Molecular Ecology (2006) **15**, 2505–2525 doi: 10.1111/j.1365-294X.2006.02956.x

Lineage origin and expansion of a Neotropical migrant songbird after recent glaciation events

MARYLÈNE BOULET* and H. LISLE GIBBS*†
Department of Biology, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada, †Department of Evolution, Ecology and Organismal Biology, Ohio State University, 300 Aronoff Laboratory, 318 W. 12th Avenue, Columbus, OH 43210-1293, USA

Abstract

Birds of the Northern Hemisphere often harbour the genetic signature of postglaciation expansion but analyses identifying the location of refugia and the directionality of expansions are rare. Here we explore the evolutionary history of yellow warbler lineages, focusing on how these lineages recolonized their current range. We genotyped samples from 696 yellow warblers via direct sequencing of a 333-bp control region I mitochondrial DNA fragment or lineage-specific genotyping. Phylogenetic analysis revealed two monophyletic clades: a highly migratory group including previously identified eastern and western lineages and a less migratory group including a lineage consisting of tropical residents and a new 'southern' lineage localized in southwest United States. We then modelled the expansion of the eastern and western lineages, identified the location of potential refugia and assessed the importance of migration as a historical factor promoting gene flow. The expansion of the eastern lineage proceeded from a main refugia in the eastern United States, with possible contribution of an additional local refugia. In the western lineage, the expansion proceeded from a single refugia possibly located in western United States. Because two lineages overlapped to varying degrees in central North America, we suggest that the Canadian Prairies offered a bridge of riparian habitats where the lineages met after glacier retreat, while the US Central Great Plains acted as a barrier that limited secondary contact. Finally, gene flow was more important along the north–south axis of migration than away from it, suggesting spring migration played a role in the dispersal of lineages.

Keywords: *Dendroica petechia*, expansion, DNA lineages, Pleistocene glaciations, yellow warbler

Received 23 September 2005; revision received 10 February 2006; accepted 6 March 2006

Introduction

Recently, a number of studies have explicitly tested the contributions of refugia as historical determinants of the genetic differentiation between populations (Holder *et al*. 1999; Griswold & Baker 2002; Flagstad & Røed 2003). Not surprisingly, several vertebrates of the Northern Hemisphere harbour the signature of postglaciation expansion, as reflected by star-like phylogenies and a wave-shaped distribution of pairwise differences between sequences (Zink 1997).Yet explicit analyses of the number of refugia implied in the isolation process and the directionality of expansions are rare. Such analyses utilize the spatial distribution of alleles and geographic coordinates of sampling sites and are a form of landscape genetics (Manel *et al*. 2003).

A simple test to detect the possibility of multiple refugia is to correlate genetic diversity with geographic location of sampling sites. The assumption is that populations at the limits of their ranges are less diverse. If genetic diversity remains high or increases at the limits of the range, then these populations may be composed of two or more lineages that differentiated in distinct glacial refugia (Hewitt 1996, 2000; Griswold & Baker 2002). More sophisticated methods, such as geographic information system (GIS) may be applied to genetic data to model trait clines and examine spatial patterns in genetic variability or differentiation (Ritchie *et al*. 2001;

Correspondence: Marylène Boulet, PhD, Present address: Département de biologie, Pavillon Charles-Eugène Marchand (lab 1133), Université Laval, Québec, QC G1K 7P4 Canada. Fax: 418-656-7176; E-mail: marylene.boulet@bio.ulaval.ca

Hoffmann *et al*. 2003). By mapping the occurrence of differential traits or genetic variation, this technique can identify barriers to expansions and potential refugia.

The mechanisms by which expansions from refugia may occur are not always well known, especially in species that are not restricted to a specific habitat such as migratory birds. Helbig (2003) recently proposed that in birds, migratory species will show less phylogeographic structure along the main migration axis than away from it because gene flow is facilitated by migration. Helbig found support for this hypothesis in the northern yellow warbler (*Dendroica petechia*, group *aestiva*), a Neotropical bird migrant with an extremely wide breeding distribution. In this species, there was stronger genetic differentiation (inferred to be less gene flow) along the east–west axis than along the north–south axis of migration (Milot *et al*. 2000). However, Helbig did not take into account that a historical factor (Pleistocene glaciation) was primarily responsible for east–west differentiation between populations, which could bias estimates of gene flow.

Here we reexamine the geographic distributions of the eastern and western lineages of the yellow warbler initially described by Milot *et al*. (2000) using a geographically more widespread set of samples and geostatistical analyses. We then build on these new results to trace the evolutionary history of the yellow warbler lineages, and in particular the two lineages that were likely impacted by Pleistocene glaciations, i.e. the eastern and western lineages. Using phylogenetic analyses, we begin by determining whether the eastern and western lineages originated from a newly identified southern group during Pleistocene glaciations. We test whether the eastern and western lineages expanded from single or multiple lineages by examining geographic patterns of diversity. We assume that genetic diversity will remain high or increase at the limits of the range if populations were isolated in multiple refugia and later became in contact and merged. Thus we predict that diversity will remain: (i) high at high latitudes for both lineages, (ii) high at westerly longitudes for the eastern lineage, (iii) high at easterly longitudes, due to colonization from multiple refugia as observed in northern Europe in some species (Hewitt 1996, 2000; Griswold & Baker 2002). Finally, we analyse geographic patterns of gene flow to determine whether migration is a historical factor of gene flow in the yellow warbler. We predict that gene flow will be more important along the north–south migration axis than along the east–west axis in each of the lineages.

Materials and methods

Data collection

We obtained 684 yellow warbler samples from 56 breeding sites all across Canada and the United States (Appendix I). Among these samples, 155 were from seven locations included in a previous study (Milot *et al*. 2000). In addition, we obtained 12 samples of the yellow warbler resident forms: eight mangrove warblers *Dendroica petechia eritachorides* and four golden warblers *Dendroica petechia petechia* (Appendix I). Blood and feather quills were acquired from field trips and banding stations. Blood was sampled by puncturing the brachial vein, preserved in lysis buffer (Seutin *et al*. 1991) and later extracted using a salt extraction protocol (L. DeSousa, personal communication). Feather quills were obtained by pulling out 2–4 rectrices from live animals and DNA was obtained using the QIAamp Tissue kit (QIAGEN) or the Genelute mammalian genomic DNA kit (Sigma-Aldrich). Additional details regarding collection and extraction of samples can be found in Boulet (2004).

DNA sequencing and lineage genotyping

We used two methods to identify the mtDNA lineage of samples. First, we sequenced a 333-bp mtDNA control region I sequence from 390 northern yellow warbler, eight mangrove warblers, and four golden warblers using primers DPdl-L5 and DPdl-H4 or primers ND6L303 and DLOOP-H700 when sequences where less clear. The polymerase chain reaction (PCR) conditions are described in Milot *et al*. (2000) and Boulet (2004). We sequenced the amplified fragments with primers DPdl-L5 and DPdl-H4 using the Thermosequenase kit (Amersham Biosciences) or using an ABI3100 Genetic Analyser (Applied Biosystems). We aligned sequences in BIOEDIT (Hall 1999) and trimmed the first 12 bases of the original 344-bp sequence because the first bases at the 5′ end were sometimes missing in manually sequenced samples. We renumbered the locus position 1–333 because we found an indel at the 3′ end of the sequence. The trimming procedure resulted in the reclassification of haplotype E47 (see Milot *et al*. 2000) into haplotype E01.

A previous study identified two fixed nucleotide differences in the mitochondrial control region I that could differentiate eastern from western haplotypes: all eastern haplotypes had a GCG at sites 234–236, whereas all western haplotypes had a CCA at sites 234–236 (Milot *et al*. 2000). We took advantage of this variation to develop lineage-specific primers that allowed us to rapidly screen 294 additional individuals for variation at these lineage-specific positions. We used three primers: (1) primer DPdl-EAST annealing to the east-specific GCG motif and primer DPdl-HS3; (2) primer DPdl-WEST annealing to the west-specific CCA motif; and (3) primer DPdl-HS3 (see Boulet 2004 for details of PCRs). Samples that showed a clear bright band on agarose gels with primers DPdl-WEST and DPdl-HS3 were classified as western haplotypes. Samples that did not amplify were further tested using the DPdl-EAST primer and

were classified as eastern samples if we obtained a clear amplification. Samples that had weak and ambiguous band with both DPdl-EAST and DPdl-WEST primers were sequenced with DPdl-L5 and DPdl-H4 to determine their lineage. This revealed that poorly amplifying birds were southern individuals with an ATG mutation at sites 234–236 and a single eastern individual with a GTG mutation at sites 234–236.

Phylogenetic analysis of yellow warbler lineages

We used the program COLLAPSE in TCS version 1.13 (Clement *et al*. 2000) to identify the number of unique yellow warbler haplotypes. We verified the neutrality of the control region fragment using Tajima's *D* test (Tajima 1989) as implemented in ARLEQUIN 2.000 (Schneider *et al*. 2000). We constructed a minimum spanning network for each lineage of the northern yellow warblers using TCS 1.13. We also performed a Bayesian phylogenetic analysis on mtDNA sequences to determine the historical relationships of the lineages. We determined the appropriate substitution model and parameters using a hierarchical likelihood ratio test in MODELTEST 3.06 (Posada & Crandall 1998). We entered the selected model (HKY85, transition/transversion ratio = 9.1, proportion of invariable sites = 0.73, gamma parameter = 0.64) into MRBAYES 3.0 (Huelsenbeck & Ronquist 2003) to construct a phylogenetic tree. We ran the analysis for 2 000 000 generations and sampled every 100 generations. Burn-in frequency was set to the first 20% of the sampled trees. Direct examination of the sampled log-likelihood values showed that values had reached a stationary equilibrium by this point. We rooted the tree with two sequences of chestnut-sided warbler (*Dendroica pennsylvanica*) generated with primers ND6L303 and DLOOP-H700.

Estimates of divergence between lineages

To determine when the eastern and western lineages diverged from the southern or resident lineages, we employed the program MDIV (Nielsen & Wakeley 2001) which simultaneously estimates theta ($\theta = 2N_{ef}\mu$), immigration rate ($M = N_{ef}m$), lineage population divergence ($T = t/N_{ef}$), and time to most recent common ancestor ($TMRCA = t/N_{ef}$), where N_{ef} refers to the effective population size of females and t refers to a generation time of 1 year for yellow warblers. We ran two series of analyses consisting of (i) all southern individuals ($n = 10$), all resident yellow warblers from Mexico ($n = 9$), 25 randomly chosen eastern and 25 randomly chosen western individuals; (ii) two sets containing half of the eastern and western sequences for an east–west comparison. All east–west comparisons gave similar results and we will only present estimates for one half split data set. We translated the time of population divergence (T) and time to most common ancestor ($TMRCA$) into a time in year units (t) using different mutation rates to account for uncertainties in the rate of evolution of the avian control region. Following Pérez-Tris *et al*. (2004), we used a low mutation rate (0.1 substitutions/site/million years or s/s/Myr), an intermediate mutation rate (0.2 s/s/Myr), and a high mutation rate (0.3 s/s/Myr). The low mutation rate is equivalent to the mutation rate used by Milot *et al*. (2000). We assumed a generation time of one year.

