Thalassarche melanophris* and *T. impavida*) exhibit population genetic structure which may be due to the distinct foraging areas (Burg and Croxall 2001). During the breeding season, Cassin's auklets throughout the range do not forage far from their breeding colony (Adams *et al*. 2004, Boyd *et al*. 2008). For example, the marine distribution of breeding Cassin's auklets at Triangle Island during the 1999-2001 breeding seasons was approximately 650-1400 km^2 (Boyd *et al*. 2008). Similarly, on average, Cassin's auklets breeding on Triangle Island were found foraging 50-80 km from the breeding colony, in either a southwest or northwest direction (Boyd *et al.* 2008). Moreover, Adams *et al*. (2004) used radio-telemetry to conclude that Cassin's auklets breeding at the northern Channel Islands foraged under 30 km on average from the colony, usually in a northward direction to the Santa Barbara Channel. The restricted foraging ranges may lead to more pronounced population genetic structure within the species (Friesen *et al.* 2007). For example, Xantus's murrelet forages at a shorter geographic distance than the distance between Guadalupe Island and other sampling sites, suggesting the geographic distance may have resulted in isolation of the Guadalupe Island population (Birt *et al.* 2012).

Although Cassin's auklets exhibit natal philopatry, this strategy could become non-adaptive if food resources decrease near the natal colony (Pyle 2001). In El Niño years when the upwellings were reduced, many Cassin's auklets at Southeast Farallon Island appeared to skip breeding for the year, decreasing population numbers at the colony (Pyle *et al.* 2001). Individuals may disperse to breed at neighbouring colonies if food availability is limited (Pyle *et al.* 2001). In addition, many studies show that dispersal in birds is female-biased (Hall *et al.* 2009). Female Cassin's auklets have a larger home range in comparison to males as females have a higher energetic cost at the beginning of the breeding season and may need to forage at a farther

distance (Adams *et al*. 2004). Therefore, estimates of female-mediated gene flow may overestimate the overall population migration rate.

Conservation implications for Cassin's auklets and Baja California

Evidence from the control region and microsatellite analyses suggest that the two genetic groups of Cassin's auklets should not be considered separate ESUs. Mitochondrial haplotypes were not reciprocally monophyletic in the haplotype network and thus not consistent with the most common requirement for delineating ESUs (Moritz 1994). In addition, gene flow has occurred since the time of divergence, indicating the two genetic populations are not reproductively isolated. However, based on previous ecological evidence and the mitochondrial and nuclear analyses in this study, the two genetic populations are likely separate management units. A common rule used to define management units is that a migration rate less than 0.1 means the two populations are demographically independent (Waples and Gaggiotti 2006). The genetic differentiation of the two populations with some asymmetrical on-going gene flow (northern into the southern population) suggests that secondary contact may have facilitated gene flow between two separate management units.

Traditionally, individuals breeding in the Channel Islands were part of the northern subspecies *P. a. aleuticus*. However, based on the control region and microsatellite analyses, individuals breeding at the Channel Islands are more likely to belong to the southern subspecies *P. a. australe*. Haig and Winker (2010) summarized the issue of conservation practices based on avian subspecies status and concluded that management decisions can be made on the subspecies level if the distinction is based on multiple characters. Therefore, the subspecies designations should be revised to match the northern and southern genetic groups diagnosed in this study.

The population genetic differentiation of Cassin's auklet individuals breeding in southern California and Baja California adds to the knowledge that Baja California is a biodiversity hotspot. However, in Cassin's auklets, a semi-permeable barrier to gene flow seems to occur farther north, thus southern California may need to be included in the biodiversity hotspot designation. The differentiation of individuals in Baja California and southern California occurs across many taxa (e.g. see Riemann and Ezcurra 2005, Allen 2008, Garcillán *et al*. 2008), suggesting a potential consistent barrier occurs in this area.

Future directions

To fully understand how and why the two genetic populations of Cassin's auklets are diverging, sampling or tracking Cassin's auklets in central California during the non-breeding season could indicate population assignments of individuals in the over-wintering region. The existence of population-specific non-breeding areas could support the hypothesis that non-breeding distribution strongly influences population genetic structure in seabirds (Friesen *et al.* 2007). Investigating the population assignments of non-breeding individuals (north versus south) may help in determining where the barriers to gene flow exist, important for understanding why individuals in Baja California are genetically differentiated.

Using neutral markers to provide insight into the interplay between genetic drift and gene flow during the divergence of populations can be useful because population genetic structure is the variation on which selection can act on (Funk *et al.* 2012). However, sometimes conserving contemporary adaptive differences as well as remnants of historical variation is critical (Crandall *et al.* 2000). Studies using microsatellites have the power to delineate ESUs or MUs but cannot characterize adaptive variation (Funk *et al.* 2012). A genomics approach could characterize both

neutral and adaptive variation and is becoming more popular in population genetic studies with the advances in next-generation sequencing techniques (Funk *et al.* 2012).

One limitation to traditional population genetics studies is that the evolutionary history of the species is often inferred based on contemporary population genetic variation and structure. Analyzing markers with high mutation rates and high rates of homoplasy could mislead the inferences about the often complex evolutionary histories of species. Approximate Bayesian Computation (ABC) can simulate genetic data in complex scenarios and compare summary statistics to the observed data (Beaumont *et al*. 2002, Cornuet *et al.* 2008). One advantage to ABC is that it can incorporate more complex demographic models (e.g. constant gene flow since the time of divergence versus gene flow only during a certain period of time), providing a potentially clearer explanation of how and why populations diverge.

Literature Cited

Abraham CL, Sydeman WJ (2004) Ocean climate, euphausiids and auklet nesting: inter-annual trends and variation in phenology, diet and growth of a planktivorous seabird, *Ptychoramphus aleuticus*. *Marine Ecology Progress Series,* **274**:235-250.

Adams J, Takekawa JY, Carter HR (2004) Foraging distance and home range of Cassin's auklets nesting at two colonies in the California Channel Islands. *Condor,* **106**:618-637.

Ainley D, Manuwal DA, Adams J, Thoresen AC (2011) Cassin's Auklet (*Ptychoramphus aleuticus*), *The Birds of North America Online* (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/050.

Ainley DG, Boekelheide RJ (Eds.) (1990) Seabirds of the Farallon Islands: Ecology, dynamics, and structure of an upwelling-system community. Stanford University Press, Stanford, California.

Allen GR (2008) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**:541-556.

Baker AJ, Marshall HD (1997) Mitochondrial control region sequences as tools for understanding evolution. In: Mindell DP (Ed.) *Avian Molecular Evolution and Systematics*. Academic Press, San Diego, California.

