PHYLOGEOGRAPHY AND GENETIC DIVERSITY IN RED KNOTS (*CALIDRIS CANUTUS*)

by

Deborah M. Buehler

A thesis submitted in conformity with the requirements

for the degree of Master of Science

Graduate Department of Zoology

University of Toronto

ABSTRACT

Phylogeography and genetic diversity in Red Knots (*Calidris canutus*). M.Sc. thesis. 2003. Deborah M. Buehler. Graduate Department of Zoology. University of Toronto

This thesis examines phylogeography and genetic diversity in Red Knots (*Calidris canutus*). I sequenced and characterized the complete control region of knots and verified that the sequence was mitochondrial in origin. I then used control region sequences and coalescent analysis to reconstruct the demographic history and biogeography of knots. My analysis indicated that knots probably expanded eastwards across the arctic from a severely bottlenecked population in eastern Eurasia as Pleistocene ice sheets melted. Population divergence times suggested that all six subspecies of knots arose within the last 20,000 years or so, and evolutionary effective population sizes of females were small (N_{ef} = 2,000 – 14,000). To examine genetic diversity, I analyzed control region sequences, amplified fragment length polymorphisms and microsatellites, which indicated that knots are genetically depauperate relative to Dunlins and other bird species. I likened knots to an avian cheetah, with little genetic variation throughout a global range.

ACKNOWLEGEMENTS

This thesis would not have been possible without the help and support of many people. Oliver Haddrath and Mark Peck provided excellent technical support and insight in the lab and the field. Mark Engstrom gave many helpful suggestions as the second member of my committee. Theunis Piersma and Patricia González provided encouragement and a wealth of information on knots. The members of the Ornithology Lab at the Royal Ontario Museum: Maryann Burbidge, Nicola Chong, Andy Given, Annette Greenslade, Tara Paton, Sergio Perreira and Gustavo Ybazeta gave hours of intellectual conversation and help with troubleshooting of all sorts. Cathy Ayley, Sue Chopra, Chris Pankewycz and Josie Valotta offered plenty of help with the administrative trials and tribulations of graduate school. My parents, Norma and Willy Buehler, and my friends Tamara Brodey, Meri Perra and Kristin VanRooijen provided nourishment, encouragement and understanding through many late nights and weekends at the lab. They also kept me afloat during emotional battles with immigration at the beginning of this endeavor, supporting me the entire time. My supervisor Allan Baker was remarkable, offering not only brilliant academic guidance, but also almost parental understanding and support from start to finish. He believed in me, encouraged me, literally would not let me quit despite very trying personal circumstances at the start of this thesis. For that I will always be grateful. Finally, I thank my husband and soul mate Alberto Castillo, who supported all parts of this endeavor, even if it meant umpteen visa applications and several plane tickets to come visit me in Canada. He provided love and support both in Canada and long distance from his native Panama when circumstance

kept us apart. This thesis could not have been completed without his unconditional love and support.

An NSERC PGS A Scholarship (PGSA-231177-2000) to myself, and grants from NSERC (200-02) and the Royal Ontario Museum Foundation to Allan Baker financially supported this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

General Introduction

In this thesis I examine phylogeography and genetic diversity in Red Knots (*Calidris canutus*). This introduction provides relevant background information for the research described in subsequent chapters. I begin with a general description of the study species, and then give a brief history of developments in molecular ecology with relation to birds, shorebirds and knots. I conclude with a description of the molecular markers and theory used in this research and lay out how these concepts are distributed into chapters.

Red Knots

Red Knots are long distance migrant shorebirds. They breed in the high arctic and each year embark on migrations of up to 15,000 km to marine staging and wintering areas worldwide. Their spectacular migrations make them the long distance athletes of the ornithological world, and their population biology and physiology in relation to these migrations have long fascinated ornithologists.

Breeding plumage in knots varies among the six subspecies but in general the species has russet red under parts and a gray back spangled with red and black chevrons. This plumage allows birds to blend with the reds, browns and grays of the arctic spring tundra. During the non-breeding season knots are soft gray in color with their behavior on wintering grounds making up for their dull appearance. Away from breeding areas

knots congregate in huge flocks on intertidal flats rich in mollusk prey, and these flocks are some of the largest and densest of all wader species (Davidson and Piersma 1992).

Knots are globally distributed into six subspecies, each with different morphologies, migratory routes and annual cycles (Figure 1-1). *C.c. roselaari*, the least studied subspecies, is thought to breed in northwest Alaska and Wrangel Island and winters in the southeast United States; *C.c. rufa* breeds in the central Canadian arctic and winters in southern Patagonia and Tierra del Fuego; *C.c. rogersi* breeds on the Chukotsky Peninsula in eastern Russia and winters in south east Australia and New Zealand; *C.c. piersmai* breeds on the New Siberian Islands in north central Russia and winters in northwest Australia; *C.c. islandica* breeds in northern Greenland and northeast Canada and winters in north west Europe; finally, *C.c. canutus* breeds on the Taymyr Peninsula in western Siberia and winters in west and south west Africa (Tomkovich 1992; Tomkovich 2001). This global distribution makes knots an ideal species for the study of population structure and genetic diversity on a worldwide scale.

A Brief History of Molecular Ecology in Birds

Information on the genetic structure and diversity of bird species can result in valuable management guidelines for conservation. For example, species with highly differentiated populations and low levels of gene flow should not be managed as single units because mixing of genetically differentiated populations would cause introgression and the dilution of adaptations. In addition, species that are fragmented into genetically distinct populations have lower effective population sizes than panmictic species,

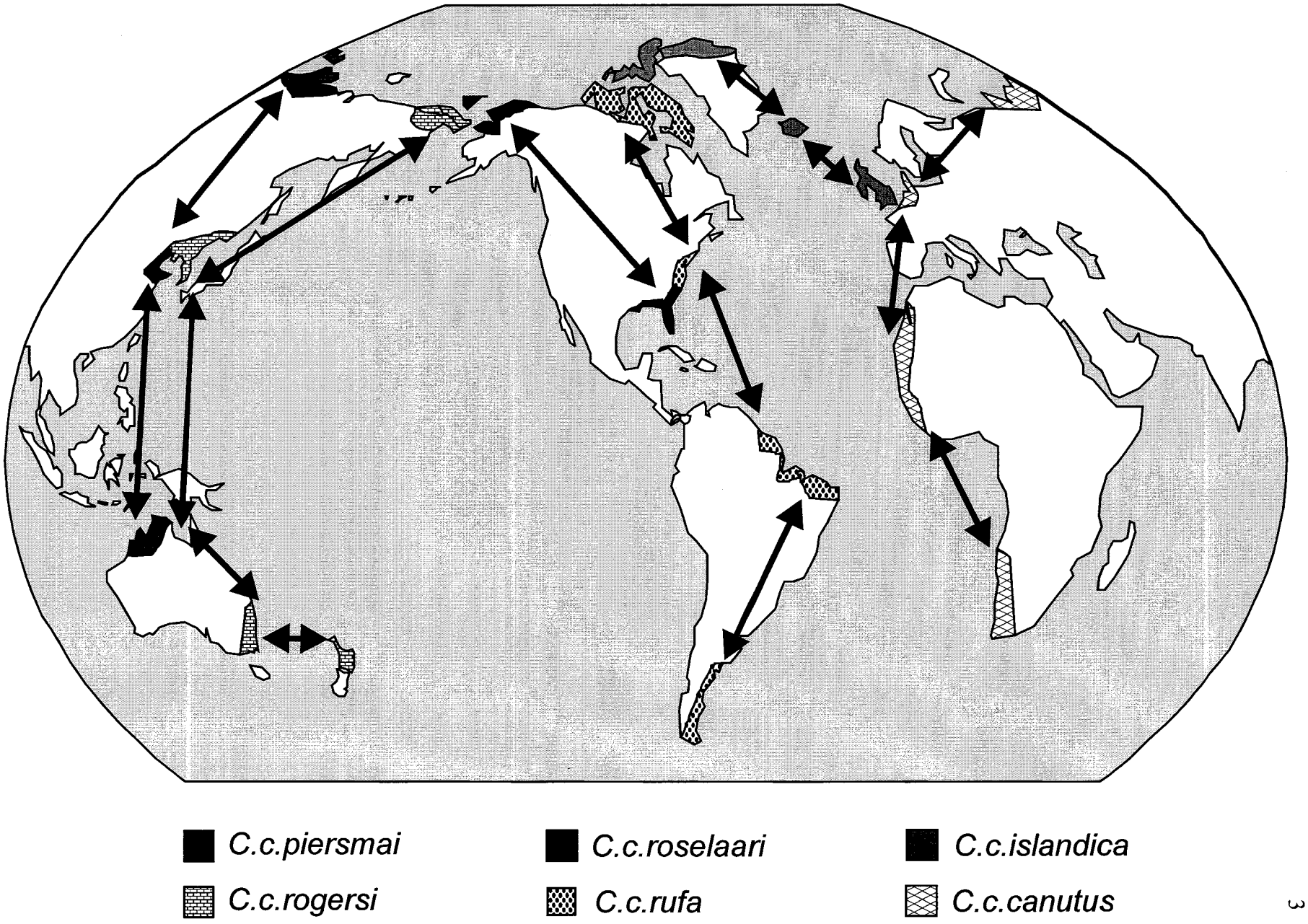


Figure 1-1: Map showing the global distribution of the six knot subspecies.

making them a higher conservation priority. Likewise, species with very little genetic diversity are vulnerable to changes in environmental conditions such as disease, global warming, or habitat loss (Webster et al. 2002). Conservation measures to protect the specific environments to which these species are adapted are important as they lack the genetic diversity necessary for evolutionary responses to change.

In comparison with other vertebrate species, birds present a puzzle for population geneticists. On the one hand, many species show high fidelity to their breeding grounds implying constraints on gene flow. Furthermore, regional populations of many species, including knots, show phenotypic differences in plumage and morphometrics. On the other hand, birds have an amazing potential for dispersal by virtue of their ability to fly and their tendency to migrate (Avise 2000). Furthermore, recent range expansions as glacial ice receded in northern temperate zones may cause a lack of genetic differentiation due to rapid and recent population growth (Baker et al. 1994; Milá et al. 2000).

Early work on the genetics of bird populations using allozyme allele frequencies in temperate regions showed that conspecific avian populations were far less structured than were conspecific populations of freshwater fish, small mammals, reptiles and amphibians (Avise 1983). More recently, studies on mitochondrial DNA (mtDNA) have demonstrated a variety of structural patterns. Some species such as the Red-winged Blackbird (*Agelaius phoeniceus*), the Short-tailed Shearwater, *Puffinus tenuirostris* (Austin et al. 1994), the Northern Pintail Duck, *Anas acuta* (Cronin et al. 1996), the Greenfinch *Carduelis chloris* (Merilä et al. 1997), the Bluethroat, *Luscinia svecica*

(Questiau et al. 1998) and the Swamp Sparrow, *Melospiza georgiana* (Greenberg et al. 1998) have common haplotypes distributed over moderate to large geographic areas. In contrast, other bird species show marked mtDNA phylogeographic structure at varying geographic scales. These species include the Seaside and Sharp-tailed Sparrow *Ammodramus maritimus* and *A. caudacutus* (Avise and Nelson 1989; Rising and Avise 1993), the Yellow Warbler, *Dendroica petechia* (Klein and Brown 1995), the Common Guillemot, *Uria aalge* (Friesen et al. 1996), the Common Chaffinch, *Fringilla coelebs* (Marshall and Baker 1997), the LeConte's Thrasher, *Toxostoma lecontei* (Zink et al. 1997) and the Adelaide's Warbler, *Dendrocia adelaidaea* (Lovette et al. 1998). The most striking example of an avian species with phylogeographic structure is found in New Zealand's flightless Brown Kiwi, *Apterx australis*. In this species virtually every population possesses private mtDNA haplotypes, with birds from the southern part of the South Island so divergent in mtDNA sequence and other genetic characters that they warrant species recognition (Baker et al. 1995).

Although mtDNA has proved to be the workhorse of phylogeographic studies in birds, dominant marker approaches using randomly amplified polymorphic DNA (RAPDs) and amplified fragment length polymorphisms (AFLPs) have also been useful in recent years. These methods target arbitrary DNA fragments in the total genome and require relatively little adjustment to new study species. By varying the primers used, these methods offer an almost infinite number of potential markers (Webster et al. 2002). RAPDs have been used to distinguish different species of shorebirds and were able to uncover population specific markers in two species (Haig et al. 1997). Population

differentiation has also been uncovered using AFLPs in southwestern Willow Flycatchers (Busch et al. 2000) and between the Common Chiffchaff *Phylloscopus collybita collybita* and the Iberian Chiffchaff *Phylloscopus c. brehmii* (Bensch et al. 2002). Although the avian genome appears to be relatively deficient in microsatellites (Neff and Gross 2001; Primmer et al. 1997) these markers have also been used to uncover weak population structure and information on gene flow in, for example, Yellow Warblers, *Dendroica petechia* (Gibbs et al. 2000).

In comparison to other avian species, shorebirds had very little genetic variation when examined with allozymes (Baker et al. 1985; Baker and Strauch 1988). More recent studies, however, have found differing levels of variation and structure within the shorebird group. MtDNA work on the control region of Dunlins for example, uncovered high levels of genetic diversity and clear phylogeographic structure with five major groups (Wenink and Baker 1996; Wenink et al. 1993). Genetic markers can now be used to distinguish the breeding origin of individual Dunlins in staging and wintering areas, and are helping to answer questions regarding large-scale migration patterns (Wenink and Baker 1996; Wennerberg 2001). In contrast, work using mtDNA on Curlew Sandpipers, White-rumped Sandpipers and Turnstones found very little differentiation among sampling sites (Wenink et al. 1994; Wennerberg and Burke 2001; Wennerberg et al. 2002).

Red Knots have presented a phylogeographic puzzle for years. Although they have morphologically different races, and banding studies show that the wintering grounds of these races are distinct (A. Baker and T. Piersma pers. comm.), little is known

about genetic differences among subspecies. Early work on knots using allozymes and mtDNA restriction endonucleases revealed small-scale divergence between subspecies and showed that knots are one of the oldest lineages in the Calidrine sandpipers (genus *Calidris*) (Baker 1992; Borowik and McLennan 1999). More recent work on the hyper-variable control region of mtDNA indicated very low variability, both in a low number of haplotypes and in low sequence divergence. The low genetic variability on a global scale suggests that the present knot subspecies have recently evolved from a severely bottlenecked population (Baker et al. 1994).

Background Information on the Molecular Markers and Theory used in this Research

In my thesis research I use three markers to examine genetic structure and diversity in Red Knots. These markers are mitochondrial control region sequences, AFLPs and microsatellites.

The mitochondrial genome consists of a circular DNA molecule within the mitochondria of cells. It is a streamlined molecule which codes for two ribosomal RNA's, 22 transfer RNA's and 13 polypeptides. MtDNA has many advantages as a marker for phylogenetic analyses (Moore 1995), and is one of the most frequently used markers in avian molecular systematics because of its maternal inheritance, haploidy and rapid rate of evolution (Baker and Marshall 1997). For intraspecific investigations, the control region is particularly useful as it evolves three to ten times faster than the average for the whole molecule (Brown et al. 1979) and, in most cases, is the most variable part

of the molecule (e.g. Baker and Marshall 1997; Greenberg et al. 1983, but see Kvist et al. 1999; Ruokonen and Kvist 2002). These characteristics make the control region highly suitable for the study of population structure within a species (Wenink et al. 1993).

One drawback to the use of control region sequences for population analysis is the existence of numts (nuclear homologs of mitochondrial DNA; Sorenson and Quinn 1998). The streamlined nature of the mitochondrial genome is a product of the gradual transfer through time of information from the organelle to the nucleus (Perna and Kocher 1996); thus copies of mtDNA sequence are transferred to the nucleus creating nuclear pseudogenes. Because recent nuclear transfers retain high sequence identity with mtDNA, these pseudogenes may be inadvertently amplified during PCR targeting the mitochondrial genome. The existence of numts requires that researchers take precautions against the amplification of nuclear pseudogenes when using mtDNA for systematic or population genetics research.

In Chapter 2 of this thesis I address the problem of numts in knots. I use long template sequencing to obtain and characterize the complete mitochondrial control region. This sequence aligns well with the control regions of other Charadriiformes and possesses the expected conserved sequence blocks, and functional elements of a mitochondrial copy. The flanking tRNA Glutamate (Glu) and Phenylalanine (Phe) show correct secondary structure and a relative rate test indicates no significant difference between substitution rates in the sequence I obtained versus the known mitochondrial sequence of turnstones. As a final verification, sequence obtained from purified mtDNA was identical to sequence obtained using total genomic DNA. These characteristics

verify that the sequence presented is mitochondrial in origin, and I use the protocol as well as the sequence itself as a reference for further sequences which I use to examine genetic structure and diversity in Chapters 3 and 4.

Chapter 3 of this thesis is dedicated to the examination of knot genetic structure and historical demography using control region sequences. In addition, patterns of historical demography in knots are contrasted with those of well-differentiated Dunlins. This provides a unique opportunity to examine genetic evolution at two time intervals, the first after 20,000 years in knots and the second after nearly 200,000 years in Dunlins. My examination of control region sequences is based on a statistical framework known as coalescent theory, described below.

Coalescent analysis arose when modern techniques for DNA sequencing gave researchers access to measures of genetic variation at the nucleotide level (see Rosenberg and Nordborg 2002 for a review). This greater variation allowed better resolution at the tips of evolutionary trees – the intraspecific level. Traditional statistical methods however, were not appropriate for dealing with sequence variation within species. Traditional summary statistics such as average heterozygosity and F_{ST} traditionally used in population genetics ignored much of the new information; namely how sequences were related historically. On the other hand, sequence based tree building methods borrowed from interspecific phylogenetics did not account for uncertainties caused by the randomness of gene evolution at the intraspecific level.

Trees produced from the sequence variation of a particular gene are referred to as gene trees. The structure of these trees tells the evolutionary history of the haplotypes or

alleles sampled. At the interspecific level where species form the taxonomic units of trees, gene trees very closely approximate species trees. This is because long time intervals between species branching events far exceed gene lineage branching events within species. Any haplotypes or alleles sampled from one species will be more closely related to haplotypes or alleles within that species than haplotypes or alleles found in a different species. In other words ancestral haplotypes or alleles shared between species have been pruned away over time and species have reached equilibrium between mutation and genetic drift. At the intraspecific level this is rarely the case and shared ancestral polymorphisms create gene trees that do not match population trees. The root of the problem is that evolution of neutral loci is a stochastic process. Within a species many gene trees exist, and each may depict a different evolutionary history. Therefore, to use sequence data at the intraspecific level where populations are not likely to be in equilibrium, models that consider the existence of many gene trees are needed. Coalescent theory allows the development of such models.

The basic premise behind coalescent theory is that, without selection, lineages in any sample will sort randomly into parental lineages looking backward in time (Kingman 1982). Whenever two lineages sort into the same ancestral lineage they are said to coalesce. When all lineages coalesce to a single ancestral lineage, the most recent common ancestor (MRCA) has been reached. The time needed to reach this ancestor depends on the size of the population (i.e. more ancestral lineages slow time to MRCA) and the number of lineages stemming from each parent lineage (i.e. characters with high mutation rates that generate many lineages speed time to MRCA). On average, the time

needed to reach the MRCA for nuclear, diploid, bi-parentally passed genes is $4N_e$ generations (Page and Holmes 1998), where N_e is the effective population size of the study population. Because mitochondrial DNA is haploid and maternally inherited, the time to MRCA using mtDNA is one quarter that for nuclear genes (N_e generations).

Coalescent theory is valuable for the examination of sequences in a historical framework; however, in this research I also wished to examine the present status of knots in terms of genetic diversity. To examine genetic diversity across the genome, markers from the nuclear genome in addition to control region sequences were necessary. Chapter IV describes research using control region sequences, AFLPs and microsatellites as a genome wide survey of knot genetic diversity.

AFLPs are a relatively new molecular technique first developed and used in plants (Vos et al. 1995). AFLPs are similar to restriction fragment length polymorphisms (RFLPs) in their ability to detect changes at restriction sites. However, they are also sensitive to nucleotide substitutions at selective positions determined by random base addition to pre-selective and selective PCR primers. Thus their mutation rate is more comparable to randomly amplified polymorphic DNA's (RAPDs) than to RFLPs (Busch et al. 2000). AFLPs are highly suited to genetic surveys of birds for several reasons. First, AFLPs do not require a previous knowledge of the avian genome of interest, nor do they require a large amount of development time. Second, a large number of markers can be quickly generated and analyzed using the AFLP method. Third, AFLPs are a PCR-based approach and thus require only a small amount of DNA as a template source (Busch et al. 2000). One drawback to the method is that AFLPs generate dominant

markers and allele frequencies must therefore be inferred indirectly assuming Hardy-Weinberg equilibrium. Nevertheless, these markers provide a genome wide survey and have the potential to uncover population-specific loci that can be used to identify the breeding origin of birds on migratory flyways and wintering areas.

The technique is based on a double amplification via PCR of a subset of restriction fragments from a total digest of genomic DNA. The DNA is cut with two restriction enzymes (a rare cutter and a frequent cutter). Each fragment is then ligated to adaptors that serve as the binding site for primers with one selective nucleotide added at the 3' end. The selected fragments are amplified in a pre-selective PCR. This PCR product is then used as a template for a selective amplification using primers with three additional nucleotides at the 3' end. Restriction enzymes applied to genomic DNA produce a banding pattern too dense to be scored. The purpose of the selective amplifications in the AFLP method is thus two fold, first selective amplifications decrease the number of bands on gels, resulting in a multi-locus fingerprint-like pattern, and second, these amplification make AFLPs sensitive to random nucleotide substitutions outside of cut sites. A schematic representation of AFLP methodology is found in Figure 1-2.

AFLPs have been used in bird species for a number of research purposes including the determination of extra-pair paternity in Bluethroats (Questiau et al. 1999), the determination of genetic variation in endangered Southwestern Willow Flycatchers

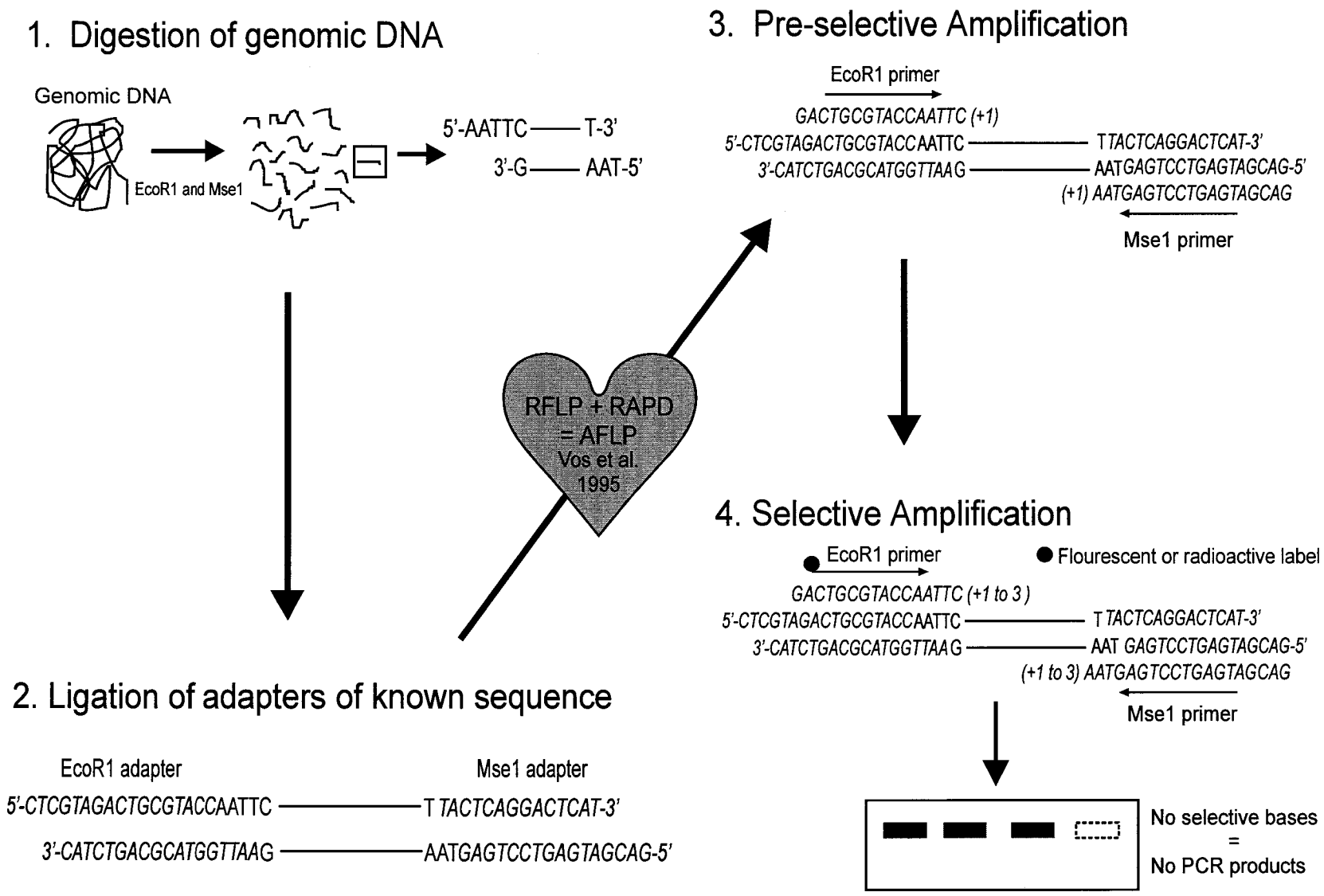


Figure 1-2: A schematic overview of the AFLP protocol used in this research (C. Ritland see GDC website http://www.forestry.ubc.ca/gdc/). The DNA is cut with two restriction enzymes. Each fragment is then ligated to adaptors that serve as the binding site for primers with one selective nucleotide added at the 3' end. The selected fragments are amplified in a pre-selective PCR. This PCR product is then used as a template for a selective amplification using primers with three additional nucleotides at the 3' end. The result is a multi-locus, fingerprint-like pattern.

