Population Structure and Gene Flow in Two Long-distance Migrant Birds, the Bicknell's Thrush (*Catharus bicknelli*) and Veery (*C. fuscescens*)

by

Walter G. Ellison

A Dissertation

Submitted to the University at Albany, State University of New York

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

College of Arts & Sciences

Department of Biology

2001

Population Structure and Gene Flow in Two Long-distance Migrant Birds, the Bicknell's Thrush (*Catharus bicknelli*) and Veery (*C. fuscescens*)

Walter G. Ellison

Ph.D. Advisor, Dr. Kenneth P. Able

ABSTRACT

Gene flow counteracts the effects of genetic drift and local natural selection regimes. As such, it tends to make populations genetically homogenous and promotes the retention of genetic diversity. Birds, especially migratory species, are among the most mobile and active of organisms and generally show relatively high levels of gene flow, although varying habitat requirements result in varying distributions. The Bicknell's Thrush and the Veery are two closely-related long-distance migratory birds with very different distributional patterns. The former is largely restricted to upper elevation, conifer-dominated habitats in a handful of northeastern North American mountain ranges, and the latter is nearly uniformly distributed across northeastern North America. I examined the effect of distributional pattern and migratory behavior on gene flow by comparing the population structures of these two species.

I used mitochondrial DNA control region III sequences to examine genotypic variation in Bicknell's Thrush and Veery. In addition, I obtained sequences of Gray-cheeked, Hermit, and Swainson's thrushes to examine

phylogenetic relationships in the genus *Catharus*. Phylogenetic analysis suggested that the Gray-cheeked Thrush, Veery, and Bicknell's Thrush form a species group within *Catharus* and that Bicknell's Thrush and Veery may be sister species. The latter finding is counterintuitive because of the Bicknell's Thrush's long-term historical status as a subspecies of the Gray-cheeked Thrush.

Sequencing of Veery samples revealed 10 mitochondrial genotypes. and Bicknell's Thrush samples revealed seven mitochondrial lineages. Population genetic analysis with the programs AMOVA, HAPLO2, and Fluctuate showed no significant population structure in either species. Comparable results were produced with the coalescence-based gene flow program Migrate. These results indicate that there is effective panmixis among the geographically isolated populations of Bicknell's Thrush as well as in the more uniformly distributed Veery.

In spite of the larger number of mitochondrial variants in the Veery, the Bicknell's Thrush had higher overall genetic diversity and deeper lineage structure than its congener. The pattern of variation in Veery lineages appears to have arisen from a glacial population bottleneck followed by rapid population growth at the onset of the current interglacial. In contrast, Bicknell's Thrush has apparently had a more stable but much smaller population throughout its history.

The results of this study indicate that Bicknell's Thrush populations are well-connected by movement of birds among them. As such, conservation of this scarce species should concentrate on habitat preservation and maintenance of

large source populations to recolonize small habitat islands rather than become occupied with concern over genetic diversity or subdivision of the species into management units.

ACKNOWLEDGEMENTS

Quoting Robert Hunter is irresistible, because it *has* been a long and strange trip through mountain forests, lab benches draped with protective paper, and the twisted pathways of computer programs. I could not have finished this project without a lot of advice, material aid, technical assistance, encouragement, and good old fashioned hand-holding. The award show speech analogy also comes to mind, but credit must be given where due, this was not an easy task and no graduate student is an island.

First, I gratefully acknowledge the guidance, forbearance, and contributions of my Graduate Committee. My chair, Ken Able, provided much material and intellectual aid, and was indeed a benevolent overseer through a longer-than-expected process. Caro-Beth Stewart went the extra mile by hosting me in her laboratory and paying for the sequencing that is the rock upon which this dissertation was built. I thank Jerram Brown for sharing his wealth of knowledge of birds and their behavior, and for many an illuminating discussion. Timothy Gage has been a valued and reliable member of my committee; it is not an easy thing to be a contributor at a distance. Finally, I sincerely thank Niles Lehman for coming on board late in the game, and for sharing his sound instincts and expertise most generously.

The members of the Stewart Lab were generous in sharing their time, expertise, and efforts on my behalf. Past lab technicians who gave much advice, and support, and who ran the sequences I relied on to do this study were Randall

Collura, Susan Jenks and Deidre Gilday. Others who aided me in many ways and simply commiserated with me when things were balky were Shou-hsien Li, Walter Messier, Jason de Koning, Praveen Karanth, and Rusty Gonser. Several other biology department graduate students contributed to the pleasantness of my sojourn here. My officemates, present and former, have been great company and provided much intellectual stimulus for me. I am grateful to have known Isabella Scheiber, Peggy Friar, Bill Johnson, and Kevin Omland. I am also thankful I had the pleasure of knowing Diana Weber, David Lemmon, Don LaFrancois, Nirmal Bhagabati, and Christine Gervasi. My worthy field assistants in 1995 were University at Albany undergraduates Nathan Schucker and Shannon Indoe.

I would not have gotten very far without a lot of help from people outside of the University at Albany. I received much support and help afield and in formulating my ideas about Bicknell's Thrush from the director and staff of the Vermont Institute of Natural Science's Conservation Biology Department, especially Chris Rimmer and Kent McFarland. I gained much of my initial lab training at the National Zoological Park's Molecular Genetics Lab. I am particularly appreciative of the help offered by director Rob Fleischer, head technician Carl McIntosh, and Cheryl Tarr. Other long-distance colleagues who provided critical material aid were Tom Alworth of the Edmund Niles Huyck Preserve, Rensselaerville, New York; H. Lisle Gibbs of McMaster University, Hamilton, Ontario, Canada; and Gilles Seutin of Environment Canada.

Two people who helped me learn much more about myself, and about my dedication to this project were Carl Walker and Stephen Giordano. I might not have stayed the course without the sage advice of these insightful gentlemen, I am eternally grateful to them.

My family has been a rock to cling to in stormy seas. I am most grateful to my father George Ellison and stepmother Molly Lofgren for all of their material and spiritual aid. Conversations with my brother Robin and sister Gayle reinvigorated my desire to survive and succeed in this effort.

I was generously supported by a number of grants and awards during this project. I owe a debt of gratitude to the Frank M. Chapman Memorial Fund of the American Museum of Natural History; the Smithsonian Institution Visiting Scholars Program; the University at Albany Benevolent Association; the Kathleen Anderson Award of the Manomet Observatory for Conservation Science; and the University at Albany Graduate Student Organization. A grant to Caro-Beth Stewart from the National Science Foundation (#DEB-9453469) supported my work in her laboratory.

Finally, I can never thank enough my endlessly patient and wise spouse, Nancy L. Martin, for the infinite array of contributions she has made to this project and to me personally. This dissertation is as much hers as mine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

m

102

119

The moment I heard it I said, "There is a new bird, a new thrush", for the quality of all thrush songs is the same, a moment more, and I knew it was Bicknell's thrush. The song is in a minor key, finer, more attenuated, and more under the breath than that of any other thrush. It seemed as if the bird was blowing in a delicate, slender, golden tube, so fine and yet so flutelike, and resonant the song appeared. At times it was like a musical whisper of great sweetness and power. The birds were numerous about the summit but we saw them nowhere else. No other thrush was seen, though a few times during our stay I caught a mere echo of the hermit's song far down the mountainside — John Burroughs on his encounter with Bicknell's Thrush at its type locality, Slide Mountain, Ulster County, New York.

Chapter 1

Introduction

Background

Living things are distributed across the Earth in disjunct populations. The consequence of this is that these populations vary in genetic composition. Thus, populations of organisms tend to show greater or lesser partitioning of genetic variation across landscapes, with many showing allele and genotype frequency shifts among sub-populations, and others being genetically homogenous. In a great many cases, differences may be profound enough to lead to the naming and classification of discrete taxa. Many taxa are both morphologically and genetically distinct. However, "genetic species" may show little difference in outward appearance. Such sibling species have long been of interest to evolutionary biologists because they give clear testimony to the decoupling of speciation and morphological divergence.

Most evolutionists agree with Mayr (1963) that genetic, if not geographic, isolation is required for speciation to come about. It is evident that interruption of gene flow is an important factor in the origin of species. Among incipient species, lack of gene flow might promote genomic reorganization and speciation (Mayr 1963). Lack of gene flow is also implicated in the loss of genetic diversity in populations through genetic drift (reviewed in Hartl and Clark 1997). Genetic drift is the tendency for genetic variation to be lost through time via random

processes, particularly the probability that the owner of a genetic variant will not breed in a given generation. Inbreeding is the mating of genetically closely related individuals. The greater the likelihood of the union of gametes containing a large number of alleles or genotypes from a common ancestor, the higher the rate of inbreeding. Such mating patterns lead inevitably to the loss of genetic variation within populations. Within a species, such loss of genetic diversity may compromise the ability of the species to respond to environmental change by limiting its arsenal of alternative alleles (O'Brien *et al.* 1985). Inbreeding depression might also so severely affect a population's breeding vitality that the threat of extinction may be increased (Saccheri *et al.* 1998). It follows that lack of gene flow has considerable bearing on issues of biological conservation because it has the capacity to increase the likelihood of extinction for small populations.

Gene flow is intimately bound to the natural history of an organism and may be broadly related to two facets of natural history: 1) behavioral propensity to disperse, and 2) physical dispersal capability, particularly for long distance dispersal. The first is related to an organism's degree of adaptation to specific conditions, in other words, there is a steep decline in fitness or a high fitness penalty associated with long-range movement. The second is related to an organism's physical adaptations for making long-range movements. Examples of organisms with low gene flow include heavy-seeded or insect-pollinated terrestrial plants, and lacustrine freshwater fish. Examples of organisms with high gene flow include wind-pollinated plants with lightweight seeds, spore producing organisms, volant insects, and some mobile vertebrates. Birds in particular show

a considerable capacity for long-range movement and often have a predisposition for natal dispersal (Greenwood 1987). A few particularly strong flying and dispersive birds have virtually cosmopolitan breeding ranges, e.g., Peregrine Falcon (*Falco peregrinus*) and Barn Owl (*Tyto alba*) (Sibley and Monroe 1990).

Other factors that affect genetic contact among populations include distance between populations and the suitability for dispersal of the intervening territory among populations (e.g., most fish are not going to cross *terra firma* to contact other fish). Given these factors, when one chooses to study gene flow within a species one must consider that organism's method and timing of dispersal, its behavior, its habitat requirements, and its biotic interactions. It is clear that direct measurement of gene flow via tracking of individual organisms and recording their breeding success from field data requires a lot of hard-to-obtain information that also requires many years of fieldwork – often a full research career – to develop an assessment of the necessary parameters. Molecular markers can produce a reasonable estimate of key population genetic parameters within a much shorter time (Avise 1994).