Demographic expansions of eastern and western lineages

We first verified that the eastern and western lineages had the genetic signature of a sudden expansion by constructing mismatch distributions of pairwise nucleotide differences among haplotypes. A unimodal distribution of pairwise differences suggests a population expansion, whereas a multimodal distribution suggests stable equilibrium or possible geographic structure (Rogers 1995). We ran mismatch analyses for the eastern and western lineages separately and then for sites comprising at least nine individuals (eastern lineage, $n = 9$; western lineage, $n = 7$). We split admixed sites into subsites of a single lineage (e.g. Last Mountain became Last Mountain East and Last Mountain West). To increase the sample size of a few sites, we pooled Cuyahoga, Oak Harbor and Vicksburg to create the South Great Lake group and Fletcher Lake with Tatlayoko Lake to create the Central BC group (Appendix II). We used a test developed by Schneider & Excoffier (1999) to compare the observed pairwise distributions with a distribution predicted by a sudden-expansion model. In addition, we used Fu's test which detects populations that have expanded (Fu 1997). Both tests of expansions were done in ARLEQUIN 2.0 (Schneider *et al*. 2000).

Geographic distribution of lineages and possible location of refugia

We initially mapped the frequency of eastern, western, and southern lineages in the 56 breeding sites using the ARCGIS 8.0 software (ESRI Inc.). We then used a multilogit model to statistically describe the spatial distribution of the three lineages (Proc LOGISTIC, SAS 2002). A multicategory logit model is a generalization of a logistic regression that handles more than two response categories (Agresti 1996). In our model, the response categories were the number of eastern, western, and southern haplotypes/number of samples within a site ($n = 684$ individuals in total, Appendix I). The tested variables were the latitude, the longitude, and the interaction between the latitude and longitude of the sampling sites. We used a backward procedure to identify significant variables and we evaluated the fit of the model using deviance statistics (Hosmer & Lemeshow 1989). We calculated the predicted probabilities of observing a specific lineage based on selected multilogit model.

We then performed a geostatistical analysis on the predicted probabilities of our model to graphically illustrate the distributions of lineages. We used a kriging technique in the Geostatistical Analyst module of the ARCGIS 8.0 software (Johnston *et al*. 2001) and we constructed five genetic contours (10%, 25%, 50%, 75% and 90% probability). The 90% genetic contours defined the limits of an abundance zone that presumably include the glacial refugia from which a given lineage expanded, the 10% contour defined an exclusion zone where the lineage becomes extremely rare, and the 25–75% contours defined an overlap zone colonized by more than one lineage. For each lineage, we overlaid the ice sheet limits at 18 000 years before present (ybp) (eastern lineage: Laurentide Ice Sheet, western lineage: Cordilleran Ice Sheet) available from published maps (Dyke & Prest 1987) to delineate an ice-free region that would have included the refugia. We assumed eastern refugia would be located east of the 90% eastern abundance zone and south of the Laurentide Ice Sheet including the east plains and glacial islands, while western refugia would be located west of the of the 90% western abundance zone and south or northwest (Beringia) of the Cordilleran Ice Sheet. For continental refugia, our criteria are based on the assumption that habitats available for warblers oscillated latitudinally during climatic changes. Thus continental refugia would be located anywhere (latitudinally) south of the glaciers and within a zone where the lineage is very abundant.

Genetic diversity and possible number of refugia

Hewitt (1996, 2000) demonstrated that the last glaciations had a profound effect on genetic diversity: the genetic diversity in populations located in the northernmost areas was often reduced when compared to populations located in areas that remained unglaciated. Therefore, the presence of high levels of genetic diversity in previously glaciated northern areas would suggest colonization from multiple refugia (Griswold & Baker 2002). For this analysis, we grouped sites with few individuals to increase sample size. Those sites were usually located < 300 km from each other without any major barriers (Appendix II). We split sites composed of individuals of two lineages into groups of the same lineage and we obtained 19 eastern groups and 20 western groups having at least two sequences (Appendix II). To test whether the eastern and western lineages expanded from a single refugia or from multiple refugia, we estimated correlations between two indices of genetic diversity (π, nucleotide diversity and H, haplotype diversity) vs. the longitude and latitude of sampling sites. When correlations where significant ($P > 0.05$), we quantified the relationship between the index of genetic diversity and the significant variable using linear or quadratic regressions.

Geographic patterns of gene flow

We assessed the relative role of drift and historical factors (isolation in refugia) as determinants of differentiation within lineages. We quantified the relationship between pairwise estimates of immigration M and geographic distance between group sites to determine if lineages are in equilibrium between drift and gene flow (Slatkin 1993; Hutchinson & Templeton 1999). A significant negative relationship between M and pairwise geographic distance of sites indicates equilibrium between gene flow and drift, whereas a nonsignificant relationship between these variables indicates that equilibrium has not yet been reached. In the latter case, diverse patterns can be obtained: little variance of M values around mean at any geographic distances suggests panmixia where gene flow is more important than drift, whereas a large variance of M values around mean at any geographic distances suggests fragmentation events, where drift is more important than gene flow (Hutchinson & Templeton 1999). Using the program MDIV, we calculated pairwise immigration rates M between 9 sites in the eastern lineage and 7 sites in the western lineage ($n \geq 9$ individuals/site, Appendix II). We conducted Mantel tests of isolation-by-distance in IBD 1.52 (Bohonak 2002) using the log (M) and the log (distance in km between sites) (Slatkin 1993). We then tested whether immigration was more important along the main migration axis (i.e. latitudinally) than away from the migration axis (i.e. longitudinally) by computing Mantel tests between M values (calculated for the isolation-by-distance analysis) and the difference in degrees of latitude or longitude of sites in eastern and western lineages separately.

Results

Polymorphisms and phylogenetic analysis

We found 124 unique haplotypes within the 398 yellow warbler sequences including the golden and mangrove warblers (Appendix III). Tajima's D did not differ from zero ($D = -0.34$, $P = 0.38$) suggesting that this high variation in the control region segment was selectively neutral. The most common haplotypes were W02 ($n = 49$), W08 ($n = 35$), E01 ($n = 48$), E02 ($n = 19$), and E19 ($n = 11$). The network analysis showed that within the eastern lineage, the haplotype E01 occupied a central position and had the greatest number of links leading to other haplotypes (Appendix IV). In the western lineage, both W02 and W08 were central to the network. Finally, haplotype S01 was the most central haplotype of the southern lineage and was present in the three sites where the southern lineage was observed.

The Bayesian phylogenetic tree showed two highly supported monophyletic clades: (i) a small clade including

four haplotypes (S01 to S04) found in yellow warblers breeding in southwest United States and seven haplotypes (S05 to S11) in the Neotropical residents and (ii) a large clade including all eastern and western haplotypes (Fig. 1). Within the small clade, the southern haplotypes tended to be basal, while the haplotypes from Neotropical residents in Mexico, Venezuela, and Puerto Rico (S05 to S11, thereafter named resident haplotypes) tended to be nested within the southern haplotypes (Fig. 1). There was no support for the taxonomic division of golden and mangrove warblers despite significant differences in head plumage coloration between these taxa (Browning 1994). Within the large clade, all the western haplotypes were nested within the eastern haplotypes with 100% support of posterior probabilities, confirming that the western lineage derived from the eastern lineage (Milot *et al*. 2000). The analysis clearly identified haplotypes E20 (found in birds from Gros Morne NF, Germantown NB), E39 (Germantown NB), and E54 (Last Mountain SK) as ancestral to the western lineage.

Divergence of the eastern and western lineages

The pairwise MDIV estimates of lineage divergence confirmed the patterns of divergence suggested by the phylogenetic analysis. The eastern and western lineages diverged from the southern and resident lineages during the same broad period, i.e. between 154 000 and 164 000 years ago (Table 1). This corresponds to the end of an Illinoian glaciation (Pagé 1999). The divergence between the eastern and western was much more recent and occurred about 85 000 years ago. Estimates of lineage *TMRCA* were older, ranging from 131 000 ybp (eastern vs. western lineage) to 205 400 ybp (eastern vs. resident lineage, Table 1). Older estimates of *TMRCA* relative to estimates of lineage divergence are expected if lineages continued to exchange genes after being isolated (Edwards & Beerli 2000).

Demographic expansion in eastern and western lineages

Both eastern and western lineages conformed to the sudden-expansion model (sum of squared deviations $SSD < 0.01$, $P \geq 0.2$, Table 2, Fig. 2). This was further confirmed by the strongly negative F_S statistics (all < -20, $P < 0.001$), which is indicative of expansion. The age of expansion in units of mutational time (τ) was twice as old in the eastern lineage ($\tau = 4.08$) than in the western lineage ($\tau = 1.98$). Although these τ values were not significantly different due to the large 95% confidence intervals around the point estimates, these results suggest that the expansion of the eastern lineage preceded the expansion of the western lineage (Table 2). Time of expansion was about 37 400 ybp in the east lineage, and 18 200 ybp in the western lineage, when averaging across mutations rates (Table 2).

Within the eastern lineage, all sites showed evidence of a stepwise sudden expansion (*SSD* tests, all $P > 0.05$, Fig. 3). However, some sites had distinct pairwise distributions: while the easternmost sites like Germantown, Gros Morne and Trois-Rivières had multiple peaks and a wide range of mutation steps, Fort Riley in Kansas had a distinctive and typical unimodal distribution. In the western lineage, all sites also conformed to the null hypothesis of sudden expansion (*SSD* test, all $P > 0.05$; see Fig. 3) and nearly all of them had distinctive unimodal distributions. Fairbanks (AK) had an exceptionally high proportion of similar pairs of sequences and showed no evidence of expansion according to the F_S test (Fig. 3).

Geographic distribution of lineages

The three lineages clearly had distinct geographic distributions. Eastern haplotypes were present from Newfoundland to Alberta and Montana, while western haplotypes were present from Michigan to Alaska (Fig. 4). The southern

Table 1 MDIV estimates of immigration rates ($M = N_{ef}m$), population divergence times (T), and time to most recent common ancestor (*TMRCA*) between pairs of yellow warbler lineages. Population and *TMRCA* times are translated in year units before present (ybp). An average estimate across the three mutations rates (Low: 0.1 s/s/Myr; Intermediate: 0.2 s/s/Myr; High: 0.3 s/s/Myr, see Pérez-Tris *et al*. 2004) is given for each comparison

Lineages	*M* (95% CI)	*T*	*tT* (ybp)	*TMRCA*	*tTMRCA* (ybp)
East-South	0.05 (0–0.18)	2.53	Ave.: 153 900 Low: 251 900 Int.: 126 000 High: 84 000	3.15	Ave.: 191 600 Low: 313 600 Int.: 156 800 High: 104 500
East-Resident	0.008 (0–0.10)	2.4	Ave.: 164 300 Low: 268 900 Int.: 134 400 High: 89 600	3.00	Ave.: 205 400 Low: 336 000 Int.: 168 000 High: 112 000
West-South	0.002 (0–0.03)	4.47	Ave.: 164 100 Low: 268 500 Int.: 134 200 High: 89 500	4.91	Ave.: 180 200 Low: 294 900 Int.: 147 400 High: 98 300
West-Resident	0.002 (0–0.03)	3.5	Ave.: 161 200 Low: 263 800 Int.: 131 900 High: 87 900	4.33	Ave.: 199 500 Low: 326 400 Int.: 163 200 High: 108 800
East-West	0.008 (0–0.4)	0.61	Ave.: 85 000 Low: 139 000 Int.: 69 500 High: 46 300	0.94	Ave.: 131 000 Low: 214 300 Int.: 107 100 High: 71 500