Baker AJ, Pereira SL, Paton TA (2007) Phylogenetic relationships and divergence times of Charadriiformes genera: multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. *Biology Letters*, **3**:205-210.

Bailey E (1993) Introduction of foxes to Alaskan Islands – History, effects on avifauna, and eradication. *United States Department of the Interior Fish and Wildlife Service: Resource Publication*. Washington, D.C.

Barton DC, Lindquist KE, Henry RW III, Luna Mendoza LM (2004) Landbird and waterbird notes from Isla Guadalupe, Mexico. *Western Birds,* **35**:186-196.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in population genetics. *Genetics,* **162**:2025-2035.

Bertram DF, Mackas DL, McKinnell SM (2001) The seasonal cycle revisted: interannual variation and ecosystem consequences. *Progress in Oceanography,* **49**:283-307.

Bertram DF, Harfenist A, Hedd A (2009) Seabird nestling diets reflect latitudinal temperature-dependent variation in availability of key zooplankton prey populations. *Marine Ecology Progress Series*, **393**:199-210.

Birt TP, Carter HR, Whitworth DL, McDonald A, Newman SH, Gress F, Palacios E, Koepke JS, Friesen VL (2012) Rangewide population genetic structure of Xantus's murrelet (*Synthliboramphus hypoleucus*). *Auk*, **129**:44-55.

Boyd WS, Tranquilla LM, Ryder JL, Shisko SG, Bertram DF (2008) Variation in marine distribution of Cassin's auklets (*Ptychoramphus aleuticus*) breeding at Triangle Island, British Columbia. *Auk*, **125**:158-166.

Briggs KT, Ainley DG, Spear LB, Adams PB, Smith SE (1987) Distribution and diet of Cassin's auklet and common murre in relation to central California Upwellings. In: Ouellet H (Ed.) *Acta XIX Congress on International Ornithology*. University of Ottawa Press, Ottawa.

Byrd GV, Renner HM, Renner M (2005) Distribution patterns and population trends of breeding seabirds in the Aleutian Islands. *Fisheries Oceanography*, **14**:139-159.

Burg TM, Croxall JP (2001) Global relationahips amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Molecular Ecology*, **10**:2647-2660.

Castoe TA, Poole AW, de Koning APJ, Jones KL, Tomback DF, Oyler-McCance SJ, Fike JA, Lance SL, Streicher JW, Smith EN, Pollock DD (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, **7**:e30953.

Chakraborty R (1990) Mitochondrial DNA polymorphism reveals heterogeneity within some Asian populations. *American Journal of Human Genetics*, **47**:87-94.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**:1657-1660.

Cockerham CC (1969) Variance of gene frequencies. *Evolution*, **23**:72-84.

Congdon BC, Piatt JF, Martin K, Friesen VL (2000) Mechanisms of population differentiation in marbled murrelets: Historical versus contemporary processes. *Evolution*, **54**:974-986.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**:2001-2014.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, Guillemaud T, Estoup A (2008) Inferring population history with *DIY ABC*: a user-friendly approach to approximate Bayesian computation. *Bioinformatics,* **24**:2713-2719.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution,* **15**:290-295.

Dawson MN (2001) Phylogeography in coastal marine animals: a solution from California? *Journal of Biogeography*, **28**:723-736.

Dawson DA, Hunter FM, Pandhal J, Buckland R, Parham A, Jones IL, Bradshaw M, Jehle R, Burke T (2005) Assessment of 17 new whiskered auklet (*Aethia pygmaea*) microsatellite loci in 42 seabirds identifies 5-15 polymorphic markers for each of nine Alcinae species. *Molecular Ecology Notes,* **5**:289-297.

Day RH (2006) Seabirds in the northern Gulf of Alaska and adjacent waters, October to May. *Western Birds,* **37**:190-214.

De Jong MA, Wahlberg N, van Eijk M, Brakefield PM, Zwaan BJ (2011) Mitochondrial DNA signature for range-wide populations of *Bicyclus anynana* suggests a rapid expansion from recent refugia. *PLoS ONE,* **6**:e21385.

Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome: A novel gene order in higher vertebrates. *Journal of Molecular Biology*, **212**:599-634.

Dewoody J, Nason JD, Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes,* **6**:951-957.

Estoup A, Jarne P, Cornuet J (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**:1591-1604.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**:2611-2620.

Ewens WJ (1972) Sampling theory of selectively neutral alleles. *Theoretical Population Biology,* **3**:87-112.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform populationgenetics analyses under Linux and Windows. *Molecular Ecology Resources,* **10**:564-567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**:1567–1587.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology,* **16**:1149-1166.

Fraser D, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**:2741-2752.

Friesen VL, Birt TP, Piatt JF, Golightly RT, Newman SH, Hébert PN, Congdon BC, Gissing G (2005) Population genetic structure and conservation of marbled murrelets (*Brachyramphus marmoratus*). *Conservation Genetics*, **6**:607-614.

Friesen VL, Burg TM, McCoy KD (2007) Mechanisms of population differentiation in seabirds. *Molecular Ecology*, **16**:1765-1785.

Friesen VL, Piatt JF (2003) A genetic study to aid in restoration of murres, guillemots and murrelets to the Gulf of Alaska. *Exxon Valdez* Oil Spill Restoration Project Final Report (Restoration Project 00169). Alaska Biological Sciences Center, U.S. Geological Survey, Anchorage, Alaska.

Friesen VL, Piatt JF, Baker AJ (1996) Evidence from cytochrome B sequences and allozymes for a 'new' species of Alcid: The long-billed murrelet (*Brachyramphus perdix*). *Condor*, **98**:681-690.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics,* **147**:915-925.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, **27**:489-496.

Garcillán PP, Ezcurra E, Vega E (2008) Guadalupe Island: Lost paradise recovered? Overgrazing impact on extinction in a remote oceanic island as estimated through accumulation functions. *Biodiversity and Conservation,* **17**:1613-1625.

Gaston AJ, Bertram DF, Boyne AW, Chardine JW, Davoren G, Diamond AW, Hedd A, Montevecchi WA, Hipfner JM, Lemon MJF, Mallory ML, Rail J-F, Robertson GJ (2009) Changes in Canadian seabird populations and ecology since 1970 in relation to changes in oceanography and food webs. *Environmental Reviews,* **17**: 267-286.

Gaston AJ, Jones IL (1998) The auks: Alcidae. *In* Bird Families of the World. Oxford University Press, New York.

