YELLOW WARBLER POPULATION GENETICS

POPULATION GENETIC STRUCTURE AND PHYLOGEOGRAPHY OF YELLOW WARBLERS (*DENDROICA PETECHIA*) INFERRED FROM MITOCHONDRIAL DNA SEQUENCE DATA

By

EMMANUEL MILOT, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (1997) McMaster University

(Biology) Hamilton, Ontario

TITLE: Population Genetic Structure and Phylogeography of Yellow Warblers (*Dendroica petechia*) Inferred from Mitochondrial DNA Sequence Data.

AUTHOR: Emmanuel Milot, B.Sc. (Université du Québec à Trois-Rivières)

SUPERVISOR: Professor H. Lisle Gibbs

NUMBER OF PAGES: x, 101

ABSTRACT

The Yellow Warbler (*Dendroica petechia*) is a highly polytypic bird species with a vast breeding range in the Americas. To assess the level of population structuring within the northern part of its range, I surveyed the nucleotide variation present in a 344bp segment of the mitochondrial DNA (mtDNA) control region I (CR-I) from 155 breeding individuals. These birds were caught at seven locations in Canada and Alaska. Fifty-nine haplotypes were observed in this sample, with pairwise distances between haplotypes ranging from 0.29 to 4.35%. The number of nucleotide sites with multiple hits indicates a high rate of evolution in this region. A homologue to the CR-I was also identified and likely originated from a paralogous duplication event, as suggested by the comparison of sequences from the two regions. Significant population structuring in Yellow Warblers across North America was revealed by analyses of nucleotide diversity and molecular variance, which demonstrated the existence of a major subdivision between eastern (Manitoba to Newfoundland) and western (Alaska and British Columbia) warbler populations. This finding provides evidence for very low levels of gene flow between these two groups. Fifteen out of 21 pairs of populations differ significantly in their genetic composition, indicating further structuring at a smaller geographic scale. Within the eastern group of populations, both the high mutation rate of the CR-I and an isolation-by-distance process seem to be responsible for differences between locations. Phylogenetic analyses indicate that western birds form a monophyletic group whereas

eastern birds are paraphyletic with respect to the western ones. However this conclusion remains hypothetical because of a lack of statistical support for the monophyly of western haplotypes. Nevertheless, this situation is consistent with a historical splitting of warbler populations by a vicariant event, possibly of Pleistocene origin, and provides intraspecific support for vicariance as a mechanism leading to isolation and speciation of western warbler taxa, as hypothesized by Mengel (1964). However, other scenarios, such as a founder event in the west from an eastern stock, cannot be excluded, although they are less likely based on mtDNA data. The absence of phylogeographic structure in the East suggests a recent expansion of Yellow Warbler populations from a restricted geographic range. These findings demonstrate that populations of continentally distributed North American passerine species can show high level of population structuring when assayed with an hypervariable molecular marker such as the mtDNA control region I.

ACKNOWLEDGMENTS

I am grateful to many people without whose help I could not have completed this work and overcome my deep ignorance about genetics. I am particularly grateful to my supervisor Lisle Gibbs, for giving me the opportunity to do my Masters research in his laboratory. Dr. Gibbs provided constant guidance, encouragement and good criticism throughout these years, and was very patient towards a student who was quite far from mastering Shakespeare's language at the beginning. I also thank him for numerous passionate discussions, the main reason for enjoying science. I am thankful to Robert Dawson from whom I learned much about birds and genetics. Robert also taught me about field techniques and sampling. I wish to give special thanks to: Lisa Tabak who processed samples from Manitoba and Alaska; Beeta Matthews who taught me about various aspects and subtleties of lab work; Liliana De Sousa for answering my numerous lab questions and for helping me with cloning. Lorie Collins also provided important assistance during field work and helped me on many occasions by reviewing my written English. She also made sure that no day spent in the lab would be boring, and I thank her for that. I am also grateful to Brian Golding whose countless advice played a major role in the building of a new perception of molecular evolution, and who wiped out my old naïve beliefs on that topic (all of them , I hope !). Fariborg Yadzani guided me through the uncertain world of computers, and explained me the disconcerting caprices of several

programs. We also had many interesting discussions about genetics, and I sincerely thank him for these.

Field work was conducted in collaboration with many people, and I wish to thank Dr. Keith Hobson from the Canadian Wildlife Service, Saskatoon, and his team for collecting samples in Alaska, British Columbia and Manitoba, as well as the staff of the Queen University Biology Station, Ontario, for their help. I thank Reynald Rivard who designed for me a highly efficient system to call birds, which is largely responsible for the success of my field season in 1995. I also want to thank all one and two-days assistants who helped me during my field work in Québec, especially Karine Bernier. Thanks to Nedra Klein for providing me with mtDNA samples from North American Yellow Warblers and sequences from the tropical specimens, to Christopher Wills for the topiary pruning program and to Laila Bastedo for computer assistance. I am also obliged to Dr. Richard Morton who kindly agreed to be on thesis defense committee.

Finally, I am very grateful to my parents who provided constant support during those years, and who saved me from starving more than once.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Tables

Figures

INTRODUCTION

Molecular markers can be used to estimate relatedness between conspecific individuals and, by extrapolation, between populations to which they belong. Therefore, knowing the geographic distribution of genotypes allows one to assess the genetic structure of natural populations. From this information, inferences can be made about demographic processes within species, such as gene flow and genetic drift. In addition, the study of intraspecific genetic variation allows the testing of evolutionary models in the context of coalescent theory. This theory makes testable predictions about phylogenetic patterns expected under certain evolutionary hypotheses (Edwards, 1993a). Its constant refinement (e.g., Slatkin & Maddison, 1989; Nee *et al.*, 1995; Slatkin, 1996) provides useful analytical methods for reconstructing population histories. For example, the genetic patterns revealed in some species have been used to assess issues as diverse as the colonization of islands (Tarr & Fleisher, 1993; Klein & Brown, 1994), the consequences of Pleistocene glaciation on species ranges (Zink & Dittman, 1993a; Wenink *et al.*, 1996), behavioral effects on population subdivision (Lougheed *et al.*, 1993; Friesen *et al.*, 1996) and speciation processes (Bermingham *et al*, 1992).

Perhaps the main type of molecular marker used for such analyses is mitochondrial DNA (mtDNA). The rapid rate at which mtDNA evolves, and its simple mode of inheritance (maternal, non-recombining) have made it a powerful tool to assess levels of polymorphism within species (Avise, 1994). Restriction site analysis, or direct

sequencing using PCR, permit researchers to survey mtDNA patterns in relatively large numbers of individuals in a short period of time. As a result, a more comprehensive portrait of genetic processes in some groups of organisms is emerging, because data now exist for a growing number of species (Avise, 1994; Zink, 1996). This is the case for North American birds, which have received attention from many population geneticists during the past decade.

Population structuring in North American birds

Birds are characterized by their high dispersal capabilities owing to their ability to fly. However many species also show a strong fidelity to their breeding sites (Greenwood & Harvey, 1982), a behavioral trait limiting dispersal and known as philopatry. Due to the contrasting effect of these attributes, whether or not birds exhibit significant levels of genetic structuring remained an "enigma" (Avise, 1996). Early surveys of allozyme variation in birds were reviewed by Barrowclough (1983), who compared inbreeding coefficients (F_{st}) estimated for a number of vertebrate taxa. He reported that F_{st} values in birds are often an order of magnitude lower than for other taxa such as amphibians, reptiles and mammals. Low levels of geographic variation detected with allozymes suggest that gene flow is important between bird populations, presumably as a consequence of their high mobility. However, this observation may also reflect the lack of sensitivity of this type of molecular marker in detecting genetic differentiation. This possibility is illustrated by the comparison of genetic distances

calculated from allozyme versus mtDNA data for four pairs of avian species: two rails, two dowitchers, two grackles and two titmice (Avise & Zink, 1988). For all of these taxa, mtDNA distances between sister species are much greater than those based on allozymes. In addition, higher levels of polymorphism and population structuring have been uncovered from mtDNA than from allozymes in some species (e.g., Van Wagner & Baker, 1990). Therefore, mtDNA may be a more appropriate marker to resolve the enigma of population structuring in birds.

In fact, most recent studies of genetic variation in North American birds have involved mtDNA restriction site analysis. The patterns documented in different species are quite diverse, but indicate that some birds can, indeed, exhibit significant structuring, although generally at a larger geographic scale than similarly distributed but less mobile organisms (Avise & Zink, 1988; Avise, 1994). Yet, several species show limited or no geographic partitioning of mtDNA diversity (Ball & Avise, 1992; Zink *et al.*, 1991; Zink, 1996). An extreme example of this occurs in the prairie grouse complex (genus *Tympanuchus*) composed of three largely allopatric species, two of them having highly fragmented ranges. Nevertheless, a predominant mtDNA haplotype is shared by all species, while other less frequently occurring haplotypes are closely related to that common type, revealing little genetic structuring in grouse species (Ellsworth *et al.*, 1994). Similarly, in studies of the migratory Red-winged Blackbird (*Agelaius phoeniceus*; Ball *et al.*, 1988) and the resident Downy Woodpecker (*Picoides pubescens*; Ball & Avise, 1992), widespread mtDNA haplotypes were found to occur across the

continent. While these situations may reflect high levels of gene flow between different parts of the species ranges, they could also result from a recent expansion of populations from a restricted geographic area (Zink, 1994). In comparison, mtDNA data revealed the existence of five phylogeographic groups in the Deer Mouse (*Peromyscus maniculatus*; Lansman *et al.*, 1983) over a geographic range which is comparable to that of these two birds. A situation close to that observed in the Red-winged Blackbird and the Downy Woodpecker was documented for the Black-capped Chickadee (*Parus atricapillus*) and the Boreal Chickadee (*Parus hudsonicus*). In these species, common haplotypes are widespread across North America, except that they are absent from Newfoundland, and also from Nova Scotia in *P. hudsonicus* (Gill *et al.*, 1993). Moreover, because mtDNA phylogenies in these chickadees are shallow, the division between maritime and continental populations is thought to be recent, and to reflect low ongoing gene flow between these regions. To explain the phylogenetic pattern observed, Gill *et al.* (1993) suggest that chickadee populations have recently expanded their geographic range from a single area. As these birds are resident, some topographic barrier could account for actual haplotypic differences between regions. A contrasting situation occurs in the lesser Snow Goose (*Chen caerulescens caerulescens*), a subspecies breeding in arctic North America and Siberia, for which two important clades were uncovered from mtDNA sequence data (Quinn, 1992). These clades are not geographically partitioned, and are thought to represent an ancient Pleistocene splitting of populations that are now experiencing a secondary contact.

The cases of chickadees and Snow Goose illustrate how mtDNA data can serve to discriminate between the effects of ongoing and long term (historical) gene flow, by providing an historical perspective on genetic differentiation through the reconstruction of intraspecific phylogenies. The ability to distinguish between these is critical to the inference of population structuring, because the genetic pattern observed in a particular species may reflect some historical associations of populations, and not necessary contemporary processes (Strand *et al.*, 1996). However, to make realistic conclusions about ongoing population processes, it is equally important to document adequately the geographic distribution of genotypes. This is because sampling effect can prevent the observation of certain haplotypes or lineages that are nonetheless present in populations, and can lead to wrong conclusions about current levels of gene flow within a species. The risk of making such erroneous interpretations may be reduced by collecting a sufficient number of samples in order to obtain good estimates of haplotypes frequencies in each population assayed. However, that condition has rarely been fulfilled in mtDNA studies of North American birds, as many of them involve small sample sizes (often less than 10 individuals per location; e.g., Van Wagner and Baker, 1990; Gill *et al.*, 1993; Zink, 1993; 1994). Therefore, a more complete survey of DNA polymorphism within populations of a widespread avian species could give a more accurate portrait of actual demographic processes at a continental scale.

While the capacity to discriminate between ongoing and historical gene flow represents an important issue in some species, the existence of distinct mtDNA

phylogeographic units in other species clearly reflects a long history of population isolation. For example, the Fox Sparrow (*Passerella iliaca*) exhibits four major mtDNA lineages, geographically partitioned across North America (Zink, 1994). Although these lineages come into contact where their distributions meet, hybridization between these phylogenetic groups seems to be limited to narrow contact zones. Another example is provided by the Canada Goose (*Branta canadensis*) which is divided into two distinct mtDNA clades partitioned according to a North-South break crossing the continent (Van Wagner & Baker, 1990). This pattern was interpreted as resulting from the fragmentation of goose populations during a glacial advance. Significant population structuring has also been found in other species (some of which will be presented below), but this brief review demonstrates the diversity of population patterns existing in North American birds, as well as the role played by contemporary processes and long term isolation conditions in shaping them. Long term isolation is often attributed to the action of vicariant events, which are environmental changes (e.g., climatic or geological) causing modifications in the geographic distribution of organisms (Avise, 1994). As shown below, vicariance hypotheses involving Pleistocene glaciations have been frequently used to account for genetic patterns within North American species (e.g., Avise, 1992; Quinn, 1992).

Phylogeographic structures resulting from Pleistocene events

Climatic and habitat changes associated with glacial events are thought to have induced important modifications in bird ranges throughout the Pleistocene epoch, since at present, many species occupy previously glaciated areas (Pielou, 1991). Fluctuations in the distribution of the habitat of these species likely created opportunities for the fragmentation of populations and for subsequent allopatric differentiation (Mengel, 1964; Selander, 1965; Hubbard, 1973). The example of sparrows illustrates well some vicariant scenarios invoked to account for genetic patterns within North American birds.

Intensive surveys of mtDNA restriction site polymorphism have been completed for several sparrow species. These species vary in their habitat attributes and distribution over the continent. The Seaside Sparrow (*Ammodramus maritimus*) is distributed along the US Atlantic and Gulf coasts and breeds in salt marsh habitats. Avise and Nelson (1989) found that populations from both coasts form two distinct phylogeographic entities, and hypothesized that these clades have originated as a consequence of the emergence of the Florida peninsula during the Pleistocene, which fragmented the species habitat in two geographically disconnected segments. Rising and Avise (1993) examined mtDNA variation within the Sharp-tailed Sparrow (*Ammodramus caudacutus*), a species with a discontinuous breeding range located along the Atlantic, the southern coast of Hudson Bay and in the Prairies. They found that the birds from the Atlantic coast, south of Maine, belong to a phylogeographic unit distinct from those inhabiting the Prairies, Hudson Bay and eastern Canada. Morphological and behavioral differences between

these two groups were also documented previously (Greenlaw, 1993). Rising and Avise (1993) suggest that the concordance between genetic, morphological and behavioral variation supports Greenlaw's hypothesis (1993) that pro-northern and pro-southern clades were separated during an earlier glacial advance. As a consequence of these studies, "northern" and "southern" Sharp-tailed Sparrows were ranked at the species level (*A. nelsoni* and *A. caudacutus*, respectively) by the American Ornithologists' Union (A.O.U., 1995).

While the geographic ranges of Seaside and Sharp-tailed Sparrows are limited and discontinuous, those of the Song Sparrow (*Melospiza melodia*) and the Chipping Sparrow (*Spizella passerina*) cover roughly all the continent, extending northward to the tree line. Studies of mtDNA variation across North America revealed very little phylogeographic structuring in both species (Zink & Dittman, 1993a; 1993b). However, in the Song Sparrow, there is weak evidence for the splitting of populations into distinct eastern and western refugia during a glacial maximum. This evidence is provided by the rooting of the mtDNA phylogeny with a putative ancestor haplotype, which results in a branching pattern suggesting Newfoundland and the Queen Charlotte Islands as potential refugial areas. But it is not possible to readily distinguish between this two-refugia hypothesis and that of a recent expansion of sparrow populations from a single area. The examples mentioned above illustrate how glaciations can cause long-term isolation of populations which sometimes persists, as in the Sharp-tailed Sparrow, after the retreat of ice barriers. In other cases, these vicariant factors can confine populations to a single

refugium, and hence, limit their level of geographic structuring as hypothesized for the Chipping Sparrow. These examples also show how such historical events can be traced from the actual geographic distribution of genotypes.

Zink (1996) proposed an additional refinement in that type of genetic analysis, which consists of searching for convergence among codistributed species with diverse ecological attributes. If a convergent phylogeographic pattern was uncovered in a number of species, this would suggest that they have responded in a similar way to isolating barriers or vicariant events. This author reviewed studies that used RFLPs analyses to document mtDNA variation in North American birds, but did not find any important convergence. Some species show evidence for a phylogeographic distinction between eastern and western populations, but the sampling effort and geographic scale vary importantly among studies, and such comparisons are of limited value until more complete surveys of mtDNA variation are undertaken for the species considered. Zink (1996) enumerates several factors that could explain the absence of convergence among species, including "differences in: response to barriers or selective gradients, level of gene flow, rates of molecular evolution, effective population size or generation time." Another possible cause is that the species compared may have been codistributed only since a short time, and could therefore have experienced different vicariant histories. Moreover, whether or not a congruent phylogeographic pattern should be expected in different codistributed species is not clear. This is because some glaciation events are relatively recent, and the time scale involved may be too short for intraspecific genetic

structuring and potential convergence to develop. Therefore, the use of a molecular marker evolving faster than mtDNA RFLPs might be more powerful in detecting population structuring caused by recent vicariant events.