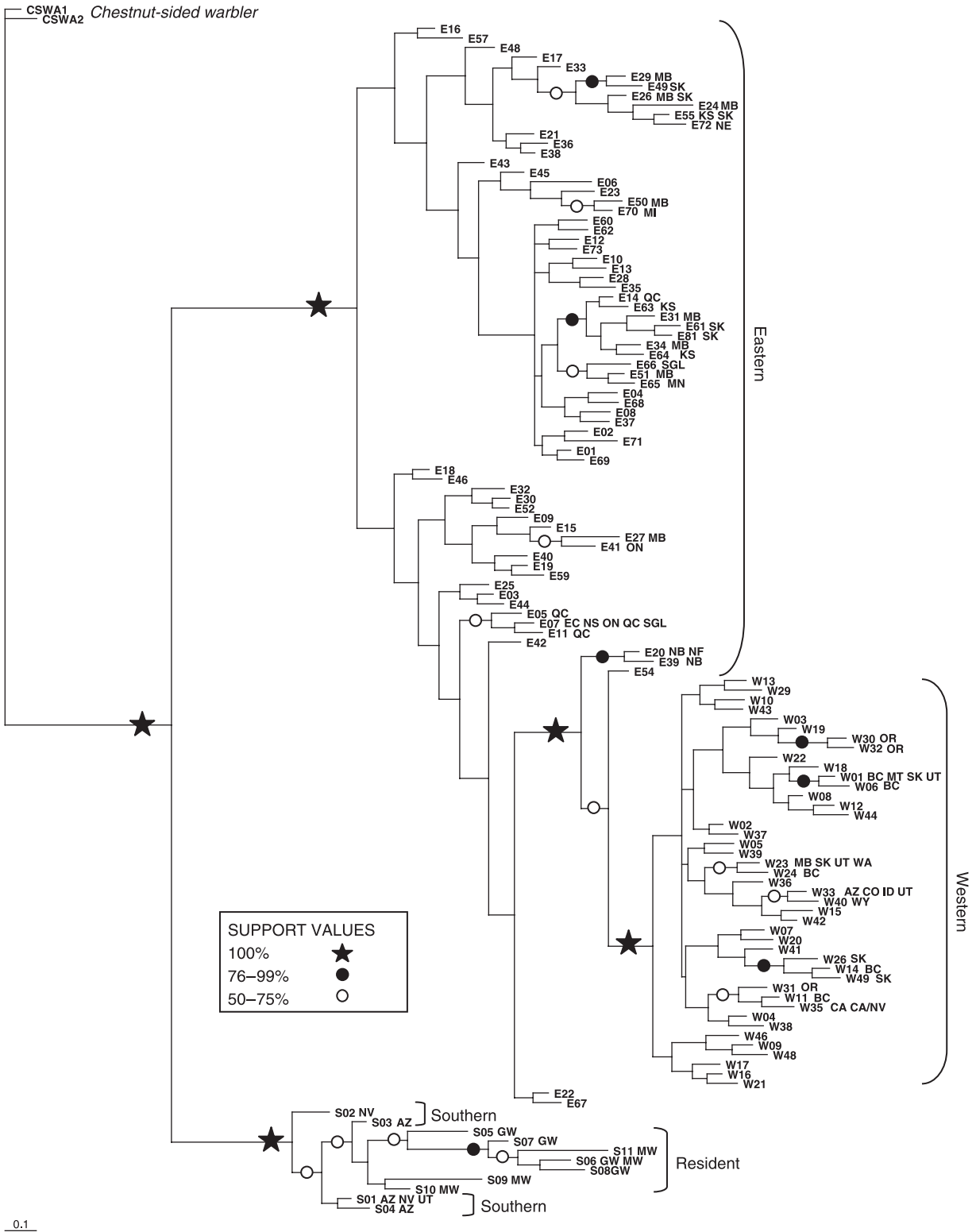

Fig. 1 Phylogenetic Bayesian tree showing the relationships between the 124 yellow warbler mitochondrial haplotypes and their corresponding lineages. The tree was rooted with two chestnut-sided warbler sequences. Stars refer to 100% support in posterior probabilities, closed circles to 76–99% support in posterior probabilities, and open circles to 50–75% in posterior probabilities. MW designates mangrove warblers and GW designates golden warblers. Abbreviations beside significant clusters refer to states or provinces where haplotypes were found.

Table 2 Demographic parameters describing population expansions for eastern and western lineages obtained from mismatch analyses, corresponding 95% confidence intervals (CI), and results of the two expansion tests [*SSD*, Schneider & Excoffier's test of sudden expansion (Schneider & Excoffier 1999); F_S: Fu's test of neutrality (Fu 1997)]. The parameters τ, t, θ_0, θ_1 are age of expansion in units of mutational time, age of expansion in years before present (ybp), population size before expansion, and population size after expansion in units of mutational time. The time of expansion was averaged across the three mutations rates tested for the control region (Low: 0.1 s/s/Myr; Intermediate: 0.2 s/s/Myr; High: 0.3 s/s/Myr, see Pérez-Tris *et al*. 2004)

Parameters	East	West
τ (95% CI)	4.08 (1.87–6.10)	1.98 (1.13–2.32)
Average t in ybp	37 400	18 200
Low	61 200	29 800
Intermediate	30 600	14 900
High	20 400	9900
θ_0 (95% CI)	0.01 (0.00–1.86)	0.00 (0.00–0.87)
θ_1 (95% CI)	14.45 (6.67–5624.45)	3277.50 (25.55–8065.00)
Sudden expansion test *SSD* (H_0 = expansion)	0.002; $P = 0.6$	0.001; $P = 0.2$
Fu's expansion test F_S (H_0 = no expansion)	−25.8; $P < 0.001$	−27.5; $P < 0.001$
Evidence of expansion	Very strong	Very strong

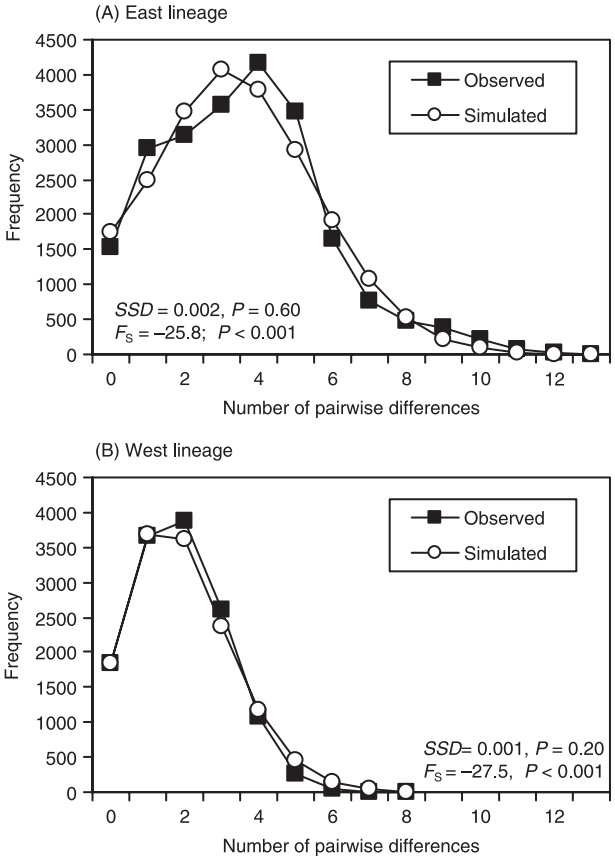

Fig. 2 Mismatch distribution of yellow warbler haplotypes in the (A) eastern lineage and (B) western lineage. The line with open circles indicates the expected distribution under a model of sudden expansion and the line with closed squares indicates the observed distribution.

haplotypes were restricted to the southwest part of the breeding range in Lake Mead NV ($n = 6$), Flagstaff AZ ($n = 3$), and Vernal UT ($n = 1$). The spatial distributions of the three lineages were best explained by the latitude, the longitude, and the interaction between the latitude and longitude of the sampling sites (multicategory logit model, longitude: d.f. = 2, $\chi^2 = 67.05$, $P < 0.001$; latitude: d.f. = 2, $\chi^2 = 47.76$, $P < 0.001$, longitude*latitude: d.f. = 2, $\chi^2 = 52.12$, $P < 0.001$, Fig. 4), meaning the overall distributions of the lineages were generally but not always parallel to each other. In fact, the longitude*latitude interaction was essentially caused by the southwest distribution of southern haplotypes because only longitude explained the spatial distribution of eastern and western lineages after removal of the southern lineage from the analysis (logistic model: d.f. = 1, $\chi^2 = 66.30$, $P < 0.001$, map not shown).

The geostatistical analysis defined three main regions: (i) an exclusion zone, where a lineage was nearly absent (10% genetic contour); (ii) an abundance zone, where a lineage was extremely abundant (90% genetic contour); (iii) an overlap zone, where a specific lineage shared the breeding area with another lineage (area between 10% to 90% genetic contours). For the eastern lineage, the abundance zone was located east of –95° longitude in the United States but diverted eastward to –85° longitude in Ontario (Fig. 5). The western abundance zone broadly followed –110° longitude but diverted westward when it reached –40° latitude in the Rocky Mountains in the United States (Fig. 6). In contrast, the abundance zone (with a 75% genetic contour in this case) was restricted to New Mexico (Fig. 7). The east–west overlap zone was particularly extensive and ranged from 1300 km in the US Central Great Plains to

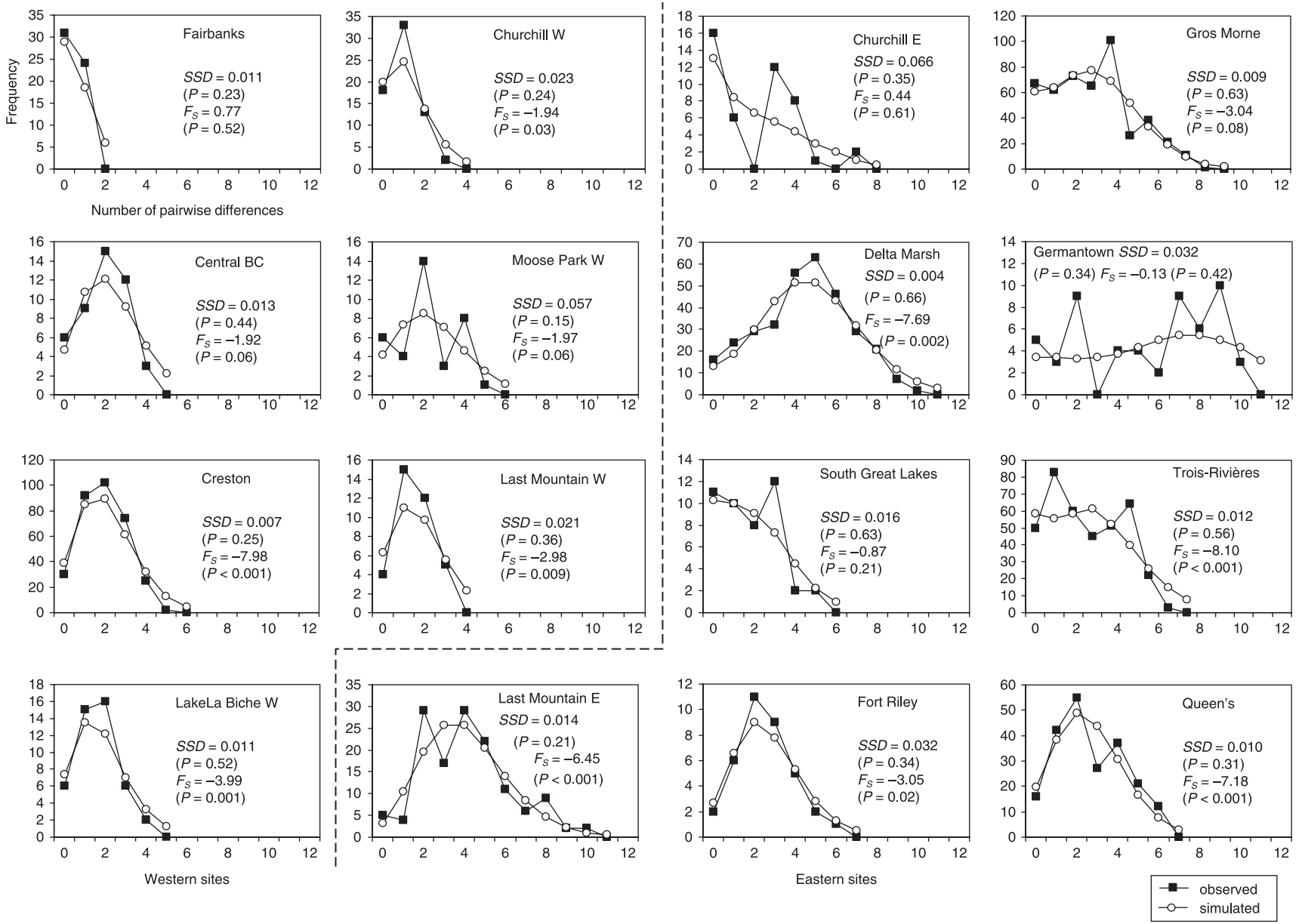

Fig. 3 Mismatch distribution of yellow warbler haplotypes in seven western and nine eastern lineage sites. The line with open circles indicates the expected distribution under a model of sudden expansion and the line with closed squares indicates the observed distribution. *SSD* refers to Schneider & Excoffier's test of sudden expansion (Schneider & Excoffier 1999) and F_S refers to Fu's test of neutrality (Fu 1997).