In this study I used mitochondrial DNA sequence data to develop estimates of gene flow and other population-genetic parameters for two long-distance migratory bird species. The Bicknell's Thrush (*Catharus bicknelli*) has a sub-divided breeding range that is limited to pockets of high elevation, stunted conifers. In contrast, a closely related (near sibling) species, the Veery (*C. fuscescens*), occupies lower elevation deciduous habitats and has not been

highly sub-divided into widely separated pockets of habitat over much of its recent history. Because both are long-distance migrants, these birds have the capacity to move decisively and quickly over many miles, thus there is little question of an individual's ability to move the relatively shorter distances among sub-populations within its breeding range. However, both species also show considerable faithfulness to breeding territories (Rimmer *et al.* 2001; Moskoff 1995). Such philopatry limits the amount of potential genetic exchange among populations. Most gene flow in migratory birds apparently hinges on natal philopatry, breeding at or near the natal place (Greenwood 1987). If young birds readily move large distances to settle for their first breeding attempt, for instance among mountain ranges in the case of Bicknell's Thrush, then gene flow should be high and population genetic structure minimal (Figure 1.1A). The large number of isolated, often very small, populations of Bicknell's Thrush implies habitat saturation in many locations, in turn suggesting that young birds should prospect widely before the start of their first breeding season. There is some indirect evidence from stable isotope analysis of tail feathers and instances of inter-population movement between color-marked populations that suggest that Bicknell's Thrush does indeed show a strong propensity for dispersal (Hobson *et al.* 2001). As such I propose that Bicknell's Thrush should show at least moderate levels of gene flow, especially among closely adjacent mountain ranges such as the Adirondacks and Green Mountains. Populations in more distant ranges, such as the White Mountains and Catskills, could show some isolation by distance with concomitantly lower gene flow between them.

However, a relatively small rate of exchange of breeding individuals, approximately one per generation per population, is all that is necessary for effective panmixia among sub-populations (Wright 1969). Alternatively, if young Bicknell's Thrushes disperse only short distances, largely within mountain ranges, the species should show strong genetic sub-division among its populations (Figure 1.1B). The Veery is so numerous and mobile it should experience few barriers to gene flow.

In order to study the population genetics of Bicknell's Thrush and Veery, I collected blood from at least 10 birds of each species from each of the four major northeastern United States mountain ranges; the Adirondacks of northeastern New York State, the Catskills of southeastern New York, the Green Mountains that run the length of Vermont, and the White Mountains of northern New Hampshire. I then sequenced and analyzed mitochondrial DNA (mtDNA) amplified from these samples by the polymerase chain reaction (PCR) and subjected these sequences to population genetic analyses. By adopting the above sampling scheme I examined samples of Veery and Bicknell's Thrush that were geographically clustered in each mountain range. These samples differed only in the elevation and habitat wherein each species was sampled. The most important distinction between the samples from the two species presumably was the presence of bridging populations of Veery among the mountain ranges and the lack of intervening Bicknell's Thrush populations. This study is the first explicitly to choose two migratory bird species with differing distributions and to examine their respective population structures with an eye toward demonstrating

the capacity of long-distance migratory birds to effect gene flow among even geographically isolated populations. My study suggests that gene flow is sufficient to maintain contact among isolated populations. I have further used the results of these analyses to examine the evolutionary and conservation implications of the patterns observed.

Population Structure and Gene Flow

Population genetics is concerned with the apportionment, maintenance and loss of genetic variation among groups of organisms. The core of this study involves the estimation of population genetic parameters. Concepts pertinent to this study include population structure, effective population size (N_e), genetic drift, inbreeding, gene flow (also called migration) parameterized as Nm, population growth, and coalescence theory. As such, it is germane to give an overview of these concepts, define their parameters, discuss their estimation from mtDNA sequence data, and to relate these topics to my study.

Population Structure

Population structure refers to the apportioning of genetic variation among sub-populations nested within a larger population. A population shows high structure or genetic architecture when a large proportion of its genetic variability is sub-divided *among* its constituent sub-populations rather than *within* sub-populations (Wright 1978). The more genetically distinct sub-populations are, the greater the genetic structure.

Population structure is measured by the fixation index (F), proposed by Sewall Wright (1931). Wright defined a number of fixation indices that measure various facets of population structure. The among-populations element of this structure is defined by F_{ST}. This represents the apportionment of alleles among sub-populations (S) relative to the total population (T). F_{ST} may also be viewed as a variance statistic that measures the amount of genetic variance among sub-populations relative to within sub-populations and total population variance. A

simplified formula given by Slatkin (1987) is $F_{ST} = \sigma^2 / p\,(1 - p)$, where p is the mean frequency of a given genetic variant and σ^2 reflects the variance in an allele's frequency among sub-populations. Partitioning of genetic variation implies restriction of movement by individuals among sub-populations and lack of structure suggests high levels of gene flow. Thus, Wright (1978) was able to derive a relationship between gene flow expressed as Nm (number of migrants per generation), and F_{ST} in the form $F_{ST} = 1 / 1 + 4Nm$ for diploid transmission. Takahata and Palumbi (1985) presented the equivalent measure for organellar genomes such as mtDNA: $F_{ST} = 1 / 1 + 2Nm$, where N is the number of migrant females per generation. Thus it is possible to derive an indirect estimate of gene flow from measures of population structure. It is also possible to estimate gene flow from patterns of common ancestry in gene trees (Beerli and Felsenstein 1999); the use of coalescence will be discussed in a following section.

Conditions that tend to increase population structure include: 1) inbreeding, which increases the probability of identical alleles occurring in the same individual; 2) genetic drift, which randomly removes genetic variation among sub-populations; 3) selection, which can increase or decrease the frequency of alleles having phenotypic effects; and 4) small effective population sizes, which promotes genetic drift. Genetic structure is lowered by high gene flow and large effective population sizes.

Two recently developed methods were employed in this study for estimating analogs of F_{ST} from mtDNA sequence data: ϕ_{ST} of Excoffier, Smouse, and Quattro (1992) and N_{ST} of Lynch and Crease (1990). Because mitochondrial

DNA is derived almost entirely from the mother's mitochondria and is thus from a haploid source, mitochondrial alleles are usually referred to as haplotypes to distinguish them from standard mendelian diploid alleles, I use haplotype henceforward in discussions of mitochondrial alleles. While there are reasonable objections to this usage it has become so widespread that to attempt to rectify this here would confuse more readers than edify them. The fixation statistic Φ_{ST} is derived from the variance definition of F_{ST} and uses Euclidean distance among DNA sequences to calculate within, among, and total sub-population variation in haplotype distances for a sample of sub-populations. The equation takes the form, $\phi_{ST} = \sigma^2_S / (\sigma^2_W + \sigma^2_S)$, where σ^2_S represents the variance among sub-populations and σ^2_W is the variance within sub-populations (Smouse 1998). N_{ST} is also derived from the variance definition of F_{ST} but uses variance in nucleotide diversity derived from sequence data within and among sub-populations to derive its fixation statistic (Lynch and Crease 1990). As such, N_{ST} directly measures differences among nucleotide sequences whereas Φ_{ST} is derived indirectly from sequence data via the distance (i.e. dissimilarity) among sequences.

Gene Flow

Gene flow is the movement of genetic variants among sub-populations. The effects of gene flow run counter to those caused by drift, in that gene flow brings genetic variation into a sub-population, thus homogenizing variation and maintaining genetic diversity among sub-populations. Population geneticists often refer to gene flow as migration (e.g., Hartl and Clark 1997) and use the symbol *m* to denote the effective genetic migration rate among populations. I

prefer to use the term gene flow because this study concerns birds, and the word "migration" has a different meaning in avian ecology, that of seasonal movement of individuals.

Gene flow is central to this study. The rate of gene flow is not calculated and presented as *m* alone, the effective migration rate, but is instead scaled by effective population size as *Nm* , which denotes the number of individuals moving between sub-populations each generation (Hartl and Clark 1997). Another way to measure gene flow is to examine the distance dispersing individuals move each generation. Neigel and others (1991) have proposed such a measure based on mtDNA sequence data.

I estimate gene flow in the current study with two methods. First, indirect estimation from F_{ST} analogs via the relationship $Nm = 1-F_{ST}/4\ F_{ST}$ (Slatkin 1987). and second, maximum likelihood estimation using coalescence theory (Beerli and Felsenstein 1999). These estimates allow me to compare values of *Nm* derived from each approach, which should provide different insights into gene flow among Veery and Bicknell's Thrush populations.

Effective Population Size (N_e)

Effective population size refers to the ideal average population that breeds in each generation producing reproductively successful offspring within a larger census population. Given that not all individuals breed or produce reproductively competent offspring in each generation, the effective population is almost always smaller than the census population, often considerably so.

Effective population size forms a bridge among population genetic concepts because the discipline is concerned with that portion of a population that breeds each generation. As such, N_e appears in most of the major equations defining population genetic parameters. Effective population size affects genetic drift, amounts of gene flow, and inbreeding. Small N_e increases drift, promoting inbreeding, and may reduce gene flow, especially if a small N_e is combined with physical isolation, or reduced probability of dispersal among sub-populations. Large N_e generally ameliorates these patterns. Organisms with a large N_e tend to have more gene flow via dispersal, and are less prone to drift due to their large numbers of breeders. Effective population size may be estimated from measures taken directly via fieldwork or indirectly through measures of population structure, such as F_{ST} and its analogs, and by analysis of gene trees with coalescence theory (Neigel 1996). A gene tree is a hypothesis of genealogical relationships among alleles (or haplotypes) derived from DNA sequence data using techniques based either upon parsimony or distance methods and thus should represent the historical pattern of allele diversification in the study organism(s) (Avise 1994). Gene trees should provide a more objective measure than F_{ST}, which is burdened by assumptions about drift equilibrium, and the symmetry of *Nm* among sub-populations (Beerli 1998).

Direct measures of N_e require fieldwork to determine the mating system, average generation time, estimates of the numbers of individuals that reproduce each generation, fitness variance among individuals, and an accurate census. Given an organism amenable to such analysis – *i.e.*, one easily counted and

studied – direct measures are more accurate and preferred over indirect measures derived from population sampling and molecular marker analysis (Bossart and Prowell 1998). Of course many species, including the birds studied here, are not amenable to extraction of key measures due to long life expectancy, high mobility, retiring behavior, and difficult habitats. Direct measures also may be less expensive than molecular analysis, although they often require many years of effort by investigators that negate this advantage.

Indirect measures with molecular markers, such as mtDNA sequences, are more capable of quickly and efficiently extracting estimates from carefully sampled populations. A reasonable estimate of key population parameters, such as N_e, can be made in less than a year given the existence of applicable molecular methods. Other analyses that are not possible with field data may also be attempted with molecular techniques. These analyses include estimation of phylogeny (Swofford 1996), inferences about biogeographic history (phylogeography, Avise 1994), detection of historic processes such as past reduction of genetic diversity through population bottlenecks (Hewitt 1996), and estimation of population growth through coalescence analysis (Kuhner *et al.* 1998b). Molecular techniques also trade off rapidity of analysis with financial considerations. It is very expensive to use molecular methods. It is thus important to focus on a specific question that has little likelihood of solution with direct field measurements.

Population Bottlenecks and Founder Effects

Population bottlenecks (Nei *et al.* 1975) and founder effects (Mayr 1963) are topics related to genetic drift. A population bottleneck is a period of very small population size relative to the long-term effective population size, and founder effects are those associated with the establishment of small satellite populations apart from the body of a species' population. The effect of small population size on drift and the concomitantly greater likelihood of identical alleles eventually coming together in such populations define the effects of these phenomena. The smaller the population, the greater the probability that alleles will be lost to drift. Genetic variation also is lost to founder populations by sampling effects related to founder population membership. Add the increased likelihood of ongoing drift. which also accelerates loss of genetic variation and it becomes clear that populations that have passed through a recent founder event or bottleneck will show less genetic variation than populations that have remained large. The longer an isolated population stays near its founding size or a bottlenecked population remains at that low level, the greater the likelihood that further variation will be lost to drift. Thus population fecundity and growth rates are of great importance in maintenance of genetic variation when populations become small and isolated (Nei *et al.* 1975).