(Busch et al. 2000), and the identification of hybrids in the zone of contact between two subspecies of chiffchaff (Bensch et al. 2002). In my research I use AFLPs as a novel approach to the study of the nuclear genome of knots.

Microsatellites are short segments of nuclear DNA with a specific sequence of 1-6 bases repeated up to a maximum of 60 times. Mutation rates at microsatellite loci have been estimated to range between 10^{-3} and 10^{-5} mutations per locus per generation (Edwards et al. 1992; Hearne et al. 1992; Weissenbach et al. 1992) and these high mutation rates are probably caused by the addition or subtraction of perfect repeats (Goldstein and Pollock 1997; Neff and Gross 2001). Microsatellites are co-dominant, fast evolving and because they are diploid and bi-parentally inherited they offer a larger effective population size than mtDNA, making them a convenient marker for studies in population structure (Goldstein and Pollock 1997).

Although they are considered a powerful molecular tool, microsatellites have never been examined in knots. I developed microsatellite loci using a genomic DNA library enriched for microsatellites with universal linker and Strepavidin-Biotin probing (Hamilton et al. 1999). In this procedure genomic DNA is isolated and then digested with restriction enzymes to create a genomic library of fragments 400-1000 bp in length. The library is enriched for GT and CT repeats using biotinylated oligonucleotides $(CT)_{15}$ and $(GT)_{15}$ and ligated into vectors. Vectors are transformed into bacteria for cloning and positive colonies are selected and sequenced (Figure 1-3).

To summarize, my research is an attempt to clarify knot population genetics using a multifaceted approach. The thesis is broken up into five chapters: the first is this

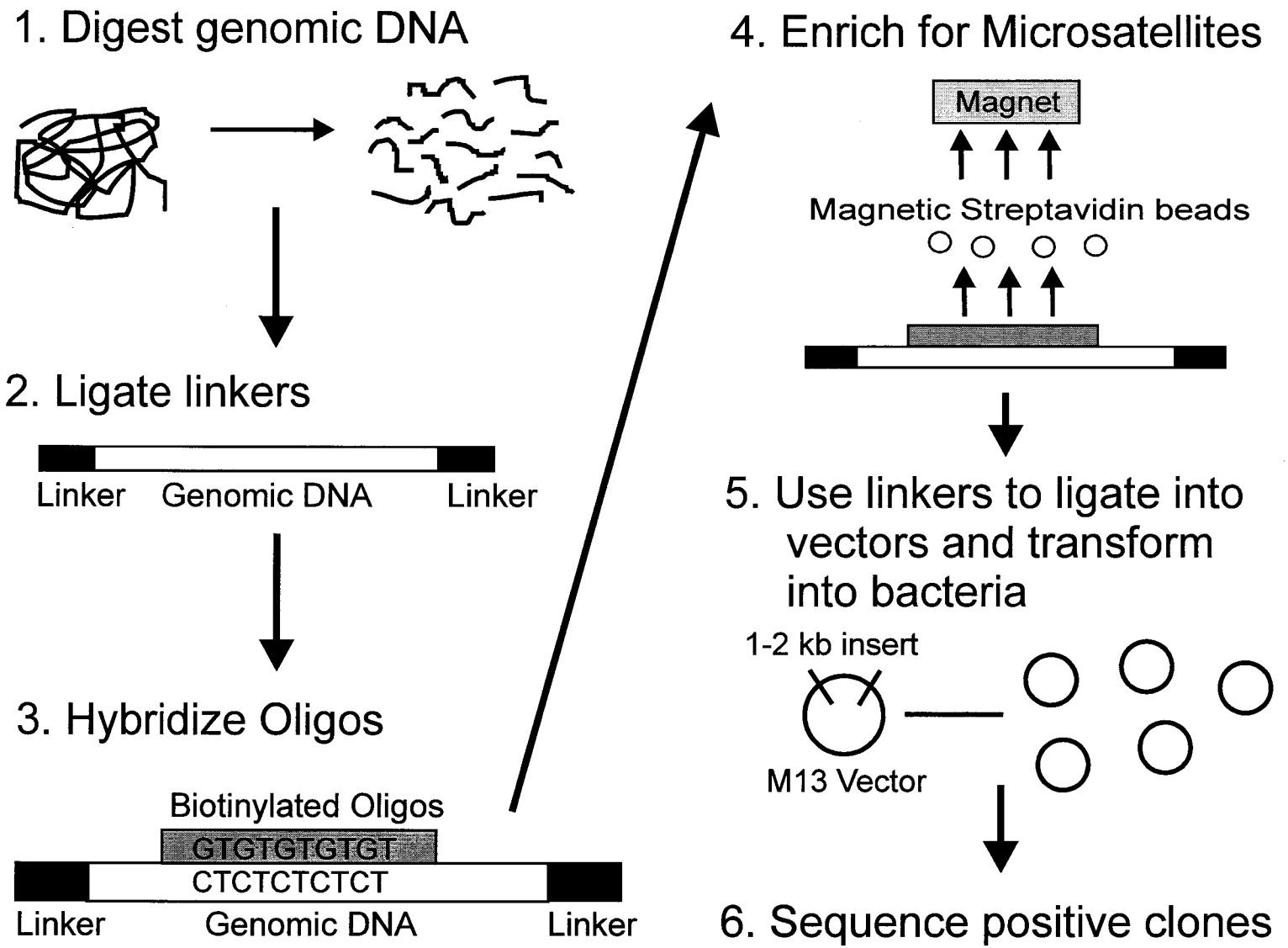


Figure 1-3: A schematic overview of the microsatellite enrichment protocol used to develop microsatellite loci in this research (Hamilton et al. 1999). Genomic DNA is digested with restriction enzymes to create a genomic library of fragments 400-1000 bp in length. Linkers are attached to the fragments and biotinylated oligonucleotides $(CT)_{15}$ and $(GT)_{15}$ in conjunction with streptavidin beads and a magnet are used to enrich the library for microsatellite repeats. Fragments are then ligated into vectors, and transformed into bacteria for cloning. Positive colonies are picked and sequenced.

introduction, the second is an examination of the complete knot control region, the third is an examination of population structure and historical demography, the fourth is an examination of genetic diversity across the genome and the fifth is a synthesis of my findings. This research strives to answer the questions (1) What is the population structure and historical demography of knots? (2) Can the biogeographic history and phylogenetic relationships of currently recognized subspecies be reconstructed? (3) What is the genetic diversity of knots and how does this diversity compare to that in other bird species?

References

Austin, J. J., R. W. G. White, and J. R. Ovenden. 1994. Population genetic structure of philopatric, colonially nesting seabird, the short-tailed shearwater (*Puffinus tenuirostris*). Auk 111:70-70.

Avise, J. C. 1983. Commentary. Pp. 262-270 *in* A. H. Brush and G. A. Clark, eds. Perspectives in Ornithology. Cambridge University Press, New York.

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

Avise, J. C., and W. S. Nelson. 1989. Molecular genetic relationships of the extinct dusky seaside sparrow. Science 243:646-648.

Baker, A. J. 1992. Molecular genetics of *Calidris* with special reference to knots. Wader Study Group Bulletin 64 Supplement:64-70.

Baker, A. J., C. H. Daughtery, R. Colbourne, and J. L. McLennan. 1995. Flightless brown kiwis of New Zealand posses extremely subdivided population strucutre and cryptic species like small mammals. Proceedings of the National Academy of Sciences of the U.S.A. 92:8254-8258.

Baker, A. J., A. M. Lynch, and C. E. Edwards. 1985. Biochemical genetics studies of shorebirds: methods and applications. Wader Study Group Bulletin 44:34-39.

Baker, A. J., and H. D. Marshall. 1997. Mitochondrial control region sequences as tools for understanding evolution. Pp. 51-82 *in* D. P. Mindell, ed. Avian Molecular Evolution and Systematics. Academic Press, San Diego.

Baker, A. J., T. Piersma, and L. Rosenmeier. 1994. Unraveling the intraspecific phylogeography of knots (*Calidris canutus*) - A progress report on the search for genetic markers. Journal fur Ornithologie 135:599-608.

Baker, A. J., and J. G. Strauch. 1988. Genetic variation and differentiation in shorebirds. Pp. 1639-1645 *in* H. Ouellet, ed. Acta XIX Congressus Internationalis Ornithologici, vol II. Ottawa University Press, Ottawa.

Bensch, S., A. J. Helbig, M. Salomon, and I. Seibold. 2002. Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. Molecular Ecology 11:473-481.

Borowik, O., and D. A. McLennan. 1999. Phylogenetic patterns of parental care in Calidrine sandpipers. Auk 116:1107-1117.

Brown, W. M., M. B. George, and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the U.S.A. 76:1967-1971.

Busch, J. D., M. P. Miller, E. H. Paxton, M. K. Sogge, and P. Keim. 2000. Genetic variation in the endangered southwestern willow flycatcher. Auk 117:586-595.

Cronin, M. A., J. B. Grand, D. Esler, D. V. Derksen, and K. T. Schriber. 1996. Breeding populations of northern pintails have similar mitchondrial DNA. Canadian Journal of Zoology 74:992-999.

Davidson, N. C., and T. Piersma. 1992. Introduction. Wader Study Group Bulletin 64 Supplement:9-13.

Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty. 1992. Genetic variations at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241-253.

Friesen, V. L., W. A. Montevecchi, A. J. Baker, R. T. Barret, and W. S. Davidson. 1996. Population differentiation and evolution in the common guillemont Uria aalge. Molecular Ecology 6:1047-1058.

Gibbs, H. L., R. J. G. Dawson, and K. A. Hobson. 2000. Limited differentiation in microsatellite DNA variation among northern populations of the yellow warbler: Evidence for male based gene flow? Molecular Ecology 9:2137-2147.

Goldstein, D. B., and D. D. Pollock. 1997. Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. Journal of Heredity 88:335-342.

Greenberg, B. D., J. E. Newbold, and A. Sygino. 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondral DNA. Gene 21:33-49.

Greenberg, R., P. J. Cordero, S. Droege, and R. C. Fleischer. 1998. Morphological adaptation with no mitochondrial DNA differentiation in the costal plain swamp sparrow. Auk 115:706-712.

Haig, S. M., C. L. Gratto-Trevor, T. D. Mullins, and M. A. Colwell. 1997. Population identification of western hemisphere shorebirds throughout the annual cycle. Molecular Ecology 6:413-427.

Hamilton, M. B., E. L. Pincus, A. DiFiore, and R. C. Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27:500: 502: 504-507.

Hearne, C. M., S. Ghosh, and J. A. Todd. 1992. Microsatellites for linkage analysis of genetic traits. Trends in Genetics 8

Kingman, J. F. C. 1982. On the geneaology of large populations. Journal of Applied Probability 19A: 27-43.

Klein, N. K., and W. M. Brown. 1995. Intraspecific molecular phylogeny in the yellow warbler (*Dendroica petechia*) and implications for avian biogeography in the West Indies. Evolution 48:1914-1932.

Kvist, L., M. Ruokonen, J. Lumme, and M. Orell. 1999. Different population structures in northern and southern populations of European blue tits (*Parus caeruleus*). Journal of Evolutionary Biology 12:798-805.

Lovette, I., J,, E. Bermingham, G. Seutin, and R. E. Ricklefs. 1998. Evolutionary differentiation in three endemic West Indian warblers. Auk 115:890-903.

Marshall, H. D., and A. J. Baker. 1997. Strucutral conservation and variation in the mitochondrial control region of Fringilline finches (*Fringilla spp.*) and the greenfinch (*Carduelis chloris*). Molecular Biology and Evolution 14:173-184.

Merilä, J., M. Björkland, and A. J. Baker. 1997. Historical demography and present day population structure of the greenfinch, *Carduelis chloris* - an analysis of mtDNA control-region sequences. Evolution 51:946-956.

Milá, B., D. J. Girman, M. Kimura, and T. B. Smith. 2000. Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. Proceedings: Biological Sciences (The Royal Society) 267:1033-1040.

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

Neff, B. D., and M. R. Gross. 2001. Microsatellite evolution in vertebrates: Inference from AC dinucleotide repeats. Evolution 55:1717-1733.

Page, R. D. M., and E. C. Holmes. 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell Science Ltd., Oxford.

Perna, N. T., and T. D. Kocher. 1996. Mitochondrial DNA: Molecular fossils in the nucleus. Current Biology 6:128-129.

Primmer, C. R., T. Raudsepp, B. P. Chowdhary, A. P. Moller, and H. Ellegren. 1997. Low frequency of microsatellites in the avian genome. Genome Research 7:471-482.

Questiau, S., M.-C. Eybert, A. R. Gaginskaya, L. Gielly, and P. Taberlet. 1998. Recent divergence between two morphologically differentiated subspecies of bluethroat (Aves: *Musicicapidae: Luscinia svecica*) inferred from mitochondrial DNA sequence variation. Molecular Ecology 7:239-245.

Questiau, S., M. C. Eybert, and P. Taberlet. 1999. Amplified fragment length polymorphism (AFLP) markers reveal extra-pair parentage in a bird species: the bluethroat (*Luscinia svecica*). Molecular Ecology 8

Rising, J. D., and J. C. Avise. 1993. An application of geneological concordance principles to the taxonomy and evolutionary history of the sharp tailed sparrow (*Ammodramus caudacutus*). Auk 110:844-856.

Rosenberg, N. A., and M. Nordborg. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nature Reviews: Genetics 3:380-390.

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

Sorenson, M. D., and T. W. Quinn. 1998. Numts: A challenge for avian systematics and population biology. Auk 115:214-221.

Tomkovich, P. S. 1992. An analysis of the geographic variability in knots *Calidris canutus* based on museum skins. Wader Study Group Bulletin 64 Supplement:17-23.

Tomkovich, P. S. 2001. A new subspecies of Red Knot *Calidris canutus* from the New Siberian Islands. Bulletin of the British Ornithologists Club 121:257-263.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and e. al. 1995. AFLP - A new technique for DNA fingerprinting. Nucleic Acid Research 23:4407-4414.

Webster, M. S., P. P. Marra, S. M. Haig, S. Bensch, and R. T. Holmes. 2002. Links between worlds: unraveling migratory connectivity. Trends in Ecology and Evolution 17:76-83.

Weissenbach, J., G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix, and M. Lathrop. 1992. A second-generation linkage map of the human genome. Nature (London) 359:794-801.

Wenink, P. W., and A. J. Baker. 1996. Mitochondrial DNA lineages in composite flocks of migratory and wintering dunlins (*Calidris alpina*). Auk 113:744-756.

Wenink, P. W., A. J. Baker, and M. G. L. Tilanus. 1993. Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpina*). Proceedings of the National Academy of Sciences of the U.S.A. 90:94-98.

Wenink, P. W., A. J. Baker, and M. G. L. Tilanus. 1994. Mitochondrial control region sequences in two shorebird species, the Turnstone and the Dunlin, and their utility in population genetic studies. Molecular Biology and Evolution 11:22-31.

Wennerberg, L. 2001. Breeding origin and migration pattern of dunlin (*Calidris alpina*) revealed by mitochondrial DNA analysis. Molecular Ecology 10:1111-1120.

Wennerberg, L., and T. Burke. 2001. Low genetic differentiation between Curlew Sandpiper (*Calidris ferruginea*) populations with highly divergent migratory directions shown by mitochondrial DNA and microsatellite analysis. Pp. 59-72. Genetic Variation and Migration of Waders. Lund University, PhD Thesis.

Wennerberg, L., M. Klaassen, and Å. Lindstrom. 2002. Geographic variation and population structure in the white-rumped sandpiper *Calidris fusicollis* as shown by morphology, mitochondrial DNA and carbon isotope ratios. Oecologia 131:380-390.

Zink, R. M., R. C. Blackwell, and O. Rojassoto. 1997. Species limits in the LeConte's trasher. Condor 99:132-138.

CHAPTER 2

Characterization of the mitochondrial control region of Red Knots (*Calidris canutus*)

Abstract

I sequenced and characterized the complete mitochondrial control regions of 11 Red Knots *Calidris canutus*. The control region is 1168 bp in length and is flanked by tRNA Glutamate (Glu) and the gene ND6 at its 5' end and tRNA Phenylalanine (Phe) and the gene 12S on its 3' end. The sequence aligns well with the control regions of other Charadriiformes as well as chicken, and possesses conserved sequence blocks F, E, D, C and the bird similarity box (BSB) in domain II as expected for a mitochondrial copy. The sequence contains a cytosine string and TAS elements in domain I, and a tetranucleotide microsatellite at its 3' end. Flanking tRNA Glu and tRNA Phe regions show correct secondary structure, and a relative rate test indicated no significant difference between substitution rates in the sequence I obtained versus the known mitochondrial sequence of turnstones. These characteristics indicate that the sequence is mitochondrial in origin. To confirm this I sequenced the control region of a single individual using both purified mitochondrial DNA and genomic DNA. Using both methods the sequences were identical, confirming that amplification product from genomic DNA produces mitochondrial and not nuclear pseudogene sequence. The sequence and methods presented in this paper may now serve as a reference for future studies using the knot control region. Furthermore, the discovery of an additional five

variable sites in 11 individuals, towards the 3' end of the control region, highlights the importance of this area as a source of variation for future studies in knots.

Introduction

The control region is a non-coding sequence of variable length found in the mitochondrial genome. In vertebrates this region contains promoters for transcription, the heavy strand replication origin (O_H), and the displacement loop (D-loop; Chang and Clayton 1986). Research based on the distribution of variable nucleotide positions and differential frequencies of the nucleotides shows that the mitochondrial control region is divided into three domains (Saccone et al. 1991). Most of the variability in the control region is concentrated in domains I and III. These domains show variation in length and an elevated number of nucleotide substitutions. The central domain II is conserved compared to the flanking I and III domains (Baker and Marshall 1997).

Mitochondrial DNA (mtDNA) has many advantages as a marker for phylogenetic analyses (Moore 1995) and is one of the most frequently used markers in avian molecular systematics because of its maternal inheritance, haploidy and rapid rate of evolution (Baker and Marshall 1997). For intraspecific investigations, the control region is particularly useful as it evolves three to ten times faster than the average for the whole molecule (Brown et al. 1979) and is thus a highly polymorphic region of mtDNA providing a rich source of variation to determine population structure within a species (Wenink et al. 1993).

One drawback to the use of control region sequences for population analysis is the existence of numts (nuclear mitochondrial DNA segment; Sorenson and Quinn 1998). The streamlined nature of the mitochondrial genome is a product of the gradual transfer of information from the organelle to the nucleus (Perna and Kocher 1996). Through this process copies of mtDNA sequence can be transferred to the nucleus creating nuclear pseudogenes. Because recent nuclear transfers retain high sequence identity with mtDNA, these pseudogenes may be inadvertently amplified during PCR targeting the mitochondrial genome. The existence of numts requires that researchers take precautions against the amplification of nuclear pseudogenes when using mtDNA for systematic or population genetics research.

Red Knots (*Calidris canutus*) are long distance migrant shorebirds. They breed in the high arctic tundra and each year embark on migrations of up to 15,000 km to marine staging and wintering areas at different latitudes throughout the world. Their population biology and physiology in relation to their circumpolar migrations have long fascinated researchers. The control region of Red Knots has been used in population genetics analysis by Baker et. al. (1994), who amplified and sequenced 255 bp of domain I for 25 knots from four subspecies. They found very low variability, both in number of haplotypes and in sequence divergence. This study has been cited as an example of a species with very little phylogeographic structure (Avise 2000; Baker and Marshall 1997; Wennerberg et al. 2002). Although precautions were taken by amplifying some sequences from purified mtDNA, the mitochondrial origin of sequences obtained from genomic DNA isolated from blood was never definitively confirmed.

The purpose of this paper is to sequence and characterize the complete control region of Red Knots in order to verify that the sequence is of mitochondrial origin. The verification is done through a series of precautions and tests including the amplification of a relatively long fragment for sequencing, the location of all expected structural components, the confirmation of proper secondary structure in surrounding tRNA's and a relative rate test of sequence evolution between the sequence and that of a known mitochondrial sequence in turnstones. To further check that the product amplified from genomic template produces the same sequence as that from purified mitochondrial DNA I use both methods on the same individual. Additionally, I examine diversity in a sample of 11 complete control regions to locate areas of high variability.

Materials and Methods

Red Knots were sampled using ground traps in breeding areas and cannon nets in wintering and staging areas globally. A few drops of blood were taken from each bird and stored in 50mM EDTA and 70% ethanol. For a single individual (MKP311), both total genomic DNA and purified mitochondrial DNA isolated from liver and heart tissue using a caesium chloride gradient method described in Van Wagner and Baker (1990) were used as template. All of the samples used are part of the Royal Ontario Museum collections. DNA was isolated using standard phenol-chloroform extractions (Sambrook et al. 1989).

The following protocol was used to obtain sequence from genomic DNA and purified mtDNA from the same individual (MKP 311) and genomic DNA alone in 10

other knots. A 2000 bp fragment encompassing the control region as well as the flanking tRNA Glutamate (Glu), tRNA Phenylalanine (Phe), and ND6 regions was amplified. Although numts longer than 2000 bp have been found in the human genome (Tourmen et al. 2002; Woischnik and Moraes 2002), very few numts (2.76%) were found to encompass the entire control region. (Woischnik and Moraes 2002). The initial long fragment was amplified using primers 12SH (5'- GGA TAA CAA TTT CAC ACA GGG TGA ACC TTC CGG TAC ACT TAC C –3') and LproT (5'- CAC GAC GTT GTA AAA CGA CGC TCC CAA AGC TGG TAT TTC –3') (T. Paton). Amplification was performed as specified by the manufacture using an Expand Long Template PCR Kit (Roche), and using the PCR profile: 94°C for a 2 minute denaturation; 10 cycles of 92°C for 30 seconds, 63°C for 30 seconds and 68°C for 12 minutes; 25 cycles of 92°C for 30 seconds, 63°C for 30 seconds and 68°C for 12 minutes + 20 seconds/cycle; 68°C for 7 minutes hold and 4°C soak. Nested amplifications were performed to obtain different sections using primers: L98, H401, L438 (Wenink et al. 1993), and H1537 (5'-TGA CCG CGG TGG CTG GCA CAA G –3') (O. Haddrath) and were carried out in a reaction volume of 25 µl containing 1.0 µl of DNA, 2.5 µl of 10X EH buffer, 0.5 units of Taq DNA polymerase (Qiagen), 5 mM dNTPs, and 2.5 pmoles of each primer. The amplification protocol was: 95°C for a 2 minute denaturation; 36 cycles of 95°C for 45 seconds, 50°C for 45 seconds and 72°C for 90 seconds; 72°C for 5 minutes hold and 4°C soak. Sequences were obtained using overlapping primers: H401, L98, L438, H772 (Wenink et al. 1993) and a newly designed primer KnotMidCRL (5'- GCA ACG GGT GAA TAC AAT CTA AGA C –3'). The presence of a cytosine string at the 5' end of

domain I and an imperfect tetranucleotide microsatellite at the 3' end of domain III made it necessary to obtain sequence from tRNA Glu and tRNA Phe with primers ND6Lend (5'– AAA CTY AAC AAC CAC CCA CA –3') (O. Haddrath) and H1537 respectively. All sequences were obtained using the Sanger dideoxy chain termination method (Sanger et al. 1977) and the Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit (USB) with P^{33} on a 6% polyacrylamide gel.

Knot sequences were aligned with Genebank control region sequences of Dunlins *Calidris alpina* (Wenink et al. 1993), Ruddy Turnstones *Arenaria interpres*, Blackish Oystercatchers *Haematopus ater* (Paton et al. 2002) and chickens *Gallus gallus* (Valverde et al. 1994) using ClustalW (Thompson et al. 1994), and final alignment adjustments were performed manually. The boundary of domains I and II was defined 20 bp in the 5' direction from the conserved F-box. In the alignment of knot, dunlin, turnstone, oystercatcher and chicken a high degree of sequence similarity is seen after this point (Figure 2-1). Domain III was defined beginning with conserved sequence block (CBS-1) (Ruokonen and Kvist 2002).

The proper folding structure of the tRNA Phe and tRNA Glu sequences was determined by mapping the sequences onto the model structure presented for chicken tRNA folding (Desjardins and Morais 1990). Finally, a relative rate test for sequence evolution was performed using the program Phyltest (Kumar 2001) between the control region sequences of knots and turnstones, with the oystercatcher sequence as an outgroup. This analysis is a two-cluster test of significant difference in branch lengths (number of substitutions per site) between the taxa of interest, using a two-tailed normal

deviate test (Takezaki et al. 1995). The turnstone sequence is of confirmed mitochondrial origin and was taken from Paton et. al. (2002) who used purified mitochondrial DNA from tissue and sequenced from 15 kb long templates.

Results and Discussion

Complete control region sequences were obtained from 11 individuals. Sequence obtained from genomic DNA was identical to that obtained from purified mtDNA in individual MKP311. The alignment of the control region of knots with dunlin, turnstone, oystercatcher and chicken control regions is shown in Figure 2-1. The control region is 1168 bp in length, slightly shorter than the turnstone (1172) and oystercatcher (1296) control regions, and is flanked by tRNA Glu and the gene ND6 at its 5' end, and tRNA Phe and the gene 12S on its 3' end. The arrangement of genes surrounding the control region matches that of other Charadriiformes and Galliformes, but differs from some Picidae, Cuculidae and Passeriformes suboscines (Mindell et al. 1998). Figure 2-2 shows a schematic representation of the control region of knots and its characteristics.