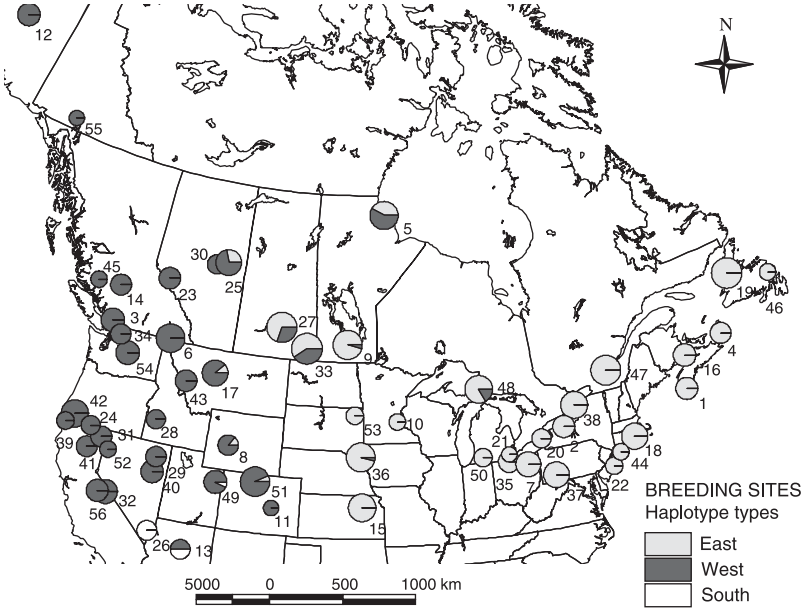

Fig. 4 Distribution of eastern, southern, and western haplotypes found in yellow warblers captured on 56 breeding grounds ($n = 684$ individuals). The size of pie charts is proportional to the number of samples for a specific site. The number below each pie chart corresponds to an identification number (refer to Appendix I for details): #1 Bon Portage Island NS, #2 Braddock Bay NY, #3 Burnaby area BC, #4 Cape Breton Island NS, #5 Churchill MB, #6 Creston BC, #7 Cuyahoga OH, #8 Deep Creek WY, #9 Delta Marsh MB, #10 Dodge Nature Center MN, #11 El Paso county CO, #12 Fairbanks AK, #13 Flagstaff AZ, #14 Fletcher Lake BC, #15 Fort Riley KS, #16 Germantown NB, #17 Great Falls MT, #18 Great Swamp RI, #19 Gros Morne National Park NF, #20 Haldimand ON, #21 Holiday Beach ON, #22 ISBP Ocean county NJ, #23 Jasper AB, #24 KLAM CA, #25 Lake La Biche AB, #26 Lake Mead NV, #27 Last Mountain SK, #28 Lucky Peak ID, #29 Mary's River NV, #30 Meanock AB, #31 Modoc CA, #32 Mono county CA, #33 Moose Park SK, #34 Mount Baker WA, #35 Oak Harbor OH, #36 Ponca State Park NE, #37 Powdermill NR PA, #39 RFSL CA, #40 Ruby Lake NV, #41 Sarc CA, #42 South Oregon OR, #43 Stevensville MT, #44 Suffolk county NY, #45 Tatlayoko BC, #46 Terra Nova National Park NF, #47 Trois-Rivières QC, #48 Vermilion MI, #49 Vernal UT, #50 Vicksburg MI, #51 Walden CO, #52 Washoe county NV, #53 Waubay SD, #54 Wenatchee WA, #55 Whitehorse YU, #56 Yosemite CA.

> 3000 km in the Canadian Prairies (Figs 5 and 6). The west–south overlap zone was sharper: only 500 km wide (Fig. 7).

Possible location of refugia

We overlay the maximum extent of the glacier advance at 18 000 ybp to the 90% abundance zone to delineate unglaciated areas where putative eastern and western refugia could be located. Based on our criteria, the eastern refugia was likely located east of –95° longitude in the United States and south of 40° latitude. This zone includes the three regions proposed as possible refugia: the mid-latitude ice-free lands, the east coast plains, and glacial islands (Fig. 5). For the western lineage, the west abundance zone was possibly west of –109° longitude in the United States to –116° longitude in northern Canada, south to –40° latitude in the Rocky Mountains and –36° in California. The sectors where the western refugia could be located included two main regions: Beringia and western US states south of the Cordilleran glacier (Fig. 6).

Genetic diversity and possible number of refugia

Eastern and western lineages had comparable levels of haplotype diversity (eastern: $H = 0.93$, $n = 212$ haplotypes; western: $H = 0.86$, $n = 164$ haplotypes) but differed in their levels of nucleotide diversity with the eastern lineage being nearly twice more diverse than the western lineage (east: $\pi = 0.0108$; west: $\pi = 0.0057$). In the eastern lineage, the diversity indices showed no decrease at northern latitudes (H: $r = -0.03$, $P = 0.90$; π: $r = 0.92$, $P = 0.11$) or at western longitudes where the lineage was becoming less abundant (H: $r = -0.37$, $P = 0.13$; π: $r = -0.27$, $P = 0.29$). This raises the possibility that the eastern lineage has survived in at least two refugia. The western lineage showed a different pattern: both indices negatively correlated with latitude (H: $r = -0.47$, $P = 0.04$; π: $r = -0.58$, $P = 0.008$, Fig. 8) but not with longitude (H: $r = 0.31$, $P = 0.18$; π: $r = 0.08$, $P = 0.74$). Closer inspection between diversity indices and latitude revealed a quadratic relationship: diversity was lowest at low and high latitudes and peaked at 45° latitude for H ($F_{2,17} = 8.0$, $R^2 = 0.49$, $P = 0.004$) and 42° for π ($F_{2,17} = 11.5$, $R^2 = 0.57$, $P < 0.001$). This

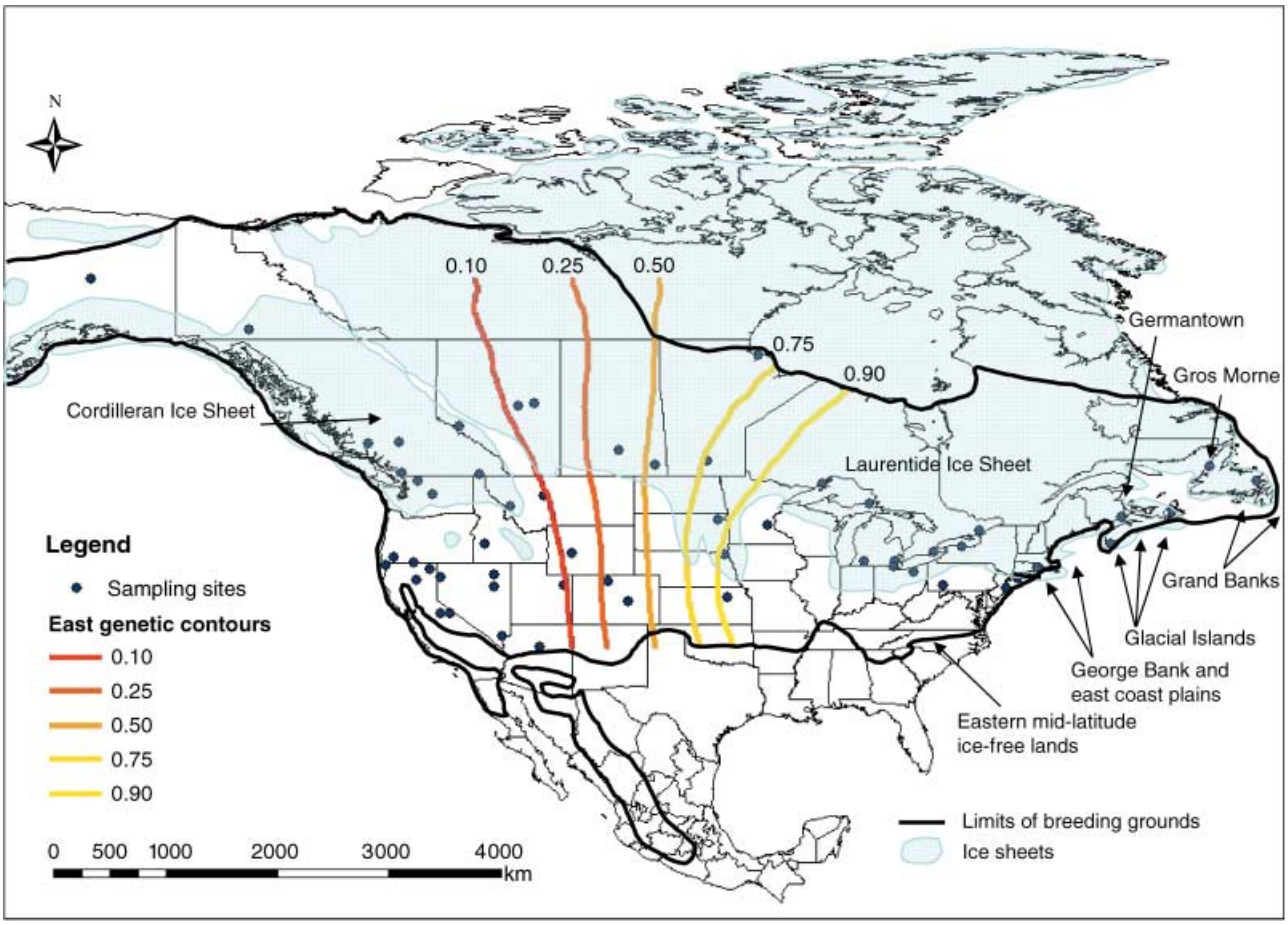

Fig. 5 Map of predicted probabilities of observing eastern haplotypes on the breeding grounds. Genetic contours were obtained by performing a geostatistical analysis (kriging) on the predicted probabilities of the multicategory logit model. The location of refugia in eastern North America according to geological evidence is indicated by arrows (Pielou 1991). We also show the location of Germantown and Gros Morne sites. The limits of the Cordilleran and Laurentide Ice Sheets were obtained from Dyke & Prest (1987).

result suggests the existence of an area of high diversity located in the western United States (California/Nevada: $H = 0.96$, $\pi = 0.9643$: Oregon: $H = 1.0000$, $\pi = 0.0100$; Idaho: $H = 0.90$, $\pi = 0.0084$).