The following discussion conflates founder effects with population bottlenecks because of their similar effects and lack of mutual exclusivity. Because we may only infer loss of variation through small population size from the historical record (e.g., as from DNA sequence data), we cannot tell which sort

of event caused the condition observed in an extant population. The Veery and Bicknell's Thrush appear to have had different recent population histories. The distribution at the Wisconsin glacial maximum 20,000 YBP (years before present) of tree species characteristic of current Veery habitat suggests a relatively small hypothetical Veery range somewhere along the coast of the Gulf of Mexico between the Florida panhandle and the Mississippi delta (Delcourt and Delcourt 1987). Thus the Veery probably went through a population bottleneck at that time. Furthermore, during their 8-10,000 year advance to their present range position ca. 10,000 YBP, they probably went through multiple founder events along the expanding range front (Hewitt 1996). On the other hand, Bicknell's Thrush was likely fairly widespread ca. 100-300 miles south of the glacial front based on patterns of pollen deposition for fir (*Abies* spp.) and spruce (*Picea* spp.) 20,000 YBP (Delcourt and Delcourt 1987). Bicknell's Thrush may never have entered a true population bottleneck in the past. Current trends in winter habitat loss may be precipitating a bottleneck (Rimmer and McFarland 1999), but this would be undetectable with any presently known molecular marker system. Genetically isolated populations could have existed over the long-term in Bicknell's Thrush unless gene flow has been high among the isolated habitat patches that have probably characterized the species' range throughout its history. I expect it to be more likely for the Veery to show clear evidence of bottlenecking than Bicknell's Thrush because the latter presumably had a larger range at glacial maximum and less latitudinal range expansion at the start of the current interglacial.

Coalescence

A relatively recent approach to analysis of population genetics among sub-populations is examination of the pattern of common ancestry or coalescence along gene trees (Kingman 1980; Kingman 1982). Generating a true gene tree (but not a tree that represents all individuals) should enable one to trace the history of populations backward through time from the branch tips through coalescent nodes (Kingman 1982; Hartl and Clark 1997). By way of example, Tree A in Figure 1.2 has three branch points indicating common ancestry and the number of coalescent events in the tree ($n_{[tips]} - 1 = n_{[coalescences]}$). Tree B shows seven such events and also shows progressively more splitting towards the branch tips. These trees each represent events taking place in an equivalent time frame. There are two patterns of interest in this comparison of trees. First, trees with more coalescence arise from larger effective populations. The reasons for this relationship with population size are twofold: 1) larger populations are more resistant to genetic drift, and 2) mutations are more numerous and likely to be maintained within a larger effective population. Also, because the rate of splitting increases moving up Tree B, one may infer the population is growing, producing more variation with the passage of time, as opposed to Tree A where the population is at best stable because variation is static through time.

Gene flow also may be mapped onto a gene tree via the logic of parsimony (Slatkin and Maddison 1989). In Figure 1.3, Tree A has no detectable movement among populations because each allele/haplotype is either geographically restricted or ubiquitous, whereas in Tree B there are two

possible explanations of the illustrated pattern of allele/haplotype distribution, both requiring four gene flow events. Either Region 1 is the origin of all variants and individuals carrying each one dispersed to new regions (2 and 3), or each variant originated in Regions 2 or 3 and all eventually dispersed to Region 1. By superimposing these inferences on estimates of effective population size derived via coalescence analysis one may develop estimates of the amount of gene flow among sub-populations within a species (Beerli and Felsenstein 1999).

Selection of a Molecular Marker

I decided at the outset of this study to use molecular markers to analyze population genetic structure in Bicknell's Thrush and Veery. Direct measurement of gene flow and effective population size in the field has advantages in being a conservative measure with fewer invalidating assumptions than indirect measurement with molecular genetic data (Bossart and Prowell 1998). However, field measurement is laborious, taking much time except in ideal study species (e.g., those with short generation times or that are easily counted), may be too conservative in failing to identify important variation among sub-populations, and cannot reveal underlying historical patterns of phylogeny, bottlenecks, and phylogeography.

Molecular markers have multiple advantages including relatively quick (if expensive) measurement of genetic variation, and less intrusive sampling of target populations such as less disturbance of habitat and breeding behavior. Another advantage of molecular markers is the drastically lowered need for collection of the study species. Molecular markers can be used to analyze population history. They can provide a relatively quick and reasonably accurate assessment of key population parameters for making timely conservation decisions. In most cases, these measures should be sufficiently accurate to make conservation decisions that might be delayed by the acquisition of field-determined measures.

There are a large and growing number of molecular markers from which to choose. Some of the most frequently used include: 1) protein electromorphs,

better known as allozymes; 2) restriction-fragment length polymorphism maps (RFLPs); 3) variable number of tandem-repeat loci (VNTRs) including so-called minisatellite and microsatellite DNA "fingerprints"; 4) randomly amplified polymerase chain reaction (PCR) products (also called randomly amplified polymorphic DNA [RAPDs]); and 5) direct sequences of organellar and nuclear DNA. Each of these markers have advantages and disadvantages when applied to problems related to population genetics (e.g., gene flow), birds as the taxon of interest, time frame, expense, and ease of use. For instance, allozymes have been used for over thirty years, are relatively easy to use, inexpensive, and easily interpreted as co-dominant Mendelian markers. However, they tend not to be diverse in birds (Avise 1983), and likely hide much useful variation in synonymous sites in the DNA coding for the proteins assayed.

DNA sequence level variation has the major advantage of being the source of all the variation measured via the indirect analyses like allozymes. Homoplasy is an issue when one interprets DNA sequences, but the sources of this convergent variation are less varied, and somewhat easier to identify and analyze than is the case for indirect marker analyses. Different types of sequence, such as genes and pseudogenes have different mutation rates and thus varying timeframes for investigating aspects of population genetics and systematics. For instance, in a study of gene flow one should take into account the likely timing of gene flow events measured by the chosen marker or sequence. VNTRs, with their high mutation rates, would be expected to detect more recent gene flow than virtually any direct sequence. However, meaningful

signal from VNTR data might be obscured by high variability making it difficult to sample enough individuals to detect movement of numerous scarce alleles, or homoplasy due to high mutation rates making it impossible to assure identity by descent for outwardly similar alleles.

Generating sequence data often requires the technically difficult steps of cloning and PCR primer development, but many well-worked chromosomal regions, nuclear genes, and organelle genomes have published conserved primers (e.g., mtDNA, Kocher *et al.* 1989). There are clear advantages to using sequence data, especially from mtDNA, for genealogical analysis of lineages through coalescence theory (e.g., Kuhner *et al.* 1995).

The analysis of mtDNA in population genetics and systematics has a long tenure in the flowering of molecular evolution reaching back to the 1970s (Brown and Vinograd 1974; Avise 1998). As such there is an excellent diversity of studies to form the basis of new research. The animal mitochondrial genome is small and compact, possessing coding sequence for 13 proteins, two ribosomal RNAs, and 22 transfer RNAs, plus the non-coding control region that serves a regulatory function and is the origin of mitochondrial replication (reviewed in Brown 1985). The genetic code of mtDNA is also slightly different from the well-known standard code of most eukaryotic nuclear DNA (Barrell *et al.* 1979; Bibb *et al.* 1981). The animal mitochondrial genome also exhibits a ca.10 times higher rate of nucleotide substitution than does single-copy nuclear DNA (Brown *et al.* 1982; Ingman *et al.* 2000). Mitochondrial DNA is almost exclusively inherited as a single non-recombining unit from the mother (Brown 1985). Published exceptions

to maternal inheritance (Gyllensten *et al.* 1991) and lack of recombination exist (Eyre-Walker *et al.* 1999; Hagelberg *et al.* 1999). However, these cases appear to be exceptions or based on inappropriate analysis of linkage disequilibrium in the case of recombination (Kumar *et al.* 2000), such instances do not invalidate conclusions drawn from prior research.

High substitution rates in mtDNA make it valuable in population-genetic analyses because multiple mtDNA genotypes usually exist within a species. Maternal inheritance and lack of recombination make lineage analysis and tree-building more straightforward (Avise *et al.* 1987). Maternal inheritance also means mtDNA has an essentially haploid effective population. Small N_e leads to shorter time intervals to points of common ancestry in phylogenies than in nuclear markers, making it more likely to recover true phylogenies with mtDNA data (Moore 1995).

The control region has conserved regions involved with mitochondrial gene regulation. These appear to be largely associated with a conserved central domain called the displacement loop ("D-loop") because of its observed secondary structure (Brown 1985; Quinn 1997). However the 5' and 3' flanking regions, referred to as regions I and III by Baker and Marshall (1997), are more variable and are useful for population-genetic studies of closely related individuals. Reasons for greater variability in the areas flanking the D-loop seem to be lack of regulatory function leading to selective neutrality, and strand displacement effects surrounding the D-loop during replication and transcription (Brown 1985; Quinn 1997). The control region, especially regions I and III, has a

higher, sometimes much higher, mutation rate than coding mtDNA. I chose to use sequences from the non-coding mtDNA control region in this study primarily because its high mutation rate promised more variable sites for population analysis of the Bicknell's Thrush and Veery, which appeared to be very closely related. My second reason was the reasonable assumption of relative selective neutrality in the regions flanking the D-loop. Neutrality provides assurance that the distribution of haplotypes among sub-populations is attributable to genetic drift and gene flow rather than selection. Most of the analyses I discuss in this study were based on 395 base pairs of sequence from control region III.

Portions of the mitochondrial genome have been transferred to the nucleus where they exist as pseudogenes, these pseudogenes are often referred to in the current literature as NUMTs – nuclear mitochondrial DNA (e.g., Collura and Stewart 1995; Quinn 1997). Such nuclear copies pose a problem for investigators because misleading phylogenies can arise from them. Because correct tree topologies are necessary to use many of the inference tools I have mentioned above, especially those based on coalescence, the presence of NUMTs could lead to incorrect conclusions. For instance, NUMTs may show much slower rates of change than mtDNA, leading to shorter branch lengths on trees and shorter estimates of the time of divergence among taxa (e.g., Price *et al.* 1998; Lovette and Bermingham 1999).

I the succeeding chapters I will be discussing results of mtDNA sequencing within the control region of my two focal species, Bicknell's Thrush and Veery, with some reference to sequences from three other *Catharus*

thrushes, Gray-cheeked (*C. minimus*), Hermit (*C. guttatatus*), and Swainson's (*C. ustulatus*).

Study Species

I chose to study two closely related thrushes of the genus *Catharus*, Bicknell's Thrush (*C. bicknelli*) and Veery (*C. fuscescens*). These birds were chosen primarily for their contrasting distributional patterns in northeastern North America. The Bicknell's Thrush occurs largely at relatively high elevation (>1000 m in the United States, and usually >600 m in southeastern Canada) in sufficiently lofty mountain ranges, whereas the Veery is widespread at lower elevations nearly throughout the region.