Base composition in the control region resembles that found in other avian control regions (Baker and Marshall 1997). There is a lack of L-strand guanines in all domains but it is most pronounced in domain III. Adenines and cytosines are most prevalent in domain I, cytosines and thymines in domain II, and adenines and cytosines in domain III (Table 2-1).

```
                                                                                                                C-string      [80]
C.canutus     ---------- ---------- ---------- ---------- ---------- ---------- -----CACCG CCCC----C-      10
C.alpina      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------      N/A
A.interpres   ---------- ---------- ---------- ---ATAAACC ACCAAGCCCC ATAAACATTC GCGGC.C..C ....----.C      43
H.ater        AACGCGTAGA CCATACATAA TTCAGCTTTA TGTATACCTA TCCATGCCAT GCTCCCCCTC ATATT....T .T..GTAC.C      80
G.gallus      ---------- ---------- ---------- AATTTTATTT TTTAACCTAA CTCCCCTACT AAGTGT...C ....----TT      46
                                                                                                                [160]
C.canutus     CCCCCGCGCC GC-ATGTAAT TTGGGCATTT TTTGCAA-CC AGT--CATGC GCTCTGCCAC ATATAATTGT CCCCCC-AT      84
C.alpina      ---------- ---------- ---------- -----TGT.. .A.--T...T .G.GC..TG. .....C.... ......CC..      43
A.interpres   .......... ..-....... .......... ....TGG-.T .AC--T...T A..GGTTTG. ...C.C.... ......CC..      119
H.ater        ....TA.C.. C.C.C.CG.G CC.--...G. .A..GGGT.T ..CAG....T A..T.ATT.. ..TA.TC.A. ATG..AC-..      157
G.gallus      T....AG.GG .G---...TA C.AT....AA .-C.T-GCAT .CATTT..AT A-----.... .....-..A. GGTA..G-G.      114
                                                                                                                [240]
C.canutus     ACTACATACC A-TCCATGTT CCAAATCCAT TA--ATATAC A-----ACCG GGCTATACCC TCCCCCGCCA CTCACCCTCC      156
C.alpina      ........... .-........ .......... ..--..T... .-----.... ........A. CT.T..A..C .......G..      115
A.interpres   ......A. T-........ ..TT...... ..--..T.T. .-----C... ........-. C..T..T..C .A.....A..      190
H.ater        .-......AA TA.......A .TG.G..-.. ..CT..T.TG .-----T.T. .CA..CC.-. C..T...A.G .AT.T.T-..      228
G.gallus      .A..T...T .-.AT....A .T...C.... ..-T..G..T .CGGGC.TTA AC.....TT. CA.ATTT.TC .CA.TGTC.A      192
                                                                                                                [320]
C.canutus     TCCAGAGAAC TACCGAGCCA ATGAACCTAG GAATATTCAC ACATTCTGTA CTAAACC--C ATC-----AA ACTGTTAGGA      229
C.alpina      .......GGT A.....AG.. .......... .......... ...CA..... .......--. ...-----.. CT........      188
A.interpres   .......... A......A.. .......... ....T..... ....CT.... .......--. ..-------.C CTCAC.....      262
H.ater        CA...G...T G.AATTTGTC ...G.TTGT. ......CTC. TT.AA.-... .....A.--. ...-----C. .G.......-      299
G.gallus      .T.TAT.C.T G.T.T..GAC ..-.CT.ATT T.CCC.C.C. .T.GA.A..T .C.....ACT ...AAGCC.C CTAAC..T..      271
                                                                                                                [400]
C.canutus     -TTAT-ACAT ACAA-ACTCT C-AAAAT-GT ACGGCAGTGC TTGAATA--- CAC---ACCA TGAATGGTCT AAGTACAAAC      298
C.alpina      -....-.... TA..-..... .T...GC-.. .......... ..T..C.--- ...---G... ...T....T. ....G..G..      258
A.interpres   -.....G... .G..-....C TCG.TT.A.C .......... C.T..C.--- ...---..T. ..CC.A.... ...GT..T..      333
H.ater        -....-.... .ACTA.GCT. T--.-..GA. .......... ......G--- ...---..T. ...T...A.. CC.CC..T-T      367
G.gallus      A.GG.T...G GAC.T.AATC TC.CTC.--C .T.TTCTCC. CCC..C.AGT ...CTA..T. ........TA C..G...T..      349
```

Figure 2-1: L-strand (5' – 3') control region sequences for four Charadriiformes and chicken. Alignment generated by CLUSTAL W (Thompson *et al.* 1994). Domain I C-string, domain II T-string and domain III CAAA-microsatellite are bold/italic, putative TAS elements and transcription promoters are underlined, conserved sequence blocks (F, E, D, C, BSB and CSB-1) are highlighted.

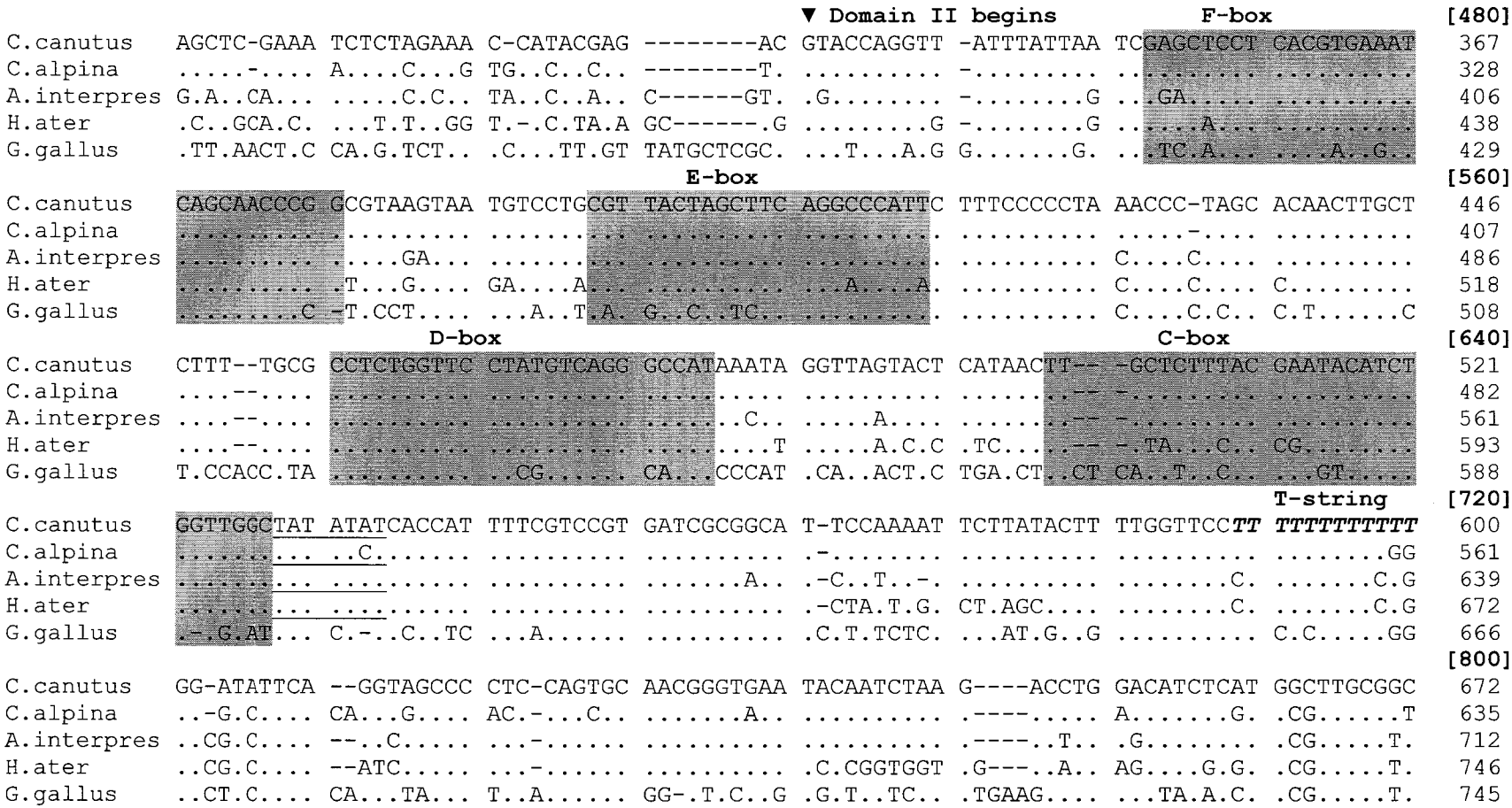


Figure 2-1 continued

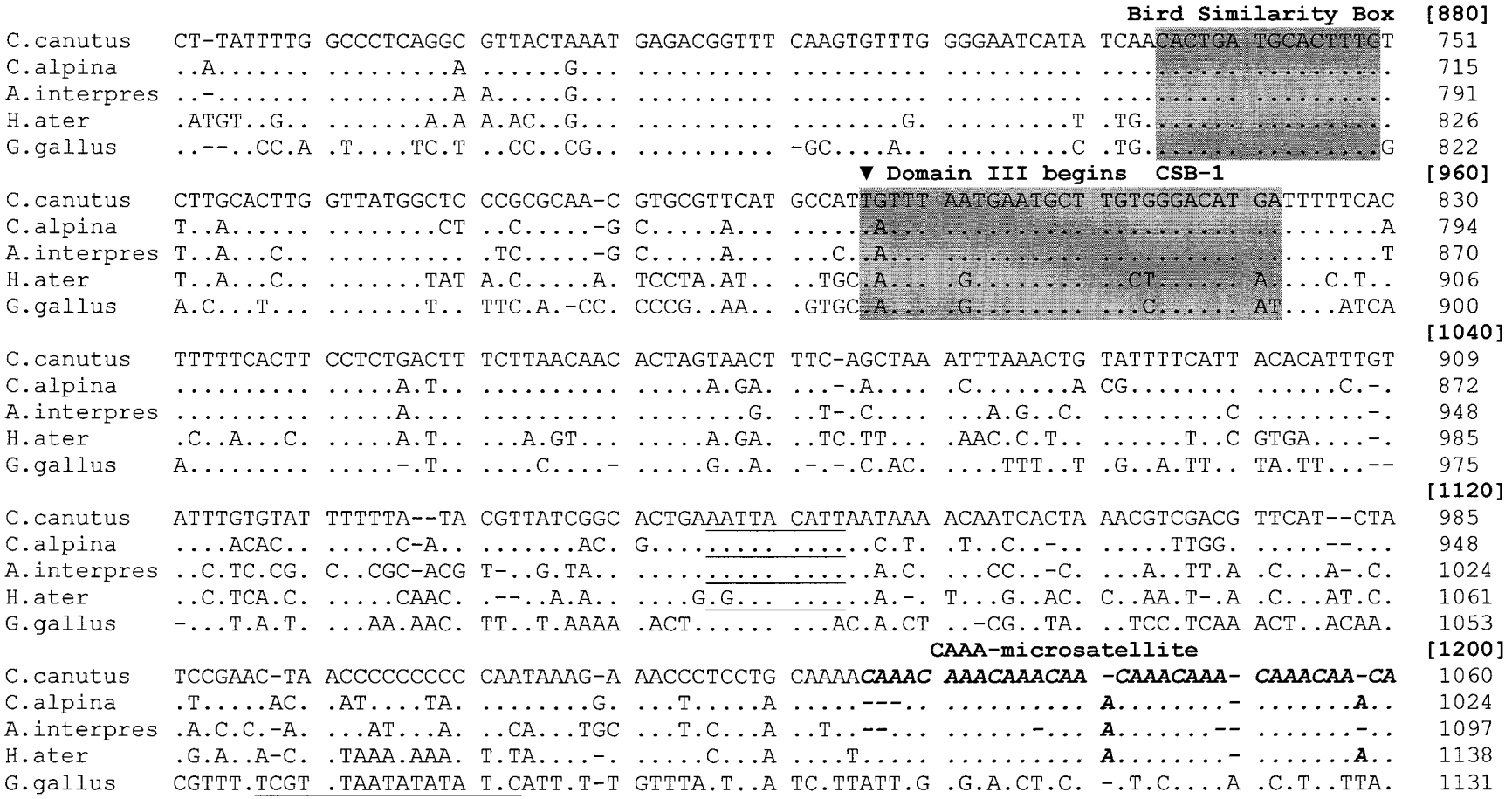


Figure 2-1 continued

```
                                                                                                                [1280]
C.canutus     AACAAACAAA CAA-CAAACA AACAAACAA- -CAAACAAAC AAACAA-CAA ACAAACAAAC AA-CAAACAA ACAAACAA-C  1135
C.alpina      .......... ...A...... .........A -......... ...AT.-.CC ---------- ---------- ----------  1072
A.interpres   .....-.... ...-...... .-......-A -.......-. .......-... ....-..... ..A...-... -.......A.  1166
H.ater        .......... ...C...C.. .........A -......... ......A... .......... ..A....... ........A.  1217
G.gallus      .....--... TTTA..TG.C .CTT...TCC C.TC...... ..T.GT-T.T TT.T.TTGTT ..-TT.G... ...C.A..C.  1207
                                                                                                                [1359]
C.canutus     AAACAAACAA ACAA-CAAAC AAACAAACCC CGCCC----- ---------- ---------- ---------- ---------  1168
C.alpina      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------  1072
A.interpres   ....CT---- ---------- ---------- ---------- ---------- ---------- ---------- ---------  1172
H.ater        .......... .......... .......... ..AAACAAAC AAACAAACAA ACAAACAAAC AAACAAACAA ACAAACAAA  1296
G.gallus      CGC.TTCT.C CACT.T..A- ---------- ---------- ---------- ---------- ---------- ---------  1227
```

Figure 2-1 continued

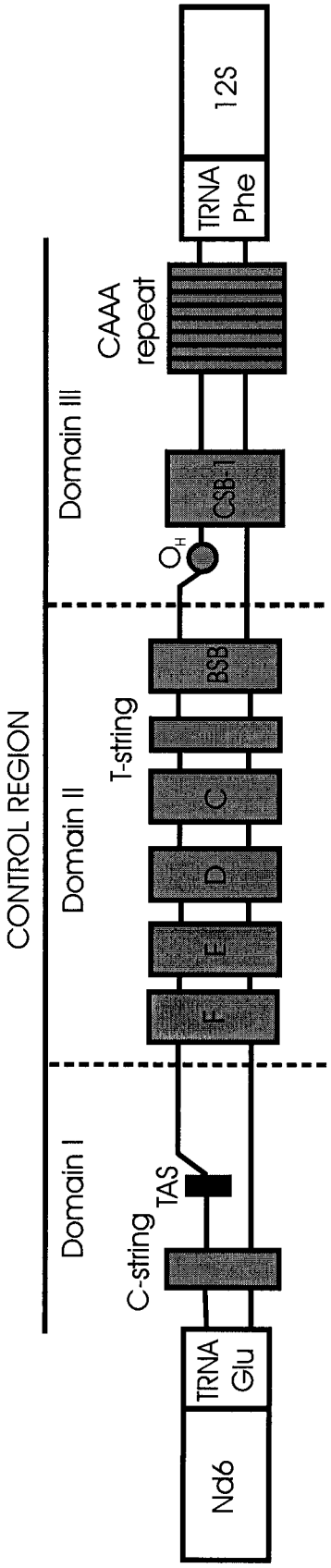


Figure 2-2: A schematic representation of the control region of red knots and its characteristics.

Table 2-1: Base composition (%) in the control region of Red Knots.

	A	C	G	T	Total bp
Domain I	30.8	32.6	13.8	22.8	325
Domain II	21.8	25.3	19.7	33.2	467
Domain III	46.2	25	6.7	22	372
Complete D-loop	32.1	27.2	13.9	26.7	1164

I found domain I of the control region to be 325 bp in length, and located two conserved elements within this domain. A cytosine string at the 5' end is widely conserved among birds (Ruokonen and Kvist 2002), and forms a hairpin structure with a string of guanines a short distance downstream in Galliformes and Anseriformes. In knots, however, this string forms a loop. The reason behind the conservation of this cytosine string remains unknown. Termination associated sequences called TAS elements involved in the termination of DNA synthesis are also found in domain I. These elements contain TATAT or TACAT motifs that are conserved to a variable degree (Foran et al. 1988). In knots as well as dunlins, turnstones and oystercatchers only a single TACAT motif was found (Figure 2-1).

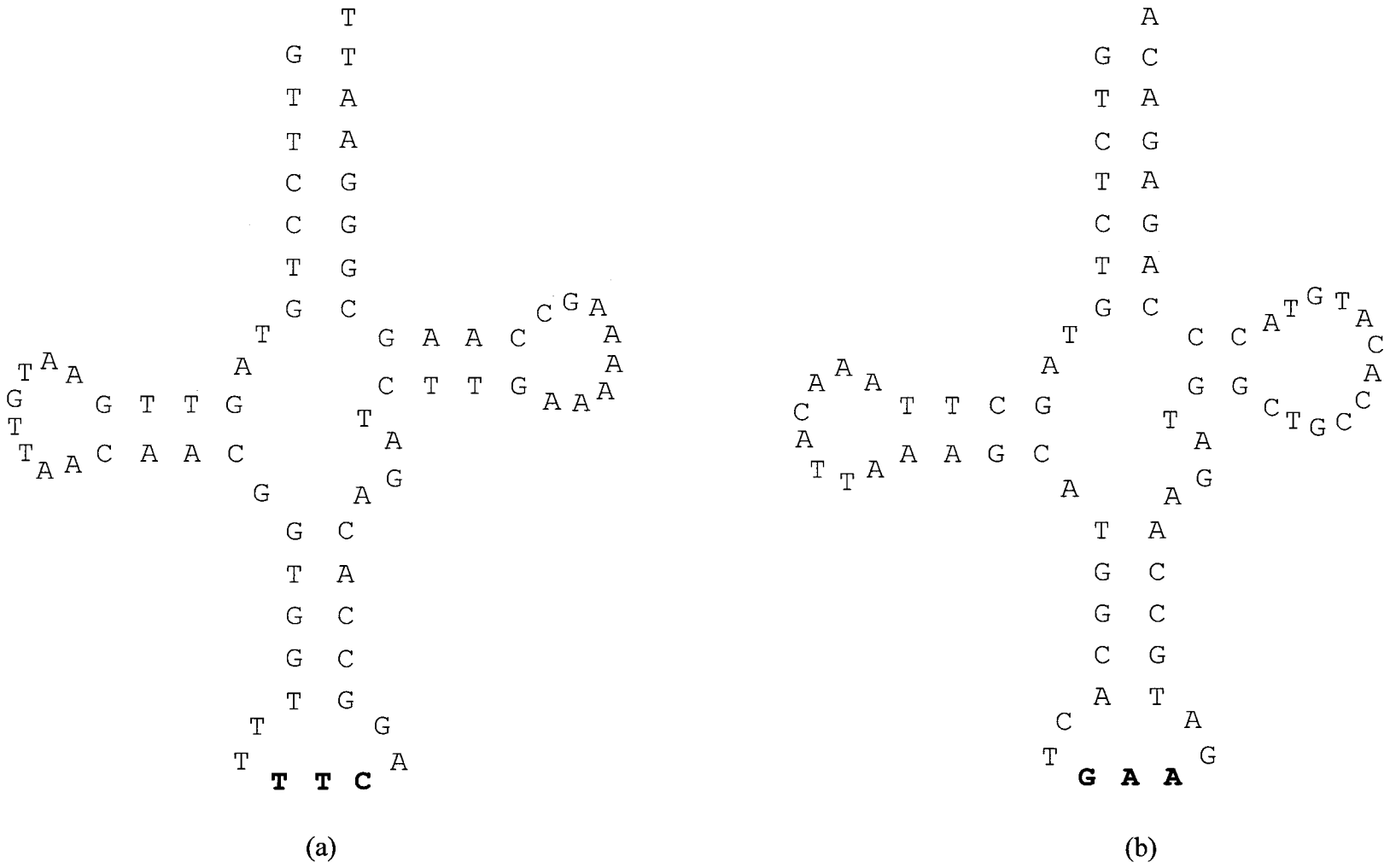
Domain II was found to be 467 in length and contained five conserved sequence blocks, identified as F, E, D, C and the bird similarity box (BSB) when compared to sequences presented in Ruokonen and Kvist (2002). All Charadriiformes and chicken had identical sequence in the BSB. Dunlin and knot sequences were identical in all boxes, Charadriiformes were identical in box D, and turnstones dunlins and knots were identical in boxes C and E. A 3 bp deletion was found in box C for all Charadriiformes when aligned to chicken. A T-string was located between the C box and the BSB in knots. This string was interrupted with cytosines to a varying degree in turnstones, oystercatchers and chickens, and was slightly shorter in dunlins (Figure 2-1).

Domain III was found to be 372 bp in length and the CSB-1 sequence was located at its 5' end. The origin of heavy strand replication (O_H) is thought to occur close to this conserved sequence block (Ghivizziani et al. 1994). In addition, knots possessed a tetra

nucleotide microsatellite repeat at the 3' end of domain III. This type of repeat has also been found in dunlins, turnstones, and oystercatchers (Charadriiformes), as well as in Adélie penguins (*Pygoscelis adeliae,* Sphenisciformes; Ritchie and Lambert 2000). In knots, it is not a perfect repeat but instead follows a pattern of $(CAAA)_3(CAA)$. Stutter patterns produced at the far end of this microsatellite on sequencing gels indicate the possibility of heteroplasmy (intra-individual length variation) in this repeat. This type of heteroplasmy was first reported in the order Ciconiiformes (Berg et al. 1995).

Domain III of the chicken control region contains a single bi-directional promoter for the transcription of both light and the heavy strands (LSP/HSP) (L'Abbé et al. 1991). This promoter sequence (5'-GTATAATATATATACA –3') is found towards the end of domain III just before the 5' end of the CAAA repeat in the alignment with Charadriiformes. In Adélie penguins, a putative LSP/HSP promoter (5'CATTAATATATAATAG –3') also precedes the CAAA repeat (Ritchie and Lambert 2000). In knots a palindromic motif with a cytosine string is aligned with the chicken transcription promoter sequence. The TATA promoter sequence is also absent from the other Charadriiformes in the alignment. Knots, dunlins, turnstones and oystercatchers do possess a well-conserved TACAT motif approximately 50 bp upstream from the chicken LSP/HSP promoter in the alignment. When folded this motif forms a cruciform structure slightly smaller than those formed by the chicken and penguin motifs.

Flanking the knot control region are tRNA's Glu and Phe. To verify the authenticity of the sequence I tested the secondary structure of these tRNA regions. In order to function correctly these tRNA's must fold into a clover-leaf motif. Figure 2-3



(a) (b)

Figure 2-3: Cloverleaf folding pattern shown by the tRNA Glu (a) and tRNA Phe (b) which flank the knot control region.

shows that the tRNA sequences folded correctly as predicted by the tRNA folding model published for the chicken (Desjardins and Morais 1990).

Finally, the relative rate test of sequence evolution found no significant difference in rates of substitutions between the control region sequences of knots and turnstones (knot = 3.2×10^{-3} substitutions/site/million years, turnstone = 2.8×10^{-3} substitutions/site/million years, $Z = 0.67$). This finding supports the argument that the sequence is mitochondrial in origin. The absolute rate of mutation in the nucleus is much lower than that in the mitochondria, despite the fact that lack of functionality allows pseudogenes to acquire more mutations over time (Arctander 1995). Thus, if the knot sequence were a nuclear pseudogene the rate of substitution relative to turnstones would be much lower.

Preliminary screening for genetic variation using the full control region sequence of 11 individuals showed eight variable sites (a single transversion and seven transitions), producing eight haplotypes (Table 2-2). This is more variation than was found in 25 individuals using only the 5' end of the control region (Baker et. al. 1994). In this study five of the eight variable sites were located towards the 3' end of the control region highlighting this area as a source of variability and possible markers for future population genetics work in this species.

To summarize, this study presents the first complete control region sequence of Red Knots. The control region is 1168 bp in length and contains all of the expected conserved sequence blocks, and repeat structures. Furthermore, substitution rate in the sequence did not vary significantly from that of the control region in turnstones and

Table 2-2: Variable sites found in the control regions (excluding the 3' tetranucleotide repeat) of a sample of 11 knots representing five subspecies. Variation is represented with respect to the first individual listed and identities are indicated by dots. Haplotypes are identified by number in the third column. Sites are numbered according to the complete control region sequence presented in Figure 2-1.

Individual	Subspecies	Haplotype	Variable Sites 112 67889 398 08029 885 62779
M5315	*C.c. rufa*	1	TTT GATAT
REKN5B	*C.c. rufa*	1	... .
11428006	*C.c. roselaari*	2	C
11428176	*C.c. roselaari*	3	A.. .GC..
MKP311	*C.c. roselaari*	4	..C C
AJB6050	*C.c. piersmai*	5	... A..G.
AJB6051	*C.c. piersmai*	1	... .
80261066	*C.c. islandica*	1	... .
80261067	*C.c. islandica*	6	C
KN12	*C.c. canutus*	7	.C. A..G.
KN13	*C.c. canutus*	8	... A...C

flanking tRNA regions, folded correctly into secondary structure. Finally, sequence obtained from purified mtDNA is identical to sequence obtained using total genomic DNA. These characteristics verify that the sequence presented is mitochondrial in origin and indicate that this protocol as well as the sequence itself can serve as a reference for future population genetics studies in knots. Furthermore, this study shows the value of the 3' end of the control as a source of variation and possible markers, setting the stage for a large-scale phylogeographic survey using a larger sample size and sequence from both the 5' and 3' ends of the control region.

Acknowledgements

This research was supported by an NSERC PGS A Scholarship (PGSA-231177-2000) to DMB and grants from NSERC (200-02) and the Royal Ontario Museum Foundation to AJB. I thank Oliver Haddrath for technical support in the lab and stimulating discussion in the analysis. I also thank Dr. David Irwin and Tara Paton for helpful comments and suggestions on the manuscript.

References

Arctander, P. 1995. Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proceedings of the Royal Society of London, Series B 262:13-19.

