THE PHYLOGEOGRAPHY AND POST-GLACIAL EXPANSION OF THE BOREAL CHICKADEE (*POECILE HUDSONICUS*)

LINDA AMY LAIT
BSc Biology, University of Lethbridge, 2009

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

[MASTER OF SCIENCE]

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

ABSTRACT

Using a combination of both mitochondrial DNA and microsatellites, as well as spatio-geographic modelling, this study examines how the most recent glaciations may have affected the population genetic structure of the boreal chickadee (*Poecile hudsonicus*), a small resident passerine of the North American boreal forests. The mtDNA data support a separation between eastern and western populations, with central populations containing a mixture of haplotypes from both the east and west. Estimated dates place the divergence during the Wisconsin (56.2-129.7 kya). Microsatellite analyses support the separation of Newfoundland from all mainland populations, indicating that the eastern straits are restricting gene flow. No evidence for reduced gene flow was found for the Rocky Mountains. The results suggest the use of multiple glacial refugia, one in Beringia and one in the east, followed by stepping-stone colonisation and secondary admixture.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my family (Mum, Dad and Susan) for all of their support, especially for reading this thesis cover to cover. Without their help and distractions this project would have been much harder and less enjoyable to get through.

I would like to thank my supervisor, Theresa Burg, for all of her help over the past few years – from the field to the final product. From her I have learned all of the fundamentals that will help carry me through the next stage of my career. My committee members, John Bain and Sergio Pellis, have been paramount in pushing me in the right direction and giving me new ideas. My lab mates, Brendan, Paulo, John, and Kim, for all of their help both in the field and in the lab, and for all of the laughs along the way. A special thank you to Brendan for teaching me all I know about fieldwork.

Many thanks to all of the field assistants who braved the rain to catch chickadees, to Ken, Angelique, Emily, and Kevin for catching birds for me, as well as to all of the kind people who allowed us to catch birds at their feeders. Thank you to all of the museums who provided tissue samples for this research.

This research would not have been possible without all of the funding agencies who supported this project: Natural Sciences and Engineering Research Council (NSERC; TMB and LAL), Alberta Innovates (TMB and LAL), Canada Foundation for Innovation, and the University of Lethbridge.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine
AFLP	amplified fragment length polymorphism
all	complete ENM occurrence dataset
AMOVA	analysis of molecular variance
AN	number of alleles
A_R	allelic richness
ATP	ATPase 6-8 coding region
AUC	area under the curve
BAPS	Bayesian analysis of population structure
bp	base pair
C	cytosine
ca	circa
cpDNA	chloroplast DNA
CR	control region
cytb	cytochrome b
D-loop	displacement loop
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
ENM	ecological niche modelling
exo	exonuclease I
FCA	factorial correspondence analysis
FDR	false discovery rate
G	guanine
GIS	geographic information system
GPS	global positioning system
h	haplotype diversity
H	heavy strand
H_E	expected heterozygosity
H_O	observed heterozygosity
IAM	infinite alleles model
K	number of groups
km	kilometre
km^2	square kilometre
kya	thousand years ago
L	light strand
LGM	last glacial maximum
Ln Pr (X|K)	log likelihood probability of K clusters
m	metre
M	molar (moles/litre)
McMC	Markov chain Monte Carlo
$MgCl_2$	magnesium chloride
mg	milligram
min	minutes

MIROC	model for interdisciplinary research on climate
mL	millilitre
mm	millimetre
mM	millimolar
mod	modified ENM occurrence dataset
mRNA	messenger RNA
mtDNA	mitochondrial DNA
My	million years
n	sample size
NaCl	sodium chloride
NADH	nicotinamide adenine dinucleotide (reduced)
nd2	NADH dehydrogenase subunit 2
N_E	effective population size
ng	nanogram
nm	nanometre
NP	National Park
nuDNA	nuclear DNA
PA_R	private allelic richness
PCO	principal coordinates analysis
PCR	polymerase chain reaction
P_{crit}	critical p value
PMIP2	Paleoclimate Modelling Intercomparison Project Phase II
PP	Provincial Park
precip	precipitation
Q	ancestry coefficient
r^2	correlation coefficient (linear regression analysis)
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
ROC	receiver operating characteristic
ROS	reactive oxygen species
rRNA	ribosomal RNA
rpm	revolutions per minute
s	seconds
SAMOVA	spatial analysis of molecular variance
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate
SMM	stepwise mutation model
t	divergence time
T	thymine
T_1	annealing temperature 1
T_2	annealing temperature 2
temp	temperature
TPM	two-phase model
tRNA	transfer RNA
U	unit
w/v	weight per volume

°C	degrees Celsius
°N	degrees north of the equator
°W	degrees west of the meridian
δ	corrected average pairwise differences
μ	mutation rate
μL	microlitre
μM	micromolar
π	nucleotide diversity

Populations:

AKA	Alaska Anchorage
AKF	Alaska Fairbanks
AKW	Alaska Wrangell St. Elias
CAB	central Alberta
CBC	central British Columbia
LAB	Labrador
NBC	northern British Columbia
NL	Newfoundland
NON	northern Ontario
NQC	northern Québec
NSNB	Nova Scotia & New Brunswick
NY	New York
SAB	southern Alberta
SK	Saskatchewan

Museums:

AMNH	American Museum of Natural History
BMNHC	Burke Museum of Natural History and Culture
CMN	Canadian Museum of Nature
NBM	New Brunswick Museum
NYSM	New York State Museum
RABM	Royal Alberta Museum
RBCM	Royal British Columbia Museum
ROM	Royal Ontario Museum
SMNH	Smithsonian Museum of Natural History

CHAPTER ONE – GENERAL INTRODUCTION

Ecological patterns and processes shape the genetic patterns seen in species today. Phylogeography aims to uncover how paleogeographic and environmental processes may have influenced genetic variation, and to identify this signature in the current geographic distribution of species (Avise *et al.* 1987; Avise 2000). Intraspecific phylogeography specifically looks at how a species responds to an array of factors, such as land formation (e.g. islands), major geological events (e.g. volcanic activity), and climatic events (e.g. glaciations). Topographic features (e.g. mountain ranges) can also influence the geographic distribution of a species.

The recurring glaciations during the Pleistocene impacted the genetic structure and biodiversity seen today, particularly in temperate species. The rapid contraction of habitat led to loss of diversity, both in terms of number of species and within-species variation (Hewitt 2004a). Species which were physically isolated in separate locations by the ice sheets underwent genetic divergence and, if prolonged, resulted in speciation (Stewart & Lister 2001; Hewitt 2004a).

1.1 The Pleistocene Epoch

The Earth's biological history is full of change. Large volcanic eruptions, resulting in global warming, precipitated mass extinction events in the Paleozoic. The Mesozoic was a time of speciation and extinction events due to the break-up of the Pangaea supercontinent, as well as climate change following the impact of asteroids (Hallam & Wignall 1997). The current era, the Cenozoic, began approximately 65 million years ago.

Although species have not suffered extinctions to the same degree as in earlier eras, the most recent period, the Quaternary, has been inundated by severe geological events (Gradstein *et al.* 2004).

1.1.1 The last ice age

The Pleistocene epoch occurred from approximately 1.8 million years ago to 12,000 years ago, and consisted of long glacial periods interspersed by shorter interglacials (Nilsson 1983). This oscillation of freezing and warming greatly influenced the evolution of species in a wide range of habitats – from the arctic species that thrived during colder periods, to the temperate and tropical species that suffered huge losses as the climate cooled at the onset of glaciation (Hewitt 2004a; Stewart & Dalén 2008). These extreme changes in climate are believed to be a result of the Milankovitch cycle, a combination of variations in the three cycles of the Earth's orbit: the eccentricity (100,000 year cycle), the axial tilt (40,000 year cycle) and the precession or wobble (26,000 year cycle; Bennett 1990; Hewitt 1996; Hewitt 2000). At their most extreme, the global insolation (solar radiation) is sufficiently reduced to cause cooler summers to the extent that the winter ice-pack does not melt (Clark *et al.* 2009). It is this lowering of summer temperatures, in combination with large amounts of polar ice, which trigger the onset of glaciations.

The most recent glaciation occurred approximately 110-12 kya (thousand years ago): the Würm (Eurasia) or Wisconsin (North America) glaciation. During this time much of the northern hemisphere was covered by large ice sheets, sea levels dropped, and climate conditions changed considerably (Flint 1947; Hewitt 1996). The last glacial

maximum (LGM) was the largest in the Pleistocene, perhaps the most extreme in history, with over 17 million km^2 of ice covering North America (Barendregt & Irving 1998; Dyke *et al.* 2003; Clark *et al.* 2009). The ice reached its maximum extent between 21 and 18 kya, and sea level drops of 120 m uncovered 20% more land than at present (Pielou 1991; Hewitt 2004a).

In North America two main ice sheets were present (Figure 1.1): the Laurentide ice sheet in the centre and east, and the Cordilleran ice sheet in the west. The Laurentide ice sheet originated from centres over Keewatin (Northwest Territories) and Labrador, and spread to cover an area from the Rocky Mountains to the Atlantic coast, extending as far south as 40°N in some regions (Pielou 1991; Hewitt 1996; Barendregt & Irving 1998). The Cordilleran ice sheet extended along the west side of the Rocky Mountains from northern Washington up to Alaska (Pielou 1991). A third, the Innuitian ice sheet, covered Greenland and the Arctic (Dyke *et al.* 2002).

1.1.2 Glacial refugia

With much of the land covered with snow and ice, temperate fauna and flora underwent dramatic range shifts and extinctions, while many arctic and high-latitude species flourished (Barrowclough *et al.* 2004; Hewitt 2004a). Temperate species were unable to survive in their pre-glaciation range, and persisted only in ice-free regions known as refugia. A combination of fossil evidence, pollen data and sediment cores has been used to identify the location of several glacial refugia (Pielou 1991). While these habitable areas tended to be found south of the ice sheets, abundant evidence exists for

more northerly 'cryptic' refugia (Stewart & Lister 2001; Hewitt 2004a; Provan & Bennett 2008).

In North America there has been much debate over the number and location of glacial refugia (Figure 1.1). Much of mid-latitude United States was ice-free, and while the climate was much different than today, this region is thought to have supported most of the plants and animals of North America (Pielou 1991; Jaramillo-Correa *et al.* 2009). There existed as many as six or more refugia south of the ice sheets, with eastern and western refugia separated by a large swath of desert (Pielou 1991; Swenson & Howard 2005 and references therein). Each supported a small array of species not generally co-associated (Graham *et al.* 1996; Stewart & Lister 2001). Boreal and temperate trees were widespread in the southeast, providing habitat for forest species such as birds and mammals (Jackson *et al.* 2000).

A second widely-accepted ice-free area existed in the far north-west, in the area between Alaska and the Yukon across to Siberia. In the early 20th century Hultén (1937) named this region Beringia after the land that emerged as the sea levels dropped. The Bering Strait is a mere 50 m deep, and the land bridge was quite extensive (Brubaker *et al.* 2005). Originally believed to harbour mostly arctic shrubs in a steppe-tundra environment (Brubaker *et al.* 2005; Elias & Crocker 2008), evidence now supports the use of Beringia by a multitude of plants (Bain & Golden 2005; Anderson *et al.* 2006; Gerardi *et al.* 2010), insects (Scudder *et al.* 1993; Reiss *et al.* 1999), fish (Crossman & Harington 1970; Cumbaa *et al.* 1981), birds (Holder *et al.* 1999; Scribner *et al.* 2003; Jones *et al.* 2005; Pruett & Winker 2005), and mammals (Flagstad & Røed 2003; Eddingsaas *et al.* 2004; Cegelski *et al.* 2006; Aubry *et al.* 2009). The land bridge may

also have acted as a migration corridor, allowing terrestrial species to travel freely between North America and Asia (Elias & Crocker 2008).

A number of additional, smaller refugia are purported to have existed along the periphery of the ice sheets. On the east coast of Canada, the Atlantic seabed was exposed for much of the LGM (Pielou 1991). This 'Atlantic shelf' refugium is thought to have supported numerous tree (Schauffler & Jacobson 2002; Brubaker *et al.* 2005) and bird (Gill *et al.* 1993; Zink & Dittmann 1993; Colbeck *et al.* 2008) species. A number of now-submerged banks and islands may have combined to form a larger area, or may have been emerged at different times (Pielou 1991). This dynamic environment became even more unstable as the ice sheets began to recede and the sea levels rose once again, isolating previously connected areas. Along the west coast of North America, both the Alexander Archipelago (Alaska) and Haida Gwaii (also known as the Queen Charlotte Islands, British Columbia) have been shown to support at least a small number of species (Byun *et al.* 1997; Soltis *et al.* 1997; Janzen *et al.* 2002; Burg *et al.* 2005). Vancouver Island and several isolated coastal refugia, as well as nunataks (exposed mountain terrain above the ice sheets), may also have played a role in Pleistocene species survival (Pielou 1991).

1.1.3 Colonisation

As the climate warmed, the ice sheets began to melt and the sea levels rose. The recession of the ice sheets was asynchronous; the Laurentide ice sheet started to recede about 16 kya and the Cordilleran ice sheet 12 kya (Harris 1996). As land became available, the return of life was rapid – shrubs and trees, closely followed by the animals they sustain, moved out of refugia (Hewitt 2001). Many high-latitude regions in North

America were colonised from Beringia, while the general trend was a northerly movement from the south (Hewitt 2001). In general, genetic diversity decreased with distance from the refugia due to sequential bottlenecks and founder events (Hewitt 2000; Petit *et al.* 2003; Hewitt 2004b). This genetic depauperation led to the term "southern richness, northern purity", often referred to in glacial phylogeography (Hewitt 1996, 2000, 2001).

There are several ways in which species could have colonised previously glaciated areas, and accordingly, a number of post-glacial colonisation models have been proposed: the phalanx, pioneer, and stepping-stone models (Figure 1.2a; Nichols & Hewitt 1994; Ibrahim *et al.* 1996). The phalanx model is based on Fisher's wave, and is characterised by a gradual spread of individuals into new habitat. The advancing wave is slow enough that gene flow is maintained among the populations, analogous to a large panmictic population, allowing genetic diversity to be retained. This model results in a homogeneous pattern, with high within-population diversity and low among-population differences (Nichols & Hewitt 1994; Hewitt 1996). The pioneer model includes both short and long-distance dispersal. The short-distance movement retains much of the genetic diversity, while the long-distance colonisers tend to be few in number, and result in founder effects and pockets of genetically distinct clusters of individuals. In contrast to the phalanx model, the pioneer model shows modest among-population differences, with high among-population distances around the pocket populations, and moderate within-population variation throughout most of the range except in pocket populations where it is low (Nichols & Hewitt 1994; Ibrahim *et al.* 1996). The stepping-stone model, similar to an island model, allows short-distance dispersal between neighbouring populations.

Movement tends to be primarily into new habitat, although subsequent bidirectional gene flow may exist. As short-distance dispersal allows a reasonable number of migrants, genetic diversity remains relatively high within populations, and moderate among populations, with neighbouring populations showing few differences (Ibrahim *et al.* 1996).

Many species persisted in multiple glacial refugia; consequently, diversity patterns are more complex than these models predict (Figure 1.2b). As refugial populations expanded into virgin territory, their reproduction was exponential. If an area was already inhabited, however, one of two things happened: (1) the advance stopped, or (2) the expansion continued, although at a lesser rate (Hewitt 2000, 2004b). In either case, the merging of two (or more) refugial populations resulted in an increase in diversity due to secondary admixture (Taberlet *et al.* 1998; Petit *et al.* 2003).

1.2 Physical Barriers

Barriers to dispersal can come in many forms: physical barriers, intraspecific competition, behavioural barriers, resource availability and climatic conditions. While some of these obstacles may be easier to overcome than others, they all play a major role in shaping species biodiversity. Physical barriers are perhaps the largest contributors to population genetic structure due to their permanency.

1.2.1 Mountain ranges

Every continent has mountain ranges which act as barriers. In Europe, the Alps have been shown to restrict dispersal in plants (Thiel-Egenter *et al.* 2011), while in Asia,

the Himalayas have reduced gene flow in rhesus macaques *Macaca mulatta* (Kyes *et al.* 2006). In Canada and the United States there are five major mountain chains, all running north-south through the continent: the Coastal, Cascade, Rocky, Sierra Nevada and Appalachian Mountains (Figure 1.3). While the Rocky Mountains are by far the largest, all of these ranges restrict dispersal in some species. Both the mountain ranges themselves, and the area between them, act as barriers in many tree species (Soltis *et al.* 1997 and references therein; Jaramillo-Correa *et al.* 2009 and references therein). Gene flow in crustaceans (Crease *et al.* 1997), reptiles (Fontanella *et al.* 2008), amphibians (Austin *et al.* 2002; Jones *et al.* 2006), birds (Burg *et al.* 2005; Peters *et al.* 2005), and mammals (Rueness *et al.* 2003; Geffen *et al.* 2004) is also restricted.

1.2.2 Waterways

Previous studies have shown that large bodies of water (Figure 1.3) can stop movement in both aquatic and terrestrial species. The Mississippi and Tombigbee Rivers prevent movement in freshwater fish (Bermingham & Avise 1986; Avise 1992), reptiles (Burbrink 2002; Brandley *et al.* 2010), amphibians (Hoffman & Blouin 2004; Moriarty & Cannatella 2004), birds (Gill *et al.* 1999), and mammals (White *et al.* 2000; Brant & Orti 2003). Straits surrounding islands may preclude colonisation, or gene flow between existing populations. This may be the case with rock ptarmigan *Lagopus mutus* (Holder *et al.* 1999) and moose *Alces alces* (Broders *et al.* 1999) in Newfoundland, as well as the many endemic bird species in Haida Gwaii (Topp & Winker 2008). Unidirectional barriers such as waterfalls can also cause genetic discontinuities (e.g. in fish; Castric *et al.* 2001; Crispo *et al.* 2006).

1.2.3 Human-mediated barriers

With the onset of anthropogenic development, many species suffered various types of habitat disturbance. Habitat fragmentation, often via logging, road-building, or agriculture, has had a profound influence on genetic structure, namely gene flow and levels of genetic variation. Studies have shown that genetic discontinuities have arisen from 'patchy' habitat in insects (Vandergast *et al.* 2006), reptiles (Stow *et al.* 2001), birds (McDonald *et al.* 1999; Uimaniemi *et al.* 2000), and mammals (Coulon *et al.* 2004). Roads prevent movement both as a foreign surface (Baur & Baur 1990; Vos *et al.* 2001) and due to animal-vehicle collisions (Riley *et al.* 2006; Frantz *et al.* 2010). This introduction of unfamiliar barriers may also contribute to the connectivity of a species, and has been shown to precipitate the onset of genetic drift.

1.3 Molecular Markers

The introduction of the polymerase chain reaction (PCR) and improved laboratory techniques have opened new doors to answer more complex evolutionary questions. The most common marker in animal molecular studies remains mitochondrial DNA (mtDNA), although studies employing nuclear markers are increasing (Avise 2004). A multi-locus study is the preferred approach, and two complementary markers are mtDNA and microsatellites (Zhang & Hewitt 2003). The main differences between these two markers are the effective population size, the mutation rate and the mode of inheritance (Hare 2001; Zink & Barrowclough 2008; Brito & Edwards 2009). The genetic patterns detected by a molecular marker is dependent upon both the effective population size and the mutation rate (Hare 2001; Zink & Barrowclough 2008).

1.3.1 Mitochondrial DNA

Animal mtDNA is a maternally inherited marker, in most species, which does not generally undergo recombination. It is a circular molecule containing 15 coding genes (two rRNAs and 13 mRNAs), 22 tRNAs and a non-coding control region (Figure 1.4a; Bensch & Harlid 2000; Kvist 2000). As mtDNA is a single, continuous genome, all markers are physically linked. There are four main gene orders in avian mtDNA, each of which is fairly conserved within taxa (Desjardins & Morais 1990; Gibb *et al.* 2006). Most avian species have the standard 'chicken' gene order, allowing easy discovery of marker-specific primers (Singh *et al.* 2008). MtDNA is a highly variable marker, with an average mutation rate of ~2% per million years, 5-10 times faster than in nuclear DNA (nuDNA). The higher mutation rate is primarily due to an inefficient repair mechanism, higher copy number per cell, and shorter generation time (i.e. mtDNA replicates more often than nuDNA; Brown *et al.* 1979; Baker & Marshall 1997). Given its uniparental inheritance and haploid nature, mtDNA has an effective population size (N_E) four times smaller than nuclear DNA. This makes it more sensitive to genetic drift, and it will likely show monophyly much more rapidly, allowing relatively recent events (e.g. the LGM) to be studied (Hare 2001; Palumbi *et al.* 2001; Zink & Barrowclough 2008; Brito & Edwards 2009).

Initially, fairly conserved coding regions such as cytochrome b (cytb) or NADH dehydrogenase subunit 2 (nd2) were used to study both inter- and intraspecific relationships. These regions tend to have a relatively low mutation rate, and may be under functional constraints (Avise 2004). Recently, studies have moved towards using the non-coding control region. The avian control region consists of three domains (Figure 1.4a): I,

II, and III. Domains I and III tend to be the most variable (domain III in *Parus*), and domain II contains conserved functional boxes responsible for regulating transcription and replication in the mtDNA genome (Kvist 2000; Ruokonen & Kvist 2002). The displacement loop (D-loop), or origin of replication, is located in domain III (Ruokonen & Kvist 2002). It was initially believed that the control region had a mutation rate of 15-20% per million years, considerably higher than that of the coding regions (Baker & Marshall 1997). This figure was primarily based on a short fragment of domain I found in the snow goose (Quinn 1992). It has since been shown that the mutation rate varies among groups, and that rates can be as low as 2-5% in both birds and mammals (Cann *et al.* 1984; Ruokonen & Kvist 2002; Pereira *et al.* 2004; Päckert *et al.* 2006).

Mitochondrial DNA has two main limitations. First, each cell contains multiple (e.g. 50-100) mitochondria, and within each mitochondrion there also exist more than one copy (e.g. 2-10) of the genome. While most of the mtDNA genomes will likely be identical, it is possible that an individual can contain multiple haplotypes. This phenomenon, termed heteroplasmy, is fairly rare, and is detected by the presence of multiple peaks in a sequence chromatogram (Avise *et al.* 1987). The second limitation is perhaps slightly more common, and less identifiable. Two individuals with the same sequence are assumed to be identical by descent. Instead, it may be that a base pair has mutated back to its original state (e.g. A → T → A), and thus there should actually be two mutations between the individuals (Avise 2004; Zink & Barrowclough 2008). Homoplasy is difficult to detect; however, presence of a strong geographic pattern should overcome this limitation.

1.3.2 Microsatellites

Microsatellites are short tandem repeats, generally of 2-6 base pairs (Figure 1.4b), found throughout the nuclear genome as well as in chloroplast DNA (cpDNA; Jarne & Lagoda 1996; Avise 2004). The repeats can be pure (e.g. AGAGAG), compound (e.g. AGAGCTCT) or interrupted (AGAGTTAGAGAG) (Jarne & Lagoda 1996). Pure repeats are generally the most commonly used in population studies, likely due to their high variability and ease of scoring (Luikart & England 1999). As nuclear markers are biparentally inherited, both the male and female histories impact the results. Microsatellites are codominant markers, and are inherited in a standard Mendelian manner (Jarne & Lagoda 1996).

Microsatellites have an extremely high mutation rate, estimated to be from 10^{-6} to 10^{-2} mutations per generation (Schlötterer 2000). Based on an average generation time of two years (common in passerines; Uimaniemi *et al.* 2003), this rate is up to 1000 times higher than that of mtDNA. This level of polymorphism creates a high degree of resolution, and it is argued that more recent events may be revealed with microsatellites that are not evident with mtDNA (Flanders *et al.* 2009; Kempf *et al.* 2009). The high mutation rate is due to the method of mutation – DNA replication slippage. During replication, DNA polymerase often detaches from the strand it is copying, especially if the sequence contains a number of repeats. Given the repetitive nature of microsatellites, when the enzyme tries to reattach, it does so in the wrong place, either upstream or downstream by a few base pairs. This causes the microsatellite to grow or shrink by one or a few repeats. A repair mechanism recognises and fixes some of these errors, but many are not caught and remain in the DNA (Jarne & Lagoda 1996; Schlötterer 2000).

Two hypothesized mutation models in microsatellites are the stepwise mutation model (SMM) and the infinite alleles model (IAM). While it is unlikely that either of these models is strictly followed, they are good representations of what is happening. The SMM states that the mutations occur one repeat unit at a time, either increasing or decreasing in number (Jarne & Lagoda 1996). A variation on the SMM, the two-phase stepwise model (TPM), allows rare multiple repeat changes as well as single repeat changes (Di Rienzo *et al.* 1994). The IAM states that every mutation creates a new allele (Goldstein & Pollock 1997).

As with all markers, microsatellites have a number of limitations. The initial issue is with primer development. This can be a lengthy and costly process if loci are not already developed in the species of interest or a closely related species (Jarne & Lagoda 1996; Goldstein & Pollock 1997). Microsatellites are also believed to be heavily plagued by homoplasy. The high level of variation, and similarity due to convergence rather than descent, may make genetic patterns difficult to interpret or completely overwrite historical signatures (Primmer & Ellegren 1998; Estoup *et al.* 2002; Wang 2010).

1.4 Study Species

1.4.1 Paridae

The Paridae family (tits, titmice and chickadees) is a group of small, familiar songbirds found throughout Europe, Asia, Africa and North America. Parids are generally found in treed areas, and are known for caching seeds and insects (Gill *et al.* 2005). Their feeding behaviour and reliance on temperate forests may explain the absence of these birds in the warmer southern continents (e.g. reliance on cool

temperatures for food storage). Worldwide there are ~68 species of Paridae separated into nine genera (Lepage 2003; Gill *et al.* 2005). In North America there are seven chickadee (*Poecile*) and five titmouse (*Baeolophus*) species (American Ornithologists' Union 1998; Chesser *et al.* 2010). All but two of these species are restricted either to the eastern (Carolina chickadee *P. carolinensis* and tufted titmouse *B. bicolor*) or western (Siberian tit or grey-headed chickadee *P. cinctus,* mountain chickadee *P. gambeli,* chestnut-backed chickadee *P. rufescens,* Mexican chickadee *P. sclateri,* oak titmouse *B. inornatus*, juniper titmouse *B. ridgwayi*, bridled titmouse *B. wollweberi*, and black-crested titmouse *B. atricristatus*) half of the continent (Peterson 1990; Sibley 2003). The two wide-spread species, black-capped chickadee *P. atricapillus* and boreal chickadee *P. hudsonicus*, are found coast to coast.

1.4.2 The boreal chickadee

This study looks at the boreal chickadee, *Poecile hudsonicus*. The boreal chickadee is found primarily in the boreal forests of North America. It is generally associated with spruce and fir trees which it uses for both food and shelter (Gill *et al.* 1993; Ficken *et al.* 1996). Its range extends from Alaska to Newfoundland, as far south as Montana in the west and Wisconsin/New York in the east (Figure 1.3; Sibley 2003). The distribution closely resembles that of both black spruce *Picea mariana* and white spruce *Picea glauca* ranges, to which it is closely connected (Anderson *et al.* 2006; Gerardi *et al.* 2010).

The boreal chickadee is a permanent resident and does not undergo seasonal migration (Ficken *et al.* 1996), although occasional irruptions occur in response to food supply (e.g. spruce budworm outbreaks; Yunick 1984; Bolgiano 2004). Banding records

(Canadian Bird Banding Office) are sparse, but all recaptured birds were found within 40 km from the initial banding site (n = 21). Little is known about natal dispersal in this species. As with other chickadee species, individuals form territorial pairs in the breeding season and small flocks in the winter (n = 4; McLaren 1975). They are cavity nesters, and can either excavate their own nest in a soft stump or tree, or use one previously made by a woodpecker or other primary excavator (McLaren 1975; Ficken *et al.* 1996). Several studies have examined the breeding biology and vocalisations in this small songbird (McLaren 1975, 1976), yet little is known about the population structure.

A previous study by Gill *et al.* (1993) examined the relationship of North American chickadees. They found two clades: a 'black-capped' group consisting of *P. atricapillus, P. carolinensis*, and *P. gambeli* and a 'brown-capped' group made up of *P. hudsonicus, P. rufescens*, and *P. sclateri*. A later study (Gill *et al.* 2005) included *P. cinctus* in the brown-capped group and moved *P. sclateri* to the black-capped clade. It has been shown that *P. rufescens* is the sister-species to *P. hudsonicus*, likely diverging in the early Pleistocene (Gill *et al.* 2005). As part of the original study, Gill *et al.* also looked at the within-species differences. In the boreal chickadee they found seven haplotypes in the five continental populations sampled, and an eighth haplotype restricted to Nova Scotia and Newfoundland. The highest observed diversity was in the central populations, however only 37 individuals were included in the study. The molecular markers used by Gill *et al.* (1993) were restriction fragment length polymorphisms (RFLPs), common at the time, but since replaced by more variable sequence data and nuclear markers (Avise 2004).

1.4.3 Subspecies

As with many birds, subspecies have been described based on morphological features and geographical boundaries (Avise 1992; Zink 2004). There has been much debate over the value of subspecific designation unless a strict definition can be agreed upon (Ball & Avise 1992; Zink 2004). In many species there is no genetic evidence to support described subspecies (Ball & Avise 1992). A good example of this is the dark-eyed junco, *Junco hyemalis*. There are dramatic plumage differences between the five recognised subspecies, yet no distinct genetic pattern supports this (Milá *et al.* 2007). This may be due to phenotypic plasticity and longer coalescent time required in genes.

The 1957 AOU checklist, the last to designate subspecies, described five subspecies in boreal chickadees (American Ornithologists' Union 1957). Using a combination of the AOU checklist (1957), Pyle (1997), a widely used bird identification guide, and Birds of North America (Ficken *et al.* 1996), I extracted the approximate distributions of the five subspecies (Figure 1.5). *P. hudsonicus littoralis* is described as extending from Québec east to the Maritimes and the north-eastern United States, while *P. hudsonicus rabbittsi* was described for Newfoundland and a nearby island. The widespread *P. hudsonicus hudsonicus* extends from Alaska to Ontario, *P. hudsonicus columbianus* along the Rocky Mountains and Coastal Mountains, and *P. hudsonicus cascadensis* is found exclusively in the northern Cascade Mountains (southern British Columbia/northern Washington). Gill *et al.* (1993) found no support for these subspecies, nor those described in *P. atricapillus* or *P. carolinensis*. However, as mentioned above, the intraspecific study required greater sampling, and a more sensitive molecular marker may provide greater resolution.

1.5 Thesis Aims

The aims of this study are to understand where the boreal chickadee survived the last glacial maximum, how it colonised once the ice sheets receded, and what role physical barriers play both in the historical and contemporary population genetic structure. While many studies have looked at phylogeography in North American birds (e.g. Barrowclough *et al.* 2004; Burg *et al.* 2005; Spellman *et al.* 2007), few focus on widespread species (Ball *et al.* 1988; Colbeck *et al.* 2008), and fewer still on widespread high-latitude resident species (Klicka *et al.* 2011). The studies there are tend to lack samples from the northern extent of the range, which is important when evaluating Pleistocene glacial refugia and post-glacial expansion.

In order to study the phylogeography and postglacial expansion of the boreal chickadee, I incorporated both spatial modelling and genetic analysis. Spatial modelling allows the potential species distributions to be estimated based on current climate data (ecological niche modelling) and past climate estimates (paleogeographical modelling). The modelled distribution can be used in conjunction with genetic data to identify putative refugial locations. I employed molecular markers with complementary modes of inheritance and different mutation rates. The combination of mtDNA, which is especially useful at examining historical patterns (Uimaniemi *et al.* 2003; Zink & Barrowclough 2008), and microsatellites, reported to be useful at extracting patterns of contemporary movement (Costello *et al.* 2003; Chiucchi & Gibbs 2010), will allow a more complete picture of the chickadee's history than is currently available. The use of both markers also avoids the limitations of single gene inferences (Zhang & Hewitt 2003; Brito & Edwards 2009). Samples were collected from 14 populations covering most of the

chickadee's distribution, and sampling locations were chosen such that physical barriers (Figure 1.3) and putative refugia (Figure 1.1) could be tested.

1.6 Predictions

Given the widespread distribution of the boreal chickadee (Figure 1.3), and the fact that it includes a number of possible refugia (Figure 1.1; Pielou 1991), I predict that there will be evidence of expansion from multiple glacial refugia – one (or more) in the east, and one (or more) in the west. If a single refugium was used, genetic diversity would decrease with increasing distance from the refugial population, and some alleles would likely be shared among distant populations. If multiple refugia were used a slightly different pattern would be expected; high genetic diversity in (or near) all refugial populations, and where the different expansions meet (i.e. a contact zone; Taberlet *et al.* 1998). Few alleles would be shared among refugial populations, depending on the duration of isolation.

As the boreal chickadee had to wait until suitable habitat became available (i.e. for trees to colonise the land), and most of their current range was glaciated, I predict that boreal chickadees expanded following a phalanx model. Genetic diversity will be high within-populations and low among-populations. This mirrors the limited genetic structure seen with RFLPs (Gill *et al.* 1993). Physical barriers will likely act to prevent dispersal in this sedentary songbird. The Rocky Mountains have been shown to act as a barrier to gene flow in many birds (Milot *et al.* 2000; Lovette *et al.* 2004; Burg *et al.* 2005; Peters *et al.* 2005), and different subspecies of boreal chickadees are found on either side of the mountains. Island populations often have reduced gene flow to the mainland, likely due

the presence of an impassable water barrier. Newfoundland has a number of endemic species and subspecies (Broders *et al.* 1999; Holder *et al.* 1999; Cronin *et al.* 2005; Hearn *et al.* 2006), including a subspecies of boreal chickadee (American Ornithologists' Union 1957), suggesting that gene flow between the island and the mainland will likely be reduced.

1.7 Thesis Organisation

The thesis is written in four chapters. The first chapter provides a general background of the biological processes which drive population genetic structure in temperate species (the Pleistocene glaciations and barriers to dispersal), and the molecular methods which can be used to study them. Chapter 2 examines the effects of the Wisconsin glaciation on the boreal chickadee, how genetic diversity may have been affected, and the resulting population genetic structure. Analysis of mtDNA data, in conjunction with spatio-geographical modelling, highlights possible glacial refugia and a likely colonisation model. The third chapter looks at population structure using microsatellite markers. I address whether physical barriers (e.g. mountain ranges and bodies of water) act to restrict dispersal and gene flow in this species. The final chapter summarises the main results of Chapters 2 and 3, focusing on the similarities and differences seen with the two molecular markers. I address possible causes for any discrepancies between the two markers. The subspecies designations are compared to the genetic findings. Potential future work is suggested.

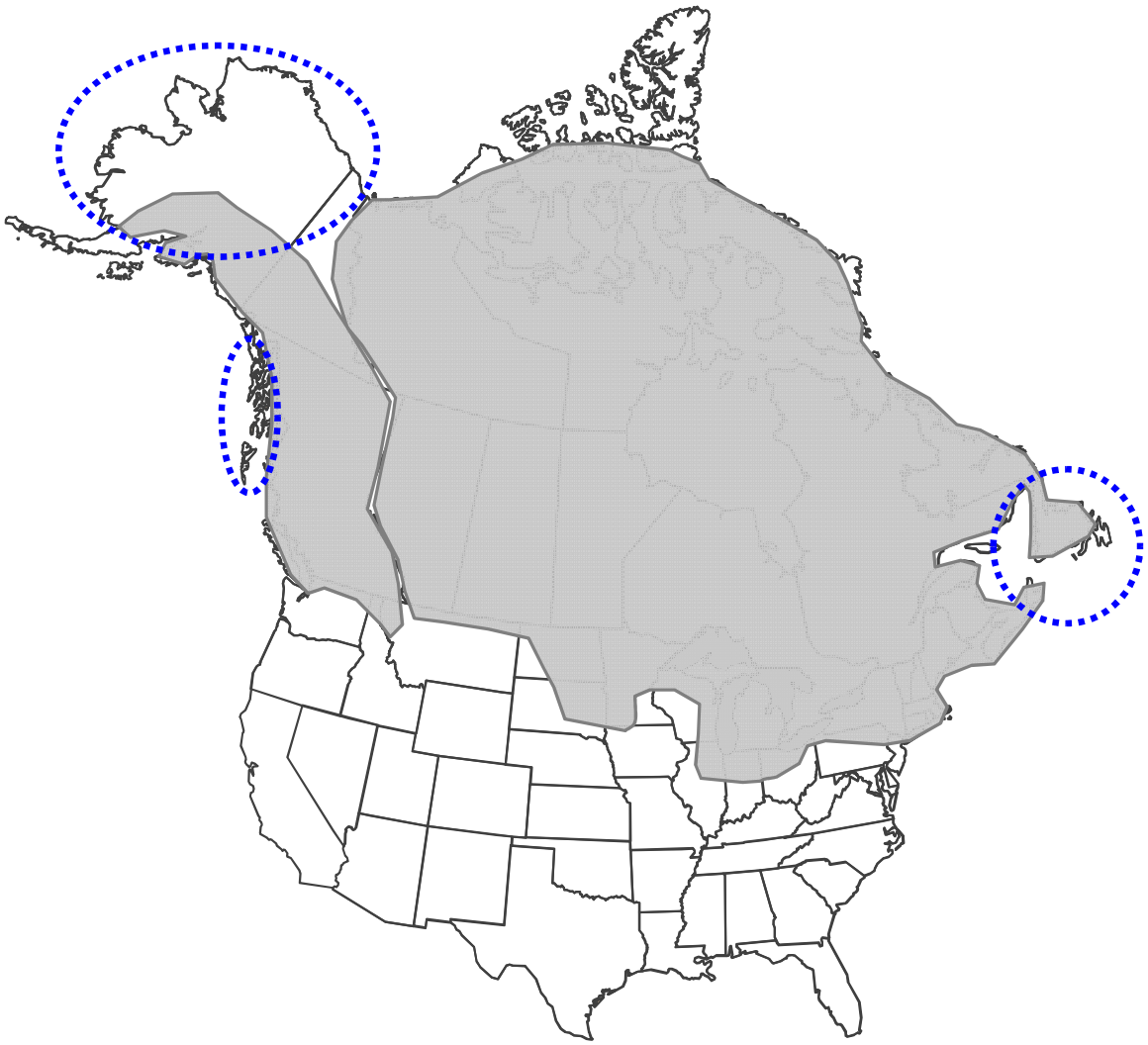

Figure 1.1. Approximate extent of the Laurentide (right) and Cordilleran (left) ice sheets (grey) at the last glacial maximum. Dashed blue lines represent the location of three putative northern refugia (from left to right: Beringia, Haida Gwaii/Alexander archipelago and Atlantic Coast). Areas south of the ice sheets also acted as glacial refugia. Figure modified from Pielou (1991).