Bicknell's Thrush was first recognized and collected by Eugene P. Bicknell, a New York City naturalist, on Slide Mountain in the high Catskills of Ulster County, New York, in 1881. No one had known of its existence although it resided on mountains not far from the continent's largest city. Robert Ridgway provided the first description of Bicknell's Thrush (Ridgway 1882), largely from Mr. Bicknell's specimens. He named it *Hylocichla aliciae bicknelli* in honor of its collector. As the trinomial indicates, Ridgway considered it a subspecies of another thrush, the widespread sub-arctic Gray-cheeked Thrush, formerly *H. aliciae*, now *Catharus minimus*. Recent observations by several ornithologists have provided evidence that Bicknell's Thrush is a full species subtly distinct from Gray-cheeked Thrush. This was argued by Alan Phillips (1991) and Joe Marshall (*in litt.*) and persuasively synthesized by Henri Ouellet (1993). Ouellet's presentation highlighted plumage, skeletal, vocal, behavioral, and biochemical (mtDNA restriction digests provided by G. Seutin) evidence in favor of the distinctiveness of Bicknell's Thrush; his argument was accepted by the American

Ornithologist's Union in 1995 (American Ornithologists' Union 1995). The species occurs largely at upper elevations from northeastern Nova Scotia and the north shore of the St. Lawrence River in Québec westward in the Laurentian Mountains to just north of Montréal and southward to the Catskills of New York State, the southern Green Mountains of Vermont, and the White Mountains of New Hampshire (American Ornithologists' Union 1998). It nests in stunted conifers and mixed stands of regenerating conifers and hardwoods. The recent elevation of Bicknell's Thrush to full species combined with its limited distribution and scarcity have made it the topic of a great deal of recent research into its natural history by biologists in the northeastern United States and southeastern Canada. The Canadian Wildlife Service maintains a website on the species with information about current research (http://www.ns.ec.ca/wildlife/bicknells_thrush).

The Veery is closely related to Bicknell's and Gray-cheeked thrushes. It ranges widely in North America, mostly north of 40°N latitude from north central Alberta east to southwestern Newfoundland and south to eastern Oregon, northern Utah, Colorado, Iowa, the Great Lakes States, in the Appalachians to the Great Smokies, and to the piedmont of Maryland. A few populations also occur in the mountains of northern Arizona and New Mexico (Moskoff 1995; American Ornithologists' Union 1998). The Veery occurs in densely overgrown, wet places including shrubby or forested swamps, and regenerating or shrubby hardwood forest under canopy gaps of varying size. Its preference for overgrown, low-stature habitats is similar to the habitat preferences of Bicknell's and Gray-cheeked thrushes. The Veery is also similar to Bicknell's and Gray-cheeked in

appearance and voice although it can be consistently distinguished from them with a little practice (Peterson 1980).

Both the Veery and Bicknell's Thrush are long-distance migrants. Bicknell's Thrush winters in the Greater Antilles, apparently with the vast majority on Hispaniola (Phillips 1991, C. C. Rimmer, pers. comm.). As with its breeding range, the Veery's wintering range is much more extensive and is much farther to the south, essentially encompassing the western Amazon basin of Brazil with a few occurring in adjacent South American countries (Ridgely and Tudor 1989).

The Veery and Bicknell's Thrush are morphologically, vocally, ecologically, and behaviorally similar (see Table 1.1). They are allopatric only in the sense of being found at different elevations and floristically distinct habitats within the same geographic region. The major difference of interest for the purpose of this study is the profound difference in their distribution across the landscape with Bicknell's Thrush limited to scattered habitat islands and the Veery almost pan-regional. A study of the partitioning of genetic variation among regional populations of Veery and Bicknell's Thrush should provide an illuminating test of the effect of an essentially insular distribution on one hand (Bicknell's) and continuous distribution on the other (Veery) on gene flow. The limitation of Bicknell's Thrush to isolated mountain summits and ranges versus the Veery's widespread occurrence among and within mountain ranges amounts to a "natural experiment" (Diamond 1986) testing the effect of 10,000 years of habitat isolation on Bicknell's Thrush as Veery acts as a natural control.

Conservation

The Bicknell's Thrush has a small global population, certainly less than 50,000. It is limited to habitat islands in a generally low productivity habitat on the breeding grounds (sub-alpine forest) (Sabo 1980). The wintering range is literally insular, largely confined to Hispaniola, and humans heavily exploit the wintering habitat (montane forest and woodland). These factors imply the potential for the endangerment of Bicknell's Thrush in the near future, in spite of reasonably healthy numbers given the limitations of its small and specialized range.

The conservation implications of this study are twofold. First, because the Bicknell's Thrush has been limited to habitat islands for at least 10,000 years, habitat isolation may have led to loss of genetic variation as well as to small demographically unstable sub-populations. Such loss of genetic variation should be detectable in the comparison of Bicknell's Thrush to Veery. A result demonstrating loss of genetic variation due to a fragmentary distribution might in turn be related to the issue of the effects of habitat fragmentation on migratory birds in general. Second, the Bicknell's Thrush itself is the subject of some conservation concern in both the United States and Canada (Rimmer *et al.* 1993, G. Seutin, pers. comm.). Thus, my results regarding gene flow and population structure of this species may find application among policy makers and habitat managers on the government-held lands that Bicknell's Thrush occupies during the breeding season in the United States and on the privately-held forestry lands it inhabits in southeastern Canada.

There is far less reason for concern for the much more common Veery. However, it is a forest-dwelling neotropical migrant experiencing accelerating habitat loss and degradation (Askins *et al.* 1990). Data from the National Biological Survey's Breeding Bird Survey from 1966-1993 show a 29.4% decline with a 19.6% decline since 1984 for the Veery (Price *et al.* 1995). Thus, given negative population and habitat trends, the widespread and common Veery may not remain so indefinitely. As the species' habitat becomes more fragmented, the potential issues regarding gene flow or random genetic drift proposed above for Bicknell's Thrush could also come into play for the Veery.

It is reasonable to propose lowered gene flow and higher inbreeding among isolated habitat islands. This is particularly true for organisms with low dispersal capabilities and rates. How true this might be for volant birds is not clear. Most studies of avian gene flow suggest that it is fairly high (Rockwell and Barrowclough 1987; Ball and Avise 1992; Zink and Dittmann 1993). Therefore, habitat isolation or habitat fragmentation may have few genetic consequences among birds. Nonetheless the long-term isolation of breeding populations of Bicknell's Thrush for at least 10,000 years provides an opportunity to establish the species' dispersal capabilities or to document the effects of such a regime if it does indeed establish genetic isolates. *If* gene flow were severely limited among sub-populations we might expect to see 1) loss of genetic diversity through drift and metapopulation dynamics, and 2) a possible tendency for local adaptation due to potential differences in natural selection among mountain ranges. The first condition implies lowered ability to deal with extreme environmental variation,

and loss of population viability due to inbreeding depression. Both of these processes could hasten extinction (Fleischer 1998; Saccheri *et al.* 1998). Local adaptation could also have implications for management of Bicknell's Thrushes on different mountain ranges. If certain sub-populations of the species are discrete management units, this affects the promulgation of a recovery plan, should one be needed, since inter-population cross-fostering might be limited by the potential for adaptation to local conditions by members of such genetically isolated units. The same observations would apply also to the Veery, but are probably moot given that species' current status and distribution. If gene flow is sufficient to prevent genetic isolation of populations, then this would greatly simplify the management of Bicknell's Thrush.

	Bicknell's Thrush	Veery
Range	Isolated on mountaintops (Andrle and Carroll 1988)	Widespread at lower elevations (Andrle and Carroll 1988)
Breeding Habitat	Thick low conifers and deciduous shrubs (Ouellet 1993)	Deciduous & mixed woodlands with thick shrubs, shrub wetlands, clearcuts (Bertin 1977)
Migration	Long distance (American Ornithologists' Union 1998)	Long distance (Moskoff 1995)
Site Fidelity	High (Rimmer and McFarland 1999)	High (Moskoff 1995)
Mating System	Polygamy (Rimmer and McFarland 1999)	Monogamy (Dilger 1956b)
Nest Site	Small tree (Wallace 1939)	Shrubs/herbs (off ground) (Moskoff 1995)
Food	Insects, fruit (Wallace 1939)	Insects, fruit (Moskoff 1995)
Feeding Location	Ground (Sabo 1980)	Ground (Sabo and Holmes 1983)

Table 1.1. Comparison of life-history traits of Bicknell's Thrush (*Catharus bicknelli)* to Veery (*C. fuscecscens*).

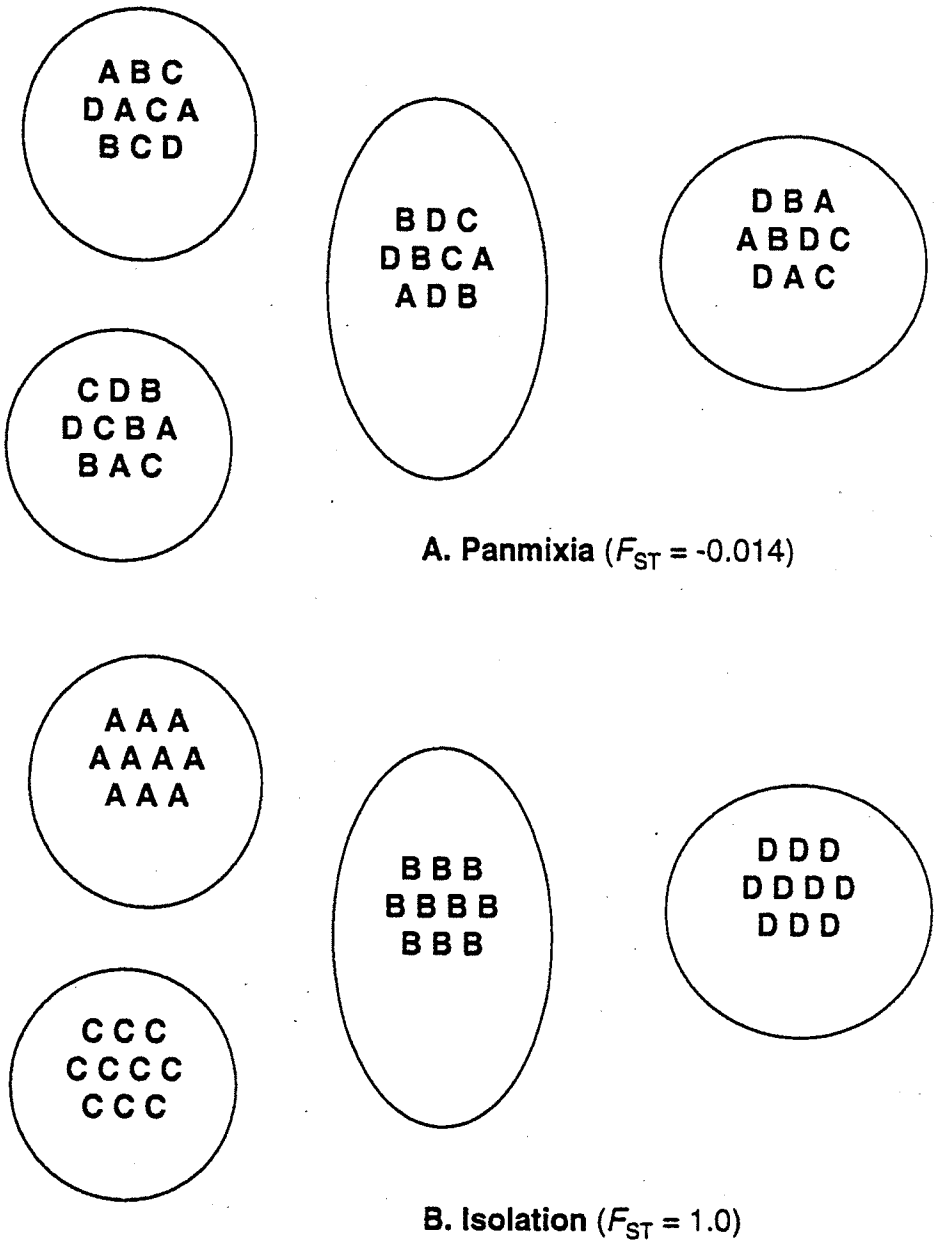


Figure 1.1. Population structure alternatives. Possible extreme examples of population structure within geographically isolated populations. Letters represent alleles or genotypes of individual organisms. **A.** Panmixia, no barriers to genetic exchange between populations. All alleles/genotypes are shared among all populations. F_{ST} is negative because within population variation exceeds among population variation slightly. **B.** Complete isolation of populations, leading to complete fixation and no exchange of alleles/genotypes among populations.