Another recent study also used a comparative approach to investigate the role of both Pleistocene events and physical barriers in inducing diversification in birds. As opposed to Zink who reviewed intraspecific patterns in North American species, Brumfield and Capparella (1996) looked at genetic variation at the intrageneric level in several neotropical bird taxa, from three neighbor geographic regions: southern Central America, the Chocó area (South America) and Amazonia. These authors intended to assess whether the Chocó avifauna originally derived from Central American or Amazonian stocks, and also to test for different mechanisms that have been proposed to explain its emergence. These include the isolation of bird populations caused by the Andean mountain uplift, the repeated splitting of the forested range during glacial (or interglacial) periods, and the across Andes dispersal hypothesis. Within a number of genera, Brumfield and Capparella selected a representative species or subspecies for each of the three regions. Using the frequencies of alleles at different allozyme loci, they estimated the relationships between congeneric birds. A majority of these independent phylogenetic comparisons shows that birds from the Chocó area are more closely related to those from Central America than to those from Amazonia. This supports a Central American origin of most Chocó bird taxa examined. So, contrary to what Zink observed in North American birds, some convergence is present among tropical taxa. However, the

prevalent pattern is consistent with the three hypotheses regarding the diversification of birds in that region, and therefore not informative about the mechanism that caused it. But this study provides an example of how phylogenies can serve to infer conclusions about scenarios of differentiation not only at the intraspecific level, but also at the interspecific level. Another example is that of North American warblers for which a genetic approach was used to test a speciation model developed by Mengel (1964) from biogeographic considerations. This model is presented in the following section.

Warbler speciation during the Pleistocene

Interspecific patterns of geographic variation have sometimes been interpreted as evidence for the segregation of North American bird populations into disjoint refugia during Pleistocene glaciations (Selander, 1965; Hubbard, 1973). Palynological records indicate that during these cold periods, forested refugia were located mostly along the Atlantic and Pacific coasts, and in the southernmost regions of the continent. (Pielou, 1991). Mengel (1964) noted that many groups of closely related warbler species, notably the Black-throated Green Warbler (*Dendroica virens*) and the Nashville Warbler (*Vermivora ruficapilla*) groups, show a specific continental pattern in the distribution of species ranges. Each group is composed of a species inhabiting mainly eastern coniferous forests (but occurring in some cases as far west as British Columbia), and of one or more species (or subspecies) with a more restricted range in the western mountain forests. Mengel proposed a vicariance model to explain that distribution and the origin of many

of these wood warblers. He suggested that several modern groups derived from ancestral species that lived in eastern deciduous forests. An initial Pleistocene glacial advance forced these species southward into refugia. Within their new reduced range, some populations adapted to coniferous habitats that were partly confined to the same geographic area. During the following interglacial, these populations were able to expand their distribution more northward and westward than before, owing to their habitat shift. A subsequent glaciation isolated birds living at the western edge of these expanded ranges, and confined them to a distinct refugium, creating an opportunity for allopatric evolution. According to Mengel, successive glacial advances would have led to repeated divisions of ancestral warbler species, and resulted in the emergence of several new species.

This model implies a specific pattern of phylogenetic relationships between members of a group of closely related species. According to Bermingham *et al.* (1992), if species inhabiting western North America evolved independently from an eastern stock, as proposed by Mengel, then each of them should branch off the lineage leading to the modern representatives of that ancestral stock. Using mtDNA RFLPs data, these authors reconstructed the relationships between four of the five species composing the *D. virens* group. Assuming that the gene tree is representative of the species tree, the topology observed is only partly consistent with the model. Not all species seem to derive from an eastern ancestor, and the phylogeny rather suggests a more complex history involving further differentiation in at least one of the western daughter species. From examples

introduced earlier, I showed that the type of Pleistocene events invoked by Mengel to explain warbler speciation can also account for genetic patterns within species. Thus, a further refinement would be to test his hypothesis at the intraspecific level. If support is found for it, then the model could be extended to account for diversification within warbler species as well.

General portrait of population structuring in North American birds and objectives of the present study

Some general features about North American bird populations begin to emerge from the studies presented above. First, because many species exhibit little or no structuring, it appears that high levels of gene flow across continental ranges may represent a common situation. Second, since in species that do show significant population subdivision, genetic breaks often do not obviously match physical barriers, then such barriers may have little effect on dispersal of individuals. Rather, vicariant events seem to have played a more important role in shaping these patterns, although the absence of convergence among species suggests that they have responded idiosyncratically to such events (Zink, 1996). But as mentioned earlier, the absence of geographic pattern in some species may also reflect a lack of sensitivity of mtDNA restriction site analyses in detecting population structuring. Direct sequencing of highly variable mitochondrial regions might represent a more powerful approach to uncover

genetic patterns within species and may be more suitable to resolve intraspecific phylogenies (Edwards, 1993a).

Edwards (1993b) examined mtDNA control region I sequence variation in an Australasian bird, the Grey-crowned Babbler (*Pomatostomus temporalis*), and uncovered substantial levels of polymorphism. He found 86 haplotypes in a sample of 163 birds collected across Australia and in Papua New Guinea, a number more important than that found in most studies of mtDNA restriction site polymorphism. He also detected low levels of gene flow between distant populations of babblers. In a second study, Edwards (1993a) found that three geographically close and isolated babbler populations, but known to have been in contact during the Wisconsin glaciation, have probably exchanged migrants since their isolation have begun, that is, around 10 000 years ago. Recently, Wenink *et al.* (1996) assessed the global population structure in Dunlins (*Calidris alpina*) from a worldwide sample of 155 birds. Based on control region sequences, these birds are clearly subdivided into five major phylogeographic units, which are thought to have originated from repeated population fragmentation during the Pleistocene. Moreover, an analysis of molecular variance shows that more than 75% of the genetic variance in the sample is partitioned among these groups, an observation suggesting that population structure is currently maintained through breeding site phylopatry. Sequence data also indicate subdivision of Dunlin populations at a smaller geographic scale in Europe (Wenink *et al.*, 1996). Therefore, control region sequences may constitute a useful marker to infer population structuring at different geographic

scales. However, except for the Dunlin and the Lesser Snow Goose (Quinn, 1992) whose breeding ranges are limited to the arctic zone, no study of mtDNA sequence variation has been conducted for North American widespread birds.

Here I present a study of control region sequence variation in a North American passerine, the Yellow Warbler (*Dendroica petechia*). The Yellow Warbler has one of the largest breeding distribution among New World passerines, covering all North America up to the northern limit of forests, Central America, the West Indies, and the coastal regions of northwestern South America (including northern Peru, Ecuador, Columbia and Venezuela; Browning, 1994). The species is also found in the Galapagos Archipelago. It is taxonomically divided in three groups, each one composed of several subspecies (A.O.U., 1957). North American populations are migratory and spend the winter in neotropical regions. They belong to the *aestiva* group and are referred to as typical "Yellow Warblers". Birds from Central and South America, "Mangrove Warblers", are resident and form the *erythachorides* group. Finally the third and nominal group, *petechia*, includes resident birds from the West Indies ("Golden Warblers"). I will adopt here this terminology in order to distinguish North American birds (Yellow Warblers) from tropical *D. petechia*.

The species was recently the subject of a population genetics investigation by Klein and Brown (1994). These authors looked at mtDNA restriction site polymorphism in a sample of birds from a large part of the species range, with a particular attention to the West Indies. Their main objective was to infer the colonization pattern of these

islands by the species, and their results are consistent with the occurrence of two colonization events. Limited sampling in North America indicates no sharing of haplotypes between individuals from the Washington state and Michigan and North Carolina suggesting a potential subdivision somewhere along a East-West axis.

Six subspecies of Yellow Warblers are recognized over North America (A.O.U., 1957), but Browning (1994) proposed to increase that number to eight. These taxonomical divisions are based on slight variations in breeding plumage coloration (brightness, density of streaking). Yellow Warblers are continuously distributed over the continent and breed in thickets and shrubs along forest, marsh, bog and lake edges, and also in riparian habitats, gardens and orchards (Gauthier, 1995).

I have documented the polymorphism in the hypervariable mtDNA control region I in northern populations (Canada and Alaska) of Yellow Warblers in order 1) to assess whether or not a continentally distributed migratory bird exhibits population structuring, 2) to infer phylogeographic patterns within that species and 3) to evaluate the usefulness of control sequence data in assessing intraspecific genetic patterns.

MATERIALS AND METHODS

1. Sample collection and processing

Sampling: During the 1994 and 1995 breeding seasons, sampling was carried out at seven locations[1] throughout Canada and in Alaska (Fig. 1; Table 1). A total of 155 Yellow Warblers were caught by setting up mist nets in their territories and attracting them in with playback tapes. The breeding status of individuals was inferred on the base of territorial behaviour (males) or the presence of a brood patch (females). Blood samples were collected from the brachial vein of the birds using a 30 ga syringe needle and 75μL capillary tubes, and kept in lysis buffer (10 mM Tris Base, 10 mM EDTA, 10 mM NaOH and 1% n-lauryl-sarcosine [Seutin *et al.*, 1991]) at 4°C until processed.

DNA extraction: I isolated total genomic DNA by phenol/chloroform extraction according to the following protocol: I added 500 μL of lysis buffer and 30 μL of proteinase K to 100 μL of the blood/lysis buffer mixture and left the samples rotating at 37°C overnight. Then, I added 600 μL of phenol/chloroform to the samples and shook the tubes and subsequently spin them in a microfuge for 5 minutes at 16,000g. After centrifugation, I transferred the supernatant into a new tube and repeated the procedure twice by adding an extra 600 μL of phenol/chloroform the first time and only chloroform

[1] Abbreviations frequently used in this paper are: ALA: Alaska; BCO: British Columbia; MAN: Manitoba; NBW: New Brunswick; NFL: Newfoundland; ONT: Ontario; QUE: Québec; mtDNA: mitochondrial DNA; CR-I: control region I.

the second time. Next, I precipitated the DNA with 3M NaOAC (about one tenth of the volume in the tubes) and ~1000 μL of 100% Ethanol. I finally washed the pellet with 100 μL of 70% Ethanol and resuspended it in distilled deionized water.

Primer development and sequencing strategy: I developed primers to amplify the regions I (CR-I) and II (CR-II) of the Yellow Warbler mitochondrial control region (Fig. 2). These regions are located at the 5' and 3' ends, respectively, of the control region, and have been described as 'hypervariable' because of the high level of polymorphism they exhibit in some species (Wenink *et al.*, 1993). I designed specific primers for CR-I from Yellow Warbler sequences amplified and sequenced with two other primers (L16743 and H417) from Tarr (1995). I used primers (Miller, 1993) annealing to the 12S RNA gene (GSH-12S) and the central domain (GSL-GLU) of the control region to amplify a 247bp section of the central domain and also to develop Yellow Warbler specific primers for the amplification of 133bp of the CR-II. A preliminary survey of nucleotide polymorphism uncovered little variation in the central domain and the CR-II (Table 2). Therefore, I decided to focus on the more variable CR-I and to sequence that region for the whole set of 155 samples.

DNA amplification and sequencing: I used the extracted DNA as a template and primers LDPdl5 and HDPdl4 (Fig. 2) in 50μL PCR amplifications of a 344 bp fragment of the CR-I. PCR mixes were composed of 100ng of DNA, 2.5mM of buffer (100mM Tris-

HCL + 500mM KCL), 2.5 mM of $MgCl_2$, 0.5μM of primers, 0.2μM of dNTPs and 1.25 Unit of AmpliTaq™ polymerase (Perkin-Elmer). Each reaction consisted of 35 cycles of denaturation at 94°C, annealing at 60°C and extension at 72°C, all steps being of one minute duration. I assessed the size of the products obtained by running 5μL of sample next to a ladder on an Ethidium Bromide-stained 1% Agarose gel made of 1x TAE buffer. I then purified products of successful amplifications by transferring them into polysulfone membrane filter units (Millipore Corp., 30 000 NMWL) and washing them three times by adding 2-300μL of water and spinning down for 6-8 minutes at 5.2g. I sequenced both strands of the purified double-stranded DNA using the AmpliCycle™ (Perkin-Elmer) cycle-sequencing kit and the same primers used for the amplification, end-labeled with ^{33}P. I used a third primer (LDPdl6, Fig. 2), annealing 88 bp downstream of LDPdl5, to clarify a small number of ambiguities present in some sequences. All sequencing reactions consisted in 25 cycles of 1min denaturation at 94°C and 1min annealing at 60°C. I ran sequencing products on 6% polyacrylamide denaturing gels and visualized them by gel autoradiography.

2. Control region homologues

I originally conducted amplification and sequencing of the CR-I with an annealing temperature of 50°C rather than 60°C. PCR product run on Agarose gels resulted in a single band of the expected size when compared to the ladder, suggesting that only one product was synthesized during the reactions. However, many nucleotide

positions on sequencing gels showed double bands, indicating that the primers might anneal at more than one place on the genome. I solved this problem by increasing the annealing temperature to 60°C, both in the original PCRs and sequencing reactions. This resulted in much clearer sequences with low ambiguities. To verify that the region amplified under these new conditions was located on the mitochondrial genome, I did the following check. Four *D. petechia* mtDNA samples, collected for a previous study by Klein and Brown (1994), and which had been purified with CsCl gradient ultracentrifugation, were amplified and sequenced using the method described above (annealing at 50°C). These amplifications were successful and the results revealed that the four birds had sequences very similar or identical to others found in the present study, based on amplifications from blood-extracted DNA.

Nuclear homologues of mitochondrial genes have been found in a number of organisms (Zhang & Hewitt, 1996) and may represent a real problem in studies using PCR to document DNA polymorphism. To assess the origin of double bands on my sequencing gels, four samples (two from Québec and two from British Columbia) were selected for cloning. All were amplified according to the method described above, but using an annealing temperature of 50°C. I expected that, under these conditions, both CR-I and homologues to be amplified and hence identifiable when sequenced. PCR products were then ligated into pBluescript®II SK plasmid vectors (Stratagene). These vectors had been previously digested with a blunt-cutting restriction enzyme (*Eco*RV) and modified (following Hoelzel & Green, 1992) to include an overhang T-residue to

their 3'-end, in order to facilitate the cloning of PCR products. *Echerichia coli* strain XL1 cells were then transformed with the ligation mixtures and these cells were grown on agar plates for 18 hours, at 37°C. The agar medium contained ampicillin (50mg/mL) to prevent the growth of non-transformed cells, and X-gal and IPTG in order to identify recombinant (white) colonies. For each sample, six positive clones were collected and grown up in Terrific Broth (Tartof & Hobbs, 1987) for 18 hours, at 37°C. The cells were then harvested, and a alkaline lysis mini-prep procedure was performed (Sambrook *et al.*, 1989) to purify the plasmid DNA. Each clone was sequenced with the Sequenase kit v.2.0 (United States Biochemicals) following the manufacturer instructions and T3 and T7 universal primers, using the ^{35}S-dATP incorporation procedure.

3. Statistical analyses

Test for neutral evolution: An important assumption in inferring population processes from mtDNA data is that the mitochondrial genome evolves neutrally, implying that migration and genetic drift are the primary factors affecting haplotype frequencies in populations (Ballard & Kreitman, 1995). Violation of that assumption may lead to incorrect interpretation of mtDNA polymorphism. Therefore, I applied Tajima's test (1989) to the *D. petechia* data to test for the neutral evolution of the control region I.

Haplotypic diversity and population structuring: Haplotypic diversity in each sampled locality was calculated from the index $h = (1-\Sigma x_i^2)n / (n-1)$, where n is the sample size and x_i^2 the frequency of the i^{th} haplotype in the locality (Nei, 1987).

To assess the population structuring of Yellow Warblers across North America, I used a hierarchical analysis of nucleotide diversity as described by Holsinger and Mason-Gamer (1996). This procedure involves the following steps: first, pairwise estimates of the genetic diversity partitioned between sampled locations are computed using the $\hat{g}_{ST}$ index. This statistic corresponds to the probability that two haplotypes drawn at random differ at a nucleotide site, and corrects for average nucleotide diversity within populations (Holsinger & Mason-Gamer, 1996); it is computed from a matrix of pairwise distances between haplotypes. Second, the two localities showing the smallest $\hat{g}_{ST}$ value (hence, the smallest genetic divergence) are then grouped together, and pairwise $\hat{g}_{ST}$ indexes are then recomputed between the remaining localities and the new cluster. The analysis proceeds until all locations have been added, and a tree of the relationships between them is obtained. Statistical support for $\hat{g}_{ST}$ values was assessed by creating, for each node in the populations tree, 1000 replicates of random resampling of haplotypes, and keeping sample size in each population equal to that in the original data set. Null distributions of $\hat{g}_{ST}$ were generated from these replicates and the values obtained from the original data compared to them. I wrote a program in TurboPascal in order to perform the hierarchical analysis of nucleotide diversity, to create bootstrap replicates and to conduct Tajima's test of neutrality.