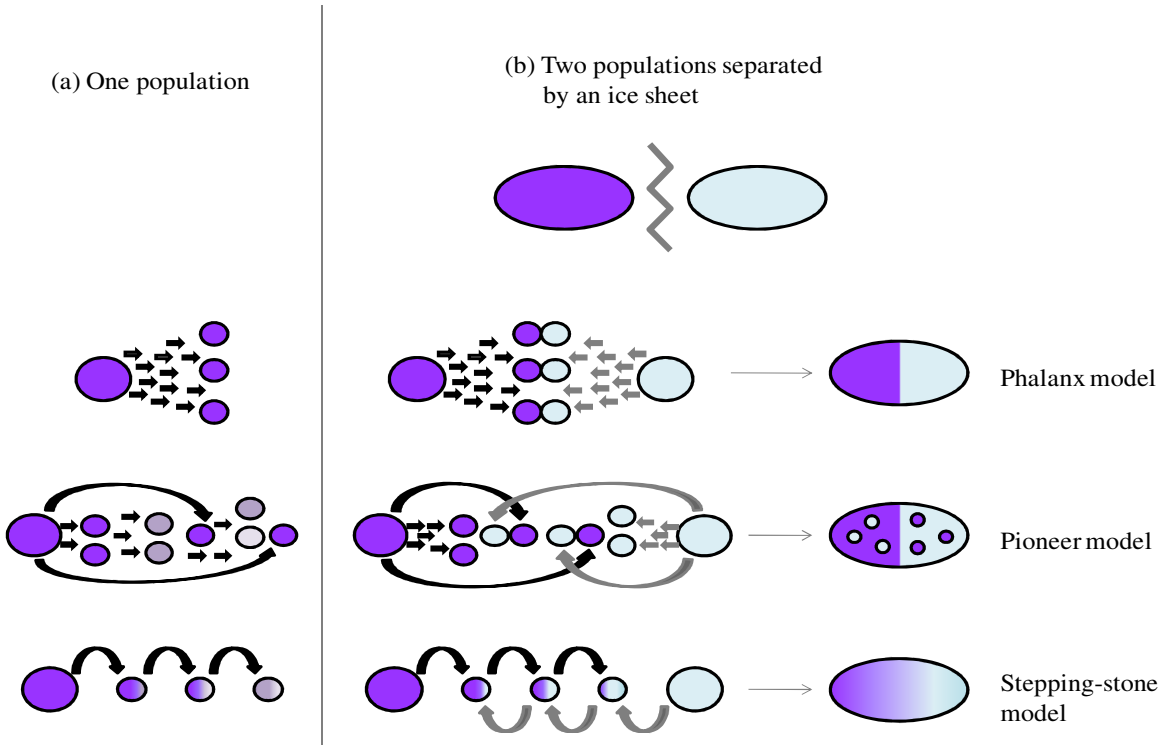

Figure 1.2. Examples of three potential colonisation models and the resulting patterns of diversity if individuals originated from (a) one glacial refugium or (b) two isolated glacial refugia which subsequently underwent secondary contact. In the phalanx model, an advancing wave causes low inter-population diversity and an abrupt transition between populations. The pioneer model is characterised by long-distance dispersal events intermingled with a slow spread of individuals, resulting in pocket populations. In the stepping-stone model, founders were exchanged between neighbouring populations, forming a gradient. In each case, the presence of an existing population either prevents or reduces further movement. Figure modified from Johansen & Latta (2003), and based on colonisation patterns described by Ibrahim *et al.* (1996).

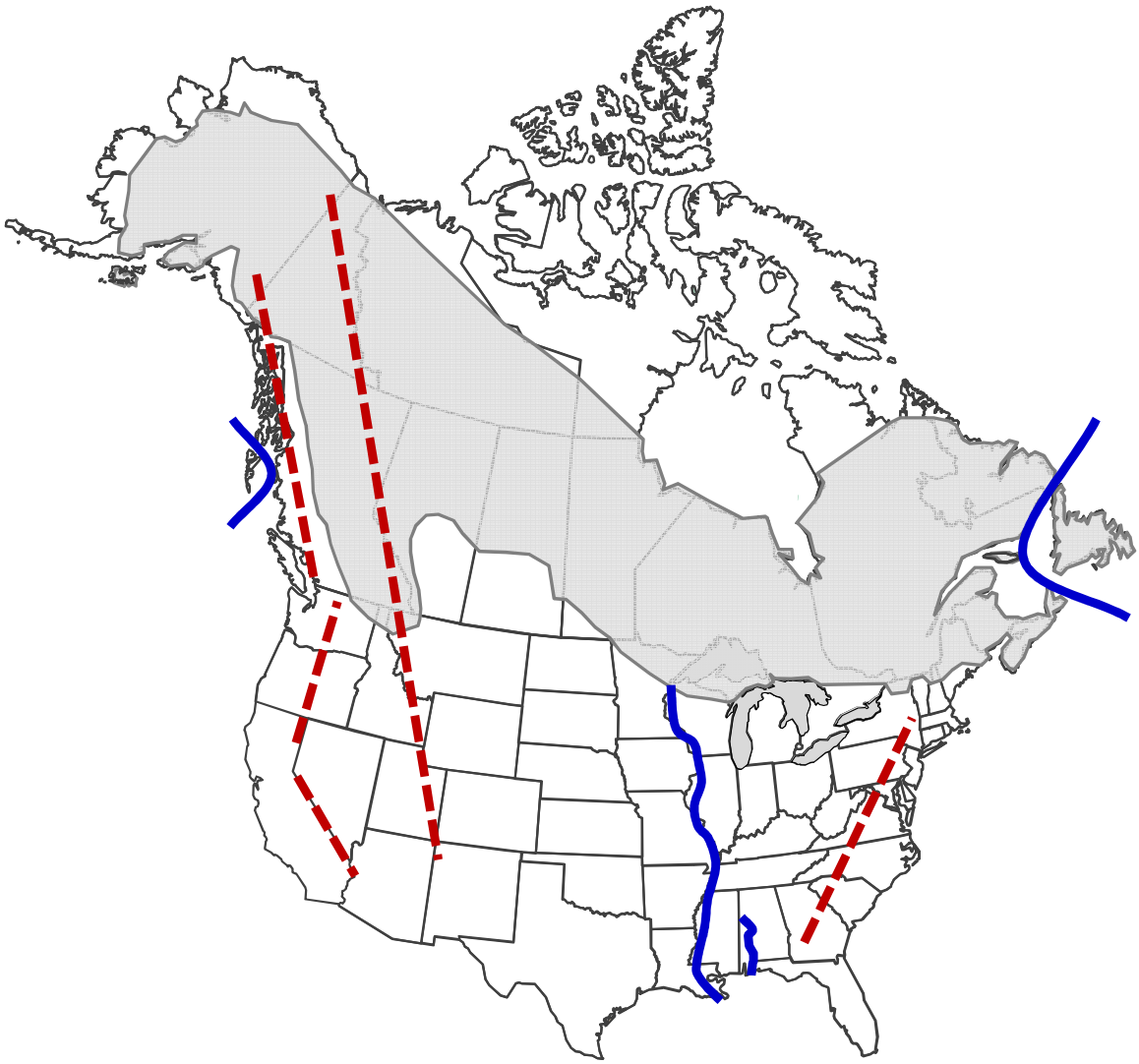

Figure 1.3. Large physical barriers such as mountain ranges (red, dashed; from left to right: Coastal (top), Cascade (middle), Sierra Nevada (bottom), Rocky and Appalachian Mountains) and large bodies of water (blue, solid; from left to right: Hecate Strait, Mississippi and Tombigbee Rivers, and Gulf of St. Lawrence) may prevent dispersal and gene flow in North American species. The distribution of the boreal chickadee (light grey) is modified from Sibley (2003).

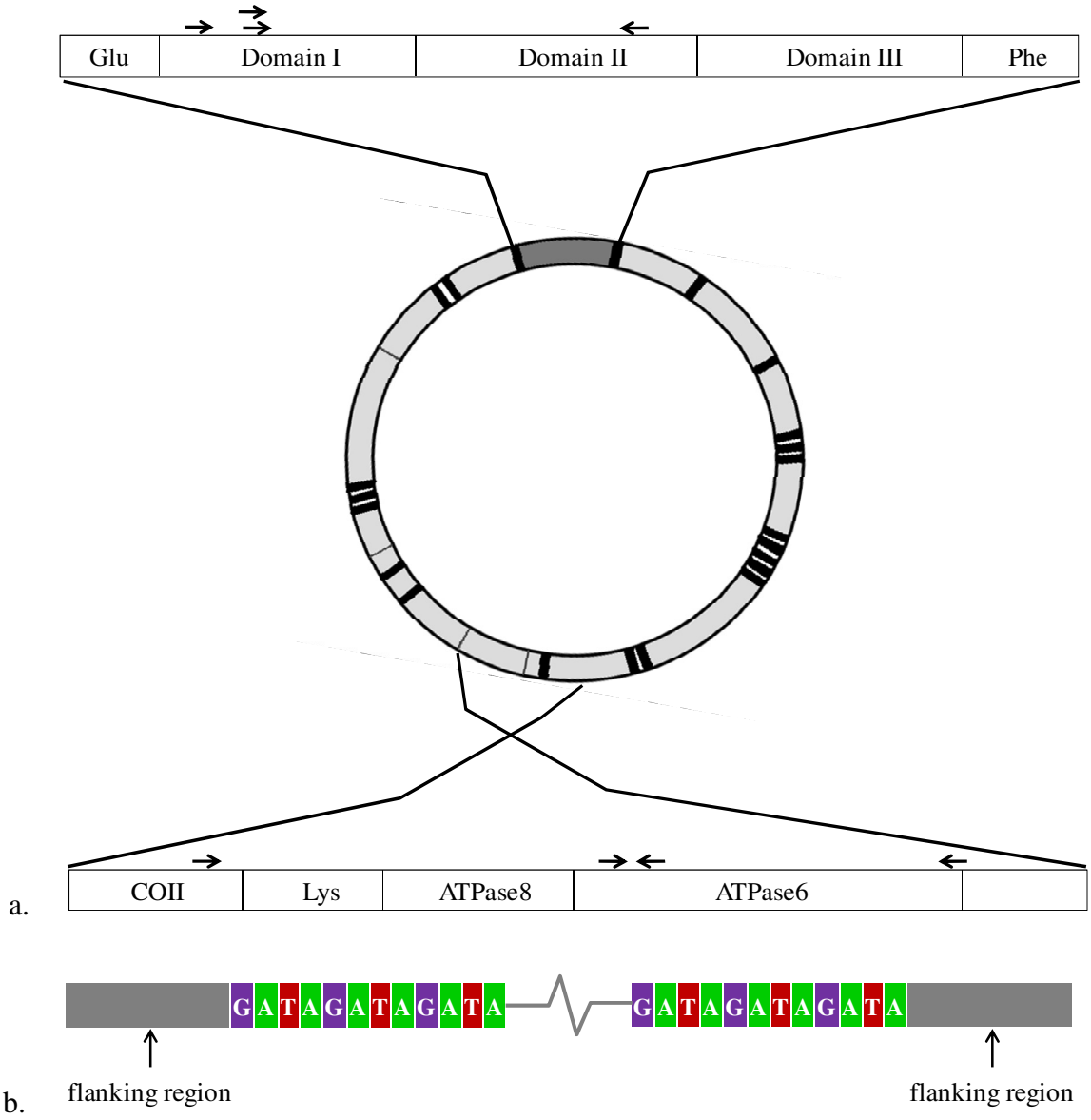

Figure 1.4. (a) The general structure of avian mitochondrial DNA (mtDNA) showing the control region (dark grey), 15 coding regions (light grey) and 22 tRNAs (black lines). The callouts (not drawn to scale) show the areas sequenced in the study (control region, top, and ATP coding region, bottom), and the arrows depict primer binding sites (refer to Chapter 2 for names). Figure modified from Desjardins & Morais (1990) and Kvist (2000). (b) A typical microsatellite sequence. Primers bind in the flanking regions which surround short tandem repeats (i.e. GATA). These short sequences can be dozens of repeats long.

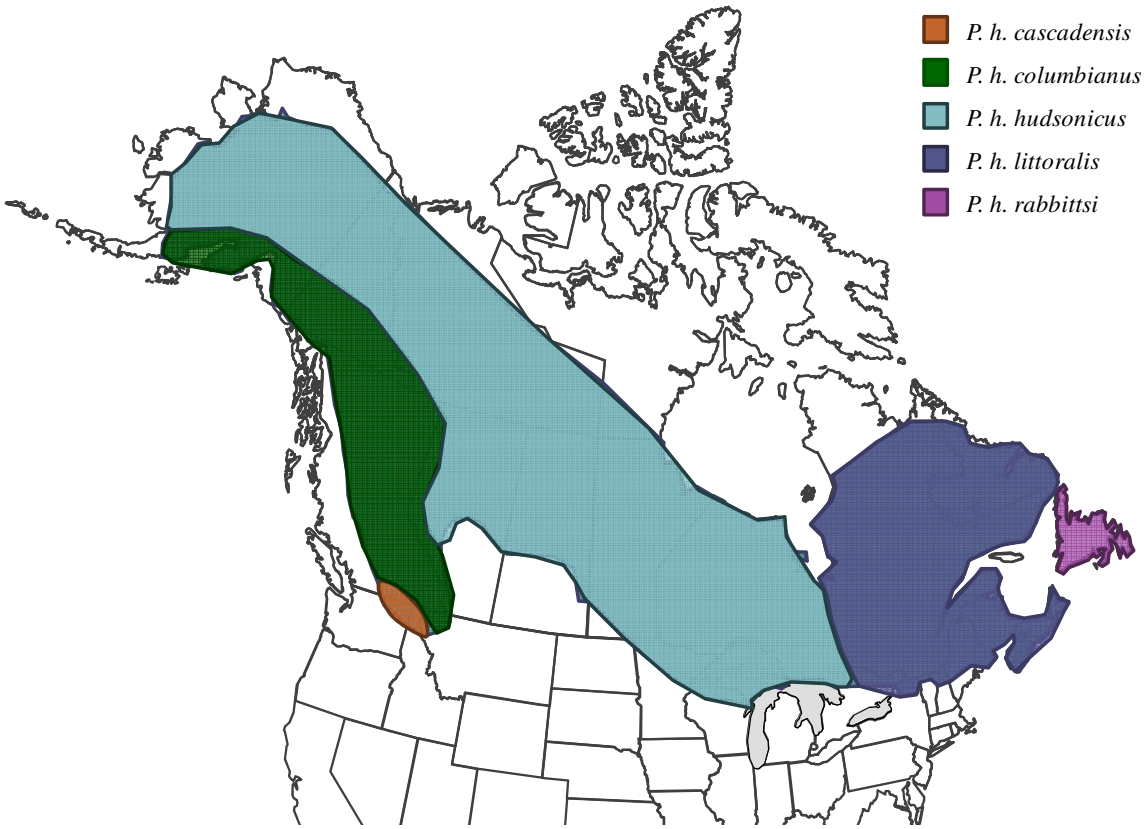

Figure 1.5. Approximate distribution for five described subspecies or boreal chickadee (American Ornithologists' Union 1957) modified from Pyle (1997) and Birds of North America (Ficken *et al.* 1996).

CHAPTER TWO

The Phylogeography of the Boreal Chickadee (*Poecile hudsonicus*)

L.A. Lait[1] and T.M. Burg[1]

[1]Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, T1K 3M4

2.1 Introduction

The Wisconsin glaciation, the most recent major glacial event of the Pleistocene, took place approximately 110 to 12 kya, with the maximum extent of ice occurring between 21 and 18 kya (Pielou 1991). Throughout the Pleistocene, the climate and vegetation in North America was greatly modified, with large ice sheets present in Canada and the northern United States, and much cooler, drier conditions in the southern United States than are typically found today (Adams & Faure 1997). Fossil evidence supports the presence of a number of potential refugia: the United States south of the ice sheets, Beringia, and several putative coastal and mountain refugia (Pielou 1991).

Previous studies have shown that many of the species now found in the Nearctic survived in one or more of these refugia (Avise 1992; Zink 1996). The migratory yellow-rumped warbler *Dendroica coronata* shows little genetic differentiation between eastern and western populations, suggesting a single southern refugium (Milá *et al.* 2006). This pattern is also evident in the downy woodpecker *Picoides pubescens* (Pulgarin-Restrepo 2011), mourning dove *Zenaida macroura* (Ball & Avise 1992), and red-winged blackbird *Agelaius phoeniceus* (Ball *et al.* 1988). Many plant and animal species persisted in both coastal and inland refugia in the south (e.g. black spruce *Picea mariana* (Jaramillo-Correa *et al.* 2004), yellow warblers *Dendroica petechia* (Boulet & Gibbs 2006), and eastern chipmunks *Tamius striatus* (Rowe 2004)). There is a growing body of literature supporting Beringia as a glacial refugium (e.g. white spruce *Picea glauca* (Anderson *et al.* 2006; de Lafontaine *et al.* 2010), rock ptarmigan *Lagopus mutus* (Holder *et al.* 1999), and bighorn sheep *Ovis canadensis* (Loehr *et al.* 2006)). More recently, evidence for the use of a north-eastern cryptic refugia (e.g. Sable Island and the Grand Banks off

Newfoundland; Pielou 1991) has been found in several species of trees (Walter & Epperson 2001; Jaramillo-Correa *et al.* 2004) and birds (Zink & Dittmann 1993; Zink *et al.* 2003; Colbeck *et al.* 2008). The American redstart *Setophaga ruticilla* (Colbeck *et al.* 2008), common yellowthroat *Geothlypis trichas* (Ball & Avise 1992), white-breasted nuthatch *Sitta carolinensis* (Spellman & Klicka 2007), hairy woodpecker *Picoides villosus* (Graham 2011), and loggerhead shrike *Lanius ludovicians* (Vallianatos *et al.* 2001) all show evidence of having persisted in multiple glacial refugia.

As the ice sheets receded and the climate warmed, colonisation of the previously glaciated regions occurred. Three general expansion models have been suggested: the pioneer, phalanx and stepping-stone models (Figure 1.2; Nichols & Hewitt 1994; Ibrahim *et al.* 1996), each producing a different population genetic pattern.

The boreal chickadee (*Poecile hudsonicus*) is a small, specialist songbird found throughout Canada and the northern United States (Figure 2.1). The current range of the boreal chickadee is nearly exclusively in the northern boreal forests, and was almost entirely covered by the ice sheets during the last glacial maximum (Pielou 1991; Ficken *et al.* 1996). Unlike many passerines, the boreal chickadee is a permanent resident, with little or no migration occurring between breeding seasons. This lack of seasonal movement should allow for a simpler genetic structure as there is less opportunity for gene flow to occur.

A previous study by Gill *et al.* (1993) examined the population genetics of North American chickadees. Mitochondrial DNA restriction fragment analysis was used to examine the genetic structure both within and between species. In the boreal chickadee, 37 samples were collected from seven populations. Little genetic differentiation was seen

in the continental populations, which roughly correspond to the previously glaciated region, and a unique haplotype was found in Atlantic Canada, in the Nova Scotia and Newfoundland populations. The low diversity seen may be a result of the low sensitivity of restriction fragment length polymorphisms (RFLPs).

The present study examines the patterns of genetic diversity in the boreal chickadee using an intensive, range-wide sampling regime and more variable mitochondrial DNA (mtDNA) sequencing. MtDNA is an excellent marker for phylogeographic studies as it is uniparentally inherited and therefore does not regularly undergo recombination (Avise 1992, 2004). This allows it to retain historical patterns. As mtDNA is maternally inherited, it will only track the movement and history of females (Avise 2004); however, given the sedentary nature of this species, the maternal lineage should be representative of the species. We used both the non-coding control region and the ATP coding region.

The aim of this study is to evaluate the population genetic structure in the boreal chickadee to determine: (1) Did the postglacial expansion originate from a single or multiple glacial refugia? (2) Did the colonisation occur as a wave (phalanx), stepping-stone or long-distance (pioneer) dispersal? As the current distribution of boreal chickadees includes both Alaska (Beringia) and Nova Scotia and Newfoundland (the putative Atlantic shelf refugia), we predict that the populations expanded from multiple glacial refugia. If the boreal chickadee did survive in multiple, isolated refugia during the last glacial maximum (LGM), multiple divergent groups will be evident having arisen through mutation and drift. We predict that the population genetic structure will show evidence of expansion from multiple glacial refugia – one on the Atlantic coast and one

in Beringia. Within-population genetic diversity will be high, with evidence of a phalanx model of colonisation (Nichols & Hewitt 1994; Ibrahim *et al.* 1996).

2.2 Materials and Methods

2.2.1 Spatio-geographic modelling

2.2.1.1 The model

Both contemporary (ecological niche modelling; ENM) and historical (paleo-distributional modelling) patterns of distribution were estimated using spatio-geographic data. Environmental data (e.g. climatic variables, soil data, vegetation cover, elevation and sunlight) and species occurrence records are combined to estimate the potential distribution of a species (i.e. where they could survive) by assuming that the species is found in the preferred environmental conditions (Phillips *et al.* 2006; Richards *et al.* 2007). This data can then be used in conjunction with paleo-climatic data to extrapolate past distribution, or with climatic predictions to estimate future distributions (Richards et al. 2007). The models rely on the key assumption of niche fidelity – that the current niche requirements are reflective of those in the past and/or future.

The distributions were predicted in MAXENT v3.3.3 which uses a maximum entropy statistical model on presence-only occurrence data (Phillips *et al.* 2006; Phillips & Dudík 2008). The program uses the environmental data in combination with the occurrence data to train the model as to what habitat is considered suitable versus unsuitable for the species in question. The model then assigns the most likely distribution as the posterior distribution which meets all of the functional constraints (based on the environmental data) while maximising entropy (chaos; the most spread out distribution)

(Phillips *et al.* 2006). Ten replicates were run using the cross-validation method and default settings.

2.2.1.2 Occurrence data

Occurrence records were comprised of sampling locations from this study and sightings downloaded from the Global Biodiversity Information Facility (GBIF) data portal (Global Biodiversity Information Facility data portal) and the Avian Knowledge Network (Peterson 1990). Only those with accompanying GPS coordinates were included. Several problems are known to arise with occurrence data: sightings tend to be positively correlated with population density and habitat accessibility (e.g. roads); geographic locations may be incorrect, and species may be misidentified (Phillips *et al.* 2006). In order to test for the effect of incorrect occurrence data, the model was run using all occurrence records (all) and a modified (mod) dataset. For the modified dataset, the records were carefully examined for unlikely sightings. These sightings may have been erroneous (e.g. misidentified or transcribed incorrectly), or may represent rare long-distance movements that do not follow the species' typical habitat requirements. All recorded sightings below 40°N in the east (east of 95°W), and below 46°N in the west, were discarded. This boundary is well outside of the normal expected range (Figure 2.1), and eliminated 239 of 32,379 records. Similar predicted distributions were seen with the full and modified datasets, with slight differences in the contribution of each environmental variable (see Results).

2.2.1.3 Environmental data

The distribution model was based on 19 WorldClim climatic variables (see Results for full list) extrapolated from GIS layers as described in Carstens *et al.* (2007). The variables included average, extreme and ranges of temperature (temp) and precipitation (precip), and were available for both the contemporary timescale and estimates for the last glacial maximum (ca. 21 kya). The MIROC (a Model for Interdisciplinary Research on Climate) climate layers used as the past climate estimates were provided by the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2; Waltari *et al.* 2007).

2.2.2 Sample collection

Two hundred and thirty samples were collected over four breeding seasons from 11 sampling locations across the chickadees' range (Figure 2.1). Sampling locations, hereafter referred to as populations, were limited to a 50 km radius whenever possible, with no obvious barriers to dispersal, with the exception of the NSNB and NL sites. In these two cases, multiple sites were combined after confirming there were no population differences. Birds were caught using 12 m mist nets and call playback. A 100 µL blood sample was collected from the brachial vein and stored in ethanol. All birds were banded and released. The samples were returned to the lab and stored at -80°C.

Fifty-three museum samples (Appendix 1) augmented sample sizes from field sites and added an additional three sampling locations (Figure 2.1; 20 samples were from existing sites and 33 in new locations). Two samples were not sequenced. Samples were provided by the American Natural History Museum (n = 3), Burke Museum (n = 2), Canadian Museum of Nature (n = 11), New Brunswick Museum (n = 3), New York State

Museum (n = 3), Royal Alberta Museum (n = 1), Royal Ontario Museum (n = 16), Royal British Columbia Museum (n = 1), and the Smithsonian Institution National Museum of Natural History (n = 13).

2.2.3 DNA extraction

DNA was extracted from blood, feather and tissue samples using a modified chelex procedure (Walsh et al. 1991). A small sample (10 µL blood or a 3 mm slice of tissue/feather) was added to 300 µL DNA extraction buffer (0.1 M Tris buffer (pH 8), 0.05 M EDTA, 0.5 M NaCl, 1% SDS and 5% w/v chelex), 3 µL proteinase K (20 mg/mL) and 2.5 µL RNase (10 mg/mL). After incubating overnight at 50°C, the extracted DNA was vortexed, centrifuged at 10,000 rpm for two minutes, transferred into 300 µL 1X low TE buffer (0.01 M Tris buffer (pH 8), 0.0001 M EDTA and 5% w/v chelex) and stored at -20°C.

2.2.4 DNA amplification

Two fragments of mtDNA were amplified using the polymerase chain reaction (PCR). A 766 base pair (bp) fragment of the control region (CR) contained part of domain I (295 bp) and most of domain II (471 bp). The control region is the only major non-coding region found in avian mtDNA (Ruokonen & Kvist 2002), and as such is not under functional constraint. A 923 bp coding fragment (ATP) contained three coding regions and a transfer RNA (tRNA): part of the cytochrome oxidase II gene (61 bp), the tRNA lysine (71 bp), and the overlapping ATPase 6 (168 bp) and ATPase 8 (627 bp) genes (Desjardins & Morais 1990). All PCR reactions were performed in an Eppendorf

Mastercycler ep gradient thermocycler. PCR products were visualised on a 0.8% agarose gel.

The control region was amplified using the primers LmochCR1 and H1015chCR (Table 2.1). For some of the museum samples, the DNA was degraded and a semi-nested PCR was required for successful amplification. For these samples, the first round amplification used L26chCR and H1015chCR, while the second round used the internal primer LmochCR2 with H1015chCR (Table 2.1).

The PCR reaction (25 µL) contained 1X GoTaq Flexi buffer (Promega), 2.5 mM $MgCl_2$ (Promega), 0.2 mM dNTP (Fisher Scientific), 4 µM forward and reverse primer, 1 U GoTaq Flexi DNA polymerase (Promega) and genomic DNA. All PCRs were done using the following thermal cycling program: an initial cycle of 2 min at 94°C, 45 s at 54°C, and 60 s at 72°C; followed by 37 cycles of 30 s at 94°C, 45 s at 54°C, 60 s at 72°C; and one final extension at 72°C for 5 min followed by 4°C for 20 s.

The ATP gene fragment was amplified using universal avian primers L8929 COII and H9855 ATP6 (Table 2.1; Sorenson *et al.* 1999). For some museum samples, semi-nested amplifications were done using the internal primers H534chATP and L298chATP (L8929/H534 and L298/H9855; Table 2.1). The ATP gene region was amplified in a 25 µL reaction which contained 1X Crimson buffer (New England Biolabs), 1.5 mM $MgCl_2$ (New England Biolabs), 0.2 mM dNTP (Fisher Scientific), 4 µM of an H and L strand primer and 1 U Crimson Taq DNA polymerase (New England Biolabs). A similar PCR program to the control region amplification was used with the following modification: initial reactions were performed with an annealing temperature of 58°C and the second round of semi-nested PCRs were run with an annealing temperature of 48°C.

2.2.5 Sequencing

All PCR products (5 µL) were purified with an Exo-SAP clean up. This consisted of a 15 min incubation step at 37°C with 0.1 U exonuclease I (USB Corporation) to remove excess primers and 0.1 U shrimp alkaline phosphatase (Promega) to dephosphorylate unincorporated dNTPs, followed by a 15 min step at 80°C to denature both enzymes. The sequencing reaction (10 µL) contained 0.25 µL Big Dye v3.1 (Applied Biosystems), 1X sequencing buffer (McLab), 0.3 µM of one primer and 1 µL purified, Exo-SAP-treated PCR product (approximately 100 ng DNA). Sequencing reactions consisted of one cycle at 96°C for 2 min; 25 cycles of 96°C for 30 s, 50°C for 15 s; and a final extension at 60°C for 4 min. The sequences were cleaned using an ethanol precipitation (removes excess primer, RNA, proteins and sequencing reagents) and sequenced using capillary electrophoresis on a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA). Some PCR products from blood samples and most museum samples were sent to Genome Quebec for sequencing (McGill University, QC).

2.2.6 Genetic analyses

As the mitochondrial genome is a single continuous strand of DNA, all analyses were run on a concatenated sequence of the CR and ATP fragments because the genes are physically linked and therefore inherited as a single unit. The analyses were also performed on the two fragments separately; the results were similar and are not shown. The DNA sequences were checked by visual inspection of the chromatograms and aligned in the program MEGA 4.0 (Tamura *et al.* 2007). All variable sites were double checked against the chromatograms. Haplotypes (unique DNA sequences) were assigned

to individuals manually, and confirmed using both MultAlin (Corpet 1988) and TCS v1.21 (Clement *et al.* 2000).

Nucleotide and haplotype diversity were calculated in DnaSP v5 (Rozas & Rozas 1999; Rozas *et al.* 2003; Librado & Rozas 2009) as a measure of genetic diversity in the populations; nucleotide diversity was calculated using the equation $\pi = \sum_{ij} x_i x_j \pi_{ij}$ where x_i and x_j are the frequency of the i^{th} and j^{th} sequence (Nei & Li 1979), and haplotype diversity as $h = \left(\frac{n}{n-1}\right)(1 - \sum_i x_i^2)$ where n is the sample size and x_i is the frequency of the i^{th} haplotype (Nei & Tajima 1981). A statistical parsimony network was constructed to visualise the relationship between haplotypes. The network was constructed in TCS v1.21 (Clement *et al.* 2000) with a 95% connection limit, and all connections were confirmed by visual inspection.

Pairwise genetic differences (Φ_{ST} values), based on Wright's fixation index (Wright 1965), were calculated in Arlequin v3.11 (Excoffier *et al.* 2005). Pairwise Φ_{ST} values between two populations range from 0 (panmixia) to 1 (completely isolated; Wright 1965). In order to correct the critical significance value for multiple tests (multiplicity; Benjamini & Hochberg 1995), a modified false discovery rate (FDR) procedure (Benjamini & Yekutieli 2001) was applied, as described in Narum (2006; critical p-value = $\frac{\propto}{\sum_i^1}$; i = 1 to k where k is the number of tests). Traditional Bonferroni corrections are good at preventing type I errors (false positives), however they substantially reduce the power of the tests by introducing type II errors (false negatives; Moran 2003; García 2004; Nakagawa 2004; Verhoeven *et al.* 2005). The FDR procedure was first described by Benjamini and Hochberg (1995) to increase test power, and a modified FDR was described by Benjamini and Yekutieli (2001). The modified FDR is intermediate between

Bonferroni and FDR in controlling for both type I and type II errors, and has been shown to be the most effective (Narum 2006). Critical p-values and the number of significant p-values were calculated using all three methods.

A Mantel's test was performed to examine the correlation between genetic and geographic distances in Genepop v4.0.10 (Raymond & Rousset 1995; Rousset 2008). A significant correlation between genetic and geographic distances suggests that there is isolation-by-distance (Slatkin 1993). Geographic distances were calculated using the weighted average GPS coordinates for each population and the Geographic Distance Matrix Generator v1.2.3 (Ersts 2010), and linearised Φ_{ST} values were used for genetic distances. Significance was tested using 10,000 permutations.

The allocation of genetic variation within and among populations was tested using an analysis of molecular variance (AMOVA; 100,000 permutations) in Arlequin v3.11 (Excoffier *et al.* 2005). A spatial analysis of molecular variance (SAMOVA; Appendix 2; Dupanloup *et al.* 2002) uses both geographical and genetic data to detect population genetic structure. The SAMOVA partitions the populations into groups in order to maximise Φ_{CT} (between-group variance; Excoffier *et al.* 1992) for a specified K (K = 2 to 13; 100 iterations). As a result it identifies both maximally differentiated groups and potential genetic barriers, although the process is fairly sensitive to isolation-by-distance (Dupanloup *et al.* 2002).

A principal coordinates analysis (PCO; Appendix 2) was performed in GenAlEx v6.3 (Peakall & Smouse 2006) on both individuals and population pairwise differences. The PCO allows the visualisation of patterns in a set of random data by assigning eigenvalues (or principal components) to the data (see Appendix 2). Although the PCO

on summary statistics may lose some of its power, the resulting eigenvectors are much easier to see due to the overlapping nature of the individual samples.

Cluster analysis was run using Bayesian Analysis of Population Structure (BAPS) v5.2 (Corander *et al.* 2008). BAPS uses Bayesian analysis to assign individuals to K clusters based on genetic data, with no *a priori* population information. To my knowledge, this is the only clustering analysis software that allows haploid sequence data to be used; all other programs use nuclear allele frequencies (e.g. amplified fragment length polymorphisms (AFLPs) and microsatellites). As per Corander *et al.* (2008), the analysis used the clustering with linked loci option (Corander & Tang 2007) and variable K (K = 1 to 20). The control region and ATP sequences were concatenated and the number of clusters (K) determined based on maximising the log marginal likelihood of the best visited partitions.

2.2.7 Molecular clock calibration

Many avian species are not well represented in the fossil record (Lovette 2004). Molecular calibrations, therefore, are often made using a combination of paleo- and biogeographical data (e.g. volcanic activity, the formation of islands, and the introduction of geographical barriers; Gill *et al.* 2005; Päckert *et al.* 2007; Weir & Schluter 2008).

Traditionally, mtDNA has been believed to have a standard 'molecular-clock' rate of 2%/My (Brown *et al.* 1979). This rate was originally calibrated in primates (Brown *et al.* 1979), but has also been reported in the snow goose *Chen caerulescens* (Shields & Wilson 1987) and the Hawaiian honeycreepers (subfamily Drepanidinae; Fleischer *et al.* 1998). Since then, the divergence rate for mtDNA cytochrome b has been calibrated in a

number of avian lineages (Weir & Schluter 2008). It has been shown to vary from 0.95% to 3.74%/My (Weir & Schluter 2008) in passerines, and, more specifically, 0.7% to 1.8%/My in Paridae (*Poecile*; Päckert *et al.* 2007), with an overall average rate in birds of 2.1%/My (Weir & Schluter 2008). These approximate rates can be used to calibrate the divergence times for some of the North American chickadees.

Cytochrome b sequences were downloaded from GenBank (*P. rufescens*: EU075470-075501 (Topp & Winker 2008), AF347948 (Gill *et al.* 2005); *P. hudsonicus*: AF347949 (Gill *et al.* 2005); *P. cinctus*: AF347950 (Gill *et al.* 2005)) and aligned in MEGA 4.0 (Tamura *et al.* 2007). Using the lowest (0.7%), average (2.1%) and highest (3.7%) estimated mutation rates for cytochrome b, approximate divergence dates were calculated according to Wilson *et al.* (1985) using the formula $t = \delta/\mu$, where δ is the corrected average pairwise difference $\left(\delta = \delta_{xy} - 0.5\left(\delta_x + \delta_y\right)\right)$ (Arlequin v3.11; Excoffier *et al.* 2005) and μ is the divergence rate / locus / million years (Table 2.2).

The approximate divergence dates match those calculated by Gill *et al.* (2005) using both cytochrome b sequences and DNA-DNA hybridisation distances (approximately 1.5 million years ago). The calculated dates were then used to compute a range of approximate divergence rates for the ATP and control region fragments using $\mu = \delta/t$ (Table 2.2). For the ATP fragment, only *P. hudsonicus* (this study) and *P. rufescens* (JF514157-514176; Lait *et al.* unpublished) sequences were available. For the control region, *P. cinctus* (AF257185; Uimaniemi *et al.* 2003), *P. hudsonicus* (this study), and *P. rufescens* (JF514177-514215; Lait *et al.* unpublished) sequences were used.

The calculated divergence rate for ATP was 0.59% to 3.17%/My, similar to the 2%/My widely used for mtDNA coding regions (Lovette 2004). For the control region,

we found the rate to be 0.53% to 3.75%/My, with slightly higher rates in domain I (0.82% to 4.83%/My) than in domain II (0.20% to 2.02%/My). While these figures are considerably lower than traditionally considered for this non-coding region (Quinn 1992; Baker & Marshall 1997; Lambert *et al.* 2002; Tris *et al.* 2004; Milá *et al.* 2007), similar rates have recently been calibrated in a number of parids (Kvist *et al.* 2001; Ruokonen & Kvist 2002; Pereira & Baker 2006; Päckert *et al.* 2007), cracids (Pereira *et al.* 2004), goldcrests (Päckert *et al.* 2006), and the fox sparrow *Passerella iliaca* (Zink *et al.* 2003). While using the cytb sequences to calculate divergence times is not ideal, the congruence with Gill *et al.*'s (2005) divergence date estimates provides additional support.

Using the divergence rates calculated from the average 2.1%/My cytochrome b rate, I calculated a weighted average divergence rate for the boreal chickadee sequences as follows: 295 bp of CR domain I, 471 bp of CR domain II and 923 bp of ATP. This weighted rate of 1.70%/My, with low (0.52%/My) and high (3.14%/My) weighted rates as ranges, was used to calculate approximate divergence dates using $t = \delta/\mu$ (Wilson *et al.* 1985). Corrected average pairwise difference were calculated in Arlequin v3.11 (Excoffier *et al.* 2005).