Goudet J (1995) Fstat version 1.2: A computer program to calculate Fstatistics. *Journal of Heredity,* **86**:485-486.

Haig SM, Winker K (2010) Avian subspecies: Summary and prospectus. *Ornithological Monographs*, **67**:172-175.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**:95-98.

Hall LA, Palsbøll PJ, Beissinger SR, Harvey JT, Bérubé M, Raphael MG, Nelson SK, Golightly RT, McFarlane-Tranquilla L, Newman SH, Peery MZ (2009) Characterizing dispersal patterns in a threatened seabird with limited genetic structure. *Molecular Ecology*, **18**:5074-5085.

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

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences,USA*, **104**:2785–2790.

Hipfner JM (2008) Matches and mismatches: Ocean climate, prey phenology and breeding success in a zooplanktivorous seabird. *Marine Ecology Progress Series*, **368**:295-304.

Hipfner JM, McFarlane-Tranquilla LA, Addison B (2010) Experimental evidence that both timing and parental quality affect breeding success in a zooplanktivorous seabird. *Auk*, **127**:195-203.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, **10**:639-650.

Hubbs CL (1960) The marine vertebrates of the outer coast. *Systematic Zoology*, **9**:134-147.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**:1322-1332.

Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**:1898-1914.

Huyer A (1983) Coastal upwelling in the California Current System. *Progress in Oceanography*, **12**:259-284.

Hyrenback KD, Veit RR (2003) Ocean warming and seabird communities of the southern California Current System (1987-98): Response at multiple temporal scales. *Deep-Sea Research II*, **50**:2537-2565.

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

Kidd MG, Friesen VL (1998) Sequence variation in the guillemot (Alcidae: *Cepphus*) mitochondrial control region and its nuclear homolog. *Molecular Biology and Evolution*, **15**:61-70.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**:429-434.

Lande R, Engen S, Sæther B (2003) Stochastic Population Dynamics in ecology and conservation. Oxford University Press, New York.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**:2947-2948.

Lee DE, Nur N, Sydeman WJ (2007) Climate and demography of the planktivorous Cassin's auklet *Ptychoramphus aleuticus* off northern California: Implications for population change. *Journal of Animal Ecology*, **76**:337-347.

Legaard KR, Thomas AC (2006) Spatial patterns in seasonal and interannual variability of chlorophyll and sea surface temperature in the California Current. *Journal of Geophysical Research*, **3**:1-21.

McDonald DB, Potts WK (1997) DNA microsatellites as genetic markers for several scales. In: Mindell, DP (Ed.), Avian Molecular Evolution and Systematics. Academic Press, San Diego, CA.

Milot E, Weimerskirch H, Bernatchez L (2008) The seabird paradox: dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. *Molecular Ecology*, **17**:1658-1673.

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

Moritz C (1994) Defining 'evolutionary significant units' for conservation. *Trends in Ecology and Evolution*, **9**:373-375.

Moum T, Árnason E (2001) Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. *Molecular Ecology*, **10**:2463-2478.

Moum T, Bakke I (2001) Mitochondrial control region structure and single site heteroplasmy in the razorbill (*Alca torda*; Aves). *Current Genetics*, **39**:198-203.

Moum T, Willassen NP, Johansen S (1994) Intragenic rearrangements in the mitochondrial NADH dehydrogenase subunit 6 gene of vertebrates. *Current Genetics*, **25**:554-557.

Narum SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, **7**:783-787.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Neushul M, Clarke WD, Brown DW (1967) Subtidal plant and animal communities of the Southern California Islands. *Proceedings of the Symposium on the Biology of the California Islands*, published by Santa Barbara Botanic Garden, California edited by Ralph N. Philbrick.

Oosterhout CV, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**:535-538.

Pereira SL, Baker AJ (2008) DNA evidence for a Paleocene origin of the Alcidae (Aves: Charadriiformes) in the Pacific and multiple dispersals across northern oceans. *Molecular Phylogenetics and Evolution*, **46**:430-445.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**:945-959.

Pyle P (2001) Age at first breeding and natal dispersal in a declining population of Cassin's auklet. *The Auk*, **118**:996-1007.

Pyle P, Sydeman WJ, Hester M (2001) Effects of age, breeding experience, mate fidelity and site fidelity on breeding performance in a declining population of Cassin's auklets. *Journal of Animal Ecology*, **70**:1088-1097.

Renner M, Hunt GL Jr, Piatt JF, Byrd GV (2008) Seasonal and distributional patterns of seabirds along the Aleutian Archipelago. *Marine Ecology Progress Series*, **357**:301-311.

Riemann H, Ezcurra E (2005) Plant endemism and natural protected areas in the peninsula of Baja California, Mexico. *Biological Conservation*, **122**:141-150.

Rodway MS, Lemon MJF (2011) Use of permanent plots to monitor trends in burrow-nesting seabird populations in British Columbia. *Marine Ornithology*, **39**:243-253.

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**:137-138.

Rozen S, Skaletsky HJ (2000) PRIMER3 on the WWW for general users and for biologist programmers. In Krawetz S, Misener S (Eds.) *Bioinformatics Methods and Protocols: Methods in Molecualr Biology*. Humana Press, Totowa, NJ.

Ruokonen M, Kvist L (2002) Structure and evolution of the avian mitochondrial control region. *Molecular Phylogenetics and Evolution,* **23**:422-432.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory, New York.

Shields GF, Wilson AC (1987) Calibration of mitochondrial DNA evolution in geese. *Journal of Molecular Evolution,* **24**:212-217.

Sinnock P (1975) The Wahlund effects for the two-locus model. *American Naturalist*, **109**:565-570.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**:457-462.

Stephensen SW, Irons DB (2003) Comparison of colonial breeding seabirds in the eastern Bering Sea and Gulf of Alaska. *Marine Ornithology*, **31**:167-173.

Sydeman WJ, Bradley RW, Warzybok P, Abraham CL, Jahncke J, Hyrenbach KD, Kousky V, Hipfner JM, Ohman MD (2006) Planktivorous auklet *Ptychoramphus aleuticus* responses to ocean climate, 2005: Unusual atmospheric blocking? *Geophysical Research Letters,* **22**: L22S09.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**:437-460.

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

Templeton AR, Crandall KA, Sing CF (1992) A cladistics analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**:619-633.

Thorne RF (1969) The California Islands. *Annals of the Missouri Botanical Garden,* **56**:391-408.

U. S. Fish and Wildlife Service [Online] (2006) Alaska Seabird Information Series. <http://alaska.fws.gov/mbsp/mbm/seabirds/pdf/caau.pdf> (December 2010).