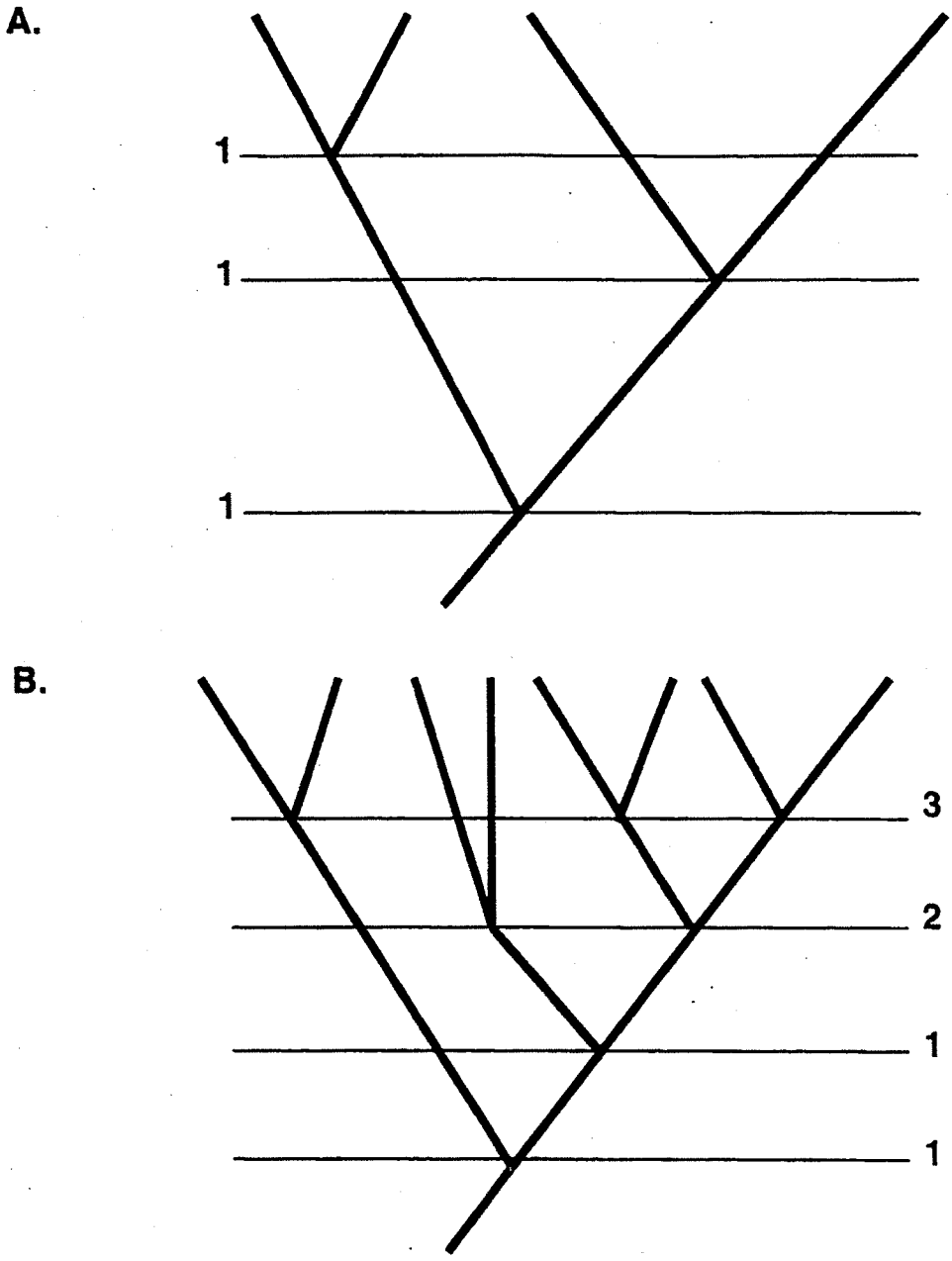


Figure 1.2. Estimating N_e and population growth via coalescence. Dendrograms illustrating coalescence patterns in two hypothetical genealogies. Numbers adjacent to each tree indicate the number of lineage splittings at each level of the tree. At top (A) is a sparse genealogy with little branching representing a low population with little or no growth. At bottom (B) is a genealogy with much branching increasing higher in the tree representing a large population with a high growth rate. See text for further explanation.

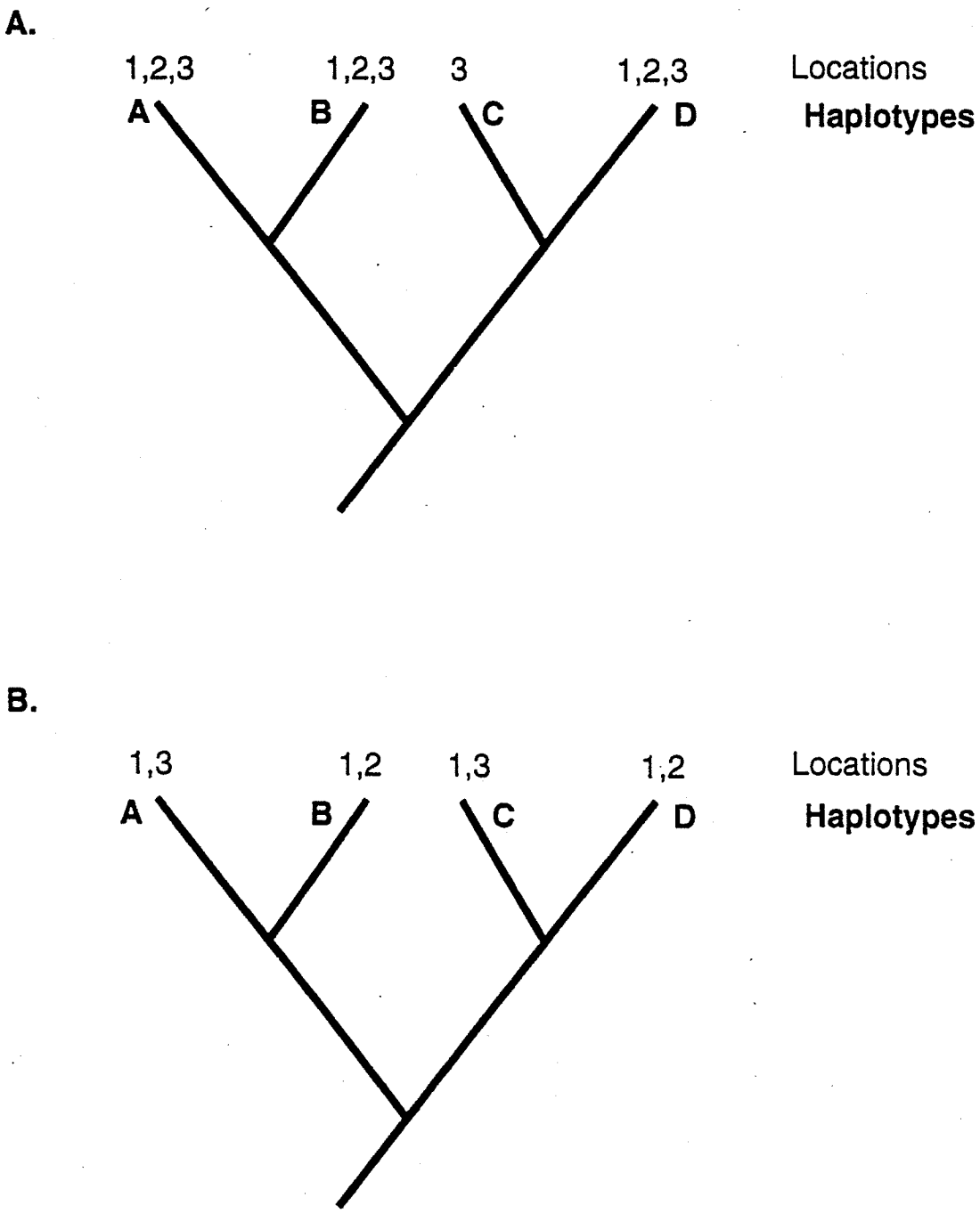


Figure 1.3. Tracing gene flow on genealogies. Hypothetical combined area and haplotype dendrograms illustrating, at top (A), four haplotypes in three regions with no migration events, and, at bottom (B), the same number of haplotypes and regions with four gene flow events. See text for further explanation (Beerli 1998).

Chapter 2

General Methods

Field Methods: Sample Collection

I envisioned this study as a comparison of population-genetic parameters from two thrush species sampled in discrete regions. I decided to obtain several individuals from each region to establish within and among sub-population variation. My field experience suggested that 10 birds per sub-population would be logistically feasible. The tissue used in the study was blood because it is easy to obtain from birds, its removal has little effect on the survival of birds, and it is a source of abundant DNA in birds (although not as much for mtDNA as for nuclear DNA) (Arctander 1988). I was concerned that my sampling of thrushes – especially the scarce Bicknell's Thrush – be as non-destructive as possible. Blood seemed the best alternative because muscle biopsy (Baker 1981) might be more destructive than blood sampling.

Because Bicknell's Thrush occurs at difficult-to-access upper elevations in northeastern United States mountain ranges, I collected blood samples from 10 birds each from the Catskill Mountains of southeastern New York, the Adirondack Mountains of northeastern New York, the Green Mountains of central Vermont, and the White Mountains of northern New Hampshire. I then required a similar sampling scheme for a closely related thrush species that was more continuously distributed over its range. After some initial work on the Gray-cheeked Thrush in

Labrador and Newfoundland in 1995 and 1996, I decided to turn to another thrush, the Veery, in autumn 1996. This allowed me to sample a species in the same mountain ranges at sites close to those where I had obtained Bicknell's Thrush samples. The mean distance between the paired Veery and Bicknell's Thrush sampling locations was 11.2 km (SD = 7.2). The Veery proved much less difficult to locate and capture than Gray-cheeked Thrush. I caught all of the remaining birds I required in 1997. See Tables 2.1 and 2.2 for information on date of capture, locality, and whether a sequence was produced for the 58 Bicknell's Thrushes (n_{seq}=43) and 43 Veeries (n_{seq}=40) I captured. See also Figure 2.1, which gives sampling locations for both species. Full data on collection locations, elevations and morphometrics of the birds sampled can be found in Appendix 1.