2.3 Results

2.3.1 Spatio-geographic modelling

The maximum entropy model performed significantly better than random, as shown by both the binomial test of omission and the receiver operating characteristic (ROC) analysis (Figure 2.2). The fact that the tested omission rate was close to the predicted omission rate is indicative of a low number of occurrences in areas deemed unsuitable by

the model ($p < 0.0001$; Phillips *et al.* 2006). The area under the ROC curve (AUC) was 0.829 (±0.004). The higher the AUC, the better the model is at discriminating between unsuitable and suitable habitat (Phillips *et al.* 2006). As we are using presence-only data, the highest possible AUC value is 1 and random predictions would have an AUC of 0.5 (Carstens *et al.* 2007).

The present distribution predicted by the model (Figure 2.3) closely matches the current range of the boreal chickadee found in many field guides and other resources (Figure 2.1, e.g. Birds of North America, Sibley's, National Geographic). The potential distribution at the last glacial maximum (ca. 21 kya) shows suitable habitat in Alaska, along the western coast of North America, south of the ice sheets , and in Newfoundland in the east (Figure 2.3). Similar patterns were seen with both the complete and modified occurrence dataset.

The contribution of each of the environmental layers varied considerably and the layers are not independent of each other (Table 2.3). The highest contributions were from the annual mean temperature (34.9%) and the isothermality (14.1%). When all occurrence values were used, the mean temperature in the coldest quarter (23.5% all) was an important factor, and no longer appeared as a large contributor in the modified dataset.

2.3.2 Population structure

Two fragments of mtDNA were successfully amplified for 281 samples from 14 populations. A 766 bp fragment of the non-coding control region contained 56 variable sites, 34 of which were parsimony informative (found in more than one sequence; Appendix 3). Almost two-thirds of the variable sites in the control region were found in

domain I (34 of the 56), and of those found in domain II, only eight were located in one of the conserved boxes (Appendix 4). The average nucleotide composition of the light strand was 28.1% T, 29.7% C, 24.6% A and 17.6% G. A 923 bp fragment of coding DNA (ATP) contained 53 variable sites (21 parsimony informative; Appendix 3). The variable sites were distributed as follows: 18 in the first codon position, four in the second, 28 in the third, and three in the tRNA (Appendix 4). There were no unexpected stop codons. The average nucleotide composition of the light strand was 22.8% T, 39.2% C, 28.4% A and 9.6% G. Both fragments of mtDNA contained fixed differences between the two groups (Appendix 3): site 586 in the CR is a 'C' in almost all eastern birds, with only a single NL bird having a 'T'; and site 1444 in the ATP is an 'A' in most eastern birds, and again only a single NL bird (the same individual) had a 'G' at this site.

When the fragments were concatenated, there were a total of 109 variable sites: one insertion/deletion (in the non-coding control region fragment), 105 transitions and three transversions. No site had more than two base variants. There were 127 haplotypes, 26 shared between two or more individuals and 101 unique haplotypes (Table 2.4). Of the 26 shared haplotypes, 12 were restricted to a single population, seven were found in two or more western populations, four in two or more eastern populations, and three were present in populations across the range. No haplotypes were shared among the most eastern populations (LAB, NY and NL) and the western populations (Alaska, British Columbia and Alberta).

All of the populations shared a haplotype with at least one other population. Haplotype diversity was high in all populations (0.600 to 1.000) and nucleotide diversity ranged from 0.0009 to 0.0032 (Table 2.5). Some of the highest diversity values were

found in AKF and NON, while one of the lowest (π) was in NL (Table 2.5). Although both haplotype and nucleotide diversity were lower in NL, it had the second highest number of private haplotypes (Table 2.4).

The statistical parsimony network (Figure 2.4) showed that although there was little population genetic structure in the boreal chickadee, what there was supported some geographic structure. There was very little overlap among the eastern (NSNB, LAB, NY and NL; haplotypes P and R to Z) and western (AKA, AKF, AKW, NBC, CBC, CAB and SAB; haplotypes A to O) populations, and the central populations (SK, NON and NQC) were present throughout the network. Similar patterns were seen when the ATP and CR fragments were analysed separately (not shown).

2.3.3 Genetic analyses

Pairwise Φ_{ST} values (Table 2.6) showed NL to be significantly different from all other populations, and AKA and AKW were significantly different than all but each other. The few values that were not significant were between neighbouring populations or a series of adjacent sites (e.g. NQC, NSNB, LAB and NY). The smallest Φ_{ST} values were between CBC and CAB; LAB and NQC and LAB and NSNB. All negative values were treated as zero. Using the modified FDR critical p-value (P_{crit} = 0.0098), 70 of 91 pairwise differences were significant. Little difference was seen with the more powerful original FDR method (P_{crit} = 0.0418; 76 of 91 were significant), and notably fewer were significant with the more conservative sequential Bonferroni method (P_{crit} = 0.0015; 57 of 91).

A significant correlation between genetic and geographic distance was found when all populations were considered ($p < 0.001$; $r^2 = 0.51$; Figure 2.5). The AMOVA found that the majority of the variation was within populations (72.2%) rather than among populations (27.8%; $\Phi_{ST} = 0.278$, $p < 0.001$). The SAMOVA detected the presence of two groups ($\Phi_{CT} = 0.305$, $p < 0.001$), although there was little difference seen between two and three groups ($\Phi_{CT} = 0.301$, $p < 0.001$). The two groups found by the SAMOVA were a western group (AKA, AKF, AKW, NBC, CBC, CAB, SAB, SK and NON) and an eastern group (NQC, NSNB, LAB, NY and NL), and with three groups NY (n = 6) was separated from the eastern populations. As the SAMOVA uses a simulated annealing process (Dupanloup *et al*. 2002), all possible groups of neighbouring populations are tested, and it is unlikely to mistake a local maximum for the global maximum.

Both the PCO on the individual samples and on the population Φ_{ST} values supported the separation of the eastern and western populations, with central populations (SK, NON and NQC) falling intermediately (Figure 2.6). In the PCO on the populations, most of the variation was explained by the first coordinate (74.0% coordinate 1, 10.0% coordinate 2 and 8.8% coordinate 3). The variation in the individual PCO (not shown) was explained by both coordinates 1 (56.0%) and 2 (20.4%). Individuals clustered in groups (e.g. east versus west), but not strictly in populations, and showed less well defined structure.

Bayesian clustering analysis separated the samples into three clusters: one primarily in the east (blue), one primarily in the west (red), and a third almost exclusively in the west (yellow; Figure 2.7). The five highest log maximum likelihood values ranged from -2376 (K = 3) to -2557 (K = 5), and the probability of three clusters was one.

2.3.4 Divergence time estimates

Using the calculated divergence rate estimates (see Materials and Methods), approximate divergence times were calculated among populations (Table 2.7). The estimated dates were based on the rate calculated using the 2.1%/My cytochrome b divergence rate, and the ranges are based on the low and high calculated rates. The two most geographically distinct populations (AKA and NL) diverged approximately 77.8 kya (42.0-253.1 kya), while AKA and AKF diverged approximately 13.0 kya (7.0-42.3 kya). The greatest divergence times are found between CBC and NL and CBC and NY, and there is no evidence of divergence between CAB and CBC or between NSNB and LAB (Table 2.7). These dates must be used with caution as the estimation of divergence rates is uncertain. However, the fact that both the ATP and CR fragments, when analysed separately, gave similar dates is encouraging (not shown). When the estimated divergence times were compared to the geographic distance from putative refugia (NL and AKF; Figure 2.8), a significant strong positive correlation was found ($p < 0.001$). A similar pattern was seen if LAB or NSNB was tested as the eastern putative refugium.

2.4 Discussion

2.4.1 Evidence for multiple glacial refugia

Two main scenarios could explain the pattern of variation seen in the haplotype network (Figure 2.4): (1) the use of multiple isolated refugia with subsequent mixing following colonisation; or (2) the use of a single central refugium followed by outward movement as the ice receded. In both cases genetic diversity would be high in the central populations. However, in the case of a single refugium, diversity would be expected to

decrease with increasing geographic distance from the centre due to sequential founder effects (Hewitt 2004b), whereas with multiple refugia, diversity should be high in both the refugial populations and a secondary contact zone would be present containing a mixture of a subset of haplotypes from each refugia (Taberlet *et al.* 1998; Petit *et al.* 2003). Evidence rejecting the central refugium hypothesis includes diversity and haplotype distribution.

One piece of evidence for the use of multiple refugia and not a central refugium comes from the lack of shared haplotypes between the eastern and western groups, while the central populations contained a subset of each (Figure 2.4). If the central refugium hypothesis was correct, then the western and eastern populations should contain a subset of haplotypes found in central populations. Paleogeographic modelling showed multiple patches of suitable habitat at the LGM (Figure 2.3), consistent with putative refugia in boreal tree species (Jaramillo-Correa *et al.* 2009 and references therein). The estimated divergence times place the separation towards the beginning of the Wisconsin glaciation, with the populations likely becoming physically isolated as the Laurentide ice sheet expanded and genetically isolated some time thereafter.

With the extremely high haplotype diversity in AKF, a third possibility is that there was a single western refugium that expanded eastwards as the ice receded, with a long distance dispersal event forming the eastern group (pioneer model). A single Beringian refugium has been suggested for many arctic plants (Hultén 1937; Eidesen *et al.* 2007) as well as several endemic bird species (Pruett & Winker 2005). However, given the pattern of diversity (high in west, centre and east), and the presence of unique eastern haplotypes that are absent in AKF, this scenario is unlikely. If there was a single western refugium,

we would expect to find some shared haplotypes between eastern and western populations, and lower diversity in the east (NSNB and LAB), particularly if it was founded by long distance dispersal (see Table 2.4 and Figure 2.4).

2.4.2 Possible refugia locations

Studies on both black and white spruce, the boreal chickadee's preferred habitat, found evidence of both species in Alaska and the north-east during the LGM using both DNA and pollen data (Anderson *et al.* 2006; de Lafontaine *et al.* 2010; Gerardi *et al.* 2010). In fact, the population genetic pattern of the boreal chickadee closely resembles that of these tree species: both the present-day distribution and the likely historical range. Evidence of white spruce has been found in Alaska, the south-central plains, and in the north-east (Anderson *et al.* 2006; de Lafontaine *et al.* 2010), and black spruce in Alaska, multiple southern refugia and a north-eastern refugium near Newfoundland and Labrador (Jaramillo-Correa *et al.* 2004; Gerardi *et al.* 2010). These putative refugia are congruent with those found by the paleogeographical modelling (Figure 2.3).

The Beringia land bridge between North America and Asia, which included much of Alaska and the Yukon, supported a number of plant and animal species through the last glacial maximum (e.g. Hultén 1937; Fedorov & Stenseth 2002; Fleming & Cook 2002; Flagstad & Røed 2003; Hewitt 2004a; Brubaker *et al.* 2005). The extremely high diversity in AKF boreal chickadees, as well as the high number of private haplotypes found in this population, indicates that this area was used by the boreal chickadee during the most recent glacial period. A number of other birds are also believed to have used this area including the dunlin *Calidris alpina* (Wenink *et al.* 1996), rock sandpiper *Calidris*

ptilocnemis (Pruett & Winker 2005) and rock ptarmigan *Lagopus mutus* (Holder *et al.* 1999).

The role of Newfoundland or an Atlantic shelf refugium during the Wisconsin is still debated, though there have been both macrofossils and pollen data found in the area (Pielou 1991). Although the haplotype diversity in NL was lower than other populations (due to many birds containing an identical haplotype), the surrounding populations (LAB and NSNB) exhibited relatively high diversity while containing only the eastern haplotypes. This could be due to the use of the nearby Atlantic shelf as glacial refugium followed by a bottleneck in Newfoundland as the ice sheets melted and the sea levels rose. The entire area surrounding the eastern Newfoundland peninsula was exposed with the drop in sea levels associated with the glaciations, and as the sea levels increased the islands were again separated from the mainland (Pielou 1991). The large number of private haplotypes in this area suggests that NL may be an older population.

Several passerines are reputed to have survived in the Newfoundland area: the song sparrow *Melospiza melodia* (Zink & Dittmann 1993), fox sparrow *Passerella iliaca* (Zink *et al.* 2003), American redstart *Setophaga ruticilla* (Colbeck *et al.* 2008), and black-capped chickadee *Poecile atricapillus* (Gill *et al.* 1993). The two spruce species mentioned above, as well as the red pine *Pinus resinosa* (Walter & Epperson 2001), also support the use of a north-eastern refugium. The paleogeographical model for boreal chickadees supports suitable habitat both south of the ice sheets in mid-latitude United States and in Newfoundland. While there is currently no definitive way to distinguish the exact location of the north eastern refugium, recent studies confirm the presence of trees

further north than previously assumed (Jaramillo-Correa *et al.* 2009; Godbout *et al.* 2010).

The complex pattern found in the western populations raises the possibility of additional glacial refugia – either in Alaska and/or in the south. The separation of AKA/AKW from the rest of the western populations may suggest an isolated coastal refugium, followed by limited gene flow with AKF. Alternatively, an early dispersal event from a Beringian refugium may prove a more probable scenario: the AKA/AKW populations have slightly lower haplotype diversity than AKF (Table 2.5), and contain a subset of haplotypes found in AKF (Table 2.4, Figure 2.4). The designation of different subspecies of boreal chickadees from Anchorage and Fairbanks supports long term isolation of these populations.

The third cluster identified by BAPS (Figure 2.7), and the pattern seen in the principal coordinates analysis (Figure 2.6), suggest that a southern or mountain refugium may have existed (CBC/CAB/SAB). It has been shown in a number of plants (Soltis *et al.* 1997; Gavin 2009; Gugger *et al.* 2010) and amphibians (Carstens *et al.* 2005) that there was an inland refugium in Idaho. The use of this refugium, or one nearby, has also been discovered in birds (Burg *et al.* 2006; Krosby & Rohwer 2009) and mammals (Hird & Sullivan 2009). Whether the boreal chickadee survived in this inland refugium, or a more distant southern refugium, cannot be determined with the current data. Additional samples along the very southern edge of the range may help to clarify this, but given the current distribution, the absence of many southern populations may make this question unanswerable.

2.4.3 Colonisation pattern

A stepping-stone colonisation is the most likely model to have occurred in this species. Population genetic differences were not significant between neighbouring populations (Table 2.6), with the exception of NBC/CBC and NL/all other populations. If the populations had expanded in a gradual wave, little among-population diversity would be expected (Ibrahim et al. 1996), whereas most of the non-adjacent chickadee populations exhibit significant genetic differences. Similarly, if long distance dispersal had occurred, some geographically distant populations would be more genetically similar than some relatively close populations. The pattern of isolation-by-distance supports movement between neighbouring populations, but no long distant migration (Figure 2.5).

2.4.4 Population genetic structure

The population genetic structure of the boreal chickadee supports the separation of eastern and western groups with fixed differences between them. Gene flow between populations is low as shown by the large number of significant pairwise Φ_{ST} values (Table 2.6). This is similar to the pattern seen in many North American birds (Milot *et al.* 2000; Bull *et al.* 2010; Manthey *et al.* 2011). In the central populations a gradient was found: the two populations west of Hudson's Bay (SK and NON) contained 57-73% western haplotypes (those found in Alaska, British Columbia and Alberta) while NQC contained only 36-45% (some of the individuals had only one of the two fixed differences). This cline is likely the result of a secondary contact zone. When compared to the earlier work on North American chickadees, the eastern populations represent Gill

et al.'s Atlantic haplotype and the central and western populations the continental haplotypes.

The geographic distribution of haplotypes in the two groups differed considerably. Boreal chickadee populations in Atlantic Canada underwent a recent population expansion as indicated by the starburst pattern (Figure 2.4). A similar pattern is found in other Atlantic forest species (de Lafontaine *et al.* 2010; Gerardi *et al.* 2010). The intricate pattern observed in the western group (Figures 2.4 and 2.7) may result from isolated Pleistocene glacial refugia or separation following dispersal from a western refugium with limited subsequent gene flow. Evidence of multifaceted structure in western North America is found in both plants (Brunsfeld *et al.* 2007) and animals (Krosby & Rohwer 2009). In both groups gene flow to other areas is reduced (Tables 2.4 and 2.6, Figure 2.7).

The east and west groups were only separated by the two fixed differences, indicative of either a low mutation rate or a short separation time. The mutation rate in parids, specifically in the control region, has been shown to be much lower than previously reported in birds (see Materials and Methods; also Kvist *et al.* 2001; Ruokonen & Kvist 2002; Päckert *et al.* 2007). Using the lower mutation rate estimate of 1.7% per site per million year, divergence estimates between east and west populations were 56.2-129.7 kya (Table 2.7), suggesting that incomplete lineage sorting rather than post-Pleistocene separation is responsible for the shallow structure.

2.4.5 Conclusions

The results of this study support the use of multiple glacial refugia by the boreal chickadee and a stepping-stone model of colonisation. A cline is evident in the central

populations, suggestive of a contact zone between two isolated groups. The sequence data has allowed considerably more resolution, suggesting patterns of dispersal and the potential locations of refugia. All of the refugial locations are supported by paleogeographical modelling (Figure 2.3), with evidence for suitable habitat in Beringia, along the west coast and in the east.

Table 2.1. Primers used for amplification of mitochondrial DNA fragments. Each reaction uses one L strand and one H strand primer.

Primer name	**Sequence (5' to 3')**	**Source**
Control Region		
H1015chCR	CGC GGG TTT AAC GAA TGT GG	1
LmochCR1	CAG GGT ATG TAT GTC TTT GCA TTC	1
L26chCR	GGR TTG GAT GCA ACT GCC AGC	2
LmochCR2	TTT TTC ATG GTT TAC AGG GTA TG	2
ATP coding region		
L8929 COII	GGM CAR TGC TCA GAA ATC TGY GG	3
H9855 ATP6	ACG TAG GCT TGG ATT AKG CTA CWG C	3
L298chATP	CTT GAC CAT GAA CYT AAG CT	2
H534chATP	ATT AGG GAT GTT ARG ATK AGG GC	2

1 Lait *et al.* (submitted)

2 This study

3 Sorenson *et al.* (1999)

Table 2.2. Approximate divergence times (top) for three *Poecile* species based on cytochrome b divergence rates of 0.7%/My, 2.1%/My and 3.7%/My. Approximate divergence rates (bottom) for the mtDNA ATPase 6-8 coding region and the control region were calculated based on these dates. The rates used in subsequent calculations are shaded.

	0.7%/My	**2.1%/My**	**3.7%/My**
Divergence Times (My)			
P. cinctus-hudsonicus	4.5	1.5	0.8
P. cinctus-rufescens	5.5	1.8	1.0
P. hudsonicus-rufescens	4.9	1.6	0.9
Divergence Rates (%/My)			
ATPase 6-8	0.59	1.78	3.17
Control Region	0.53	1.92	3.75
domain I	0.82	2.62	4.83
domain II	0.20	0.95	2.02

Table 2.3. The environmental layers used in the maximum entropy modelling, as described in Carstens and Richards (2007), and the percent contribution towards the predicted distribution. The values were averaged over 10 cross-validated replicate runs. Both the modified (mod) and complete (all) occurrence records were tested.

#	**Environmental variable**	**% mod**	**% all**
1	Annual mean temp	34.9	8.0
2	Mean diurnal temp range	1.3	1.6
3	Isothermality	14.1	13.7
4	Temp seasonality	1.6	1.5
5	Max temp of warmest month	11.8	5.8
6	Min temp of coldest month	4.1	5.2
7	Annual temp range	1.5	1.5
8	Mean temp of wettest quarter	0.1	0.1
9	Mean temp of driest quarter	0.3	1.5
10	Mean temp of warmest quarter	0.1	0.1
11	Mean temp of coldest quarter	0.8	23.5
12	Annual precipitation	2.0	3.8
13	Precip of wettest month	0.5	0.9
14	Precip of driest month	1.9	2.4
15	Precip seasonality	3.1	2.4
16	Precip of wettest quarter	0.1	0.1
17	Precip of driest quarter	9.6	15.5
18	Precip of warmest quarter	1.9	2.6
19	Precip of coldest quarter	10.3	9.9

Table 2.4. Distribution of shared haplotypes among 14 boreal chickadee populations. Refer to Figure 2.1 for locations.

Haplotype	AKA	AKF	AKW	NBC	CBC	CAB	SAB	SK	NON	NQC	NSNB	LAB	NY	NL	TOTAL
A	12	1	2												**15**
B	2														**2**
C	8	1	9												**18**
D		1		2											**3**
E	1	1													**2**
F				3											**3**
G		1	1												**2**
H		1		5											**6**
I		2													**2**
J	2	1													**3**
K					2										**2**
L					2										**2**
M	2	3	2	1	5	5		6	1						**25**
N	2	3		7	4	1	1	2							**20**
O		2			1	2			1	2	1				**9**
P								3	2	3	1	5		16	**30**
Q								3							**3**
R									1		2				**3**
S											3				**3**
T											1	2			**3**
U											6	3	4		**13**
V												2			**2**
W														2	**2**
X														2	**2**
Y														3	**3**
Z														2	**2**
# unique[1]	**7**	**17**	**11**	**4**	**3**	**4**	**4**	**5**	**10**	**6**	**7**	**10**	**2**	**11**	**101**
# private[2]	8	18	11	5	5	4	4	6	10	6	8	11	2	15	113
# haplotypes	**14**	**28**	**15**	**9**	**8**	**7**	**5**	**9**	**14**	**8**	**13**	**14**	**3**	**16**	**127**
Sample size	**36**	**34**	**25**	**22**	**17**	**12**	**5**	**19**	**15**	**11**	**21**	**22**	**6**	**36**	**281**

[1] only found in one individual

[2] only found in one population

Table 2.5. Sample size (n), haplotype diversity (h), and nucleotide diversity (π) in 14 chickadee populations for the concatenated CR and ATP sequence. Refer to Figure 2.1 for locations.

	n	**h**	**π**
AKA	36	0.844	0.0016
AKF	34	0.986	0.0020
AKW	25	0.873	0.0021
NBC	22	0.848	0.0015
CBC	17	0.868	0.0018
CAB	12	0.833	0.0021
SAB	5	1.000	0.0013
SK	19	0.871	0.0023
NON	15	0.990	0.0032
NQC	11	0.927	0.0018
NSNB	21	0.910	0.0020
LAB	22	0.935	0.0018
NY	6	0.600	0.0022
NL	36	0.800	0.0009
ALL	**281**	**0.965**	**0.0024**

Table 2.6. Population pairwise Φ_{ST} values based on 100,172 permutations (above diagonal), and corresponding p-values (below diagonal). Significant values after modified FDR correction are shown in bold (P_{crit} = 0.0098). Refer to Figure 2.1 for locations.

	AKA	AKF	AKW	NBC	CBC	CAB	SAB	SK	NON	NQC	NSNB	LAB	NY	NL
AKA	*	0.108	0.021	0.157	0.292	0.238	0.250	0.164	0.118	0.271	0.391	0.361	0.469	0.506
AKF	**<0.001**	*	0.078	0.020	0.101	0.066	0.051	0.077	0.121	0.238	0.378	0.348	0.436	0.489
AKW	0.129	**0.004**	*	0.123	0.237	0.179	0.212	0.131	0.101	0.241	0.361	0.335	0.417	0.491
NBC	**<0.001**	0.111	**<0.001**	*	0.222	0.185	0.090	0.158	0.154	0.298	0.418	0.391	0.493	0.563
CBC	**<0.001**	**0.006**	**<0.001**	**<0.001**	*	-0.026	0.235	0.129	0.252	0.408	0.494	0.483	0.545	0.631
CAB	**<0.001**	0.039	**0.003**	**<0.001**	0.611	*	0.194	0.045	0.166	0.320	0.427	0.410	0.462	0.579
SAB	**0.004**	0.095	**0.007**	0.066	0.014	0.044	*	0.186	0.153	0.342	0.450	0.437	0.512	0.645
SK	**<0.001**	0.011	**0.003**	**<0.001**	0.019	0.147	0.026	*	0.035	0.101	0.228	0.193	0.295	0.349
NON	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.003**	**0.003**	0.111	*	0.018	0.141	0.111	0.183	0.255
NQC	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.002**	0.043	0.173	*	0.047	0.000	0.153	0.117
NSNB	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.091	*	0.000	0.030	0.130
LAB	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.003**	0.388	0.379	*	0.070	0.090
NY	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.005**	**0.003**	**0.004**	0.028	0.222	0.105	*	0.327
NL	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.001**	**<0.001**	**<0.001**	**<0.001**	*

Table 2.7. Approximate divergence times among populations based on a divergence rate estimate of 1.70%/My and the corrected average pairwise differences calculated in Arlequin v3.11. Times are given in kya. Refer to Figure 2.1 for locations.

	AKA	AKF	AKW	NBC	CBC	CAB	SAB	SK	NON	NQC	NSNB	LAB	NY
AKF	13.0												
AKW	2.1	10.2											
NBC	17.4	2.4	14.8										
CBC	40.4	13.0	35.7	26.8									
CAB	31.2	8.3	26.5	21.9	0.0								
SAB	33.0	9.9	34.8	9.2	32.0	29.4							
SK	20.8	10.2	19.0	20.2	17.7	6.2	32.5						
NON	14.5	17.9	15.5	22.1	47.9	32.3	37.4	5.3					
NQC	36.1	36.6	37.4	38.8	71.3	53.8	51.8	14.4	3.9				
NSNB	66.2	71.8	67.5	72.7	108.4	88.4	93.1	36.7	23.3	5.8			
LAB	56.2	60.9	57.6	61.8	97.9	77.3	80.8	28.2	16.4	0.1	0.0		
NY	86.1	92.3	87.7	89.8	129.7	107.2	112.9	55.7	41.7	19.3	2.8	6.8	
NL	77.8	83.0	79.5	86.7	120.1	97.0	103.6	43.2	29.5	7.2	11.0	6.9	26.5

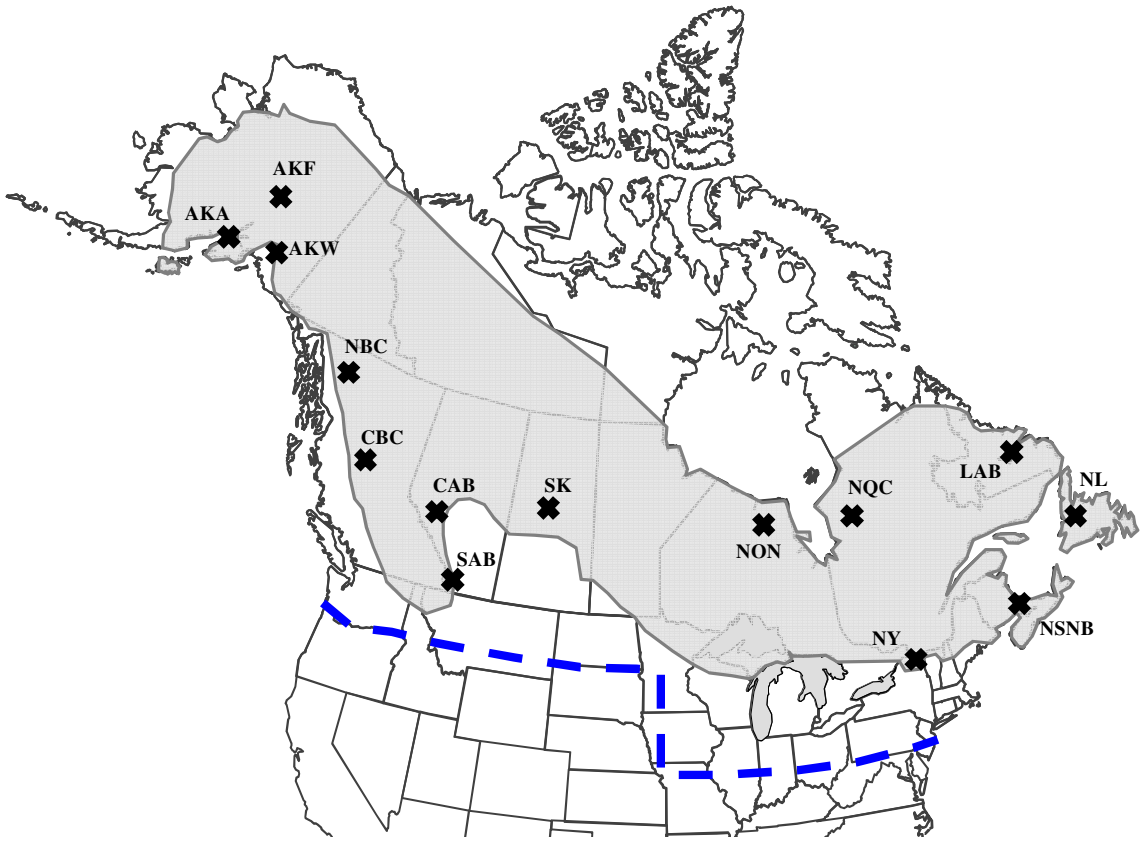

Figure 2.1. Distribution of boreal chickadees across North America (shaded). Winter irruptions may extend slightly further south in the east. Sampling locations are shown by the crosses: AKA (Alaska Anchorage), AKF (Alaska Fairbanks), AKW (Alaska Wrangell St. Elias), NBC (northern British Columbia), CBC (central British Columbia), CAB (central Alberta), SAB (southern Alberta), SK (Saskatchewan), NON (northern Ontario), NQC (northern Québec), NY (New York), NSNB (Nova Scotia & New Brunswick), LAB (Labrador), and NL (Newfoundland). The dashed line represents the cut-off for the modified ENM occurrence dataset. Figure modified from Sibley's field guide (Sibley 2003).

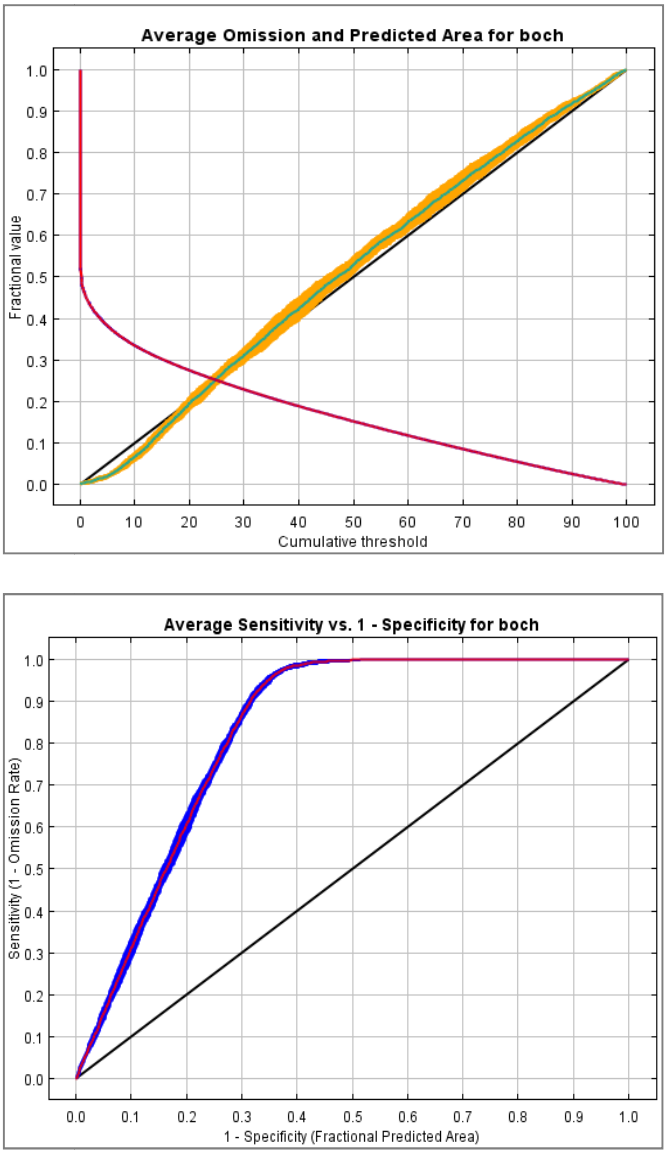

Figure 2.2. The test omission data (top) and ROC analysis (bottom) for the modified boreal chickadee spatial modelling in MAXENT v3.3.3. The test omission (green line), ± one standard deviation (yellow) closely follows the predicted data (black line, mostly covered). The ROC is shown by the red line, with the blue representing one standard deviation.

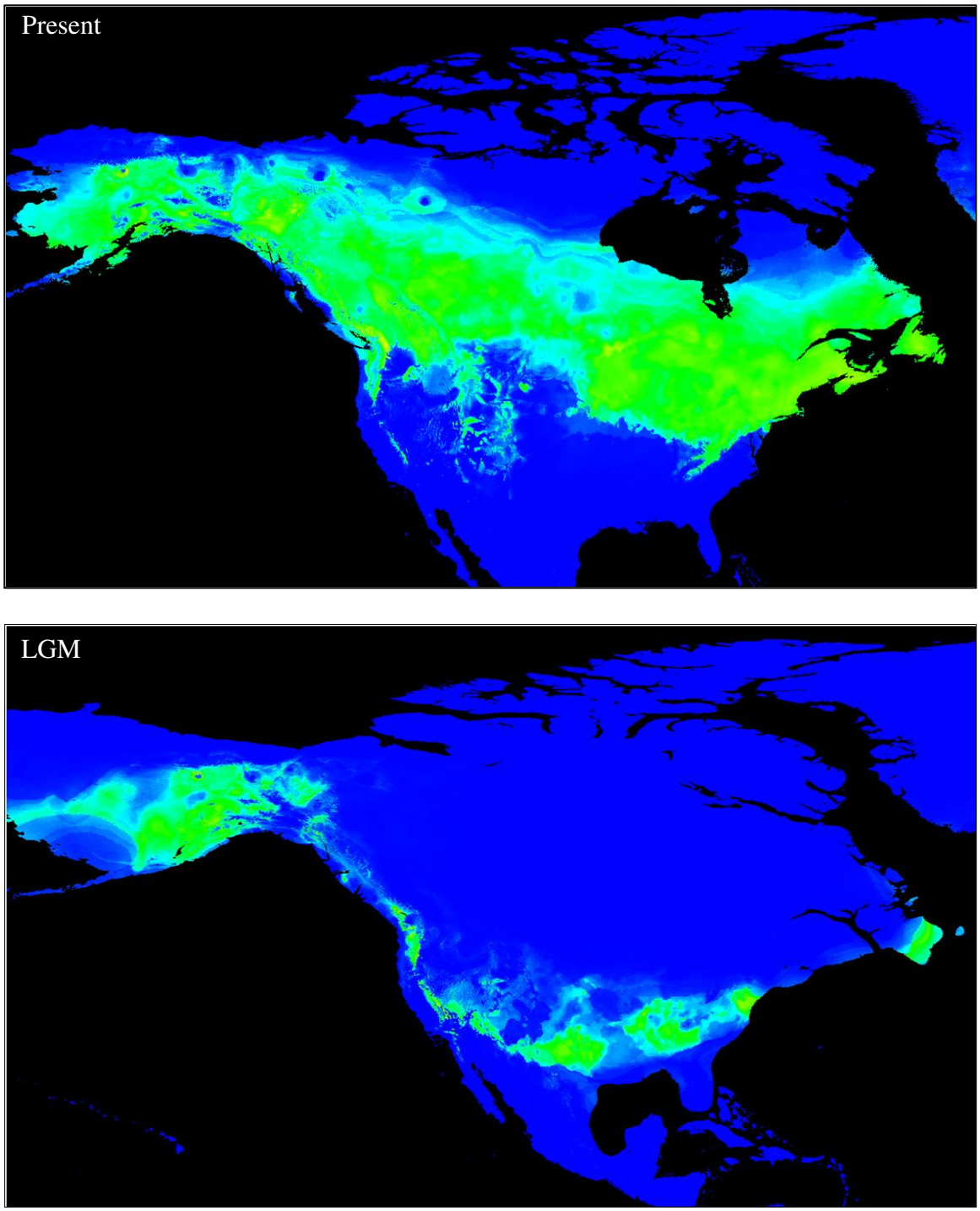

Figure 2.3. The predicted distribution of the boreal chickadee using maximum entropy in MAXENT v3.3.3 for the present (top) and LGM (bottom). The warmer colours (yellow and green) represent the most probable locations. Standard deviation was low for both timeframes.