Veit ML, Robertson RJ, Hamel PB, Friesen VL (2005) Population genetic structure and dispersal across a fragmented landscape in cerulean warblers (*Dendroica cerulea*). *Conservation Genetics*, **6**:159-174.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**:1419-1439.

Watterson GA (1978) Homozygosity test of neutrality. *Genetics*, **88**:405-417.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**:2321-2328.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST} ≠1/(4Nm+1). *Heredity*, **82**:117-125.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**:1177-1191.

Wolf S, Keitt B, Aguirre-Munoz A, Tershy B, Palacios E, Croll D (2006) Transboundary seabird conservation in an important North American marine ecoregion. *Environmental Conservation,* **33**:294-305.

Wolf SG, Snyder MA, Sydeman WJ, Doak DF, Croll DA (2010) Predicting population consequences of ocean climate change for an ecosystem sentinel, the seabird Cassin's auklet. *Global Change Biology,* **16**:1923-1935.

Wolf SG, Sydeman WJ, Hipfner JM, Abraham CL, Tershy BR, Croll DA (2009) Range-wide reproductive consequences of ocean climate variability for the seabird Cassin's auklet. *Ecology,* **90**:742-753.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**:97-159.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**:323-354.

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, **19**:395-420.

Appendix

Appendix A1. A sequence (S) of the most common haplotype (individual: CAsb32) of the mitochondrial control region in Cassin's auklet. Location of domains (CRI = domain I, CRII = domain II, and CRIII = domain III) and conserved blocks (L) are mapped based on the mitochondrial control sequence of razorbills (*Alca torda*; Moum and Bakke 2001) and primer positions (P) are mapped onto the sequence. An asterisk (*) represents fragments that were sequenced in all individuals. Variable sites indicated in bold font.

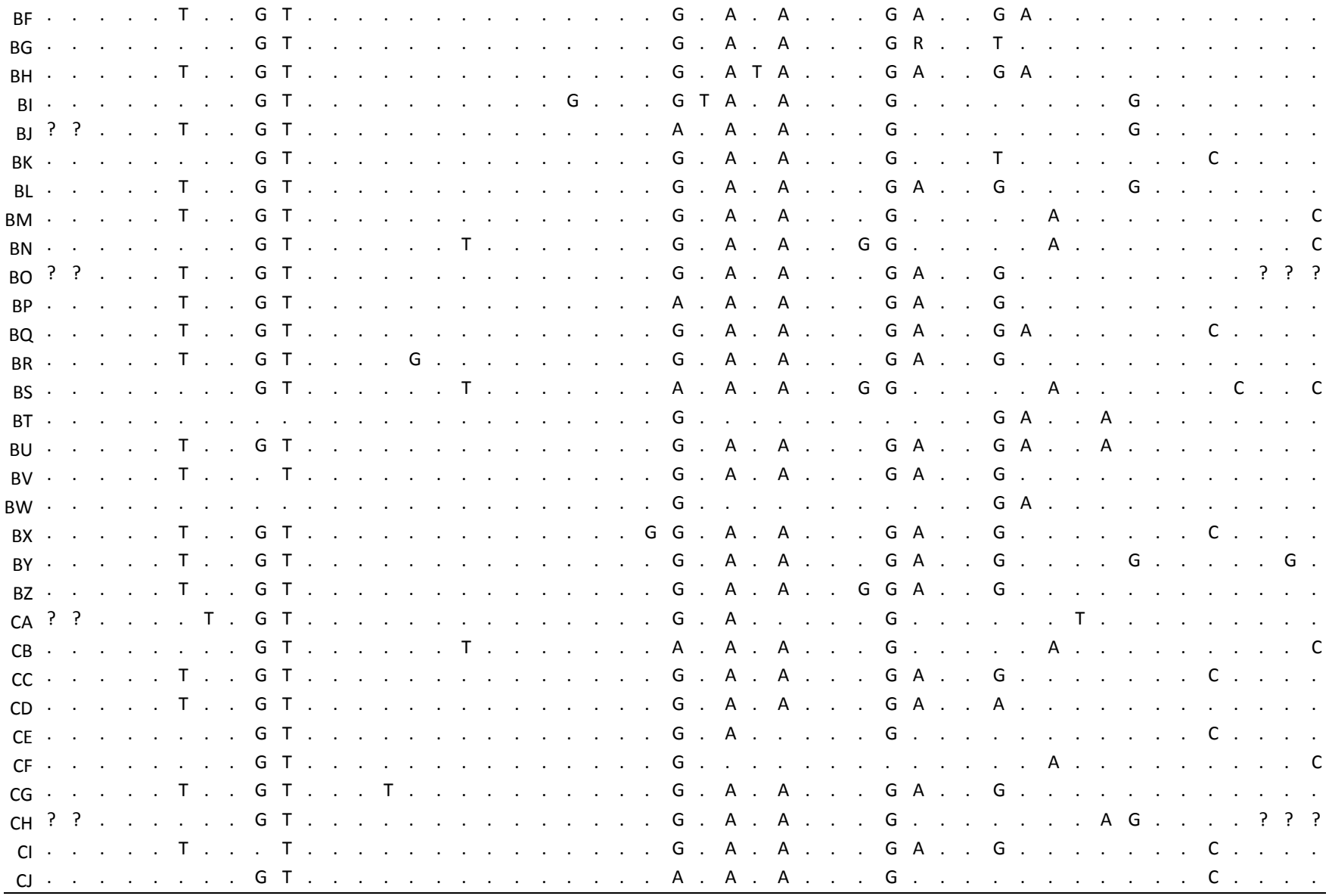
```
S:atgcacgagc gtgtctcggc atgtacgggt gtacattaac ctattacccc ca**t**a**g**tacat 60
L:<-CRI----- ---------- ---------- ---------- ---------- ----------
P:|ca_CR11L~~~~~~~~|                                       ********

S:acaacccatg t**tc**caaaacc attaa**c**a**c**a**c** aaa**c**ag**a**cat a**c**a**c**ca**c**gat tttccacatt 120
L:---------- ---------- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** *********

S:**c**ccac**c**tcca g**a**ggac**c**tc**c** c**g**cccaatgg aatccgaaat ccattacaat at**a**cg**ta**cta 180
L:---------- ---------- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** **********

S:ataccattta **c**tct**c**taggt tat**a**cata**r**t attattaatg **c**atacg**g**cag tgcttgt**c**t**g** 240
L:---------- ---------- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** **********

S:caccttt gaa tggaa**c**tagg acaaggccat ccaaaatctc ttggg**ga**cac aaagcttcgt 300
L:---------- ---------- ---------- ---------- ---------[ CRII------
P:********** ********** ********** ********** ********** **********

S:actagatgga tttatta**a**tc gtgctcctca cgtgaaatca gcaacccggt gtt**g**gtaaga 360
L:---------- ---------- |Fbox----- ---------- ---------|- ----------
P:********** ********** ********** ********** ********** **********

S:tcctacg**t**ta ccagcttcag gaccattctt tccccctaca cccctagccc atcttgctct 420
L:---------- ---------- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** **********

S:tttgcgcctc tggttcctat gtcagggcca taacttgatt aatcctttca acttgtactt 480
L:------|Dbo x--------- ---------- --|------- ---------- --|Cbox---
P:********** ********** ********** ********** *****

S:caccgataca tctggtcggc tctttgtctt tcatctcacc cgtgatcgcg acatccgacc 540
L:---------- ---------| ---------- ---------- ---------- ----------
P:                                        |ca_cr600H~~~~ ~~~~~|

S:gttttggcac ctttggttcc ttttttttt ggggcgtctt caggtaaccc ctccagtgca 600
L:---------- ---------- ---------- ---------- ---------- ----------
P:                                           |ca_cr6 52L~~~

S:ccgaggtaaa tacaatctaa aacctggacc ctccctggta ttcgtccggt ttttggccct 660
L:---------- ---------- ---------- ---------- ---------- ----------
```

```
P:~~~~~|

S:caagaatacc tgaatgagac ggttg**a**agtg ttgggggaat catttttgca ttgatgcact 720
L:---------- ---------- ---------- ---------- --------|B irdSimilarity
P:      **** ********** ********** ********** ********** **********

S:ttgctttaca tctggttatg gtg**cg**tccgc a**g**gc**c**actat **g**tatgctgct atttta**a**tgaa 780
L:~~|------- ---------- ---------- ---------- ---------] [CRIII;CSB1
P:********** ********** ********** ********** ********** **********

S:tg**c**tt**gt**tgg acatgattta tcattc**t**tca tttcctctaa ctttctaaac aacactagta 840
L:---------- -----|---- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** **********

S:gattctaact gaaaaatgaa ccgtattttc atcacacatt ttatcatcac cttcgttact 900
L:---------- ---------- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** **********

S:tgtaaaagac agtaaagttt cattaaaaaa ggacaaacac aa**t**ccacgtc cacac**a**c**t**cg 960
L:---------- ---------- ---------- ---------- ---------- ----------
P:********** ********** ********** ********** ********** *********

S:accaacacga cacaaactat aaaagaaccc ccctgaaaaa at                   1002
L:---------- ---------- ---------- ---------- ---(CAACA) ->
P:ca_cr960H~ ~~~~~~~~~~|
```