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

Baker, A. J., T. Piersma, and L. Rosenmeier. 1994. Unraveling the intraspecific phylogeography of knots (*Calidris canutus*) - A progress report on the search for genetic markers. Journal fur Ornithologie 135:599-608.

Baker, A. J., and H. D. Marshall. 1997. Mitochondrial control region sequences as tools for understanding evolution. Pp. 51-82 *in* D. P. Mindell, ed. Avian Molecular Evolution and Systematics. Academic Press, San Diego.

Berg, T., T. Moum, and S. Johansen. 1995. Variable numbers of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. Current Genetics 27:257-262.

Brown, W. M., M. B. George, and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the U.S.A. 76:1967-1971.

Chang, D., and D. Clayton. 1986. Identification of primary transcription start sites of mouse mitochondrial DNA: Accurate in vitro initiation of both heavy- and light-strand transcriptions. Molecular Cell Biology 6:1446-1453.

Desjardins, P., and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome: A novel gene order in higher vertebrates. Journal of Molecular Biology 212:559-634.

Foran, D. R., J. E. Hixson, and W. S. Brown. 1988. Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop synthesis. Nucleic Acid Research 166:5841-5861.

Ghivizziani, S. C., C. S. Madsen, M. R. Nelen, C. V. Ammini, and W. W. Hauswirth. 1994. In organello footprint analysis of human mitochondrial DNA: Human mitochondrial transcription factor A interactions at the origin of replication. Molecular Cell Biology 14:7717-7730.

Kumar, S. 2001. PHYLTEST: A program for testing phylogenetic hypothesis. Computer software distributed by author.

L'Abbé, D., J.-F. Duhaime, B. F. Lang, and R. Morais. 1991. The transcription of DNA in chicken mitochondria initiates from one major bidirectional promoter. Journal of Biological Chemistry 266:10844-10850.

Mindell, D. P., M. D. Sorenson, and D. E. Dimcheff. 1998. Multiple independent origin of mitochondrial gene order in birds. Proceedings of the National Academy of Sciences of the U.S.A. 95:10693-10697.

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

Paton, T., O. Haddrath, and A. J. Baker. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proceedings of the Royal Society of London, Series B 269:839-846.

Perna, N. T., and T. D. Kocher. 1996. Mitochondrial DNA: Molecular fossils in the nucleus. Current Biology 6:128-129.

Ritchie, P. A., and D. M. Lambert. 2000. A repeat complex in the mitochondrial control region of Adelie penguins from Antarctica. Genome 43:613-618.

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

Saccone, C., G. Pesole, and E. Sbisá. 1991. The main regulatory region of the mammalian mitochondrial DNA: structure-function model and evolutionary pattern. Journal of Molecular Evolution 33:83-91.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA Sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the U.S.A. 74:5463-5467.

Sorenson, M. D., and T. W. Quinn. 1998. Numts: A challenge for avian systematics and population biology. Auk 115:214-221.

Takezaki, N., A. Razhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized tress. Molecular Biology and Evolution 12:823-833.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalities and weight matrix choice. Nucleic Acid Research 22:4673-4680.

Tourmen, Y., O. Baris, P. Dessen, C. Jacques, Y. Malthiery, and P. Reynier. 2002. Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics 80:71-77.

Valverde, J. R., R. Marco, and R. Garesse. 1994. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria. Proceedings of the National Academy of Sciences of the U.S.A. 91:5368-5371.

Van Wagner, C. E., and A. J. Baker. 1990. Association between mitochondrial DNA and morphological evolution in Canada Geese. Journal of Molecular Evolution 31:373-382.

Wenink, P. W., A. J. Baker, and M. G. L. Tilanus. 1993. Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpina*). Proceedings of the National Academy of Sciences of the U.S.A. 90:94-98.

Wennerberg, L., M. Klaassen, and Å. Lindstrom. 2002. Geographic variation and population structure in the white-rumped sandpiper *Calidris fusicollis* as shown by morphology, mitochondrial DNA and carbon isotope ratios. Oecologia 131:380-390.

Woischnik, M., and C. T. Moraes. 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Research 12:885-893.

CHAPTER 3

Snapshots of genetic evolution: Using coalescent theory to examine population divergence times and biogeographic history in Red Knots and Dunlins

Abstract

Coalescent analysis is a non-equilibrium based modeling approach that allows researchers to make inferences about the demographic histories of species and populations based on the shape of gene genealogies. Herein I employ Bayesian coalescent modeling of samples of mitochondrial control region sequences in two species of shorebird, the Red Knot (*Calidris canutus*) and the dunlin (*Calidris alpina*) to estimate evolutionary effective population size, population divergence times and time to most recent common ancestor (TMRCA) of genes in the samples. The gene trees for the two species contrast sharply: knot haplotypes are connected in a shallow, star phylogeny whereas dunlin haplotypes are related in a deeper bifurcating genealogy. Divergence times of populations representing all six subspecies of knots are shown to have occurred within the last 20,000 years or so, and evolutionary effective population sizes of females are small (N_{ef} = 2,000 – 14,000). This implies that knots were bottlenecked in a small refugial population during the Last Glacial Maximum (LGM) 18,000 – 21,000 ybp, and subsequently expanded to their worldwide range as the ice sheets retreated from the arctic breeding grounds. I hypothesize that breeding knots were restricted to unglaciated eastern Eurasia arctic regions during the LGM, and gradually expanded eastwards into Alaska, the high Canadian arctic and Greenland as the ice melted. Population divergence

times in Dunlins are much older (58,000 – 194,000 ybp) and effective population size has historically been higher in major lineages (N_{ef} = 12,000 – 44,000). I conclude that dunlin populations were not severely reduced in size in the last 200,000 years, and major lineages have differentiated under restricted gene flow for a much longer time than knots. Knots offer us a snapshot of genetic evolution in the last 20,000 years, whereas Dunlins show us genetic evolution over an order of magnitude longer time frame.

Introduction

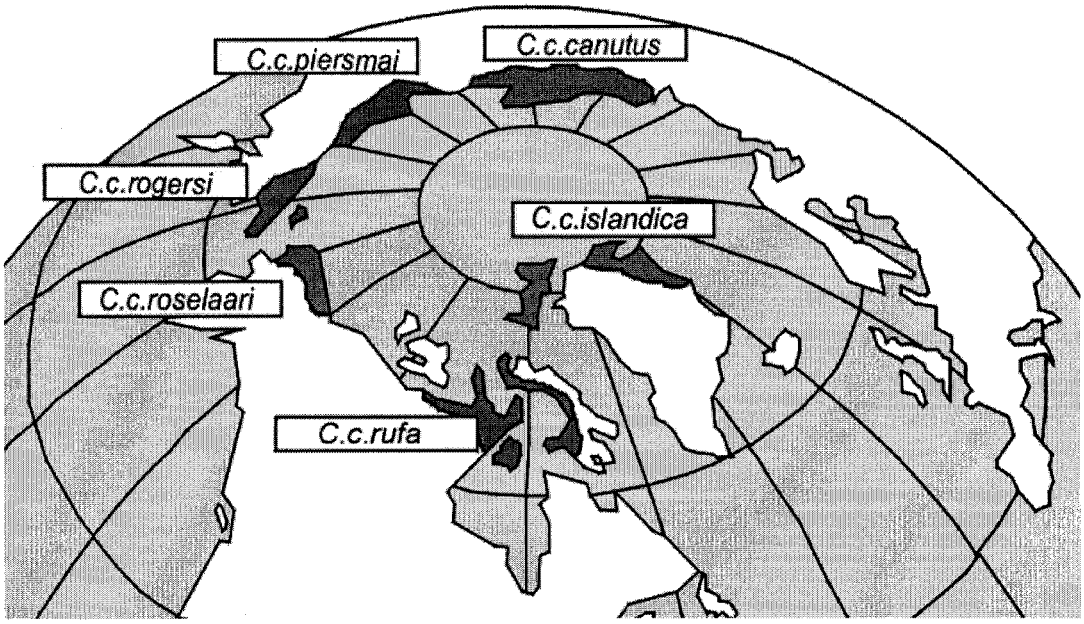
In recent years molecular analysis using gene trees has allowed researchers to make inferences about the demographic histories of various extant species (see Baker in press, Fedorov and Stenseth 2002 and Griswold and Baker 2002 for examples). In this study I use Bayesian coalescent modeling to examine population divergence times and to make inferences about the historical demography of two species of shorebird, the Red Knot (*Calidris canutus*) and the dunlin (*Calidris alpina*). Knots and Dunlins are an excellent pair of closely related species for comparison of demographic history. They are both long-distance migrants with similar life history traits, yet all previous molecular investigations have shown radically different phylogeographic patterns, with knots showing apparent panmixia caused by a late Pleistocene bottleneck (Baker et al. 1994), and Dunlins showing clearly distinguishable phylogeographic lineages (Wenink et al. 1996; Wenink et al. 1993).

Red Knots are currently classified as six subspecies, each with distinct morphological traits, migration routes and annual cycles: *C.c. roselaari, C.c. rufa, C.c. piersmai, C.c. islandica, C.c. canutus* and *C.c. rogersi* (Tomkovich 1992; Tomkovich

2001). Previous research on the population genetics of knots failed to distinguish geographically isolated groups (Baker et al. 1994). In this paper I use additional data to further examine this problem. Dunlins comprise five morphologically and genetically distinct mitochondrial DNA (mtDNA) lineages found in Canada, Europe, central Siberia, Beringia, and Alaska (Wenink et al. 1996). The Canadian lineage comprises *C.a. hudsonia*, the European linage contains *C.a. arctica, C.a. schinzii,* and *C.a. alpina*, the central Siberian lineages comprises *C.a. centralis,* the Beringian lineage represents *C.a. sakhalina*, and finally the Alaskan lineage encompasses *C.a. articola* and *C.a. pacifica.* For comparative purposes I will consider the mtDNA lineages in Dunlins rather than population samples. Maps showing the breeding areas of knots and Dunlins are shown in Figure 3-1a and 3-1b respectively.

The study of evolution at the tips of evolutionary trees, the intraspecific level, brings new challenges to the researcher. When studying intraspecific phylogenies rather than species phylogenies there is a higher probability that haplotypes drawn from the populations will not have had time to sort into distinct lineages (Avise 1994). The result is polyphyly or paraphyly within species. The problem with polyphyly or paraphyly is that it becomes difficult to distinguish whether populations are genetically similar due to high levels of current gene flow or because they retain shared ancestral polymorphisms from recent range expansions. Traditional models of population subdivision assume that populations have reached equilibrium between migration and drift; however, a species undergoing lineage sorting is not in equilibrium thus models that do not assume equilibrium are necessary. Coalescent theory allows the development of such models.

(a) Knots



(b) Dunlins

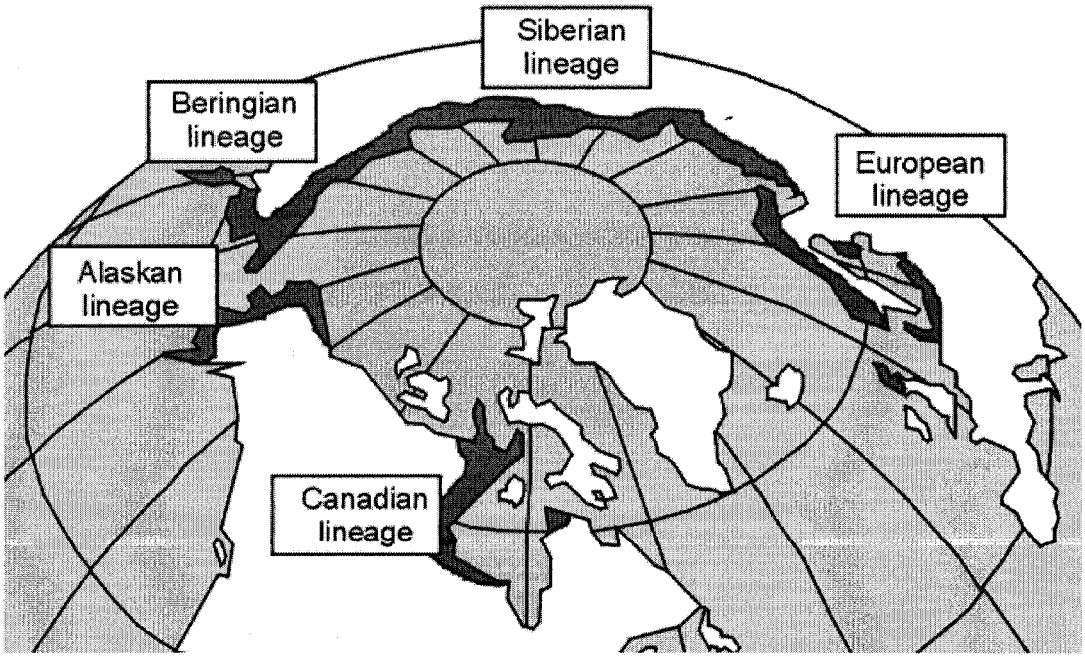


Figure 3-1: Arctic breeding grounds for knot subspecies (a) and dunlin mtDNA lineages (b).

Coalescent theory has experienced rapid development over the last few years (Fu and Li 1999; Griffiths and Tavaré 1997; Hudson 1990; Kingman 1982), and out of this period of development many analytical methods have emerged. Most of these methods strive to make inferences about either population divergence times based on migration data, or inferences about the migration of individuals between populations based on divergence data. Nielsen and Wakeley (2001) describe a coalescent model that jointly estimates divergence time and migration rates for populations using DNA sequence data.

I use the Nielsen and Wakeley model in this study to analyze mitochondrial DNA sequences from the control regions of knots and Dunlins. The sequences are first used to make inferences about the historical demography of the two species. I then examine the relative importance of non-equilibrium conditions (or shared ancestral polymorphisms) among populations versus gene flow on the polyphyly or paraphyly of lineages in the species. Finally, I combine classic phylogeography with population divergence times obtained via coalescent analysis to make inferences about the historical biogeography of these species.

Materials and Methods

Sample Collection

Red Knots were caught using ground traps in breeding areas, and cannon nets in non-breeding areas globally (Appendix 3-1). A few drops of blood were taken from each bird and stored at room temperature in 50mM EDTA and 70% ethanol, then subsequently frozen at –80˚C in the Royal Ontario Museum's Ornithology collection.

Knot subspecies separate on the breeding grounds, intermingle in certain staging areas (i.e. *C.c. rufa* and *C.c. roselaari* in the southeastern USA, *C.c. islandica* and *C.c. canutus* in the Wadden Sea and *C.c. piersmai* and *C.c. rogersi* in southeastern Australia) and separate again in wintering areas. The subspecific assignment of birds not caught on the breeding grounds was determined using the timing of passage in staging areas and global location in wintering areas. For example, knots found in staging areas on the Wadden Sea in mid-October are most certainly *C.c. islandica* birds as *C.c. canutus* knots often skip the Wadden Sea altogether during fall migration, and in years that they do appear they leave for Africa long before October (Nebel et al. 2000). Knots found in South Carolina and Georgia in mid-April or October are *C.c. roselaari* birds as *C.c. rufa* knots arrive later in the spring, and fall birds molt their feathers making it very unlikely that they are *C.c. rufa* birds en route to the Caribbean Sea (B. Harrington pers. comm.). Wintering areas are widespread and banding studies have shown that subspecies are not mixing on the wintering grounds (A. J. Baker unpublished data).

DNA Sequencing

Total DNA from each bird was isolated using standard phenol/chloroform extractions (Sambrook et al. 1989). For knots, 91 individuals representing all six morphologically recognized subspecies were sequenced. Product from the 5' end of the control region was obtained by standard polymerase chain reaction (PCR) amplification (Mullis and Faloona 1987) using primers L98 and H401 (Wenink et al. 1993) and was sequenced with H401. Product from the 3' end of the control region was obtained using

primers L438 (Wenink et al. 1993) and H1537 (5'-TGA CCG CGG TGG CTG GCA CAA G –3') (O. Haddrath) and was sequenced with a newly designed primer KnotMidCRL (5'- GCA ACG GGT GAA TAC AAT CTA AGA C –3'). The amplification protocol was a 2 minute denaturation at 94°C, followed by 36 cycles of 94°C for 45 seconds, 50°C for 45 sec, 72°C for 90 sec, with a final extension at 72°C for 5 min (Perkin-Elmer 480 thermocycler). Sequences were obtained using the Sanger dideoxy chain termination method (Sanger et al. 1977) and the Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit (USB) with P^{33} on 6% polyacrylamide gel. To confirm that the sequences were mitochondrial and not nuclear in origin, all sequences were aligned to the complete control region sequence presented in Chapter 2 and were obtained using the same primers. The 208 sequences from Dunlins worldwide were obtained from datasets published in Wenink et al. (1996) and Wenink and Baker (1996).

Analysis

Preliminary sequence analyses were performed using Modeltest 3.06 (Posada and Crandall 1998) to determine which model of sequence evolution best fitted the data. Variation in knot control region sequences was modeled best by either the HKY (Hasegawa et al. 1985) or the TrN (Tamura and Nei 1993) model. Dunlin sequences best fitted a variation of the TrN model which allowed for two different rates for transitions and two separate rates for transversions. To correct for rate variation among sites, the model for dunlins had a gamma shape parameter of $\alpha = 0.6091$. Minimum spanning

networks were created using the TrN model in the program Arlequin (Schneider et al. 2000). Arlequin was also used to calculate Tajima's test of departure from neutrality (Tajima 1989) and nucleotide diversity (π) = $\Sigma \pi_{ij}/(n(n\text{-}1)/2)$, where π_{ij} equals the proportion of the nucleotide differences between the *i*th and the *j*th sequences and *n* is the number of individuals (Nei 1987). Pairwise mismatch distributions were created using the program DNAsp3.5 (Rozas and Rozas 1999).

Coalescent analyses of knot subspecies and dunlin mtDNA lineages were performed using the HKY model of sequence evolution (Hasegawa et al. 1985) and the program Mdiv (Nielsen and Wakeley 2001). Mdiv is based on a coalescent model that jointly estimates divergence time and migration rates for populations using DNA sequence data. Input for the program is DNA sequence and output is four inferred parameters: θ, M, T and TMRCA. These parameters are measures of mutation rate, migration and divergence time respectively and can be used to calculate effective population size (N_e), the number of migrants exchanged between populations per generation ($N_{ef}m$) and population or gene divergence time (t) though the following equations (modified for mitochondrial DNA data): $\theta = 2 N_{ef}\mu$, $M = N_{ef}m$ and T (or TMRCA)= t/N_{ef}. Markov Chain Monte Carlo simulations generate posterior probability distributions whose modes represent the parameter estimates. Credibility intervals for the estimates are taken as the interval that contains 95% of the posterior distribution (Nielsen and Wakeley 2001).

I used Mdiv to estimate the long term effective population size of females (N_{ef}), the number of migrants exchanged between populations per generation ($N_{ef}m$), and

population and gene divergence times. The mutation rate (μ) needed to calculate N_{ef} was obtained using the Chaffinch mutation rate for the 5' end of the control region (Marshall 1997). This mutation rate was adjusted for sequence length by dividing by the number of base pairs sequenced in Chaffinches and multiplying by the number of base pairs sequenced in knots. Domain I (5' end) mutation rates were then adjusted to reflect lower rates of substitution in domain III (3' end) by multiplying by the ratio of variable sites in the two regions. Finally, the mutation rates for domains I and III were summed and then multiplied by a generation time (g) to obtain $\mu = 9.39 \times 10^{-5}$ substitutions/locus/generation. Generation time in both knots and Dunlins is 2 years.

Results

Neutrality Tests

In the knot control region sequences, Tajima's test showed a significant deviation from neutral expectations in the subspecies *C.c. rogersi* ($D = -1.715$, $p = 0.031$) and when all of the subspecies were pooled ($D = -1.983$, $p = 0.003$). These negative D statistics indicate that there are a greater number of rare alleles present than would be expected if the species were at equilibrium. I take this negative D as the signature of a bottlenecked population and not as a violation of neutrality. Research presented in this paper and earlier research by Baker et al. (1994) indicates that knots were severely bottlenecked in the late Pleistocene. In a population expanding rapidly after a bottleneck, an excess of rare alleles is expected as population growth preserves rare alleles from elimination by

genetic drift. In Dunlins none of the Tajima's tests were significant, and thus the sequences appear to be selectively neutral.

Sequence Variation

In knots I obtained control region sequences of 675 bp for 91 individuals containing 21 variable sites with 20 transitions and a single transversion. Twenty-five haplotypes were identified worldwide in the six named subspecies (Appendix 3-2). In Dunlins, control region sequences of 608 bp were obtained for 208 individuals and contained 52 variable sites comprised of 42 transitions, 8 transversions and two indels. The sequences yielded 53 haplotypes summarized in Appendix 3-3. Nucleotide diversities for both species are given in Table 3-1. Within-population nucleotide diversities in knots are on average two times lower than those of Dunlins, and are eight times lower when populations are pooled together in each species.

When pairwise differences among sequences are presented as mismatch distributions (Figure 3-2a and 3-2b), the pattern in knots clearly resembles the pattern expected from a population experiencing rapid growth (Avise 2000). This graph shows that most of the sequences differ by a single base pair change. In contrast, Dunlins show a curve, which fits neither population growth nor constant population size, but instead is a multi-modal pattern characteristic of a species with several well differentiated groups. The two major modes in the dunlin distribution represent within and between group differences.

Genetic Structure

Knot and dunlin haplotypes are summarized in minimum spanning networks shown in Figures 3-3 and 3-4. Most haplotypes in the knot network differ by a single base change producing a star-like pattern characteristic of a species that has undergone a recent bottleneck (Slatkin and Hudson 1991). Furthermore, the most common haplotypes in the network, Rog1, Isl1 and Can1 are shared among subspecies, with the only evidence of population genetic structure coming from the clustering of *C.c. canutus* haplotypes near the top of the network. In sharp contrast the dunlin network shows very clear sorting into lineages. The Canadian haplotypes are clearly the most divergent, differing from all other haplotypes by at least 14 mutational steps. European haplotypes are differentiated from Siberian, Beringian and Alaskan haplotypes by at least six mutational steps. And Siberian, Beringian and Alaskan haplotypes differ by at least three to four substitutions.

Despite the apparent lack of sorting in the minimum spanning network, knots showed low but significant population differentiation using both conventional F-statistics and Exact tests (Table 3-2 for F_{ST} summary). Four genetically distinct groups were found corresponding to *C.c. canutus, C.c. piersmai, C.c. rogersi* and a North American group containing *C.c. roselaari, C.c. rufa* and *C.c. islandica* (Pooled Exact test, $p < 0.001$). Dunlins showed highly significant population differentiation among all lineages (F_{ST} and Exact test, $p < 0.0001$)

Table 3-1: Nucleotide diversities ± standard deviation, for knots (a) and Dunlins (b).

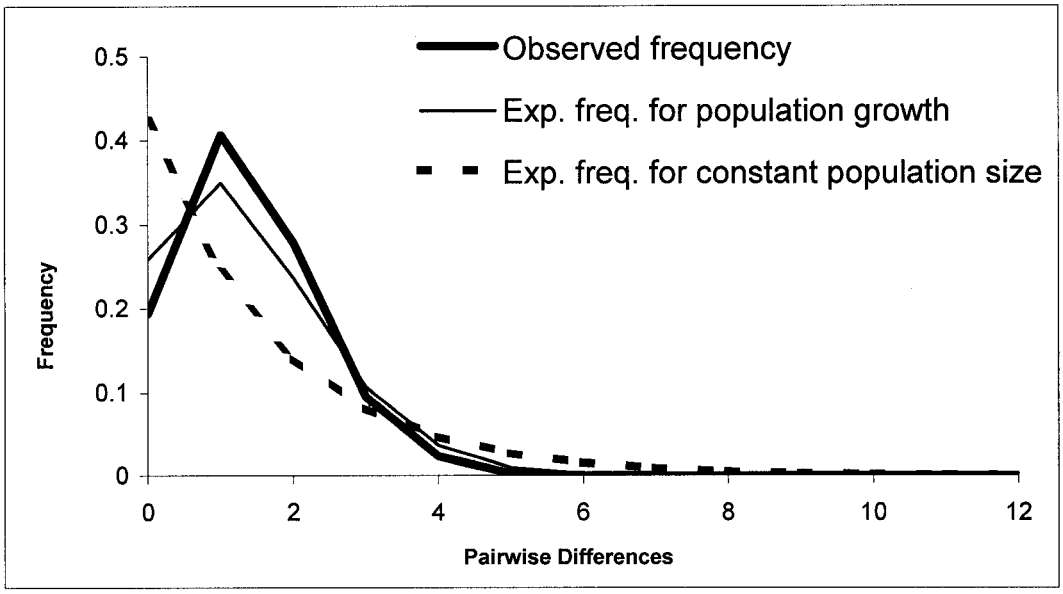
(a) Knots

	C.c. canutus (n=12)	*C.c. islandica* (n=15)	*C.c. piersmai* (n=15)	*C.c. rogersi* (n=19)	*C.c. roselaari* (n=15)	*C.c. rufa* (n=15)	Pooled (n=91)	Average per Population
Nucleotide Diversity	0.0023 ± 0.0017	0.0016 ± 0.0012	0.002 ± 0.0014	0.0012 ± 0.0010	0.0021 ± 0.0015	0.0011± 0.001	0.002 ± 0.0014	0.0020

(b) Dunlins

	Alaskan lineage (n=33)	Beringian lineage (n=7)	Canadian lineage (n=16)	European lineage (n=110)	Siberian lineage (n=42)	Pooled (n=208)	Average per Population
Nucleotide Diversity	0.0024 ± 0.0016	0.0046 ± 0.0031	0.0027 ± 0.0019	0.0030 ± 0.0020	0.0053 ± 0.0031	0.0160 ± 0.008	0.0036

(a) Knots



(b) Dunlins

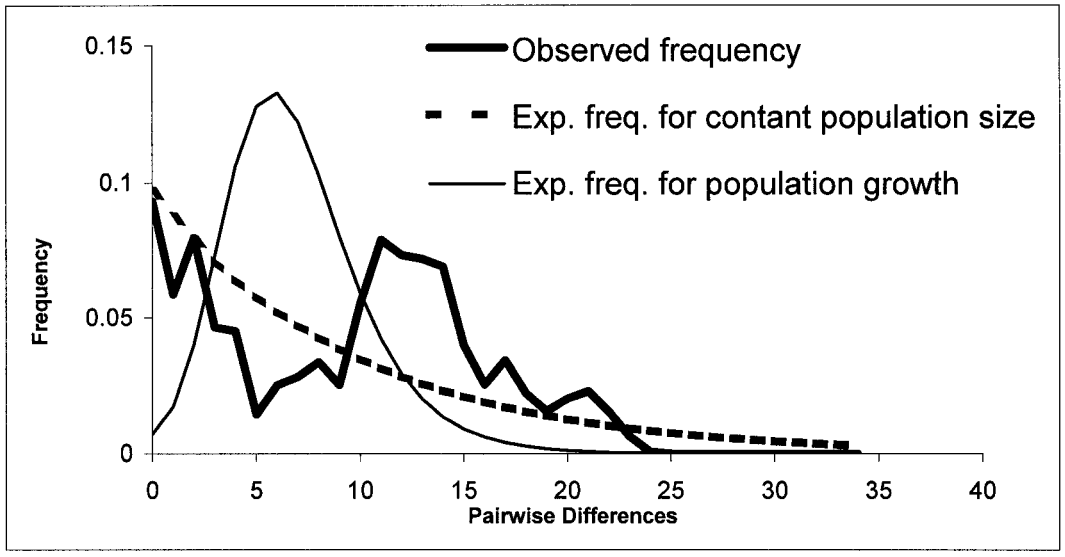


Figure 3-2: Observed and expected mismatch distributions in mitochondrial control region sequences for knots (a) and dunlins (b). Knots closely match the pattern expected under growth conditions. Dunlins show multiple modes indicative of within and between population peaks.