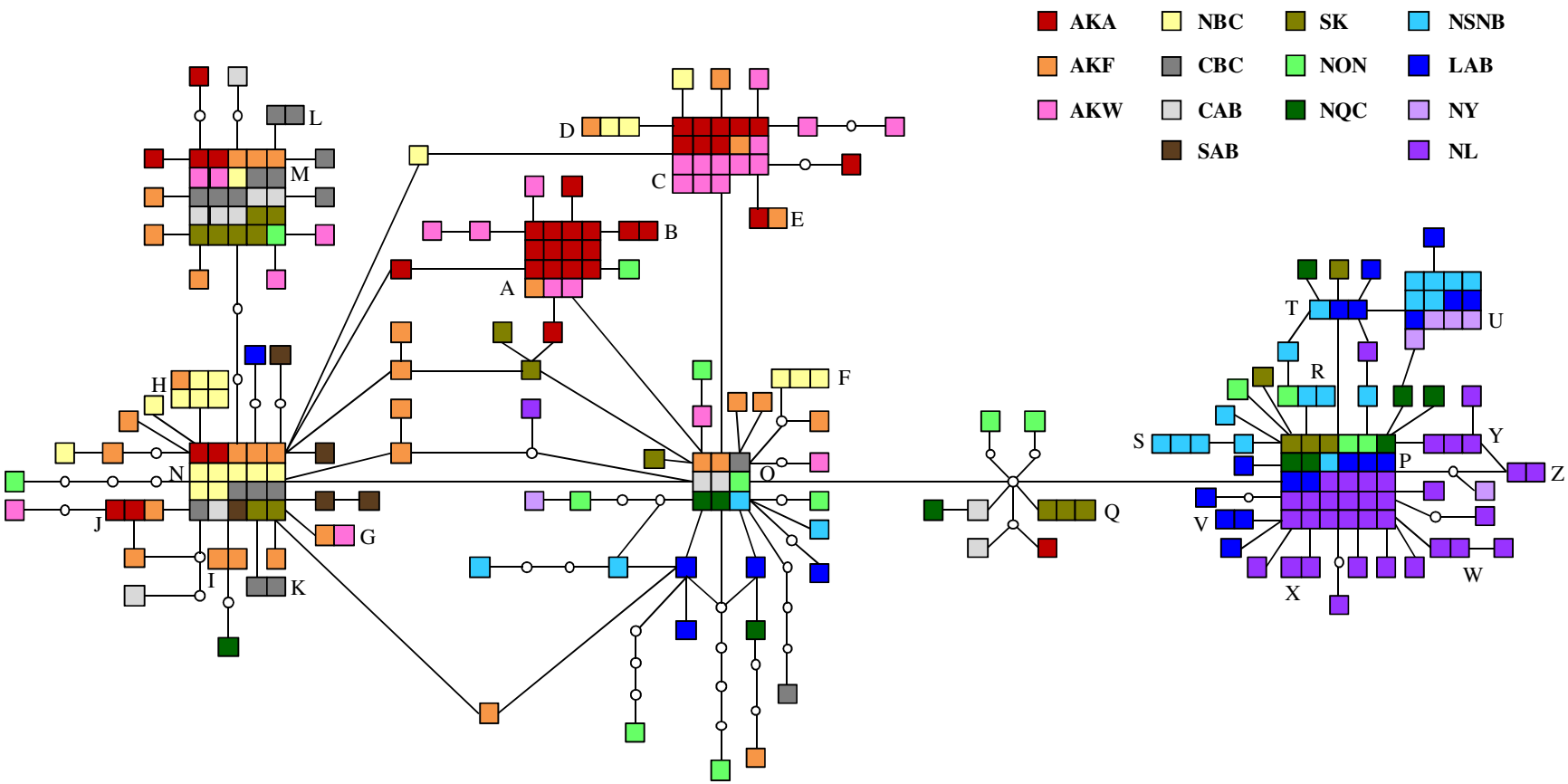

Figure 2.4. Statistical parsimony network of the concatenated mtDNA sequences. Each square represents an individual, the groups of squares (A to Z) represent shared haplotypes, and each individual is colour-coded by population. Each connection is one nucleotide change, and the circles represent inferred haplotypes. Refer to Figure 2.1 for locations.

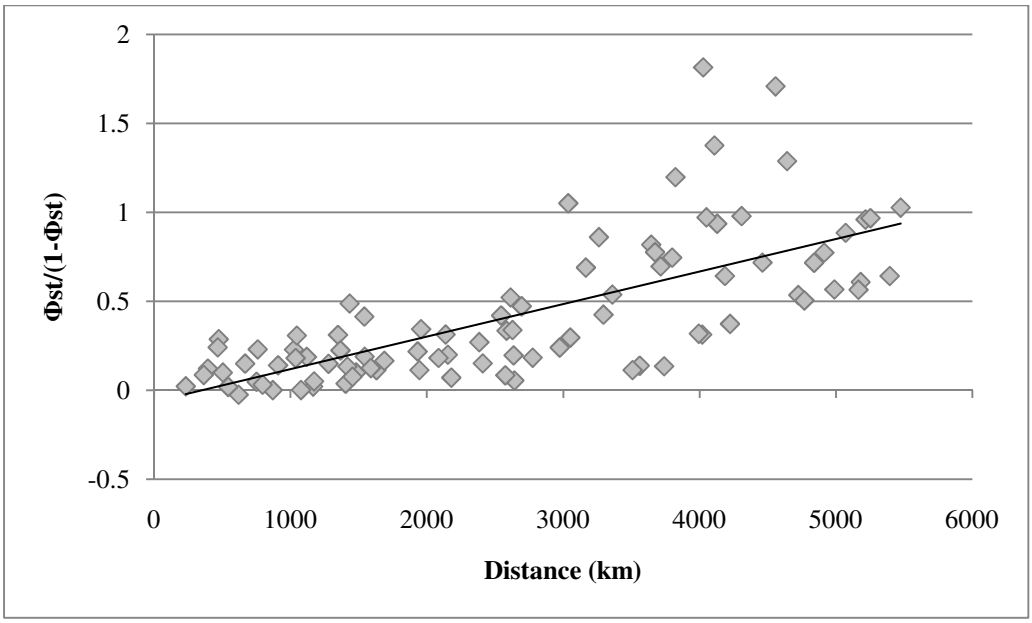

Figure 2.5. A Mantel's test showing the relationship between genetic (Φ_{ST}/ (1-Φ_{ST})) and geographic (km) data. The positive correlation indicates isolation-by-distance ($p < 0.001$).

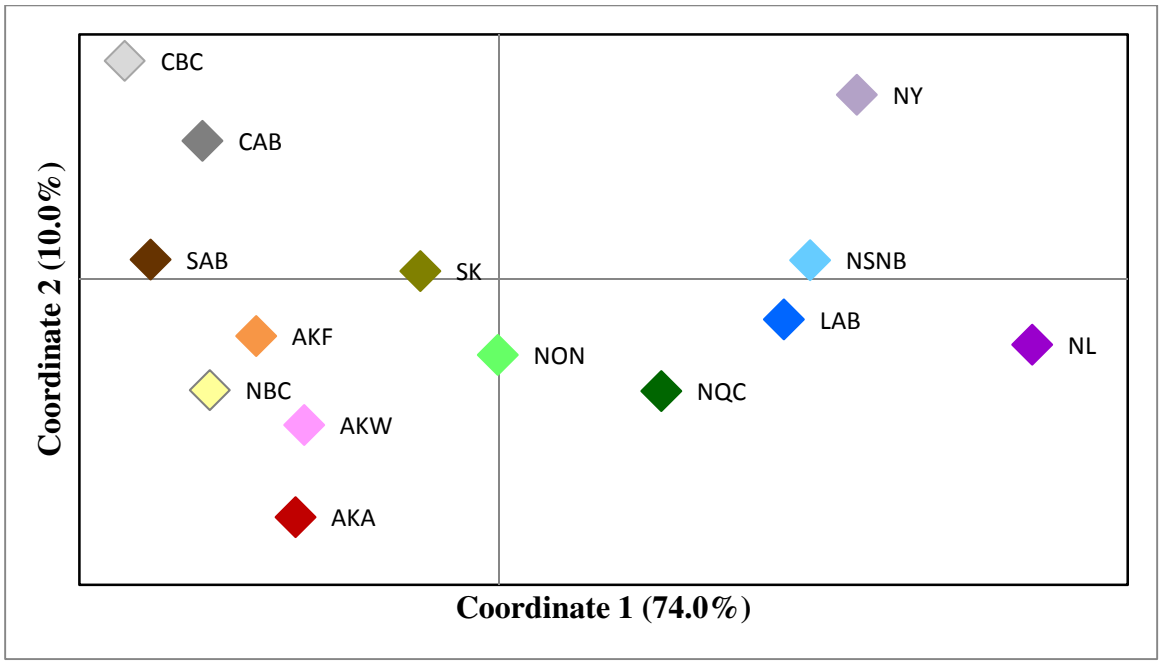

Figure 2.6. Principal coordinates analysis (PCO) based on the population pairwise Φ_{ST} values. Populations are colour-coded as per Figure 2.4. Refer to Figure 2.1 for locations.

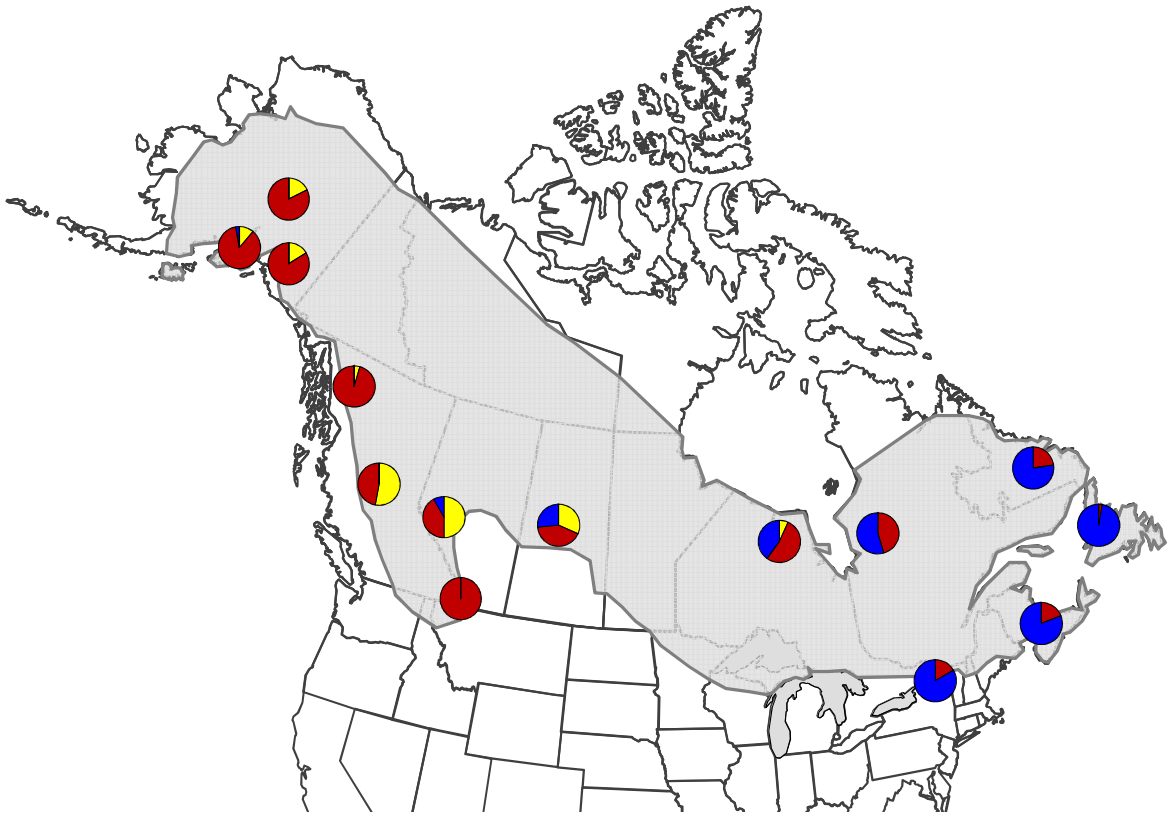

Figure 2.7. The geographic distribution of the three clusters found in BAPS v5.2: eastern (blue), western (yellow) and primarily western (red) groups. The current distribution of the boreal chickadee is shown in grey.

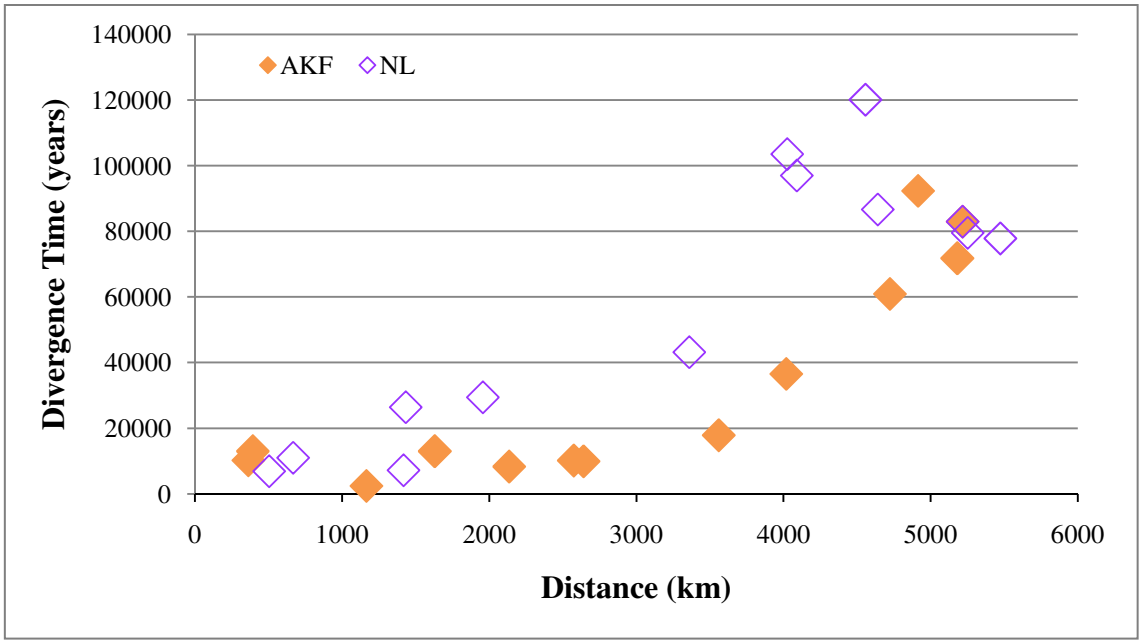

Figure 2.8. Comparison of geographic distance (km) and divergence time (years) from two putative refugia: AKF (orange, filled) and NL (purple, open). In both cases the correlation was significant ($p < 0.001$).

CHAPTER THREE

Population Structure and Microsatellite Analysis of a Resident Passerine

L.A. Lait[1] and T.M. Burg[1]

[1]Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, T1K 3M4

3.1 Introduction

Population genetic structure is a result of both historical and contemporary processes. Past events such as tectonic movements, volcanic eruptions and the Pleistocene glaciations have all contributed to the genetic makeup of species (Benzie 1999; Hewitt 2004a; Vandergast *et al.* 2004; Holland & Cowie 2007; Coykendall *et al.* 2011). Features such as physical barriers and anthropogenic interference also play a role in species' diversity and gene flow (Janzen 1967; Baur & Baur 1990; Trizio *et al.* 2005). Understanding the population genetic structure of a species informs potential decisions on management and development, and the identification of a distinct population or subspecies may prove crucial in conservation planning.

The last glacial maximum occurred approximately 21 kya to 18 kya (Pielou 1991). At this time much of the northern hemisphere was covered by large ice sheets, forcing native fauna and flora into multiple glacial refugia (Taberlet *et al.* 1998; Hewitt 2004a). These impenetrable barriers are known to have prevented gene flow in all but a few species, and played a major role in shaping today's population genetic structure (Stewart & Lister 2001).

Existing physical barriers may also prevent or reduce dispersal in both plants and animals. Large areas such as deserts, mountain ranges, or bodies of water often act as barriers to gene flow. The Rocky Mountains have previously been shown to effectively prevent gene flow in both resident (Crease *et al.* 1997; Rueness *et al.* 2003; Burg *et al.* 2005) and migratory (Milot *et al.* 2000; Lovette *et al.* 2004; Peters *et al.* 2005) species. Large rivers such as the Mississippi can prevent dispersal in terrestrial species (Phillips *et al.* 1972; Taulman & Robbins 1996; Gill *et al.* 1999; Howes *et al.* 2006; Alexander Pyron

& Burbrink 2009), while land masses (e.g. the Florida Peninsula) can act as a barrier for aquatic and coastal species (Avise 1992). Ecological barriers, such as established populations or prey availability, may also affect the level of gene flow between populations and movement of individuals into new habitat (Braillet *et al.* 2002; Musiani *et al.* 2007).

The boreal chickadee (*Poecile hudsonicus*) is a small, boreal songbird found throughout Canada and the northern United States (Figure 3.1; Ficken *et al.* 1996). It is a resident species, with little movement occurring between years. The current distribution includes a number of potential physical barriers: the Rocky Mountains, a number of smaller mountain ranges in Alaska, and both the Cabot Strait (between Newfoundland and Nova Scotia) and the Strait of Belle Isle (between Newfoundland and Québec/Labrador) in the east. The Cabot Strait is approximately 95 km wide at its narrowest point (Dawson 1915) with very inhospitable conditions (e.g. few intermediate islands to act as stopping places, high winds and cold temperatures). On the north-western coast of NL, the Strait of Belle Isle is only 15 km at the narrowest point (Dawson 1915); however, suitable habitat is not immediately available on the Québec/Labrador side (personal observation). The boreal chickadee range also encompasses a number of historical barriers (e.g. the Cordilleran and Laurentide ice sheets covered more than 80% of the current distribution).

With the advent of highly variable molecular markers, as well as advances in laboratory techniques, large multi-locus studies are becoming more feasible. Microsatellites are highly variable markers which have been shown to be useful for answering questions about contemporary processes in intraspecific studies (Costello *et al.*

2003; Flanders *et al.* 2009; Chiucchi & Gibbs 2010). These markers are short tandem repeats found throughout the nuclear genome in animals (Jarne & Lagoda 1996). In contrast to organellar DNA (mitochondrial and chloroplast DNA), nuclear markers are biparentally inherited, thus both the male and female movements contribute to patterns of genetic variation (Avise 2004).

While a number of studies have examined the genetic structure of boreal forest trees (Walter & Epperson 2001; Gamache *et al.* 2003; Jaramillo-Correa *et al.* 2004; Anderson *et al.* 2006), few have looked at their avian inhabitants, especially at the continental scale. This study aims to use hypervariable microsatellite markers to assess the population genetic structure of the boreal chickadee across its range. Given the limited structure found in North American chickadee species using mtDNA (Gill *et al.* 1993), we predict that the population structure will be shallow with a separation of eastern and western populations. The impact of physical barriers on this population structure will also be tested. Both mountain ranges and large bodies of water will likely act to prevent dispersal and gene flow in this small songbird.

3.2 Materials and Methods

3.2.1 Sample collection and DNA extraction

Samples were collected from 11 sampling locations during the 2007-2010 breeding seasons (Figure 3.1). Sampling sites were restricted to 50 km areas where possible, with no obvious dispersal barriers. All populations are derived from a single sampling site, with the exception of NSNB and NL where multiple nearby locations were combined to increase sample size after confirming genetic similarity. Mist nets and call playbacks

were used to capture birds. The birds were banded and a small blood sample was collected from the brachial vein. Samples were stored in 95% ethanol and upon return to the laboratory at -80°C. Sample sizes were supplemented and three additional populations were added using tissue samples provided by nine museums (Appendix 1). Samples were taken from either side of physical barriers (i.e. the Rocky Mountains, Cabot Strait, and Strait of Belle Isle) so that direct comparisons could be made. A modified chelex procedure was used to extract DNA from blood, feather and tissue samples (Walsh *et al.* 1991). Once extracted, the DNA was stored in 1X low TE buffer at -20°C.

3.2.2 DNA amplification and genotyping

Four individuals from distant populations (two AKA, two NL) were screened using 20 microsatellite primer pairs developed in a number of other passerines. Eight loci were successfully amplified and were polymorphic (Table 3.1). All forward primers were labelled with an M13 tag (5' – CAC GAC GTT GTA AAA CGA C – 3') which allowed the integration of a fluorescently labelled primer (700 nm or 800 nm) directly into the PCR product.

A 10 µL reaction mixture contained 1X Crimson buffer (New England Biolabs), 1-2 mM $MgCl_2$ (New England Biolabs), 0.2 mM dNTP (Fisher Scientific), 1 µM of the forward and reverse primers (Table 3.1), 0.05 µM of the M13 fluorescent primer (Eurofins MWG Operon) and 0.5 U Crimson Taq DNA polymerase (New England Biolabs). The PCR used a two-step annealing process (Table 3.1). The program was: one cycle of 94°C for 2 min, T_1 for 45 s, and 72°C for 1 min; seven cycles of 1 min at 94°C,

30 s at T_1, and 45 s at 72°C; 31 cycles of 30 s at 94°C, 30 s at T_2, and 45 s at 72°C; and a final elongation at 72°C for 5 min. In four of the loci (Escu6, Pat14, Ppi2 and Titgata39), the second step was reduced from 31 cycles to 25 cycles.

The PCR products were mixed with LI-COR stop solution (95% formamide, 20 mM EDTA and bromophenol blue), denatured for 3 min at 94°C, and run on a 6% polyacrylamide gel using a LI-COR 4300 DNA Analyzer (LI-COR Inc., Lincoln, NE). Alleles were scored by visual inspection, and genotypes were independently confirmed by a second person. Where possible, both a 700 nm and 800 nm labelled PCR of the same locus were run on the same gel to aid in scoring. Four controls of known allele sizes (both a size standard and pre-screened individuals) were included on each gel to ensure consistent scoring between gels.

3.2.3 Data analyses

Individuals with two or more missing loci (25%) were excluded from analyses. The exception to this were the museum samples from NON, NWQC and NY where samples with up to four (50%) missing loci were included in order to maintain reasonable sample sizes. This level of missing data may affect results, and has been accounted for where possible. The SAB population was not genotyped due to small sample size and late arrival of samples.

MICRO-CHECKER v2.2.3 was used to detect input errors, allelic dropout, slippage stutter or null alleles (Van Oosterhout *et al.* 2004). The presence of null alleles has been shown to strongly influence both Hardy-Weinberg equilibrium and F-statistics (Dakin & Avise 2004), while having less of an effect on Bayesian clustering analysis (Orsini *et al.*

2008). Exact tests were run in GENEPOP v4.0.10 (Raymond & Rousset 1995; Rousset 2008) to check for linkage disequilibrium and deviations from Hardy-Weinberg equilibrium. Tests were run using modified Markov chain parameters (1,000 batches, 10,000 iterations, and 10,000 dememorisation steps). BOTTLENECK v1.2.02 (Cornuet & Luikart 1996) was used to test for a recent reduction in effective population size. When a population undergoes a bottleneck, a heterozygosity excess will occur for a short period of time at all loci as allelic diversity decreases more quickly than heterozygosity through the loss of rare alleles (Cornuet & Luikart 1996). The presence of a recent bottleneck was tested under both the stepwise mutation model (SMM) and the infinite alleles model (IAM), the two extreme mutation models for microsatellites. The ratio of expected to observed heterozygosity was tested using both a standard sign test and a Wilcoxon sign-rank test. If no bottleneck has occurred, there is an equal probability of heterozygote excess or deficiency. The sign and Wilcoxon sign-rank tests calculate whether there were significantly more loci with an excess than expected given the data (Cornuet & Luikart 1996; Luikart *et al.* 1998; Piry *et al.* 1999). The Wilcoxon sign-rank test is the most powerful test for a small number of loci (≤ 20), but requires at least four loci to detect significance (Piry *et al.* 1999). All p-values were corrected for multiple tests using the modified FDR method (Benjamini & Yekutieli 2001).

Genetic diversity was compared using expected heterozygosity and allelic richness. As the number of alleles detected in the population is related to sample size, a larger sample size would be expected to have a larger number of alleles (Kalinowski 2005). Allelic richness was calculated using rarefaction, a statistical method that accounts for differences in sample size by repeatedly sampling n alleles (where n is the smallest

sample size) at random from the population (Hurlbert 1971). Allelic richness (A_R) and private allelic richness (PA_R), a measure of population distinctiveness (Kalinowski 2004), were calculated in HP-Rare v1.1 (Kalinowski 2005). Observed (H_O) and expected (H_E) heterozygosities were calculated in Microsatellite Analyser v4.05 (Dieringer & Schlötterer 2003).

3.2.4 Statistical analyses

Multilocus genotype analysis is often performed using a series of summary statistics (F-statistics). While these can be useful at measuring population differences, they rely on *a priori* population grouping which can decrease their power when detecting groups (Latch *et al.* 2006). In this study we combine traditional F-statistics with the more contemporary Bayesian clustering analyses.

3.2.4.1 Population structure

A number of summary statistics have been developed over the years, mostly derived from Wright's original fixation index (F_{ST}; Wright 1965). Weir and Cockerham (1984) modified the original F_{ST} description from biallelic data to include multiallelic markers (Meirmans & Hedrick 2011). Slatkin's R_{ST} (1995), specifically designed for use with microsatellites, and Nei's G_{ST} (1973) are both based on the ratio of within- and between-population genetic diversity (H_E), while the newer Jost's D (2008) is derived from expected number of alleles (Meirmans & Hedrick 2011). All of these methods have limitations: F_{ST} and G_{ST} are highly dependent on the within-population variation (Hedrick

1999), R_{ST} is only accurate when loci strictly follow the stepwise mutation model, and Jost's D is very slow to reach equilibrium (Meirmans & Hedrick 2011).

For this study both global and population-pairwise F_{ST} and R_{ST} values were run in Arlequin v3.11, and significance was tested using 100,000 permutations (Excoffier *et al.* 2005). While F_{ST} is more accurate with small sample sizes ($n_s \leq 10$) and few loci ($n_l < 20$), R_{ST} incorporates the potential mutation model into the calculations (Gaggiotti *et al.* 1999). Due to the level of missing data in some populations, all comparisons involving NON and NQC were run using four and five loci respectively (excluding Escu6, Ppi2 and Pdo5 in both, as well as Escu4 in NON), and all other populations were analysed using both six (excluding Ppi2 and Pdo5) and eight loci. A modified FDR correction (Benjamini & Yekutieli 2001) was applied.

3.2.4.2 Bayesian clustering analyses

Clustering analyses were done using the programs STRUCTURE v2.3 (Pritchard *et al.* 2000; Falush *et al.* 2003) and BAPS v5.2 (Corander *et al.* 2008). While both programs implement a Bayesian approach, they employ different search strategies, and thus allow independent confirmation of results. STRUCTURE applies a Markov chain Monte Carlo (McMC) model to test for likelihood (Pritchard *et al.* 2000) whereas BAPS uses a stochastic, greedy optimization algorithm (Corander *et al.* 2003). Both programs rely on deviations from Hardy-Weinberg equilibrium and linkage disequilibrium to determine the number of clusters present in the data, and will struggle when genetic population differentiation is low ($F_{ST} \leq 0.03$; Latch *et al.* 2006).

STRUCTURE v2.3 was used to delineate groupings based on individual genotypes. While the program can be run both with or without admixture (i.e. each sample can be assigned either to a single (no-admix) or multiple (admix) clusters), it is strongly recommended that the admixture model be used; simulations have shown that the admixture model is much more robust to deviations from assumptions (François & Durand 2010). The clustering analysis uses the genotype frequencies to assign individuals to a predefined number of clusters (K). STRUCTURE was run using uncorrelated allele frequencies and the admixture model, both with and without sampling locations as priors. Prior sampling information may be needed in small datasets to detect genetic structure when F_{ST} is low, and using this prior information does not find superfluous structure. The population information is discarded by the program if it does not agree with the genetic data or is not necessary (Hubisz *et al.* 2009). The program was run from K = 1 to 13 (10 runs each) for 100,000 burn-in and 300,000 post burn-in McMC steps. The results of the 10 runs were averaged using Structure Harvester v0.6.6 (Earl 2011), and the most likely value of K was calculated using two methods: the highest penalised log likelihood which is used to calculate a standard Bayes factor (Pritchard *et al.* 2000; Hubisz *et al.* 2009), and ΔK (Evanno *et al.* 2005), a method purported to provide a more accurate estimate of K as it is less reliant on the variance at high K values (Latch *et al.* 2006). All runs were repeated for 12 populations excluding NL (see Results) to test for additional substructure.

BAPS v5.2 was run first using 'clustering of individuals' and then 'clustering of groups of individuals' for K_{MAX} = 13. Rather than fixing the number of clusters (K), BAPS searches all values of K from 1 to K_{MAX} and reports K for the maximum log marginal likelihood value. The clustering of groups option allows the incorporation of population

information by defining groups, and assumes that all population members belong to the same cluster (Corander *et al.* 2003; Hubisz *et al.* 2009). The mixture results were then run using admixture analysis for 50 iterations (Corander & Marttinen 2006).

3.2.4.3 Graphical analyses

In order to summarise the population genetic structure such that the pattern can be visualised, two graphical statistical methods were used. While both methods perform similar analyses, they use different underlying statistical algorithms. A factorial correspondence analysis (FCA), a multivariate analysis used to describe variation in a set of random data, was run in GENETIX 4.05 (Belkhir *et al.* 1996-2004). The FCA uses individual genotypes to assign points based on their contribution to the overall variation among populations. Each population is represented as a cloud of points which are then averaged to provide a single population point. A principal coordinates analysis (PCO) assigns eigenvalues to summary statistics (F_{ST}) as described in Appendix 2. The PCO was performed in GenAlEx v6.3 (Peakall & Smouse 2006), and significance was tested in PCA-GEN v1.2.1 (Goudet 1999). Both of these analyses are useful methods of verifying the Bayesian clustering results (François & Durand 2010).

3.2.4.4 Dispersal barriers

The presence of geographic structure, specifically evidence of genetic or geographical barriers, was tested using three methods. To begin with, a Mantel's test was performed to compare straight-line geographic distances and linearised genetic differences to test for isolation-by-distance. The geographical distances were calculated

using the Geographic Distance Matrix Generator v1.2.3 (Ersts 2010) based on the weighted average GPS coordinates for each location. For the genetic differences, population pairwise F_{ST} values were tested.

A spatial analysis of molecular variance (SAMOVA; Dupanloup *et al.* 2002) uses Delaunay triangulation and individual genotypes to assign individuals to groups such that the F_{CT} (among-group variation) is maximised (see Appendix 2). This 'clumping' analysis tests all possible groups for a given K, starting with K−1 groups of a single individual and one group containing the rest. Individuals are then moved between neighbouring groups, and the group with the highest F_{CT} is reported (Dupanloup *et al.* 2002). The SAMOVA was run for K = 1 to 13, for 100 iterations, and F_{CT} values were calculated in Arlequin v3.11 (100,000 permutations; Excoffier *et al.* 2005).

Monmonier's algorithm is a 'splitting' analysis which uses Voronoi tessellation and Delaunay triangulation (see Appendix 2) to construct a map of sample locations, and then introduces genetic barriers based on a matrix (or matrices) of genetic distances (Manni & Guerard 2004). This algorithm, as implemented in BARRIER v2.2 (Manni *et al.* 2004), was employed to detect potential barriers using population pairwise F_{ST} values as genetic distances.

3.3 Results

3.3.1 Data analyses

After excluding individuals due to missing data, 260 individuals from 13 populations were successfully genotyped. Genotyping of Pdo5 and Ppi2 was not possible for birds from NON, NQC and many from NY, most likely due to the degraded nature of

these samples. Of the 10 NON samples, only four were genotyped at Escu4 and five at Escu6; and of the 10 NQC samples, only two were genotyped at Escu6. These four loci were removed from population-specific analyses where possible (e.g. F_{ST} and R_{ST} pairwise comparisons involving NON and NQC), and must be treated with caution where left in. All other populations had little missing data, mostly found in Pdo5. Exact tests showed departures from Hardy-Weinberg, after correction for multiple tests ($P_{crit} = 0.01$), in two loci: Pdo5 (AKA, AKF, NL; $p < 0.01$) and Ppi2 (AKA, AKW, CBC, NL, NS; $p < 0.01$). When testing for linkage, only four of 338 tests were significant ($P_{crit} = 0.008$; AKA Ppi2/Pdo5, AKW Pat14/Escu6, NBC Pat43/Pdo5, and SK Pat14/Ppi2; $p < 0.001$). As such, we felt that the loci were not in linkage disequilibrium. MICRO-CHECKER v2.2.3 suggested the presence of null alleles in at least one of Pdo5 or Ppi2 for seven of the 10 populations available (excluded NON, NQC and NY). All subsequent analyses excluded Pdo5 and Ppi2, though in some instances analyses were run with all eight loci for comparison.

Observed heterozygosity (H_O) was similar to expected heterozygosity (H_E) across all populations for five of the eight loci (Table 3.2). Pat14, Pdo5 and Ppi2 had significantly lower heterozygosity than expected (student's t-test; $p < 0.01$). For Pdo5 and Ppi2 this may be caused by the presence of null alleles. Observed heterozygosity ranged from 0.4-1.0 across all loci (where sample size ≥ 5; Table 3.2). Allelic richness was relatively similar across populations (3.01-3.37, average 3.20; Table 3.2), with a slightly higher value in NSNB. Private allelic diversity was highest in NL (0.50) and NSNB (0.51) relative to an average 0.34.

There was no evidence to support a recent bottleneck in any of the populations after correcting for multiple tests. Before corrections, all populations except for NON and NY exhibited significant differences from expected ($0.05 > p > 0.016$). This may imply either that a bottleneck occurred in the more distant past, and the populations are now returning to mutation-drift equilibrium, or that additional polymorphic loci are required to detect the signal (Piry *et al.* 1999). Results were similar between the IAM and SMM models, and between the two statistical tests ($p \geq 0.016$, $P_{crit} = 0.016$). Lower p-values were seen with the Wilcoxon sign-rank test than with the sign test, as expected given the increased power over a small number of loci.

3.3.2 Population structure

Population pairwise F_{ST} values showed significant differences between NL and all other populations (Table 3.3). NBC was significantly different from AKA, AKF and the eastern populations (NSNB, LAB, and NL). Pairwise values between other populations were generally not significant (18 of 78 comparisons were significant after modified FDR correction). The global F_{ST}, based on six loci, was 0.015 ($p < 0.001$), low relative to standard microsatellite F_{ST} values which range from 0 to about 0.1 (Meirmans & Hedrick 2011). The high expected heterozygosity will decrease the maximum possible F_{ST} and R_{ST} values (Meirmans & Hedrick 2011). The maximum theoretical F_{ST} value was 0.16 (Hedrick 1999). Comparisons of NON and NQC with the other populations, when based on six loci, showed themselves to be sensitive to missing data (i.e. when F_{ST} was calculated with six loci, NON and NQC showed significant differences from all other populations based solely on the loci with missing data). All pairwise comparisons

involving these two populations had loci with missing data removed. A similar pattern was seen when all eight loci were analysed (global $F_{ST} = 0.013$, $p < 0.001$).

When analysed separately, each locus had a slightly different impact on the overall population structure. Pat14 separated NON from all but NY and LAB; Pat43, Escu4, Escu6, Pdo5 and Ppi2 separated NL from most other populations; Pdo5 isolated NBC from all but CAB and NY; and Ppi2 separated NSNB and SK from the other populations (data not shown). Titgata02 and Titgata39 had only a few significant values, and did not show a discernible pattern.

R_{ST} values were similar to F_{ST} ($R_{ST} = 0.032$, $p < 0.001$; Table 3.4), but showed less structure (11 of 78 comparisons were significant after modified FDR correction). This is not surprising given the moderate sample sizes and number of loci, R_{ST} performs best with sample sizes > 50 and > 20 loci (Gaggiotti *et al.* 1999). NL was found to be significantly different from Alaska (AKA, AKF and AKW), British Columbia (NBC and CBC), NON, and LAB.

3.3.3 Bayesian clustering analyses

As the global F_{ST} was 0.015, and all but one pairwise comparison was below 0.05, population information was required as a prior. The initial STRUCTURE analysis, with no additional priors, detected no structure. When the population information was included, STRUCTURE identified two clusters supported by both log likelihood penalised tests (Bayes factor = 1.00, ln Pr (X | K) = -7,162) and ΔK (Figure 3.2). Individuals were assigned to a cluster based on the ancestry coefficient (Q; François & Durand 2010). The groups consisted of NL and all other populations (Figure 3.3). All Q values were > 0.50,

and 95% of individuals were assigned with $Q > 0.75$. When K was fixed at K = 3, NSNB formed a third cluster with $Q > 0.50$ (Figure 3.3). The same pattern was seen when the clustering analysis was run with eight loci, although the log likelihood penalised test supported three clusters (NL, and all other populations split into two clusters) while ΔK suggested two groups (NL and all others). There was no additional substructure seen when NL was removed from the analysis (data not shown).

BAPS was initially run using clustering of individuals; eight clusters were identified with little discernible structure (Figure 3.4). When BAPS was run with clustering of groups of individuals, which assigns all members of a single population to the same cluster, two clusters were found (NL and the rest of the populations; Figure 3.4). Similar results were seen when run with six or eight loci (data not shown).

3.3.4 Canonical analyses

Both the FCA and PCO showed similar patterns: little genetic structure among the populations with the exception of NL. In the FCA (Figure 3.5), NL and NON were separated along axis 1 (23.6% of the variation), NON along axis 2 (11.0%) and NSNB separated along axis 3 (9.9%). The separation of NON is likely due to relatively large amounts of missing data for the six loci analysed. In the PCO (Figure 3.6), the partition of NL from the rest of the populations was along coordinates 1 (42.5%) and 2 (24.8%), while coordinate 3 explained 13.0% of the variation. The broken-stick method employed in PCA-GEN v1.2.1 (Goudet 1999) found only coordinate 1 to be significant ($p < 0.05$).

3.3.5 Barriers

A weak but significant relationship was present between genetic and geographic distances ($p = 0.002$; $r^2 = 0.050$), suggesting isolation-by-distance (Figure 3.7). The SAMOVA detected the presence of two groups ($F_{CT} = 0.032$, $p = 0.011$; Figure 3.8): one group consisted of NL and the other group contained all other populations. There was a large difference in among-group variance between the different values of K ($K = 3$, $F_{CT} = 0.021$, $p = 0.011$; $K = 4$, $F_{CT} = 0.017$, $p = 0.015$). The optimum group for each K value separated NL as its own group.

Monmonier's algorithm detected one barrier separating NL from the continental populations ($F_{CT} = 0.032$, $p = 0.011$). When forced to add additional barriers, the second isolated NY ($F_{CT} = 0.031$, $p < 0.001$) and the third formed a north-western group (AKA, AKF, AKW and NBC; $F_{CT} = 0.013$, $p = 0.001$; Figure 3.8). F_{CT} values were calculated in Arlequin v3.11 (100,000 permutations; Excoffier *et al.* 2005).

3.4 Discussion

3.4.1 Genetic structure

The combination of F-statistics and Bayesian clustering analyses confirmed the presence of two genetic groups: NL and all other populations (a continental group). The presence of a distinct NL population is seen in many species: the resident moose *Alces alces* (Broders *et al.* 1999) and rock ptarmigan *Lagopus mutus* (Holder *et al.* 1999), as well as the migratory song sparrow *Melospiza melodia* (Zink & Dittmann 1993) and American redstart *Setophaga ruticilla* (Colbeck *et al.* 2008). The boreal chickadee is closely tied to a number of boreal tree species which also show evidence of distinct

eastern clades, often including the maritime provinces as well as Newfoundland (Walter & Epperson 2001; Jaramillo-Correa *et al.* 2004; Boys *et al.* 2005; Gerardi *et al.* 2010). This divergence may result from an early separation event, such as colonisation following the last glacial maximum, or isolation during the last glacial maximum. The higher private allelic richness seen in NL and NSNB (Table 3.2) refutes the idea of a postglacial founder event and multiple colonisations would be unlikely given the genetic isolation of NL (Table 3.3). The mtDNA results also support isolation in the LGM (see Chapter 2). A number of studies provide evidence for the presence of a glacial refugium in the Northeast, potentially in the area of Newfoundland (Pielou 1991; Holder *et al.* 1999; Jaramillo-Correa *et al.* 2004; Gerardi *et al.* 2010). Additional polymorphic loci may provide greater resolution in genetic structure, and may elucidate the cause of this separation of eastern populations.