Appendix A2. Variable sites in the mitochondrial control region. See Appendix A1 for location of variable sites and Appendix A3 for shared haplotypes between sampling sites. The vertical line in the top row separates domain I from domain III.

	53	55	72	73	86	88	90	94	97	102	104	107	121	126	132	137	140	142	173	176	177	191	195	204	209	221	227	238	240	256	286	287	318	354	358	684	744	745	752	755	761	776	783	786	787	807	943	956	958
A	T	G	T	C	C	C	C	C	A	C	C	C	C	C	A	C	C	G	A	T	A	C	C	A	R	C	G	C	G	C	G	A	A	G	T	A	C	G	G	C	G	A	C	G	T	T	T	A	T
B	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	.	R	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D	.	.	C	.	.	T	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
E	.	.	C	.	.	T	.	.	G	.	.	T	Y	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.
F	.	.	C	.	.	T	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
G	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	.	.	.	.	.
H	.	.	C	.	.	T	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	Y	A	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
I	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
J	.	.	C	.	.	T	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
L	.	.	C	.	.	T	.	.	G	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.
M	.	.	C	T	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.
N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
O	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	?	?	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
S	.	.	C	T	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	G	.	.	.	.	T	.	.	.	.	.	.	.	A	C	.	.	.	.
T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
U	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
V	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
W	.	.	N	.	.	T	.	.	G	T	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
X	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Y	.	.	.	.	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

```
Z  . . C . . T . . G . T T . . . . . . . . . . . . . . A . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AA . . C . . T . . G . . T . T . . . . . . . . . . . . A . A . . T . . . . . . . . . . . . . . . . . . C . . . . .
AB . . . . . . . . G . . . . . . . . . . . Y . . . . G T . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AC . . C . . T . . . . . T . . . . . . . . . . . . . . A . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AD . . . . . . . . G . . . . . . . . . . . . . . . . . . G . . . . . . . G . . . . . . . . . . S . . . . . . . . . . .
AE . . . . . . . . . . . . . . . . . . . A . . . . . . . G . . . . . . . . . . . . . . . . . . T . . . . . . . . . . .
AF . . C . . T . . G . . T . . . T . . . . . . . . . . A . A . . . . . . . . . . T . . . . . . . . . C . . . . .
AG . . C . . T . . G . . T . . . . . . . . . . . . . . A . A . . . T . . . . . . . . . . . . . . . . . . . . . . .
AH . . C . . T . . G . . T . . . . . . . . . . . . . . A . A . . . . . . . . . . . . . . . . . . . . . . . C . . .
AI . . . . . T . . . . . . . . . . . . . . . . . . . . G . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AJ . . C . . T . . G . . T . . . . . . . G . . . . . . A . A . . . . . . . . . . . . . . . . . . . . . C . . . . .
AK . . . . . T . . . . . T . . . . . . . . . . . . . . G . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AL . . C . . T . . G . . T . . . . . . . . . . . . . . A . A . . T . . . . . G . . . . . . . . . . . . . . . . .
AM . . . . . . . . . . . . . . . . . . . . . . . . . . G . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AN . . . . . . . . . . . . . . . . . . . . . . . . . . A . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AO . . . . . . . . G . . . . . . . . . . . . . . . . . A . A . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AP C A . . . . . . . G . . . . . . . . . . . . . . . . G . . . A . . . . . . . . . . . . . . . . . C . . . . .
AQ . . . . . . . . . . . . . . . . . . . . . . . . . . G . . . . . . . . . . . . . . . . . . . . . . C . . . . .
AR . . C . . T . . G . . T . . . . . . . T . . . . . . A . A . . T . . . . . . . . . . . . . . . C . . . . .
AS . . C . . T . . G . . T . . . . . . . . . . . . . . A . A . . T . . . . . . . . . . . . . . . C . . . . .
AT . . C . . T . T G . . T . . . . T . . . . . . . . A . A . . . . . . . . . T . . . . . . . . C . . . . .
AU . . . . . . . . . . . . . . . . . . . . . . . . . . A . . . . T . . . . . . . . . . . . . . . . . . . . . . . .
AV . . . . . . . . G . . . . . . . . . . . . . . . . . G . . . A . . . . . . . . . . . . . . . . . . . . . . . . .
AW . . . . . . . . G T . . . . . . . . . . . . . . . . G . A . A . . G G . . . . . A . . . . . . . . . . C
AX . . . . . . . . G T . . . . . . . T . . . . . . . . G . A . . . . G G . . R . . R . . . . . . . . . . Y
AY . . . . . . . . G T . . . . . . . . . . . . . . . . G G . A . A . . . G . . R Y . . . . . . . . Y . . . .
AZ . . . . . . . . G T . . . . . . . . . . . . . . . . G . A . A . . . G . . . . . . . . . . . . . C . . . .
BA . . . . . T . . G T . . . . . . . . . . . . . . . S . G . A . A . . . G A . . . . . . . . . . . . . . . . .
BB . . . . . . . . G T . . . . . . . . . . . . . . . . G . A . A . . . . . . . T . . . . . . . . C . . . .
BC . . . . . T . . G T . . . . . . . . . . . . . . . . G . A . A . . . . A . . S . . . . . . . . . . . . . .
BD . . . . . . . . G . . . . . . . . . . . . . . . . . . . R . . . . . . . . . T . . . . . . . . C . . . .
BE . . . . . . . . G . . . . . . . . . . . . . . . . . . A . R . . . R . . . G . . . . . . . . . . . . . . . .
```