I also obtained samples from three additional thrush species, Gray-cheeked Thrush (*C. minimus minimus*, *C. m. aliciae*), Hermit Thrush (*C. guttata faxoni*) and Swainson's Thrush (*C. ustulatus swainsoni*), largely for phylogenetic work. I captured two Gray-cheeked Thrushes (presumably *C. m. aliciae*), in central Labrador in June 1995, and two more near St. Anthony, Newfoundland (presumably *C. m. minimus*), in June 1996. Hermit Thrush samples originated from two non-breeding birds I captured at the Edmund Niles Huyck Preserve, in Rensselaerville, Albany County, New York, in October 1997. Eight Swainson's Thrush samples came from the laboratory of H. Lisle Gibbs of McMaster University, Hamilton, Ontario. The Gibbs lab also provided a Gray-cheeked Thrush sample from Gros Morne National Park, Newfoundland. Details of this

work may be found below in Chapter Three. Information on these three additional thrush samples may be found in Table 2.3.

I used a fairly straightforward technique to capture birds for sampling. I set one or two six-meter mist nets in suitable habitat where a bird had been previously located. For Bicknell's Thrush this was generally along a hiking trail because sub-alpine spruce-fir is often almost impenetrable. I usually set up nets much farther from a trail or established roadway to capture Veeries. My nets were mounted on lightweight collapsible poles constructed from polyvinyl chloride pipes making the traps lightweight and portable in off-road situations (design courtesy Nancy L. Martin). I placed a portable tape recorder next to the net and positioned a carved and painted thrush model (courtesy Nancy L. Martin) near the net. I then played a recording of Bicknell's Thrush or Veery territorial song. The former were acquired through my own field recordings of songs and aggressive calls, the latter were taken from the Federation of Ontario Naturalists recording "Sounds of Nature, 8: Thrushes, wrens and mockingbirds" recorded by Donald J. Borror and William W. H. Gunn. These recorded songs usually drew agitated males to the net. In the majority of cases thrushes were active enough to eventually strike the net and become entangled. I would then quickly extract the bird, take several morphometric measurements (wing chord, tarsus, tail, culmen, and mass), and place a United States Geological Survey bird band on its leg. Ten additional Bicknell's Thrushes were captured with similar methods in cooperation with researchers from the Vermont Institute of Natural Science of Woodstock, Vermont.

Holding the wing between thumb and index finger at its wrist joint I would make a small puncture with a 27 gauge needle (1993-1995) or a lancet (1995-1997) in the ulnar wing vein and collect the resulting venous blood flow in two heparinized capillary tubes to a total of approximately 100 µl for most birds (some yielded less). In most cases I immediately placed the blood into two 1.5 ml microfuge tubes containing 1.0 ml of Queens Lysis Buffer (Seutin *et al.* 1991), 0.1 M Tris-HCl , 0.1 M EDTA, 0.2 M NaCl, and 1% n-lauroylsarcosine (pH 8.0), ca. 50 µl of blood to each tube. I released the bird after its bleeding had ceased, usually after less than three minutes. In the first two years of my study I color marked Bicknell's Thrushes with unique combinations of three plastic bands but I discontinued this after 1994 because the Vermont Institute of Natural Science was conducting a similar color marking project in several of the places I had worked. Nonetheless, re-sightings of birds I had bled and color-marked in prior years on four mountains, including one bird that had acted sluggishly after his release, reassured me that birds generally recovered from the bleeding procedure.

DNA Isolation

The blood samples were stored at –20 ° C in lysis buffer. Over the course of this study, two DNA extraction methods were used. Both relied on proteinase K digestion of the protein associated with the DNA. I initially used a fairly standard phenol-chloroform isolation technique modified from Sambrook et al. (1989) by David Westneat and communicated to me by Shou-hsien Li and Elaina

Tuttle (see Table 2.4A). I eventually stopped using this protocol because it proved inconsistent in yield and may have produced DNA contaminated with PCR inhibitors. I turned to the protocol of Gemmel and Akiyama (1996), suggested by Shou-hsien Li, with modifications as indicated in Table 2.4B. This method uses Lithium Chloride (LiCl) as the salt required for DNA precipitation at an earlier stage than in standard protocols to aid in deproteinization (Miller *et al.* 1988; Johns and Paulus-Thomas 1989). It also avoids phenol extraction, instead employing only one chloroform:isoamyl alcohol extraction. This method proved less time-consuming, provided good and consistent yields, and reasonably pure DNA precipitates. The mean concentration of DNA from the Gemmel and Akiyama method measured by analysis with a Sim Aminco 3000 Array spectrophotometer set to 260 nm absorbance was 220 ng/µl (n = 110).

Polymerase Chain Reaction Protocols

I used nine primers to amplify the control region of the five thrush species I worked on in this study (see Table 2.5). When I initiated this line of research the primers I used had been designed for other passerine birds, particularly Mexican Jay (*Aphelocoma ultramarina*). In order to avoid or minimize amplification of NUMTs, I adopted a semi-nested PCR approach suggested by Shou-hsien Li. Thus, for the first two years of the project I employed outer primers located in the nitrogen dehydrogenase 6 (ND6) gene and the 12s ribosomal RNA gene in an initial PCR followed by amplification with a pair of primers, either L437 or L817 and H1248 (see Table 2.5), whose priming sites were inside the former two. This

produced a ca. 420 base pair product from control region III. Figure 2.2 is an example of an agarose gel of PCR reaction products derived from amplification of control region III in four Bicknell's Thrushes and four Veeries, all produced with the primers BTL800 and H1248. See Figure 2.3 for a map of the control region in relation to neighboring genes and the locations of priming sites. During August 1998 I designed primers directly from Bicknell's Thrush sequences that permitted clean amplification of control region III without recourse to semi-nested PCR. The sequences derived from the semi-nested and direct procedures were identical.

I used a PCR cycling protocol derived from Shou-hsien Li's Mexican Jay research, optimized for amplification of *Catharus* thrush DNA, from October 1996 to July 1998 (see Table 2.6A). I adopted a different cycling profile and reaction volume at the suggestion of Niles Lehman in August 1998 (Table 2.6B). I generated PCR product for both light and heavy strands for 43 Bicknell's Thrushes, 40 Veeries, four Gray-cheeked Thrushes, two Hermit Thrushes, and two Swainson's Thrushes.

Sequencing Protocols

Once PCR product of the correct size has been obtained it must be prepared for DNA sequencing. First, one must clean the product by removing extraneous fragmentary PCR products and excess unbound primer DNA. I used Millipore Ultrafree®-MC filters and a protocol modified from the manufacturer's to purify the PCR products (Table 2.7A). I concentrated the filtrate based upon estimates of DNA quantity derived from 1.5% agarose gels. I estimated the

amount of purified product to use in the subsequent sequencing reaction from another quantification gel. See Table 2.7B for reaction reagent volumes and cycling parameters specified for the dye terminator reaction by ABI (Applied Biosystems, Inc.). ABI Ready Reaction mix contains Taq DNA polymerase (AmpliTaq®), ddNTPs lableled with fluorescent dyes, PCR buffer, and Mg^{++}. Sequencing product was purified in Sephadex® columns with a protocol suggested by Randall Collura (see Table 2.6C). I stored purified sequencing product at – 20 °C until the sequencing gel was ready to be run on the ABI 373A automated sequencer in Caro-Beth Stewart's laboratory. Laboratory technicians ran the samples on a 6% polyacrylamide gel run at 2500 volts, 30 watts and 40 millamperes in 1X TBE buffer for 12 hours (usually overnight).

Sequence Analysis

The ABI 373 sequencer automatically calls bases from the chromatographic peaks registered from the nucleotide-specific dye incorporated into the PCR product and stores these data as a simple text file with the base calls, and as a graphical file representing the sequence chromatogram showing chromatographic peaks and base calls. I aligned the sequence chromatograms by eye based on my knowledge of primer sequences and unique conserved motifs I discovered among all of the sequences. I edited sequences with the ABI editing program, SeqEd®, through comparison of individual sequences and comparison of light and heavy strands of the same sequence. Some of the more ambiguous examples were amplified and sequenced a second or third time. This

included a number of sequences initially run in only one direction from the earlier semi-nested PCR runs (L437/H1248). An alignment of the complete sequences from all specimens is given in Appendix 2. I will discuss analytical methods applied to these sequences, e.g. phylogenetic inference, in subsequent Chapters.

Sample ID	Capture Date	Location	Sequenced (yes/no)*
GB917	2 June 1993	Mt. Mansfield, VT	No
GB715a	3 June 1993	Mt. Mansfield, VT	Yes
GB716	4 June 1993	Mt. Mansfield, VT	No
GB717g	5 June 1993	Mt. Mansfield, VT	AF373787
CB718a	9 June 1993	Hunter Mt., NY	AF373783
CB719c	11 June 1993	Slide Mt., NY	AF373784
CB720d	14 June 1993	Slide Mt., NY	Yes
GB722	17 June 1993	Shrewsbury Peak, VT	No
GB723	17 June 1993	Shrewsbury Peak, VT	No
GB724a	19 June 1993	Shrewsbury Peak, VT	Yes
GB725a	19 June 1993	Shrewsbury Peak, VT	Yes
CB726	26 June 1993	Black Dome Mt., NY	No
CB727b	26 June 1993	Black Dome Mt., NY	Yes
CB728f	27 June 1993	Plateau Mt., NY	AF373786
CB729	27 June 1993	Plateau Mt., NY	No
CB730b	27 June 1993	Plateau Mt., NY	Yes
CB731	27 June 1993	Plateau Mt., NY	No
AB732e	30 June 1993	Whiteface Mt., NY	AF373789
AB733a	30 June 1993	Whiteface Mt., NY	Yes
AB734a	1 July 1993	Whiteface Mt., NY	Yes
GB921	1 June 1994	Mt. Mansfield, VT	No

Table 2.1. Bicknell's Thrush sample data. Table continues on following pages. Sample ID numbers are derived from a single letter mountain range designation (G=Green Mt; C=Catskill; A=Adirondack; W=White Mt), a single letter species identification (B=Bicknell's), the last three digits of the USGS Bird Banding Lab band identification number for each bird, and for those sequenced, the single letter haplotype code for its species. Out-of-sequence numbers were 10 birds captured by researchers from the Vermont Institute of Natural Science of Woodstock, VT. * A representative of each unique sequence identified was submitted to GenBank, these are indicated by GenBank accession numbers. More details of locations and morphometric data for each bird may be found in Appendix 1. Aligned sequences for all individuals sequenced may be found in Appendix 2.