3.4.2 Physical barriers

The distribution of the boreal chickadee contains a number of physical barriers which may act to prevent dispersal and gene flow. Genetic evidence reveals the role of the eastern straits as effective barriers to dispersal in boreal chickadees and other resident species (Benkman 1989; Holder *et al.* 1999). The genetic isolation of NL is likely a result of inhospitable conditions (e.g. large expanses of open water) and a break in suitable habitat. A similar pattern of variation is seen in the eastern range of the black spruce (Gerardi *et al.* 2010) and the rock ptarmigan (Holder *et al.* 1999). Other seemingly small barriers (e.g. Isthmus of Panama) also prevent gene flow in birds (Steeves *et al.* 2005).

Contrary to the pattern found in many passerines (Milot *et al.* 2000; Lovette *et al.* 2004; Burg *et al.* 2005), the Rocky Mountains are not acting as a barrier to gene flow in this species. There exists continuous habitat from Alaska to the east coast (Anderson *et al.* 2006; Gerardi *et al.* 2010), and although boreal chickadees do not tend to be found at extremely high elevations, there are potential dispersal corridors throughout the Rocky Mountains (e.g. the Crowsnest , Kicking Horse and Yellowhead passes) as well as many smaller valleys. The genetic pattern seen in the mtDNA also supports gene flow between CBC and CAB (see Table 2.6).

3.4.3 Levels of diversity

Microsatellite analysis of the boreal chickadee uncovered significant differences in population genetic structure. As would be expected from such hypervariable markers, the heterozygosity levels were high in all loci tested. This high level of within-population diversity has a direct impact on the maximum possible F_{ST} and R_{ST} values (Hedrick 1999). Compared to the average F_{ST} value of 0.1 often seen in microsatellites (Meirmans & Hedrick 2011), a global F_{ST} of 0.015 (less than 10% of the maximum theoretical F_{ST} value of 0.16), while significant ($p < 0.001$), suggests a low level of structure in this species. Of course global F_{ST} values rely not only on the level of genetic structure, but also on the size and distribution of samples as well as the mutation rate and number of markers used (Meirmans & Hedrick 2011).

3.4.4 Conclusions

The population genetic structure found in the boreal chickadee suggests isolation of the eastern populations, specifically NL. While the NL population is distinct from the mainland, the overall genetic differentiation in this species is low. The separation is likely a result of separate glacial populations during the most recent Pleistocene glaciations followed by subsequent gene flow between regions. Gene flow across the eastern straits is reduced, while the Rocky Mountains do not appear to be obstructing gene flow. The different impact of the physical barriers highlights the importance of including matrix quality as well as habitat features when looking at dispersal barriers.

Table 3.1. Microsatellite primers, PCR conditions, species of origin and source reference. Forward primers (F suffix) had a M13 sequence added to the 5' end (sequence not shown). Pat MP 2-14 and Pat MP 2-43 are referred to as Pat14 and Pat43 in the text.

Primer Name	**Sequence (5' to 3')**	**Focal species**	**$MgCl_2$ (mM)**	**T_1** (°C)	**T_2** (°C)	**Source**
Escu4F Escu4R	TTC CCT CAC AAT TTT CCG AC TAT GTG CTG AAG TGA ACC ATC C	reed bunting *Emberiza schoeniclus*	2.0	45°C	48°C	1
Escu6F Escu6R	CAT AGT GAT GCC CTG CTA GG GCA AGT GCT CCT TAA TAT TTG G	reed bunting *Emberiza schoeniclus*	1.0	50°C	52°C	1
Pat MP 2-14F Pat MP 2-14R	GAA CAG ATA AAG CCA AAT TAC TAG TGA ATG CTT GAT TTC TTT G	black-capped chickadee *Poecile atricapillus*	2.0	50°C	52°C	2
Pat MP 2-43F Pat MP 2-43R	ACA GGT AGT CAG AAA TGG AAA G GTA TCC AGA GTC TTT GCT GAT G	black-capped chickadee *Poecile atricapillus*	2.0	50°C	52°C	2
Pdo5F Pdo5R	GAT GTT GCA GTG ACC TCT CTT G GCT GTG TTA ATG CTA TGA AAA TGG	house sparrow *Passer domesticus*	1.0	45°C	48°C	3
Ppi2F Ppi2R	CAC AGA CCA TTC GAA GCA GA GCT CCG ATG GTG AAT GAA GT	black-billed magpie *Pica pica*	1.5	50°C	52°C	4
Titgata39F Titgata39R	CAT GTA TTT TCC AAA AGT AAA TAT CTG CTA TTC TGC AAA CTT GTG G	green-backed tit *Parus monticolus*	2.0	50°C	52°C	5
Titgata02F Titgata02R	ATT GCT TGA TAT TTG AAA GCA TA TTG TCT TTT GGG TTG CCT GA	green-backed tit *Parus monticolus*	2.0	50°C	52°C	5

1 (Hanotte *et al.* 1994)

2 (Otter *et al.* 1998)

3 (Griffith *et al.* 1999)

4 (Martinez *et al.* 1999)

5 (Wang *et al.* 2005)

Table 3.2. Sample size (n), observed (H_O) and expected (H_E) heterozygosities, total number of alleles (AN), allelic richness (A_R) and private allelic richness (PA_R) for 13 boreal chickadee populations. Both A_R and PA_R were estimated using rarefaction for the first six loci only.

	Pat14	**Pat43**	**Escu4**	**Escu6**	**Titgata02**	**Titgata39**	**Pdo5**	**Ppi2**	**Average**
AKA (n=35)									
H_O	0.912	0.824	0.765	0.882	0.765	0.912	0.840	0.765	**0.83**
H_E	0.932	0.850	0.770	0.919	0.766	0.875	0.953	0.938	**0.88**
AN	20	15	5	18	9	12	22	25	**15.75**
A_R	3.62	3.20	2.83	3.54	2.82	3.34			**3.23**
PA_R	0.66	0.43	0.02	0.51	0.21	0.35			**0.36**
AKF (n=34)									
H_O	0.882	1.000	0.706	0.882	0.941	0.853	0.640	0.824	**0.84**
H_E	0.926	0.877	0.752	0.939	0.819	0.839	0.945	0.930	**0.88**
AN	18	14	6	19	13	9	20	28	**15.88**
A_R	3.58	3.33	2.76	3.65	3.07	3.16			**3.26**
PA_R	0.38	0.39	0.06	0.44	0.44	0.07			**0.30**
AKW (n=24)									
H_O	0.833	0.792	0.708	0.792	0.667	0.833	0.833	0.833	**0.79**
H_E	0.912	0.849	0.762	0.950	0.801	0.878	0.951	0.948	**0.88**
AN	15	11	5	21	10	10	19	25	**14.50**
A_R	3.51	3.22	2.79	3.71	2.99	3.32			**3.26**
PA_R	0.32	0.30	0.01	0.67	0.26	0.11			**0.28**
NBC (n=22)									
H_O	0.909	0.955	0.636	1.000	0.909	0.773	0.857	0.864	**0.86**
H_E	0.926	0.851	0.697	0.936	0.808	0.833	0.883	0.904	**0.85**
AN	16	12	5	17	9	10	17	15	**12.63**
A_R	3.58	3.21	2.57	3.63	3.03	3.13			**3.19**
PA_R	0.48	0.25	0.01	0.55	0.50	0.29			**0.35**
CBC (n=16)									
H_O	0.933	0.938	0.688	0.813	0.750	0.875	0.769	0.714	**0.81**
H_E	0.933	0.815	0.732	0.921	0.821	0.806	0.945	0.937	**0.86**
AN	13	8	5	12	9	8	15	16	**10.75**
A_R	3.62	3.07	2.67	3.55	3.08	3.06			**3.17**
PA_R	0.37	0.12	0.11	0.50	0.42	0.06			**0.26**
CAB (n=12)									
H_O	0.917	0.833	0.417	0.917	0.750	0.833	0.727	0.909	**0.79**
H_E	0.946	0.877	0.699	0.953	0.855	0.862	0.935	0.961	**0.89**
AN	13	10	4	14	10	7	13	16	**10.88**
A_R	3.68	3.34	2.54	3.72	3.23	3.24			**3.29**
PA_R	0.52	0.32	0.01	0.55	0.54	0.04			**0.33**

SK (n=19)									
H_O	0.895	0.895	0.895	0.889	0.789	0.684	0.933	0.842	**0.85**
H_E	0.922	0.906	0.743	0.948	0.755	0.886	0.956	0.964	**0.89**
AN	17	15	5	15	7	10	17	24	**13.75**
A_R	3.57	3.49	2.73	3.69	2.81	3.38			**3.28**
PA_R	0.64	0.67	0.09	0.40	0.19	0.22			**0.37**
NON (n=10)									
H_O	0.778	0.800	0.750	1.000	1.000	0.900	n/a	n/a	**0.87**
H_E	0.869	0.895	0.750	0.733	0.837	0.868	n/a	n/a	**0.83**
AN	9	12	4	4	7	8	n/a	n/a	**7.33**
A_R	3.31	3.44	2.77	2.66	3.13	3.28			**3.10**
PA_R	0.67	0.72	0.00	0.28	0.36	0.08			**0.35**
NQC (n=10)									
H_O	0.900	0.900	0.900	0.000	0.700	0.800	n/a	n/a	**0.70**
H_E	0.947	0.905	0.800	0.667	0.726	0.868	n/a	n/a	**0.82**
AN	12	9	5	2	5	8	n/a	n/a	**6.83**
A_R	3.69	3.46	2.96	2.00	2.66	3.28			**3.01**
PA_R	0.57	0.54	0.03	0.20	0.02	0.22			**0.26**
NY (n=5)									
H_O	0.750	0.800	0.400	1.000	0.800	1.000	1.000	0.333	**0.76**
H_E	0.964	0.644	0.533	0.929	0.889	0.844	1.000	0.867	**0.83**
AN	7	4	2	6	6	6	6	4	**5.13**
A_R	3.79	2.46	1.92	3.57	3.37	3.20			**3.05**
PA_R	0.71	0.10	0.00	0.71	0.45	0.03			**0.34**
NSNB (n=19)									
H_O	0.789	0.947	0.789	0.895	0.895	0.789	0.857	0.667	**0.83**
H_E	0.902	0.935	0.780	0.929	0.841	0.902	0.947	0.954	**0.90**
AN	14	16	5	17	11	13	15	21	**14.00**
A_R	3.47	3.62	2.87	3.60	3.19	3.45			**3.37**
PA_R	0.48	0.79	0.04	0.62	0.72	0.40			**0.51**
LAB (n=20)									
H_O	0.950	0.800	0.650	0.850	0.700	0.800	0.900	0.950	**0.83**
H_E	0.940	0.855	0.682	0.947	0.735	0.813	0.951	0.976	**0.86**
AN	15	12	4	18	5	7	18	29	**13.50**
A_R	3.65	3.24	2.49	3.70	2.71	3.03			**3.14**
PA_R	0.36	0.45	0.00	0.63	0.02	0.03			**0.25**
NL (n=34)									
H_O	0.848	0.970	0.788	0.848	0.788	0.818	0.714	0.704	**0.81**
H_E	0.913	0.891	0.793	0.858	0.778	0.855	0.926	0.934	**0.87**
AN	18	15	6	14	9	11	15	23	**13.88**
A_R	3.51	3.41	2.94	3.24	2.87	3.24			**3.20**
PA_R	0.65	0.95	0.49	0.33	0.24	0.34			**0.50**

Table 3.3. Population pairwise F_{ST} comparisons based on 100,172 permutations. Values are based on data from six loci for 11 populations. To account for missing data, all pairwise comparisons involving NQC are based on five loci, and four for NON. F_{ST} values are given above the diagonal and p-values below. Significant values after modified FDR correction are in bold (P_{crit} = 0.010).

	AKA	AKF	AKW	NBC	CBC	CAB	SK	NON	NQC	NY	NSNB	LAB	NL
AKA	*	0.002	-0.002	0.015	0.007	0.000	0.017	0.022	-0.005	0.022	0.007	0.009	0.043
AKF	0.196	*	-0.001	0.012	-0.001	-0.005	0.003	0.008	-0.001	0.020	0.007	0.005	0.034
AKW	0.721	0.603	*	0.008	0.005	-0.006	0.007	0.011	-0.011	0.009	0.003	0.002	0.040
NBC	**0.003**	**0.006**	0.108	*	0.014	-0.002	0.010	0.021	0.019	0.038	0.016	0.018	0.046
CBC	0.101	0.524	0.269	0.032	*	-0.009	0.011	0.019	0.021	0.015	0.021	0.020	0.046
CAB	0.511	0.824	0.911	0.591	0.895	*	-0.005	-0.001	-0.006	0.015	0.007	0.006	0.039
SK	**0.002**	0.202	0.146	0.041	0.064	0.783	*	0.003	0.003	0.037	0.008	0.012	0.030
NON	0.017	0.101	0.184	0.024	0.045	0.543	0.403	*	0.007	0.002	0.016	-0.003	0.030
NQC	0.711	0.480	0.932	0.042	0.044	0.731	0.355	0.303	*	0.042	0.001	0.006	0.023
NY	0.070	0.050	0.371	0.018	0.169	0.252	0.019	0.386	0.032	*	0.017	-0.005	0.074
NSNB	0.079	0.063	0.354	**0.009**	**0.003**	0.240	0.108	0.057	0.477	0.122	*	0.005	0.025
LAB	0.045	0.149	0.463	**0.007**	0.016	0.311	0.042	0.592	0.282	0.634	0.237	*	0.049
NL	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.002**	**0.008**	**<0.001**	**<0.001**	**<0.001**	*

Table 3.4. Population pairwise R_{ST} comparisons based on 100,172 permutations. Values are based on data from six loci for 11 populations. To account for missing data, all pairwise comparisons involving NQC are based on five loci, and four for NON. R_{ST} values are above the diagonal and p-values below. Significant values after modified FDR correction are in bold (P_{crit} = 0.010).

	AKA	AKF	AKW	NBC	CBC	CAB	SK	NON	NQC	NY	NSNB	LAB	NL
AKA	*	-0.003	-0.001	-0.004	-0.008	-0.016	0.024	0.093	-0.017	-0.047	0.018	0.020	0.054
AKF	0.530	*	-0.010	0.001	0.017	-0.018	0.018	0.102	-0.009	-0.041	0.009	-0.007	0.065
AKW	0.427	0.821	*	-0.009	-0.005	-0.015	0.033	0.067	0.001	-0.078	0.018	-0.007	0.089
NBC	0.485	0.395	0.668	*	0.000	-0.027	0.018	0.074	0.010	-0.061	0.002	0.012	0.064
CBC	0.624	0.156	0.564	0.386	*	0.005	0.069	0.122	-0.010	-0.043	0.064	0.051	0.115
CAB	0.787	0.868	0.752	0.969	0.368	*	-0.013	0.075	-0.022	-0.051	-0.016	-0.005	0.030
SK	0.048	0.106	0.046	0.108	**0.006**	0.616	*	0.076	-0.024	-0.020	-0.005	0.018	0.009
NON	**0.006**	**0.008**	0.057	0.047	0.013	0.049	0.029	*	0.174	-0.015	0.039	0.078	0.145
NQC	0.673	0.568	0.455	0.268	0.521	0.712	0.762	**0.007**	*	0.037	0.031	0.044	0.067
NY	0.889	0.865	0.991	0.925	0.756	0.840	0.491	0.459	0.135	*	-0.011	-0.051	0.038
NSNB	0.076	0.217	0.121	0.314	0.013	0.724	0.450	0.085	0.096	0.437	*	0.009	0.015
LAB	0.080	0.678	0.617	0.191	0.040	0.525	0.143	0.023	0.068	0.867	0.223	*	0.083
NL	**0.001**	**0.001**	**<0.001**	**0.003**	**<0.001**	0.090	0.201	**0.002**	0.021	0.163	0.131	**0.001**	*

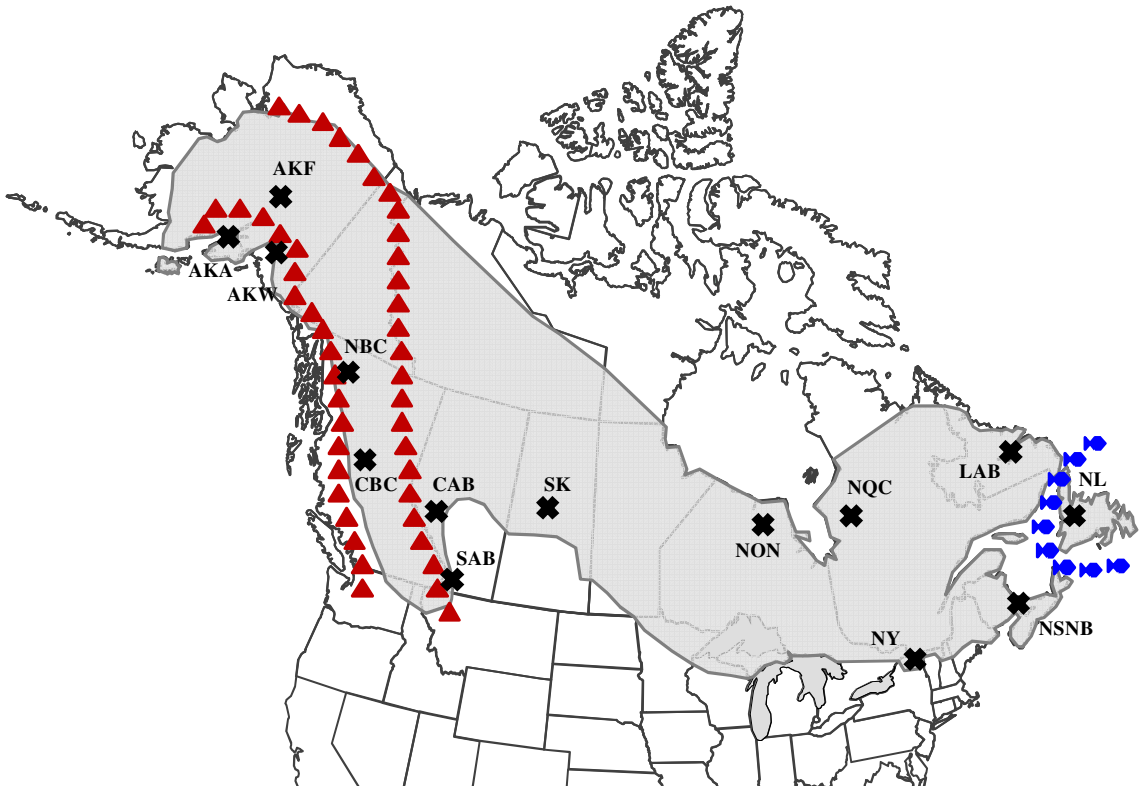

Figure 3.1. The distribution of the boreal chickadee across North America (shaded) modified from Sibley's field guide (Sibley 2003). Sampling sites are denoted by black crosses, with location names as in Figure 2.1. Physical barriers are shown in red (mountains) and blue (bodies of water).

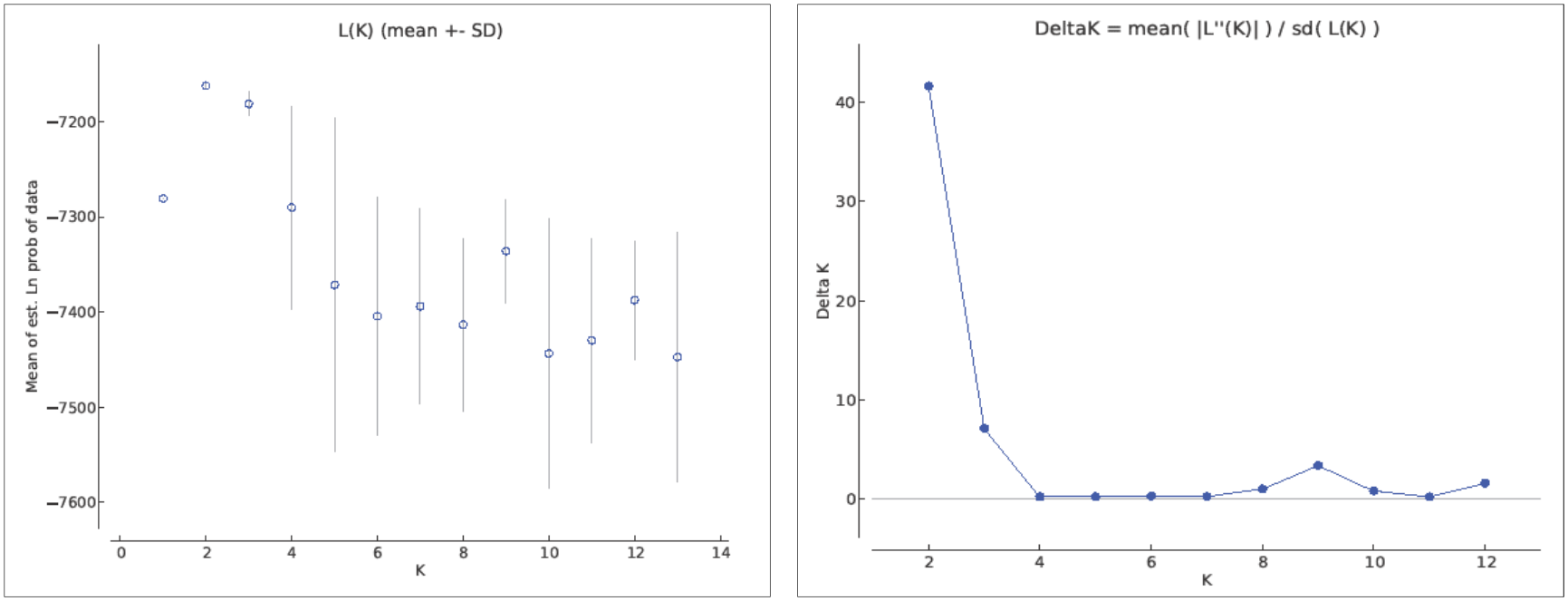

Figure 3.2. The penalised log likelihood test (left) and ΔK (right) depicting the two clusters found in STRUCTURE v2.3. The penalised log likelihood test takes the maximum ln Pr (X | K) as the correct number of clusters. ΔK infers the number of clusters from the difference between the different ln Pr (X | K).

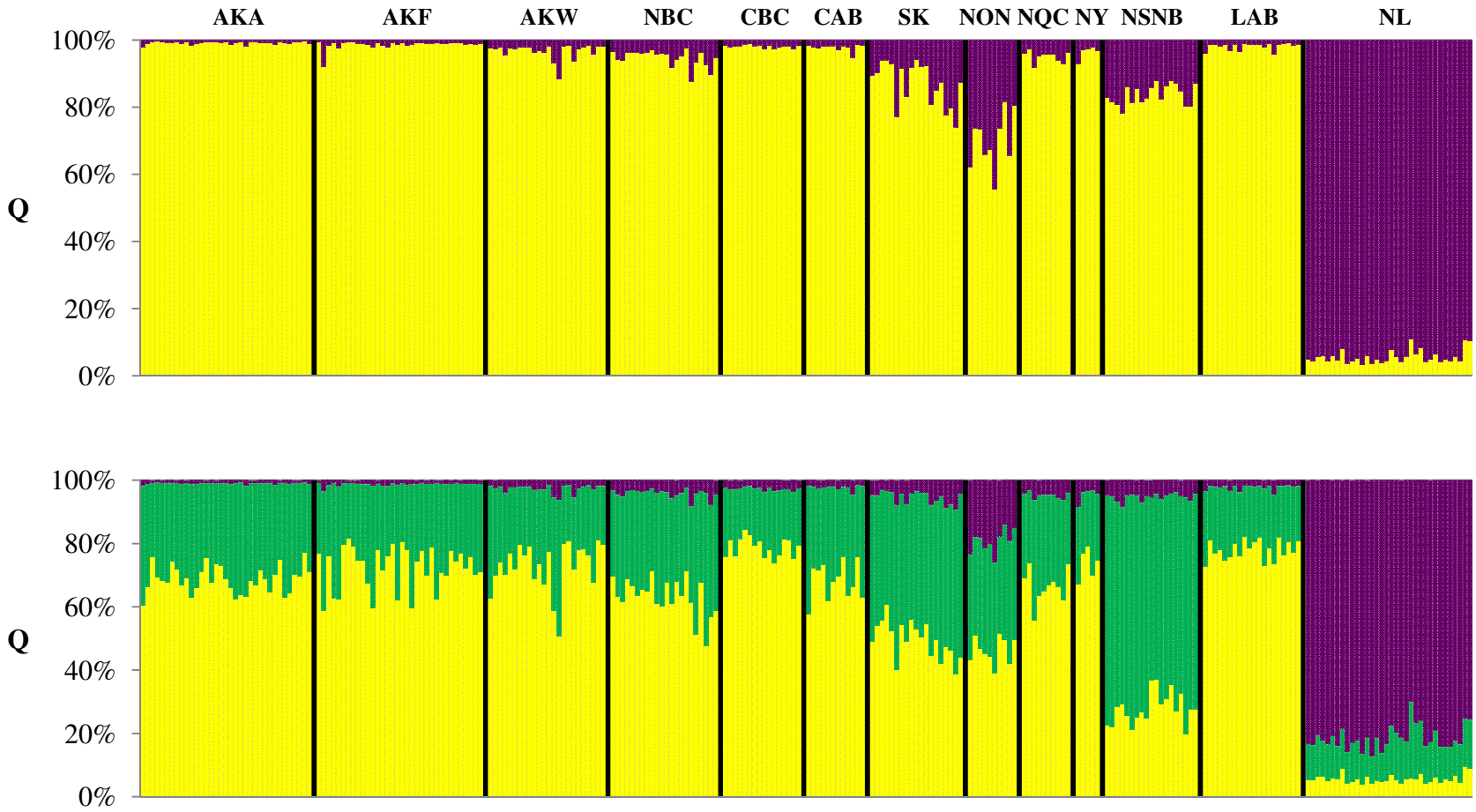

Figure 3.3.Bayesian clustering analysis run in STRUCTURE v2.3 with six microsatellite loci for K = 2 (top) and K = 3 (bottom). Each vertical line represents an individual, and the y-axis denotes the cluster membership (Q). Colours represent the different clusters.

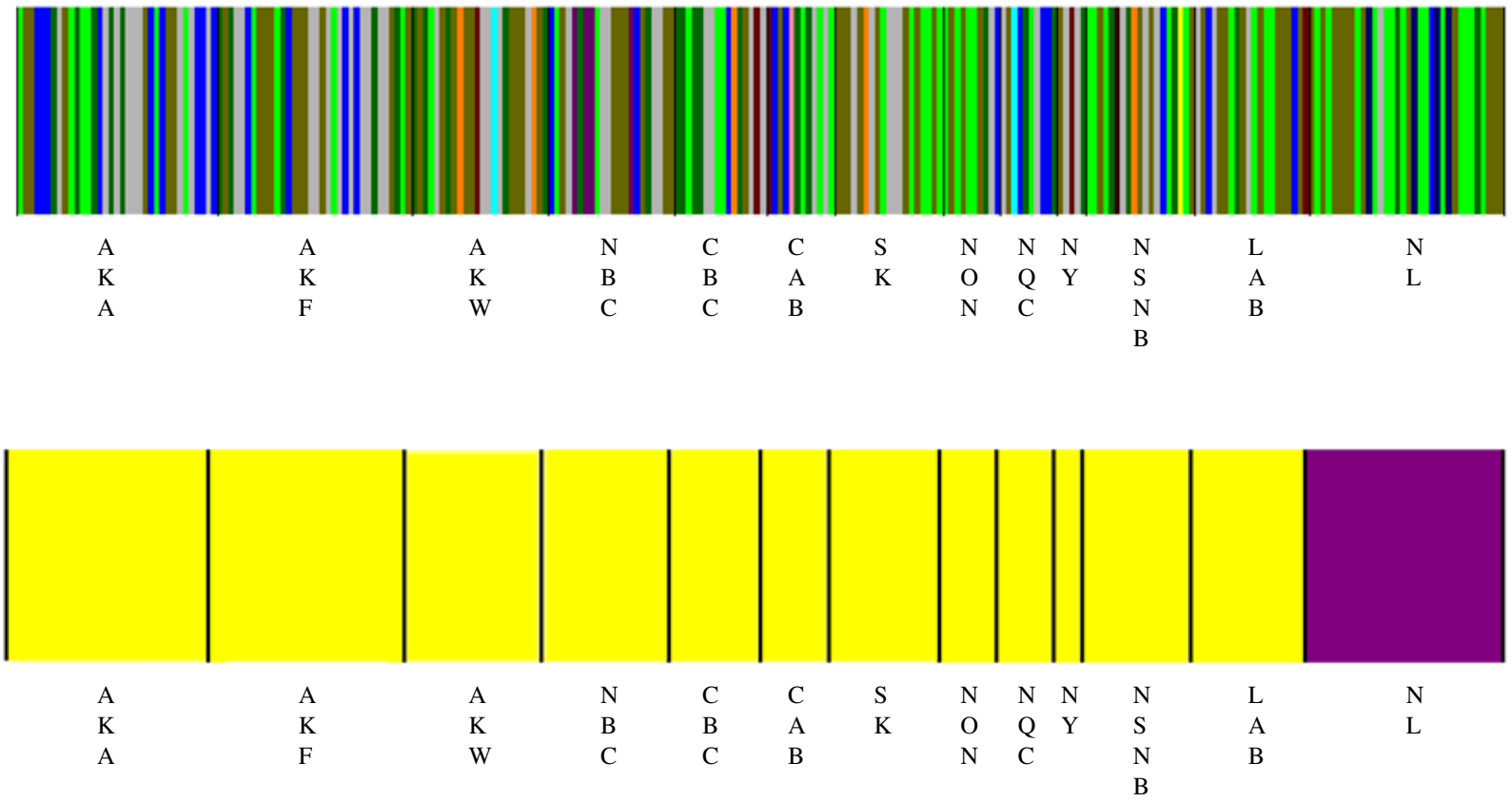

Figure 3.4. Bayesian clustering analysis run in BAPS v5.2 on six microsatellite loci. Runs consisted of individuals without (top, K = 13) and with (bottom, K = 2) prior population information. The location prior was needed due to the small F_{ST} values seen in the boreal chickadee ($F_{ST} < 0.05$), and forced individuals from the same population to be assigned to the same cluster.

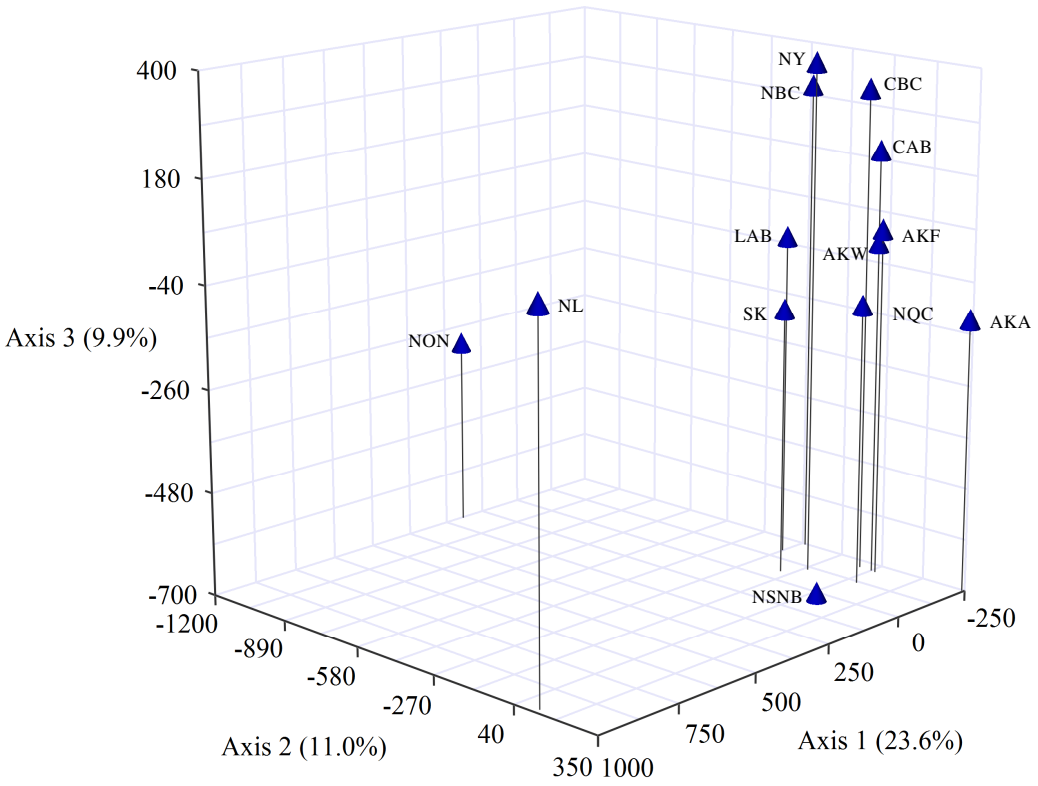

Figure 3.5. Factorial correspondence analysis (FCA) based on six microsatellite loci across 13 boreal chickadee populations. Axis 1 explained 23.6% of the variation, axis 2 11.0% and axis 3 9.9%.

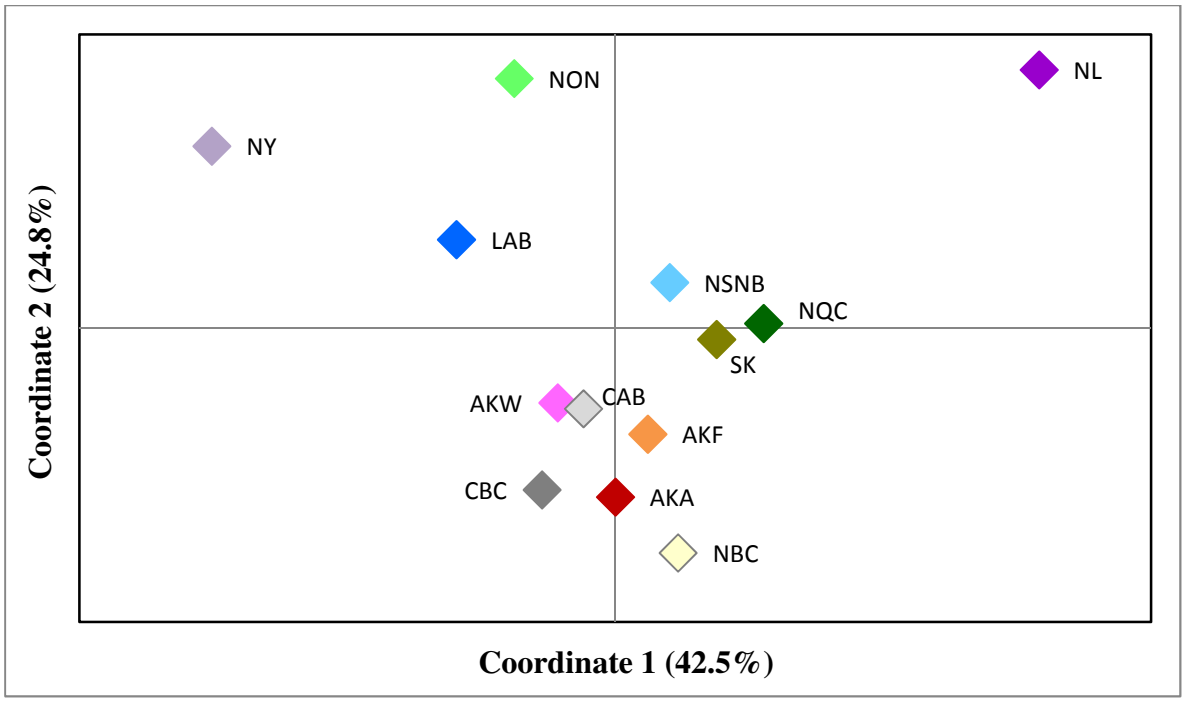

Figure 3.6. Principal coordinates analysis (PCO) showing the genetic structure of the boreal chickadee run on population pairwise F_{ST} values. Coordinate 1 explained 42.5% of the variation ($p < 0.05$), coordinate 2 24.8% (n.s.) and 3 13.0% (n.s.).