For an additional assessment of population structuring, I also conducted an analysis of molecular variance (AMOVA; Excoffier *et al.*, 1992). In contrast to the previous method, this one requires that we define groups of populations *a priori*, based on considerations external to the analysis. From a matrix of pairwise distances between haplotypes, the analysis computes three components of the total genetic variance in the sample: the within-populations variance, the variance partitioned among populations within groups, and the variance among groups. This information is then used to measure haplotypic correlations expressed by "Φ"-statistics (defined in Table 5), in a way similar to *F*-statistics (Excoffier *et al.*, 1992). Based on the result of the hierarchical analysis of diversity, I divided the Yellow Warbler populations into two groups and conducted a nested analysis of variance on them. The first group is composed of ALA and BCO, and the second one includes all other localities (MAN, ONT, QUE, NBW, NFL). Corrected distances, accounting for a biased transistion-transversion ratio and for evolutionary rate variation among nucleotide sites (see next section), were used in the AMOVA and analysis of nucleotide diversity.

Phylogenetic analyses: Because a bias towards transitional over transversional changes has been documented for the mitochondrial genome (Wakeley, 1996), I estimated the ratio, *R*, of transitions to transversions in the Yellow Warbler control region sequences using the following method: Kimura 2-parameters distances between haplotypes were computed (DNADIST in PHYLIP; Felsenstein, 1993) using the empirical ratio ($R = 6.0$)

observed in the data set. I used these distances to reconstruct the relationships between all haplotypes with the neighbor-joining method (Saitou & Nei, 1987; NEIGHBOR in PHYLIP). The tree was rooted with three tropical *D. petechia* sequences (see below). Then, to estimate the total number of changes that occurred since warblers have diverged from a common ancestor, I assigned nucleotide states at polymorphic sites to the phylogeny branches according to a parsimony criterion. This resulted in a global estimate of 98 substitutions, and in a new transition-transversion ratio of 13.0, roughly twice that observed originally.

In addition, the number of substitutions *per se* estimated from the neighbor-joining tree allowed me to build a distribution of the number of sites affected by different number of changes (Fig. 3). Since variation in the rate of substitution among sites are reported for several mitochondrial and nuclear genes (Yang, 1996), distances between haplotypes were corrected assuming a gamma distribution of rates. I calculated the alpha (α) parameter of the gamma function from the equation $\mu^2/Var(\mu) = 1/\alpha^2$ where μ is the average number of substitutions per site. Kimura 2-parameters distances were recomputed using the new transition-transversion ratio and the value for the gamma correction of 0.51. These distances also served for the hierarchical analysis of nucleotide diversity and the AMOVA.

I reconstructed a neighbor-joining phylogeny from these corrected distances. The tree was rooted with five outgroup sequences. Three are from tropical *D. petechia* (two Mangrove Warblers and one Golden Warbler, kindly provided by Nedra Klein) and

two from other *Dendroica* species (*D. striata* and *D. pensylvanica*). The Mangrove Warblers are from the Pacific coast of Costa Rica and from eastern Venezuela, and the Golden Warbler is from Puerto Rico. Statistical support for the nodes in the phylogeny was evaluated by the bootstrap test (bootstrap replicates generated with SEQBOOT in PHYLIP) and by the branch-length test (Rzhetsky & Nei 1992; performed with MEGA; Kumar *et al.* 1993).

The large number of taxa (64 including the outgroups) prevented me from using the maximum parsimony approach to estimate relationships between all haplotypes, owing to a lack of computer facility. To cope with this problem, I reduced the number of taxa with the topiary pruning method recently proposed by Wills (1996). This approach aims to eliminate homoplastic subsitutions occurring along shallow lineages that represent uninformative "noise", partly hiding the phylogenetic signal needed to infer the basal structure of the tree. The method works as follows: a consensus sequence is built using all taxa by selecting the commonest bases at each nucleotide site. Then, the complete data set is duplicated in a number of identical copies, corresponding to the "pruning level" desired. All pairs of sequences in the original data set are then compared to determine the number of nucleotide differences between each of them. If, for example, a pair differs at three sites, then the two sequences are compared with the consensus at these three sites. Variant bases, i.e., those different from the consensus sequence, are removed (considered as unknown) from the taxa in the third copied data set (corresponding to pruning level 3). These variants are also deleted from the same taxa in

higher-numbered data copies (>3 in this example), while lower copies are unmodified by the current pairwise comparison. Note that the original data set remains intact. The outcome of this procedure is that in each pruned set, some sequences will have become identical, so that the number of different taxa will be smaller than in the original set. Each taxon will represent either a single sequence or a number of merged ones. This approach is less drastic than others involving the complete deletion of certain sites rather than specific bases at some sites (Wills, 1996).

A True Basic computer program, written by C. Wills, was used to perform topiary pruning on *D. petechia* sequences. For each resulting data set, the program provides the number of duplicated (removed) taxa and the number of informative sites per remaining taxon. This information serves to evaluate the efficiency of the pruning procedure at different levels. The procedure was carried out up to pruning level 10, which happened to be more than sufficient, as 60 taxa out of 64 were eliminated once this final level was reached.

I built maximum-parsimony phylogenies using PAUP 3.0 (Swofford, 1991) from data sets obtained by pruning levels 2 and 3. For the first set, an heuristic search was done using the tree bisection-reconnection branch swapping option, and keeping all minimal trees. Unfortunately, computer facility was lacking to conduct the bootstrap test within a reasonable time. For the second set, a branch-and-bound search was performed. Two hundred bootstrap replicates were computed to estimate the support of nodes in the phylogeny.

RESULTS

1. Polymorphism in *D. petechia*

A total of 32 polymorphic nucleotide sites (9.3%) were identified within the 344bp Yellow Warbler control region I sequences (excluding the outgroup taxa). This value increases to 44 (12.8%) when the sequences from the three tropical *D. petechia* are added in the analysis, and to 54 (15.7%) when all outgroups are included. Figure 4 illustrates the CR-I sequences from two birds. Nucleotide frequencies show a deficiency in the G + C content, a situation previously reported for the mitochondrial (Tamura, 1992). The average frequencies for each nucleotides are: A=30.6%, T=23.9%, G=14.2% and C=31.2%. A total of 59 haplotypes were found among the 155 Yellow Warblers sequenced (Fig. 5). All variants were identified on the base of nucleotide substitutions, as no insertions or deletions were observed. Pairwise distances, calculated as the proportion of sites at which two haplotypes differ, average 1.88%, and range from 0.29 to 4.35%. The maximal value occurs between a haplotype found in Québec (E24) and one from British Colombia (W1). This magnitude of intraspecific variability is comparable to that detected in control region data in other avian studies (Edwards, 1993a, 1993b; Wenink *et al.*, 1996).

Tajima's *D* statistic value (-2.35) reveals a significant departure from the neutral evolution expectation ($p<0.01$; Table 2 in Tajima, 1989). The negative *D* indicates an

excess in the number of segregating sites relative to the average pairwise difference between haplotypes.

2. Control region homologues

For two of the four individuals selected for cloning, all of the clones (6) examined had identical sequences to those obtained from the same samples by cycle-sequencing. For each of the two other samples cloned (one from Québec and one from British Columbia), two different sequences were obtained: one identical to the cycle-sequencing product, corresponding to the CR-I, and another which differed at a number of sites. This confirms that two different DNA fragments were amplified when an annealing temperature of 50°C was used in the PCRs.

The two homologue sequences were aligned with the control regions from the same individuals using the program CLUSTAL (Higgins & Sharp, 1989), with penalties for gap insertion and extension of 5 and 1, respectively (Fig. 6). The homologues differ from each other by 1.2% (2 transitions and 2 transversions; excluding gaps), and the two CR-I sequences by 2.9%. However, the 'within-individual' distances between the CR-I and the homolog are much greater: 14.5% for the Québec bird and 17.0% for the British Columbia bird, excluding gaps. This is roughly twice the highest divergence observed (8.4%) between a *D. petechia* and an outgroup species CR-I sequence.

3. Geographic distribution of mtDNA variation

Most haplotypes are represented by a single individual, and therefore occur as private haplotypes (i.e., present at only one location; Appendix I). This results in high values of haplotypic diversity within all locales, except for Alaska (Table 1). This location has, with New Brunswick, the smallest sample size (n = 11), although the number of observed haplotypes is 3.5 times larger in New Brunswick. While Alaska haplotypes are found in British Colombia as well, these two localities share only one haplotype with any of the five other localities; it is found in a single bird from Manitoba. Similarly, Manitoba has a unique haplotype in common with the four other provinces (ONT, QUE, NBW and NFL). These eastern populations share their haplotypes to a varying extent. This geographic partitioning of the genetic diversity suggests significant population structuring. But the great proportion of private haplotypes and the small divergence between some of them (many differ at two or three nucleotide sites), probably also reflects the effect of the high mutation rate of the CR-I (Quinn, 1992; Vigilant *et al.*, 1991, Wenink *et al.*, 1993).

Table 3 shows three measures of the proportion of the total genetic diversity in the sample that is partitioned among locations. The G_{st} statistic (Nei, 1987) is computed from haplotype frequencies data, while Φ_{st} and $\hat{g}_{st}$ also incorporate the extent of divergence (pairwise distances) between haplotypes. The Φ_{st}-statistic represents the proportion of the genetic variance partitioned among populations (Excoffier *et al.*, 1992), and the $\hat{g}_{st}$ index the fraction of the nucleotide diversity (π) occurring among them (Holsinger & Mason-

Gamer, 1996). The G_{st} value is four times smaller than that obtained for the two other statistics, hence the level of population structuring inferred from the molecular data is much higher than when haplotype frequencies alone are used. This difference results, on one hand, from the fact that haplotypic diversity within localities is high (Table 1), limiting the relative importance of the diversity partitioned among them. On the other hand, the high Φ_{st} and $\hat{g}_{st}$ values indicate that, overall, the extent of sequence divergence between haplotypes is much lower within than among populations.

Figure 7 shows the partitioning of the genetic variation among the seven localities, as inferred from the hierarchical analysis of nucleotide diversity. The analysis identifies a major split between a "eastern" set of populations (NFL, NBW, QUE, ONT and MAN) and two populations (BCO and ALA) found on the western side of the Rocky Mountains, as more than 50% of the total genetic diversity in the sample is partitioned between these groups.

Within the eastern group, the geographically distant Newfoundland and Manitoba localities cluster together, while New Brunswick branches out of the group composed of MAN, ONT, QUE and NFL. However, pairwise differences (not shown) within the eastern group are relatively small ($\hat{g}_{st}$ values range from 0.025 between NFL and MAN to 0.137 between QUE and NBW) compared to those between "eastern" and "western" populations ($\hat{g}_{st}$ ranging from 0.382 between NBW and BCO to 0.738 between ONT and ALA). The negative $\hat{g}_{st}$ observed for the QUE-ONT cluster in figure 7 can be interpreted as due to sampling effect, and provides evidence for a very high (undefined) level of gene

flow between localities (Hudson *et al.*, 1992). This suggests that the two sampling sites are part of the same panmictic population.

Pairwise Φ_{st} values provide a very similar description of genetic differences between locations (Table 4), with the only non significant values occurring between Québec and Ontario, and between Alaska and British Columbia. However, the application of the sequential Bonferonni technique (Rice, 1989) indicates that four extra pairwise comparisons are non-significant at the 0.05 level. If we select the highest p-values and consider them as non-significant, then all four comparisons involve pairs of eastern locations: NFL-MAN, NFL-NBW, QUE-MAN and MAN-NBW (Table 4). Again, the values between eastern and western locales (Φ_{st} ranging from 0.63 to 0.82) are larger than those observed within these two groups (-0.01 to 0.33). The nested analysis of variance on these groups indicates that a large proportion of the genetic variance (66.6%) is partitioned among them (Table 5), a result concordant with the hierarchical analysis of nucleotide diversity. Comparatively, the variance among populations within groups is very small (4.4%), but reveals some significant structuring. Within groups, the genetic variance occurs mainly within populations (29.0% of the total). This situation is similar to that observed between Dunlin populations (Wenink *et al.*, 1996) from a sample of birds collected across the northern hemisphere.

4. Phylogenetic analyses

Figure 8 shows the neighbor-joining reconstruction of phylogenetic relationships between all *D. petechia* haplotypes. Two important patterns are evident: First, haplotypes found in the western populations form a distinct clade and are thus monophyletic. Second, no obvious geographic structuring of mitochondrial lineages is observed in the East, as haplotypes occurring in populations are spread throughout the different lineages. A minor exception consists of the cluster including E17, E26, E29, E33, E36 and E38. The latter five haplotypes are from Manitoba only, but that clade is weakly supported (bootstrap value < 50%). The western haplotypes coalesce with a smaller cluster composed of two eastern ones (E20 from NFL and NBW, and E39 from NBW); hence, eastern birds are paraphyletic with respect to those from the West. Other neighbor-joining trees were constructed using different transition/transversion ratios and gamma corrections. Tamura-Nei distances (Tamura & Nei, 1993; in MEGA) were also used because of the unequal nucleotide frequencies. All these phylogenies show the same clustering of western haplotypes with E20 and E39. This result suggests the existence of two phylogeographic units at the continental scale. The first composed of the western Yellow Warblers while the second one comprising the paraphyletic birds inhabiting the East. However the bootstrap support for the western clade is low (78.8%), as well as for other nodes in the phylogeny, except for the one grouping haplotypes E20 and E39 (93.2%) and the node distinguishing the outgroup cluster form Yellow Warbler haplotypes (99.4%). This latter value indicates that all taxa used to root the tree are true outgroups to North American *D.*

petechia with regards to control region I sequences. The branch-length test result cannot be directly compared to the bootstrap test, because the program MEGA does not allow the user to enter a pre-selected transition-transversion ratio for computing pairwise distances. Thus, these distances are slightly different than those used for the neighbor-joining reconstruction with DNADIST in PHYLIP, for which I used a ratio of 13, as mentioned above. Hence, the trees obtained from both programs slightly differ in their branch lengths and topologies. Nevertheless, the support for the western clade from the branch-length test is 85%, thus higher than that provided by bootstrap replicates, but still below the significance level of 95%. Thus, the monophyly of western haplotypes remains hypothetical because of this lack of statistical support. A general and important feature visible in the topology is that a greater diversity of presumably ancient lineages, or at least "deeper" in terms of sequence divergence, is represented by eastern haplotypes. This does not constitute an artifact of the larger sample size in the East (n=118) than in the West (n=37), because by selecting any single eastern location, one still finds a larger variety of these deeper lineages, as pictured in Figure 8.

The topiary pruning procedure had a radical effect in reducing the number of taxa for subsequent phylogenetic analyses using maximum parsimony (Fig. 9). At pruning level 2, 25 taxa are remaining and the number of informative sites per taxa increases from 5.6 (level 1) to 12. At level 3 there are only 13 taxa left and 19 informative sites per taxa. Moreover, all western haplotypes have been merged into a single one (W1; Fig. 10) and no eastern haplotype is included in that western representative. The maximum parsimony

trees reconstructed from pruned data sets show the same basal structure as the neighbor-joining trees: the western haplotype representatives cluster with a taxon resulting of the merging of E20 and E39. The bootstrap value for that clade is 82% compared to 72.2% in the neighbor-joining phylogeny. Only the tree for pruning level 3 is shown in Figure 10, as the one for the second level has the same structure, and since bootstraps could not be obtained for that tree (see Materials and Methods).

DISCUSSION

The present study provides evidence for the rapid evolution of the mitochondrial control region I and its usefulness in population genetics studies. In addition, my results show that a continentally distributed passerine species can exhibit important population structuring, which contrasts with the situation documented in other widespread species such as the Red-winged Blackbird (Ball *et al.*, 1988) and the Downy Woodpecker (Ball & Avise, 1992). These findings will be discussed here and compared to those of a recent survey of microsatellite variation in the Yellow Warbler (Dawson *et al.*, in preparation). Their implications for the significance of subspecies divisions in *D. petechia* and conservation purposes will also be presented.

1. Evolution of the control region I

Three different aspects of the CR-I evolution can be addressed from the *D. petechia* sequence data: the mutation rate, the assumption of neutral evolution, and the existence of homologous sequences.