Spatial patterns in gene flow

We found a significant relationship between the immigration rate M and the geographic distance between sites in the eastern lineage (Mantel test, $z = 31.8$, $r = -0.60$, $P < 0.001$, $y = 5.41 - 1.61x$, $R^2 = 0.36$, Fig. 9), indicating equilibrium between gene flow and drift in this lineage. The isolation by distance was not biased by the easterly location of New Brunswick: when New Brunswick was removed from the analysis, the isolation-by-distance pattern remained significant and the strength of the relationship increased ($z = 32.22$, $r = -0.73$, $P < 0.001$, $y = 5.29 - 1.54x$, $R^2 = 0.53$). The overall isolation-by-distance pattern observed within the eastern lineage was mainly due to longitudinally (east–west) restricted gene flow: there was a significant relationship between M and distance along an east–west axis ($z = 10.49$, $r = -0.43$, $P = 0.005$, $y = 1.58 - 1.09x$, $R^2 = 0.19$) but not along a north–south axis ($z = 6.20$, $r = -0.22$, $P = 0.12$, $R^2 = 0.05$). Within the western lineage, the relationship between M and distance was negative but not significant ($z = 29.82$, $r = -0.42$, $P = 0.10$, $y = 8.96 - 2.74x$, $R^2 = 0.17$; Fig. 9B). M values tended to be very variable at most geographic distances, a situation that occurs when drift is more important that gene flow (case III, Hutchinson & Templeton 1999). Gene flow was more restricted away from the migration axis (east–west) ($z = 9.94$, $r = -0.44$, $P = 0.08$, $R^2 = 0.20$) than along the latitudinal migration axis (north–south) ($z = 4.98$, $r = -0.28$, $P = 0.20$, $R^2 = 0.08$), suggesting that recent gene flow and colonization events occurred along the latitudinal axis.

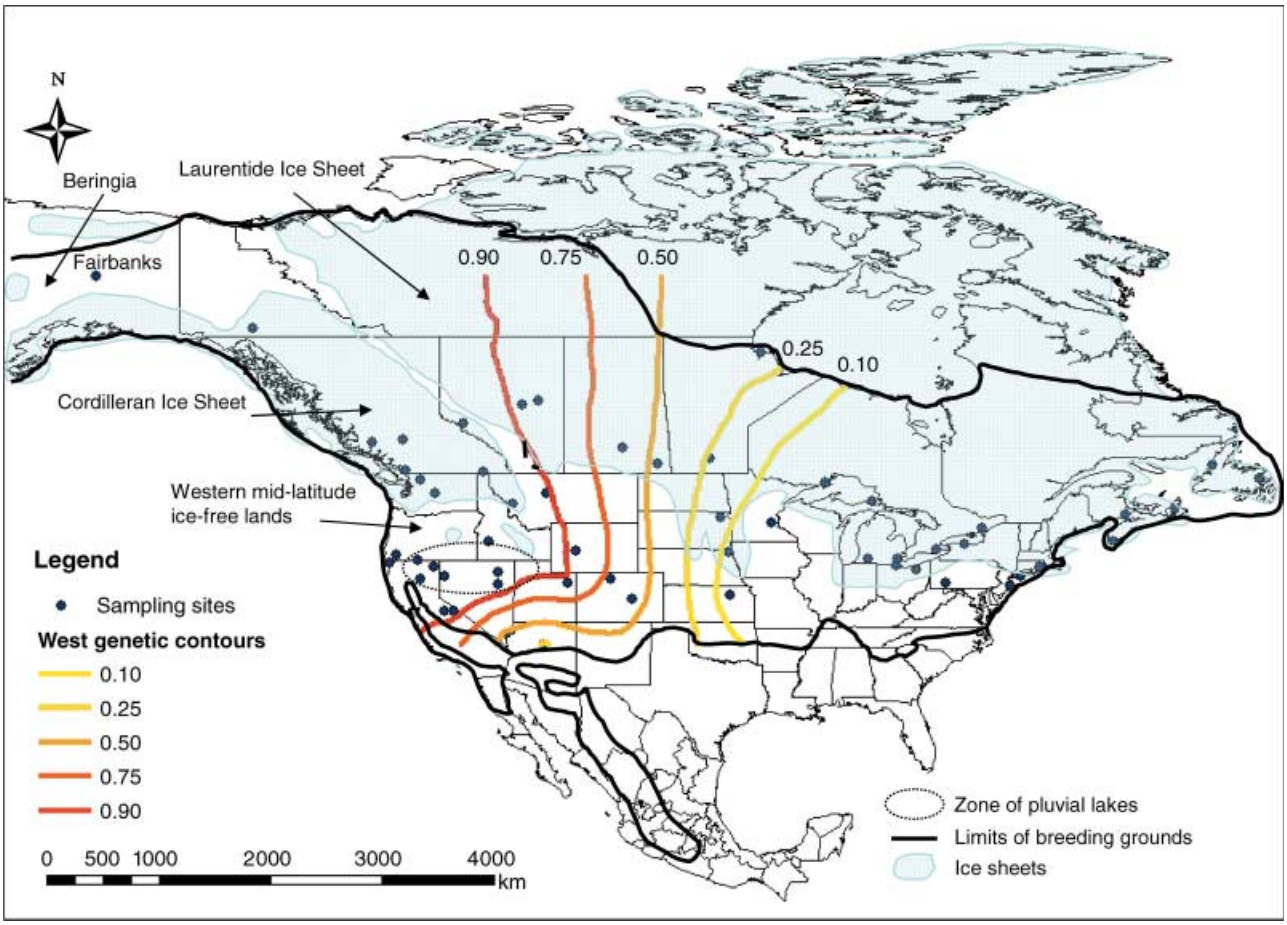

Fig. 6 Map of predicted probabilities of observing western haplotypes on the breeding grounds. Genetic contours were obtained by performing a geostatistical analysis (kriging) on the predicted probabilities of the multicategory logit model. The location of refugia in western North America according to geological evidence is indicated by arrows (Pielou 1991). The limits of the Cordilleran and Laurentide Ice Sheets were obtained from Dyke & Prest (1987). Note that western contours are not the exact mirror images of the eastern contours because of the presence of the southern lineage in southwest United States (see Fig. 7).

Discussion

Origin and expansion of eastern and western lineages

Our phylogenetic analysis clearly showed that the eastern and western lineages clustered together and apart from another group formed by the southern and resident lineages. Although the southern individuals are phylogenetically more related to resident forms, these individuals probably accomplish short-distance migrations based on the annual distribution of this lineage (Boulet 2004). We therefore consider that the eastern and western lineages together form a highly migratory group, while the southern and resident lineages form a less migratory group. These groups diverged about 160 000 years ago, or less than that if intermediate or high mutation rates are used (Table 1), corresponding to the end of the Illinoian glaciation and the following warming period (Pagé 1999). Assuming the ancestor of the highly migratory birds was a partial migrant or even a short-distance migrant, we propose that as the environment warmed up and early succession habitats moved northward into deglaciated areas, the yellow warbler extended its breeding distribution northward to occupy these new habitats and became increasingly migratory. Then, during another glaciation event, the migratory form became isolated into eastern and western refugia, and the eastern and western lineages diverged from each other, as initially showed by Milot *et al*. (2000). Assuming an intermediate mutation rate, the east–west divergence occurred about 69 500 ybp, during a glaciation event of the Wisconsin (79 000–65 000 ybp, Pagé 1999). This scenario of divergence is concordant with Mengel's model of warbler differentiation, where eastern warbler taxa repeatedly gave rise to western taxa during four glacial periods (Mengel 1964, 1970). Our phylogenetic analysis brings an additional element to this scenario: it clearly shows that the western lineage was

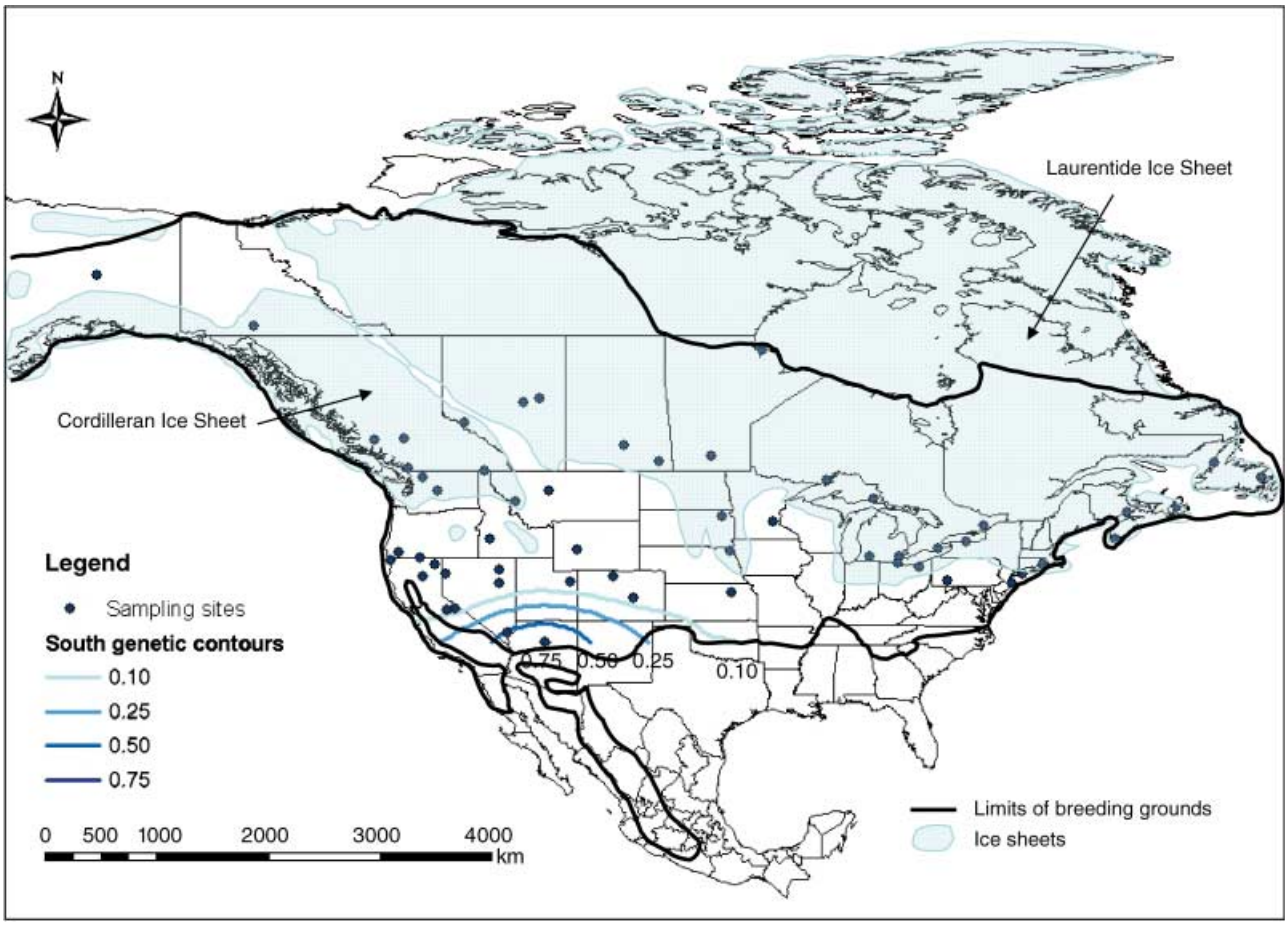

Fig. 7 Map of predicted probabilities of observing southern haplotypes on the breeding grounds. Genetic contours were obtained by performing a geostatistical analysis (kriging) on the predicted probabilities of the multicategory logit model. The limits of the Cordilleran and Laurentide Ice Sheets were obtained from Dyke & Prest (1987).

nested within the eastern lineage, and thus derived from the eastern lineage. Based on this result, we can speculate that the lineage that progressively invaded North America after the Illinoian glaciation was in fact an ancient eastern lineage.