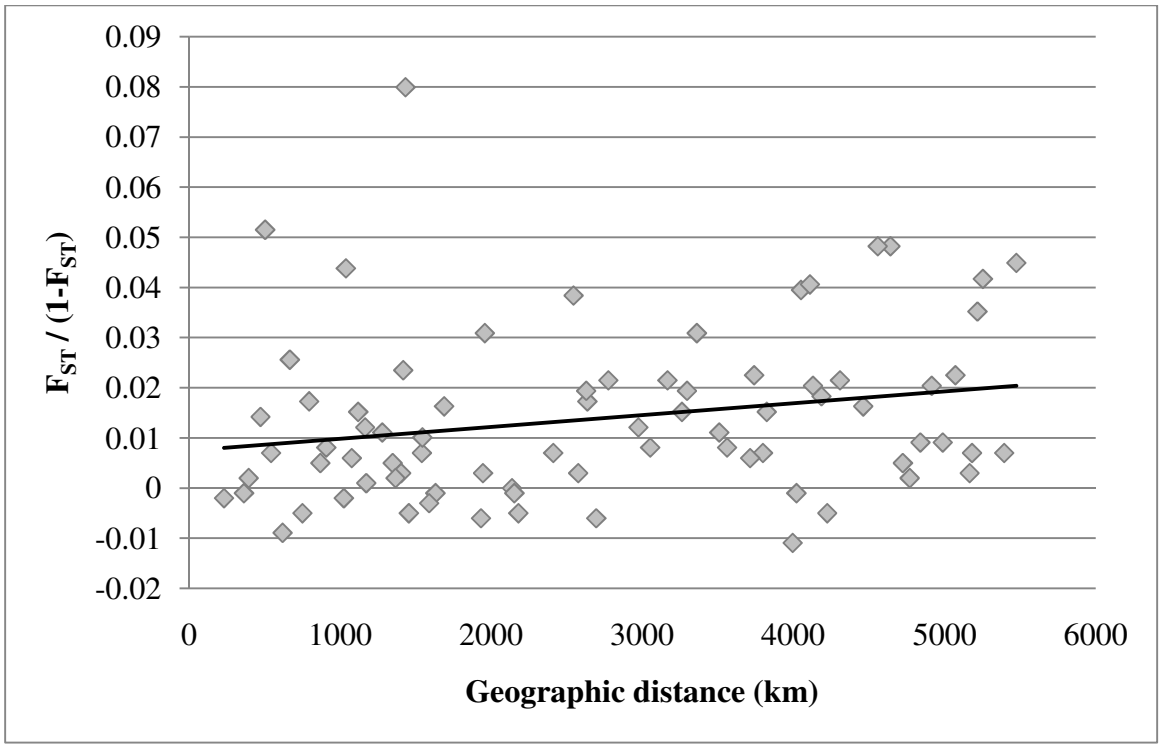

Figure 3.7. A weak but significant relationship between genetic and geographic distances as calculated in a Mantel's test ($r^2 = 0.050$, $p = 0.002$).

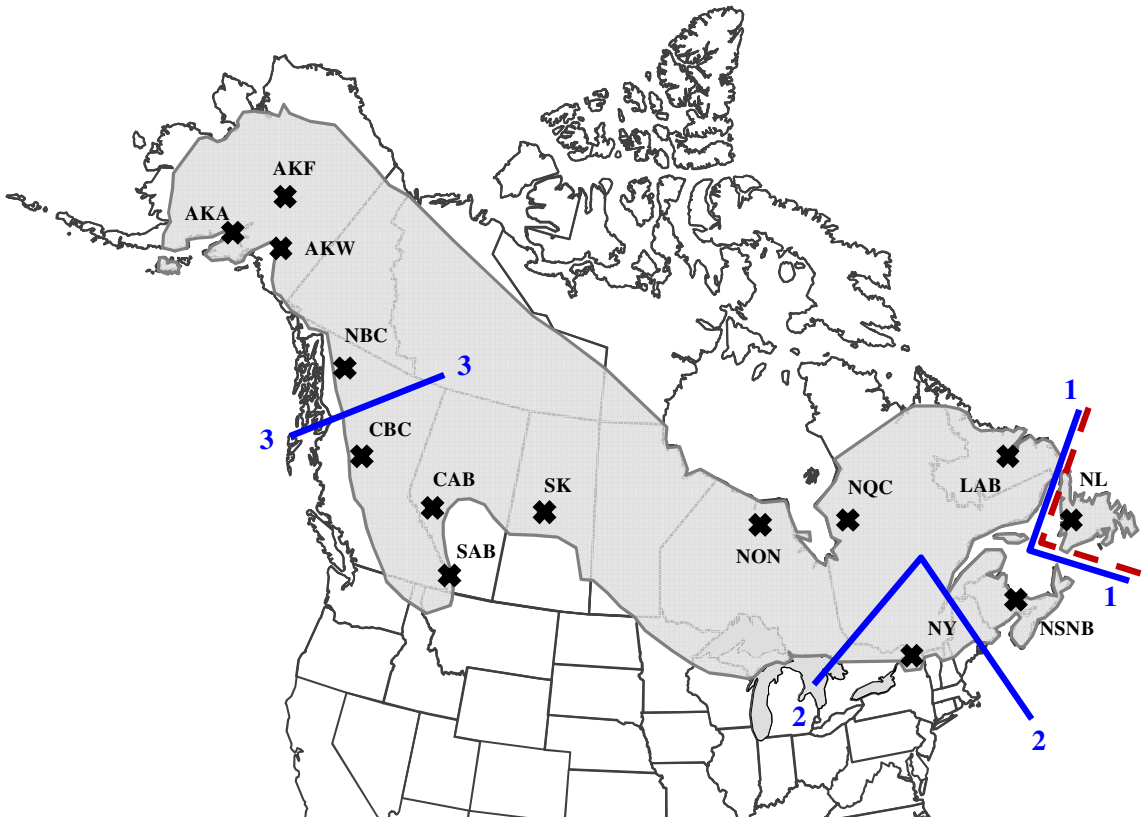

Figure 3.8. The genetic barriers determined by Monmonier's algorithm (solid, blue), and the groups found with SAMOVA (dashed, red), superimposed upon the distribution map of the boreal chickadee (grey). The numbers denote the order in which the barriers were identified. The crosses represent sampling locations (see Figure 2.1 for location names).

CHAPTER FOUR – GENERAL DISCUSSION

4.1 Population Genetic Structure

Both mtDNA fragments (CR and ATP) supported two main groups: an eastern group consisting of Québec, Atlantic Canada, and New York, and a western group found from Alaska to Ontario. Estimated divergence times between eastern and western populations (56.2-129.7 kya) place the separation during the last glaciation, suggesting that boreal chickadee populations were isolated by large ice sheets in multiple refugia. A gradient is evident in the central populations, indicative of the merging of individuals from genetically distinct populations. Similarly, boreal tree species, which provide habitat for the boreal chickadee, commonly show evidence of a phylogenetic break between Ontario and Québec, south of the Hudson Bay (Jaramillo-Correa *et al.* 2009). The congregation of contact zones in this region is likely a result of the manner in which the ice sheets receded. The Hudson Bay area was one of the last regions to melt (Pielou 1991), and as such was one of the last areas colonised. The presence of a conspecific population as colonisation occurred would then prevent exponential expansion (Hewitt 2000). For trees, this may be the fact that dispersing into an established area is difficult as seedlings do not grow well in shade (Johansen & Latta 2003).

A number of studies on North American species have detected a general east/west split, likely resulting from glacial vicariance. The jack pine *Pinus banksiana* shows evidence of a western group and multiple eastern groups (Godbout *et al.* 2005), as do the black spruce *Picea mariana* (Gerardi *et al.* 2010) and white spruce *Picea glauca* (de Lafontaine *et al.* 2010). The pattern is congruent with several bird (Milot *et al.* 2000; Bull

et al. 2010; Manthey *et al.* 2011) and mammal (Arbogast 2007; Aubry *et al.* 2009) species.

The mtDNA further divides the western boreal chickadee group, with isolated populations north and east of the Wrangell and Chugach Mountains and south of the Alaskan Mountain Range (AKA and AKW). Bayesian clustering analysis suggests the presence of two groups in the west. This more complicated pattern is a common feature in the Pacific Northwest, where numerous barriers and refugia have together played a role in structuring populations (see Soltis *et al.* 1997; Brubaker *et al.* 2005; Gavin 2009). Western coniferous trees such as the lodgepole pine *Pinus contorta* (Godbout *et al.* 2008) and Douglas fir *Pseudotsuga menziesii* (Gugger *et al.* 2010) show evidence of multiple western lineages. The mountain goat *Oreamnos americanus* (Shafer *et al.* 2010a), American pika *Ochotona princeps* (Galbreath *et al.* 2010), and fox sparrow *Passerella iliaca* (Zink 2008) show similarly complex histories.

Uncovering different patterns with mtDNA and microsatellite data is not uncommon in phylogeographic and population genetic studies. The different pattern seen in the boreal chickadee mtDNA and microsatellite markers could be a result of the difference in N_E between the markers, or due to male-biased gene flow. The different effective population sizes could explain the larger number of significant pairwise Φ_{ST} values with the mtDNA than F_{ST} values with microsatellites. The NL/mainland break could be an older separation, with the general east/west pattern only beginning to emerge. The FCA and PCO show very slight association among eastern populations and western populations (Figures 3.5 and 3.6). Alternatively, if gene flow is higher in males, either through increased dispersal or greater breeding success in a new population, this could

explain the disappearance of an east/west split in the eastern mainland populations for nuclear markers, while dispersal to NL is prevented by a physical barrier. In the mtDNA, pairwise Φ_{ST} values are significant, suggesting very low levels of female gene flow, while nuclear F_{ST} comparisons indicate much higher movement. Although male-biased dispersal is believed to be uncommon in birds (Greenwood & Harvey 1982), it has been revealed in several corvid species (Williams & Rabenold 2005; Li & Merilä 2010), as well as the great bustard *Otis tarda* in Europe (Alonso & Alonso 1992), the red-billed quelea *Quelea quelea* in Africa (Dallimer *et al.* 2002), and the sandhill crane *Grus canadensis* (Jones *et al.* 2005), yellow warbler *Dendroica petechia* (Gibbs et al. 2000), and common eider *Somateria mollissima* (Sonsthagen *et al.* 2009) in North America. Band recoveries from boreal chickadees for both males and females showed limited dispersal with a maximum distance of 37 km between captures (Canadian Bird Banding Office).

The population genetic structure seen in the boreal chickadee is shallow, with slight but distinctive differences between eastern and western populations. A number of studies on *Parus* spp. have also identified low population genetic structure (Kvist *et al.* 1998; Uimaniemi *et al.* 2003). It has been noted that genetic differentiation in birds tends to be low relative to fish, reptiles, amphibians, and mammals (Avise & Aquadro 1982; Kessler & Avise 1985; Johns & Avise 1998), with many species lacking reciprocal monophyly between otherwise distinct groups, as is seen with the boreal chickadee. A number of alternative explanations have been proposed: the avian constraint hypothesis suggests that birds may be less tolerant of non-synonymous mutations than other vertebrates (Stanley & Harrison 1999); the Hill-Robertson theory states that regions of DNA with low

recombination, such as the W and Y chromosomes and mtDNA, will show lower genetic diversity due to smaller N_E (Berlin *et al.* 2007), and that linkage may occur between mtDNA and the W chromosome, further reducing genetic diversity; and it has been suggested that birds produce less reactive oxygen species (ROS), resulting in lower rates of DNA damage despite their faster metabolic rates (Hickey 2008). All of these factors may contribute to the lower variation seen in avian species.

4.2 Postglacial Expansion

4.2.1 Glacial refugia

The genetic pattern seen in the mtDNA supports the use of multiple glacial refugia in the boreal chickadee. It has been suggested that widespread species, especially generalists, are more likely to have survived in multiple isolated refugia (Bhagwat & Willis 2008; Shafer *et al.* 2010b). Evidence for multiple glacial refugia has been found in a number of songbirds (Ball & Avise 1992; Spellman & Klicka 2007; Colbeck *et al.* 2008), while others suggest a single refugium (Ball *et al.* 1988; Ball & Avise 1992; Milá *et al.* 2006). Diversity levels are as expected if two (or more) refugia merged following glacial isolation, with high diversity in or near glacial refugia (AKF and NL) and evidence of secondary mixing in the centre (i.e. NON and NQC; Taberlet *et al.* 1998; Petit *et al.* 2003). Spatial modelling supports the presence of multiple putative refugia: Beringia, the western coast of North America, the southern United States, and Newfoundland.

In contrast with other species, the boreal chickadee appears to have used Beringia as one of its main refugium. Both nuclear and mitochondrial DNA diversity are high in

AKF. While the majority of species thought to have persisted in Beringia are either restricted to the Northwest or Arctic regions (Holder *et al.* 1999; Fedorov & Stenseth 2002; Fleming & Cook 2002; Pruett & Winker 2005), evidence supports the use of this refugium by the wolverine *Gulo gulo* (Chappell *et al.* 2004) and red fox *Vulpes vulpes* (Aubry *et al.* 2009), two widespread temperate mammals, as well as the black spruce and white spruce, two boreal trees upon which the boreal chickadee relies heavily (Jaramillo-Correa *et al.* 2004; Anderson *et al.* 2006; Gerardi *et al.* 2010). Many, if not all, birds that exist outside of the extreme north-west persisted mainly in a southern refugium (Scribner *et al.* 2003; Jones *et al.* 2005; Saitoh *et al.* 2010).

A second glacial refugium existed east of the ice sheets, likely in north-eastern North America. The persistence of the boreal chickadee in an Atlantic shelf refugium is concordant with patterns seen in a number of plants (Walter & Epperson 2001; Schauffler & Jacobson 2002; Jaramillo-Correa *et al.* 2004; de Lafontaine *et al.* 2010; Godbout *et al.* 2010) and birds (Gill *et al.* 1993; Zink & Dittmann 1993; Zink *et al.* 2003; Colbeck *et al.* 2008). Both mtDNA and microsatellite analyses support the separation of Newfoundland as a distinct population. Given the lower nucleotide and haplotype diversities in this population, it is unlikely that the island alone acted as a source population. However, it may be that the boreal chickadee persisted on the Atlantic shelf or the north eastern edge of the United States, on the periphery of the ice sheet, and as the ice sheets melted and the island became isolated a number of individuals were separated on the island of Newfoundland (Pielou 1991). This could explain the high diversity in NSNB and LAB, the high private allelic diversity in NSNB and NL, and the high number of private haplotypes in NL.

Additional refugia may have been utilised in the west – either south of the ice sheets, in the Rocky Mountains, or on the Alaska coast. The significant mtDNA differences seen between coastal (AKA, AKW) and central (AKF) Alaska populations may suggest an isolated refugium, evident in a number of Beringian species (Eddingsaas *et al.* 2004; Galbreath *et al.* 2010), or an early dispersal event. As the ice sheets receded, the areas around the coast melted first, providing suitable habitat for early migrants (Pielou 1991; Harris 1996). Subsequent gene flow may be reduced, either through a behavioural barrier or the coastal Alaska Mountains. MtDNA divergence estimates (10.2-13.0 kya) support an early dispersal event towards the end of the LGM.

4.2.2 Colonisation patterns

Colonisation patterns are best inferred from non-recombinant markers such as mtDNA. These markers retain historical patterns longer due to uniparental inheritance. The strong, significant isolation-by-distance patterns, non-significant differences between adjacent populations, and fairly wide central cline (extending from SK to NQC), are indicative of a stepping-stone colonisation model. If a gradual expansion had occurred, a much narrower cline would likely be evident in the secondary contact zone, as seen in the ponderosa pine *Pinus ponderosa* (Johansen & Latta 2003). Stepping-stone colonisation is commonly found in island models (Harbaugh *et al.* 2009), and has been seen in fish (Pogson *et al.* 2001), birds (Reeves *et al.* 2008), and mammals (Wisely *et al.* 2004).

4.3 Physical Barriers

Physical barriers such as mountain ranges and rivers are known to shape population structure in both plants and animals. In some cases, the change in habitat between the mountain ranges can act as the barrier to be crossed (e.g. drier conditions found between the Rocky and Cascade Mountains due to rain shadow; DeChaine & Martin 2004). The boreal chickadee is closely tied to the boreal forests of Canada and the northern United States, particularly to spruce and fir trees (Gill *et al.* 1993; Ficken *et al.* 1996). It is the presence or absence of this habitat that determines the connectivity in the boreal chickadee. Neither the mtDNA nor the microsatellite data show any evidence that the western mountain ranges are preventing gene flow in the boreal chickadee; in fact, the CBC and CAB populations showed no genetic differentiation (Tables 2.6, 3.3 and 3.4). This is likely due to the prevalence of treed dispersal corridors through the multitude of valleys and passes. These high mountains have been shown to act as a barrier in several songbirds (Milot *et al.* 2000; Burg *et al.* 2005), while not affecting gene flow in others (Ball & Avise 1992; Zink *et al.* 2003; Colbeck *et al.* 2008). It is likely a difference in habitat requirements, as well as the ability to adapt to high altitude, that determines whether a species can cross this geographical barrier.

The Cabot Strait and the Strait of Belle Isle which surround Newfoundland are acting as barriers to gene flow in the boreal chickadee. Both mtDNA and microsatellites support the separation of NL as a distinct population. As with many island ecosystems, Newfoundland hosts a number of endemic species and subspecies (Cronin *et al.* 2005; Hearn *et al.* 2006), as well as genetically distinct populations (Zink & Dittmann 1993; Broders *et al.* 1999; Holder *et al.* 1999; Zink *et al.* 2003; Colbeck *et al.* 2008). A

combination of geographic distance and inhospitable conditions likely contributes to the maintained genetic divergence seen in this species.

4.4 Subspecies

In birds, subspecies are often arbitrary divisions which are not supported by genetic patterns (Ball & Avise 1992; Haavie *et al.* 2000). Descriptions tend to be based on either morphological differences, which can vary from significant (e.g. dark-eyed junco; Milá *et al.* 2007) to subtle (e.g. downy woodpecker; Ball & Avise 1992), or purely on geographical distribution. A general explanation for morphological differences is that local adaptation and phenotypic plasticity can happen rapidly, in a few generations, with genetic concordance taking hundreds of generations to catch up (Zink 2004). At the other extreme, some cryptic species show no morphological differences, are not separated into subspecies, and yet have well-supported genetic differences (Bickford *et al.* 2007; Manthey *et al.* 2011).

The mtDNA data provide limited support for the five subspecies designations (American Ornithologists' Union 1957), although geographic areas do not coincide exactly. The two eastern subspecies, *P. hudsonicus littoralis* and *P. hudsonicus rabbittsi*, have the most support. There is some evidence for the widespread *P. hudsonicus hudsonicus*; similarities exist with the mtDNA western group, although the genetic data combine NBC and AKF. *P. hudsonicus columbianus* could be associated with the slight differences we saw in CBC and CAB, but again the distributions do not concur. No samples were collected from the putative range of *P. hudsonicus cascadensis*, found in southern British Columbia and northern Washington.

4.5 Future Work

Further sampling from areas in the extreme south of the boreal chickadee's range would be beneficial, both to clarify if there was a southern glacial refugium, and to test the fifth subspecies, *P. hudsonicus cascadensis*. Current sampling was not able to reject the hypothesis of a southern refugium in the west, nor could it comment on the existence of a Cascade Mountains subspecies.

Additional microsatellite markers may increase the resolution provided by the nuclear markers, perhaps finding more comparable structure to the mtDNA. The nuclear markers are expected to have lower structure given the larger effective population size, unless the mutation rate is sufficiently high to provide higher resolution. Microsatellites commonly have sufficiently high mutations rates; however, the markers used were all developed in other species, and variability is species-specific (Pleines *et al.* 2008). The development of microsatellite primers is costly and time consuming, but would be beneficial both in answering phylogeographic questions, and also for looking at potential hybridisation. Hybridisation with other chickadee species is believed to occur, but does not seem to be frequent. One of the birds we caught appeared (morphologically) to be a hybrid with a black-capped chickadee, and was caught with a group of black-capped chickadees in an area where we had not seen boreal chickadees; however, the mtDNA revealed a boreal chickadee mother and the microsatellites were inconclusive. This implies that it may have been an F2 or later hybrid, or that the morphology was deceiving. More variable markers may help to decipher the history of this and other possible hybrids. Alternatively, amplified fragment length polymorphisms (AFLPs) or nuclear coding genes could be employed.

There are no published data on natal dispersal in the boreal chickadee. Learning more about how this species disperses once it leaves the nest could be invaluable, both to clarifying population genetic structure, and also to help understand important dispersal corridors and habitat. It was observed that once fledged the young are not generally found in the parents' territory (McLaren 1975), but where they go is unknown. There are several methods which could be invoked to study this question. Radio telemetry has often been employed in studying natal dispersal (Shirer & Downhower 1968; Koenig *et al.* 2000; Forsman *et al.* 2002; Wiens *et al.* 2006), as well as foraging (Ostrand *et al.* 1998) and bird-mediated seed dispersal (Pons & Pauses 2007). Radio tracking is a cost-effective and useful method for tracking an animal's dispersal, but it is limited by long-distance dispersal (Koenig *et al.* 2000). An alternative method is the more recently-publicised satellite telemetry. Satellite tracking has been used to study breeding and post-breeding dispersal in birds (Steenhof *et al.* 2005; Weimerskirch *et al.* 2006), to establish habitat use in marine animals (Hatch *et al.* 2000; Elwen *et al.* 2006), and to detect migration routes and wintering grounds in songbirds, raptors, and waterfowl (Martell *et al.* 2001; Kenow *et al.* 2002; Stutchbury *et al.* 2009). In both cases, technological advances in battery life are allowing birds to be tracked over longer periods of time. This would be required for tracking juveniles (e.g. over the course of a year).

4.6 General Conclusions

In conclusion, the population genetic structure of the boreal chickadee reveals isolation in multiple glacial refugia – one in Beringia and one on the north-east coast of North America, either in or near the Atlantic shelf refugium. Both the mtDNA and

paleogeographic modelling support this pattern. Colonisation of unglaciated habitat generally followed a stepping-stone model, likely tracking the spread of spruce species. A secondary contact zone is present south of Hudson Bay between Ontario and Québec, similar to that seen in many boreal tree species. The Rocky Mountains did not contribute to the current geographical pattern, likely due to prevalence of dispersal corridors and relatively continuous habitat; however, Newfoundland remains an isolated island population. The differences seen between the two markers suggest male-biased dispersal in the boreal chickadee. All of these patterns are congruent with black spruce and white spruce, signifying niche conservatism in this songbird. The close association between these species suggests that they may be useful bioindicators for each other.

REFERENCES

Adams JM, Faure H (1997) Preliminary vegetation maps of the world since the last glacial maximum: An aid to archaeological understanding. *Journal of Archaeological Science* **24**, 623-647.

Alexander Pyron R, Burbrink FT (2009) Lineage diversification in a widespread species: Roles for niche divergence and conservatism in the common kingsnake, *Lampropeltis getula*. *Molecular Ecology* **18**, 3443-3457.

Alonso JC, Alonso JA (1992) Male-biased dispersal in the great bustard *Otis tarda*. *Ornis Scandinavica* **23**, 81-88.

American Ornithologists' Union (1957) *Check-list of North American Birds*, 5th edn. American Ornithologists' Union, Washington, D.C.

American Ornithologists' Union (1998) *Check-list of North American Birds*, 7th edn. American Ornithologists' Union, Washington, D.C.

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12447-12450.

Arbogast BS (2007) A brief history of the New World flying squirrels: Phylogeny, biogeography, and conservation genetics. *Journal of Mammalogy* **88**, 840-849.

Aubry KB, Statham MJ, Sacks BN, Perrine JD, Wisely SM (2009) Phylogeography of the North American red fox: Vicariance in Pleistocene forest refugia. *Molecular Ecology* **18**, 2668-2686.

Austin JD, Lougheed SC, Neidrauer L, Check AA, Boag PT (2002) Cryptic lineages in a small frog: The post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Molecular Phylogenetics and Evolution* **25**, 316-329.

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna - A case-history with lessons for conservation biology. *Oikos* **63**, 62-76.

Avise JC (2000) *Phylogeography: The History and Formation of Species* Harvard University Press, Cambridge, MA.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer & Associates, Sunderland, MA.

Avise JC, Aquadro CF (1982) A comparative summary of genetic distances in the vertebrates. *Evolutionary Biology* **15**, 151-185.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography - The mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Bain JF, Golden JL (2005) Chloroplast haplotype diversity patterns in *Packera pauciflora* (Asteraceae) are affected by geographical isolation, hybridization, and breeding system. *Canadian Journal of Botany* **83**, 1039-1045.

Baker A, Marshall D (1997) *Avian Molecular Evolution and Systematics* Academic Press, San Diego, CA.

Ball RM, Avise JC (1992) Mitochondrial-DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *The Auk* **109**, 626-636.

Ball RM, Freeman S, James FC, Bermingham E, Avise JC (1988) Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 1558-1562.

Barendregt RW, Irving E (1998) Changes in the extent of North American ice sheets during the late Cenozoic. *Canadian Journal of Earth Sciences* **35**, 504-509.

Barrowclough GF, Groth JG, Mertz LA, Gutiérrez RJ (2004) Phylogeographic structure, gene flow and species status in blue grouse (*Dendragapus obscurus*). *Molecular Ecology* **13**, 1911-1922.

Baur A, Baur B (1990) Are roads barriers to dispersal in the land snail *Arianta arbustorum*? *Canadian Journal of Zoology* **68**, 613-617.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations* Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165-1188.

Benkman CW (1989) On the evolution and ecology of island populations of crossbills. *Evolution* **43**, 1324-1330.

Bennett KD (1990) Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology* **16**, 11-21.

Bensch S, Harlid A (2000) Mitochondrial genomic rearrangements in songbirds. *Molecular Biology and Evolution* **17**, 107-113.

Benzie JAH (1999) Genetic structure of coral reef organisms: Ghosts of dispersal past. *American Zoologist* **39**, 131-145.

Berlin S, Tomaras D, Charlesworth B (2007) Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome. *Heredity* **99**, 389 - 396.

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in southeastern United States. *Genetics* **113**, 939-965.

Bhagwat SA, Willis KJ (2008) Species persistence in northerly glacial refugia of Europe: A matter of chance or biogeographical traits? *Journal of Biogeography* **35**, 464-482.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**, 148-155.

Bolgiano NC (2004) Cause & effect: Changes in boreal bird irruptions in eastern North America relative to the 1970s spruce budworm infestation. *American Birds* **104**, 26-33.

Boulet M, Gibbs HL (2006) Lineage origin and expansion of a Neotropical migrant songbird after recent glaciation events. *Molecular Ecology* **15**, 2505-2525.

Boys J, Cherry M, Dayanandan S (2005) Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of Botany* **92**, 833-841.

Braillet C, Charmantier A, Archaux F, Dos Santos A, Perret P, Lambrechts MM (2002) Two blue tit *Parus caeruleus* populations from Corsica differ in social dominance. *Journal of Avian Biology* **33**, 446-450.

Brandley MC, Guiher TJ, Pyron RA, Winne CT, Burbrink FT (2010) Does dispersal across an aquatic geographic barrier obscure phylogeographic structure in the diamond-backed watersnake (*Nerodia rhombifer*)? *Molecular Phylogenetics and Evolution* **57**, 552-560.

Brant SV, Orti G (2003) Phylogeography of the northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): Past fragmentation and postglacial recolonization. *Molecular Ecology* **12**, 1435-1449.

Brito P, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Broders HG, Mahoney SP, Montevecchi WA, Davidson WS (1999) Population genetic structure and the effect of founder events on the genetic variability of moose, *Alces alces*, in Canada. *Molecular Ecology* **8**, 1309-1315.

Brown WM, George MJ, Wilson AC (1979) Rapid evolution of mitochondrial DNA. *Proceedings of the National Academy of Sciences* **76**, 1967-1971.

Brubaker LB, Anderson PM, Edwards ME, Lozhkin AV (2005) Beringia as a glacial refugium for boreal trees and shrubs: New perspectives from mapped pollen data. *Journal of Biogeography* **32**, 833-848.

Brunsfeld SJ, Miller TR, Carstens BC (2007) Insights into the biogeography of the Pacific Northwest of North America: Evidence from the phylogeography of *Salix melanopsis*. *Systematic Biology* **32**, 129-139.

Bull RD, McCracken A, Gaston AJ, Birt TP, Friesen VL (2010) Evidence of recent population differentiation in orange-crowned warblers (*Vermivora celata*) in Haida Gwaii. *The Auk* **127**, 23-34.

Burbrink FT (2002) Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses. *Molecular Phylogenetics and Evolution* **25**, 465-476.

Burg TM, Gaston AJ, Winker K, Friesen VL (2005) Rapid divergence and postglacial colonization in western North American Steller's jays (*Cyanocitta stelleri*). *Molecular Ecology* **14**, 3745-3755.

Burg TM, Gaston AJ, Winker K, Friesen VL (2006) Effects of Pleistocene glaciations on population structure of North American chestnut-backed chickadees. *Molecular Ecology* **15**, 2409-2419.

Byun SA, Koop BF, Reimchen TE (1997) North American black bear mtDNA phylogeography: Implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution* **51**, 1647-1653.

Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* **106**, 479-499.

Carstens BC, Degenhardt JD, Stevenson AL, Sullivan J (2005) Accounting for coalescent stochasticity in testing phylogeographical hypotheses: Modelling Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. *Molecular Ecology* **14**, 255-265.

Carstens BC, Richards CL, Crandall K (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution* **61**, 1439-1454.

Castric V, Bonney F, Bernatchez L (2001) Landscape structure and hierarchical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution* **55**, 1016-1028.

Cegelski CC, Waits LP, Anderson NJ, Flagstad O, Strobeck C, Kyle CJ (2006) Genetic diversity and population structure of wolverine (*Gulo gulo*) populations at the southern edge of their current distribution in North America with implications for genetic viability. *Conservation Genetics* **7**, 197-211.

Chappell DE, Van Den Bussche RA, Krizan J, Patterson B (2004) Contrasting levels of genetic differentiation among populations of wolverines (*Gulo gulo*) from northern Canada revealed by nuclear and mitochondrial loci. *Conservation Genetics* **5**, 759-767.

Chesser RT, Banks RC, Barker FK, Cicero C, Dunn JL, Kratter AW, Lovette IJ, Rasmussen PC, Remsen JV, Jr., Rising JD, Stotz DF, Winker K (2010) Fifty-first supplement to the American Ornithologists' Union *Check-list of North American Birds*. *The Auk* **127**, 726-744.

Chiucchi JE, Gibbs HL (2010) Similarity of contemporary and historical gene flow among highly fragmented populations of an endangered rattlesnake. *Molecular Ecology* **19**, 5345-5358.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The last glacial maximum. *Science* **325**, 710-714.

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Colbeck GJ, Gibbs HL, Marra PP, Hobson K, Webster MS (2008) Phylogeography of a widespread North American migratory songbird (*Setophaga ruticilla*). *Journal of Heredity* **99**, 453-463.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* **15**, 2833-2843.

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**, 19-31.

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

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.

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881-10890.

Costello AB, Down TE, Pollard SM, Pacas CJ, Taylor EB (2003) The influence of history and contemporary stream hydrology on the evolution of genetic diversity within species: An examination of microsatellite DNA variation in bull trout, *Salvelinus confluentus* (Pisces: Salmonidae). *Evolution* **57**, 328-344.

Coulon A, Cosson JF, Angibault JM, Cargnelutti B, Galan M, Morellet N, Petit E, Aulagnier S, Hewison AJM (2004) Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: An individual-based approach. *Molecular Ecology* **13**, 2841-2850.

Coykendall DK, Johnson S, Karl S, Lutz R, Vrijenhoek R (2011) Genetic diversity and demographic instability in *Riftia pachyptila* tubeworms from eastern Pacific hydrothermal vents. *BMC Evolutionary Biology* **11**, 96.

Crease TJ, Lee S-K, Yu S-L, Spitze K, Lehman N, Lynch M (1997) Allozyme and mtDNA variation in populations of the *Daphnia pulex* complex from both sides of the Rocky Mountains. *Heredity* **79**, 242-251.

Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* **15**, 49-62.

Cronin MA, MacNeil MD, Patton JC (2005) Variation in mitochondrial DNA and microsatellite DNA in caribou (*Rangifer tarandus*) in North America. *Journal of Mammalogy* **86**, 495-505.

Crossman EJ, Harington CR (1970) Pleistocene pike, *Esox lucius*, and *Esox* sp., from the Yukon Territory and Ontario. *Canadian Journal of Earth Sciences* **7**, 1130-1138.

Cumbaa SL, McAllister DE, Morlan RE (1981) Late Pleistocene fish fossils of *Coregonus*, *Stenodus*, *Thymallus*, *Catostomus*, *Lota*, and *Cottus* from the Old Crow Basin, northern Yukon, Canada. *Canadian Journal of Earth Sciences* **18**, 1740-1754.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* **93**, 504-509.

Dallimer M, Blackburn C, Jones PJ, Pemberton JM (2002) Genetic evidence for male biased dispersal in the red-billed quelea *Quelea quelea*. *Molecular Ecology* **11**, 529-533.

Dawson SE (1915) *Canada & Newfoundland* E. Stanford, Ltd., CA.

de Lafontaine G, Turgeon J, Payette S (2010) Phylogeography of white spruce (*Picea glauca*) in eastern North America reveals contrasting ecological trajectories. *Journal of Biogeography* **37**, 741-751.

DeChaine EG, Martin AP (2004) Historic cycles of fragmentation and expansion in *Parnassius smintheus* (Papilionidae) inferred using mitochondrial DNA. *Evolution* **58**, 113-127.

Delaunay B (1934) Sur la sphère vide. *Bulletin of the Academy of Sciences of the USSR* **7**, 793-800.

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**, 599-635.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences* **91**, 3166-3170.

Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): A platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3**, 167-169.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**, 2571-2581.

Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ (2002) The Laurentide and Innuitian ice sheets during the last glacial maximum. *Quaternary Science Reviews* **21**, 9-31.

Dyke AS, Moore A, Robertson L (2003) Deglaciation of North America. In: *Geological Survey of Canada: Open File 1574.*

Earl DA (2011) Structure Harvester v0.6.5. Available at http://taylor0.biology.ucla.edu/structureHarvester/.

Eddingsaas AA, Jacobsen BK, Lessa EP, Cook JA (2004) Evolutionary history of the arctic ground squirrel (*Spermophilus parryii*) in nearctic Beringia. *Journal of Mammology* **85**, 601-610.

Eidesen PB, Carlsen T, Molau U, Brochmann C (2007) Repeatedly out of Beringia: *Cassiope tetragona* embraces the Arctic. *Journal of Biogeography* **34**, 1559-1574.

Elias SA, Crocker B (2008) The Bering Land Bridge: A moisture barrier to the dispersal of steppe-tundra biota? *Quaternary Science Reviews* **27**, 2473-2483.

Elwen S, Meÿer MA, Best PB, Kotze PGH, Thornton M, Swanson S (2006) Range and movements of female Heaviside's dolphins (*Cephalorhynchus heavisidii*), as determined by satellite-linked telemetry. *Journal of Mammalogy* **87**, 866-877.

Ersts P (2010) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation.

Estoup A, Jarne P, Cornuet J-M (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.

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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

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.

Fedorov VB, Stenseth NC (2002) Multiple glacial refugia in the North American Arctic: Inference from phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 2071-2077.

Ficken MS, McLaren MA, Hailman JP (1996) Boreal Chickadee (*Parus hudsonicus*). In: *Birds of North America, No. 254 (A. Poole and F. Gill, eds)*. The Academy of Natural Sciences, Philadelphia, PA and The American Ornithologists' Union, Washington D.C.

Flagstad Ø, Røed KH (2003) Refugial origins of reindeer (*Rangifer tarandus l.*) inferred from mitochondrial DNA sequences. *Evolution* **57**, 658-670.

Flanders JON, Jones G, Benda P, Dietz C, Zhang S, Li G, Sharifi M, Rossiter SJ (2009) Phylogeography of the greater horseshoe bat, *Rhinolophus ferrumequinum*: Contrasting results from mitochondrial and microsatellite data. *Molecular Ecology* **18**, 306-318.

Fleischer RC, McIntosh CE, Tarr CL (1998) Evolution on a volcanic conveyor belt: Using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Molecular Ecology* **7**, 533-545.

Fleming MA, Cook JA (2002) Phylogeography of endemic ermine (*Mustela erminea*) in southeast Alaska. *Molecular Ecology* **11**, 795-807.

Flint RF (1947) *Glacial Geology and the Pleistocene Epoch* John Wiley & Sons, Inc, New York, NY.

Fontanella FM, Feldman CR, Siddall ME, Burbrink FT (2008) Phylogeography of *Diadophus punctatus*: Extensive lineage diversity and repeated patterns of historical demography in a trans-continental snake. *Molecular Phylogenetics and Evolution* **46**, 1049-1070.

Forsman ED, Anthony RG, Reid JA, Loschl PJ, Sovern SG, Taylor M, Biswell BL, Ellingson A, Meslow EC, Miller GS, Swindle KA, Thrailkill JA, Wagner FF, Seaman DE (2002) Natal and breeding dispersal of northern spotted owls. *Wildlife Monographs*, 1-35.

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* **10**, 773-784.

Frantz AC, Pope LC, Etherington TR, Wilson GJ, Burke T (2010) Using isolation-by-distance-based approaches to assess the barrier effect of linear landscape elements on badger (*Meles meles*) dispersal. *Molecular Ecology* **19**, 1663-1674.

Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* **8**, 1513-1520.

Galbreath KE, Hafner DJ, Zamudio KR, Agnew K (2010) Isolation and introgression in the intermountain West: Contrasting gene genealogies reveal the complex biogeographic history of the American pika (*Ochotona princeps*). *Journal of Biogeography* **37**, 344-362.

Gamache I, Jaramillo-Correa JP, Payette S, Bousquet J (2003) Diverging patterns of mitochondrial and nuclear DNA diversity in subarctic black spruce: Imprint of a founder effect associated with postglacial colonization. *Molecular Ecology* **12**, 891-901.

García LV (2004) Escaping the Bonferroni iron claw in ecological studies. *Oikos* **105**, 657-663.

Gavin DG (2009) The coastal-disjunct mesic flora in the inland Pacific Northwest of USA and Canada: Refugia, dispersal and disequilibrium. *Diversity and Distributions* **15**, 972-982.

Geffen ELI, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology* **13**, 2481-2490.

Gerardi S, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2010) From glacial refugia to modern populations: New assemblages of organelle genomes generated by differential cytoplasmic gene flow in transcontinental black spruce. *Molecular Ecology* **19**, 5265-5280.

Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D (2006) Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. *Molecular Biology and Evolution* **24**, 269-280.

Gibbs HL, Dawson RJG, Hobson KA (2000) Limited differentiation in microsatellite DNA variation among northern populations of the yellow warbler: Evidence for male-biased gene flow? *Molecular Ecology* **9**, 2137-2147.

Gill FB, Mostrom AM, Mack AL (1993) Speciation in North-American chickadees: I. Patterns of mtDNA genetic-divergence. *Evolution* **47**, 195-212.

Gill FB, Slikas B, Agro D (1999) Speciation in North American chickadees: II. Geography of mtDNA haplotypes in *Poecile carolinensis*. *The Auk* **116**, 274-277.

Gill FB, Slikas B, Sheldon FH (2005) Phylogeny of titmice (Paridae): II. Species relationships based on sequences of the mitochondrial cytochrome-b gene. *The Auk* **122**, 121-143.