1.1. mutation rate

The extent of sequence polymorphism within *D. petechia*, measured from pairwise distances and haplotypic diversity, confirms the high evolutionary rate of the CR-I documented elsewhere, relative to other mitochondrial markers. For example, in a

species of marine turtle, the CR-I substitution rate estimated from sequence data is about eight times higher than the average mtDNA rate inferred from RFLPs (Norman *et al.*, 1994). Likewise, Cann *et al.* (1987) documented mtDNA restriction site polymorphism in humans from 147 samples collected from five populations throughout the Old World. They identified 133 haplotypes in that sample and the average pairwise divergence between haplotypes within populations ranged from 0.01% to 0.47%, with the highest average value observed between individuals from Africa. Vigilant *et al.* (1991) assayed control region sequence variation (1122bp) from 189 human mtDNAs collected over a similar range. Although the number of haplotypes (135) was very close to that detected by Cann *et al.*, average haplotypic distances were much greater (over 2% among African individuals). The comparison of sequence data from the present study with mtDNA RFLP data reported by Klein and Brown (1994) indicates that CR-I evolves faster than the rest of the mitochondrial genome. Klein and Brown found 37 haplotypes in a sample of 194 birds (0.19 hapl./ind.) from the entire species range (tropical and temperate regions), while I detected 59 (0.38 hapl./ind.) in the present data set (155 birds, excluding outgroups). In addition, the range of pairwise distances between haplotypes was similar for both studies, despite the fact that Klein and Brown' sample included presumably more divergent individuals from both tropical and North American *D. petechia* populations.

However, whether or not the high rate of evolution of the CR-I makes it a more suitable marker to detect population structuring remains uncertain. This is because such a rate can increase the haplotypic diversity within populations so high that the relative

importance of the diversity partitioned among them decreases. That problem can occur when one uses analytical methods that compare genetic variation within and between populations to infer structuring, such as the analysis of molecular variance or the G_{st} index. This situation was observed by Wenink *et al.* (1993) who examined the level of genetic variation in the cytochrome *b* gene, the CR-I and the control region II (CR-II), present in a sample of 73 Dunlins from around the world. Among these regions, the CR-I was the most variable, including 71% of the polymorphic sites surveyed in that study. Wenink *et al.* detected a smaller level of structuring from CR-I data than from CR-II and cytochrome *b* gene data. In the present study, however, the rate of evolution of the CR-I could partly account for the significant genetic differences between some pairs of populations, as discussed later.

1.2. Test of neutrality

The result of Tajima's test must be interpreted with caution, because one of its assumption is violated by the present data set. This assumption is that the CR-I evolves under the infinite-sites model, and that new mutations always arise at new nucleotide positions (Fu, 1996). According to the estimate of the total number of substitutions that occurred along *D. petechia* lineages, it is evident that several sites have mutated more than once (Fig. 3). Fu (1996) suggests that, when multiple hits are obvious, one may correct for this by estimating from a phylogeny the parameters required for neutrality tests such as Tajima's. Therefore, I performed Tajima's test by substituting the number of

polymorphic sites (a parameter of the test) with the total number of substitutions, estimated from the *D. petechia* phylogeny (excluding all outgroups). Gamma distances were also used instead of non-corrected distances. However, the new *D* value obtained is even lower (-3.10) than the first one, indicating an even greater departure from the neutral model expectation. Because of the sharp and important subdivision between eastern (MAN, ONT, QUE, NBW, NFL) and western (ALA, BCO) warbler populations revealed by the analyses of population structure, I also applied Tajima's test independently to both groups. A comparable departure from neutrality is observed for the two groups (eastern group: $D = -2.55$, western group: $D = -3.02$; $p<0.01$ in both cases).

A possible explanation for the excess of segregating sites detected by Tajima's test is that slightly deleterious mutations may persist at low frequencies in the populations. As Tajima (1989) explains, the number of segregating sites observed in a sample of genes is affected by the existence of deleterious mutants, but is independent of their frequency. However, the presence or absence of these mutants does not have a large effect on the average pairwise distance between sequences, because they occur at low frequencies. Thus, their existence increases the proportion of segregating sites relative to that expected under the neutral model.

Departure from neutrality has also been shown for the mitochondrial genome of other organisms, including *Drosophila* (Ballard & Kreitman, 1994; Rand *et al.*, 1994), rodents (Nachman *et al.*, 1994), human and chimpanzees (Nachman *et al.*, 1996). For these particular cases, conclusions were based on differences in the proportion of replacement

to silent changes within species relative to between species, in different coding genes. But in *D. petechia*, even if the control region I is non-coding, selection at other mitochondrial linked genes could result in nonneutral evolution of that region (Ballard & Kreitman, 1995). However, the outcome of the neutrality test could also be an artifact of the high substitution rate of the CR-I. If new neutral mutations arise frequently in populations, this would be expected to result in a great number of haplotypes occurring at low frequencies and differing at a few nucleotide sites from each other, which is to some extent similar to the pattern caused by slightly deleterious mutants. This could be what happens in *D. petechia*, as shown by the high values of haplotype diversity within populations, and by the large number of private haplotypes. Therefore, Tajima's test may not be much informative about patterns of evolution when applied to a hypervariable region such as the control region I, if it fails to distinguish between the effect of a high substitution rate and a true departure from neutrality. Finally, the negative *D* values observed for both eastern and western subsamples (and also for the whole data set) could be due to further population subdivision within these two regions. This is because Tajima's test assumes that all genes are drawn from a single population. When this is not the case, the existence of population structuring will result in a higher proportion of segregating sites in the sample, relative to pairwise genetic distances, than that expected under the neutral model.

1.3. Homologous sequences

The relationships between the homologues and the CR-I from the same individuals (Figure 11) suggest that the homologous region originated from a paralogous duplication event, that is, prior to the coalescence of the CR-I lineages of the two individuals compared. The distance between the homologues and the CR-I is about twice the maximal one observed between a Yellow Warbler and an outgroup species CR-I sequences. Consequently, it appears plausible that the duplication event occurred before the emergence of *D. petechia*. It is not possible to tell if the homologous region is located on the nuclear or the mitochondrial genome. Nuclear copies of mitochondrial genes, including the control region, have been found in many species (for a review see Zhang & Hewitt, 1996). Heteroplasmy has also been documented in a wide variety of organisms (Hartl & Clark, 1989; Brown *et al.*, 1996) including at least three bird species (Avise & Zink, 1988; Ball & Avise, 1992). However, Klein and Brown (1994) found no evidence for it from their survey of mtDNA restriction site variation in *D. petechia*. Groth and Barrowclough (1996) recently uncovered in the Barred Owl (*Strix varia*) the existence of two control regions on the same mtDNA molecule. The two sequences reported here may represent such a duplication. Therefore, neither a nuclear nor mitochondrial origin can be excluded without further investigation. A method that could be used to verify if the homologue is located on the mitochondrial genome would consist of conducting a set of PCR amplifications from purified mtDNA at a lower annealing temperature (i.e., 50°C), in order to see if the multiple bands on sequencing gels are still present. If this is the case,

then this would indicate that two distinct products are still being amplified, even if mtDNA is the only template in the PCR reactions. Another methodology involves the use of a radioactively labeled CR-I homologue probe in a southern blotting of DNA samples extracted from diverse tissues, containing different proportions of mtDNA relative to nuclear DNA (Quinn & White, 1989). While keeping the DNA concentration constant in all blottings, a stronger band in mtDNA-rich tissues than in other tissues would support a mitochondrial origin of the homologous fragment. A similar band intensity across tissues however, would indicate that the probe anneals to a nuclear region. Finally, a more direct but more time-consuming and expensive procedure consists of sequencing the whole mitochondrial genome, as done by Groth and Barrowclough (1996), to see if any duplicated control region segment exists.

2. Population subdivision in *Dendroica petechia*

The major East-West subdivision of Yellow Warbler populations is the first significant pattern of this type to be documented for intraspecific variation within a North American bird. This pattern implies very low levels of gene flow between the eastern and western localities sampled. Pairwise estimates of the number of migrants per generation (*Nm*; not shown) were obtained by using the method of Hudson *et al.* (1992), which is based on the mean number of sequence differences within and between populations. The *Nm* value averaged 0.2 between eastern and western populations, which is an order of

magnitude smaller than what has been found in other avian species from allozyme data (Barrowclough, 1983).

Within the eastern region, the hierarchical analysis of nucleotide diversity provides two interesting results. First, the close relationship between Québec and Ontario indicates a significant exchange of migrants and a possible panmixia of populations in that area. This is not surprising given their relative geographic proximity and the lack of a major barrier between them. On the other hand, it appears unlikely that Newfoundland and Manitoba populations are more related, as suggested by the analysis, given that the species is continuously distributed in the provinces between these two locations. A likely explanation for this pattern is that a convergence of haplotypes from these populations has occurred through homoplastic substitutions, and that this effect is amplified because they occur along shallow lineages. This explanation seems reasonable given the frequency of multiple hit events as estimated from the data. It is also possible that, overall, genetic relationships between eastern populations are partly due to historical associations, and that they result from mtDNA lineage sorting (Strand *et al.*, 1996) during an expansion of populations from a restricted geographic area.

Isolation-by-distance could also be an important factor affecting gene flow between populations of a continuously distributed species such as the Yellow Warbler. Such a process should result in a geographic cline in the distribution of haplotypes (alleles), a situation not observed in the present study. But this would be correct only if migration was the primary factor governing haplotype frequencies, as has been argued to

be the case for allozyme loci. For variation in the CR-I, the high substitution rate seems an additional important factor causing the frequent appearance of new haplotypes in the populations. This likely explains the large proportion of private haplotypes observed in single population samples. If gene flow is insufficient to compensate for the regular appearance of haplotypes, and to homogenize their frequencies in neighbor populations, the chance of sampling the same haplotype in different locations can be greatly reduced (Edwards, 1993a). Therefore, it is possible that the observed structuring in the East reflects, to some extent, an isolation-by-distance process, but that the genetic differences between populations are also magnified by the high substitution rate of the CR-I.

Thus, rapidly evolving regions such as the CR-I may reveal a stronger population subdivision than that uncovered from other "slower" markers such as mtDNA RFLPs, and hence, make structuring easier to detect, provided that homoplastic substitutions do not overly obscure the relevant molecular information. They may be especially useful in detecting structure when natural populations have been isolated recently because, given the short time that has elapsed, such isolation may not be detectable using other markers. However, one has to be careful in interpreting intraspecific patterns inferred from a single gene, as discrepancy between population and gene histories may exist (Avise, 1994).

In the West, Alaska was presumably recolonized later than British Colombia, after the Wisconsin glaciation, owing to its higher latitude. The low haplotypic diversity in that population suggests that a small number of individuals originally reached that latitude during the colonization process. However, this observation could also constitute

an artifact of the small number of individuals sampled in Alaska relative to British Columbia, and additional sampling is required to resolve this pattern.

3. Phylogeographic structure

The continental split between eastern and western warbler populations detected by the analysis of population structure seems to correspond to a phylogeographic separation of mitochondrial lineages. East-West patterns have been uncovered in other species from mtDNA RFLPs studies conducted over more restricted geographic ranges, or involving limited sample sizes. In the Brown-headed Cowbird (*Molothrus ater*) a phylogeographic distinction exists between a subspecies inhabiting the West coast and one from the interior of the continent, although both are connected by gene flow (Fleisher *et al.*, 1991). In the Carolina Chickadee (*Parus carolinensis*) a local longitudinal split was identified in Alabama, despite the shallowness of the mtDNA tree (Gill *et al.*, 1993). But patterns in these two birds seem to occur at smaller scale than in the Yellow Warbler, although additional collection of warbler samples (e.g., in the USA) would allow to infer more accurately the latitudinal extent of the East-West subdivision in that species. However, Klein and Brown (1994) did find a pattern consistent with that uncovered in the present study from their survey of mtDNA restriction site variation in a small Yellow Warbler sample from USA. This suggests that the divergence between eastern and western populations extent further South. There is also evidence for a similar pattern in another warbler, the Common Yellowthroat (*Geothlypis trichas*), although in this species,

mtDNA variation was only assayed for a small number of birds (n=21; Ball & Avise, 1992).

The conclusion that the East-West structuring of Yellow Warbler populations also corresponds to a phylogeographic subdivision is undermined because of the low statistical support of the western clade in the neighbor-joining tree. There are several reasons to think that the pattern detected may still reflect a real branching event. First, Sitnikova *et al.* (1995) and Zharkikh and Li (1995) have shown that the bootstrap test can seriously underestimate the probability value (p-value) of nodes in a phylogeny when the number of taxa is large and the differences between them small, and that the branch-length test is closer to a real p-value. In the present case, this latter test provides a higher support, although still under the significance level, so that the monophyly of western haplotypes remains hypothetical. Second, the branching pattern defined by (eastern haplotypes, ((E20 E39), western haplotypes)) is observed in both distance and parsimony based phylogeny reconstructions. Third, the fact that all western haplotypes merged together during the pruning process is an indication that they carry a similar phylogenetic information compared to other haplotypes. In summary, the concordance between the neighbor-joining and maximum parsimony phylogenies with the population genetics analyses suggests that eastern and western birds have been isolated for a certain time and that the phylogeographic structure observed reflects that situation.

Different causes may be responsible for this pattern. These include the importance of the Rocky Mountains as a barrier to gene flow, the occurrence of a

migration event leading to the colonization of western North America from an eastern stock, and the fragmentation of warbler populations by a past vicariant event. I discuss the plausibility of these alternatives below.

If the Rocky Mountains play a role in isolating Yellow Warbler populations, it is probably minor. This is because these warblers breed throughout the Rockies at densities comparable to other regions of the continent, as shown by Breeding Bird Survey data (Sauer *et al.*, 1996). Therefore, the mountains probably have not had a significant effect on the dispersal of birds between the West coast and the rest of the continent.

The second explanation consists of the occurrence of a founder event in the West by a small number of eastern migrants. If the cladistic approach to measure gene flow is used (Slatkin & Maddison, 1989) a unique migration event is necessary to explain the origin of western haplotypes. This scenario may be oversimplistic, however, because it does not account for the fact that Yellow Warbler populations were not present on the range sampled during the last (Wisconsin) glaciation, except maybe for Newfoundland (Pielou, 1991). Since we do not know the distribution of the species prior to the present interglacial, it is not possible to assess whether this hypothesis is plausible, and difficult to discern it from the next one.

The third mechanism involves the fragmentation of a continentally distributed ancestral warbler stock due to a vicariant event, possibly a Pleistocene glacial advance, and subsequent allopatric differentiation within the isolates. Based on the branching pattern of the neighbor-joining tree, the number of "deeper" lineages is greater in the East

(a minimum of two lineages as ancient as the western one is present). Since large effective population sizes are more likely to retain a greater number of old lineages (Edwards, 1993a), the phylogenetic pattern observed is consistent with a larger historical population in the East. This suggests that, if the East-West break in *D. petechia* was caused by a vicariant event, the western clade was derived from a source stock which originally expanded from the East. Thus, the phylogeographic pattern inferred from the control region I sequences is consistent with Mengel's model of warbler differentiation, although it could originate from a different historical scenario producing a similar geographic structuring of mtDNA lineages (e.g., the founder hypothesis mentioned above). Consequently, it provides a first evidence for a pattern of intraspecific differentiation which has been postulated to lead to speciation events in North American warblers. This shows that intraspecific population studies using hypervariable mtDNA markers may prove to be useful in assessing evolutionary, in particular speciation, models.

A key issue to the Pleistocene glaciation hypothesis lies in the ability to date the coalescence of the western clade with its eastern sister cluster as occurring during the Pleistocene. Using sequence data to infer the time to a common ancestor of these lineages requires the assumption that a molecular clock exists. The closest organism for which a clock was calibrated for the control region I is the lesser Snow Goose (20.8% per million year; Quinn, 1992). But there is no reason to assume that the rate of evolution is similar in a species taxonomically as far from the Snow Goose as *D. petechia*. In addition, the

topology of the neighbor-joining tree (Fig. 8) clearly pictures a variation of that rate among lineages, which violates the clock assumption. However, independent calibrations of the rate of evolution of the mitochondrial genome as a whole provide a value around 2% sequence divergence per million year for a number of vertebrate taxa (Brown *et al.*, 1979; Wilson *et al.*, 1985). These estimates include three bird taxa: the lesser Snow Goose (Shields & Wilson, 1987), gallinaceous birds (Wilson *et al.*, 1985) and Hawaiian honeycreepers (Tarr & Fleischer, 1993). The maximal divergence observed by Klein and Brown (1994) between *D. petechia* haplotypes is 3.17%, and their sample includes both tropical and North American forms. Assuming that no major mtDNA lineage was omitted by their sampling procedure, this value sets an upper limit to the genetic distance between two individuals. This suggests that coalescence of the oldest lineages in *D. petechia* occurs near the beginning of the Pleistocene epoch (1.5-2 million years ago; Pielou, 1991) or later.

Regardless of what the cause that led to the isolation of eastern and western Yellow Warbler populations is, the presence of a western haplotype in Manitoba suggests that a secondary contact have occurred between the previously isolated mtDNA lineages, after the last post-glacial colonization of the actual range. However, more sampling is needed (e.g., in Saskatchewan and Alberta) to evaluate the geographic extent and relative importance of possible introgression in this region.

The absence of phylogeographic structuring in the eastern region (Manitoba to Newfoundland) is consistent with a recent expansion of populations from a more

restricted area (Zink, 1994). Therefore, this would support the hypothesis previously discussed that genetic relationships between populations reflect, at least partly, an historical association between them.