The western expansion occurred at $\tau = 1.98$, or 29 800–9900 ybp depending on the mutation rate, while the eastern expansion occurred at $\tau = 4.08$ or between 61 200 and 20 400 ybp (Table 2). The discrepancies between the eastern and the western lineages may be due to retained ancestral polymorphism present in the eastern lineage. The phylogenetic analysis clearly supports the idea that the eastern lineage is ancestral to the western lineage (Fig. 1). The eastern haplotype network is more complex than in the western lineage and has two centres of differentiation (Appendix IV). The eastern mismatch analysis reflects this complexity: it shows two peaks that could be associated with two expansions, one at $\tau = 1$ and one at $\tau = 4$ (Fig. 3). In contrast, there is only one clear peak in the western mismatch analysis. Six eastern sites also show this bimodality in their mismatch distributions and four of them are located in the northeast part of the continent (Fig. 4). It is therefore possible that the eastern lineage expanded twice and some ancient haplotypes of the first expansion could have been retained in populations that persisted in refugia.

The genetic contours indicated that the eastern lineage expanded westward to the foot of the Rocky Mountains, whereas the western lineage expanded eastward beyond the Rocky Mountains. Thus, the Rocky Mountains did not act as a geographic barrier that constrained expansion of western forms (Milot *et al*. 2000; Ruegg & Smith 2002).

The expansion of the eastern and western lineages was not homogenous and the width of the secondary contact zone varies throughout the landscape. The east–west cline is steeper in the US Central Great Plains (about 1300 km) than in the Canadian Prairies (> 3000 km), suggesting the US Central Great Plains acted as a much greater barrier to dispersion than did the Canadian Prairies. These

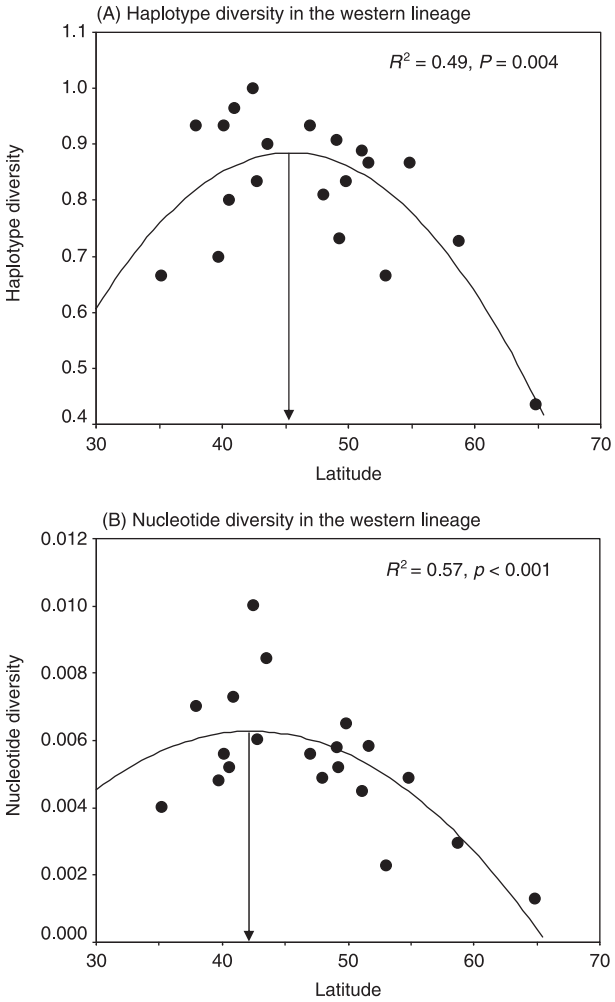

Fig. 8 Haplotype (A) and nucleotide (B) diversity in the western lineage in relationship with latitude of the sampling sites.

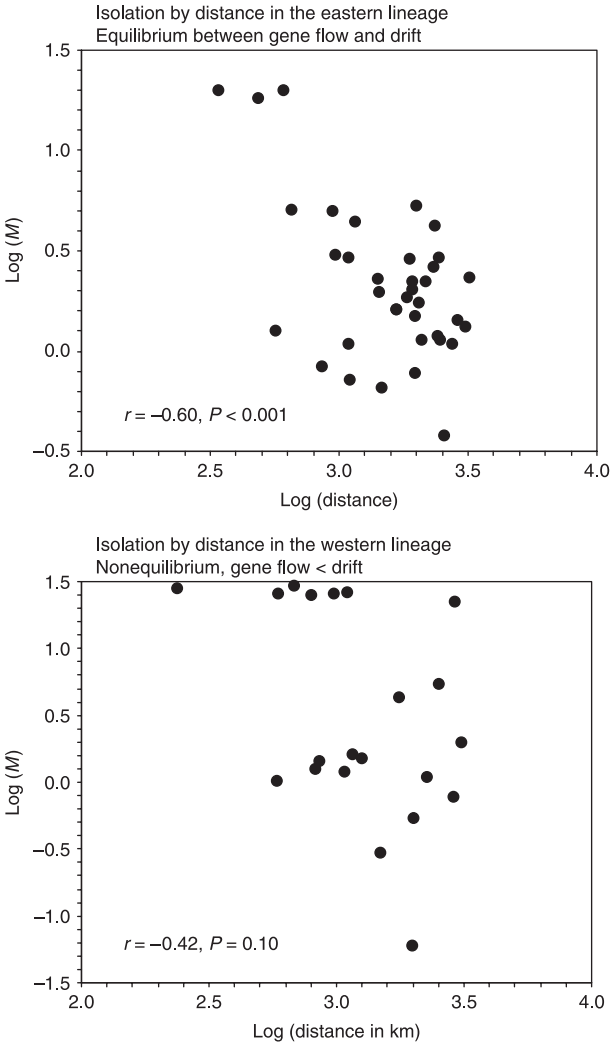

Fig. 9 Isolation-by-distance plots for all pairwise values of log (M) and log (distance in kilometre) in the eastern and western lineages.

physiographic regions differ in their geological history: while the Canadian Prairies were covered by the Laurentide Ice Sheet, the US Central Great Plains remained unglaciated and were partly covered by loess deposits creating an arid habitat not favourable to yellow warblers (Dawson 1992). Our results support Mengel's theory (Mengel 1970) that the US Central Great Plains acted as an isolating agent during full-glacial and interglacial periods by limiting the secondary contact between the lineages. The present-day abundance of yellow warblers still mirrors this isolation pattern: the density of yellow warblers is lowest the US Central Great Plains and peaks in the Canadian Prairies (Goosen & Sealy 1982; Sauer *et al*. 2003). Thus, the Canadian Prairies may have offered a bridge of riparian habitats where the expansions of the eastern and the western lineages met and hybridized in locations such as in Churchill in northern Manitoba (H. L. Gibbs, unpublished data).

Refugia of the eastern and western lineages

In eastern lineage, the indices of haplotype and nucleotide diversity were not associated with latitude or longitude of sampling locales, raising the possibility that the eastern lineage had survived in at least two refugia. The main eastern refugia was probably in eastern United States. Pollen distributions indicate that potential yellow warbler habitats were present along the US Gulf states (open woodland with deciduous trees, grassland and forbs), southeast US states (pine forest with some deciduous trees, spruce forests), and localized areas of the US Central Great Plains (spruce forests) (Gauthreaux 1980; Williams & Webb III 1996). In the east, two additional regions could have served as refugia for fauna: the east coast plains and the glacial islands including the Grand Banks near off Newfoundland (Pielou 1991). Two of our sites, Germantown and Gros Morne, are within 800 km of these localized refugia (Fig. 5). Both had

a peculiar haplotype composition, consisting in a mixture of most common haplotypes (E01, E02) and haplotypes typically found in the Maritimes (E09, E17, E19, E39, E40). In particular, Germantown had a ragged distribution with three main peaks and an exceptionally old time of expansion ($\tau = 8.79$ or 85 000 ybp using an average of the three mutation rates), the highest nucleotide diversity among eastern sites ($\pi = 0.0164$), and a significant contribution to the historical genetic structure based on isolation-by-distance analyses. Palaeoenvironment reconstructions suggest the presence of coniferous forests with dispersed deciduous trees on Georges Bank and the exposed continental shelf (Williams & Webb III 1996) as well as the presence of shrubs and herbs on Glacial Sable Island (Pielou 1991). Finally, the Grand Banks were probably a glacial refugia for the Newfoundland rock ptarmigan *Lagopus mutus welchi* during the last glaciation (Holder *et al*. 1999). We therefore conclude that the eastern lineage probably had two refugia: a main refugia located south of the Laurentide Ice Sheet in the eastern United States and a fragmented refugia possibly localized in the east coast plains and/or glacial islands.

In western North America, two regions could have harboured bird communities during Pleistocene glaciations: the mid-latitude ice-free lands and Beringia. For the western lineage, our results suggest the presence of a single refugia. First, diversity indices were lowest at low and high latitudes and peaked between 42° and 45° latitudes i.e. in northern California, Nevada, Utah, central Oregon and central Idaho. Worthy of note is that this broad region encompasses the area where the pluvial lakes were located during the last glaciation (Dawson 1992; see Fig. 6). This suggests the region had extensive riparian habitats and was likely a refugia for yellow warblers of the western lineage. In contrast, we do not have evidence that Beringia was a refugia. The Fairbanks population, located in the Beringia region, was genetically depauperate and was composed of haplotypes W02 and W08 only, i.e. the two most common haplotypes of the western lineage. Its mismatch distribution was more typical of a population that has experienced a population bottleneck. The time of expansion was recent $\tau = 0.65$ or 6300 ybp. The Beringia region was thus more likely recently colonized from southern areas after complete recession of the Cordilleran Ice Sheet in western Canada which occurred by 10 000 ybp (Dawson 1992).

Migration: a factor promoting gene flow

Helbig (2003) recently proposed that migration facilitates gene flow. He specifically predicted that migratory species will be more structured away from the migration axis than along the migration axis. For migratory wood-warblers, the migration axis is north–south. Our results show support for Helbig's (2003) hypothesis. In both lineages but especially in the eastern linage, gene flow was restricted longitudinally (i.e. along an east–west axis) but not latitudinally (i.e. along the north–south axis). The latitudinal gene flow axis parallels the migration axis. Latitudinal gene flow may be facilitated by the general wind patterns occurring in North American and by strong flows of maritime tropical air masses in spring. Migration may thus facilitate range expansion and this may have happened during colonization of lands after ice sheet retreat (Gauthreaux 1980). In such conditions, birds may overshoot their natal breeding area and even the northern limit of the species breeding range. The observation that gene flow is longitudinally restricted in yellow warblers has broader consequences for avian speciation. Migratory behaviour is often seen as a homogenizing agent preventing intraspecific differentiation and speciation (Helbig 2003). However, if dispersal is more pronounced along the migratory axis this may lead to a longitudinal differentiation in populations and the development of parallel migration systems (Salomonsen 1955). Several migratory species show such longitudinal differentiation in North America, but the Pleistocene glaciation events also played a role in shaping this east–west differentiation (Smith *et al*. 2005). Further studies in other migratory species are needed to verify the generality of Helbig's hypothesis but we predict the longitudinal restricted gene flow would be present in several other migrant species, especially in species that depend on wind patterns to accomplish their spring migration.