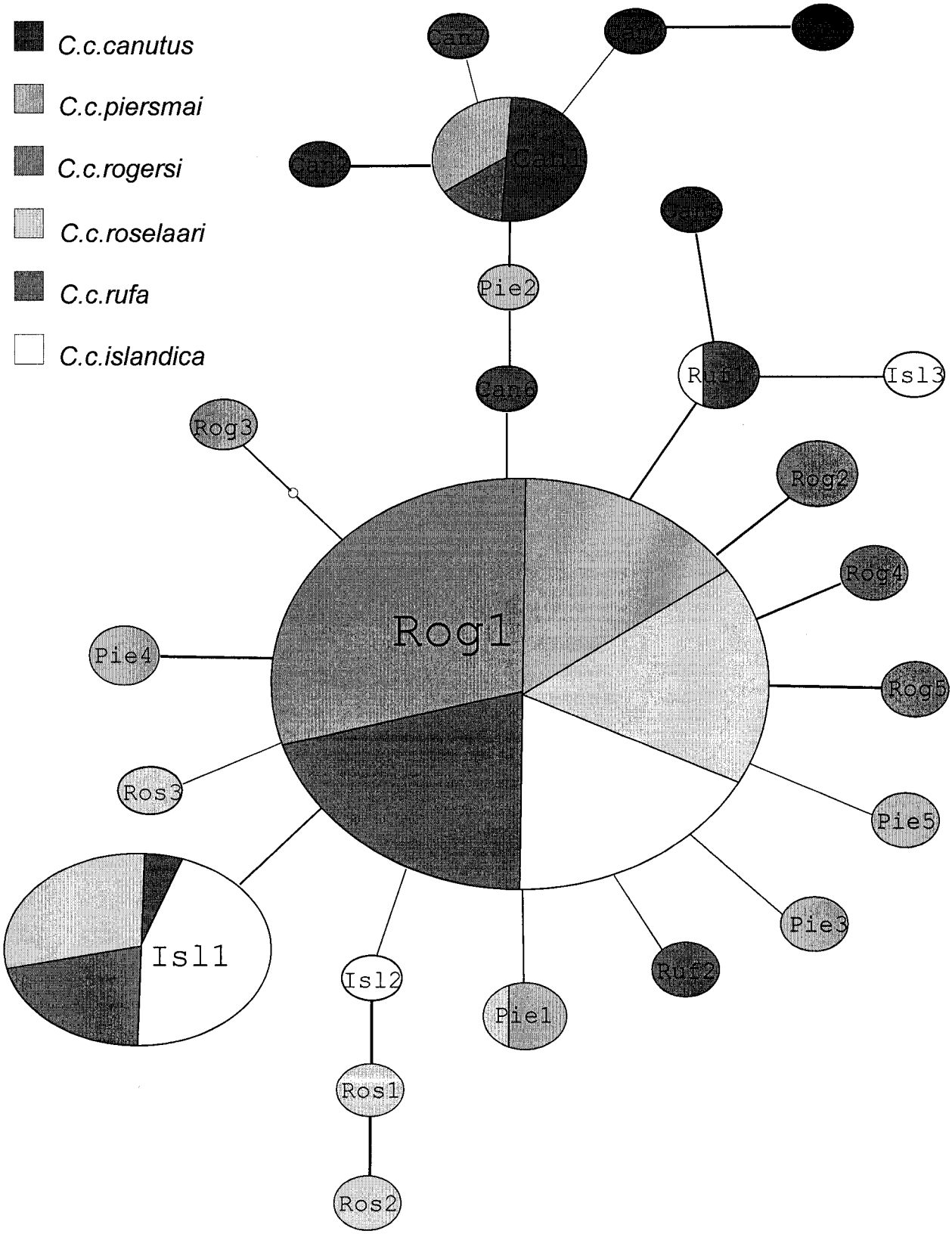


Figure 3-3: Minimum spanning network showing the relationships between haplotypes from mitochondrial control region sequences of knots. Knots show a star-like pattern characteristic of a recently bottlenecked species.

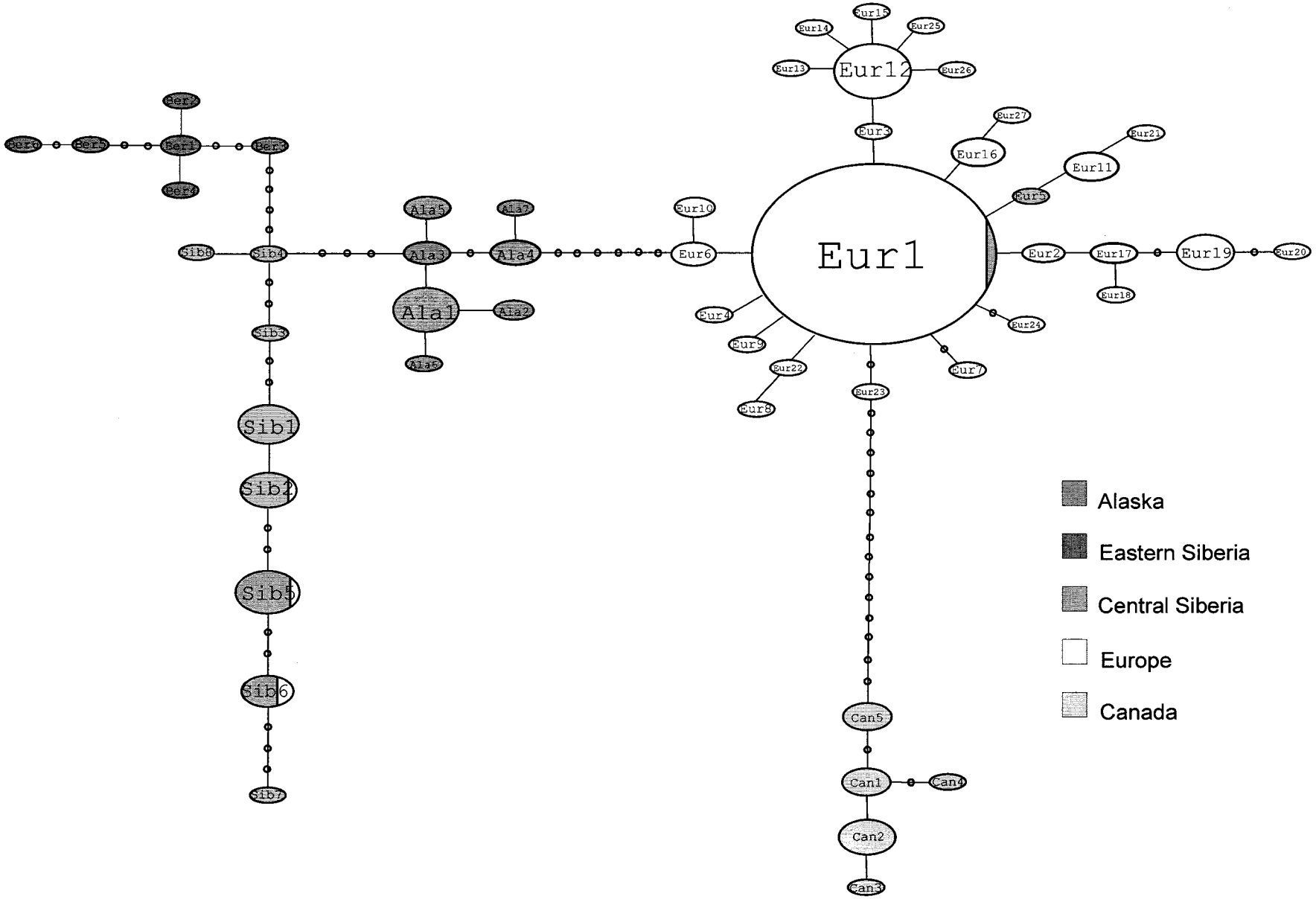


Figure 3-4: Minimum spanning network showing the relationships between haplotypes from mitochondrial control region sequences of dunlins. Dunlins show a deep bifurcating genealogy of old and well defined phylogroups.

Coalescent Analysis

Posterior probability distributions of θ for both the knot and dunlin sequences were bell-shaped, indicating a good estimation of this parameter. The θ parameter was used to calculate N_{ef} ($\theta = 2 N_{ef}\mu$), which was in turn used to calculate the population and gene divergence time estimates ($T = t/N_e*g$).

Female effective population sizes in knots were quite low and ranged from 4,100 in the *C.c. rufa/C.c. islandica* comparison to 14,058 in the *C.c. piersmai/C.c. canutus* comparison (Table 3-3). These effective population sizes were on average half those of the major dunlin lineages which ranged from 10,224 for the Canadian and Alaskan comparison to 44,462 for the European and Beringian comparison (Table 3-4). Estimates of migration ($N_{ef}m$) among subspecies of knots were on average ten times larger than those among lineages of Dunlins (all < 0.05), with knot migration estimates similar to those in European populations of Common Chaffinches (Griswold and Baker 2002) and lower than those of Grasshopper Sparrows (Bulgin et al. in press). Migration estimates were not possible between *C.c. islandica* and *C.c. rufa* or *C.c. islandica* and *C.c. roselaari* because the posterior probability distributions were ill-conditioned and did not show clear modes.

Estimates of population divergence times in knots indicate that subspecies groups diverged very recently at three time intervals (Table 3-5, above diagonal). The lineage leading to *C.c. canutus* diverged from the ancestral population approximately 20,000 (95% CI: 5,600 – 58,000) years ago, a North American group including *C.c. roselaari, C.c. rufa* and *C.c. islandica* diverged from a Siberian ancestor approximately

Table 3-2: Knot F_{ST} values for population differentiation (below diagonal) calculated using the Tamura and Nei (1993) model of sequence evolution. Above the diagonal "+" indicates significance at the 0.05 level.

	C.c. canutus	*C.c. islandica*	*C.c. piersmai*	*C.c. rogersi*	*C.c. roselaari*	*C.c. rufa*
C.c. canutus	0	+	+	+	+	+
C.c. islandica	0.3866	0	+	+	-	-
C.c. piersmai	0.1910	0.1950	0	+	+	+
C.c. rogersi	0.3713	0.2000	0.0633	0	+	+
C.c. roselaari	0.3538	-0.0162	0.1392	0.1462	0	-
C.c. rufa	0.3798	-0.0090	0.1191	0.0948	0.0136	0

Table 3-3: Mdiv estimates of θ and N_{ef} (below the diagonal), and $N_{ef}m$ (above the diagonal) in knots. The numbers in parentheses are 95% credibility intervals. Undefined indicates that Mdiv analysis produced an ill-conditioned posterior distribution and a modal estimate was not possible.

	C.c. canutus	*C.c. islandica*	*C.c. piersmai*	*C.c. rogersi*	*C.c. roselaari*	*C.c. rufa*
C.c. canutus		0.34	0.69	0.39	0.29	0.32
	-	(0.0-2.6)	(0.0-5.0)	(0.01-2.8)	(0.0-2.5)	(0.0-2.1)
C.c. islandica	2.06					
	(0.8-4.9)		0.26	0.23		
	10,969	-	(0.0-2.5)	(0.0-1.5)	undefined	undefined
C.c. piersmai	2.61	1.65				
	(1.1-6.8)	(0.7-4.5)		1.39	0.98	0.05
	14,058	8,786	-	(0.1-6.1)	(0.0-10.0)	(0.0-3.0)
C.c. rogersi	2.23	1.63	1.99			
	(1.1-5.7)	(0.7 -4.1)	(0.7-4.9)		0.61	0.24
	11,874	8,679	10,596	-	(0.0-7.1)	(0.0-6.3)
C.c. roselaari	2.39	1.26	1.87	2.10		
	(1.1-58)	(0.4-3.6)	(0.8-5.0)	(0.9-5.1)		3.68
	12,726	6,709	9,957	11,182	-	(0.0-5.0)
C.c. rufa	1.83	0.77	1.51	1.58	1.28	
	(0.6-4.2)	(0.2-2.5)	(0.6-4.7)	(0.6-4.5)	(0.6-3.6)	
	9,744	4,100	8,040	8,413	6,816	-

Table 3-4: Mdiv estimates of θ and N_{ef} (below the diagonal), and $N_{ef}m$ (above the diagonal) in Dunlins. The numbers in parentheses are 95% credibility intervals.

	Alaskan lineage	Beringian lineage	Canadian lineage	European lineage	Siberian lineage
Alaskan		0.037	0.003	0.044	0.026
lineage	-	(0.0-4.8)	(0.0-0.2)	(0.0-0.4)	(0.0-0.4)
Beringian	2.39				
lineage	(1.2-5.6)		0.004	0.078	0.038
	12,726	-	(0.0-0.4)	(0.0-0.6)	(0.0-0.6)
Canadian	1.92	2.87			
lineage	(1.0-4.3)	(1.3-6.9)		0.029	0.008
	10,224	12,282	-	(0.0-0.5)	(0.0-0.3)
European	6.32	8.35	6.85		
lineage	(4.1-10.5)	(1.3-6.9)	(4.3-12.1)		0.032
	33,653	44,462	36,475	-	(0.0-3.9)
Siberian	2.78	3.25	3.24	6.85	
lineage	(1.6-5.3)	(2.1-7.3)	(2.2-5.9)	(4.3-10.6)	
	14,803	17,306	17,252	36,475	-

12,000 (95% CI: 3,300 – 40,000) years ago and *C.c. piersmai* and *C.c. rogersi* diverged from one another approximately 6,500 (95% CI: 1,000 – 23,000) years ago. Modal estimates of divergence times for *C.c. roselaari, C.c. rufa* and *C.c. islandica* were estimated at approximately 700 years ago; however, 95% credibility intervals included zero (95% CI: 0 - 5,500).

Population divergence times in Dunlins were an order of magnitude greater than those of knots and corresponded well with previously published divergence times calculated from the percentage of sequence divergence after correction for multiple hits (Wenink et al. 1996). The Canadian lineage diverged approximately 180,000 (95% CI: 38,000 – 400,000) years ago, the European lineage diverged approximately 110,000 (94% CI: 53,000 – 370,000) years ago, and the Siberian, the Beringian and the Alaskan lineages diverged approximately 70,000 (95% CI: 21,000 – 180,000) years ago (Table 3-6). For both knots and Dunlins gene divergence times were much greater than population divergence times. This was expected, as genes are often older than the populations in which they arise.

Discussion

Demographic History

Knots and Dunlins provide classic examples of different phylogeographic patterns indicating very different demographic histories. Control region sequences from knots present the classic signature of a species that was recently and severely bottlenecked.

Table 3-5: Mdiv estimates of TMRCA and gene divergence times (below the diagonal), and T and population divergence times (above the diagonal) for pairwise comparisons in knots. The numbers in parentheses are 95% credibility intervals. A generation time of 2 years was used to translate divergence times into years before present.

	C.c. canutus	*C.c. islandica*	*C.c. piersmai*	*C.c. rogersi*	*C.c. roselaari*	*C.c. rufa*
C.c. canutus	-	1.04	0.58	0.92	0.82	0.95
		(0.3-2.7)	(0.2-2.0)	(0.2-2.3)	(0.3-2.7)	(0.2-2.9)
		22,816	16,307	21,849	20,871	18,514
C.c. islandica	1.93	-	0.61	0.73	0.052	0.012
	42,341		(0.2-2.6)	(0.2-2.5)	(0.0-0.5)	(0.0-0.2)
			10,719	12,672	698	98
C.c. piersmai	1.51	1.95	-	0.31	0.57	0.86
	42,454	34,265		(0.05-1.1)	(0.1-1.7)	(0.2-2.3)
				6,570	11,351	13,830
C.c. rogersi	1.68	2.01	1.70	-	0.46	0.64
	39,898	34,891	36,028		(0.2-2.2)	(0.2-2.0)
					10,288	10,769
C.c. roselaari	1.81	2.17	1.89	1.83	-	0.092
	46,069	29,118	37,639	40,927		(0.0-0.6)
						1,254
C.c. rufa	2.05	2.58	1.89	1.92	2.33	-
	39,952	21,157	30,393	32,307	31,761	

Table 3-6: Mdiv estimates of TMRCA and gene divergence times (below the diagonal), and T and population divergence times (above the diagonal) for pairwise comparisons in Dunlins. The numbers in parentheses are 95% credibility intervals. A generation time of 2 years was used to translate divergence times into years before present.

	Alaskan lineage	Beringian lineage	Canadian lineage	European lineage	Siberian lineage
Alaskan	-	3.08	8.76	1.62	1.97
lineage		(0.8-8.5)	(1.2-19.0)	(0.8-4.5)	(0.8-5.5)
		78,394	179,118	109,035	58,324
Beringian	3.83	-	6.36	1.29	1.95
lineage	97,483		(1.0-14.0)	(0.6-4.0)	(0.6-5.0)
			194,390	114,712	67,492
Canadian	11.1	7.74	-	2.34	5.68
lineage	226,965	236,569		(0.8-5.0)	(1.1-13.0)
				170,703	195,987
European	2.27	1.85	3.11	-	1.53
lineage	152,784	164,510	226,874		(0.7-4.1)
					111,613
Siberian	2.77	2.70	7.19	2.25	-
lineage	82,009	93,450	248,089	164,137	

This is evidenced by low genetic diversity in comparison to Dunlins, especially at the species level, a negative Tajima's D for the pooled subspecies, a mismatch distribution characteristic of rapid population expansion and a star-like minimum spanning network. In contrast, control region sequences in Dunlins show the classic signature of ancient and well-defined mtDNA matrilines corresponding to allopatrically separated populations, which have not been recently bottlenecked. This is evidenced by high levels of genetic diversity, a non-significant Tajima's D, a multi-modal mismatch distribution characteristic of two or more well defined populations, and a deep bifurcating gene tree with well defined phylogroups which are geographically separated.

Genetic Equilibrium, Gene Flow and Population Structure

A comparison between knots and Dunlins is instructive with regards to non-equilibrium conditions (due to shared ancestral polymorphisms) versus current gene flow as the cause for polyphyly and paraphyly in gene trees. In recently bottlenecked species such as knots, shared ancestral polymorphisms can result in genetic similarity among recently diverged populations, causing misleading estimates of gene flow and a lack of phylogenetic distinctiveness (Bulgin et al. in press). This problem is exemplified by the subspecies *C.c. roselaari*, *C.c. rufa* and *C.c. islandica,* which exhibit no detectable levels of genetic differentiation even in the fast-evolving control region. Furthermore, pairwise estimates of $N_{ef}m$ in these subspecies were impossible as the posterior probability surfaces were so ill-conditioned that no modes were present. This, along with negative

F_{ST} values, could be interpreted as either a product of shared ancestral polymorphisms, or as infinite amounts of gene flow.

The time (in generations) needed for F_{ST} to reach half way from an old to a new equilibrium is $t(0.5) = \ln(1/2)/\ln[(1-m)^2(1-1/2N_e)]$ (Whitlock 1992). Substituting my estimates of N_{ef} and m for my most divergent groups of knots, *C.c. canutus* and *C.c. islandica*, produced a time of 12,886 years to reach half way to equilibrium. Because the relationship between F_{ST} and time to genetic equilibrium is nonlinear, the total time to equilibrium is more than twice the time calculated above. Using the curve in Figure 3 of Whitlock and McCauley (1999), I can extrapolate that the time for F_{ST} to reach 95% equilibrium for these subspecies is approximately 130,000 years. This is much longer than any of my estimated population divergence times. Clearly knot subspecies are not in equilibrium and their lack of genetic distinctiveness and the polyphyly in their gene tree is most likely due to shared ancestral polymorphisms and not high levels of gene flow.

Contrary to knots, Dunlins exemplify a species which has probably reached genetic equilibrium. The substitution of my estimates of N_{ef} and m for the Canadian and Siberian lineages in Dunlins into Whitlock's (1992) equation produces a time halfway to equilibrium of 46,350 years. The time to 95% equilibrium can be extrapolated to approximately 200,000 years (Whitlock and McCauley 1999), which is close to my estimated population divergence time for these two lineages. These calculations indicate that Dunlins are most likely in genetic equilibrium and that enough time has passed for monophyly among lineages in this species. Because dunlin lineages are in genetic equilibrium, the small amount of mixing found in their haplotype network should be

interpreted as a small number of contemporary migrants and not as shared ancestral polymorphism. The mixing between the European and Siberian phylogroups is probably caused by occasional migrants in the current zone of overlap between these groups see (Wenink et al. 1996; Wennerberg et al. 1999 for discussion).

While the six currently recognized subspecies of knots cannot be distinguished by their control region sequences, they are beginning to sort and it would be inadvisable to lump them into a single evolutionary unit on genetic grounds alone. Given the passage of time, evolution of neutral genetic markers in knots should catch up with morphological and behavioral characters such as size and plumage differences, different migration routes, separate breeding grounds, and different molt schedules to more clearly distinguish subspecies (Davidson and Wilson 1992; Morrison and Harrington 1992; Nebel et al. 2000; Piersma and Davidson 1992; Tomkovich 1992; Tomkovich 2001).

Reconstructing Historical Biogeography

This study has examined the different demographic histories of knots and Dunlins in terms of classic phylogeographic methods. Additionally I obtained broad population divergences dates from coalescent analysis which I integrate at this point to attempt biogeographic reconstructions first in knots and then in Dunlins. My reconstructions are based on both my dates and existing palaeovegetation literature documenting possible vegetation and landform zones during the Pleistocene (Adams and Faure 1997; Levin 1992; Pielou 1991).

Coalescent analyses in knots show four emergent lineages which diverged approximately 20,000, 12,000 and 6,500 years ago, indicating that all subspecies probably originated near or after the glacial maximum of the Wisconsinan glaciation 22,000 to 18,000 years ago. Figure 3-5 presents a series of maps depicting the earth's approximate vegetation cover at 20,000, 12,000, 8,000 and 1,000 years ago. These maps are schematic reconstructions taken from much more detailed maps compiled from the synthesis palaeovegetation research by the Quaternary Environments Network (Adams and Faure 1997). During the last glacial maximum (LGM), ice sheets covered most of Europe and North America, and genetic evidence indicates that any lineages that may have existed in these areas previously were wiped out. Thus, I hypothesize that all existing populations of knots originated from a refugial group breeding in the unglaciated areas of tundra in eastern and central Eurasia. At the height of the LGM, a swath of polar desert (< 4% vegetation) separated the tundra possibly splitting the ancestral population into two groups, one breeding in central Eurasia and possibly migrating to west Africa (the current *C.c. canutus* wintering area) and one breeding in Beringia and possibly migrating to northern Australia (the current *C.c. piersmai* wintering area; Figure 3-5a).

Control region sequences indicate the divergence of a North American lineage approximately 12,000 years ago. Initial warming into the present interglacial began shortly after the LGM, and gradually melted the ice in North America to establish an ice free corridor between the Laurentide and the Cordilleran ice sheets (Adams and Faure 1997; Pielou 1991). During this initial warming period I hypothesize that the Beringian population of knots entered North America, and when the ice-free corridor

opened about 12,000 ybp, they engineered a new and shorter migratory route across North America to wintering grounds on the Gulf of Mexico. This new tradition probably offered a selective advantage (Pienkowski and Evans 1985; Helbig 1996), as it would require less energy expenditure and migratory risk than the much longer ancestral route to Australia that is still used by knots breeding in eastern Siberia today. This new migratory route behaviorally isolated the North American population from the eastern Siberian group. Furthermore, cooling during the Younger Dryas (also approximately 12,000 ybp) may have isolated these groups geographically as polar desert once again crept south (Figure 3-5b). *C.c. roselaari* birds still follow a cross-continental migratory route today, although the exact details of the route and the breeding grounds remain to be studied in detail. This hypothesis may explain why these birds do not follow a more straightforward migration route to breeding grounds in the central Canadian arctic from the Gulf shores. These birds may have been behaviorally or evolutionarily programmed to follow this cross-continental route when great ice sheets still existed in North America.