Sample ID	Capture Date	Location	Sequenced (yes/no)
GB901b	2 June 1994	Mt. Mansfield, VT	Yes
GB916a	2 June 1994	Mt. Mansfield, VT	Yes
GB937a	2 June 1994	Mt. Mansfield, VT	Yes
GB904d	3 June 1994	Mt. Mansfield, VT	AF373788
GB938a	3 June 1994	Mt. Mansfield, VT	Yes
CB737b	5 June 1994	Black Head Mt., NY	AF373785
CB738	5 June 1994	Black Head Mt., NY	No
CB739	8 June 1994	Plateau Mt., NY	No
CB740	8 June 1994	Plateau Mt., NY	No
CB741a	8 June 1994	Plateau Mt., NY	Yes
GB742	10 June 1994	Shrewsbury Peak, VT	No
WB743b	14 June 1994	Mt. Jefferson, NH	Yes
WB744b	14 June1994	Mt. Jefferson, NH	Yes
WB745b	14 June 1994	Mt. Jefferson, NH	Yes
WB746c	15 June 1994	Mt. Washington, NH	Yes
WB747a	15 June 1994	Mt. Washington, NH	Yes
AB748a	22 June 1994	Phelps Mt., NY	Yes
AB749e	22 June 1994	Phelps Mt., NY	Yes
AB750	22 June 1994	Phelps Mt., NY	No
AB751	24 June 1994	Whiteface Mt., NY	No
GB602a	5 July 1994	Haystack Mt., VT	Yes
GB603b	5 July 1994	Haystack Mt., VT	Yes
CB752	25 May 1995	Black Head Mt., NY	No
CB753b	25 May 1995	Black Head Mt., NY	Yes
CB754a	25 May 1995	Black Dome Mt., NY	Yes
AB755b	31 May 1995	Wright Peak, NY	Yes
AB735a	7 July 1993	Whiteface Mt., NY	Yes
GB736a	29 May 1994	Stratton Mt., VT	Yes
GB932a	1 June 1994	Mt. Mansfield, VT	Yes

of

Table 2.1. Bicknell's Thrush sample data. (Continued)

Sample ID*	Capture Date	Location	Sequenced (yes/no)
AB756	31 May 1995	Wright Peak, NY	No
AB757b	1 June 1995	Cascade Mt., NY	Yes
AB758b	1 June 1995	Cascade Mt., NY	Yes
AB759a	2 June 1995	Hurricane Mt., NY	Yes
WB760c	7 June 1995	Mt. Jefferson, NH	Yes
WB761c	7 June 1995	Mt. Jefferson, NH	Yes
WB762b	8 June 1995	Mt. Jefferson, NH	Yes
WB763	8 June 1995	Mt. Jefferson, NH	No
WB764a	8 June 1995	Mt. Jefferson, NH	Yes
WB765b	10 June 1995	Mt. Osceola, NH	Yes

Table 2.1. Bicknell's Thrush sample data. (Continued)

Sample ID	Capture Date	Location	Sequenced (yes/no)*
CV770a	24 May 1997	Mink Hollow, NY	Yes
CV771c	24 May 1997	Mink Hollow, NY	Yes
CV772b	24 May 1997	Mink Hollow, NY	AF373791
CV773a	24 May 1997	Mink Hollow, NY	AF373792
CV774e	26 May 1997	Stony Clove, NY	AF373793
CV775a	27 May 1997	Stony Clove, NY	Yes
CV776a	27 May 1997	Stony Clove, NY	Yes
CV777d	27 May 1997	McKenley Hollow, NY	Yes
CV778a	27 May 1997	McKenley Hollow, NY	Yes
CV779c	28 May 1997	Round Top, NY	Yes
GV780a	31 May 1997	Searsburg, VT	Yes
GV781i	2 June 1997	Woodford, VT	AF373794
GV782	3 June 1997	Stamford, VT	No
GV783a	3 June 1997	Woodford, VT	Yes
GV784	3 June 1997	Woodford, VT	No
GV785a	10 June 1997	Plymouth, VT	Yes
GV786j	10 June 1997	Plymouth, VT	AF373795
GV787a	10 June 1997	Plymouth, VT	Yes
GV788a	15 June 1997	Waterbury, VT	Yes
GV789	15 June 1997	Waterbury, VT	No
GV790a	15 June 1997	Waterbury, VT	Yes
GV791h	15 June 1997	Waterbury, VT	AF373796
GV792g	16 June 1997	Waterbury, VT	AF373797

Table 2.2. Veery sample data. Table continues on following page.
The same convention applies to sample ID numbers as in Bicknell's Thrush (Table 2.1) (V=Veery). The shift in number sequence is caused by the change in numbering between band strings (100 bands each). * A representative of each unique sequence identified was submitted to GenBank, these are indicated by GenBank accession numbers. More details of locations and morphometric data for each bird may be found in Appendix 1. Aligned sequences for all individuals sequenced may be found in Appendix 2.

Sample ID	Capture Date	Location	Sequenced (yes/no)
AV793b	17 June 1997	Elizabethtown, NY	Yes
AV794d	17 June 1997	Elizabethtown, NY	AF373798
AV795a	17 June 1997	Elizabethtown, NY	Yes
AV796a	17 June 1997	Elizabethtown, NY	Yes
AV797a	18 June 1997	Wilmington, NY	Yes
AV798a	18 June 1997	Wilmington, NY	Yes
AV799a	19 June 1997	Black Brook, NY	Yes
WV800a	22 June 1997	Woodstock, NH	Yes
WV201c	22 June 1997	Woodstock, NH	AF373790
WV202a	23 June 1997	Thornton, NH	Yes
WV203b	23 June 1997	Thornton,NH	Yes
WV204f	23 June 1997	Livermore, NH	AF373799
WV205b	24 June 1997	Carroll, NH	Yes
WV206a	24 June 1997	Carroll, NH	Yes
WV207a	24 June 1997	Bethlehem, NH	Yes
WV208a	24 June 1997	Bartlett, NH	Yes
WV209a	24 June 1997	Bartlett, NH	Yes
AV210a	30 June 1997	Black Brook, NY	Yes
AV211a	30 June 1997	Black Brook, NY	Yes
AV212a	30 June 1997	Black Brook, NY	Yes

Table 2.2. Veery sample data. (Continued)

Sample ID	Capture Date	Location	Sequenced (yes/no)*1
Gray-cheeked Thrush			
NG744a*2	30 June 1994	Gros Morne NP, NF	AF373800
LG766a	25 June 1995	Miron River, LAB	Yes
LG767b	27 June 1995	Miron River, LAB	AF373801
NG768	23 June 1996	Pistolet Bay, NF	No
NG769a	23 June 1996	Sainte Lunaire, NF	Yes
Hermit Thrush			
H213b	18 October 1997	Rensselaerville, NY	AF373803
H214a	18 October 1997	Rensselaerville, NY	AF373802
Swainson's Thrush *3			
S4002a*4	unknown	Manitoba	Yes
S9001a*4	unknown	Newfoundland	AF373804

Table 2.3. Gray-cheeked, Hermit and Swainson's thrush sample data. The same convention applies to sample ID numbers as in Bicknell's Thrush (Table 2.1) (N=Newfoundland; L=Labrador; G=Gray-cheeked; H=Hermit; S=Swainsons). ***1**: A representative of each unique sequence identified was submitted to GenBank, these are indicated by GenBank accession numbers. ***2**: Gibbs744 was sent as 200 µl of blood/lysis buffer from Dr. H. Lisle Gibbs of McMaster University, Hamilton, ONT, Canada. Collected by Robert Dawson, see photograph p. 363 in McLaren (1995). ***3**: Dates and precise locations were lacking in the information posted with the Swainson's Thrush samples from Dr. H. Lisle Gibbs. The Gibbs group sent a total of nine samples of which I list only the two I obtained sequences from. ***4**: Dr. Gibbs sent Swainson's Thrush samples in the form of isolated DNA. The DNA concentrations of Gibbs4002 and Gibbs9001 were 498 ng/µl and 226 ng/µl respectively.

More details of locations and morphometric data for Gray-cheeked Thrushes may be found in Appendix 1. Aligned sequences for all individuals sequenced may be found in Appendix 2.

A. Phenol/Chloroform DNA Extraction

1) Mix 100 µl blood/lysis buffer with 600 µl Tris-NaCl-EDTA (TNE) buffer, (*p*H 8.0); 30 µl 10% SDS; and 5 µl (20 mg/ml) proteinase K (Sigma).
2) Incubate at 65° C overnight.
3) Extract four times with phenol and once with phenol:chloroform:isoamyl alcohol (1:1(24:1)). Mix by inversion. Spin five min at < 10,000 rpm. Remove aqueous layer (ca. 300 µl) to a new 1.5 ml microfuge tube after each extraction step.
4) Add a half volume (ca. 150 µl) of 7.5 M ammonium acetate, and 2.5 volumes (ca. 1100 µl) of -20° C 100% ethanol. Mix by inversion. Let stand or store in -20° C freezer.
5) Spin 30 min at < 10,000 rpm to pellet DNA. Remove supernatant.
6) Wash DNA pellet twice in cold 70% ethanol.
7) Dry DNA pellet and resuspend DNA in 200 µl 1X TE buffer (10mM Tris-Cl, *p*H 8, 1mM EDTA).
8) Store DNA at 4° C. Quantitate with a spectrophotometer reading (see text) and store long-term in a -20° C freezer.

B. Lithium Chloride DNA Extraction

1) Mix 100 µl blood/lysis buffer in 300 µl of 1X TNE buffer (*p*H 8.0), 15 µl 10% SDS, and 2.5 µl proteinase K (Sigma). Invert gently.
2) Incubate overnight at 60° C with gentle shaking.
3) Add 300 µl 5 M lithium chloride. Invert gently for 60 sec.
4) Add 600 µl of chloroform:isoamyl alcohol (24:1). Invert gently for 30 min.
5) Spin at < 10,000 rpm for 15 min.
6) Remove supernatant (ca. 300 µl), to a new 1.5 ml microfuge tube.
7) Precipitate with 900 µl of cold 95% ethanol.
8) Spin 30 min at < 10,000 rpm to pellet DNA and then discard the supernatant.
9) Wash DNA pellet once with cold 70% ethanol.
10) Dry DNA and then resuspend the pellet in 200 µl 1X TE buffer, *p*H 7.5.
11) Store at 4° C. Quantitate with a spectrophotometer reading, and store DNA in a -20° C freezer.

Table 2.4. DNA extraction protocols. A. modified from Sambrook (Sambrook *et al.* 1989). B. modified from Gemmel and Akiyama (1996)

Name	mt Region of Origin	Sequence
GSC12	12s rRNA	5'-AAGGTTAGGACTAAGTCTTTTG-3'
L16751	tRNA glu	5'-CCAAGATCTACGGCCTGAAAAGCC-3'
ND6C	ND6	5'-CCGAGACAACCCACGCACAAG-3'
L437	Control Region	5'-CTCACGAGAACCGAGCTACT-3'
L817	Control Region	5'-CAAACTTGACACTGATGCAC-3'
H1248	12s rRNA	5'-CATCTTCAGTGTCATGCT-3'
H519	Control Region	5'-TGACCGAGGAACCAGAGG-3'
BTH519	Control Region	5'-TGAAATAGGAACCAGTG-3'
BTL800	Control Region	5'-CATACCCGGATACTGATGCAC-3'

GSC12, L16751, ND6C, L817, and H519 were all used courtesy of Shou-hsien Li (2000); L437 and H1248 were taken from Tarr (1995). I designed BTH519 and BTH800 specifically for this study.

Table 2.5. PCR primers for amplification and sequencing of control region III in thrushes.