Acknowledgements

We thank the following people and institutions for dedicated help with the collection of samples: K. A. Hobson, R. Ludkin, L. Mousseau, J. Smith, J. Miles, E. Paxton, K. B. Burk, K. Voget, J. Woods, A.-M. Barber, T. Pogson, A. Solomon, Point Reyes Bird Observatory (R. Geupel and M. Flannery), S. L. Jones, S. H. Craig, D. A. Cohrs, C. Pitman, J. Carlisle, H. den Haan, S. Yezerinac, R. Keith, D. Flath, G. Taylor, D. L. Swanson, J. MacKay, E. M. Ammon, J. Lytle, R. Dawson, E. Milot, N. Wheelwright, R. P. Yunick, D. M. Finch, J. Kelly, S. Cox, H. Walker, E. W. Brooks, D. Bonter, B. Milá, S. S. Mitra, I. Paterson, M. C. Shieldcastle, A. Chartier, B. Hall-Brooks, D. P. Vroman, R. C. Leberman, R. S. Mulvihill, G. D. Connel, J. Massey, F. Collins, J. Banta, J. Graham, K. Baer, R. C. Simpson, and any other unrecognized persons who may have plucked feathers for us. We offer our sincere gratitude to J. Hinshaw at the Bird Division, Museum of Zoology, University of Michigan, for her help with feather importation, Dr R. Zink at the Bell Museum, University of Minnesota, and Dr T. Smith, Center for Tropical Research, UCLA, for providing substantial numbers of samples. We thank L. De Sousa and L. M. Tabak for help with laboratory work and S. Dudley for statistical assistance. K. A. Hobson, R. Robertson, S. Bensch, and three anonymous reviewers provided helpful comments on manuscripts. Project funding was provided by NSERC grants and a Premier's Research Award to H.L.G. and Max Bell Foundation and STAGE (Environment Canada) grants to H.L.G. and Keith Hobson. M.B. was supported by scholarships from NSERC, FQRNT, OGSST, and the Department of Biology at McMaster University.

References

Agresti A (1996) *An Introduction to Categorical Data Analysis*. John Wiley & Sons, New York.

Bohonak AJ (2002) IBD (Isolation by Distance): a program for the analyses of isolation by distance. *Journal of Heredity*, **93**, 153–154.

Boulet M (2004) *Evolutionary history and migration patterns in the yellow warbler* (Dendroica petechia). PhD Thesis. Department of Biology, McMaster University, Hamilton, Ontario.

Browning MR (1994) A taxonomic review of *Dendroica petechia* (Yellow Warbler) (Aves: parulinae). *Proceedings of the Biological Society of Washington*, **107**, 27–51.

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

Dawson AG (1992) *Ice Age Earth*. Routledge, London, UK.

Dyke AS, Prest VK (1987) Late Wisconsin and Holocene history of the Laurentide Ice Sheet. *Géographie Physique et Quaternaire*, **41**, 237–263.

Edwards SV, Beerli P (2000) Gene divergence, population divergence, and the vicariance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

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

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

Gauthreaux SA Jr (1980) The influences of long-term and short-term climatic changes on the dispersal and migration of organisms. In: *Animal Migration, Orientation, and Navigation* (ed. Gauthreaux SA Jr), pp. 103–174. Academic Press, New York.

Goosen JP, Sealy SG (1982) Production of young in a dense nesting population of yellow warblers, *Dendroica petechia*, in Manitoba. *Canadian Field Naturalist*, **96**, 189–199.

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

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

Helbig AJ (2003) Evolution of bird migration: a phylogenetic and biogeographic perspective. In: *Avian Migration* (eds Berthold P, Gwinner E, Sonnonschein E), pp. 3–20. Springer, Berlin, Germany.

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

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

Hoffmann MH, Glaß AS, Tomiuk J, Schmuths H, Fritsch R, Bachmann K (2003) Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassiceae) with geographic information systems (GIS). *Molecular Ecology*, **12**, 1007–1019.

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

Hosmer DW, Lemeshow S (1989) *Applied Logistic Regression*. John Wiley & Sons, New York.

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

Johnston K, Ver Hoef JM, Krivoruchko K, Lucas N (2001) *Using ARCGIS Geostatistical Analyst*. ESRI, Redlands, California.

Manel S, Schwartz M, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mengel RM (1964) The probable history of species formation in some northern wood warblers (Parulidae). *Living Birds*, **3**, 9–43.

Mengel RM (1970) The North American Central Plains as an isolating agent in bird speciation. In: *Pleistocene and Recent Environments of the Central Great Plains* (eds Dort W, Jones KJ Jr), pp. 280–340. University Press of Kansas, Lawrence, Kansas.

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

Nielsen R, Wakeley JW (2001) Distinguishing migration from isolation: a Markov Chain Monte approach. *Genetics*, **158**, 885–896.

Pagé P (1999) *Les Grandes Glaciations*, 2nd edn. Guérin, Montréal.

Pérez-Tris J, Bensch S, Carbonell R, Helbig AJ, Tellería JS (2004) Historical diversification of migration patterns in a passerine bird. *Evolution*, **58**, 1819–1832.

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

Posada D, Crandall KA (1998) MODELTEST: testing model of DNA substitution. *BioInformatics*, **14**, 817–818.

Ritchie MG, Kidd DM, Gleason JM (2001) Mitochondrial DNA variation and GIS analysis confirm a secondary origin of geographical variation in the bush cricket *Ephippiger ephippiger* (Orthoptera: Tettigonioidea), and resurrect two subspecies. *Molecular Ecology*, **10**, 603–611.

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.

Ronquist & Huelsenbeck (2003) MrBayes 3: Bayseian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Ruegg KC, Smith TB (2002) Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **269**, 1375–1381.

Salomonsen F (1955) The evolutionary significance of bird-migration. *Biologiske Meddelelser*, **22**, 1–62.

SAS (2002) *SAS Online Document*, Version 8e. SAS Institute Inc, Cary, North Carolina.

Sauer JR, Hines JE, Fallon J (2004) The North American Breeding Bird Survey, Results and Analysis 1966–2003. Version 2004.1. USGS Patuxent Wildlife Research Center. www.mbr-usgs.gov/bbs/bbs/bbs2003.html.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN version 2.000: A software for population genetic data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, Switzerland.

Seutin G, White BW, Boag PT (1991) Preservation of avian blood

and tissue samples for DNA analyses. *Canadian Journal of Zoology*, **69**, 82–90.
Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
Smith TB, Clegg SM, Kimura M, Ruegg KC, Milá B, Lovette I (2005) Molecular genetic approaches to linking breeding and overwintering areas in five Neotropical migrant passerines. In: *Birds of the Two Worlds, The Ecology and Evolution of Migration* (eds Greenberg R, Marra PP), pp. 222–234. Johns Hopkins University Press, Baltimore, Maryland.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Williams TC, Webb T III (1996) Neotropical bird migration during the Ice Ages: orientation and ecology. *Auk*, **113**, 105–118.
Zink RM (1997) Phylogeographic studies of North American birds. In: *Avian Molecular Evolution and Systematics* (ed. Mindell DP), pp. 301–324. Academic Press, San Diego, California.

M. Boulet is a postdoctoral fellow in the Département de Biologie at the Université Laval working on the genetics of sea migration of brook charr. Her area of expertise includes conservation of Neotropical birds, evolution and genetic control of migration. She did her PhD thesis on migratory behaviour and migration patterns in the yellow warbler at McMaster University with H. L. Gibbs. H. L. Gibbs is an Associate Professor in the Department of Evolution, Ecology, and Organismal Biology at Ohio State University. His research focuses on the evolutionary and conservation biology of snakes and birds.

Appendix I

List of the sampling sites included in this study and their geographic coordinates in degree decimal. *Nseq*, number of individuals that were sequenced; *Nrs*, number of individuals that were genotyped using region-specific PCR amplifications; and *Ntotal* is the sum of *Nseq* + *Nrs*. The identification number (ID) refers to numbers below pie-charts on Fig. 4. The asterisk sign (*) identifies sites included in Milot *et al*. (2000)

Sampling sites	State or province, Country	*Nseq*	*Nrs*	*Ntotal*	Latitude	Longitude
Northern yellow warblers						
Bon Portage Island (1)	Nova Scotia, Canada	0	10	10	43.50	−65.70
Braddock Bay (2)	New York, USA	5	5	10	43.31	–77.71
Burnaby area (3)	British Columbia, Canada	6	7	13	49.25	–122.70
Cape Breton (4)	Nova Scotia, Canada	5	3	8	46.17	−60.75
Churchill (5)	Manitoba, Canada	22	2	24	58.73	−94.12
*Creston (6)	British Columbia, Canada	26	0	26	49.10	–116.50
Cuyahoga (7)	Ohio, USA	3	13	16	41.24	−81.55
Deep Creek (8)	Wyoming, USA	5	2	7	42.72	–109.01
*Delta Marsh (9)	Manitoba, Canada	27	0	27	50.20	−98.20
Dodge Nature Centre (10)	Minnesota, USA	2	0	2	44.98	−93.26
El Paso County (11)	Colorado, USA	1	0	1	38.83	–104.52
*Fairbanks (12)	Alaska, USA	11	0	11	64.84	–147.72
Flagstaff (13)	Arizona, USA	6	0	6	35.20	–111.65
Fletcher Lake (14)	British Columbia, Canada	8	0	8	51.67	–123.00
Fort Riley (15)	Kansas, USA	9	14	23	39.17	−96.58
*Germantown (16)	New Brunswick, Canada	11	0	11	45.68	−64.80
Great Falls (17)	Montana, USA	3	15	18	47.41	–111.37
Great Swamp (18)	Rhode Island, USA	5	14	19	41.50	–71.62
*Gros Morne (19)	Newfoundland, Canada	31	0	31	49.69	−57.74
Haldimand (20)	Ontario, Canada	0	5	5	42.81	–79.96
Holiday Beach (21)	Ontario, Canada	0	1	1	42.10	−83.12
IBSP Ocean county (22)	New Jersey, USA	3	0	3	39.93	–74.08
Jasper (23)	Alberta, Canada	7	2	9	52.98	–118.10
KLAM (24)	California, USA	0	5	5	41.99	–121.74
Lake La Biche (25)	Alberta, Canada	13	5	18	54.83	–112.05
Lake Mead (26)	Nevada, USA	6	0	6	36.02	–114.69
Last Mountain (27)	Saskatchewan, Canada	29	1	30	51.08	–105.23
Lucky Peak (28)	Idaho, USA	5	0	5	43.53	–116.05
Mary's River (29)	Nevada, USA	3	4	7	41.05	–115.28
Meanock (30)	Alberta, Canada	0	5	5	54.57	–113.33
Modoc (31)	California, USA	3	5	8	41.46	–120.52
Mono County (32)	California, USA	3	11	14	37.92	–118.87
Moose Park (33)	Saskatchewan, Canada	16	16	32	49.81	–102.42
Mount Baker (34)	Washington, USA	4	3	7	48.50	–121.50
Oak Harbor (35)	Ohio, USA	4	5	9	41.62	−83.22
Ponca State Park (36)	Nebraska, USA	7	19	26	42.60	−96.73
Powdermill NR (37)	Pennsylvania, USA	5	13	18	40.20	–79.24
*Queen's Biol. Station (38)	Ontario, Canada	21	0	21	44.58	–76.32
RFSL (39)	California, USA	0	3	3	41.77	–124.05
Ruby Lake (40)	Nevada, USA	3	8	11	40.00	–115.33
Sacramento (41)	California, USA	3	5	8	40.54	–121.52
South Oregon (42)	Oregon, USA	6	16	22	42.43	–123.44
Stevensville (43)	Montana, USA	3	6	9	46.52	–114.05
Suffolk county (44)	New York, USA	0	2	2	40.73	–73.19
Tatlayoko (45)	British Columbia, Canada	2	0	2	51.55	–125.42
Terra Nova Nat. Park (46)	Newfoundland, Canada	0	2	2	48.53	−53.93
*Trois-Rivières (47)	Québec, Canada	28	0	28	48.30	−88.93
Vermilion (48)	Michigan, USA	3	20	23	46.76	−85.15
Vernal (49)	Utah, USA	7	5	12	40.11	–109.65
Vicksburg (50)	Michigan, USA	3	1	4	42.12	−85.53
Walden (51)	Colorado, USA	5	23	28	40.50	–106.17
Washoe county (52)	Nevada, USA	2	0	2	40.75	–119.63
Waubay (53)	South Dakota, USA	3	0	3	45.43	−97.34
Wenatchee (54)	Washington, USA	3	10	13	47.42	–120.31
Whitehorse (55)	Yukon, Canada	1	0	1	60.72	–135.05
Yosemite (56)	California, USA	3	7	10	37.85	–119.57
Total		390	294	684		
Golden warbler						
Cozumel Island	Quintana Roo, Mexico	3	0	3	20.51	
Puerto Rico	Puerto Rico, USA	1	0	1	?	?
Total		4	0	4		