4. Microsatellite versus mitochondrial patterns in *D. petechia*

Dawson *et al.* (in preparation) surveyed the genetic variation at six microsatellite loci in a *D. petechia* sample including all individuals examined in the present study and some additional birds. A contingency table analysis revealed significant differences in allele frequencies between all seven populations. However, global F_{ST} values for each locus are small, similar to those obtained from allozyme data for other bird species (Barrowclough, 1983) and they translate into high levels of gene flow. These authors also found 1) that heterozygosity levels at microsatellite loci are higher in central provinces (Manitoba and Ontario) and interpreted this as the possible consequence of a secondary contact between previously isolated populations, 2) some evidence from a phylogenetic analysis for a distinction between eastern and western populations and 3) a positive relationship between geographic and genetic distances between localities, providing evidence for isolation-by-distance relationships between populations.

The patterns of genetic structuring inferred by Dawson *et al.* and in the present study are congruent to some extent. Microsatellite data suggest that eastern and western Yellow Warbler populations are experiencing a secondary contact after a certain period of isolation. However, evidence for previous isolation is not as clear as that observed for the

mtDNA data. The analysis of nucleotide diversity performed on CR-I data (Fig. 7) indicates the existence of two highly distinct clusters of populations, the first composed of ALA and BCO, and the second of MAN, ONT, QUE, NBW and NFL. But in a neighbor-joining reconstruction of relationships between locations, based on microsatellite data, the distinction between these two clusters is not strongly supported by the bootstrap test. The isolation-by-distance process detected from microsatellites may be consistent only with the mitochondrial pattern present in the eastern group of populations, although in the case of mtDNA data, this pattern does not correspond to a cline for the reasons previously discussed. Levels of gene flow (*Nm*) estimated from microsatellite markers range from 2.89 to 14.54 successful migrants per generation between pairs of populations, and are thus an order of magnitude greater than those calculated from control region I sequences.

The discrepancy between microsatellite and mtDNA data may result from different causes, including a difference between these two markers in the level of concordance between gene and population histories, the higher mutation rate of the mtDNA control region I, mutational aspects of both types of markers, or gender-biased dispersal by males. Regarding the East-West subdivision observed in the mtDNA phylogeny, the probability that a gene tree will track such a splitting event at a particular time will depend on for how long populations have been isolated. Retention of older lineages through random sorting will prevent, for a time, the emergence of a phylogeographic structure concordant with the population history. Reciprocal monophyly of populations will occur when the coalescence time of their respective haplotypes will be

shorter than the time elapsed since the separation (Moore, 1995). The more recent the coalescence in the past, the higher the chances of detecting monophyly. Similarly, shared nuclear alleles between populations will persist for a certain time after the separation. The expected coalescence time of alleles (haplotypes) is inversely proportional to the effective population size (N_e) of the gene (Moore, 1995). Autosomal loci such as microsatellites have a N_e four times greater than that of the maternally inherited mitochondrial genome (Slade *et al.*, 1995). So, regardless of the mutation rate of both types of markers, microsatellites are expected to take longer in developing a geographical distribution concordant with the history of populations. This may explain why a historical division of Yellow Warbler populations, if recent, is more clearly retraced from mtDNA data than from microsatellite data.

Using computer simulations, Hudson (1992) estimated the time necessary for mitochondrial versus nuclear markers to exhibit a phylogeny congruent with a three-species tree. His results show that, when the time since speciation events is long enough so that the probability p for a mtDNA marker to give the correct topology is equal to 0.95, p for nuclear-autosomal gene is only 0.62. Moreover, by the time $p = 0.95$ for the mtDNA tree, data from 16 nuclear loci are necessary to reach the same probability (Moore, 1995). In terms of populations, if the isolation of eastern and western population is recent enough, it is then possible that the mtDNA phylogeny can retrace that event (reciprocal monophyly of populations) while microsatellites do not yet.

At the moment, data on both control region sequences and microsatellite data are not available for other species. However, Slade *et al.* (1995) documented nucleotide polymorphism at five different single-copy nuclear DNA (scnDNA) sequences and in the mitochondrial CR-I existing between five pinniped species. Nuclear regions were selected for their high rate of evolution (three introns, one exon and one processed pseudogene). Mutation rates, estimated by plotting sequence divergence against divergence times provided by fossil records, were ~5-10 times higher for the CR-I than for scnDNA sequences. This suggests that the CR-I may represent the most sensitive marker available at present for tracing recent population histories.

Other mutational aspects should also be considered when comparing microsatellite markers to mtDNA sequence data. For example, under the microsatellite stepwise model of evolution, a mutation occurs by a gain or loss of a repeat, and all alleles are defined according to their length (Queller *et al.*, 1993). This increases the probability of homoplastic mutations and therefore, of observing alleles identical-by-state and not by descent (Jarne & Lagoda, 1996). This might result in a lower diversity of alleles in populations than for mtDNA haplotypes, and a higher level of sharing between them. My data suggest that homoplastic changes are also frequent in the control region I (Fig. 3). But the number of possible changes in a mtDNA sequence is still much higher, compared to that under the microsatellite stepwise model, as a mutation can theoretically arise at any of the nucleotide sites surveyed, and be of one of three types (one possibility of transitional change and two possibilities of transversional change). In the *D. petechia*

data set, the maximal number of alleles observed at a given microsatellite locus is 30 (for >200 birds) while the number of mtDNA haplotypes is 59 for 155 individuals. This means that the number of private types in populations is expected to be greater for mtDNA than microsatellites, and hence that a higher level of structuring might be observed with the former marker, based on haplotype (allele) frequencies. However, the appropriate mutational model for microsatellite loci (stepwise, infinite alleles, others) and the importance of homoplasies in these markers are still a matter of discussion (Jarne & Lagoda, 1996). Therefore, further research on these mutational aspects will be necessary to better understand their role in shaping the differences observed between the genetic patterns inferred from mitochondrial and microsatellite markers.

Finally, one cannot exclude the possibility that the discrepancy between mtDNA and microsatellites is due to a sex-biased dispersal pattern. If females are more philopatric to their natal or breeding sites than males, then the maternally inherited mitochondrial genome is expected to exhibit stronger geographic structuring than nuclear genes transmitted by both parents. But this would not explain the existence of a single and localized break between eastern and western Yellow Warbler populations. Moreover, band-recovery data indicate that males are in fact more philopatric to previous breeding sites than females in most birds (Greenwood & Harvey, 1982). Although natal (offspring) dispersal is more important than adult dispersal (Greenwood, 1989), it also seems generally greater for females than for males in passerine birds (Greenwood & Harvey, 1982).

5. MtDNA patterns and *D. petechia* subspecies

Intraspecific genetic patterns can serve to discuss the evolutionary significance of subspecies described on the base of phenotypic characters (Ball & Avise, 1992). I note that the mtDNA structuring observed in *D. petechia* does not match with the boundaries of subspecies recognized by the American Ornithologists' Union (1957) or by Browning (1994) in his recent review (Fig. 1). The genetic distinction between eastern and western warblers is consistent with the geographic localization of the *amnicola* and *morcomi* subspecies. However, the close mtDNA relationship between British Columbia and Alaska does not agree with the additional *banksi* division proposed by Browning to distinguish between the populations of these two regions. Moreover, the genetically closest (panmictic) locations, Québec and Ontario, are inhabited by different *D. petechia* subspecies (*amnicola* and *aestiva*, respectively), while all other eastern locations sampled belong to the same subspecies as Québec.

One drawback to the phylogenetic classification of warbler populations is that it is based on a single gene. Avise and Ball (1990) recommend that concordance among independently inherited traits should be a major guide in assigning a particular taxonomic status to groups of conspecific populations. In a few bird species (such as the Fox Sparrow; Zink, 1994), mtDNA patterns uncovered are congruent to different extent with morphological attributes. But, this does not seem to be the commonest situation. Ball and Avise (1992) surveyed mtDNA variation in six avian species in a sample including a total of 16 recognized subspecies. Among these subspecies, only two (from different species)

could be told from conspecific subspecies by their mtDNA haplotypes. For *D. petechia*, microsatellite data are even less concordant than mtDNA data with the subspecies divisions, as they indicate no clear break between eastern and western populations, while mtDNA data do.

A possible reason for the discrepancy between genetic and morphological patterns is that plumage variations in *D. petechia*, on which subspecies are defined, might be influenced by environmental factors and therefore not be completely genetically based (Ball & Avise, 1992). This could also result from differences in selective pressures exerted on the characters examined, and from the distinctive mode of inheritance of mitochondrial versus other characters. Although Yellow Warbler subspecies exhibit only very weak differences in body size (Browning, 1994), there is evidence that the geographic variation in that trait follows a cline correlated with that of several other passerine species (James, 1991).

Therefore, given that 1) little concordance exists between genetic and morphological attributes in *D. petechia*, and 2) recognized subspecies are defined on the base of slight, clinal variations in breeding plumage which do not necessarily reflect more the evolutionary history of warbler populations than DNA markers, I suggest that the present subspecific divisions in northern Yellow Warblers should be reassessed by the examination of other genetically inherited characters and/or more intensive sampling of data for currently documented attributes such as plumage. Based on my results, I would

argue that within North American breeding populations, there are two distinct evolutionary lineages that would deserve recognition at the subspecific level.

6. Conservation implications

Intraspecific genetic data can provide useful information for prioritizing conservation efforts within species. For example, Moritz (1994, 1995) has argued that phylogenetic analyses can serve to identify Evolutionary Significant Units (ESUs) within species, which should be preserved to maintain the maximal evolutionary potential of a species. ESUs are defined as historically isolated sets of populations that are detected by the study of intraspecific phylogenies. Moritz (1994) proposed that ESUs should exhibit both reciprocal monophyly for mtDNA haplotypes and also show significant differences in allele frequencies at nuclear loci. These conditions seem to be fulfilled by eastern and western *D. petechia* groups, except that a western haplotype is found in Manitoba. This minor exception indicates that additional sampling in the Prairies might reveal hybridization between these potential ESUs. The two groups can also be considered as Management Units (MUs), as their populations differ in haplotype frequencies (Moritz, 1994). MUs have an utility in short term management because they identify populations that are demographically independent of each other.

The Yellow Warbler is not an endangered species, and it is unlikely that specific conservation efforts will be directed toward it, at least in a near future. However, my results suggests that eastern and western Canada may form distinct Evolutionary

Significant Areas (Moritz, 1995; Avise, 1992) characterized by distinct vicariant histories. This point can be verified by future studies on other codistributed organisms and can represent a key information in designing area to be preserved, or in orienting conservation efforts directed towards threatened neotropical migratory birds.

Another important issue related to conservation is the identification of wintering grounds used by specific breeding populations. This is particularly critical to neotropical migrants like the Yellow Warbler, because their tropical forest wintering habitats are being seriously degraded by human activities. Therefore, preserving the species habitat requirements in the temperate zone only may be useless if the survival of populations is mostly compromised by the loss of wintering grounds. Recently, Wenink and Baker (1996) were able to assign to specific populations individual Dunlins caught at migratory stopovers and on winter territories, on the basis of their CR-I sequences. The breeding populations had been characterized in previous studies for their mtDNA lineage composition (Wenink *et al.*, 1993, 1996). A similar approach could be used to determine the eastern or western origin of Yellow Warblers caught in the tropics or during migration. Two fixed nucleotide differences exist between these groups (one transition and one transversion), and an additional three sites (two transitions and one transversion) distinguish western haplotypes from all eastern ones except E20 and E39. This array of five sites could serve as a specific subset of genetic "tags" to diagnose the breeding population origin of individual warblers. This would at least provide an indication on whether a bird is from a western or eastern origin, although filling the sampling gap

between British Columbia and Manitoba will be crucial to determine more accurately the boundary between the mtDNA lineages.

CONCLUSION

The present survey of control region I sequence variation has revealed a high level of polymorphism in *D. petechia*, greater than that previously reported from a mtDNA restriction site analysis in this species. The high mutation rate of the CR-I makes this marker useful for detecting population structuring, even when populations have been subdivided for a short time, as it might be the case for northern Yellow Warblers. This is shown by the much greater level of structuring inferred from CR-I sequences than from microsatellite data. A major division is identified both by population genetic and phylogenetic analyses between western and eastern Yellow Warbler populations, suggesting a long history of isolation between these groups. Some hypotheses that may be invoked to explain this result are subject to some reservations because they imply the monophyly of western birds, which is not statistically supported. The pattern observed is consistent either with a founder event in the West or a vicariant fragmentation of populations. It also provides intraspecific support for Mengel's (1964) model of North American warbler speciation, and therefore suggests a fruitful possibility for further genetic studies within other warbler species.

REFERENCES

American Ornithologists' Union. 1957. Check-list of North American birds. Fifth edition. A.O.U., Baltimore, Maryland.

American Ornithologists' Union. 1995. Fortieth supplement to the American Ornithologists' Union Check-list of North American Birds. *Auk* **112**: 819-830.

Avise, J. C. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos* **62**: 62-76.

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman & Hall, New York, NY.

Avise, J. C. 1996. Three fundamental contributions of molecular genetics to avian ecology and evolution. *Ibis* **138**: 16-25.

Avise, J. C. and R. M. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* **7**: 45-67.

Avise, J. C. and W. S. Nelson. 1989. Molecular genetic relationships of the extinct Dusky Seaside Sparrow. *Science* **243**: 646-648.

Avise, J. C. and R. M. Zink. 1988. Molecular genetic divergence between avian sibling species: King and Clapper Rails, Long-billed and Short-billed Dowitchers, Boat-tailed and Great-tailed Grackles, and Tufted and Black-crested Titmice. *Auk* **105**: 516-528.

Ball, R. M. and J. C. Avise. 1992. Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *Auk* **109**: 626-636.

Ball, R. M., S. Freeman, F. C. James, E. Bermingham and J. C. Avise. 1988. Phylogeographic population structure of Red-winged Blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences, USA* **85**: 1558-1562.

Ballard, J. O. and M. Kreitman. 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics* **138**: 757-772.

Ballard, J. O. and M. Kreitman. 1995. Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology & Evolution* **10**: 485-488.

Barrowclough, G. F. 1983. Biochemical studies of microevolutionary processes. Pp. 223-261 *in* A. H. Brush and G. A. Clark, Jr., editors. Perspectives in ornithology. Cambridge University Press, Cambridge.

Bermingham, E., S. Rohwer, S. Freeman and C. Wood. 1992. Vicariance biogeography in the Pleistocene and speciation in North American wood warblers: a test of Mengel's model. *Proceedings of the National Academy of Sciences, USA* **89**: 6624-6628.

Brown, J. R, K. Beckenbach, A. T. Beckenbach and M. J. Smith. 1996. Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of Sturgeon (*Acipenser*). *Genetics* **142**: 525-535.

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

Browning, M. R. 1994. A taxonomical review of *Dendroica petechia* (Yellow Warbler) (Aves: Parulinae). *Proceedings of the Biological Society of Washington* **107**: 27-51.

Brumfield, R. T. and A. P. Capparella. 1996. Historical diversification of birds in northwestern South America: a molecular perspective on the role of vicariant events. *Evolution* **50**: 1607-1624.

Cann, R. L., M. Stoneking and A. C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* **325**: 31-36.

Edwards, S. V. 1993a. Mitochondrial gene genealogy and gene flow among island and mainland populations of a sedentary songbird, the Grey-crowned Babbler (*Pomatostomus temporalis*). *Evolution* **47**: 1118-1137.

Edwards, S. V. 1993b. Long-distance gene flow in a cooperative breeder detected in genealogies of mitochondrial DNA sequences. *Proceedings of the Royal Society of London*, Series B **252**: 177-185.

Ellsworth, D. L., R. L. Honeycutt, N. J. Silvy, K. D. Rittenhouse and M. H. Smith. 1994. Mitochondrial-DNA and nuclear-gene differentiation in North American prairie grouse (genus *Tympanuchus*). *Auk* **111**: 661-671.

Excoffier, L., P. E. Smouse and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Felsenstein, J. 1993. Phylogeny inference package, version 3.5c. University of Washington, Seattle.

Fleischer, R. C., S. I. Rothstein and L. S. Miller. 1991. Mitochondrial DNA variation indicates gene flow across a zone of known secondary contact between two subspecies of the Brown-headed Cowbird. *Condor* **93**: 185-189.

Friesen, V. L., W. A. Montevecchi, A. J. Gaston, R. T. Barrett and W. S. Davidson. 1996. Molecular evidence for kin groups in the absence of large-scale genetic differentiation in a migratory bird. *Evolution* **50**: 924-930.

Fu, Y.-X. 1996. New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**: 557-570.

Gauthier, A. 1995. Paruline jaune. Pp 866-869 *in* Gauthier, J. and Y. Aubry, editors. Les oiseaux nicheurs du Québec: Atlas des oiseaux nicheurs du Québec méridional. Association Québécoise des Groupes d'Ornithologues, Société Québécoise de Protection des Oiseaux, Service Canadien de la Faune, Environnement Canada, Montréal.