A. October 1996 to July 1998:

Reaction Mix (concentration)	**Cycling Parameters**
13.1 µl ddH_2O	Denaturation:
2.0 µl 10X PCR buffer (1X)	95°C for 5 min.
0.8 µl dNTP 5mM each nt (50 µm)	35 cycles at:
1.6 µl 25 mM $MgCl_2$ (2 mM)	94°C for 30 sec.
2.0 µl Primer A 10 µM (0.2 µm)	52°C for 30 sec.
2.0 µl Primer B 10 µM (0.2 µm)	72°C for 60 sec.
0.08 µl Taq DNA Polymerase (1.25 units)	Extension:
0.5 µl Template DNA (varies)	72°C for 10 min.
Overlaid with 10 µl mineral oil	Hold at 4°C

B. August to November 1998:

Reaction Mix (concentration)	**Cycling Parameters**
20.0 µl ddH_2O	Denaturation:
5.0 µl 10X PCR buffer (1X)	94°C for 4 min.
1.0 µl dNTP 10mM each nt (50 µm)	30 cycles at:
4.0 µl 25 mM $MgCl_2$ (2mM)	92°C for 60 sec.
0.88 µl Primer A 10 µM (0.2 µm)	53°C for 60 sec.
0.88 µl Primer B 10 µM (0.2 µm)	72°C for 60 sec.
0.25 µl Taq DNA Polymerase (1.25 units)	Extension:
1.0 µl Template DNA w/37 µl ddH_2O	72°C for 10 min.
Overlaid with 10 µl mineral oil	Hold at 4°C

C. ABI Sequencing Reaction:

Reaction Mix (concentration)	**Cycling Parameters**
8 µl ABI Ready Reaction Mix (see text)	Denaturation:
1 µl Primer (in only one direction)	96 °C for 60 sec.
ddH_2O volume varies with template	26 cycles at:
Template varies with the quantity of	96 °C for 30 sec.
DNA produced/recovered from PCR	50 °C for 15 sec.
and product purification.	60 °C for 4 min.
Total reaction volume 20 µl.	Hold at 4°C
Overlaid with 10 µl mineral oil	

Table 2.6. PCR reaction mixes and cycling parameters. PCR Reactions were executed in a PT-100 thermal cycler (MJ Research) or a 9700 thermal cycler (Perkin Elmer–ABI).

A. Millipore Filter Purification

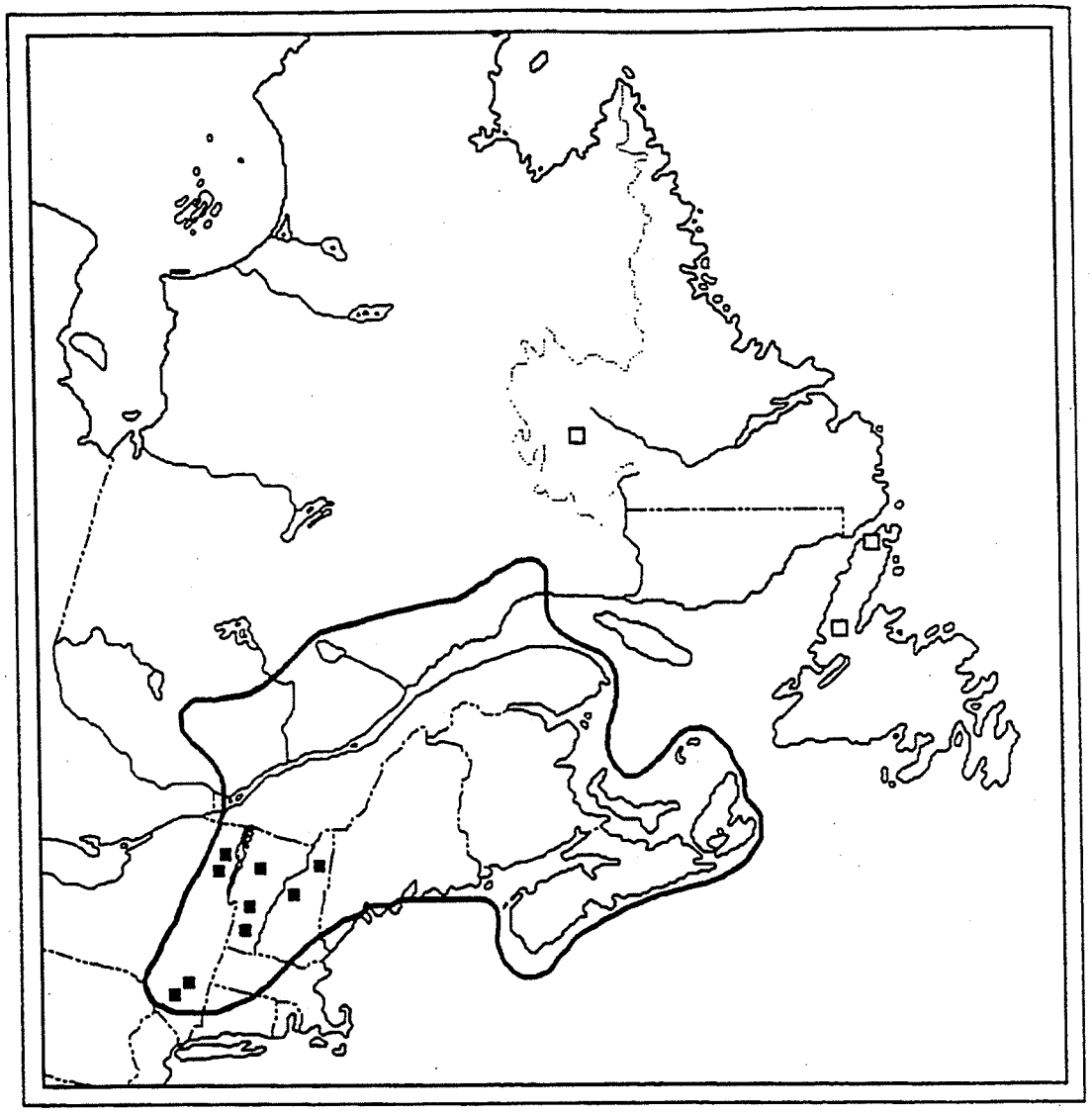
1) Add PCR product to filter. Avoid picking up any of the overlying mineral oil.
2) Add 200 µl of 1X TE buffer pH 8.0 to the product in the filter.
3) Spin from three to four min in a desktop microcentrifuge.
4) Repeat steps two to three for two more times, i.e. three total spin downs through the filter adding 200 µl of 1X TE buffer each time.
5) Remove the filtrate to a new, pre-labeled 600 µl microfuge tube. Mark as 'ultrafree' product.

B. Sephadex Column Purification of Sequencing Reaction Product

1) Place 1.5 ml aqueous Sephadex® G-50 into spin columns and wash tubes.
2) Spin for 30 sec at 5000 rpm in centrifuge, discard filtrate. Spin for a further two min and discard the remaining filtrate.
3) Place columns in labeled 1.5 ml microfuge tubes.
4) Remove PCR product from 600 µl microfuge reaction tube (avoiding mineral oil) and apply to the Sephadex® in the column. Do not touch the surface of the resin beads with the pipette tip.
5) Spin two-and-a-half min.
6) Take filtrate in the 1.5 ml microfuge tube and dry it down in spin vacuum.
7) Store dried sequencing product at -20 °C.

ere

Table 2.7. PCR product purification protocols.



	■ Bicknell's Thrush	■ Veery	□ Gray-cheeked Thrush	
Adirondack Mts, NY	13	10	W. Newfoundland	3
Catskill Mts, NY	17	10	Central Labrador	2
N. Green Mts, VT	9	5		
S. Green Mts, VT	8	8		
White Mts, NH	11	10		

Figure 2.1. Sample collection locations. Number of thrushes captured in each geographic region. Squares indicate approximate limits of sampling rather than the location of specific samples. Bicknell's Thrush range is enclosed by heavy solid line.

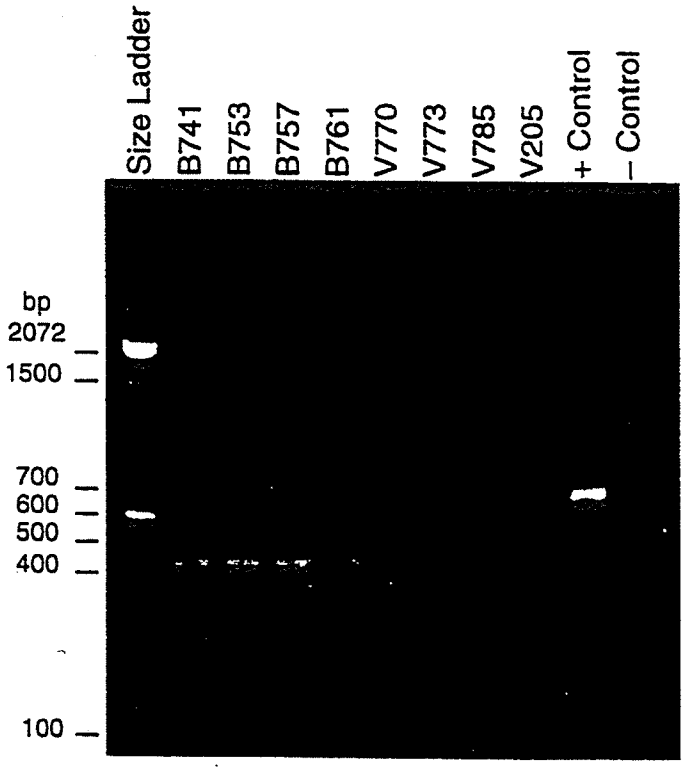
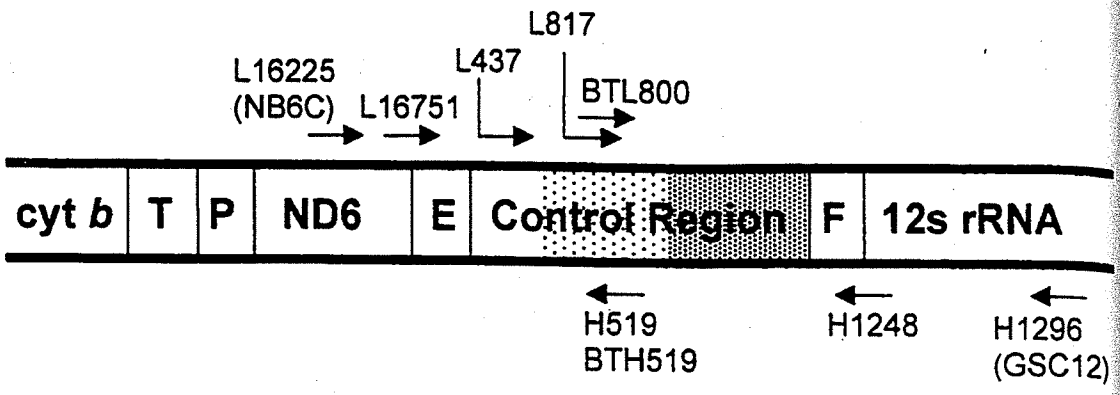


Figure 2.2. Gel of 18 August 1998. 1.5% agarose gel electrophoresed at 80 volts for 90 minutes showing ~420 base pair product of PCR amplification of mtDNA control region III of Bicknell's Thrush (B741 to B761) and Veery (V770 to V205) using primers BTL800 and H1248. (See Table 2.5 for sequences of primers and Table 2.6B for reaction conditions.) 8 µl of reaction product was loaded in each lane. At left is Gibco BRL 100 base pair sizing ladder. Positive control is a ~700 bp lysozyme fragment derived from Hanuman langur using primers FM13 and RM13, courtesy of D. Gilday.



cyt *b*: cytochrome *b*
T: tRNA Threonine
P: tRNA Proline
ND6: NADH dehydrogenase 6
E: tRNA Glutamic Acid
F: tRNA Phenylalanine

Figure 2.3. Primers for control region amplification. Schematic diagram of avian mitochondrial DNA gene sequence around the control region with positions of primers. The stippled area illustrates the 900 bp PCR product obtained via nested PCR; the darker stippling shows the 400 bp of readable sequence. L denotes light strand primer, H denotes heavy strand primer. See Table 2.5 for information on primer sequences, sources, and general locations.