Global Biodiversity Information Facility data portal (http: //data.gbif.org/datasets/resource/x): Royal British Columbia Museum (resource/475), Ontario Breeding Bird Atlas 2001-2005 (resource/63), Macaulay Library - Audio Data (resource/41), Bird Collection (resource/993), Peabody Ornithology DiGIR Service (resource/1020), Bird Tissue Collection (/resource/9148), University of Alberta Museums, Ornithology Collection (resource/773), Project FeederWatch (resource/45), eBird Bird Observation Checklist Database (resource/43), DMNS Bird Collection (resource/11617), Ontario Breeding Bird Atlas 1981-1985 (resource/62).

Godbout J, Beaulieu J, Bousquet J (2010) Phylogeographic structure of jack pine (*Pinus banksiana*; Pinaceae) supports the existence of a coastal glacial refugium in northeastern North America. *American Journal of Botany* **97**, 1903-1912.

Godbout J, Fazekas A, Newton C, Yeh FC (2008) Glacial vicariance in the Pacific Northwest: Evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. *Molecular Ecology* **17**, 2463-2475.

Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology* **14**, 3497-3512.

Goldstein DB, Pollock DD (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335-342.

Goudet J (1999) PCA-GEN (version 1.2). Lausanne, Switzerland. www.unil.ch/izea/softwares/pcegen.html.

Gradstein F, Ogg J, Smith A (2004) *A Geologic Time Scale* Cambridge University Press, Cambridge, UK.

Graham BA (2011) *The Population Genetics and Phylogeography of the Hairy Woodpecker (Picoides villosus)*, University of Lethbridge.

Graham RW, Lundelius EL, Graham MA, Schroeder EK, Toomey RS, Anderson E, Barnosky AD, Burns JA, Churcher CS, Grayson DK, Guthrie RD, Harington CR, Jefferson GT, Martin LD, McDonald HG, Morlan RE, Semken HA, Webb SD, Werdelin L, Wilson MC (1996) Spatial response of mammals to late quaternary environmental fluctuations. *Science* **272**, 1601-1606.

Greenwood PJ, Harvey PH (1982) The breeding and natal dispersal of birds. *Annual Review of Ecology and Systematics* **13**, 1-21.

Griffith SC, Stewart IRK, Dawson DA, Owens IPF, Burke T (1999) Contrasting levels of extra-pair paternity in mainland and island populations of the house sparrow (*Passer domesticus*): Is there an 'island effect'? *Biological Journal of the Linnean Society* **68**, 303-316.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: Testing hypotheses from the fossil record. *Molecular Ecology* **19**, 1877-1897.

Haavie J, Sætre G-P, Moum T (2000) Discrepancies in population differentiation at microsatellites, mitochondrial DNA and plumage colour in the pied flycatcher – Inferring evolutionary processes. *Molecular Ecology* **9**, 1137-1148.

Hallam A, Wignall PB (1997) *Mass Extinctions and their Aftermath* Oxford University Press, Oxford, UK.

Hanotte O, Zanon C, Pugh A, Greig C, Dixon A, Burke T (1994) Isolation and characterization of microsatellite loci in a passerine bird: The reed bunting *Emberiza schoeniclus*. *Molecular Ecology* **3**, 529-530.

Harbaugh DT, Wagner WL, Allan GJ, Zimmer EA, Bellwood D (2009) The Hawaiian Archipelago is a stepping stone for dispersal in the Pacific: An example from the plant genus *Melicope* (Rutaceae). *Journal of Biogeography* **36**, 230-241.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution* **16**, 700-706.

Harris S (1996) Climate-permafrost interaction between 18 ka and 8 ka bp in northwestern North America. *Erdkunde* **50**, 133-138.

Hatch SA, Meyers PM, Mulcahy DM, Douglas DC (2000) Seasonal movements and pelagic habitat use of murres and puffins determined by satellite telemetry. *The Condor* **102**, 145-154.

Hearn BJ, Neville JT, Curran WJ, Snow DP (2006) First record of the southern red-backed vole, *Clethrionomys gapperi*, in Newfoundland: Implications for the endangered Newfoundland marten, *Martes americana atrata*. *The Canadian Field-Naturalist* **120**.

Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313-318.

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

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography – Or seeing genes in space and time. *Molecular Ecology* **10**, 537-549.

Hewitt GM (2004a) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 183-195.

Hewitt GM (2004b) The structure of biodiversity – Insights from molecular phylogeography. *Frontiers in Zoology* **1**, 4.

Hickey AJR (2008) An alternate explanation for low mtDNA diversity in birds: An age-old solution? *Heredity* **100**, 443.

Hird S, Sullivan J (2009) Assessment of gene flow across a hybrid zone in red-tailed chipmunks (*Tamias ruficaudus*). *Molecular Ecology* **18**, 3097-3109.

Hoffman EA, Blouin MS (2004) Evolutionary history of the northern leopard frog: Reconstruction of phylogeny, phylogeography, and historical changes in population demography from mitochondrial DNA. *Evolution* **58**, 145-159.

Holder K, Montgomerie R, Friesen VL (1999) A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in rock ptarmigan (*Lagopus mutus*). *Evolution* **53**, 1936-1950.

Holland BS, Cowie RH (2007) A geographic mosaic of passive dispersal: Population structure in the endemic Hawaiian amber snail *Succinea caduca* (Mighels, 1845). *Molecular Ecology* **16**, 2422-2435.

Howes BJ, Lindsay B, Lougheed SC (2006) Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). *Molecular Phylogenetics and Evolution* **40**, 183-194.

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.

Hultén E (1937) *Outline of the History of Arctic and Boreal Biota During the Quaternary period* Bokförlags Aktiebolaget, Thule, Stolkholm.

Hurlbert SH (1971) The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52**, 577-586.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**, 282-291.

Jackson ST, Webb RS, Anderson KH, Overpeck JT, Webb T, Williams JW, Hansen BCS (2000) Vegetation and environment in eastern North America during the last glacial maximum. *Quaternary Science Reviews* **19**, 489-508.

Janzen DH (1967) Why mountain passes are higher in the tropics. *American Naturalist* **101**, 233-249.

Janzen FJ, Krenz JG, Haselkorn TS, Brodie ED (2002) Molecular phylogeography of common garter snakes (*Thamnophis sirtalis*) in western North America: Implications for regional historical forces. *Molecular Ecology* **11**, 1739-1751.

Jaramillo-Correa JP, Beaulieu J, Bousquet J (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology* **13**, 2735-2747.

Jaramillo-Correa JP, Beaulieu J, Khasa DP, Bousquet J (2009) Inferring the past from the present phylogeographic structure of North American forest trees: Seeing the forest for the genes. *Canadian Journal of Forest Research* **39**, 286-307.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**, 424-429.

Johansen AD, Latta RG (2003) Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. *Molecular Ecology* **12**, 293-298.

Johns GC, Avise JC (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution* **15**, 1481-1490.

Jones KL, Krapu GL, Brandt DA, Ashley MV (2005) Population genetic structure in migratory sandhill cranes and the role of Pleistocene glaciations. *Molecular Ecology* **14**, 2645-2657.

Jones M, Voss S, Ptacek M, Weisrock D, Tonkyn D (2006) River drainages and phylogeography: An evolutionary significant lineage of shovel-nosed salamander (*Desmognathus marmoratus*) in the southern Appalachians. *Molecular Phylogenetics and Evolution* **38**, 280-287.

Jost LOU (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-4026.

Kalinowski ST (2004) Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics* **5**, 539-543.

Kalinowski ST (2005) HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**, 187-189.

Kempf F, Boulinier T, De Meeûs T, Arnathau C (2009) Recent evolution of host-associated divergence in the seabird tick *Ixodes uriae*. *Molecular Ecology* **18**, 4450-4462.

Kenow KP, Meyer MW, Evers DC, Douglas DC, Hines J (2002) Use of satellite telemetry to identify common loon migration routes, staging areas and wintering range. *Waterbirds: The International Journal of Waterbird Biology* **25**, 449-458.

Kessler LG, Avise JC (1985) A comparative description of mitochondrial DNA differentiation in selected avian and other vertebrate genera. *Molecular Biology and Evolution* **2**, 109-125.

Klicka J, Spellman GM, Winker K, Chua V, Smith BT (2011) A phylogeographic and population genetic analysis of a widespread, sedentary North American bird: The hairy woodpecker (*Picoides villosus*). *The Auk* **128**, 346-362.

Koenig WD, Hooge PN, Stanback MT, Haydock J (2000) Natal dispersal in the cooperatively breeding acorn woodpecker. *The Condor* **102**, 492-502.

Krosby M, Rohwer S (2009) A 2000 km genetic wake yields evidence for northern glacial refugia and hybrid zone movement in a pair of songbirds. *Proceedings of the Royal Society B-Biological Sciences* **276**, 615-621.

Kvist L (2000) *Phylogeny and Phylogeography of European Parids*, University of Oulu.

Kvist L, Martens J, Ahola A, Orell M (2001) Phylogeography of a palaearctic sedentary passerine, the willow tit (*Parus montanus*). *Journal of Evolutionary Biology* **14**, 930-941.

Kvist L, Ruokonen M, Thessing A, Lumme J, Orell M (1998) Mitochondrial control region polymorphism reveal high amount of gene flow in Fennoscandian willow tits (*Parus montanus borealis*). *Hereditas* **128**, 133-143.

Kyes RC, Jones-Engel L, Chalise MK, Engel G, Heidrich J, Grant R, Bajimaya SS, McDonough J, Smith DG, Ferguson B (2006) Genetic characterization of rhesus macaques (*Macaca mulatta*) in Nepal. *American Journal of Primatology* **68**, 445-455.

Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, Baroni C (2002) Rates of evolution in ancient DNA from Adelie penguins. *Science* **295**, 2270-2273.

Latch E, Dharmarajan G, Glaubitz J, Rhodes O (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* **7**, 295-302.

Lepage D (2003) *Avibase - The World Bird Database*. Bird Studies Canada and Bird Life International.

Li M-H, Merilä J (2010) Genetic evidence for male-biased dispersal in the Siberian jay (*Perisoreus infaustus*) based on autosomal and Z-chromosomal markers. *Molecular Ecology* **19**, 5281-5295.

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

Loehr J, Worley K, Grapputo A, Carey J, Veitch A, Coltman DW (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 419-430.

Lovette IJ (2004) Mitochondrial dating and mixed support for the "2% rule" in birds. *The Auk* **121**, 1-6.

Lovette IJ, Clegg SM, Smith TB (2004) Limited utility of mtDNA markers for determining connectivity among breeding and overwintering locations in three neotropical migrant birds. *Conservation Biology* **18**, 156-166.

Luikart G, England PR (1999) Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* **14**, 253-256.

Luikart G, Sherwin WB, M. SB, Allendorf FW (1998) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology* **7**, 963-974.

Manni F, Guerard E (2004) Barrier vs. 2.2. Manual of the user. In: *Population genetics team, Museum of Mankind (Musee de l'Homme)*, Paris.

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by "Monmonier's algorithm". *Human Biology* **76**, 173-190.

Manthey JD, Klicka J, Spellman GM (2011) Cryptic diversity in a widespread North American songbird: Phylogeography of the brown creeper (*Certhia americana*). *Molecular Phylogenetics and Evolution* **58**, 502-512.

Martell MS, Henny CJ, Nye PE, Solensky MJ (2001) Fall migration routes, timing, and wintering sites of North American ospreys as determined by satellite telemetry. *The Condor* **103**, 715-724.

Martinez JG, Soler JJ, Soler M, Møller AP, Burke T (1999) Comparative population structure and gene flow of a brood parasite, the great spotted cuckoo (*Clamator glandarius*), and its primary host, the magpie (*Pica pica*). *Evolution* **53**, 269-278.

McDonald DB, Potts WK, Fitzpatrick JW, Woolfenden GE (1999) Contrasting genetic structures in sister species of North American scrub-jays. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 1117-1125.

McLaren MA (1975) Breeding biology of the boreal chickadee. *The Wilson Bulletin* **87**, 344-354.

McLaren MA (1976) Vocalizations of the boreal chickadee. *The Auk* **93**, 451-463.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* **11**, 5-18.

Milá B, McCormack JE, Castaneda G, Wayne RK, Smith TB (2007) Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proceedings of the Royal Society of London Series B-Biological Sciences* **274**, 2653-2660.

Milá B, Smith TB, Wayne RK (2006) Speciation and rapid phenotypic differentiation in the yellow-rumped warbler *Dendroica coronata* complex. *Molecular Ecology* **16**, 159-173.

Milot E, Gibbs HL, Hobson KA (2000) Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Molecular Ecology* **9**, 667-681.

Monmonier M (1973) Maximum-difference barriers: An alternative numerical regionalization method. *Geographical Analysis* **3**, 245-261.

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* **100**, 403-405.

Moriarty EC, Cannatella DC (2004) Phylogenetic relationships of the North American chorus frogs (*Pseudacris*: Hylidae). *Molecular Phylogenetics and Evolution* **30**, 409-420.

Musiani M, Leonard JA, Cluff HD, Gates CC, Mariani S, Paquet PC, VilÀ C, Wayne RK (2007) Differentiation of tundra/taiga and boreal coniferous forest wolves: Genetics, coat colour and association with migratory caribou. *Molecular Ecology* **16**, 4149-4170.

Nakagawa S (2004) A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology* **15**, 1044-1045.

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

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3323.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5269-5273.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145-164.

Nichols RA, Hewitt GM (1994) The genetic consequences of long-distance dispersal during colonization. *Heredity* **72**, 312-317.

Nilsson T (1983) *The Pleistocene: Geology and Life in the Quaternary Ice Age* D. Reidel Publishing Company, Dordrecht, Holland.

Orsini L, Corander J, Alasentie A, Hanski I (2008) Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology* **17**, 2629-2642.

Ostrand WD, Drew GS, Suryan RM, McDonald LL (1998) Evaluation of radio-tracking and strip transect methods for determining foraging ranges of black-legged kittiwakes. *The Condor* **100**, 709-718.

Otter K, Ratcliffe L, Michaud D, Boag PT (1998) Do female black-capped chickadees prefer high-ranking males as extra-pair partners? *Behavioral Ecology and Sociobiology* **43**, 25-36.

Päckert M, Dietzen C, Martens J, Wink M, Kvist L (2006) Radiation of Atlantic goldcrests *Regulus regulus* spp.: Evidence of a new taxon from the Canary Islands. *Journal of Avian Biology* **37**, 364-380.

Päckert M, Martens J, Tietze DT, Dietzen C, Wink M, Kvist L (2007) Calibration of a molecular clock in tits (Paridae) – Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? *Molecular Phylogenetics and Evolution* **44**, 1-14.

Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution* **55**, 859-868.

Peakall ROD, Smouse PE (2006) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Pereira S, Baker A (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Molecular Biology and Evolution* **23**, 1731-1740.

Pereira SL, Grau ET, Wajntal A (2004) Molecular architecture and rates of DNA substitutions of the mitochondrial control region of cracid birds. *Genome* **47**, 535-545.

Peters JL, Gretes W, Omland KE (2005) Late Pleistocene divergence between eastern and western populations of wood ducks (*Aix sponsa*) inferred by the 'isolation with migration' coalescent method. *Molecular Ecology* **14**, 3407-3418.

Peterson RT (1990) *Peterson Field Guides: Western Birds*, 3rd edn. Houghton Mifflin Company, New York, NY.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck GM, Demesure-Musch B, Palme A, Martin JP, Rendell S, Vendramin GG (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565.

Phillips RL, Andrews RD, Storm GL, Bishop RA (1972) Dispersal and mortality of red foxes. *The Journal of Wildlife Management* **36**, 237-248.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.

Phillips SJ, Dudík M (2008) Modeling of species distributions with MAXENT: New extensions and a comprehensive evaluation. *Ecography* **31**, 161-175.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* University of Chicago Press, Chicago, IL.

Piry S, Luikart G, Cornuet J-M (1999) Computer note. BOTTLENECK: A computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **90**, 502-503.

Pleines T, Jakob SS, Blattner FR (2008) Application of non-coding DNA regions in intraspecific analyses. *Plant Systematics and Evolution* **282**, 281-294.

Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution* **55**, 131-146.

Pons J, Pauses JG (2007) Acorn dispersal estimated by radio-tracking. *Oecologia* **153**, 903-911.

Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution* **15**, 997-1008.

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

Provan J, Bennett K (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* **23**, 564-571.

Pruett CL, Winker K (2005) Biological impacts of climatic change on a Beringian endemic: Cryptic refugia in the establishment and differentiation of the rock sandpiper (*Calidris ptilocnemis*). *Climatic Change* **68**, 219-240.

Pulgarin-Restrepo P (2011) *The Population History of the Downy Woodpecker (Picoides pubescens) in North America: Insights from Genetics, Ecological Niche Modeling and Bioaccoustics*, University of Lethbridge.

Pyle P (1997) *Identification Guide to North American Birds, Part I* Slate Creek Press, Bolinas, CA.

Quinn TW (1992) The genetic legacy of Mother Goose – Phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Molecular Ecology* **1**, 105-117.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Reeves AB, Drovetski SV, Fadeev IV (2008) Mitochondrial DNA data imply a stepping-stone colonization of Beringia by arctic warbler *Phylloscopus borealis*. *Journal of Avian Biology* **39**, 567-575.

Reiss RA, Ashworth AC, Schwert DP (1999) Molecular genetic evidence for the post-Pleistocene divergence of populations of the arctic-alpine ground beetle *Amara alpina* (Paykull) (Coleoptera: Carabidae). *Journal of Biogeography* **26**, 785-794.

Richards CL, Carstens BC, Lacey Knowles L (2007) Distribution modelling and statistical phylogeography: An integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography* **34**, 1833-1845.

Riley SPD, Pollinger JP, Sauvajot RM, York EC, Bromley C, Fuller TK, Wayne RK (2006) A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology* **15**, 1733-1741.

Rousset F (2008) Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Rowe KC (2004) Surviving the ice: Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences* **101**, 10355-10359.

Rozas J, Rozas R (1999) DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174-175.

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

Rueness EK, Stenseth NC, O'Donoghue M, Boutin S, Ellegren H, Jakobsen KS (2003) Ecological and genetic spatial structuring in the Canadian lynx. *Nature* **425**, 69-72.

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

Saitoh T, Alström P, Nishiumi I, Shigeta Y, Williams D, Olsson U, Ueda K (2010) Old divergences in a boreal bird supports long-term survival through the ice ages. *BMC Evolutionary Biology* **10**, 35.

Schauffler M, Jacobson GL (2002) Persistence of coastal spruce refugia during the Holocene in northern New England, USA, detected by stand-scale pollen stratigraphies. *Journal of Ecology* **90**, 235-250.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-371.

Scribner KT, Talbot S, L., Pearce JM, Pierson BJ, Bollinger KS, Derksen DV (2003) Phylogeography of Canada geese (*Branta canadensis*) in western North America. *The Auk* **120**, 889-907.

Scudder GGE, Ball GE, Danks HV (1993) Geographic distribution and biogeography of representative species of xeric grassland-adapted Nearctic Lygaeidae in western North America (Insecta: Heteroptera). *Memoirs of the Entomological Society of Canada*.

Shafer ABA, Côté SD, Coltman DW (2010a) Hot spots of genetic diversity descended from multiple Pleistocene refugia in an alpine ungulate. *Evolution* **65**, 125-138.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010b) Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**, 4589-4621.

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

Shirer H, W., Downhower JF (1968) Radio tracking of dispersing yellow bellied marmots. *Transactions of the Kansas Academy of Science (1903-)* **71**, 463-479.

Shlens J (2005) A tutorial on principal components analysis. In: *Systems Neurobiology Laboratory*. Salk Institute for Biological Studies, La Jolla, CA.

Sibley DA (2003) *The Sibley Field Guide to Birds of Eastern North America*, 1st edn. Random House, Inc., New York, NY.

Singh TR, Shneor O, Huchon D (2008) Bird mitochondrial gene order: Insight from 3 warbler mitochondrial genomes. *Molecular Biology and Evolution* **25**, 475-477.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**, 264-279.

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

Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* **206**, 353-373.

Sonsthagen SA, Talbot SL, Lanctot RB, Scribner KT, McCracken KG (2009) Hierarchical spatial genetic structure of common eiders (*Somateria mollissima*) breeding along a migratory corridor. *The Auk* **126**, 744-754.

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

Spellman GM, Klicka J (2007) Phylogeography of the white-breasted nuthatch (*Sitta carolinensis*): Diversification in North American pine and oak woodlands. *Molecular Ecology* **16**, 1729-1740.

Spellman GM, Riddle B, Klicka J (2007) Phylogeography of the mountain chickadee (*Poecile gambeli*): Diversification, introgression, and expansion in response to Quaternary climate change. *Molecular Ecology* **16**, 1055-1068.

Stanley SE, Harrison RG (1999) Cytochrome b evolution in birds and mammals: An evaluation of the avian constraint hypothesis. *Molecular Biology and Evolution* **16**, 1575-1585.

Steenhof K, Fuller MR, Kochert MN, Bates KK (2005) Long-range movements and breeding dispersal of prairie falcons from southwest Idaho. *The Condor* **107**, 481-496.

Steeves TE, Anderson DJ, Friesen VL (2005) The Isthmus of Panama: A major physical barrier to gene flow in a highly mobile pantropical seabird. *Journal of Evolutionary Biology* **18**, 1000-1008.

Stewart JR, Dalén L (2008) Is the glacial refugium concept relevant for northern species? A comment on Pruett and Winker 2005. *Climate Change* **86**, 19-22.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16**, 608-613.

Stow AJ, Sunnucks P, Briscoe DA, Gardner MG (2001) The impact of habitat fragmentation on dispersal of Cunningham's skink (*Egernia cunninghami*): Evidence from allelic and genotypic analyses of microsatellites. *Molecular Ecology* **10**, 867-878.

Stutchbury BJM, Tarof SA, Done T, Gow E, Kramer PM, Tautin J, Fox JW, Afanasyev V (2009) Tracking long-distance songbird migration by using geolocators. *Science* **323**, 896.

Swenson NG, Howard DJ (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* **166**, 581-591.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Taulman JF, Robbins LW (1996) Recent range expansion and distributional limits of the nine-banded armadillo (*Dasypus novemcinctus*) in the United States. *Journal of Biogeography* **23**, 635-648.

Thiel-Egenter C, Alvarez N, Holderegger R, Tribsch A, Englisch T, Wohlgemuth T, Colli L, Gaudeul M, Gielly L, Jogan N, Linder HP, Negrini R, Niklfeld H, Pellecchia M, Rioux D, Schoenswetter P, Taberlet P, van Loo M, Winkler M, Gugerli F, IntraBioDiv C (2011) Break zones in the distributions of alleles and species in alpine plants. *Journal of Biogeography* **38**, 772-782.

Topp CM, Winker K (2008) Genetic patterns of differentiation among five landbird species from the Queen Charlotte Islands, British Columbia. *The Auk* **125**, 461-472.

Tris JP, Bensch S, Carbonell R, Helbig A, Tellería JL (2004) Historical diversification of migration patterns in a passerine bird. *Evolution* **58**, 1819-1832.

Trizio I, Crestanello B, Galbusera P, Wauters LA, Tosi G, Matthysen E, Hauffe HC (2005) Geographical distance and physical barriers shape the genetic structure of Eurasian red squirrels (*Sciurus vulgaris*) in the Italian Alps. *Molecular Ecology* **14**, 469-481.

Uimaniemi L, Orell M, Kvist L, Jokimaki J, Lumme J (2003) Genetic variation of the Siberian tit *Parus cinctus* populations at the regional level: A mitochondrial sequence analysis. *Ecography* **26**, 98-106.

Uimaniemi L, Orell M, Mönkkönen M, Huhta E, Jokimäki J, Lumme J (2000) Genetic diversity in the Siberian jay *Perisoreus infaustus* in fragmented old-growth forests of Fennoscandia. *Ecography* **23**, 669-677.

Vallianatos M, Lougheed SC, Boag PT (2001) Phylogeography and genetic characteristics of a putative secondary-contact zone of the loggerhead shrike in central and eastern North America. *Canadian Journal of Zoology* **79**, 2221-2227.

Van Oosterhout C, 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.

Vandergast AG, Bohonak AJ, Weissman DB, Fisher RN (2006) Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: Phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: Stenopelmatus). *Molecular Ecology* **16**, 977-992.

Vandergast AG, Gillespie RG, Roderick GK (2004) Influence of volcanic activity on the population genetic structure of Hawaiian Tetragnatha spiders: Fragmentation, rapid population growth and the potential for accelerated evolution. *Molecular Ecology* **13**, 1729-1743.

Verhoeven KJF, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control: Increasing your power. *Oikos* **108**, 643-647.

Voronoi G (1908) Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die reine und angewandte Mathematik* **133**, 97-102.

Vos CC, Antonisse-De Jong AG, Goedhart PW, Smulders MJM (2001) Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity* **86**, 598-608.

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506-513.

Waltari E, Hijmans RJ, Townsend Peterson A, Nyari AS, Perkins SL, Guralnick RP (2007) Locating Pleisocene refugia: Comparing phylogeographic and ecological niche model predictions. *Public Library of Science One*, e563.

Walter R, Epperson BK (2001) Geographic pattern of genetic variation in *Pinus resinosa*: Area of greatest diversity is not the origin of postglacial populations. *Molecular Ecology* **10**, 103-111.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology* **19**, 2605-2608.

Wang M-T, Hsu Y-C, Yao C-T, Li S-H (2005) Isolation and characterization of 12 tetranucleotide repeat microsatellite loci from the green-backed tit (*Parus monticolus*). *Molecular Ecology Notes* **5**, 439-442.

Weimerskirch H, Corre ML, Marsac F, Barbraud C, Tostain O, Chastel O (2006) Postbreeding movements of frigatebirds tracked with satellite telemetry. *The Condor* **108**, 220-225.

Weir BS, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

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

Wenink PW, Baker AJ, Rosner H-U, Tilanus MGJ (1996) Global mitochondrial DNA phylogeography of holarctic breeding dunlins (*Calidris alpina*). *Evolution* **50**, 318-330.

White TH, Jr., Bowman JL, Leopold BD, Jacobson HA, Smith WP, Vilella FJ (2000) Influence of Mississippi alluvial valley rivers on black bear movements and dispersal: Implications for Louisiana black bear recovery. *Biological Conservation* **95**, 323-331.

Wiens JD, Reynolds RT, Noon BR (2006) Juvenile movement and natal dispersal of northern goshawks in Arizona. *The Condor* **108**, 253-269.

Williams A, Rabenold KN (2005) Male-biased dispersal, female philopatry, and routes to fitness in a social corvid. *Journal of Animal Ecology* **74**, 150-159.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helmbychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial-DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* **26**, 375-400.

Wisely SM, Buskirk SW, Russell GA, Aubry KB, Zielinski WJ (2004) Genetic diversity and structure of the fisher (*Martes pennanti*) in a peninsular and peripheral metapopulation. *Journal of Mammalogy* **85**, 640-648.

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

Yunick RP (1984) An assessment of the irruptive status of the boreal chickadee in New York state. *Journal of Field Ornithology* **55**, 31-37.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Molecular Ecology* **12**, 563-584.

Zink RM (1996) Comparative phylogeography in North American birds. *Evolution* **50**, 308-317.

Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**, 561-564.

Zink RM (2008) Microsatellite and mitochondrial DNA differentiation in the fox sparrow. *The Condor* **110**, 482-492.

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

Zink RM, Dittmann DL (1993) Gene flow, refugia, and evolution of geographic-variation in the song sparrow (*Melospiza melodia*). *Evolution* **47**, 717-729.

Zink RM, Weckstein JD, Hackett S (2003) Recent evolutionary history of the fox sparrows (Genus: *Passerella*). *The Auk* **120**, 522-527.

Appendix 1. Sample location, museum information and haplotype, listed by population. Each sample was given a project ID with location and number. The haplotype is the mtDNA haplotype from the concatenated sequence, as shown in Table 2.4, Figure 2.4 and Appendix 3. A '*' represents a unique haplotype, and 'n/a' were samples I was unable to sequence. Locations are general for simplicity, and where unavailable, GPS coordinates were approximated from location description. Museum samples are from the American Museum of Natural History, NY (AMNH), Burke Museum of Natural History and Culture, WA (BMNHC), Canadian Museum of Nature, ON (CMN), New Brunswick Museum, NB (NBM), New York State Museum, NY (NYSM), Royal Alberta Museum, AB (RABM), Royal British Columbia Museum, BC (RBCM), Royal Ontario Museum, ON (ROM), and the Smithsonian Museum of Natural History, DC (SMNH).

Project ID	Haplotype	Location	Band number	Museum	Latitude (°N)	Longitude (°W)
Alaska Anchorage						
AKA001	A	Eagle River campground, AK	2540-22802		61.3058	149.5705
AKA002	C	Eagle River campground, AK	2540-22811		61.3061	149.5673
AKA003	E	Eagle River campground, AK	2540-22812		61.3061	149.5673
AKA004	A	Eagle River campground, AK	2540-22814		61.3074	149.5692
AKA005	B	Eagle River Loop Rd, AK	2540-22822		61.2966	149.5388
AKA006	*	Eagle River Loop Rd, AK	2540-22823		61.2925	149.5388
AKA007	J	Eagle River Rd, AK	2540-22824		61.2684	149.3482
AKA008	*	Eagle River Rd, AK	2540-22826		61.2823	149.3894
AKA009	C	Eagle River campground, AK	2540-22829		61.3160	149.5749
AKA010	B	Chugach State Park, AK	2540-22830		61.2326	149.4560
AKA011	A	Chugach State Park, AK	2540-22831		61.2263	149.4560

AKA012	J	Eagle River Rd, AK	2540-22832		61.2763	149.3776
AKA013	C	Eagle River Rd, AK	2540-22833		61.2763	149.3776
AKA014	A	Eagle River greenbelt access, AK	2540-22834		61.2969	149.5322
AKA015	A	Beach Lake, AK	2540-22835		61.3921	149.5585
AKA016	*	Beach Lake, AK	2540-22836		61.3878	149.5513
AKA017	A	Beach Lake, AK	2540-22837		61.3878	149.5513
AKA018	*	Beach Lake, AK	2540-22838		61.3878	149.5513
AKA019	C	Lake Louise State Recreation Area, AK	UWBM# 53861 DAB 577	BMNHC	62.1167	146.5667
AKA020	A	Lake Louise State Recreation Area, AK	UWBM# 53862 DAB 578	BMNHC	62.1167	146.5667
AKA021	C	Elmendorf Air Force Base, AK	T#B13391 V#601804	SMNH	61.2861	149.7678
AKA022	*	Elmendorf Air Force Base, AK	T#B13431 V#601813	SMNH	61.3053	149.8125
AKA023	A	Elmendorf Air Force Base, AK	T#B13432 V#601811	SMNH	61.3053	149.8125
AKA024	M	Elmendorf Air Force Base, AK	T#B13433 V#601812	SMNH	61.3053	149.8125
AKA025	C	Elmendorf Air Force Base, AK	T#B13434 V#601810	SMNH	61.3053	149.8125
AKA026	A	Elmendorf Air Force Base, AK	T#B13435 V#622716	SMNH	61.3053	149.8125
AKA027	A	Elmendorf Air Force Base, AK	T#B13463 V#622702	SMNH	61.3053	149.8125
AKA028	N	Elmendorf Air Force Base, AK	T#B13464 V#622704	SMNH	61.3053	149.8125
AKA029	A	Elmendorf Air Force Base, AK	T#B13465 V#622703	SMNH	61.3053	149.8125
AKA030	A	Norh Fork Eagle River, AK	2540-22925		61.2969	149.5323
AKA031	*	Norh Fork Eagle River, AK	2540-22927		61.2969	149.5323
AKA032	N	Eklutna campground, AK	2540-22929		61.4073	148.1449
AKA033	*	Knik River, AK	2540-22930		61.4511	148.8211
AKA034	M	Knik River, AK	2540-22931		61.4511	148.8211

AKA035	C	Eagle River nature centre, AK	2540-22935	61.2293	149.2699
AKA036	C	Eagle River nature centre, AK	2540-22936	61.2293	149.2699
Alaska Fairbanks					
AKF001	*	Old Nanana Rd, AK	2540-22842	64.8164	148.1877
AKF002	A	Old Nanana Rd, AK	2540-22843	64.8164	148.1877
AKF003	*	Old Nanana Rd, AK	2540-22844	64.8164	148.1877
AKF004	C	Standard Creek Rd, AK	2540-22846	64.8117	148.2086
AKF005	H	Miller Hill Rd, AK	2540-22854	64.8681	147.8811
AKF006	*	Miller Hill Rd, AK	2540-22855	64.8681	147.8811
AKF007	M	Miller Hill Rd, AK	2540-22856	64.8681	147.8811
AKF008	*	Miller Hill Rd, AK	2540-22857	64.8681	147.8811
AKF009	*	Ester Dome Rd, AK	2540-22861	64.8836	148.0209
AKF010	J	Ester Dome Rd, AK	2540-22862	64.8836	148.0209
AKF011	*	Ester Dome Rd, AK	2540-22863	64.8836	148.0209
AKF012	*	UAF grounds, AK	2540-22865	64.8616	147.8291
AKF013	*	UAF grounds, AK	2540-22866	64.8616	147.8291
AKF014	I	UAF grounds, AK	2540-22867	64.8616	147.8291
AKF015	*	UAF grounds, AK	2540-22868	64.8616	147.8291
AKF016	G	Birch Hill Recreation Area, AK	2540-22870	64.8712	147.6468
AKF017	N	Birch Hill Recreation Area, AK	2540-22871	64.8712	147.6468
AKF018	N	Birch Hill Recreation Area, AK	2540-22872	64.8712	147.6468
AKF019	*	north of Fairbanks on #2, AK	2540-22873	65.0946	147.7338

AKF020	O	north of Fairbanks on #2, AK	2540-22874		65.0946	147.7338
AKF021	*	Two Rivers Rd, AK	2540-22878		64.8704	147.0428
AKF022	M	Steese Hwy, AK	2540-22881		64.2072	147.2111
AKF023	I	Tanana Valley campground, AK	2540-22884		64.8642	147.7607
AKF024	D	Delta Junction, AK	2540-22885		64.0625	145.6532
AKF025	E	Eielson Air Force Base, AK	T#B13339 V#601722	SMNH	64.7444	147.0569
AKF026	O	Eielson Air Force Base, AK	T#B13341 V#601721	SMNH	64.7444	147.0569
AKF027	*	Eielson Air Force Base, AK	T#B13460 V#622687	SMNH	64.6494	146.9789
AKF028	*	Eielson Air Force Base, AK	T#B13501 V#601754	SMNH	64.6494	146.9789
AKF029	*	Sheep Creek Rd, AK	2540-22910		64.8767	147.9071
AKF030	*	Sheep Creek Rd, AK	2540-22911		64.8767	147.9071
AKF031	*	Sheep Creek Rd, AK	2540-22914		64.8767	147.9071
AKF032	*	Sheep Creek Rd, AK	2540-22915		64.8767	147.9071
AKF033	N	Jones Rd, AK	2540-22917		64.9274	147.8963
AKF034	M	Murphy Dome area, AK	2540-22924		64.9501	148.1003
Alaska Wrangell-St. Elias						
AKW001	*	Kenny Lake, AK	n/a		61.7363	144.9541
AKW002	*	Kenny Lake, AK	2540-23164		61.7363	144.9541
AKW003	*	State Hwy 10b, AK	2540-23165		61.7069	144.8762
AKW004	A	State Hwy 10b, AK	2540-23166		61.7069	144.8762
AKW005	C	State Hwy 10c, AK	2540-23167		61.7069	144.8762
AKW006	*	State Hwy 10d, AK	2540-23168		61.7069	144.8762

AKW007	C	Old Edgerton Hwy, AK	2540-23170	61.7511	144.9898
AKW008	C	Old Edgerton Hwy, AK	2540-23172	61.7511	144.9898
AKW009	C	Old Edgerton Hwy, AK	2540-23173	61.7511	144.9898
AKW010	C	Old Edgerton Hwy, AK	2540-23181	61.7751	145.0364
AKW011	M	Old Edgerton Hwy, AK	2540-23183	61.7751	145.0364
AKW012	C	Old Edgerton Hwy, AK	2540-23185	61.7943	145.0714
AKW013	*	Old Edgerton Hwy, AK	2540-23186	61.7943	145.0714
AKW014	M	Old Edgerton Hwy, AK	2540-23189	61.7943	145.0714
AKW015	C	Old Edgerton Hwy, AK	2540-23190	61.7943	145.0714
AKW016	*	Old Edgerton Hwy, AK	2540-23192	61.7943	145.0714
AKW017	*	Old Edgerton Hwy, AK	2540-23194	61.8191	145.1415
AKW018	C	Old Edgerton Hwy, AK	2540-23195	61.8191	145.1415
AKW019	*	Richardson x Old Edgerton Hwy, AK	2540-23198	61.8245	145.2193
AKW020	*	Richardson x Old Edgerton Hwy, AK	2540-23199	61.8245	145.2193
AKW021	*	WISE Headquarters, AK	2540-22902	61.8039	145.0931
AKW022	G	WISE Headquarters, AK	2540-22903	61.8039	145.0931
AKW023	A	Old Edgerton Hwy, AK	2540-22904	61.8218	145.1714
AKW024	C	Old Edgerton Hwy, AK	2540-22905	61.8218	145.1714
AKW025	*	Old Edgerton Hwy, AK	2540-22906	61.8218	145.1714
Northern British Columbia					
NBC001	N	Dease Lake, BC	2490-57759	58.5186	130.0368
NBC002	N	Dease Lake, BC	3111-48302	58.4350	129.8940