Gill, F. B., A. M. Mostrom and A. L. Mack. 1993. Speciation in North American chickadees: 1. Patterns of mtDNA genetic divergence. *Evolution* **47**: 195-212.

Greenlaw, J. S. 1993. Behavioral and morphological diversification in Sharp-tailed Sparrows (*Ammodramus caudacutus*) of the Atlantic coast. *Auk* **110**: 286-303.

Greenwood, P. J. 1989. Inbreeding, phylopatry and optimal outbreeding in birds. Pp. 207-222 *in* F. Cooke and P. A. Buckley, editors. Avian genetics. Academic Press, London.

Greenwood, P. J. and P. H. Harvey. 1982. The natal and breeding dispersal of birds. *Annual Review of Ecology and Systematics* **13**: 1-21.

Groth, J. and G. F. Barrowclough. 1996. The complete mitochondrial genome of an owl. Communication presented at the 114th stated meeting of the American Ornithologists' Union, Boise, Idaho.

Hartl, D. L. and A. G. Clark. 1989. Principles of population genetics. Sinauer Associates Inc., Sunderland.

Higgins, D. G. and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Computer Applied Biosciences* **5**: 151-153.

Hoelzel, A. R. and A. Green. 1992. Analysis of population-level variation by sequencing PCR-amplified DNA. Pp. 159-187 *in* A. R. Hoelzel, editor. Molecular genetic analysis of populations: a practical approach. Oxford University Press, Oxford.

Holsinger, K. E. and R. J. Mason-Gamer. 1996. Hierarchical analysis of nucleotide diversity in geographically structured populations. *Genetics* **142**: 629-639.

Hubbard, J. P. 1973. Avian evolution in the aridlands of North America. *Living Bird* **12**: 155-196.

Hudson, R. R. 1992. Gene trees, species trees, and the segregation of ancestral alleles. *Genetics* **131**: 509-512.

Hudson, R. R., M. Slatkin and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583-589.

James, F. C. 1991. Complementary descriptive and experimental studies of clinal variation in birds. *American Zoologist* **31**: 694-706.

Jarne, P. and P. J. L. Lagoda. 1996. Microsatellites from molecules to populations and back. *Trends in Ecology and Evolution* **11**: 424-429.

Klein, N. K. and W. M. Brown. 1994. Intraspecific molecular phylogeny in the Yellow Warbler (*Dendroica petechia*), and implications for avian biogeography in the West Indies. *Evolution* **48**: 1914-1932.

Kumar, S, Tamura, K. and M. Nei. 1993. Molecular Evolutionary Genetics Analysis. Pennsylvania State University.

Lansman, R. A., J. C. Avise, C. F. Aquadro, J. F. Shapira and S. W. Daniel. 1983. Extensive genetic variation in mitochondrial DNA's among geographic populations of the deer mouse, *Peromyscus maniculatus*. *Evolution* **37**: 1-16.

Lougheed, S. C., P. Handford and A. J. Baker. 1993. Mitochondrial DNA hyperdiversity and vocal dialects in a subspecies transition of the Rufous-collared Sparrow. *Condor* **95**: 889-895.

Mengel, R. M. 1964. The probable history of species formation in some northern wood warblers (Parulidae). *Living Bird* **3**: 9-43.

Miller, P. C. 1993. Molecular genetic analysis of a Brown-headed Cowbird (*Molothrus ater*) population. M. Sc. Thesis, McMaster University.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**: 718-726.

Moritz, C. 1994. Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution* **9**: 373-375.

Moritz, C. 1995. Uses of molecular phylogenies for conservation. *Philosophical Transactions of the Royal Society of London*, Series B **349**: 113-118.

Nachman, M. W., S. N. Boyer and C. F. Aquadro. 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. *Proceedings of the National Academy of Sciences, USA* **91**: 6364-6368.

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro. 1996. Nonneutral mitochondrial DNA variation in Humans and Chimpanzees. *Genetics* **142**: 953-963.

Nee, S., E. C. Holmes, A. Rambaut and P. H. Harvey. 1995. Inferring population history from molecular phylogenies. *Philosophical Transactions of the Royal Society of London*, Series B **349**: 25-31.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Norman, J. A., C. Moritz and C. J. Limpus. 1994. Mitochondrial DNA control region polymorphisms: genetic markers for ecological studies of marine turtles. *Molecular Ecology* **3**: 363-373.

Pielou, E. C. 1991. After the ice age: the return of life to glaciated North America. University of Chicago Press, Chicago.

Queller, D. C., J. E. Strassmann and C. R. Hughes. 1993. Microsatellite and kindship. *Trends in Ecology and Evolution* **8**: 285-288.

Quinn, T. W. 1992. The legacy of Mother Goose - phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Molecular Ecology* **1**: 105-117.

Quinn, T. W. and B. N. White. 1989. Analysis of DNA sequence variation. Pp. 163-198 *in* F. Cooke and P. A. Buckley, editors. Avian genetics. Academic Press, London.

Rand, D. M., M. Dorfsman and L. M. Kann. 1994. Neutral and nonneutral evolution of *Drosophila* mitochondrial DNA. *Genetics* **138**: 741-756.

Rice, W. R. 1989. Analysing tables of statistical tests. *Evolution* **43**: 223-225.

Rising, J. D. and J. C. Avise. 1993. Application of genealogical-concordance principles to the taxonomy and evolutionary history of the Sharp-tailed Sparrow (*Ammodramus caudacutus*). *Auk* **110**: 844-856.

Rzhetsky, A. and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. *Molecular Biology and Evolution* **9**: 945-967.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.

Sambrook, J., E. F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual, second edition. Cold Spring Harbor Laboratory Press.

Sauer, J. R., B. G. Peterjohn, S. Schwartz and J. E. Hines. 1996. The North American Breeding Bird Survey Home Page, version 95.1. Patuxent Wildlife Research Center, Laurel, MD.

Selander, R. K. 1965. Avian speciation in the Quaternary. Pp. 527-542 *in* Wright, H. E., Jr. and D. G. Frey, editors. Quaternary of the United States. Princeton University Press, Princeton.

Seutin, G., B. N. White and P. T. Boag. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69**: 82-90.

Shields, G. F. and A. C. Wilson. 1987. Calibration of mitochondrial DNA evolution in geese. *Journal of Molecular Evolution* **24**: 212-217.

Sitnikova, T., A. Rzhetsky and M. Nei. 1995. Interior-branch and bootstrap tests of phylogenetic trees. *Molecular Biology and Evolution* **12**: 319-333.

Slade, R. W., C. Moritz and A. Heideman. 1995. Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution* **11**: 341-356.

Slatkin, M. 1996. Gene genealogies within mutant allelic classes. *Genetics* **143**: 579-587.

Slatkin, M. and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**: 603-613.

Strand, A. E., B. G. Milligan and C. M. Pruitt. 1996. Are populations islands? Analysis of chloroplast DNA variation in Aquilegia. *Evolution* **50**: 1822-1829.

Swofford, D. L. 1991. Phylogenetic Analysis Using Parsimony version 3.0. Illinois Natural History Survey.

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

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution* **9**: 678-687.

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.

Tarr, C. 1995. Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. *Molecular Ecology* **4**: 527-529.

Tarr, C. L. and R. C. Fleisher. 1993. Mitochondrial-DNA variation and evolutionary relationships in the amakihi complex. *Auk* **110**: 825-831.

Tartof, K. D. and C. A. Hobbs. 1987. Improved media for growing plasmid and cosmid clones. *Focus* **9**: 12.

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

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**: 1503-1507.

Wakely, J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology and Evolution* **11**: 158-163.

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

Wenink, P. W., Baker, A. J. and M. G. J. 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, USA* **90**: 94-98.

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

Wills, C. 1996. Topiary pruning and weighting reinforce an african origin for the human mitochondrial DNA tree. *Evolution* **50**: 977-989.

Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowsky, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biological Journal of the Linnean Society **26**: 375-400.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution* **11**: 367-372.

Zhang, D.-X. and G. M. Hewitt. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* **11**: 247-251.

Zharkikh, A. and W.-H. Li. 1995. Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Molecular Phylogenetics and Evolution* **4**: 44-63.

Zink, R. M. 1994. The geography of mitochondrial DNA variation, population structure, hybridization, and species limits in the Fox Sparrow (*Passerella iliaca*). *Evolution* **48**: 96-111.

Zink, R. M. 1996. Comparative phylogeography in North American birds. *Evolution* **50**: 306-317.

Zink, R. M. and D. L. Dittman. 1993a. Gene flow, refugia, and evolution of geographic variation in the Song Sparrow (*Melospiza melodia*). *Evolution* **47**: 717-729.

Zink, R. M. and D. L. Dittman. 1993b. Population structure and gene flow in the Chipping Sparrow and a hypothesis for evolution in the genus *Spizella*. *Wilson Bulletin* 105: 399-413.

Zink, R. M., W. L. Rootes and D. L. Dittman. 1991. Mitochondrial DNA variation, population structure, and evolution of the Commom Grackle (*Quiscalus quiscula*). *Condor* **93**: 1016-1019.

Table 1. Sample size, number of haplotypes and diversity index for each locality sampled.

Locality	Sample size	Total number of haplotypes	Number of private haplotypes	Diversity index (h)
Fairbanks, Alaska	11	2	0 (0%)	0.44
Revelstoke, British Columbia	26	13	11 (85%)	0.91
Delta Marsh, Manitoba	27	19	17 (89%)	0.96
Chaffeys Locks, Ontario	21	13	5 (38%)	0.92
Trois-Rivières, Québec	28	15	7 (47%)	0.87
Germantown, New Brunswick	11	7	2 (29%)	0.90
Gros Morne Nat. Park, Newfoundland	31	11	5 (45%)	0.86

Table 2. Variability in three segments of the control region in Yellow Warblers.

Segment	Number of sites surveyed	Sample size	Polymorphic sites	Haplotypes
control region I	344	155	32	59
central domain	247	23	1	2
control region II	133	35	1	2

Table 3. Genetic diversity partitioned among all localities sampled and probability[a] of observing higher values by chance. Diversity indices as described in the text.

Diversity index	Value	Probability	Reference
G_{st}	0.13	0.002	Nei, 1987
$\hat{g}_{st}$	0.52	<0.001	Holsinger & Mason-Gamer, 1996
Φ_{st}	0.53	<0.001	Excoffier *et al.*, 1992

[a]Probability values obtained from 1000 replicates of random redistribution of haplotypes in populations.

Table 4. Pairwise Φ_{st} values between locations (below diagonal) and probability[a] of observing higher values by chance (above diagonal). Location abbreviations as defined in the text.

Location	NFL	NBW	QUE	ONT	MAN	BCO	ALA
NFL	—	0.0053*	0.0016*	0.0036*	0.0057*	0.0000*	0.0000*
NBW	0.1547	—	0.0000*	0.0000*	0.0037*	0.0000*	0.0000*
QUE	0.1206	0.3283	—	0.6195	0.0039*	0.0000*	0.0000*
ONT	0.1236	0.3188	-0.0128	—	0.0034*	0.0000*	0.0000*
MAN	0.0823	0.1759	0.0935	0.0705	—	0.0000*	0.0000*
BCO	0.7267	0.6286	0.7760	0.7827	0.6790	—	0.0605
ALA	0.7550	0.6589	0.8058	0.8205	0.6827	0.0583	—

[a]Probabilities obtained by comparing Φ_{st} values to null distributions based on 10 000 replicates of random resampling of haplotypes among locations compared.

*Indicates comparisons significant at the 5% level. Application of the sequential Bonferroni method indicates that four of these significant comparisons should not be considered as such: NFL-MAN, NFL-NBW, QUE-MAN, MAN-NBW.

Table 5. Nested analysis of molecular variance on two *D. petechia* groups. The first group is composed of Alaska and British Columbia populations and the second group of Manitoba, Ontario, Québec, New Brunswick and Newfoundland populations. The percentage of the total genetic variance that is partitioned at different hierarchical levels of population structuring, as well as the corresponding Φ-statistics, are shown. The probability[a] of observing higher values by chance is also indicated.

Variance component	Percentage of the total variance	Corresponding Φ-statistic[b]	Probability
among groups	66.6	$\Phi_{ct} = 0.67$	<0.0001
among populations/within groups	4.4	$\Phi_{sc} = 0.13$	<0.0001
within populations	29.0	$\Phi_{st} = 0.71$	<0.0001

[a]Probability values obtained by comparing the observed values to null distributions based on 10 000 replicates of random resampling of haplotypes.

[b]Φ_{ct} = Var(among groups)/Var(total),

Φ_{sc} = Var(among populations)/(Var(among pop.)+Var(within pop.)),

Φ_{st} = (Var(among groups)+Var(among pop.))/Var(total); from Excoffier *et al.*, 1992.

Figure 1. Sampling sites and distribution of Yellow Warbler subspecies in Canada and Alaska. Subspecies as recognized by the American Ornithologists' Union (1957), except for *banksi* which represents an additional category proposed by Browning (1994). Dashed lines indicate approximate boundaries between subspecies ranges (after Browning, 1994) and the solid line shows the northern limit of the species breeding distribution. Sampling sites are: 1) Fairbanks, Alaska; 2) Revelstoke, British Columbia; 3) Delta Marsh, Manitoba; 4) Chaffeys Locks, Ontario; 5) Trois-Rivières, Québec; 6) Germantown, New Brunswick; 7) Gros Morne National Park, Newfoundland. Sample sizes given in Table 1.

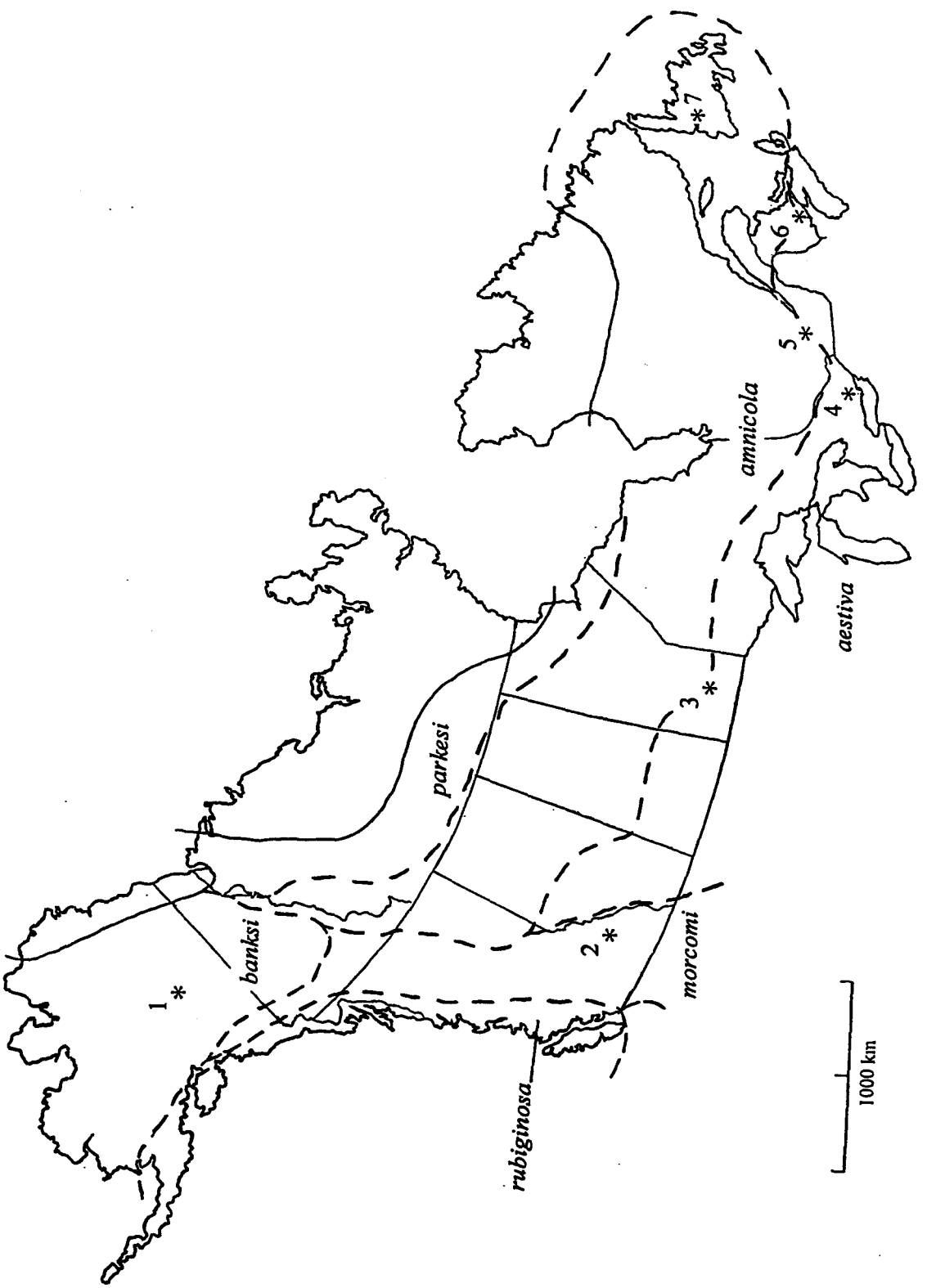
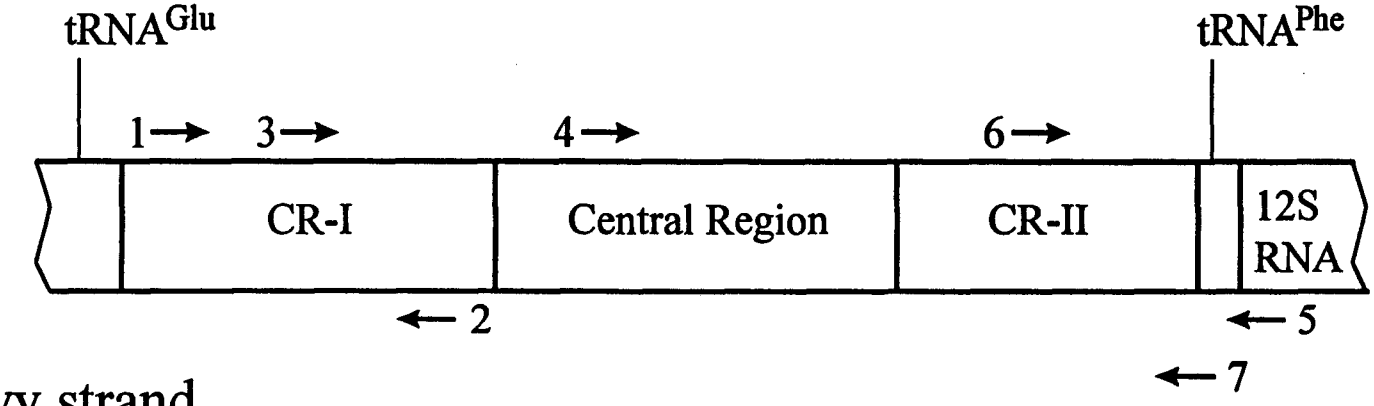