The split between *C.c. piersmai* and *C.c. rogersi* occurred approximately 6,500 years ago, soon after Eurasia reached the hypsithermal of this current interglacial (approximately 8,000-9,000 years ago; Adams and Faure 1997). During the hypsithermal, woodland crept northward over-running knot breeding habitat and covering the Chukotski peninsula, where *C.c. rogersi* birds currently breed. I propose that after the earth began to cool towards present temperatures, woodland receded and the eastern

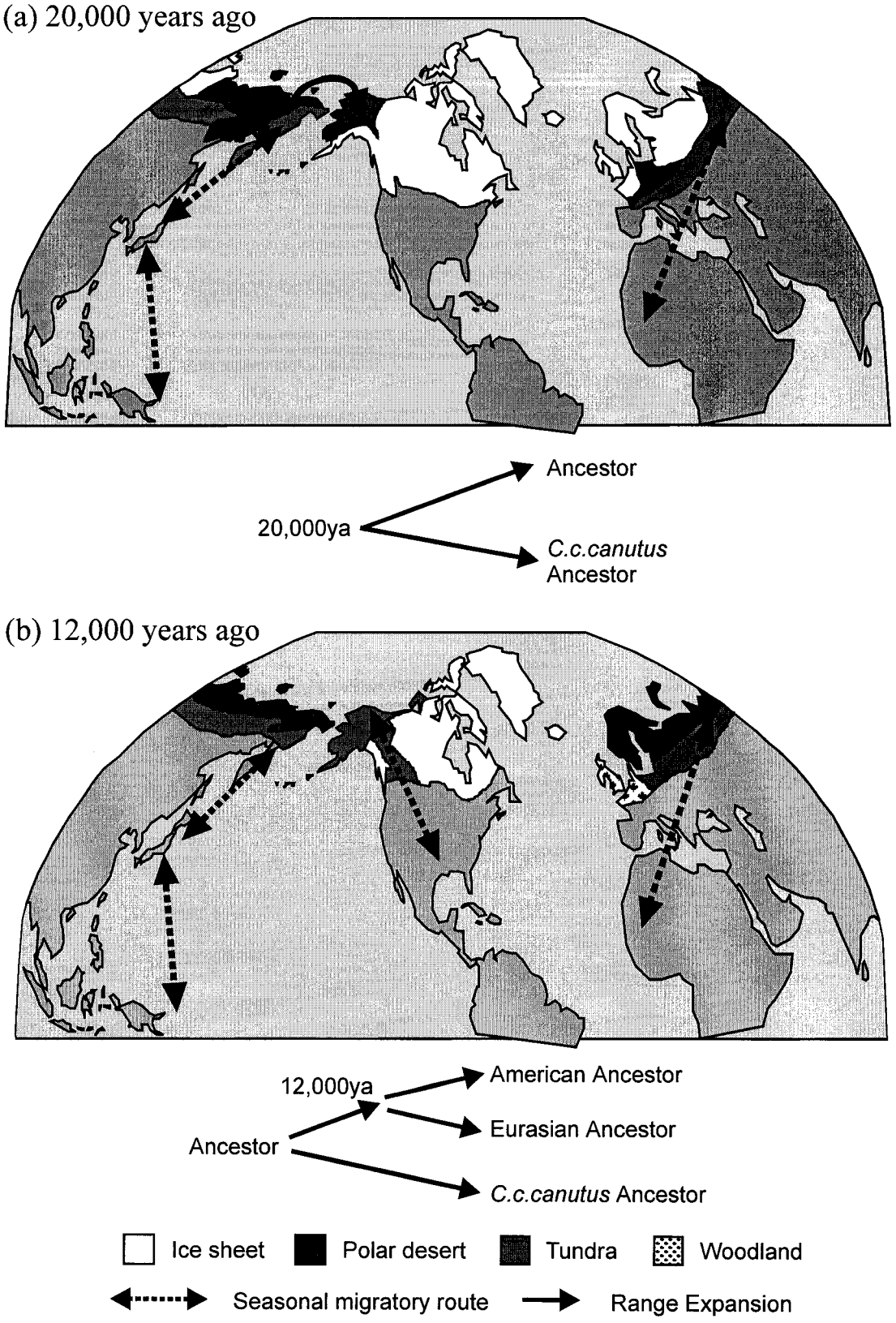


Figure 3-5: A proposed reconstruction of knot biogeographic history presented as a series of maps depicting the earth's approximate vegetation cover at 20,000, 12,000, 8,000 and 1,000 years ago. These maps are schematic reconstructions taken from much more detailed maps compiled from the synthesis palaeovegetation research by the Quaternary Environments Network (Adams and Faure 1997). Solid arrows represent ancient range expansions and dashed arrows represent proposed seasonal migration routes. The phenogram below each map shows the population divergence associated with that time period.

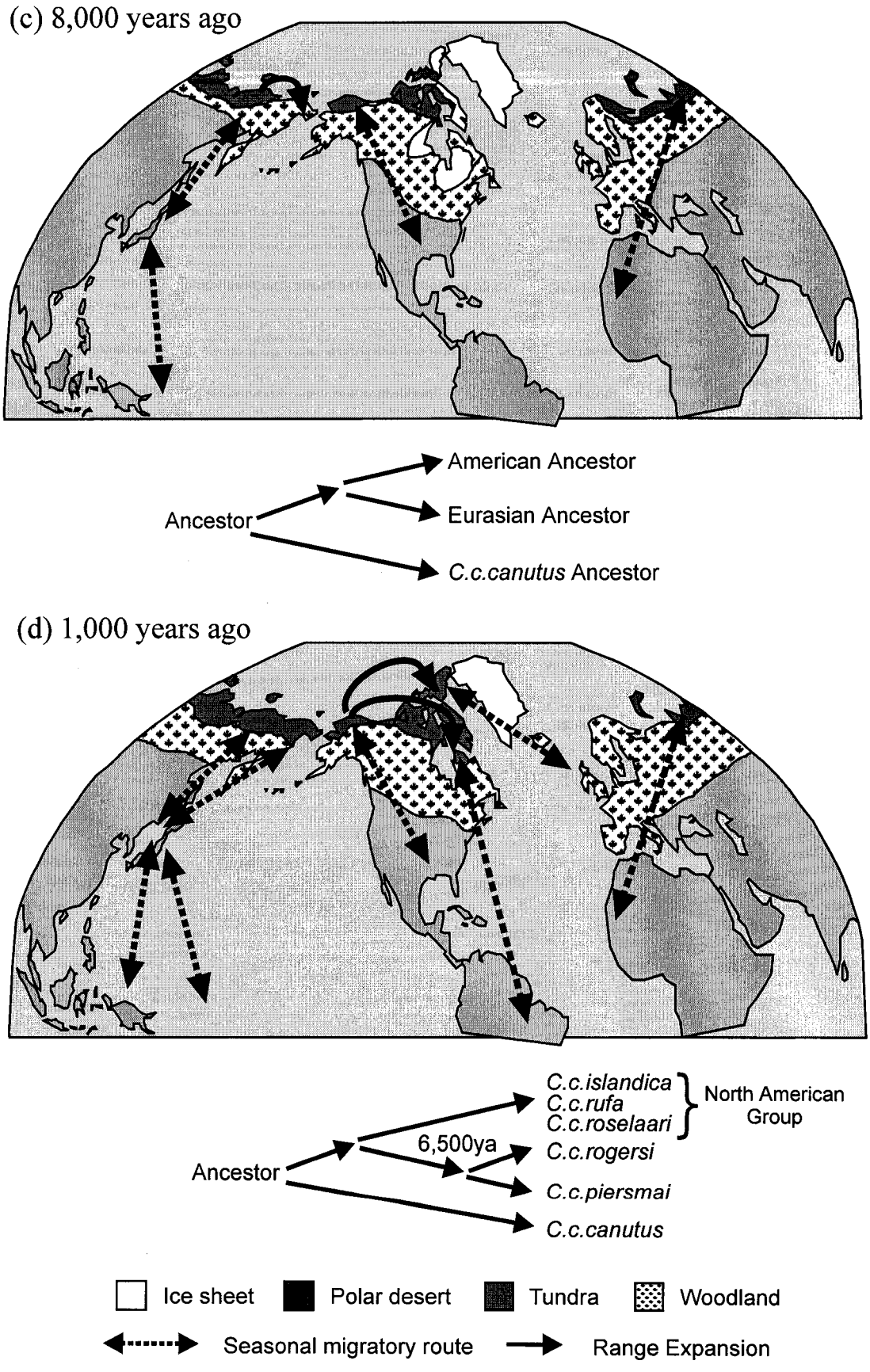


Figure 3-5 continued

Eurasian knot population split into the subspecies now known as *C.c. piersmai* and *C.c. rogersi* (Figure 3-5c).

Very recently, as tundra vegetation colonized barren land left by receding ice sheets in the Canadian Arctic, knots may have expanded their range from northern Alaska east across the lower and upper Arctic, instigating the divergence of *C.c. roselaari, C.c. rufa* and *C.c. islandica*. I propose that knots reaching breeding grounds in the Southampton Island area migrated south in winter via James Bay and the Gaspé Peninsula, and out over the Atlantic Ocean and then the Caribbean Sea on favorable winds which facilitated migration to South America (Butler et al. 1997; Gauthreaux 1991; National Geographic Society 1992). This hypothesis might explain the chain migration pattern presented by *C.c. roselaari* and *C.c. rufa* birds today. *C.c. roselaari* birds breed and winter further north than do *C.c. rufa* birds, perhaps because the *C.c. roselaari* migratory route was established while ice sheets still existed 12,000 years ago. The newer *C.c. rufa* route was engineered more recently from breeding grounds further to the east, made available after the ice retreated.

Knots currently considered *C.c. islandica* probably expanded their breeding range to the north and east as ice-free habitat emerged in the high arctic islands and Greenland. From these breeding areas, migration distances to northwest Europe via Iceland were considerably shorter than migration distances to South America. Some knots likely joined European counterparts, who breed in Greenland and winter in Europe, and any birds pioneering this new route were likely at a selective advantage in terms of energetics

and migratory risk (Pienkowski and Evans 1985; Helbig 1996), thus establishing a new flyway.

The hypothesis of sequential expansion eastward from a refugial population that survived the LGM breeding in the unglaciated regions of central and eastern Eurasia is consistent with the loss of nucleotide diversity in daughter populations. In knots, the *C.c. canutus* group shows the highest level of nucleotide diversity, consistent with its basal divergence. *C.c. roselaari* and *C.c. piersmai* have intermediate levels of genetic diversity as expected by their intermediate divergence times. Finally, *C.c. rogersi*, *C.c. islandica* and *C.c. rufa* have the lowest levels of diversity indicating that they diverged most recently, and have not had sufficient time to recover genetic variation lost through founder effects.

Dunlins have a demographic history that is very different from that of knots with coalescent times dating back approximately 180,000 years. This indicates that Dunlins were last bottlenecked to a single ancestral population during the Illinoian glaciation 200,000 to 130,000 years ago (Levin 1992). For Dunlins much of the evidence for revegetation after the Illinoian glaciation (the Sangamon interglacial) was erased when the ice sheets of Wisconsinan glaciation covered the land. For this reason the graphic illustration of dunlin biogeographical history does not included reconstructed vegetation maps (Figure 3-6).

From Table 3-6 one can see that divergence times around 180,000 years ago are between the Canadian lineage and the ancestor of all other lineages. During the interglacial, ancestral populations of Dunlins probably had a fairly widespread range.

The Illinoian glaciation was likely the vicariant event that split this population into Canadian and Eurasian lineages. If the ice sheets of the Illinoian glaciation resembled those of theWisconsinan glaciation, then these populations may have survived with widely separated breeding areas in the unglaciated tundra of Eurasia and a small refugial patch of tundra southeast of the North American ice sheets.

The next major divergence between dunlin populations occurred approximately 110,000 years ago between the ancestral populations of the European lineage and the Siberian, Beringian and Alaskan lineages. This divergence time roughly coincides with rapid cooling into the Wisconsinan glaciation following the Sangamon interglacial (Adams and Faure 1997). If Dunlins had expanded their range into Europe and Beringia during the interglacial, then the return of the ice sheets to North America and northern Europe probably fragmented this range into separate breeding populations southeast of the ice in North America, south of the ice in Europe, and west of the ice in Beringia. The final split between the Siberian, Alaskan and Beringian lineages occurred approximately 68,000 years ago. This date corresponds to a period in Eurasia in which conditions resembled those of a glacial maximum (Adams and Faure 1997). During this period dunlin populations in Beringia were probably forced south and may have bred in separate areas. These divisions were probably strengthened during the LGM 22,000 - 18,000 years ago.

My reconstructions of biogeography are hypotheses based on admittedly imprecise population divergence times that have been affected by stochastic sampling error and inaccuracies of the molecular clock. Nevertheless, the divergence times

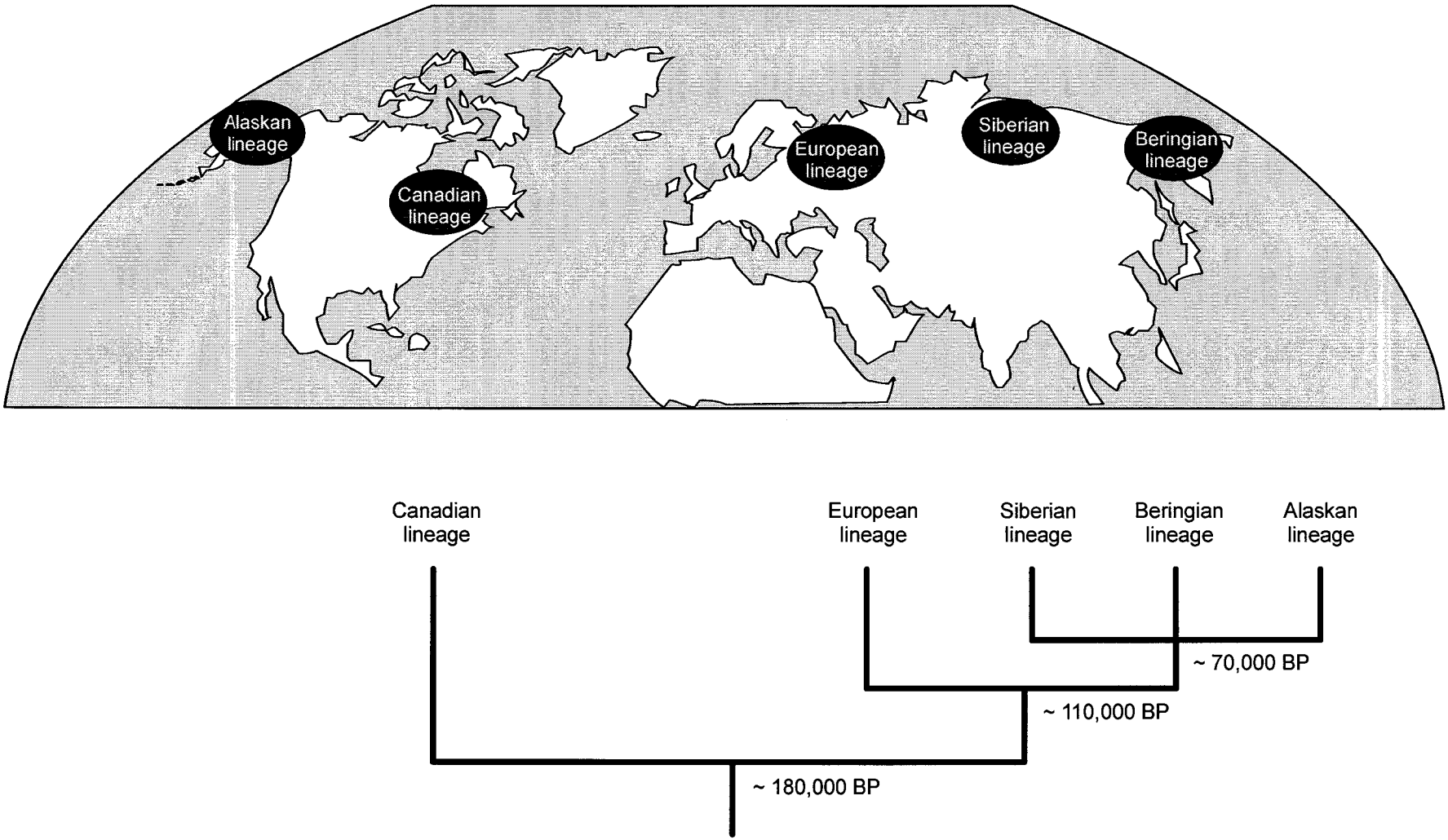


Figure 3-6: A schematic reconstruction of dunlin biogeographic history. The ovals representing populations are not meant to depict the locations of ancient populations, but are instead placed broadly around current breeding locations. Population divergence times are marked on the nodes of the phenogram below the map and date major vicariate events.

correlate well with the approximate timing of major events during the late Pleistocene. These hypotheses may be further tested by the examination of biogeography in other species with demographic histories similar to knots or Dunlins. This cross-species examination would provide either support for my hypotheses through concordance or alternative explanations. Another important test of concordance would be the examination of other genes within knots and Dunlins. I present data for the mitochondrial control region, which is only a small portion of the genome. This test may prove to be difficult however, as genes with comparable rates of evolution will be needed to resolve intraspecific groupings, especially in knots.

This study builds upon classic phylogeographic analysis of gene trees through the addition of population divergence dates using coalescent analysis. These dates along with palaeovegetation information allow biogeographic reconstructions that are more comprehensive than what was possible in the past. This type of in-depth analysis of knots and Dunlins offers two snapshots of the process of genetic evolution: one after 20,000 years and one after nearly 200,000 years.

Acknowledgements

An NSERC PGS A Scholarship (PGSA-231177-2000) to DMB and grants from NSERC (200-02) and the Royal Ontario Museum Foundation to AJB supported this research.

References

Adams, J. M., and H. Faure. 1997. Review and atlas of palaeovegetation: Preliminary land ecosystem maps of the world since the last glacial maximum. Oak Ridge National Laboratory, TN, USA, Quaternary Environments Network (QEN).

Avise, J. C. 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall, New York.

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

Baker, A. J. in press. The deep roots of bird migration: inferences from the historical record preserved in DNA. Evolution

Baker, A. J., T. Piersma, and L. Rosenmeier. 1994. Unraveling the intraspecific phylogeography of knots (*Calidris canutus*) - A progress report on the search for genetic markers. Journal fur Ornithologie 135:599-608.

Bulgin, N. L., H. L. Gibbs, P. Vickery, and A. J. Baker. in press. Ancestral polymorphisms in genetic markers obscure detection of demographically distinct populations in the endangered Florida grasshopper sparrow (*Ammodramus savannarum floridanus*). Molecular Ecology

Butler, R. W., T. D. Williams, N. Warnock, and M. A. Bishop. 1997. Wind assistance: a requirement for migration of shorebirds? Auk 114:456-466.

Davidson, N. C., and J. R. Wilson. 1992. The migration system of European-wintering knots *Calidris canutus islandica*. Wader Study Group Bulletin 64 Supplement:39-51.

Fedorov, V. B., and N. C. Stenseth. 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 269:2071-2077.

Fu, Y. X., and W. H. Li. 1999. Coalescing into the 21st Century: An Overview and Prospects of Coalescent Theory. Theoretical Population Biology 56:1-10.

Gauthreaux, S. A. J. 1991. The flight behaviour of migrating birds in changing wind fields: radar and visual analyses. American Zoologist 31:187-204.

Griffiths, R. C., and S. Tavaré. 1997. Computational methods for the coalescent. Pp. 165-182 *in* P. Donnelly and S. Tavaré, eds. Progress in Population Genetics and Human Evolutions. Springer-Verlag, New York.

Griswold, C., and A. J. Baker. 2002. Time to most recent common ancestor and divergence times of populations of common chaffinches (*Fringilla coelebs*) in Europe and North Africa: insights into Pleistocene refugia and current levels of migration. Evolution 56:143-153.

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

Helbig, A. J. 1996. Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palearctic warblers (Aves: Sylviidae). Journal of Experimental Biology 199:49-55.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Oxford Survey of Evolutionary Biology 7:1-44.

Kingman, J. F. C. 1982. The coalescent. Stochastical Procedure Applications 13:235-248.

Levin, H. L. 1992. The Earth Through Time. Saunders College Publishing, Toronto.

Marshall, H. D. 1997. Molecular Evolution and Phylogeography of Common Chaffinches (*Fringilla coelebs*). Ph.D thesis. Department of Zoology. University of Toronto, Toronto.

Morrison, R. I. G., and B. A. Harrington. 1992. The migration system of the red knot *Calidris canutus rufa* in the new world. Wader Study Group Bulletin 64 Supplement:71-84.

Mullis, K. B., and F. A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. Methods in Enzymology 115:335-350.

National Geographic Society. 1992. Atlas of the World. National Geographic Society, Washington, D.C.

Nebel, S., T. Piersma, J. Van Gils, A. Dekinga, and B. Spaans. 2000. Length of stopover, fuel storage and a sex-bias in the occurrence of red knots *Calidris canutus canutus* and *C.c. islandica* in the Wadden Sea during southward migration. Ardea 88:165-176.

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

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

Pielou, E. C. 1991. After the Ice Age: The return of life to glaciated North America. University of Chicago Press, Chicago.

Pienkowski, M. W., and P. R. Evans. 1985. The role of migration in the population dynamics of birds. Pp. 331-352 *in* R. M. Sibly and R. H. Smith, eds. Behavioral ecology: Ecological consequences of adaptive behavior. Blackwell, Oxford.

Piersma, T., and N. C. Davidson. 1992. The migrations and annual cycles of five subspecies of knots in perspective. Wader Study Group Bulletin 64 Supplement:187-197.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

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

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA Sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the U.S.A. 74:5463-5467.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555-562.

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

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10:512-526.

Tomkovich, P. S. 1992. An analysis of the geographic variability in knots *Calidris canutus* based on museum skins. Wader Study Group Bulletin 64 Supplement:17-23.

Tomkovich, P. S. 2001. A new subspecies of Red Knot *Calidris canutus* from the New Siberian Islands. Bulletin of the British Ornithologists Club 121:257-263.

Wenink, P. W., and A. J. Baker. 1996a. Mitochondrial DNA lineages in composite flocks of migratory and wintering dunlins (*Calidris alpina*). Auk 113:744-756.

Wenink, P. W., A. J. Baker, H. U. Rösner, and M. G. L. Tilanus. 1996b. Global mitochondrial DNA phylogeography of holarctic breeding dunlins (*Calidris alpina*). Evolution 50:318-330.

Wenink, P. W., A. J. Baker, and M. G. L. Tilanus. 1993. Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpina*). Proceedings of the National Academy of Sciences of the U.S.A. 90:94-98.

Wennerberg, L., N. M. A. Holmgren, P.-E. Jvnsson, and T. von Schantz. 1999. Genetic and morphological variation in Dunlin *Calidris alpina* breeding in the Palearctic tundra. Ibis 141:391-398.

Whitlock, M. C. 1992. Temporal fluctuations in demographic parameters and the genetic variance among populations. Evolution 46:608-615.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: F_{ST} is not equal to 1/(4Nm + 1). Heredity 82:117-125.

Appendix 3-1: Sample locations, dates and sample sizes for knots used in this study.

Location Collected	Date Collected	Subspecies	*N*
South Carolina, USA	16-Apr-00	*C.c.roselaari*	6
Georgia, USA	17-Oct-99	*C.c.roselaari*	9
Tierra del Fuego, Argentina	4-Dec-00	*C.c.rufa*	8
Southampton Island, Canada	5 to 7-Jul-00	*C.c.rufa*	7
Broome, Western Australia, Australia	24-Mar-96	*C.c. piersmai*	15
Ellesmere Island, Canada	01-Jun-90	*C.c.islandica*	9
Dutch Wadden Sea, Netherlands	27-Oct-95	*C.c.islandica*	1
Dutch Wadden Sea, Netherlands	8 to 9-Feb-97	*C.c.islandica*	5
Taymyr Peninsula, Russia	6 to 10-Jul-94	*C.c.canutus*	6
Cape Province, South Africa	30-Nov-91	*C.c.canutus*	6
Barry Beach, Victoria, Australia	2-Apr-02	*C.c.rogersi*	18
Corner Inlet, Victoria, Australia	25-Jul-02	*C.c.rogersi*	1

Appendix 3-2: Variable sites among 25 knot haplotypes. Variation is represented with respect to the first haplotype listed and identities are indicated by dots. Sites are numbered according to the complete control region sequence presented in Chapter 2. Haplotype names are derived from the subspecies predominantly represented: Rog, *C.c. rogersi*, Ruf, *C.c. rufa*, Isl, *C.c. islandica*, Pie, *C.c. piersmai*, Can, *C.c. canutus*.

			111
		112233333344	778889011
		395813467902	780299902
		887528969530	827799844
Haplotype	N		
1 Rog1	35	TTGTCTGTCACA	CATATTAGG
2 Rog2	3	..A.....	.....
3 Rog3	1	.C.....	......G..
4 Rog4	1	.....	.....A
5 Rog5	1	.....C....	.....
6 Ruf1	3	......A.....	.....
7 Ruf2	1	.....T...	.....
8 Isl1	18	.....	.....C...
9 Isl2	1	.....	.G.....
10 Isl3	1	......A.....	....C....
11 Ros1	1	A.....	.G.....
12 Ros2	1	A.....	.GC.....
13 Ros3	1	.....	T.....
14 Pie1	3	.....C.....	.....
15 Pie2	1	.....T.	...G.....
16 Pie3	1	...C.....	.....
17 Pie4	2	.....	.....A.
18 Pie5	1	.....G..	.....
19 Can1	9	.....	...G.....
20 Can2	1	....T.....	...G.....
21 Can3	1	......A....G	...G.....
22 Can4	1	.....G	...G.....
23 Can5	1	......A.....	...G.....
24 Can6	1	.....T.	.....
25 Can7	1	.C.....	...G.....