NBC003	*	Dease Lake, BC	2520-39842	58.4350	129.8940
NBC004	N	Dease Lake, BC	3111-48303	58.4350	129.8940
NBC005	F	Dease Lake, BC	2490-57760	58.4350	129.8940
NBC006	M	Dease Lake, BC	1501-42930	58.5069	130.0231
NBC007	D	Dease Lake, BC	2520-39843	58.5069	130.0231
NBC008	F	Dease Lake, BC	2520-39844	58.5069	130.0231
NBC009	D	Dease Lake, BC	2520-39845	58.5069	130.0231
NBC010	N	Dease Lake, BC	2520-39846	58.5069	130.0231
NBC011	H	Dease Lake, BC	2520-39847	58.5069	130.0231
NBC012	H	Dease Lake, BC	2520-39848	58.5069	130.0231
NBC013	H	Dease Lake, BC	2520-39849	58.5069	130.0231
NBC014	H	Dease Lake, BC	2520-39852	58.5069	130.0231
NBC015	F	Dease Lake, BC	2520-39853	58.5069	130.0231
NBC016	H	Dease Lake, BC	2520-39857	58.4473	130.0154
NBC017	N	Telegraph Creek, BC	2520-39864	57.9126	131.2096
NBC018	N	Dease Lake, BC	2520-39869	58.4468	130.0039
NBC019	N	Dease Lake, BC	2520-39870	58.4468	130.0039
NBC020	*	Dease Lake, BC	2520-39871	58.4468	130.0039
NBC021	*	Dease Lake, BC	2520-39872	58.4481	130.0175
NBC022	*	Dease Lake, BC	2520-39873	58.4303	129.9868
Central British Columbia					
CBC001	M	Fort St. James, BC	2350-76019	54.4431	124.2542

CBC002	N	Fort St. James, BC	2350-75935		54.4431	124.2542
CBC003	*	Fort St. James, BC	2350-76020		54.4431	124.2542
CBC004	N	Fort St. James, BC	2350-75936		54.4431	124.2542
CBC005	N	Fort St. James, BC	2520-39841		52.5276	124.1739
CBC006	M	Smithers, BC	249-57769		54.7594	127.3617
CBC007	*	Smithers, BC	2490-57770		54.7594	127.3617
CBC008	N	Smithers, BC	2490-57774		54.7413	127.3297
CBC009	K	Smithers, BC	2490-57775		54.7457	127.3399
CBC010	L	Smithers, BC	2490-57776		54.7457	127.3399
CBC011	O	Fort Nelson, BC	CN #017479	RBCM	59.4167	120.7833
CBC012	M	Smithers, BC	n/a		54.7669	127.2737
CBC013	*	Smithers, BC	2500-94905		54.7669	127.2737
CBC014	K	Smithers, BC	2500-94921		54.7445	127.3222
CBC015	M	Smithers, BC	2500-94922		54.7445	127.3222
CBC016	L	Smithers, BC	2500-94923		54.7669	127.2737
CBC017	M	Smithers, BC	2500-94924		54.7669	127.2737
Central Alberta						
CAB001	*	Hinton, AB	2520-39821		53.3828	117.6943
CAB002	*	Edson, AB	2520-39824		53.6883	116.7824
CAB003	M	Edson, AB	2520-39825		53.6883	116.7824
CAB004	M	Cadomin, AB	2520-39831		53.0342	117.3286
CAB005	M	Hinton, AB	2520-39832		53.3915	117.6757

CAB006	O	Hinton, AB	2520-39833		53.3915	117.6757
CAB007	M	Hinton, AB	2520-39834		53.3915	117.6757
CAB008	N	Hinton, AB	2520-39835		52.3848	117.6874
CAB009	M	Hinton, AB	2520-39836		53.3899	117.6772
CAB010	*	Hinton, AB	2520-39837		53.3944	117.6696
CAB011	*	Mt. Robson, BC	2520-39840		52.8571	118.6396
CAB012	O	near Bearberry, AB	Z95.11.4 cat#30512	RABM	51.8830	115.0170
Southern Alberta						
SAB001	N	Waterton Lakes NP, AB	2490-57736		49.0409	114.0331
SAB002	*	Waterton Lakes NP, AB	2490-57799		49.0580	113.9253
SAB003	*	Waterton Lakes NP, AB	2490-57800		49.0275	114.0553
SAB004	*	Waterton Lakes NP, AB	2500-94959		49.0275	114.0553
SAB005	*	Waterton Lakes NP, AB	2500-94960		49.0272	114.0436
Saskatchewan						
SK001	*	Prince Albert NP, SK	2500-94896		53.9725	106.2903
SK002	N	Prince Albert NP, SK	2500-94897		53.9725	106.2903
SK003	*	Prince Albert NP, SK	2500-94939		53.9807	106.2938
SK004	*	Prince Albert NP, SK	2500-94940		53.9807	106.2938
SK005	*	Prince Albert NP, SK	2500-94941		53.9807	106.2938
SK006	P	Prince Albert NP, SK	2500-94944		53.9807	106.2938
SK007	P	Prince Albert NP, SK	2500-94945		53.9807	106.2938

SK008	Q	Prince Albert NP, SK	2500-94947		53.9807	106.2938
SK009	P	Prince Albert NP, SK	2490-57786		53.9725	106.2903
SK010	Q	Prince Albert NP, SK	2500-94948		53.9821	106.2920
SK011	N	Prince Albert NP, SK	2500-94949		53.9821	106.2920
SK012	Q	Prince Albert NP, SK	2500-94950		53.9821	106.2920
SK013	*	Prince Albert NP, SK	2500-94951		53.9821	106.2920
SK014	M	Prince Albert NP, SK	2500-94952		53.9820	106.2957
SK015	M	Prince Albert NP, SK	2500-94953		53.9820	106.2957
SK016	M	Prince Albert NP, SK	2500-94954		53.9820	106.2957
SK017	M	Prince Albert NP, SK	2500-94955		53.9661	106.2716
SK018	M	Prince Albert NP, SK	2490-57789		54.0256	106.2429
SK019	M	Prince Albert NP, SK	2500-94956		53.9779	106.1556

Northern Ontario

NON001	n/a	near Polar Bear PP, Kenora District, ON	cat #137237	ROM	54.3402	84.5216
NON002	*	near Polar Bear PP, Kenora District, ON	cat #137238	ROM	54.3402	84.5216
NON003	*	near Polar Bear PP, Kenora District, ON	cat #137239	ROM	54.3402	84.5216
NON004	O	near Polar Bear PP, Kenora District, ON	cat #137240	ROM	54.3402	84.5216
NON005	R	near Polar Bear PP, Kenora District, ON	cat #137241	ROM	54.3402	84.5216
NON006	*	near Polar Bear PP, Kenora District, ON	cat #137243	ROM	54.3402	84.5216
NON007	*	near Polar Bear PP, Kenora District, ON	cat #137244	ROM	54.3402	84.5216
NON008	*	near Polar Bear PP, Kenora District, ON	cat #137245	ROM	54.3402	84.5216
NON009	*	near Polar Bear PP, Kenora District, ON	cat #137246	ROM	54.3402	84.5216

NON010	*	near Polar Bear PP, Kenora District, ON	cat #139944	ROM	54.3402	84.5216
NON011	*	near Polar Bear PP, Kenora District, ON	cat #139845	ROM	54.5254	84.8977
NON012	M	near Polar Bear PP, Kenora District, ON	cat #139846	ROM	54.5254	84.8977
NON013	*	near Polar Bear PP, Kenora District, ON	cat #139961	ROM	54.5117	84.9109
NON014	P	near Polar Bear PP, Kenora District, ON	cat #140073	ROM	54.5117	84.9109
NON015	*	near Polar Bear PP, Kenora District, ON	cat #139889	ROM	54.4873	84.9158
NON016	P	near Polar Bear PP, Kenora District, ON	cat #139898	ROM	54.5126	84.8591
Northern Québec						
NQC001	O	Territoire de Jamésie, Québec	80135	CMN	52.6500	76.3330
NQC002	*	Territoire de Jamésie, Québec	80136	CMN	52.6500	76.3330
NQC003	O	Territoire de Jamésie, Québec	80137	CMN	52.2420	78.5610
NQC004	P	Territoire de Jamésie, Québec	80138	CMN	53.3170	77.2830
NQC005	*	Territoire de Jamésie, Québec	80139	CMN	53.3170	77.2830
NQC006	P	Territoire de Jamésie, Québec	80140	CMN	53.3170	77.1330
NQC007	*	Territoire de Jamésie, Québec	80141	CMN	53.7330	75.9830
NQC008	*	Territoire de Jamésie, Québec	80142	CMN	53.7330	75.9830
NQC009	*	Territoire de Jamésie, Québec	80143	CMN	53.7330	75.9830
NQC010	*	Territoire de Jamésie, Québec	80144	CMN	53.7330	75.9830
NQC011	P	Territoire de Jamésie, Québec	80145	CMN	53.7330	75.9830
Nova Scotia & New Brunswick						
NSNB001	O	Economy Lake, NS	2490-57556		45.3850	63.9114

NSNB002	U	Economy Lake, NS	2490-57557		45.3850	63.9114
NSNB003	R	Economy Lake, NS	2490-57558		45.3850	63.9114
NSNB004	S	Economy Lake, NS	2490-57559		45.3850	63.9114
NSNB005	*	Economy Lake, NS	2490-57560		45.3850	63.9114
NSNB006	S	Fundy NP, NB	2490-57568		45.6150	65.0356
NSNB007	S	Fundy NP, NB	2490-57569		45.6150	65.0356
NSNB008	U	Fundy NP, NB	2490-57570		45.6150	65.0356
NSNB009	U	Fundy NP, NB	2490-57572		45.6150	65.0356
NSNB010	*	Meat Cove, NS	2490-57573		47.0428	60.5653
NSNB011	*	Mount Mitchell, NB	boch NSNB11		47.4242	66.9225
NSNB012	*	Mount Mitchell, NB	3510-55838		47.4242	66.9225
NSNB013	U	Mount Mitchell, NB	3510-55855		47.4242	66.9225
NSNB014	R	Mount Mitchell, NB	3510-55851		47.4242	66.9225
NSNB015	*	Mount Mitchell, NB	3510-55860		47.4242	66.9225
NSNB016	*	Kings Co., NB	010529	NBM	46.6500	64.8700
NSNB017	*	Kent Co., NB	007361	NBM	46.6500	64.8700
NSNB018	n/a	Kent Co., NB	007362	NBM	45.3500	66.1300
NSNB019	P	Fundy NP, NB	2500-94819		45.5657	64.9844
NSNB020	U	Middle Musquodoboit, NS	2500-94829		45.0137	63.0495
NSNB021	T	Musquodoboit Valley, NS	2500-94888		45.0141	63.0278
NSNB022	U	Antigonish, NS	2500-94839		45.7164	61.9424

Labrador

LAB001	U	Happy Valley-Goose Bay, LAB	2500-94859	53.2966	60.3742
LAB002	*	Happy Valley-Goose Bay, LAB	2500-94860	53.2966	60.3742
LAB003	*	Happy Valley-Goose Bay, LAB	2500-94861	53.3153	60.3820
LAB004	V	Happy Valley-Goose Bay, LAB	2500-94862	53.3153	60.3820
LAB005	*	Happy Valley-Goose Bay, LAB	2500-94864	53.3404	60.4109
LAB006	P	Happy Valley-Goose Bay, LAB	2500-94865	53.3404	60.4109
LAB007	P	Happy Valley-Goose Bay, LAB	2500-94866	53.3404	60.4109
LAB008	U	Happy Valley-Goose Bay, LAB	2500-94867	53.2911	60.3593
LAB009	*	Happy Valley-Goose Bay, LAB	2500-94868	53.2911	60.3593
LAB010	T	Happy Valley-Goose Bay, LAB	2500-94869	53.4057	60.4211
LAB011	*	Happy Valley-Goose Bay, LAB	2500-94870	53.4057	60.4211
LAB012	U	Happy Valley-Goose Bay, LAB	2500-94872	53.3760	60.4245
LAB013	*	Happy Valley-Goose Bay, LAB	2500-94873	53.3779	60.4199
LAB014	*	Happy Valley-Goose Bay, LAB	2500-94874	53.3779	60.4199
LAB015	P	Happy Valley-Goose Bay, LAB	2500-94875	53.3779	60.4199
LAB016	*	Happy Valley-Goose Bay, LAB	2500-94877	53.3747	60.3963
LAB017	*	Happy Valley-Goose Bay, LAB	2500-94878	53.4322	60.3755
LAB018	V	Happy Valley-Goose Bay, LAB	2500-94880	53.3928	60.3820
LAB019	P	Happy Valley-Goose Bay, LAB	2500-94881	53.3928	60.3820
LAB020	*	Happy Valley-Goose Bay, LAB	2500-94882	53.3928	60.3820
LAB021	T	Happy Valley-Goose Bay, LAB	2500-94883	53.3928	60.3820
LAB022	P	Happy Valley-Goose Bay, LAB	2500-94884	53.3928	60.3820

New York

NY001	*	Hamilton Co., NY	DOT-16803	AMNH	43.3411	74.2183
NY002	U	Hamilton Co., NY	DOT-16804	AMNH	43.3411	74.2183
NY003	U	Hamilton Co., NY	DOT-16814	AMNH	43.3411	74.2183
NY004	U	Madawaska State Preserve, NY	zo-11097	NYSM	45.1764	74.7167
NY005	U	Hurricane Mountain, NY	zo-11108	NYSM	45.1764	74.7167
NY006	*	Hurricane Mountain, NY	zo-11109	NYSM	44.1106	74.4869

Newfoundland

NL001	P	Sir Richard Squires PP, NL	2490-57578		49.3472	57.3353
NL002	P	Sir Richard Squires PP, NL	2490-57580		49.3472	57.3353
NL003	P	Sir Richard Squires PP, NL	2490-57581		49.3472	57.3353
NL004	*	Sir Richard Squires PP, NL	2490-57583		49.3472	57.3353
NL005	*	Sir Richard Squires PP, NL	2490-57584		49.3472	57.3353
NL006	*	Barachois Pond PP, NL	2490-57592		48.4536	58.4325
NL007	*	Pasadena, NL	2490-57596		49.0142	57.5981
NL008	P	Pasadena, NL	2490-57597		49.0142	57.5981
NL009	P	Pasadena, NL	2490-57598		49.0142	57.5981
NL010	*	Pasadena, NL	2490-57601		49.0142	57.5981
NL011	P	Pasadena, NL	2490-57602		49.0142	57.5981
NL012	Z	Pasadena, NL	2490-57603		49.0142	57.5981
NL013	X	Pasadena, NL	2490-57605		49.0142	57.5981
NL014	Y	Pasadena, NL	2490-57607		49.0142	57.5981

NL015	X	Pasadena, NL	2490-57608	49.0142	57.5981
NL016	P	Pasadena, NL	2490-57609	49.0142	57.5981
NL017	Y	Eagle Mountain, NL	2490-57614	49.8372	57.2575
NL018	P	Eagle Mountain, NL	2490-57615	49.8372	57.2575
NL019	P	Eagle Mountain, NL	2490-57616	49.8372	57.2575
NL020	P	Eagle Mountain, NL	2490-57617	49.8372	57.2575
NL021	Y	Eagle Mountain, NL	2490-57618	49.8372	57.2575
NL022	P	Eagle Mountain, NL	2490-57619	49.8372	57.2575
NL023	*	Eagle Mountain, NL	2490-57620	49.8372	57.2575
NL024	*	Terra Nova NP, NL	2490-57624	48.5197	53.9672
NL025	P	Terra Nova NP, NL	2490-57625	48.5197	53.9672
NL026	P	Terra Nova NP, NL	2490-57630	48.5197	53.9672
NL027	P	Sir Richard Squires PP, NL	2500-94843	49.3459	57.1705
NL028	P	Sir Richard Squires PP, NL	2500-94844	49.3459	57.1705
NL029	*	Sir Richard Squires PP, NL	2500-94845	49.3459	57.1705
NL030	W	Gros Morne NP, NL	2500-94846	49.4588	57.7596
NL031	*	Gros Morne NP, NL	2500-94847	49.4588	57.7596
NL032	*	Gros Morne NP, NL	2500-94848	49.4588	57.7596
NL033	*	Gros Morne NP, NL	2500-94849	49.4588	57.7596
NL034	W	Gros Morne NP, NL	2500-94851	49.4588	57.7596
NL035	Z	Gros Morne NP, NL	2500-94853	49.9390	57.7599
NL036	P	Gros Morne NP, NL	2500-94856	49.6216	57.9211

Appendix 2. Description of some genetic analysis programs

Spatial analysis of molecular variance (SAMOVA)

A SAMOVA assigns the most differentiated group(s) to the data using a simulated annealing approach (Dupanloup *et al.* 2002). The process begins by assigning populations to groups such that for K groups, K-1 groups have 1 individual and 1 group has the rest. Populations in a group must be geographically adjacent. The program then moves populations, one at a time, to a different neighbouring group, and the new Φ_{CT} value (among-group variation) is calculated. The process is less likely to get stuck at a local maximum than a McMC approach because the new group can be accepted whether the Φ_{CT} value increases or decreases (Dupanloup *et al.* 2002). The program continues until all possible groups from that starting point have been tested, and the group with the maximum Φ_{CT} value is reported. Each run begins with a new random starting point, and 100 starting points (iterations) is recommended by the manual.

Principal components analysis

A principal components analysis (PCA; R), or a principle coordinates analysis (PCO; GenAlEx), is an application of linear algebra whose aim is to minimise the covariance (redundancy) while maximising the variance (signal-to-noise ratio) in a set of random data (Shlens 2005). It accomplishes this by estimating a matrix P $(p_1 \quad \cdots \quad p_m)$ which converts the data matrix X $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ to an orthonormal matrix Y such that Y = PX $\begin{pmatrix} p_1x_1 & \cdots & p_mx_1 \\ \vdots & \ddots & \vdots \\ p_1x_n & \cdots & p_mx_n \end{pmatrix}$, and the variance is maximised. The eigenvectors in P are the principal components (coordinates) of X, and an orthonormal matrix (Y) is one where the transpose of the matrix (Y^T) is its inverse (Y^{-1}), and $YY^T = I$ (the identity matrix) $\begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$ (Shlens 2005). The difference between a PCA and a PCO is the program in which it is run.

Voronoi tessellation, Delaunay triangulation and Monmonier's algorithm

Voronoi diagrams (Voronoi 1908) represent the spatial neighbourhood surrounding a sample (Manni *et al.* 2004). The neighbourhood is defined such that every point is closer to its sample than any other. This results in lines being drawn equidistant between each sample (Figure A). The samples are then connected to their neighbouring points such that the Voronoi lines are bisected perpendicularly. This method, Delaunay triangulation (Delaunay 1934), allows each point to be connected to its closest neighbour.

The Monmonier's algorithm (Monmonier 1973) is then applied to the Delaunay map. The algorithm is designed to detect the maximum difference in the matrix applied (e.g. genetic distances). A barrier is drawn following the highest genetic distances as described in Manni *et al.*(2004).

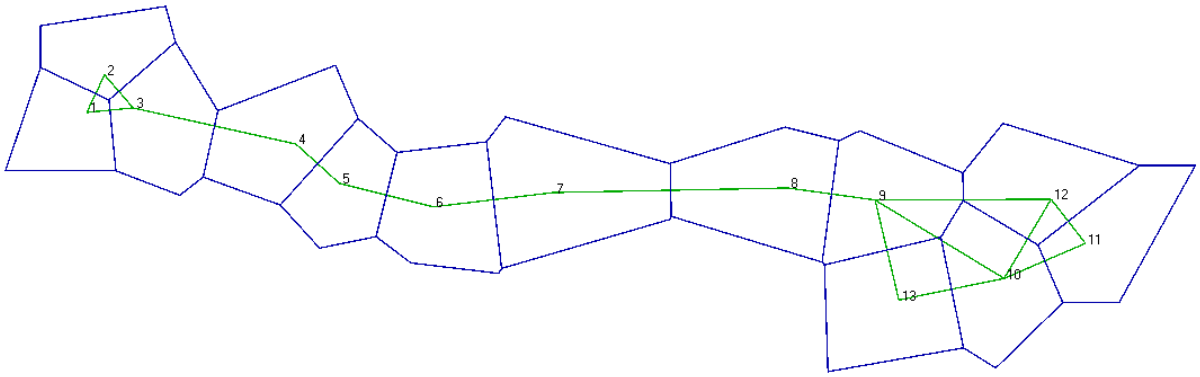

Figure A.1. A Voronoi tessellation and Delaunay triangulation of the 13 boreal chickadee populations, executed in BARRIER v2.2 (Manni *et al.* 2004). Each population is represented by a number, the Voronoi diagram is drawn in blue and the subsequent Delaunay triangulation in green.

Appendix 3. Variable sites table showing the position of all variable sites and the allocation among haplotypes. The letters represent shared haplotypes. A '.' represents a conserved base, and a '-' shows the position of the insertion/deletion.

	Control Region						ATPase 6-8					
								1111111	1111111111	1111111111	1111111111	111111111
	11	1111111222	2222222222	2222333334	4444455555	556667	7888	8990000001	1111111222	2222222333	3334444444	555566666
	2245688944	5566699000	0001111378	8889023896	7788901358	895670	8148	9250244580	0467888122	2233469012	5790123457	236900112
	5806658605	3414935012	3570123810	3462756607	6737715451	607913	3346	0033647513	5477345123	8912387308	1062163414	571747127
	TCCACGGTCT	GAGAC-CCAA	ACCCCGATCA	CCACGGTCGC	AATCCCACGT	TTTATC	TTCA	GTGCATAATT	AAGACAATCA	TAACTTTCAC	CCTTGAGGCC	TGATGCCCA
A	....	.....-....	....	....	....	......	....	....	....	....	....	....
B	....	.....-....	....	....	....	......	....	....	....	.........T	....	....
C	....	.....-T...	....	....	....	......	....	....	....	..G.......	....	....
D	....	.....-T...	....	....	....	......	....	....	....	..G.......	....	.A.......
E	....	.....-T...	....	....	....	......	....	....	....	..G.......	....	.....T...
F	....	.....-T...	...T......	....	....	......	....	....	....	....	T.........	....
G	....	.....-T...	....	...T......	....	......	....	A.........	....	....	....	....
H	....	.....-T...	....	.......T..	....	......	....	A.........	....	....	....	....
I	....	.....-T...	....	....	....	.....T	....	A.........	....	....	....	....
J	....	.....-T...	..T.......	....	....	......	....	A.........	....	....	....	....
K	C.........	.....-T...	....	....	....	......	....	A.........	....	....	....	....
L	....	.....-TT..	....	....	........A.	......	....	A.........	....T.....	....	........T.	....
M	....	.....-TT..	....	....	....	......	....	A.........	....T.....	....	........T.	....
N	....	.....-T...	....	....	....	......	....	A.........	....	....	....	....
O	....	.....-T...	....	....	....	......	....	....	....	....	....	....
P	....	.....-T...	....	....	....	C.....	....	....	....	....	.......A..	....
Q	....	.....-T...	....	T.........	....	C.....	....	....	....	....	....	....
R	..T.......	.....-T...	....	....	....	C.....	....	....	....	....	.......A..	....
S	....	.....-T...	G.........	....	........A.	C.....	....	....	....	....	.......A..	....

	TCCACGGTCT	GAGAC-CCAA	ACCCCGATCA	CCACGGTCGC	AATCCCACGT	TTTATC	TTCA	GTGCATAATT	AAGACAATCA	TAACTTTCAC	CCTTGAGGCC	TGATGCCCA
T	....	.....-T...	...T......	....	....	C.....	....	....	....	....	.......A..	....
U	....	.....-T...	...T......	....	....	C.....	....	....	....	C.........	.......A..	....
V	....	.....-T...	....	....	.G........	C.....	....	....	....	....	.......A..	....
W	....	.....-T...	....	....	....	C.....	....	....	....	....	......AA..	....
X	.....A....	.....-T...	....	....	....	C.....	....	....	....	....	.......A..	....
Y	....	.....-T...	....	....	....	C.....	....	....	....	....	....A..A..	....
Z	....	.....-T...	....	....	....	C.C...	....	....	....	....	....A..A..	....
AKA06	....	.....-TT..	.......C..	....	....	......	....	A.........	....T.....	......C...	........T.	....
AKA08	....	.....-TT..	....	........A.	....	......	....	A.........	....T.....	....	........T.	....
AKA16	....	.....-....	....	....	....	......	....	A.........	....	....	....	....
AKA18	....	A....-....	....	....	....	......	....	....	....	....	....	....
AKA22	....	A....-T...	......G...	....	....	C.....	....	....	....	....	....	....
AKA31	....	.....-....	....	....	....	......	....	....	....	....	....	.A.......
AKA33	...G......	.....-T...	....	....	....	...G..	....	....	....	..G.......	....	....
AKF01	....	.....-T...	..T.......	....	....	......	....	A.........	....	....	....	....A....
AKF03	....	.....-TT..	....	....	....	......	....	A.........	G...T.....	....	........T.	....
AKF06	....	.....-T...	....	....	....	......	....	A...G.....	....	....	....	....
AKF08	....	.....-T...	....	....	....	....C.	....	A.........	....	.......A..	....	....
AKF09	....	.....-T...	....	....	........A.	......	....	....	....	....	T.........	....
AKF11	....	A....-T...	....	....	....	......	....	A.........	....	....	....	....
AKF12	....	.....-T...	....	....	....	......	....	A......G..	....	....	....	....
AKF13	....	.....-T...	....	T.........	....	......	....	A.........	....	....	....	....
AKF15	....	.....-TT..	....	....	....	......	....	A.........	....T.G...	....	........T.	....
AKF19	.....A....	.....-T...	....	....	....	......	....	....	....	....	....	....
AKF21	....	.....-T...	....	....	C	......	....	....	....	....	....	....
AKF27	....	.....AT...	....	....	....	......	....	A.........	....	....	....	....
AKF28	....	.....-T...	....	....	....	......	....	....	.......C..	..G.......	....	....

	TCCACGGTCT	GAGAC-CCAA	ACCCCGATCA	CCACGGTCGC	AATCCCACGT	TTTATC	TTCA	GTGCATAATT	AAGACAATCA	TAACTTTCAC	CCTTGAGGCC	TGATGCCCA
AKF29	....	A....-T...	....	....	....	......	....	A.........	....	.....C....	....	....
AKF30	....	.....AT...	....	....	....	......	....	A.........	....	.....C....	....	....
AKF31	....	.....-T...	....A.....	....	....	......	....	....	....	.....C....	.......A..	....
AKF32	....	.....-TT..	....	....	....	......	....	A.........	....T.....	.....C....	........T.	....
AKW01	....	.....-T..G	....	....	....	......	....	....	....	....	....	....
AKW02	..T.......	.....-T...	....	....	....	......	....	...T......	....	..G..C....	....	....
AKW03	....	.....-T...	....	....	....	......	....	....	....	..G..C....	....	....
AKW06	....	.....-T...	..T.......	....	......G...	......	....	A.........	....	....	....	...C.....
AKW13	...G......	.....-....	....	....	....	......	....	.....C....	....	....	....	....
AKW16	...G......	.....-....	....	....	....	......	....	....	....	....	....	....
AKW17	...G......	.....-TT..	....	....	....	......	....	A.........	....T.....	....	........T.	....
AKW19	....	.....-....	....	....	..C.......	......	....	....	....	....	....	....
AKW20	....	.....-T...	....	......C...	....	......	....	......G...	....	....	....	....
AKW21	....	.....-TT..	....	....	....	......	....	A.........	....T.....	....	...C....T.	....
AKW25	....	.....-T...	....	..G.......	....	......	....	....	....	..G.......	....	....
NBC03	....	.....-T...	....	....	....	......	....	A.........	....	....	......A...	....
NBC20	....	.....-T...	....	...T......	....	......	....	....	....	..G.......	....	....
NBC21	....	.....-T...	....	....	....	......	....	A.........	....	..G.......	....	....
NBC22	....	.....-T...	....	....	....	....C.	....	A.........	.G........	.......A..	....	....
CBC03	....	.....-T...	..T......G	....	....	......	....	....	....	....C.....	......A...	....
CBC07	....	.....-TT..	....	....	....	......	....	A.........	...GT.....	....	........T.	....
CBC13	....	.....-TTG.	....	....	....	......	....	A.........	....T.....	....	........T.	....
CAB01	C	.....-T...	....	....	....	......	....	A....C....	....	....	....	....A....
CAB02	....	.....-T...	....	....	....	C.....	....	A.........	....	....	....	....
CAB10	....	.....-TT..	....	...T......	....	......	....	A.........	....T.....	....	........T.	...C.....
CAB11	....	A....-T...	....	.T........	....	C.....	....	....	....	....	....	....
SAB02	....	.....-T...	.....AG...	....	....	......	....	A.........	....	....	....	....

	TCCACGGTCT	GAGAC-CCAA	ACCCCGATCA	CCACGGTCGC	AATCCCACGT	TTTATC	TTCA	GTGCATAATT	AAGACAATCA	TAACTTTCAC	CCTTGAGGCC	TGATGCCCA
SAB03	....	.....-T...	....	....	....	......	....	A.........	....	....	....	.....T...
SAB04	....	.....-T...	....	....	....	......	....	A.........	....	....	....	.......T.
SAB05	....	.G...-T...	....	....	....	......	....	A.........	....	....	....	.....T...
SK001	.......C..	.....-T...	...T......	....	....	C.....	....	....	....	....	.......A..	....
SK003	....	A....-T...	....	....	....	......	....	....	....	....	....	....
SK004	....	A....-T...	....	....	....	......	....	..A.......	....	....	....	....
SK005	......C...	.....-T...	....	....	....	C.....	....	....	....	....	.......A..	....
SK013	....	.....-T...	....	....	....	......	....	....	.....G....	....	....	....
NON02	....	.....-T...	....	....	....	C.....	....	....	.........G	........G.	....	....
NON03	....	.....-T...	....	....	....	......	..T.	A.........	....	...T......	.T........	......T..
NON06	....	.....-....	....	....	....	......	....	....G.....	....	....	....	....
NON07	....	...G.-T...	....	T........T	.....T....	.C....	....	....	....	....	....	....
NON08	....	.....-T...	....	....	....	C.....	....	....	..A.......	....	.......A..	....
NON09	....	.....-T...	....	.....A....	G..T......	......	....	....	....	....	....	....
NON10	....	.....-T..G	........T.	....	....	......	....	....	....	....	....	....
NON11	........T.	.....-T...	....	TT........	..C....T..	......	....	....	....	....	....	....
NON13	....T.....	.....-T...	....	.T........	....	C.....	....	....	....	....	....	....
NON15	....	.....-T...	....	....	....	......	....	....	....	........G.	....	.A.......
NQC02	....	...G.-T...	....	....	....	......	....	....	....	....	.......A..	....
NQC05	.T........	.....-T...	....	....	....	.....T	....	A.A.......	....	....	....	....
NQC07	....	.....-T...	...T......	....	....	C.....	...G	....	....	....	.......A..	....
NQC08	....	.G...-T...	....	....	....	C.....	....	A.........	....	....	....	....
NQC09	....	.....-T...	....	....	....	C.....	....	....	....	C.........	.......A..	....
NQC10	....	.....-T...	....	....	....T.....	C.....	....	....	....	....	.......A..	....
NSNB05	....	.....-T...	....	T....A....	....	......	....	....	....	....	....	....
NSNB10	..T.......	.....-T...	...T......	....	....	C.....	....	....	....	....	.......A..	....
NSNB11	....	....T-T...	....	T....A....	....	......	C...	.C........	....	....	....	....

	TCCACGGTCT	GAGAC-CCAA	ACCCCGATCA	CCACGGTCGC	AATCCCACGT	TTTATC	TTCA	GTGCATAATT	AAGACAATCA	TAACTTTCAC	CCTTGAGGCC	TGATGCCCA
NSNB12	....	.....-T...	....	....	........A.	C.....	....	....	....	....	.......A..	....
NSNB15	....	.....-T...	.T........	....	....	C.....	....	....	....	....	.......A..	....
NSNB16	....	A....-T...	....	....	....	C.....	....	....	....	....	.......A..	....
NSNB17	....	.....-T...	....	....	....	.....	....	....	....	.G........	....	....
LAB02	....	.....-T...	....	T.........	....	.....	....	....	....	....	....	....
LAB03	....	.....-T...	...T......	....	....	C.....	....	....	....	....	.......A..	C........
LAB05	....	.....-T...	....	....	....	C.....	....	....	....	....	.......A..	........G
LAB09	....	.....-T...	....	....	......G...	C.....	....	....	....	....	.......A..	..G......
LAB11	....	.....-T...	....	....	....	C.....	....	....	....	....	.....G.A..	....
LAB13	....	.....-T...	....	....	....	.....	....	A.........	........T.	....	..C.......	....
LAB14	....	A....-T...	...T......	....	....	C.....	....	....	....	C.........	.......A..	....
LAB16	....	.....-T...	....	T.........	....	.....	.C..	....	....	....	....	....
LAB17	....	...G.-T...	....	....	....	.....	....	....	....	....	....	....
LAB20	....	..A..-T...	....	....	....	.....	....	...T......	....	....	....	....
NY001	....	.....-T...	....	.....A....	G..T......	.....	....	........C.	....	....	....	....
NY006	C	.....-T...	....	....	....	C.C...	....	....	....	....	.......A..	....
NL004	....	.....-T...	....	....	....	C.....	....	A.........	....	....	......AA..	....
NL005	....	.....-T...	....	....	....	C.....	....	....	....	....	.......A.T	....A....
NL006	....	.....-T...	....	....A.....	....	C.....	....	....	....	....	.......A..	....
NL007	.T........	.....-T...	....	....	....	C.....	....	....	....	....	.......A..	....
NL010	....	.....-T...	....	...T......	....	C.....	....	....	....	....	.......A..	....
NL023	....	.....-T...	....	....	....	C..G..	....	C	....	....	.......A..	....
NL024	....	.....AT...	....	....	....	.....	....	.....C....	....	....	....	....
NL029	....	.....-T...	....	.T........	....	C.....	....	....	....	....	.......A..	....
NL031	....	.....-T...	.T.T......	....	....	C.....	....	....	....	....	.......A..	....
NL032	....	.....-T.G.	....	....	....	C.....	....	....	....	....	....A..A..	....
NL033	....	.....-T..G	....	....	....	C.....	....	....	....	....	.......A..	....

Appendix 4. The consensus sequence for the control region (domains and boxes based on Kvist (2000)) and the ATPase 6-8 coding region (regions based on Desjardins & Morais (1990)). The '*' represent variable sites and the start and stop codons are underlined. Conserved boxes in the control region are shaded.

Control Region

```
Domain I
TAATGTAATGTAGGATAATCCAAATAACACGCAACTTCACCTCCACCCTAACTCAAACA
                          *  *           *               *

TTTACGCCCAAGAGATAATGTTCGGGCAGTTCCACTTCTAGGCACATTCCTATTTCAGG
      *                   *  *       *

TACCATTTAGCCCAACTGATCCTACCTCAGGCCCGAGAGGCCGCAAGCGTCACTTGAGA
                       *    *       **      *  *    *

TCGGAAACTCCCCTC-GCGCTTCAAAACCCATCCGATATACGAGGAATTGTCCCAGTAA
                *  *    **** * *  ****

CTCTTTGCATTCACGAAGTCCATAGCACATAGCCCACCTCCTAAAACCCATGACTCTCC
 *                                *        *  ** *     *
Domain 2                                                F-box
AACAGCTTTCAGGAGCTCCCAAGCCAGAGGACCTGGTTATTTATTGATCGTACTTCTCA
           *                  *          *

                                       E-box
CGAGAACCGAGCTACCCCGTGTAAGTGCTACCTTAGGTTATTGGCTTCAAGGACTTAAA
                                    *   *

                                         D-box
CTCCCCCTAAACCCCGAGCGCGACTTGCTCTTTTGCGCTATTGGTTGTAACTTCAGGAC
                                                        *

                            C-box
CATAACTTGTTGAACTCCGTCTCCCTTGCTCTTCACAGATACAAGTGGTCGGTTGAATA
   **      *   *          *   *              *

                 B-box
CTCCTCATCTCTCTCTCGTAGTTGTCGGCATCCGACCGTCTCTTCACTTTTTTTTTTTT
  *                     *                          *    *   *

AGCGTAGTCTTCAATAAACCCTTCCAGTGCGTAGCAGGAGTTATCTTCCTCTTGACATG

TCCATCATATGACCGGCGAACTGTCGTTCCCCTACCACTCAgAATGTCATGGTCTCATT
       *           * *                                  *

GGATAAGGTCGGCTCCTAATTCAACCCTGATGCACTTTGACCACATTCGTTAAACCCG
```

ATPase 6-8

COII
CATCGTAGTAGAATCTGCCCCACTCGCCAACTTCGAACACTGATCTTCCCTATCATCC**TAA**

Lys
TCACTCATTAAGAAGCTATGAAACAGCACTAGCCTTTTAAGCTAGAGACAGAGGACCACCC

ATPase 8
GCCCCTCCTTAATGAA**ATG**CCCCAACTAAATCCCAACCCTTGATTTTTTATCATGCTCGCC

TCGTGACTTACCTTCTCCCTAATCATCCAACCCAAACTCCTATCATTCGTATCCACCAACC

ATPase6
CCCCCTCTAACAAAACCCCCTCAACCACACCCTCCACCCCCTGAACCTGACC**ATG**AACC**TA**

AGCTTCTTTGACCAATTCTCCAGCCCATCCTTACTAGGAATCCCCTTAATCCTTATCTCAA

TAACATTCCCAGCACTCCTACTACCCTCCCCCAGCAACCGATGAATCACCAACCGACTCTC

AACCCTCCAACTTTGATTCATTAACCTCATCACTAAACAACTAATAATTACATTAGACAAA

AAAGGACACAAATGAGCCCTCATCCTAACATCCCTAATAATTTTCCTCCTACTAATTAACC

TACTAGGCCTACTACCCTACACATTCACCCCAACCACCCAACTATCCATAAACCTAGCCCT

GGCCTTTCCCCTATGACTTGCCACTCTCCTAACGGGCCTACGAAACCAACCCTCTGCCTCC

CTAGGGCACCTCCTACCAGAAGGCACCCCCACACCCCTAATCCCAGCCCTCATCCTAATCG

AAACAACCAGCCTCCTCATCCGCCCTCTCGCCCTAGGGGTACGACTCACAGCCAACCTCAC

AGCAGGCCACCTCCTCATCCAACTCATCTCCACAGCTACAATAGCCCTATCCTCAACAATA

CCAGCAGTATCACTCCTAACCCTCCTAGTCCTCTTCCTACTAACCATCCTAGAGGTAGCTG

TAGCTATC