1*
banksi
parkesi
rubiginosa
2*
morcomi
3*
amnicola
aestiva
4*
5*
6*
7*
1000 km

Figure 2. Primers (1-7) used to amplify and sequence the control region. Names and sequences of each primer as follows: 1 (LDPdl5): 5'-TTCTTGCTTTAAGGGTATGT-3'; 2 (HDPdl4): 5'-TCAATAGATAACCATGTCCT-3'; 3 (LDPdl6): 5'-CATTATCTCCA-AAACACGCT-3'; 4 (GSL-GLU): 5'-TTGGTTGTAACTTCAGGAAC-3'; 5 (GSH-12S): 5'-CGGAATTCAAGGTTAGGACTAAGTC-3'; 6 (LDPdl2): 5'-AAACAATTAA-ACCACATATCC-3'; 7 (HDPdl1): 5'-CGGCACGCATCTTGACATCT-3'.

Light strand



Heavy strand

100 bp

Figure 3. Number of nucleotide sites characterized by different numbers of substitutions (hits) in the 344bp fragment of the *D. petechia* control region I. Number of hits per site estimated from a neighbor-joining phylogeny, as described in the text.

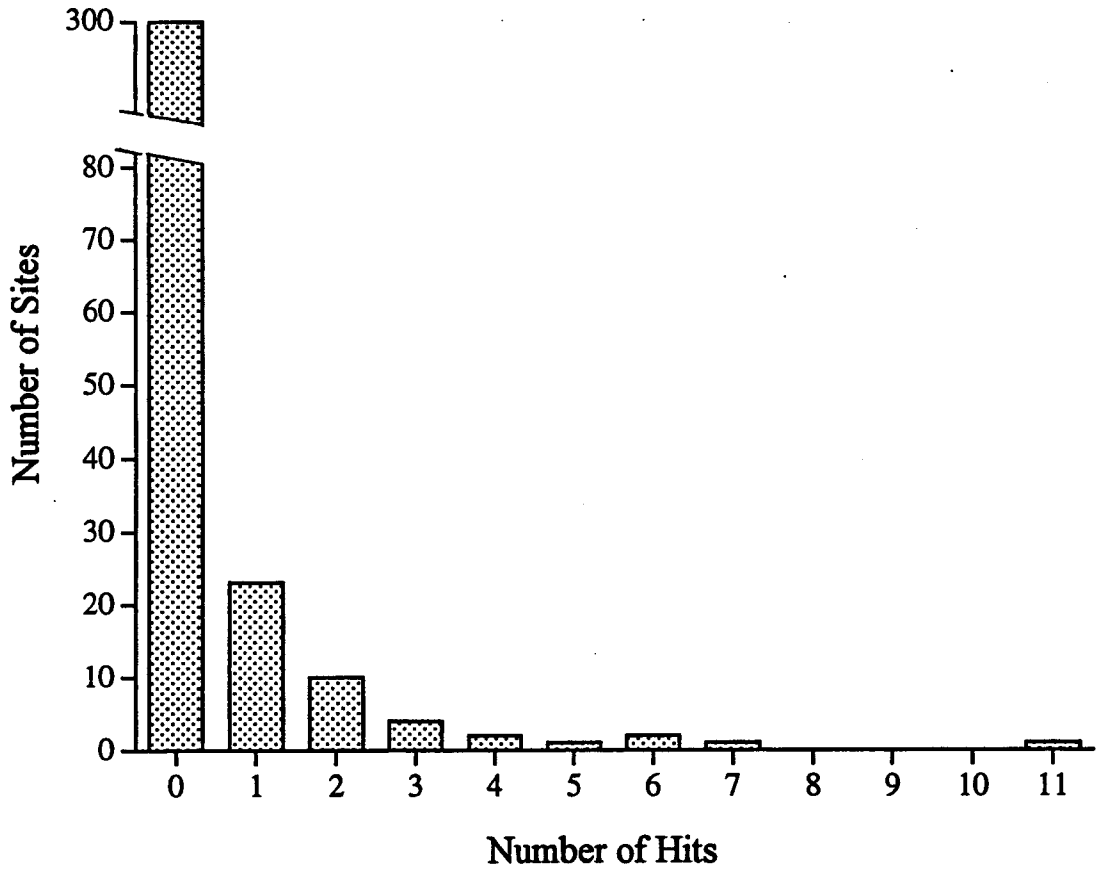

300
80
70
60
50
40
30
20
10
0
Number of Sites
0
1
2
3
4
5
6
7
8
9
10
11
Number of Hits

Figure 4. Control region I sequence for two *D. petechia* haplotypes. Dots in the W1 sequence represent identical bases to haplotype E1. Numbers indicate nucleotide positions relative to the first one in the segment sequenced.


1
E1 AATATGCATCACACTCTTTGCCCCATCAGACAGACTATGTAATGTAGGATAATCCAAGGT -60
W1 .....T.....
ATATGTAATGCTTCTCCATCATGAACCCAAACATTATCTCCAAAACACGCTCTAATTCGG -120
.....
CCAGTACGCTCCTAGGCACATCCTTGCTTCAGGTACCATATAGCCCAAATGCTCCTACCT -180
...AC..T...T.....
ACAGCCAAGCAGCAAGCGTTACCCAAAGACCCAGTAACTTATCTACTACGCTAAACGTCC -240
.....C.....T..C.....
ACCAAGCGAACGAGGAATGTCCCAGTACACCTTTGAATTCCCCTAGTCTACTGAATTCGC -300
..T..C.A.....
CCACCTCCTAGGTAATATTCTCAGCCAACAGCCTTCAAGAACTC -344
.....

Figure 5. Variable sites of the control region I in 155 Yellow Warblers. Haplotype names are shown on the left and nucleotide positions relative to the beginning of the sequence are indicated by digits at the top. Dots represent identical bases to the first haplotype. The five last haplotypes are those from the outgroups: MW1: Mangrove Warbler from Costa Rica; MW2: Mangrove Warbler from Venezuela; GW: Golden Warbler (Puerto Rico); Dest: *Dendroica striata*; Depe: *D. pensylvanica*.

```
                  1111111111111111222222222222222222222222233
      1111337777801112222333344680011122223333344444445668 22
      0134462349391674578123627860656826892567823456786141 34

E1    caacatttctggtttgtcgcctgccactatactcactacgtccaagcgacacag
E2    .........c.....................................................
E3    .........c.........................tc..........................
E4    ............................t.....................................
E5    ...t.c...c.........................tc..........................
E6    ..............................c..t...ac........................
E7    .....c...c.........................tc..........................
E8    ................................t.....................................
E9    .........c........a..............tc..a.........................
E10   ..................a.....................................................
E11   .....c...c........a..............tc..........................
E12   ..............c.....................................................
E13   .........c........a.....................................................
E14   ....g.....................................................................
E15   .........c........a.................tc...........g.....
E16   ................................t.....................................
E17   ...t.....................t.g......t.....................
E18   ................................tc..........................
E19   .........c........a..............tc..........................
E20   ...t...........ac.t..............tc..ac.....................
E22   ...t.............................tc..........................
E23   .........c.....................................c.....................
E24   .........c........a..c..t...g......t...c.....................
E25   .........c.....................tc..a.....................
E26   .........c..............t.g......t...c.....................
E27   .........c.....a..a............t...tc..a........g.....
E28   ........................t.....................................
E29   ...t.....c..............t.g......t...c.....................
E30   ..................a..............tc..........................
E31   ....g.........................g.....................
E32   ..................a..............tc.......c.....
E33   .........c..............t.g......t.....................
E34   ....g....c.....................................................
E35   ..................a.......t.....................
E36   ..................a.......t.g......t.....................
E37   ...t.....................................................
E38   ........................t.g......t.....................
E39   ...t........c..ac.t..............tc..ac.....................
E40   .........c........a.........g......tc..........................
E41   .........c.....................tc..a........g.....
E42   ...t.....c..............t.......tc..........................
E43   ..........................c.....................
E44   .........c..............t.......tc..........................
E45   ..........................g.....................
E46   ................................tc.............g...
E47   t.....................................................
W1    ...t...........ac.t.t........c.....tc.....t..c.a......
W2    ...t...........ac.t..............tc.....t..c.a......
W3    ...t...........ac.t.t............tc.....t..c.a..g...
W4    ...t.....c.....ac.t..............tc.....t..c.a......
W6    ...t...........ac.t.t........c.....tc.........c.a......
W7    ...t...........ac.t..............tc..a..t..c.a......
W8    ...t...........ac.t.t............tc.....t..c.a......
W9    ...t...........ac.t..............tc..a.....c.a......
W11   ...t...........ac.t..............tc.....t..c.a...t..
W12   ...t...........ac.t.t.....t........tc.....t..c.a......
W13   ...t...........ac.t.......t........tc.....t..c.a......
W14   ?..t...........ac.t..............tc.ta..t..c.a......
W15   ...t...........actt..c............tc.....t..c.a......

MW1   t.....cc..aa.c.a..a...at.....a.....tc..ac..tgat.......
MW2   t.....cct..a...ac.a...att....ac....t...act.tgat.......
GW    t.....cc..aa...a..a...att...ga.....tc..act.cgat.......
Dest  tggt..cct.aa...a..a..cattgtcga.....t...ac.t.c.......ta
Depe  t......ct..a...a..at.cat...cga....gt.c.ac.....a..t....
```

Figure 6. Nucleotide sequence of the homologue (Hom) and the control region I from a Québec (QUE) bird and a British Columbia (BCO) bird. Control region I haplotypes for these two individuals are E5 and W1, respectively. Dots represent identical bases to the first sequence, and dashes show gap inserted to improve alignments. Numbers indicate nucleotide positions relative to the first one in the control region I segment.

```
                1
                |
QUE  (Hom)  AATAT----CACACTCTTTGCCCCATCAGACAGCCCATGAAATGTAGGATAATCCACATT -60
BCO  (Hom)  .....TCAT....C.............................................
QUE  (E5)   .....GCAT....T...................A.....T................AGG.
BCO  (W1)   .....GCAT....T...................A.T...T................AGG.

            ACACGTAATGCTCCCCCACCAAAAACCCAAACATTATCTCGAAAACG-GATGGTATTTGG -120
            ........................C.....................T.-............
            .T.T........T.T......TG.................C.....AC.C.CTA...C..
            .T.T........T.T...T..TG.................C.....AC.C.CTA...C..

            CCAATAACCCACCAGACACATTCTTGTTTCAGGTACCATGCAGCCCAAATG-TCCTACCT -180
            ..................................................-.........
            ...G..CG.TC.T..G.....C....C............AT..........C.........
            ....C.CT.TCTT..G.....C....C............AT..........C.........

            ACAGCCGAGCCACAAGCGTCACCCAACGACCCAGGATCTTATCTACTATACGCAACACCC -240
            ......A.....................................................
            ......A...AG.......T......A.......T.A.............G.CA...GT..
            ......A...AG.......T......A.......C.A.............G.CA...GT..

            AACCTAGTGAACGAGGAATGTCCCAGTACACCTTCGAATTCCCCTAGTCTACAGGATTCG -300
            ...........................................................A.
            .-..A..C..............................T.................T.A.....
            .-.TA.CCA.............................T.................T.A.....

            CCCACCTCCTAGGTAATATTCTTAGCCAACAGCCTTCAAGCACTC -345
            .........................................
            ........................C.................A....
            ........................C.................A....
```

Figure 7. Genetic relationships between the seven locations sampled, as inferred from the hierarchical analysis of nucleotide diversity. The values of $\hat{g}_{st}$ indexes are shown and represent the proportion of the total genetic diversity, within the cluster delimited by a node, that is partitioned between sister locations. Numbers in brackets indicate the probability of observing higher values by chance. These probabilities were obtained from null distributions of $\hat{g}_{st}$ indexes, constructed in each case from 1000 replicates of random redistribution of haplotypes in the locations compared.

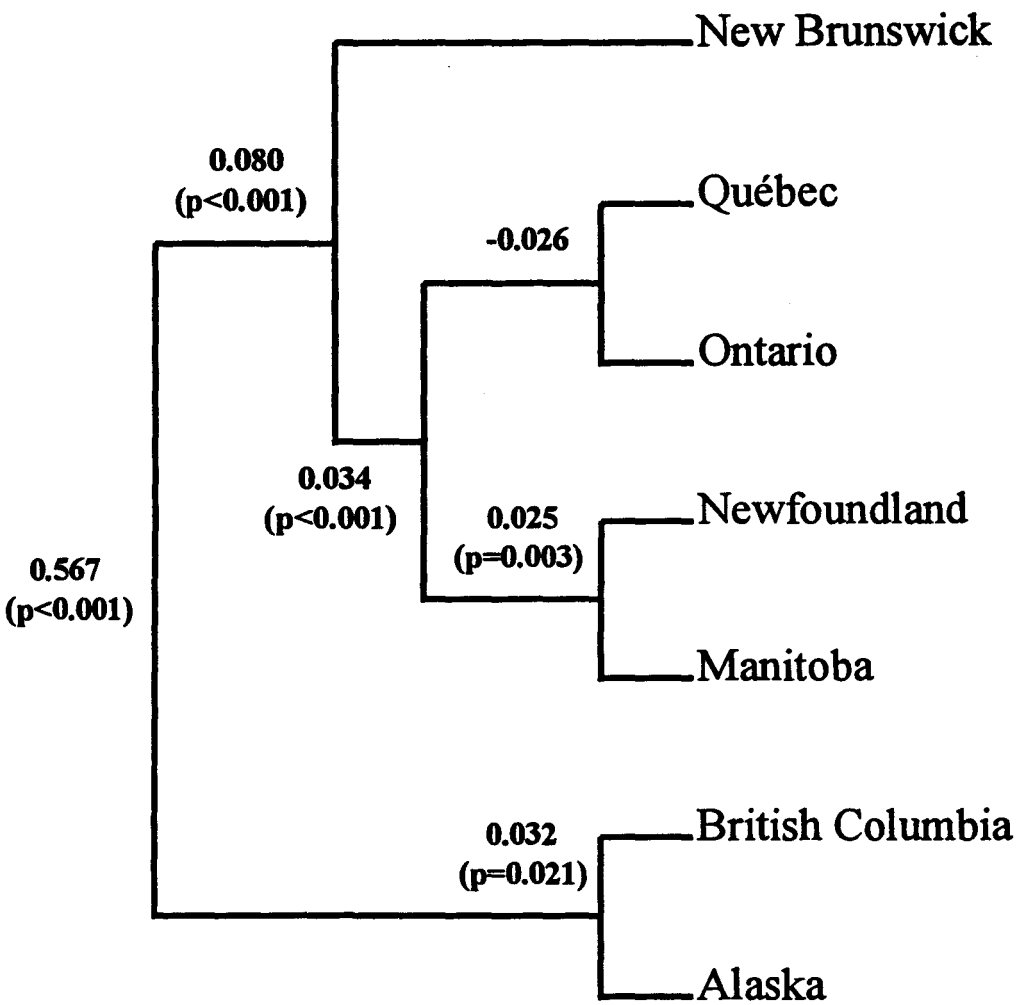

New Brunswick
0.080
(p<0.001)
Québec
-0.026
Ontario
0.034
(p<0.001)
0.025
(p=0.003)
Newfoundland
0.567
(p<0.001)
Manitoba
0.032
(p=0.021)
British Columbia
Alaska

Figure 8. Neighbor-joining reconstruction of phylogenetic relationships between all Yellow Warbler haplotypes. Haplotype names as in Figure 5. Numbers indicate the statistical support of relevant nodes as inferred from 1000 bootstraps replicates (values for most nodes not indicated owing to lack of space). Symbols following haplotype names show the populations were these haplotypes were found: ○, Alaska; •, British Columbia; ∇, Manitoba; *, Ontario; Δ, Québec; ◊, New Brunswick; ◆, Newfoundland.