Appendix A3. Mitochondrial control region haplotypes shared between sampling sites. Sampling site abbreviations are the same as listed in Table 1.

	AL	SA&FR	RE	TR	WA	SF	CH	SJ	GU	SB	Total
A	4	4	7	13	3	7		1		2	41
B	1	4		3		1					9
C	1			2	1	2	1				7
D	4	1	1								6
E	3		2	1							6
F	1	1		4							6
G	1					1	1				3
H	2										2
I	1						1				2
J	2										2
K	1				1						2
L	1		1								2
M	2										2
N	1		1								2
O	1					1					2
P	1					1					2
Q	1										1
R	1										1
S	1										1
T	1										1
U	1										1
V	1										1
W		2									2
X		1									1
Y		1									1
Z			1								1
AA			1								1
AB			1								1
AC			1	1							2
AD				1							1
AE				1							1
AF				1							1
AG				1							1
AH				1							1
AI					1						1
AJ					1						1
AK					1						1
AL						3					3
AM						2					2
AN						2					2
AO						2					2
AP						1					1
AQ						1					1
AR						1					1
AS						1					1
AT						1					1

	AL	SA&FR	RE	TR	WA	SF	CH	SJ	GU	SB	Total
AU						1					1
AV						1					1
AW						1			5		6
AX							3				3
AY							2				2
AZ							2				2
BA							2				2
BB							2				2
BC							2				2
BD							1				1
BE							1				1
BF							1	2	3	3	9
BG							2			2	4
BH							1			1	2
BI								1			1
BJ								1			1
BK								1	1		2
BL								1	1		2
BM								1		1	2
BN								1	3	3	7
BO								2	1	7	10
BP									2		2
BQ									2		2
BR									2		2
BS									1		1
BT									1		1
BU									1		1
BV									1		1
BW									1		1
BX									1		1
BY									1		1
BZ									1		1
CA									1		1
CB									1		1
CC										2	2
CD										2	2
CE										2	2
CF										1	1
CG										1	1
CH										1	1
CI										1	1
CJ										1	1
TOTAL	33	14	16	29	8	30	22	11	30	30	223

Appendix A4. The observed (H_O) and expected (H_E) heterozygosities, gene diversity and allelic richness for loci used in this study. A significant p-value (α=0.05) indicates significant departures from Hardy-Weinberg equilibrium. Sampling site abbreviations are the same as listed in Table 1.

AL

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.60	0.72	0.35	0.72	4.07	7
Pal38	0.78	0.83	0.08	0.83	5.09	7
Pal37	0.79	0.83	0.45	0.83	5.18	8
Pal42	0.71	0.71	0.06	0.71	3.96	6
Pal39	0.68	0.86	0.07	0.86	5.51	9
Apy02	0.89	0.86	0.96	0.86	5.71	10
Apy10	0.71	0.72	0.81	0.72	4.65	9
Pal11	0.62	0.75	0.21	0.75	4.49	8
Pal17	0.63	0.61	1.00	0.61	2.92	4
Pal25	0.82	0.88	0.17	0.88	6.15	10
Pal26	0.90	0.82	0.01**	0.81	4.79	11
Average				0.78	4.77	7.73

SA&FR

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.72	0.78	0.34	0.79	4.86	9
Pal38	0.79	0.83	0.52	0.83	5.04	6
Pal37	0.72	0.85	0.02*	0.85	5.57	9
Pal42	0.81	0.67	0.27	0.67	3.19	4
Pal39	0.68	0.82	0.16	0.82	4.99	8
Apy02	0.70	0.81	0.11	0.81	4.96	8
Apy10	0.63	0.71	0.11	0.71	3.96	6
Pal11	0.73	0.80	0.43	0.80	4.80	7
Pal17	0.64	0.68	0.84	0.69	3.35	4
Pal25	0.93	0.87	0.92	0.87	6.10	12
Pal26	0.87	0.77	0.65	0.77	4.37	6
Average				0.78	4.65	7.18

RE

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.75	0.72	0.88	0.71	4.47	8
Pal38	0.92	0.87	0.73	0.87	5.79	8
Pal37	0.80	0.82	0.50	0.83	5.16	7
Pal42	0.63	0.71	0.36	0.72	3.53	4
Pal39	0.69	0.76	0.31	0.77	4.49	7
Apy02	0.87	0.82	0.92	0.81	4.82	7
Apy10	0.69	0.83	0.11	0.83	5.21	8
Pal11	0.69	0.81	0.03*	0.79	4.82	8
Pal17	0.44	0.64	0.15	0.62	2.95	4
Pal25	0.81	0.87	0.21	0.88	6.00	9
Pal26	0.80	0.74	1.00	0.72	3.93	5
Average				0.78	4.65	6.82