Appendix I *Continued*

Sampling sites	State or province, Country	*Nseq*	*Nrs*	*Ntotal*	Latitude	Longitude
Mangrove warbler						
Isla del Carmen	Campeche, Mexico	1	0	1	18.70	−91.65
Isla la Peita	Nayarit, Mexico	2	0	2	?	?
Paraiso	Tabasco, Mexico	1	0	1	18.45	−93.23
Rancho Los Hermanos	Campeche, Mexico	1	0	1	?	?
Teapa	Tabasco, Mexico	1	0	1	17.55	−92.95
Costa Rica	Costa Rica	1	0	1	?	?
Venezuela	Venezuela	1	0	1	?	?
Total		8	0	8		

Appendix II

Summary of the 19 eastern group sites and the 20 western group sites used in genetic analyses based on sequence data. *N*, number of sequences. Admixed sites were split into subgroups of unique lineages

Lineage groups	*Nseq*	Corresponding sites in Appendix I
Eastern lineage		
Braddock Bay	5	Same site
Cape Breton	5	Same site
Churchill East	10	Churchill (eastern haplotypes only)
Delta Marsh East	26	Delta Marsh (eastern haplotypes only)
Dodge Nature	2	Same site
East Coast	8	Great Swamp (5), ISBP Ocean county (3)
Fort Riley	9	Same site
Germantown	11	Same site
Gros Morne Nat. Park	31	Same site
Lake La Biche East	3	Lake La Biche (eastern haplotypes only)
Last Mountain East	17	Last Mountain (eastern haplotypes only)
Moose Park East	7	Moose Park (eastern haplotypes only)
Ponca State Park East	6	Same site (eastern haplotypes only)
Powdermill NR	5	Same site
Queen's Biol. Station	21	Same site
South Great Lakes	10	Cuyahoga (3), Oak Harbor (4), Vicksburg (3)
Trois-Rivières	28	Same site
Vermilion	3	Same site
Waubay	3	Same site
Western lineage		
Burnaby area	6	Same site
Central BC	10	Fletcher Lake (8), Tatlayoko Lake (2)
Central CA	6	Mono county (3), Yosemite (3)
Churchill West	12	Churchill (western haplotypes only)
Colorado West	5	El Paso county (1), Walden (4)
Creston	26	Same site
Deep Creek West	4	Deep Creek (western haplotypes only)
Fairbanks	11	Same site
Flagstaff West	3	Flagstaff (western haplotypes only)
Jasper	7	Same site
Lake La Biche West	10	Lake La Biche (western haplotypes only)
Last Mountain West	9	Last Mountain (western haplotypes only)
Lucky Peak	5	Same site
Montana	6	Stevensville (3), Great Falls (3)
Moose Park West	9	Moose Park (western haplotypes only)
Nevada	6	Mary's River (3), Ruby Lake (3)
North CA/NV	8	Modoc (3), Sacramento (3), Washoe county (2)
South Oregon	6	Same site
Vernal West	6	Vernal (western haplotypes only)
Washington	7	Mount Baker (4), Wenatchee (3)

Appendix III

Variable sites in the 124 yellow warbler haplotypes. Sequences are compared to haplotype E01. Lineages are indicated in parentheses and dots show identical sites

	11111111	1111111111	1111222222	2222222222	222222233
	22666679	9900011111	2223335677	8899000111	2222223333	333445712
	2624012717	8924523569	0140596014	3846346047	0245670123	456453037
E01 (East)	CTATTTCTGG	CTATTGTCGC	CTGCCAACAC	ATAGTACTCC	TACGTCCCAA	GCG-AACGA
E60 (East)	....	....	....	....	...A......	....
E61 (East)	c.G....C..	....	....	.......C..	....	....
E62 (East)	....	....	....	....	...C......	....
E63 (East)	..G.......	....	....	....	....	......T..
E64 (East)	..G.......	....	....	....	...A......	....
E65 (East)	....	....	....T.....	..G.......	....	....
E66 (East)	.......C..	....	....T.....	....	....	......T..
E67 (East)	T.........	....	....	.........T	C.........	.T.......
E68 (East)	....	....	....	......A...	....	....
E69 (East)	....	....	....	....	........G.	....
E70 (East)	....	....	....	..G.......	....C.....	....
E71 (East)	...C...C..	T.........	....	....	....	....
E72 (East)	.......C..	....	.......T..	..G......T	...C......	....
E73 (East)	.C........	....	....	....	....	....
E81 (East)	..G.......	....	....	.......C..	....	....
W01 (West)	T.........	.....AC.T.	T.........	....C....T	C......T..	C.A......
W02 (West)	T.........	.....AC.T.	....	.........T	C......T..	C.A......
W03 (West)	T.........	.....AC.T.	T.........	.........T	C......T..	C.A..G...
W04 (West)	T......C..	.....AC.T.	....	.........T	C......T..	C.A......
W05 (West)	T.........	.....AC.T.	....	.........T	.......T..	C.A......
W06 (West)	T.........	.....AC.T.	T.........	....C....T	C.........	C.A......
W07 (West)	T.........	.....AC.T.	....	.........T	C..A...T..	C.A......
W08 (West)	T.........	.....AC.T.	T.........	.........T	C......T..	C.A......
W09 (West)	T.........	.....AC.T.	....	.........T	C..A......	C.A......
W10 (West)	T.........	.....AC.T.	....	.........T	C......T..	C.A....T.
W11 (West)	T.........	.....AC.T.	....	.........T	C......T..	C.A...T..
W12 (West)	T.........	.....AC.T.	T........T	.........T	C......T..	C.A......
W13 (West)	T.........	.....AC.T.	.........T	.........T	C......T..	C.A......
W14 (West)	T.........	.....AC.T.	....	.........T	C.TA...T..	C.A......
W15 (West)	T.........	.....ACTT.	.C........	.........T	C......T..	C.A......
W16 (West)	T.........	.....AC.T.	....	.........T	C.........	C.A......
W17 (West)	T.........	.....AC.T.	.........T	.........T	C.........	C.A......
W18 (West)	T.........	.....AC.T.	T.........	.........T	C......TT.	C.A......
W19 (West)	T.........	.....AC.T.	....	.........T	C......T..	C.A..G...
W20 (West)	T.........	.....AC.T.	.........T	.........T	C..A...T..	C.A......
W21 (West)	T.........	.....AC.T.	....	.........T	C.......T.	C.A......
W22 (West)	T......C..	.....AC.T.	T.........	.........T	C......T..	C.A......
W23 (West)	T.........	.C...AC.T.	....	.........T	C......T..	C.A......
W24 (West)	T......C..	.C...AC.T.	....	.........T	C......T..	C.A......
W26 (West)	T.........	.....AC.T.	T.........	.........T	C.TA...T..	C.A......
W29 (West)	t.........	..T..AC.T.	.........T	.........T	C......T..	C.A......
W30 (West)	t.........	.....ACTC.	....	.........T	C......T..	C.A..G...
W31 (West)	t......C..	.....AC.T.	....	.........T	C......T..	C.A...T..
W32 (West)	t......C..	.....ACTC.	....	.........T	C......T..	C.A..G...
W33 (West)	T......C..	.....AC.T.	.C........	.........T	C......T..	C.A......
W35 (West)	T.........	.....AC.T.	....	.C.......T	C......T..	C.A...T..
W36 (West)	T.........	.....AC.T.	.C........	.........T	C.........	C.A......
W37 (West)	T.........	.....AC.T.	........G.	.........T	C......T..	C.A......
W38 (West)	T......C..	.....AC.T.	....	.......C.T	C......T..	C.A......
W39 (West)	T.........	.....AC.T.	....	......G..T	C......T..	C.A......
W40 (West)	T......C..	.....A..T.	.C........	.........T	C......T..	C.A......

Appendix III *Continued*

	11111111 1111111111 1111222222 2222222222 222222233 22666679 9900011111 2223335677 8899000111 2222223333 333445712 2624012717 8924523569 0140596014 3846346047 0245670123 456453037
W41 (West)	T......C.. .....AC.T. ....T C..A...T.. C.A......
W42 (West)	T......... .....AC.T. .C........ ....T C......T.. C.A......
W43 (West)	T......... .....AC.T. ....T CG.....T.. C.A......
W44 (West)	T......... .....AC.T. T........T ....TT C......T.. C.A......
W46 (West)	T......C.. .....AC.T. ....T C..A...... C.A......
W48 (West)	T......... .....AC.T. T......... ....T C..A...... C.A......
W49 (West)	T......C.. .....AC.T. ....T C.TA...T.. C.A......
S01 (South)	....CC..AA ...C.A..A. ..AT...... ..G.A....T C..AC...T. AT.......
S02 (South)	.....C..AA ...C.A..A. ..AT...... ..GTA....T C..AC...T. AT.......
S03 (South)	....CC..AA ...C.A..A. ..AT...... ..G.A....T C..AC...TG AT.......
S04 (South)	....CC..AA ...C.A..A. ..AT..G... ..G.A....T C..AC...T. AT.......
S05 (Resid)	....CC..AA .....A..A. ..ATT..... ..G.A....T C..AC.T.CG AT.......
S06 (Resid)	....CC..AA .....AC.AT ..AT...... ....AC...T ...ACT..TG AT.......
S07 (Resid)	....CC..AA .....A..A. ..AT...... ....AC...T ...ACT..TG AT.......
S08 (Resid)	....CC..AA .....AC.A. ...T.....T ....AC...T ...ACT..TG AT.......
S09 (Resid)	...CCC..AA ...C.A..A. ...T.G.... .CG.A....T C.TAC...TG AT.......
S10 (Resid)	....CC..AA ...C.A..A. ..AT...... ....A....T C..AC...TG AT.......
S11 (Resid)	....CCT..A .....AC.A. ..ATT..... ....AC...T ...AC.T.TG AT.......

Appendix IV

Minimum spanning network showing the relationships between northern yellow warbler haplotypes within: A) eastern lineage, B), western lineage, C) southern lineage. Each line connecting circles or squares indicates one mutation step between haplotypes. Small open circles refer to haplotypes not observed in the yellow warbler individuals analysed but that may exist in the wild. Arrows point haplotypes found in New Brunswick.