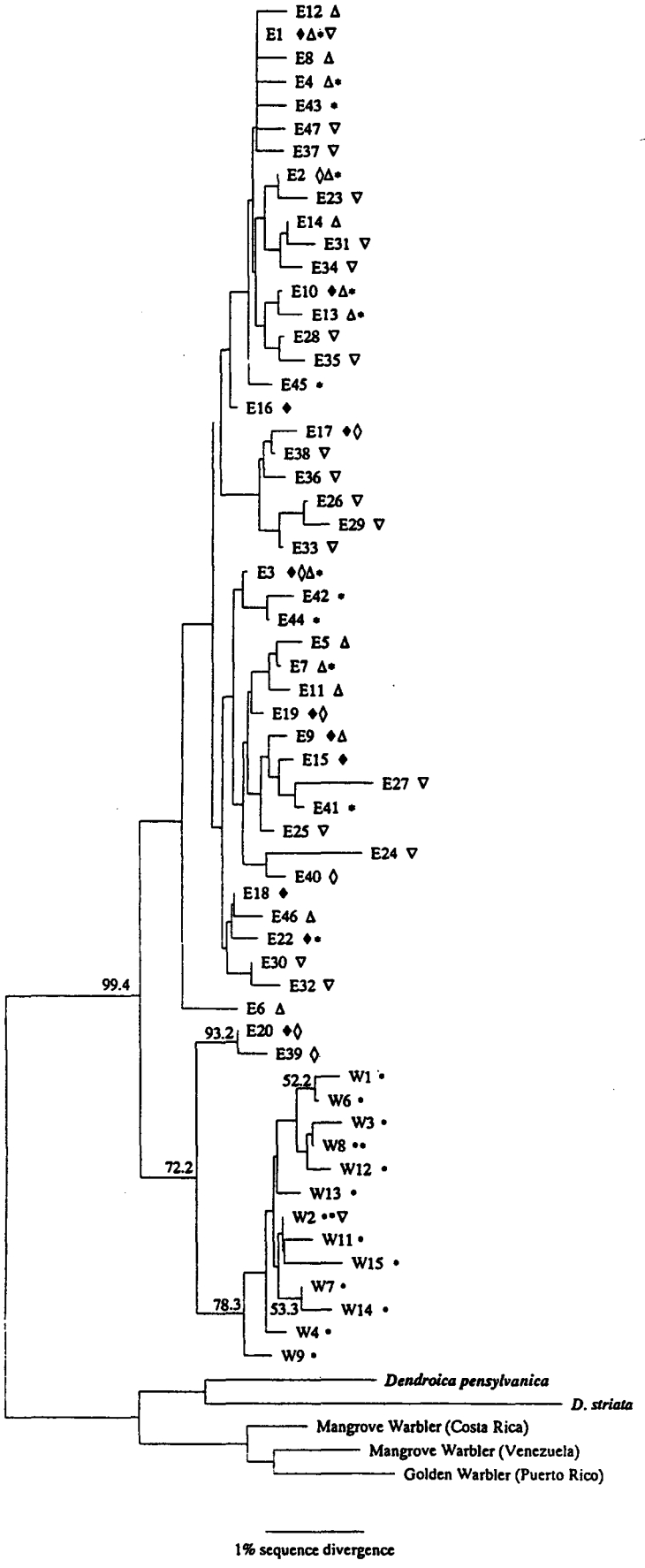

E12 Δ
E1 ♦Δ•∇
E8 Δ
E4 Δ•
E43 •
E47 ∇
E37 ∇
E2 ◊Δ•
E23 ∇
E14 Δ
E31 ∇
E34 ∇
E10 ♦Δ•
E13 Δ•
E28 ∇
E35 ∇
E45 •
E16 ♦
E17 ♦◊
E38 ∇
E36 ∇
E26 ∇
E29 ∇
E33 ∇
E3 ♦◊Δ•
E42 •
E44 •
E5 Δ
E7 Δ•
E11 Δ
E19 ♦◊
E9 ♦Δ
E15 ♦
E27 ∇
E41 •
E25 ∇
E24 ∇
E40 ◊
E18 ♦
E46 Δ
E22 ♦•
E30 ∇
E32 ∇
99.4
E6 Δ
93.2
E20 ♦◊
E39 ◊
52.2
W1 •
W6 •
W3 •
W8 ••
W12 •
72.2
W13 •
W2 ••∇
W11 •
W15 •
W7 •
78.3
53.3
W14 •
W4 •
W9 •
Dendroica pensylvanica
D. striata
Mangrove Warbler (Costa Rica)
Mangrove Warbler (Venezuela)
Golden Warbler (Puerto Rico)
1% sequence divergence

Figure 9. Number of phylogenetically informative sites per taxon for each level of the topiary pruning procedure. The number of taxa remaining in pruned data sets is: lev.1, all 64 taxa remain; lev.2, 25 taxa; lev.3, 13; lev.4, 11; lev.5, 11; lev.6, 7; lev.7, 7; lev.8, 6; lev.9, 5; lev.10, 4.

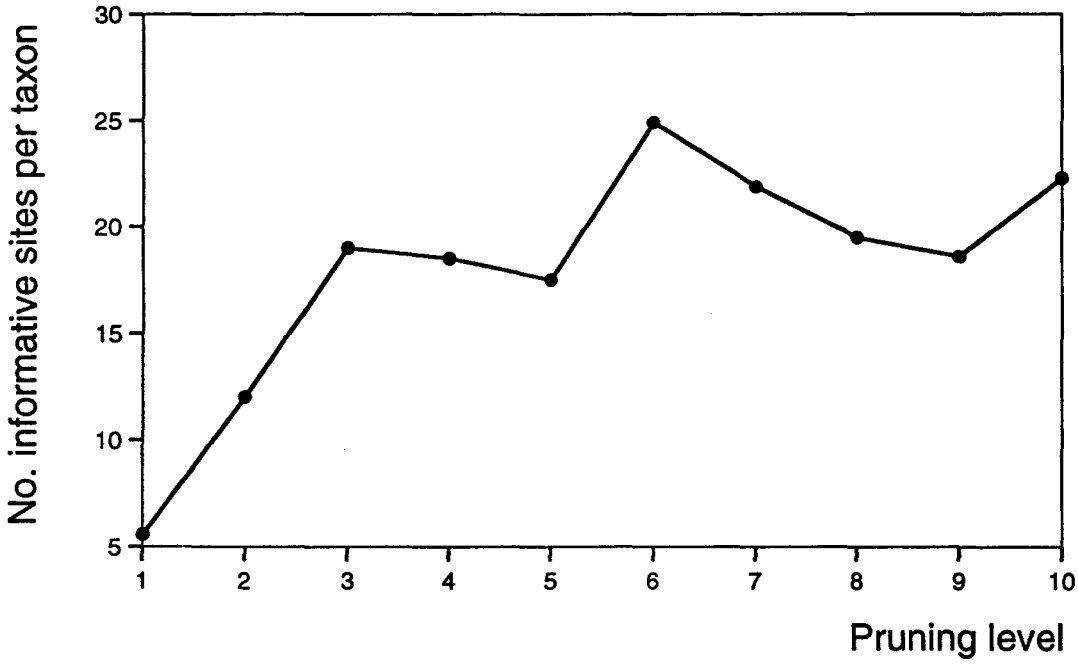

No. informative sites per taxon
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
10
Pruning level

Figure 10. Fifty percent majority rule consensus of five most parsimonious trees (69 steps; consistency index=1.0) built from pruned control region I sequences (level 3). Numbers indicate the statistical supports of nodes (when > 50%), as assessed from branch-and-bound searches done on 200 bootstrapped data sets. Taxa duplicated during the pruning procedure are as follows: E1: E2, E3, E4, E5, E7, E8, E9, E10, E11, E12, E13, E14, E15, E16, E18, E19, E22, E23, E25, E28, E30, E31, E32, E34, E35, E37, E38, E40, E42, E43, E44, E45, E46 and E47; E17: E26, E29, E33 and E36; E20: E39; W1: W2, W3, W4, W6, W7, W8, W9, W11, W12, W13, W14 and W15. Outgroup taxa are: MW1: Mangrove Warbler from Costa Rica; MW2: Mangrove Warbler from Venezuela; GW: Golden Warbler (Puerto Rico); DEST: *Dendroica striata*; DEPE: *D. pensylvanica.*

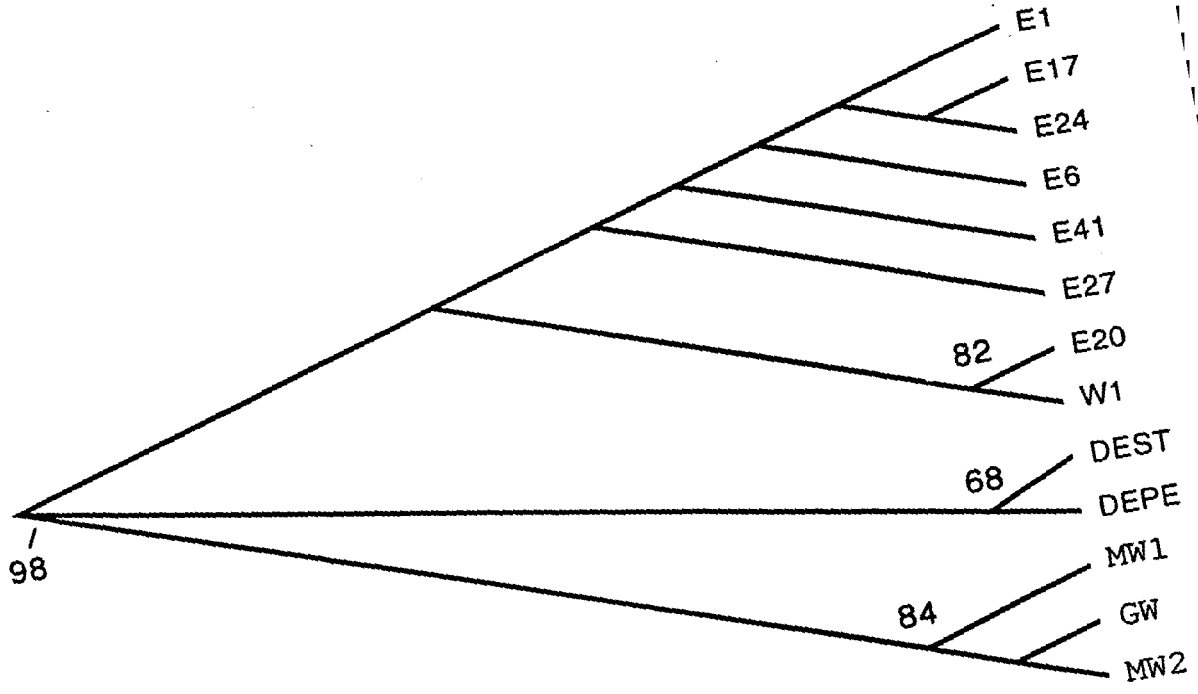
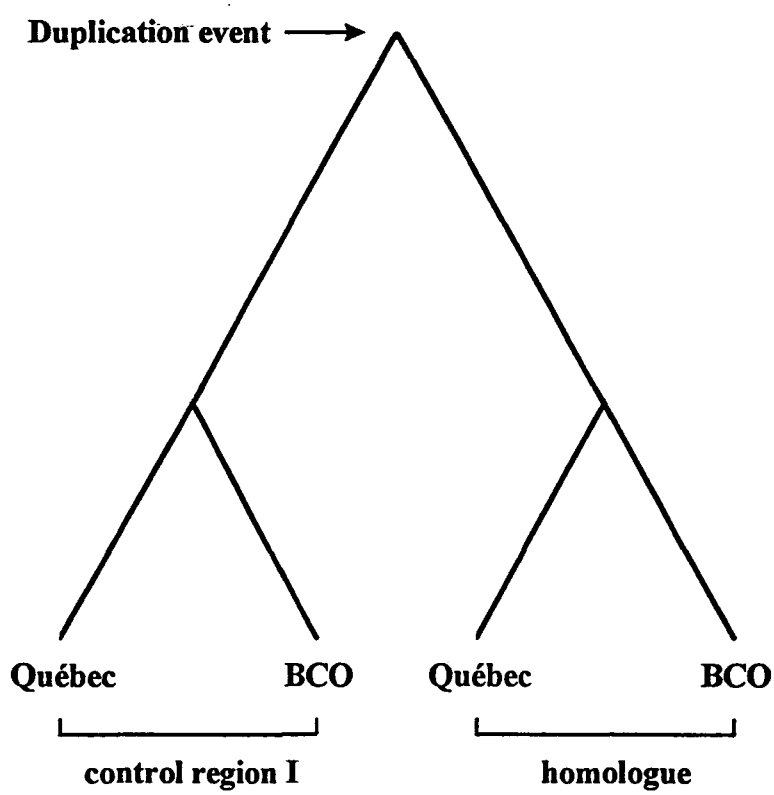

97
E1
E17
E24
E6
E41
E27
82
E20
W1
68
DEST
DEPE
98
84
MW1
GW
MW2

Figure 11. Schematic representation of the phylogenetic relationships between the two homologue sequences and the control region I from the same individuals. The coalescence point between the two regions illustrates the duplication event that has led to their divergence. Control region I haplotypes for the Québec and British Columbia (BCO) individuals are E5 and W1, respectively.


Duplication event
Québec
BCO
Québec
BCO
control region I
homologue

Appendix I. Haplotypes sampled in each locality[a].

Hap. | Loc.	NFL	NBW	QUE	ONT	MAN	BCO	ALA
E1	6	-	10	3	2	-	-
E2	-	1	2	5	-	-	-
E3	1	1	1	1	-	-	-
E4	-	-	1	3	-	-	-
E5	-	-	2	-	-	-	-
E6	-	-	1	-	-	-	-
E7	-	-	1	1	-	-	-
E8	-	-	1	-	-	-	-
E9	2	2	1	-	-	-	-
E10	1	-	1	1	-	-	-
E11	-	-	1	-	-	-	-
E12	-	-	1	-	-	-	-
E13	1	-	3	1	-	-	-
E14	-	-	1	-	-	-	-
E15	1	-	-	-	-	-	-
E16	1	-	-	-	-	-	-
E17	3	1	-	-	-	-	-
E18	3	-	-	-	-	-	-
E19	10	-	-	-	-	-	-
E20	1	1	-	-	-	-	-
E22	1	-	-	1	-	-	-
E23	-	-	-	-	1	-	-
E24	-	-	-	-	1	-	-
E25	-	-	-	-	1	-	-
E26	-	-	-	-	5	-	-
E27	-	-	-	-	1	-	-
E28	-	-	-	-	1	-	-
E29	-	-	-	-	1	-	-
E30	-	-	-	-	1	-	-

[a]Abbreviations for localities as defined in the text.

Appendix I. (continued)

Hap. | Loc.	NFL	NBW	QUE	ONT	MAN	BCO	ALA
E31	-	-	-	-	2	-	-
E32	-	-	-	-	2	-	-
E33	-	-	-	-	1	-	-
E34	-	-	-	-	1	-	-
E35	-	-	-	-	1	-	-
E36	-	-	-	-	1	-	-
E37	-	-	-	-	2	-	-
E38	-	-	-	-	1	-	-
E39	-	3	-	-	-	-	-
E40	-	2	-	-	-	-	-
E41	-	-	-	1	-	-	-
E42	-	-	-	1	-	-	-
E43	-	-	-	1	-	-	-
E44	-	-	-	1	-	-	-
E45	-	-	-	1	-	-	-
E46	-	-	1	-	-	-	-
E47	-	-	-	-	1	-	-
W1	-	-	-	-	-	2	-
W2	-	-	-	-	1	6	3
W3	-	-	-	-	-	1	-
W4	-	-	-	-	-	3	-
W6	-	-	-	-	-	1	-
W7	-	-	-	-	-	1	-
W8	-	-	-	-	-	5	8
W9	-	-	-	-	-	2	-
W11	-	-	-	-	-	1	-
W12	-	-	-	-	-	1	-
W13	-	-	-	-	-	1	-
W14	-	-	-	-	-	1	-
W15	-	-	-	-	-	1	-