TR

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.68	0.73	0.57	0.73	4.21	8
Pal38	0.86	0.85	0.88	0.85	5.42	7
Pal37	0.71	0.74	0.22	0.74	4.68	8
Pal42	0.64	0.62	0.89	0.62	3.23	4
Pal39	0.85	0.89	0.44	0.89	6.29	12
Apy02	0.79	0.80	0.65	0.80	4.89	8
Apy10	0.78	0.81	0.47	0.81	5.59	12
Pal11	0.52	0.75	0.06	0.75	4.08	6
Pal17	0.68	0.60	0.32	0.60	3.33	4
Pal25	0.89	0.91	0.10	0.91	6.73	12
Pal26	0.85	0.80	0.12	0.80	4.69	6
Average				0.77	4.83	7.91

WA

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.86	0.88	0.86	0.88	6.00	7
Pal38	0.83	0.85	0.92	0.85	5.49	6
Pal37	0.83	0.83	0.56	0.83	4.80	5
Pal42	0.60	0.69	0.62	0.70	3.00	3
Pal39	0.83	0.80	1.00	0.80	4.79	5

Apy02	0.86	0.90	0.72	0.91	6.66	8
Apy10	0.71	0.85	0.55	0.86	5.78	7
Pal11	0.71	0.84	0.65	0.85	5.29	6
Pal17	0.57	0.65	0.63	0.66	3.64	4
Pal25	0.71	0.86	0.11	0.87	5.79	7
Pal26	0.71	0.79	0.76	0.80	3.97	4
Average				0.82	5.02	5.64

SF

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.72	0.78	0.47	0.78	4.95	9
Pal38	0.58	0.81	0.01**	0.81	4.82	6
Pal37	0.61	0.83	0.02*	0.84	5.26	8
Pal42	0.50	0.64	0.11	0.64	3.09	4
Pal39	0.65	0.88	0.01	0.88	6.05	11
Apy02	0.88	0.87	0.75	0.87	5.77	11
Apy10	0.80	0.79	0.88	0.79	5.16	10
Pal11	0.67	0.71	0.36	0.71	3.86	7
Pal17	0.66	0.72	0.84	0.73	3.76	5
Pal25	0.87	0.88	0.77	0.88	6.30	12
Pal26	0.80	0.79	0.55	0.79	4.46	6
Average				0.79	4.86	8.09

CH

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.60	0.72	0.66	0.73	3.90	5
Pal38	0.95	0.85	0.56	0.85	5.58	8
Pal37	0.57	0.73	0.05	0.74	4.64	8
Pal42	0.86	0.74	0.49	0.74	4.08	6
Pal39	0.76	0.86	0.16	0.86	5.79	9
Apy02	0.84	0.89	0.07	0.89	6.39	11
Apy10	0.77	0.81	0.38	0.81	5.27	10
Pal11	0.68	0.78	0.18	0.78	4.65	8
Pal17	0.63	0.76	0.36	0.76	4.13	5
Pal25	0.90	0.91	0.78	0.91	6.85	13
Pal26	0.86	0.77	0.84	0.77	4.52	7
Average				0.80	5.07	8.18

SJ

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.64	0.78	0.03*	0.79	4.54	6
Pal38	0.89	0.79	1.00	0.79	4.54	5
Pal37	0.82	0.81	0.82	0.81	5.08	7
Pal42	0.64	0.74	0.07	0.75	3.68	4
Pal39	0.90	0.89	0.93	0.89	6.49	9
Apy02	0.91	0.94	0.73	0.94	7.52	12
Apy10	0.91	0.89	0.47	0.89	6.10	8
Pal11	0.45	0.84	0.00**	0.86	5.01	6
Pal17	0.82	0.71	1.00	0.71	3.76	5
Pal25	1.00	0.87	0.85	0.87	6.18	9
Pal26	0.91	0.81	0.99	0.81	4.87	6
Average				0.83	5.25	7.00

GU

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.73	0.69	0.37	0.69	3.50	6
Pal38	0.81	0.82	0.31	0.83	5.38	9
Pal37	0.90	0.88	0.04*	0.88	6.07	10
Pal42	0.73	0.65	0.10	0.65	3.66	6
Pal39	0.75	0.79	0.34	0.79	4.86	8
Apy02	0.96	0.93	1.00	0.93	7.29	15
Apy10	0.87	0.80	0.04*	0.80	4.85	9
Pal11	0.56	0.75	0.12	0.76	4.24	8
Pal17	0.63	0.76	0.15	0.76	4.01	5
Pal25	0.93	0.90	0.80	0.90	6.70	25
Pal26	0.80	0.82	0.20	0.82	4.89	7
Average				0.80	5.04	8.91

SB

Locus	H_O	H_E	p-value	Gene diversity	Allelic richness	Number of alleles
Pal21	0.80	0.74	0.18	0.74	4.19	6
Pal38	0.72	0.83	0.07	0.83	5.40	9
Pal37	0.70	0.77	0.20	0.77	4.83	8
Pal42	0.57	0.69	0.09	0.70	3.84	7
Pal39	0.90	0.84	0.95	0.84	5.83	12

Apy02	0.82	0.89	0.88	0.89	6.43	12
Apy10	0.80	0.82	0.74	0.82	5.22	10
Pal11	0.52	0.84	0.00**	0.85	5.24	8
Pal17	0.69	0.68	0.33	0.68	3.67	6
Pal25	0.90	0.90	0.43	0.90	6.69	13
Pal26	0.83	0.75	0.05	0.75	4.38	7
Average				0.80	5.06	8.91

Appendix A5. Exact p-values for deviations from linkage equilibrium for loci used in this study. One asterisk (*) represents a significant p-value with an α value of 0.05 whereas two asterisks (**) represents significance after a Benjamini-Yekutieli correction ($p<0.011$). Colony abbreviations are the same as listed in Table 1.