Appendix 3-3: Variable sites among 53 dunlin haplotypes. Variation is represented with respect to the first haplotype listed and identities are indicated by dots. Sites are numbered according to the dunlin control region sequence (Wenink et al. 1993). Haplotype names correspond to those in Wenink et al. 1996a and 1996b.

		111111122222222222222333333333333 556667777773777
		257799901222225555778222223345667 570661112233444
		303836817013475789569345674948127 542780290708489
Haplotype	N	
1 ALA1	12	AACATCACAGTACACCTGCTATTAACCTTCGAG GCCT-AATCCAGTGT
2 ALA2	2	...G..... .....
3 ALA3	10	.....A..... .....
4 ALA4	4	..T.....A..... .....G.....
5 ALA5	3	.....A.....G..... .....
6 ALA6	1	..T..... .....
7 ALA7	1	G.T.....A..... .....G.....
8 BER1	2	......G..A....T..A..G.....T....GA .....G.....
9 BER2	1	......G..A....T..A..G.....T....GA .....G...T.....
10 BER3	1	......G..A.......A..G.....GA .....G..T.....
11 BER4	1	....C.G..A....T..A..G.....T....GA .....G.....
12 BER5	1	G.....A....T..A..G.....GA .....G.....
13 BER6	1	G.....G..A....T..A..G.....A .....G.....
14 SIB1	10	......G..A.....C..C..T..... .....G..T.....
15 SIB2	9	......G.....C..C..T..... .....G..T.....
16 SIB3	1	......G..A.....C..... .T...G..T.....
17 SIB4	1	......G..A.....C.... .....G..T.....
18 SIB5	12	......G....G.....C.....TT..... .....G..T.....
19 SIB6	7	.G....G..A.G.....C.....TT..... .....G..T.....C
20 SIB7	1	.G....G..A.G.....G. .....G..T.....C
21 SIB8	1	......G..A.....T.C.... .....G..T.....
22 EUR1	57	..T.....A.....AC...G.....CTA.A .....G..T.....
23 EUR2	2	..T...G..A.....AC...G.....CTA.A .....G..T.....
24 EUR3	1	..T....T.A.....AC...G.....CTA.A .....G..T.....
25 EUR4	1	..T.....A...G.AC...G.....CTA.A .....G..T.....
26 EUR5	1	..T.....A.....AC...G....T..CTA.A .....G..T.....
27 EUR6	3	..T.....A.....AC...G.....CTA.. .....G..T.....
28 EUR7	1	..T.....AC..CG.....CTA.A .....G..T.....
29 EUR8	1	..T.....AC...G...G...CTA.A .....G..T.....
30 EUR9	1	.....A.....AC...G.....CTA.A .....G..TT.....
31 EUR10	2	..T.....A...G.AC...G.....CTA.. .....G..T.....
32 EUR11	4	..T.....A.....AC.....T..CTA.A .....G..T.....
33 EUR12	11	..T..T.T.A.....AC...G.....CTA.A .....G..T.....
34 EUR13	1	..T..T.T.A.....AC...G.....T.CTA.A .....G..T.....
35 EUR14	1	..T..T.T.A.....AC...G.....CTA.. .....G..T.....
36 EUR15	1	..T..T.T.A....TAC..CG.....CTA.A .....G..T.....
37 EUR16	5	..T.....A.....AC...G.....CTA.A .....G-CT.....
38 EUR17	2	..T...G..A.....AC...G.....CTA.A .....G-CT.....
39 EUR18	1	......G..A.....AC...G.....CTA.A .....G-CT.....
40 EUR19	6	..T...G.....AC...G.....CTA.A .....G-CT...G..

Appendix 3-2 continued

```
41 EUR20    1     ..T.............AC...G.......CTA.. .....G-CT...G..
42 EUR21    1     ..T......A.....AC...........CTA.A .....G..T......
43 EUR22    1     ..T......A.....AC...G...G...CTA.A .....G..T......
44 EUR23    1     .........A.....AC...G.....T.CTA.A .....G..T......
45 EUR24    1     ..T......A.....AC...G.......CTA.A .....G..T....AG
46 EUR25    1     ..T..TGT.A.....AC...G.......CTA.A .....G..T......
47 EUR26    1     ..T..T.T.A.....AC...G..G....CTA.A .....G..T......
48 EUR27    1     ..T......A.....AC...G.......CTA.A ..A..G-CT......
49 CAN1     4     ........TACCT..GCAAC.C....TAC.A.A A..CGG..T.GA...
50 CAN2     6     ........TACCT..GCAAC.C....TAC.A.A A..CGG..T..A...
51 CAN3     1     ........CACCT..GCAAC.C....TAC.A.A A..CGG..T..A...
52 CAN4     1     ........CACCT..ACAAC.C....TAC.A.A A..CGG..T.GA...
53 CAN5     4     ........TACCT..ACAAC.C....TACTA.A A..CGG..T.GA...
```

CHAPTER 4

Genetic diversity in Red Knots (*Calidris canutus*): Control region sequences, AFLPs and microsatellites reveal an avian cheetah?

Abstract

I examine genetic diversity in Red Knots (*Calidris canutus*) using control region sequences, amplified fragment polymorphisms (AFLPs) and microsatellites. I explore nuclear markers for subspecific structuring and compare levels of genetic diversity in knots with that in other bird species. My assessment of genetic variation showed that knots have very low diversity in comparison to other bird species, and that the frequency based approach used to assess relationships among subspecies was incongruent with the tree derived from control region sequences. Low diversity was found in both the mitochondrial and nuclear genomes, supporting the hypothesis that knots have undergone a recent bottleneck as opposed to a selective sweep. Due to the low levels of diversity, akin to those found in cheetahs, I recommend that populations of knots and the fragile chain of major stopover sites in their flyways continue to be protected. Habitat loss along flyways will only exacerbate the problem of low genetic diversity, as knots do not have the evolutionary resources to respond to changes in their highly specialized migratory routes and annual cycle.

Introduction

Genetic diversity within a species is the foundation of evolutionary change, providing the raw material upon which adaptation and diversification are molded. Low

levels of genetic diversity limit the ability of species and their populations to respond to short and long term threats such as disease, habitat loss and climate change (O'Brien et al. 1985; Webster et al. 2002). The cost of genetic uniformity is clearly illustrated in cheetahs through the disastrous and homogeneous sensitivity of animals from the Oregon colony to feline infectious peritonitis (O'Brien et al. 1985). Simply put, diversity allows species to adapt, while uniformity leaves species vulnerable to threats and ultimately extinction.

Red Knots (*Calidris canutus*) are long-distance migrant shorebirds. They breed in the high arctic and each year embark on migrations of up to 15,000 km to marine staging and wintering areas at different latitudes throughout the globe. Knots are distributed into six morphologically recognized subspecies: *C.c. roselaari*, *C.c. rufa*, *C.c. rogersi*, *C.c. piersmai*, *C.c. islandica* and *C.c. canutus* (Figure 4-1; Tomkovich 1992; Tomkovich 2001). As a species they are a conservation concern for a number of reasons. They are habitat specialists that are entirely coastal in the non-breeding season, and are dependent on a few essential staging areas that harbor abundant mollusk prey during their long migrations. In these staging areas the birds congregate in compact flocks making them vulnerable to local disasters such as oil spills or food chain collapse due to over-harvesting. On a global scale knots are vulnerable to climate change, which threatens both their high arctic breeding grounds through warming and their coastal non- breeding areas through rising sea levels.

Coalescent analysis of control region sequences in mitochondrial DNA (mtDNA) revealed that knots were severely bottlenecked approximately 20,000 years ago and as a

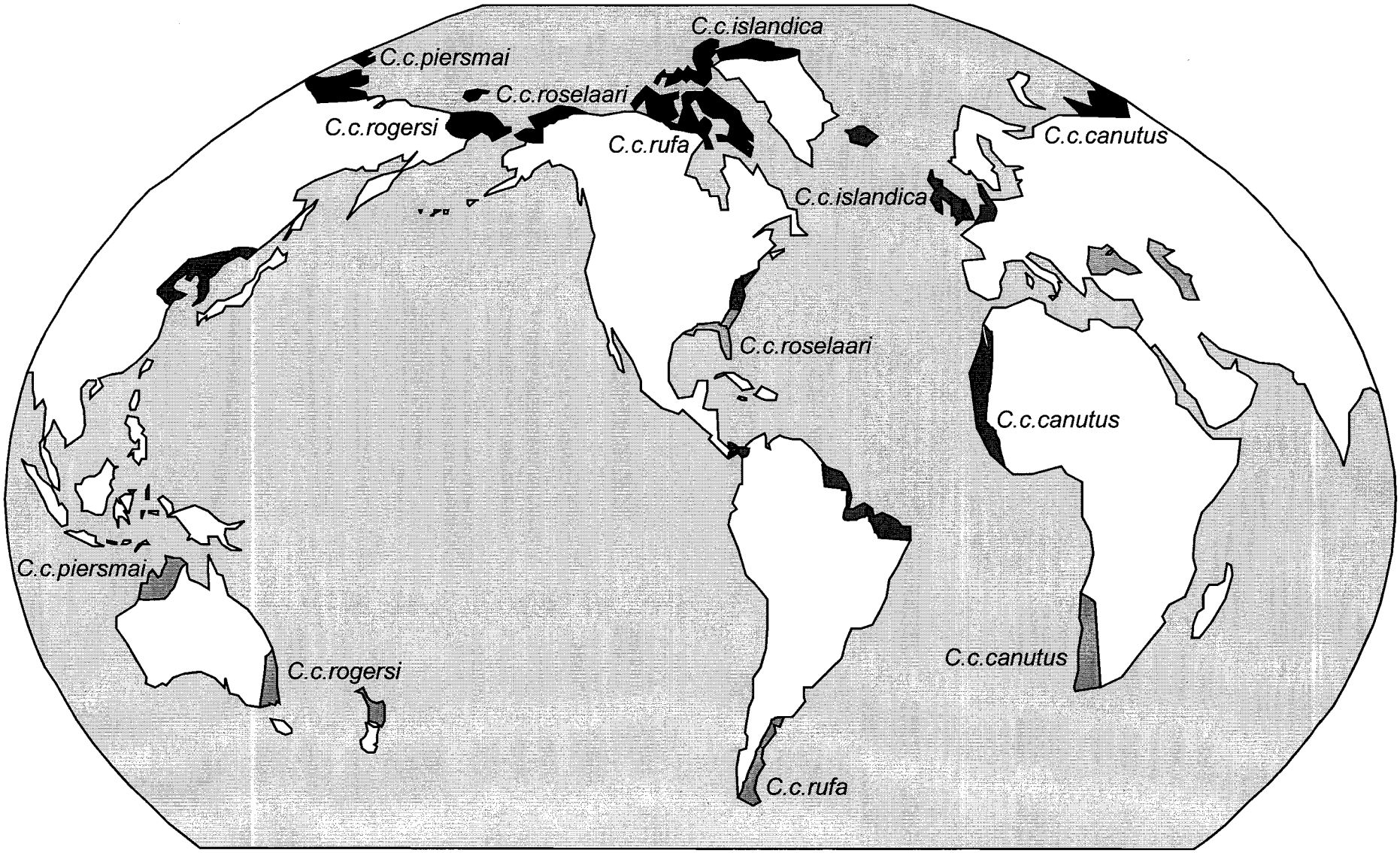


Figure 4-1: The global distribution of knots. Black shading indicates breeding areas, medium grey shading staging/wintering areas and light grey wintering areas. The term wintering refers to the northern hemisphere winter and the austral summer.

result are genetically depauperate relative to other shorebirds such as Dunlins (Chapter 3). The effect of the bottleneck is expected to be most dramatic in maternally inherited mtDNA as the effective population size is one quarter of that for biparentally inherited and diploid nuclear genes (Page and Holmes 1998). Theory predicts that biparentally inherited nuclear genes will maintain a good store of genetic variation in a species unless the population contraction was severe and prolonged. This was the case in African cheetahs (*Acinonyx jabatus*) which show a dramatic reduction in genetic variation in mtDNA, but a less severe depletion of variation in microsatellites representing the nuclear genome (Freeman et al. 2001; Menotti-Raymond and O'Brien 1995; O'Brien et al. 1985).

The objective of this study is to compare levels of genetic diversity in the fast-evolving part of the mtDNA genome (the control region) with rapidly mutating dinucleotide microsatellites, and with amplified fragment length polymorphisms (AFLPs) that survey variation across the nuclear genome. Specifically I address the question, are Red Knots the avian equivalent of the genetically impoverished cheetah?

Materials and Methods

Sample Collection

Knots were sampled globally using ground traps in breeding areas and cannon nets in wintering and staging areas (Table 4-1). A few drops of blood were taken from the brachial vein of each bird and stored at room temperature in 50mM EDTA and 70%

Table 4-1: Sample locations, dates and sample sizes for knots used in this study.

Location Collected	Date Collected	Subspecies	Control Region *N*	AFLP *N*	Microsatellite *N*
South Carolina, USA	16-Apr-00	*C.c.roselaari*	6	9	10
Florida, USA	30-Jan-86	*C.c.roselaari*	-	-	13
Georgia, USA	17-Oct-99	*C.c.roselaari*	9	7	11
Tierra del Fuego, Argentina	4-Dec-00	*C.c.rufa*	8	7	12
Southampton Island, Canada	5 to 7-Jul-00	*C.c.rufa*	7	8	21
Broome, Western Australia, Australia	24-Mar-96	*C.c. piersmai*	15	16	21
Broome, Western Australia, Australia	18-Mar-94	*C.c. piersmai*	-	-	10
Ellesmere Island, Canada	01-Jun-90	*C.c.islandica*	9	10	10
Dutch Wadden Sea, Netherlands	27-Oct-95	*C.c.islandica*	1	-	1
Dutch Wadden Sea, Netherlands	7 to 9-Feb-97	*C.c.islandica*	5	1	7
Dutch Wadden Sea, Netherlands	15-Oct-88	*C.c.islandica*	-	3	8
Dutch Wadden Sea, Netherlands	21-May-97	*C.c.islandica*	-	-	2
Taymyr Peninsula, Russia	6 to 10-Jul-94	*C.c.canutus*	6	9	9
Cape Province, South Africa	30-Nov-91	*C.c.canutus*	6	5	11
Cape Province, South Africa	19-Oct-91	*C.c.canutus*	-	-	2
Bissau, Guinea, West Africa	1-Feb-93	*C.c.canutus*	-	-	8
Barry Beach, Victoria, Australia	2-Apr-02	*C.c.rogersi*	18	16	17
Corner Inlet, Victoria, Australia	25-Jul-02	*C.c.rogersi*	1	1	1
		Total	91	92	174

ethanol, then subsequently frozen at –80°C in the Royal Ontario Museum's Ornithology collection. For all subsequent procedures total DNA was isolated using standard phenol extractions (Sambrook et al. 1989).

Control Region Sequences

Control region sequences for 91 individuals representing six subspecies, were obtained as described in Chapter 3. Sequences were aligned using ClustalW (Thompson et al. 1994) and final alignment adjustments were made manually. Arlequin (Schneider et al. 2000) was used to calculate haplotypic diversity $(h) = (n/n\text{-}1)(1\text{-}\Sigma f_i^2)$, where f_i is the frequency if the ith haplotype and n is the number of individuals sampled, and nucleotide diversity $(\pi) = \Sigma\pi_{ij}/(n(n\text{-}1)/2)$, where π_{ij} equals the proportion of the nucleotide differences between the ith and the jth sequences, and n is the number of individuals (Nei 1987).

AFLPs

AFLP loci were isolated using the protocol described in Myburg et al. (2001) and modified with M13 tailed adaptors (C. Ritland, see GDC website http://www.forestry.ubc.ca/gdc/). I digested high molecular weight DNA with the restriction enzymes EcoRI and MseI. Each fragment was then ligated to adaptors that served as the binding site for selective primers with one to three bases added at the 3' end. The ligated DNA was amplified in a pre-selective PCR using primers with one selective base, and this product was diluted and amplified again in a selective PCR with

primers having three selective bases. The selective PCR products were run out on a LI-COR 4200 sequencer resulting in a scorable, multi-locus fingerprint-like pattern.

Thirty-two primer combinations were screened for polymorphic loci on a small subset of 10 knot individuals representing five subspecies. Twelve primer combinations amplified well and produced polymorphic loci. These combinations were selected and used to screen a sample of 92 knots and 10 Dunlins. Gels were scored manually based on the presence or absence of bands. Initially I scored all amplified fragments 50 to 700 bp in length as either monomorphic or polymorphic. From this large sample I scored a subset of 29 loci for knots and 30 loci for Dunlins in detail for each individual. The chosen loci produced intense bands on gels, and were reproducible. Statistical analyses were performed using the programs Tools for Population Genetic Analysis (TFPGA; Miller 1997) and Hickory (Holsinger et al. 2002).

Microsatellites

Microsatellites markers were isolated using an enrichment protocol (Hamilton et al. 1999). Approximately 5μg of high molecular weight genomic DNA was isolated and digested using restriction enzymes HaeIII and AluI to create a genomic library with fragments 400-1000 bp in length. The library was then enriched for GT and CT repeats using biotinylated oligonucleotides $(GT)_{15}$ and $(CT)_{15}$. Fragments were ligated into the vector pBluescript (Stratagene), transformed into *E. coli* XL1-Blue Electro competent (Stratagene) cells, and grown on agar plates overnight. Colonies were lifted onto nitrocellulose filter paper, washed, and once again ligated to the biotinylated

oligonucleotides. Positive clones were detected using an NEB phototope star detection kit and captured on X-ray film. X-ray film was lined up with the original plates and positive clones were picked and extracted using the Wizard Miniprep Kit (Promega). Clones were sequenced using the Sanger dideoxy chain termination method (Sanger et al. 1977) and either the Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit (USB) on 6% polyacrylamide gel or the Thermo Sequenase DYEnamic Direct Cycle Sequencing Kit (Amersham Pharmacia) on 4% polyacrylamide gel.

One hundred and thirty-six clones were sequenced revealing 38 microsatellites. Of these 14 were located too close to the vector sequence for primer design. Primers were designed for 24 of the positive clones and an M13 tail was added to one primer of each primer pair to allow fluorescent labeling during amplification reactions. Of the 24 primer sets designed, 17 were optimized. PCR reactions were carried out in a reaction volume of 12.5 μl containing 1.0 μl of DNA, 1.25 μl of 10X EH buffer, 0.5 units of Taq DNA polymerase (Qiagen), 5 mM dNTPs, 2.5 pmoles of each primer and 2.5 pmoles of the appropriate (forward or reverse) fluorescent label. The amplification protocol was a 2 minute denaturation at 94°C, followed by 36 cycles of 94°C for 45 sec, 50°C for 45 sec, 72°C for 130 sec, with a final extension at 72°C for 5 min (Perkin-Elmer 480 thermal cycler). Amplification products were run out on a LI-COR 4200 sequencer and scored manually.

Of the 17 optimized microsatellite loci, only three proved to be polymorphic. These three loci were used as markers in a sample of 174 knots. Departure from Hardy-Weinberg equilibrium, observed and expected heterozygosities, and F_{ST} values were

obtained using the program Arlequin (Schneider et al. 2000). UPGMA, neighbor-joining and maximum-likelihood trees were created and bootstrapped using the SeqBoot, ContML, GenDist and ConSense subroutines in the program Phylip (Felsenstein 1993).

Species Comparisons

To compare knot diversity with that of other bird species, genetic diversity data were obtained from various published studies and a number of personal communications. This was relatively simple for control region sequences as genetic diversity measures are standard, and numerous studies containing these data have been published for birds.

Studies using AFLPs in birds were more difficult to obtain, and I chose to use only the percentage of polymorphic loci found during screening as a comparable measure of species level diversity. AFLPs are dominant markers. This means that the AA and Aa genotypes cannot be distinguished on gels because each appears as a single band. Allele frequencies and heterozygosities must therefore be inferred using available methods (Lynch and Milligan 1994; Weir 1990). The major problem with estimates of heterozygosity at AFLP loci is that they are generally chosen for their value as a population marker. By definition a population marker will be nearly fixed (all present or all absent). This makes within-group heterozygosities very low and biases estimates of total heterozygosity downwards. Lynch and Milligan (1994) proposed a method for inferring genotype frequencies that attempts to correct for this problem by restricting analysis to loci that show an observed frequency smaller than 1-(3/N), where N is the sample size. Although this correction does increase heterozygosities, I feel that

heterozygosities will still be biased when randomly amplified loci are not scored. Because so few avian studies using AFLPs exist in the literature, and many authors have preferentially scored population marker loci, I omit heterozygosity as a measure of diversity in AFLPs.

For microsatellites I used the percentage of polymorphic loci found during development as a measure of diversity, rather than the upwardly biased estimate obtained from heterozygosities of polymorphic loci alone. For this reason only studies that developed microsatellites directly were used for comparison. Observed and expected heterozygosities were also calculated and could be obtained from the literature for other bird species. These average heterozygosities however, are only representative of polymorphic loci. To get an idea of overall genomic levels of heterozygosity, monomorphic loci were considered as follows. By definition monomorphic loci have zero heterozygosity, so I calculated average overall genomic heterozygosity by averaging expected heterozygosities for polymorphic loci and multiplying this value by the proportion of polymorphic loci. This value was easily calculated from published data without the need for original data. For Dunlins, expected heterozygosities were not available, but the populations sampled were in Hardy-Weinberg equilibrium. Thus observed heterozygosities were used to calculate estimated genomic heterozygosity.

Results

Control Region Sequences

Table 4-2 summarizes genetic diversity measures for the control region sequences. Inferences about subspecies structure from these data were discussed in Chapter III. The table clearly shows that knots have lower control region diversity than the average for other bird species. Of the measures in the table, haplotypic and nucleotide diversities are the most comparable across studies. Both measures correct for sample size, and nucleotide diversity additionally corrects for the length of sequence and accounts for the number of differences between sequences. For these measures only Greenfinches showed lower diversity than knots. This is not entirely surprising as Greenfinches have been sequentially bottlenecked, and have a much smaller range than do knots (Merilä et al. 1997). In comparison to other arctic breeding, long-distance migrant shorebirds, knots showed slightly lower haplotypic and nucleotide diversity than Curlew Sandpipers and White-rumped Sandpipers, both species that were probably bottlenecked during the late Pleistocene (Wennerberg and Burke 2001; Wennerberg et al. 2002). A final point of interest is that cheetahs, which are famous for extremely low genetic diversity, have five times more nucleotide diversity at the species level than do knots (0.0131 versus 0.0020). Admittedly, this final point must be taken cautiously as mammals generally show more genetic diversity than birds (Avise 1983). Nevertheless, it supports the notion that knots are extremely depauperate in mtDNA control region diversity.

Table 4-2: Genetic diversity in control region sequences of knots compared with diversity in other bird species.

Species	*N*	Domain	No. of Bases	No. of Haplotypes	Pooled Haplotypic Diversity	Average Nucleotide Diversity per Population	Pooled Nucleotide Diversity	Ti	Tv	Indels	Reference
Razorbills											
Alca torda	123	I	300	43	0.920	0.0076	0.0126	35	3	0	Moum & Arnason 2001
Common Guillemot											
Uria aalge	79	I	266	29	0.720	0.0053	0.0050	28	1	0	Moum & Arnason 2001
Common Chaffinch											
Fringilla coelebs	166	I and III	598	65	0.995	0.0044	0.0074	60	22	2	Marshall & Baker 1997
White rumped Sandpiper											
Calidris fuscicollis	52	I and II	607	21	0.860	0.0037	0.0039	17	1	1	Wennerberg et al. 2002
Dunlin											
Calidris alpina	208	I and II	608	53	0.907	0.0036	0.0157	42	8	2	Wenink et al. 1996a & b
Wild Turkey											
Meleagris gallopavo	245	I	438	42	0.929	0.0036	0.0090	31	0	0	Mock et al. 2002
Curlew Sandpiper											
Calidris ferruginea	79	I and II	664	21	0.840	0.0024	0.0024	14	2	2	Wennerberg 2001

Table 4-2 continued

Song Sparrow											Fry & Zink 1998 and pers.
Melospiza melodia	95	II and III	700	55	0.940	0.0023	0.0051	43	0	0	comm.
Red Knot											
Calidris canutus	91	I and III	675	25	0.806	0.0017	0.0020	20	1	0	-
Greenfinch											
Carduelis chloris	194	I and III	637	18	0.612	0.0012	0.0013	13	0	0	Merilä et al. 1997
Average				37	0.853	0.0036	0.0064	30	4	0.7	

[1] Transition

[2] Transversion

AFLPs

I examined AFLP loci both for subspecies structure and as a genome-wide measure of genetic diversity. Table 4-3 contains a summary of AFLP primer combinations and loci used in this study. In knots, I scored a total of 836 loci and found 129 to be polymorphic. Of these polymorphic loci, 29 distinct, intense and reproducible loci were scored for all individuals and used for population structure analysis.

Analysis of AFLPs using the program Hickory indicated no significant difference between Hardy-Weinberg equilibrium ($F_{IS} = 0$) simulations and Hardy-Weinberg disequilibrium ($F_{IS} \neq 0$) simulations; thus, I carried out all further analysis assuming Hardy-Weinberg equilibrium. Using both frequentist (TFPGA) and Bayesian (Hickory) methods of analysis, knot AFLPs showed significant evidence of population structuring (TFPGA: $F_{ST} = 0.089$; 95% CI = 0.045 - 0.143; Hickory: $F_{ST} = 0.041$; 95% CI = 0.023 – 0.065). This was also true using an Exact test in TFPGA ($p < 0.0001$).

A phenogram based on Nei's (1978) genetic distance and constructed in TFPGA had a topology that was incongruent with the phenogram constructed using control region sequences. Although AFLPs clustered *C.c. islandica* and *C.c. rufa* as well as *C.c. piersmai* and *C.c. rogersi*, the AFLP data did not place *C.c. canutus* as the most divergent group. Rather *C.c. canutus* was clustered with *C.c. roselaari* and the *C.c. rufa/C.c. islandica* clade was placed in the basal position. Bootstrap values for this tree were relatively low and ranged from 30 to 78%.

Table 4-3: Summary of AFLP polymorphisms and primer combinations in knots.

Primer Combination	EcoRI Primer[1]	MseI Primer[2]	Number of amplified loci[3]	Number of polymorphic loci	Percent polymorphic loci
1	TCT	CAT	108	6	5.6
2	TCT	CGC	31	13	41.9
3	TGA	CAA	78	8	10.3
4	TGA	CAG	54	3	5.6
5	TGA	CGC	34	16	47.1
6	TGA	CGT	52	27	51.9
7	TCT	CCG	51	9	17.6
8	TCT	CCT	85	10	11.8
9	TCT	CAA	81	6	7.4
10	TGA	CTA	88	4	4.5
11	TGA	CTC	74	16	21.6
12	TGA	CTG	100	11	11.0
Total	-	-	836	129	15.4

[1] 5' CACGACGTTGTAAAACGACTGCGTACCAATTC – NNN 3'

[2] 5' GATGAGTCCTGAGTAACG - NNN

[3] Fragments in the size range of 50 bp – 700 bp

AFLPs indicated that knots possess a lower percentage of polymorphic loci than any other species studied (Table 4-4). This is despite the fact that the sample size for Dunlins in this study was very small; only 10 individuals representing 4 subspecies were examined (originally to ensure that AFLPs were clearly distinguishing knots and Dunlins at the species level). Admittedly, using such a small sample of Dunlins grossly underestimates the amount of diversity within Dunlins. However, the fact that knots possess a lower percentage of polymorphic loci than even this very small sample of Dunlins supports the notion that knots are depauperate in AFLP diversity.

Microsatellites

In this study I was able to isolate only three polymorphic microsatellite loci (Table 4-5). The number of alleles per locus was generally quite high as were the observed and expected heterozygosities per locus. Only 4.6% of genotypic frequencies were significantly different from Hardy-Weinberg equilibrium (one population by locus data set: *C.c. piersmai*/Locus PGT90, $p = 0.02$ $H_O > H_E$), a result expected by chance alone.

Phylogenetic trees constructed with a small number of microsatellite loci are not reliable (Takezaki and Nei 1996). This is true regardless of the model of evolution or the tree building method chosen. For this reason it was not surprising that I found either no significant population differentiation (using exact tests and R_{ST}), or incongruent groupings for population differentiation (using F_{ST}). Nor was it surprising that

Table 4-4: AFLP diversity in knots in comparison to that in other bird species.

Species	*N*	Amplified Loci	# P[1]	% P[2]	Reference
House Finch *Carpodacus mexicanus*	172	363	223	61.4	Z. Wang pers. comm.
Chiffchaff *Phylloscopus collybita*	69	251	141	56.2	Bensch et al. 2002
Bluethroat *Luscinia svecica*	162	232	81	34.9	Questiau et al. 1999
Willow flycatcher *Empidonax trailli extimus*	290	708	197	27.8	Busch et al. 2000
Dunlin *Calidris alpina*	10	900	203	22.6	-
Red Knot *Calidris canutus*	92	836	129	15.4	-
Average		585.4	150.2	31.4	

[1] Number of polymorphic loci

[2] Percentage of polymorphic loci

Table 4-5: Summary of microsatellite primers and loci in knots.

Locus	Microsatellite	n	Primers	N_A[1]	H_O[2]	H_E[3]
PGT61	$(GT)_{11}$	174	5'CATTACAGCTCTTAGTTTCC 3'	12	0.735	0.763
			5'CTCTTGAATCTGAAGGTGGC 3'			
PGT83	$(GT)_{10}$	170	5'AAGGGACAGGTGACAATCCG 3'	7	0.569	0.509
			5'GCCTAATGCTGACTCACACC 3'			
PGT90	$(GT)_3$ TT $(GT)_{13}$ TT $(GT)_4$	167	5'AATAGGGCACGGGGG 3'	13	0.871	0.897
			5'CACGTACATGGCTCCTTCAGGG 3'			

[1] Number of alleles

[2] Observed heterozygosity

[3] Expected heterozygosity

subspecific tree topographies were incongruent between tree building methods (i.e. maximum likelihood, neighbor-joining and UPGMA), and with topologies found using AFLPs and control region sequences. Obviously, three microsatellite loci cannot be used to infer accurate relationships among subspecies of knots.

Despite yielding no relevant information with regards to population structure, microsatellites did provide a measure of genetic diversity in the nuclear genome. Table 4-6 shows a comparison of microsatellite variation in knots relative to that of other bird species. Knots have above average observed and expected heterozygosities for polymorphic loci but the number of polymorphic loci is very low. Research has shown that knots were bottlenecked during the last glacial maximum (Chapter 3). A bottleneck would tend to eliminate rare alleles through genetic drift. This would most probably cause loci with only a few rare alleles to become monomorphic, thus decreasing the proportion of polymorphic loci in the genome. Highly polymorphic loci would retain common alleles through the bottleneck, but would lose rare alleles. Since rare alleles do not contribute greatly to the overall heterozygosity of a locus, these surviving polymorphic loci would not show markedly lower heterozygosities. To take this phenomenon into account, the overall proportion of polymorphic loci found while screening microsatellites was considered. Table 4-6 clearly shows that knots have fewer polymorphic microsatellite loci when compared to other species. Furthermore, when I estimate a measure of genomic heterozygosity, which takes monomorphic loci into account, knots show approximately one third the average heterozygosity of other bird species.

Table 4-6: Microsatellite diversity in knots in comparison to other bird species.

Species	Loci Optimized	# P [1]	% P[2]	H_O[3]	H_E[4]	Estimated Genomic Heterozygosity[5]	Reference
Dunlin *Calidris alpina*	8	8	100.0	0.428	-	0.428	Wennerberg 2001 and pers. comm.
Great Cormorant *Phalacrocorax carbo*	7	7	100.0	0.734	0.819	0.819	Piertney et al. 1998 and pers. comm.
Red Grouse *Lagopus lagopus scoticus*	10	10	100.0	0.770	0.886	0.886	Piertney & Dallas 1997
Grey-headed Albatross *Diomedea chrysostoma*	26	17	65.4	0.391	0.459	0.300	Burg et al. 1999
Black Grouse *Tetrao tetrix*	33	18	54.5	0.756	0.768	0.419	Caizergues et al. 2001; Pietney & Höglund 2001
Song Sparrow *Melospiza melodia*	14	6	42.9	0.627	0.633	0.271	Jeffrey et al. 2001
Oystercatcher *Haematopus ostralegus*	24	8	33.3	0.431	0.455	0.152	Van Treuren et al. 1999
Red Knot *Calidris canutus*	17	3	17.6	0.725	0.723	0.128	-
Average			64.2	0.608	0.678	0.425	

[1] Number of polymorphic loci

[2] Percentage of polymorphic loci

[3] Average observed heterozygosity

[4] Average expected heterozygosity

[5] Expected heterozygosity x total proportion polymorphic loci

Discussion

The major findings of this assessment of genetic variation in both mtDNA and nuclear DNA genomes are twofold. First, I found that knots have very low genetic diversity across the genome in comparison to other bird species. Second, the frequency-based approach used in the nuclear genome to assess relationships among subspecies is incongruent with and appears less robust than results obtained using control region sequences.

Mitochondrial versus Nuclear Markers

The time to coalescence for neutral nuclear markers is four times that of mitochondrial DNA (Page and Holmes 1998). This amounts to144,000 years using an estimate of 36,000 years for coalescent time in mtDNA (averaged from Table 3-4, Chapter 3). I calculated that knots were bottlenecked to a single ancestral population approximately 20,000 years ago (Chapter 3). Clearly, most nuclear genes contain variability that is much older than knot population divergences making shared ancestral polymorphism a prolonged problem in the nuclear genome. Furthermore, although high mutation rates can speed up coalescence by increasing the number of lineages stemming from ancestral nodes, mutation rates were not sufficient in nuclear markers to counter the effects of ancestral polymorphism. Mutation rate in microsatellite loci is estimated at 5×10^{-4} per locus per generation (Goldstein et al. 1995). This can be taken as a high estimate of AFLP mutation rate as these markers represent a variety of DNA segments with differing evolutionary histories. With this mutation rate and a generation time of two

years a single mutation would occur on average every 4000 years. Thus, in the 20,000 year history of knot populations only 5 new mutations are expected. This is not enough to distinguish populations accurately, making any assessment of population structure using nuclear markers in knots potentially misleading due to a backdrop of ancestral polymorphism.

Knots as the Avian Equivalent of Cheetahs in Terms of Genetic Diversity

Both the mtDNA sequences and the nuclear loci I employed show that knots have on average much lower levels of genetic diversity than the other bird species to which I compared them. Control region sequences showed less than one half the average nucleotide diversity per population, and approximately one third the pooled nucleotide diversity found in other species. Similarly, AFLPs in knots possessed about one half the percentage of polymorphic loci found in other species. This is quite a significant depletion of diversity when one considers that the AFLPs effectively provided a genome wide survey at 836 loci. Additionally, microsatellites in knots contained one quarter the average number of polymorphic loci and one third the estimated genomic heterozygosity seen in other species.

The finding of low genetic diversity across the nuclear genome is consistent with the hypothesis that knots have undergone a recent bottleneck. If a selective sweep were responsible for low levels of genetic diversity, I would expect decreased diversity to be localized to selected and closely linked genes. Broad scale sequencing of nuclear genes to obtain direct estimates of nuclear sequence diversity may provide a more in depth

means to survey nuclear diversity and further examine the possibility of a selective sweep in the future.

Regardless of the cause, low genetic diversity in knots has conservation implications. Genetic diversity is the raw material of adaptation and without it knots are unlikely to be buffered against threats such as disease, habitat loss, and global climate change. Environmental challenges along knot flyways, such as dredging in the Wadden Sea and horseshoe crab over-fishing in Delaware Bay are already causing population declines (A. Baker and T. Piersma unpublished data) that could be exacerbated by the problem of low genetic diversity in this species. Thus I recommend that knot subspecies and their global habitats be protected. In addition, I suggest that my finding of low genetic diversity in knots opens avenues for research on how adaptations may be affected. For example, immunocompetence testing and research on the knot major histocompatibilty complex (MHC) could examine whether knot immunity has been affected by low genetic diversity.

Cheetahs are renowned as a species with dangerously low levels of genetic diversity. This has been documented in allozymes (O'Brien et al. 1985; O'Brien et al. 1983), spermatozoan morphology (O'Brien et al. 1985; O'Brien et al. 1983), skin graft analysis (O'Brien et al. 1985), microsatellites (Menotti-Raymond and O'Brien 1995) and control region sequences (Freeman et al. 2001). Low diversity in cheetahs has been attributed to a putative population bottleneck approximately 10,000 years ago (Menotti-Raymond and O'Brien 1993). In many respects, the situation in knots parallels that in cheetahs. Like cheetahs, knots experienced a recent population bottleneck, and current

findings suggest that these birds are depauperate in genetic diversity across the genome. In this respect knots are much the avian equivalent of cheetahs. However knots have a global range, making the situation all the more fascinating.

Acknowledgements

I thank Oliver Haddrath for superb lab assistance especially with AFLPs. Furthermore, Melina Baumgarten, Andy Given and Maryann Burbidge provided training and guidance in microsatellite development. I am grateful to Staffan Bensch, Joseph Busch and Zhenshan Wang for insightful communications and Adam Fry, Tracey Monehan, Stuart Piertney, Liv Wennerberg and Robert Zink for unpublished data. This research was funded by an NSERC PGS A Scholarship (PGSA-231177-2000) to DMB and grants from NSERC (200-02) and the Royal Ontario Museum Foundation to AJB.

References

Avise, J. C. 1983. Commentary. Pp. 262-270 *in* A. H. Brush and G. A. Clark, eds. Perspectives in Ornithology. Cambridge University Press, New York.

Bensch, S., A. J. Helbig, M. Salomon, and I. Seibold. 2002. Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. Molecular Ecology 11:473-481.

Burg, T. M. 1999. Isolation and characterization of microsatellites in albatrosses. Molecular Ecology 8:338-341.

Busch, J. D., M. P. Miller, E. H. Paxton, M. K. Sogge, and P. Keim. 2000. Genetic variation in the endangered southwestern willow flycatcher. Auk 117:586-595.

Caizergues, A., S. Dubois, A. Loiseau, G. Mondor, and J.-Y. Rasplus. 2001. Isolation and characterization of microsatellite loci in black grouse (*Tetrao tetrix*). Molecular Ecology Notes 1:36-38.

Felsenstein, J. 1993. PHYLIP: (phylogeny inference package) version 3.5c. Computer software distributed by author, Department of Genetics, University of Washington, Seattle, Washington.

Freeman, A. R., D. E. Machugh, S. McKeown, C. Walzer, D. J. McConnell, and D. G. Bradley. 2001. Sequence variation in the mitochondrial DNA control region of wild African cheetahs (*Acinonyx jabatus*). Heredity 86:355-362.

Fry, A. J., and R. M. Zink. 1998. Geographic analysis of nucleotide diversity and song sparrow (Aves: *Emberizidae*) population history. Molecular Ecology 7:1303-1313.

Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proceedings of the National Academy of Sciences of the U.S.A. 92:6723-6727.

Hamilton, M. B., E. L. Pincus, A. DiFiore, and R. C. Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27:500: 502: 504-507.

Holsinger, K. E., P. O. Lewis, and D. K. Dey. 2002. A Bayesian approach to inferring population structure from dominant markers. Molecular Ecology 11:1157-1164.

Jeffery, K. J., L. F. Keller, P. Arcese, and M. W. Bruford. 2001. The developement of microsatellite loci in the song sparrow, *Melospiza melodia* (Aves) and genotyping errors associated with good quality DNA. Molecular Ecology Notes 1:11-13.

Lynch, M., and B. G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Molecular Ecology 3:91-99.

Marshall, H. D., and A. J. Baker. 1997. Strucutral conservation and variation in the mitochondrial control region of Fringilline finches (*Fringilla spp.*) and the greenfinch (*Carduelis chloris*). Molecular Biology and Evolution 14:173-184.

Menotti-Raymond, M., and S. J. O'Brien. 1993. Dating the genetic bottleneck of the African cheetah. Proceedings of the National Academy of Sciences of the U.S.A. 90:3172-3176.

Menotti-Raymond, M., and S. J. O'Brien. 1995. Evolutionary conservation of ten microsatellite loci in four species of Felidae. Journal of Heredity 86:319-322.

Merilä, J., M. Björkland, and A. J. Baker. 1997. Historical demography and present day population structure of the greenfinch, *Carduelis chloris* - an analysis of mtDNA control-region sequences. Evolution 51:946-956.

Miller, M. P. 1997. Tools for population genetic ananlyses (TFPGA), version 1.3: A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

Mock, K. E. 2002. Genetic variation across the historical range of the wild turkey (*Meleagris gallopavo*). Molecular Ecology 11:643-657.

Moum, T., and E. Arnason. 2001. Genetic diversity and population history of two related seabird species based on mitochondrial DNA control region sequences. Molecular Ecology 10:2462-2478.

Myburg, A. A., D. L. Remington, D. M. O'Malley, R. R. Sederoff, and R. W. Whetten. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. Biotechniques 30:348-357.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

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

O'Brien, S. J., M. E. Roelke, L. Marker, A. Newman, C. A. Winkler, D. Meltzer, L. Colly, J. F. Everman, M. Bush, and D. E. Wildt. 1985. Genetic basis for species vulnerability in the cheetah. Science 227:1428-1434.

O'Brien, S. J., D. E. Wildt, D. Goldman, C. R. Merril, and M. Bush. 1983. The cheetah is depauperate in genetic variation. Science 221:459-462.

Page, R. D. M., and E. C. Holmes. 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell Science Ltd., Oxford.

Piertney, S. B., and J. F. Dallas. 1997. Isolation and characterization of hypervariable microsatellites in the red grouse *Lagopus lagopus scoticus*. Molecular Ecology 6:93-95.

Piertney, S. B., A. Goostrey, J. F. Dallas, and D. N. Carss. 1998. Highly polymorphic microsatellite markers in the great cormorant *Phalacrocorax carbo*. Molecular Ecology 7:138-140.

Piertney, S. B., and J. Höglund. 2001. Polymorphic microsatellite DNA markers in black grouse (*Tetrao tetrix*). Molecular Ecology Notes 1:303-304.

Questiau, S., M. C. Eybert, and P. Taberlet. 1999. Amplified fragment length polymorphism (AFLP) markers reveal extra-pair parentage in a bird species: The bluethroat (*Luscinia svecica*). Molecular Ecology 8:1331-1339.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA Sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the U.S.A. 74:5463-5467.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389-399.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalities and weight matrix choice. Nucleic Acid Research 22:4673-4680.

Tomkovich, P. S. 1992. An analysis of the geographic variability in knots *Calidris canutus* based on museum skins. Wader Study Group Bulletin 64 Supplement:17-23.

Tomkovich, P. S. 2001. A new subspecies of Red Knot *Calidris canutus* from the New Siberian Islands. Bulletin of the British Ornithologists Club 121:257-263.

Van Treuren, R., R. Bijlsma, J. M. Tinbergen, D. Heg, and L. Van De Zande. 1999. Genetic analysis of the population structure of socially organized oystercatchers (*Haematopus ostralegus*) using microsatellites. Molecular Ecology 8:181-187.

Webster, M. S., P. P. Marra, S. M. Haig, S. Bensch, and R. T. Holmes. 2002. Links between worlds: unraveling migratory connectivity. Trends in Ecology and Evolution 17:76-83.

Weir, B. S. 1990. Genetic data analysis: Methods for discrete population genetic data. Sinauer Associates Inc., Sunderland Massachusetts.

Wenink, P. W., and A. J. Baker. 1996a. Mitochondrial DNA lineages in composite flocks of migratory and wintering dunlins (*Calidris alpina*). Auk 113:744-756.

Wenink, P. W., A. J. Baker, H. U. Rösner, and M. G. L. Tilanus. 1996b. Global mitochondrial DNA phylogeography of holarctic breeding dunlins (*Calidris alpina*). Evolution 50:318-330.

Wennerberg, L., and T. Burke. 2001. Low genetic differentiation between Curlew Sandpiper (*Calidris ferruginea*) populations with highly divergent migratory directions shown by mitochondrial DNA and microsatellite analysis. Pp. 59-72. Genetic Variation and Migration of Waders. Lund University, PhD Thesis.

Wennerberg, L., M. Klaassen, and Å. Lindstrom. 2002. Geographic variation and population structure in the white-rumped sandpiper *Calidris fusicollis* as shown by morphology, mitochondrial DNA and carbon isotope ratios. Oecologia 131:380-390.

CHAPTER 5

Synthesis

My thesis has been an examination of phylogeography and genetic diversity in Red Knots (*Calidris canutus)* using a suite of currently available molecular tools. I used mitochondrial control region sequences, amplified fragment length polymorphisms (AFLPs) and microsatellites to examine the questions: (1) What is the population structure and historical demography of knots? (2) Can the biogeographic history and phylogenetic relationships of currently recognized subspecies be reconstructed? (3) What is the genetic diversity of knots and how does this diversity compare to that in other bird species?

In Chapter 2 I presented the sequence and characterization of the complete mitochondrial control region. This sequence provided a reference for the control region sequences which I examined in Chapters 3 and 4. I found the control region of knots to be 1168 bp in length and flanked by tRNA Glutamate (Glu) and the gene ND6 at its 5' end, and tRNA Phenylalanine (Phe) and the gene 12S on its 3' end. The sequence aligned well with other Charadriiformes control regions as well as with the chicken control region, and possessed conserved sequence blocks F, E, D, C and the bird similarity box (BSB) in domain II as expected for a mitochondrial copy. In addition, the sequence contained a cytosine string and TAS elements in domain I, and a tetranucleotide microsatellite at its 3' end, as do most Charadriiformes. The flanking tRNA Glu and tRNA Phe demonstrated correct secondary structure and a relative rate test indicated no significant difference between substitution rates in the sequence versus the known

mitochondrial sequence of turnstones. As a final verification, sequence obtained from purified mtDNA was identical to sequence obtained using total genomic DNA. These characteristics together verified that the sequence was mitochondrial and not nuclear in origin, and could be used as a reference to verify the mitochondrial origin of an expanded sample of sequences presented in subsequent chapters.

In Chapter 3 I addressed the first and second questions presented in the introduction through the examination of historical demography, population structure and biogeographic history. I employed Bayesian coalescent modeling on samples of mitochondrial control region sequences in both knots and Dunlins (*Calidris alpina*) to estimate evolutionary effective population size, population divergence times and time to most recent common ancestor (TMRCA) of genes. The gene trees for the two species contrasted sharply, with knot haplotypes connected in a shallow, star phylogeny and dunlin haplotypes related in a deeper bifurcating genealogy. I found that divergence times for the six subspecies of knots occurred within the last 20,000 years or so, and that evolutionary effective population sizes of females were small (N_{ef} = 2,000 – 14,000). This implied that knots were bottlenecked in a small refugial population during the Last Glacial Maximum (LGM) 18,000 – 21,000 ybp, and subsequently expanded to their worldwide range as the ice sheets retreated from the arctic breeding grounds. Dunlins, in contrast, had much older population divergence times (58,000 – 194,000 ybp) and effective population sizes that were much higher (N_{ef} = 12,000 – 44,000). This indicated that dunlin populations were not severely reduced in size in the last 200,000 years, and

that major lineages had differentiated under restricted gene flow for a much longer time than knots.

The control region sequences I presented in Chapter 3 uncovered four emergent lineages of knots. These lineages corresponded to the *C.c. canutus* subspecies, the *C.c. piersmai* subspecies, the *C.c. rogersi* subspecies and a North American group comprised of *C.c. roselaari*, *C.c. rufa* and *C.c. islandica*. Although, control region sequences were unable to genetically distinguish the North American breeding subspecies, I argued that these groups diverged too recently to be sorted into discrete lineages and that the current lack of genetic distinctiveness is caused by shared ancestral polymorphisms rather than current gene flow. *C.c. roselaari, C.c. rufa* and *C.c. islandica* are effectively isolated populations that have non-overlapping breeding grounds, different flyways and different annual cycles, and I argued that these groups should be accorded subspecific status despite genetic similarity.

The significance of the study presented in Chapter 3 was that I used population divergence times obtained through coalescent analysis to build upon classic phylogeographic analysis of gene trees. These dates along with palaeovegetation information allowed me to hypothesize biogeographic reconstructions that are more comprehensive than what was possible in the past. I proposed that an ancestral population of breeding knots was restricted to unglaciated tundra regions in the eastern Eurasian arctic during the LGM. The lineage now recognized as the *C.c. canutus* subspecies diverged from the ancestral population approximately 20,000 ybp. Later, about 12,000 ybp, the sister eastern Siberian lineage (now made up of *C.c. piersmai* and

C.c. rogersi) split to produce the North American lineage (now recognized as *C.c. roselaari, C.c. rufa* and *C.c. islandica*). This North American group likely spread into Alaska and may have migrated through the ice free corridor between the North American glaciers to wintering areas on the Gulf of Mexico. In eastern Siberia *C.c. piersmai* and *C.c. rogersi* diverged from one another approximately 6,500 years ago as woodland receded after the hypsithermal. The North American subspecies diverged only within the last 1,000 years expanding eastward from Alaska to the high arctic islands and Greenland.

In Chapter 4 I addressed the second question presented in the introduction by examining genetic diversity in Red Knots using control region sequences, AFLPs and microsatellites to compare measures of genetic diversity in knots with that in other bird species. This assessment of genetic variation showed that knots have very low diversity. Control region sequences showed less than one half the average nucleotide diversity per population and approximately one third the pooled nucleotide diversity found in other species. Similarly, AFLPs in knots possessed about one half the percentage of polymorphic loci found in other species. This is quite a significant depletion of diversity when one considers that the AFLPs provided a genome wide survey. Additionally, microsatellites in knots contained one quarter the average number of polymorphic loci and one third the estimated genomic heterozygosity seen in other species.

My finding of low genetic diversity in both the mitochondrial and nuclear genomes support the hypothesis that knots have undergone a recent bottleneck as opposed to a selective sweep, and necessitate recommendations that knots and the fragile

chain of major stopover sites along their flyways continue to be protected. I likened knots to the avian equivalent of a cheetah, a species with surprisingly low levels of genetic diversity, which is made even more impressive by the fact that the loss of genetic variation occurs throughout their global range.

The data that I presented in this thesis is significant in terms of knot biology for many reasons. First, I presented the complete and characterized control region, which can now be used as a reference sequence for future research. Second, I found for the first time emerging lineages of knots that can be distinguished genetically, and argued that genetic similarity between indistinguishable knot subspecies is most likely due to shared ancestral polymorphisms rather than to contemporary gene flow. Finally, I brought to light that fact that knots possess very low levels of genetic diversity, across the genome, in comparison to other bird species. My findings have conservation implications for knots, with population structure and low genetic diversity emphasizing the need for the continued recognition of subspecies as conservation units. This is especially important in light of the current population declines affecting most subspecies.

In addition, my findings suggest avenues for further research on Red Knots. For example, the impact of low genetic diversity on adaptations could be examined on the immune system through immunocompetence studies and the direct examination of diversity in the major histocompatibility complex (MHC). Furthermore, since low diversity and shared ancestral polymorphisms make knot DNA unsuitable for molecular markers, future research could examine other ways to track subspecies on their migratory routes. These might include an examination of stable isotopes, or the use of genetic

markers isolated from parasites unique to certain knot flyways. With short generation times and high levels of mutation, parasites could provide a rich source of genetic differentiation for molecular markers and may provide an important tool for the ongoing study of global migrations in Red Knots.