AL

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.00**										
Pal37	0.41	0.03*									
Pal42	0.02*	0.03*	0.04*								
Pal39	0.18	0.08	0.00**	0.90							
Apy02	0.01*	0.05	0.03*	0.00**							
Apy10	0.08	0.02*	0.02*	0.03*	0.11	0.00**					
Pal11	0.08	0.3	0.09	0.16	0.00**	0.02*	0.13				
Pal17	0.07	0.05	0.25	0.01*	0.07	0.00**	0.00**	0.04*			
Pal25	0.03*	0.04*	0.01*	0.01*	0.05	0.14	0.11	0.42	0.01**		
Pal26	0.22	0.02*	0.05	0.04*	0.7	0.00**	0.33	0.12	0.00**	0.70	

SA&FR

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.04*										
Pal37	0.04*	0.11									
Pal42	0.26	0.52	0.67								
Pal39	0.89	0.02*	0.48	0.31							
Apy02	0.20	0.17	0.31	0.88	0.91						
Apy10	0.36	0.13	0.76	0.29	0.076	0.17					
Pal11	0.35	0.66	0.04*	0.43	0.37	0.32	0.90				
Pal17	0.75	0.50	0.44	0.48	0.03*	0.19	0.02*	0.10			
Pal25	0.18	0.38	0.94	0.34	0.42	0.02*	0.28	0.02*	0.00**		
Pal26	0.48	0.37	0.94	0.90	0.14	0.90	0.17	0.35	0.00**	0.13	

RE

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.05										
Pal37	0.43	0.90									
Pal42	0.82	0.06	0.75								
Pal39	0.13	0.40	0.89	0.44							
Apy02	0.39	0.06	0.235	0.96	0.35						
Apy10	0.00**	0.18	0.33	0.93	0.76	0.27					
Pal11	0.20	0.03*	0.89	0.64	0.04*	0.39	0.89				
Pal17	0.44	0.36	0.34	0.44	0.80	0.67	0.46	0.34			
Pal25	0.51	0.051	0.28	0.42	0.29	0.23	0.62	0.01**	0.58		
Pal26	0.14	0.43	0.00**	0.73	0.78	0.62	0.62	0.63	0.047	0.79	

TR

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.65										
Pal37	0.72	0.07									
Pal42	0.83	0.32	0.09								
Pal39	0.03*	0.23	0.16	0.31							
Apy02	0.23	0.55	0.52	0.23	0.84						
Apy10	0.94	0.16	0.66	0.40	0.86	0.64					
Pal11	0.35	0.37	0.21	0.60	0.61	0.22	0.64				
Pal17	0.48	0.47	0.11	0.24	0.95	0.87	0.65	0.50			
Pal25	0.56	0.83	0.26	0.79	0.36	0.51	0.30	0.90	0.25		
Pal26	0.34	0.53	0.22	0.89	0.41	0.81	0.88	0.72	0.73	0.20	

WA

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.92										
Pal37	0.25	0.24									
Pal42	0.62	0.26	0.50								
Pal39	0.18	0.41	0.36	0.21							
Apy02	0.99	0.67	0.02*	0.46	0.04*						
Apy10	0.64	0.85	0.63	0.67	0.29	0.02*					
Pal11	0.16	0.14	0.24	0.33	0.24	0.61	0.16				
Pal17	0.96	0.83	0.55	0.05	0.26	0.04*	0.053	0.34			
Pal25	0.18	0.17	0.38	0.09	0.47	0.81	0.79	0.40	0.60		
Pal26	0.20	0.23	0.10	0.77	0.28	0.85	0.32	0.61	0.38	0.03*	

SF

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.67										
Pal37	0.55	0.53									
Pal42	0.66	0.45	0.00**								
Pal39	0.16	0.01**	0.20	0.73							
Apy02	0.64	0.16	0.89	0.49	0.25						
Apy10	0.50	0.64	0.06	0.03*	0.02*	0.83					
Pal11	0.80	0.83	0.58	0.73	0.68	0.79	0.68				
Pal17	0.20	0.02*	0.24	0.87	0.41	0.43	0.58	0.99	.		
Pal25	0.36	0.81	0.83	0.11	0.96	0.71	0.43	0.86	0.35		
Pal26	0.07	0.89	0.25	0.16	0.69	0.44	0.69	0.09	0.43	0.83	

CH

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.01*										
Pal37	0.10	0.52									
Pal42	0.11	0.50	0.57								
Pal39	0.45	0.25	0.00**	0.81							
Apy02	0.28	0.48	0.21	0.92	0.34						
Apy10	0.14	0.23	0.19	0.93	0.43	0.98					
Pal11	0.60	0.49	0.51	0.20	0.70	0.79	0.96				
Pal17	0.20	0.36	0.16	0.86	0.23	0.21	0.18	0.52			
Pal25	0.95	0.71	0.56	0.76	0.44	0.54	0.37	0.14	0.13		
Pal26	0.84	0.96	0.64	0.91	0.13	0.19	0.79	0.45	0.97	0.57	

SJ

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.54										
Pal37	0.28	0.30									
Pal42	0.30	0.89	0.91								
Pal39	0.15	0.93	0.54	0.29							
Apy02	0.23	0.51	0.99	0.28	0.62						
Apy10	0.02*	0.81	0.93	0.18	0.22	0.60					
Pal11	0.69	0.64	0.88	0.04*	0.73	0.51	0.09				
Pal17	0.09	0.62	0.51	0.37	0.60	0.13	0.40	0.22			
Pal25	0.46	0.39	0.79	0.29	0.62	0.83	0.17	0.00**	0.83		
Pal26	0.36	0.29	0.31	0.42	0.80	0.79	0.04*	0.55	0.41	0.31	

GU

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.36										
Pal37	0.01**	0.36									
Pal42	0.32	0.35	0.76								
Pal39	0.19	0.049	0.42	0.15							
Apy02	0.50	0.21	0.16	0.09	0.55						
Apy10	0.61	0.10	0.08	0.08	0.48	0.35					
Pal11	0.21	0.85	0.10	0.53	0.01	0.44	0.37				
Pal17	0.53	0.61	0.28	0.01*	0.06	0.48	0.22	0.91			
Pal25	0.65	0.12	0.03*	0.77	0.01**	0.01*	0.01*	0.41	0.42		
Pal26	0.53	0.59	0.14	0.38	0.14	0.40	0.74	0.94	0.34	0.55	

SB

	Pal21	Pal38	Pal37	Pal42	Pal39	Apy02	Apy10	Pal11	Pal17	Pal25	Pal26
Pal38	0.59										
Pal37	0.98	0.31									
Pal42	0.93	0.63	0.83								
Pal39	0.93	0.90	0.67	0.32							
Apy02	1.00	0.67	0.11	0.27	0.87						
Apy10	0.06	0.36	0.93	0.46	0.97	0.83					
Pal11	0.46	0.24	0.40	0.40	0.91	0.14	0.41				
Pal17	0.63	0.89	0.63	0.65	0.43	0.14	0.48	0.44			
Pal25	0.06	0.00**	0.37	0.20	0.14	0.01*	0.71	0.66	0.77		
Pal26	0.25	0.42	0.42	0.57	0.26	0.43	0.87	0.99	0.06	0.16	