Molecular Ecology (2001) **10**, 371–386

Colonization, extinction, and phylogeographic patterning in a freshwater crustacean

ANDREA J. COX and PAUL D. N. HEBERT
Department of Zoology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Abstract

Phylogeographic analyses have revealed the importance of Pleistocene vicariance events in shaping the distribution of genetic diversity in freshwater fishes. However, few studies have examined the patterning of variation in freshwater organisms with differing dispersal syndromes and life histories. The present investigation addresses this gap, examining the phylogeography of *Sida crystallina*, a species whose production of diapausing eggs capable of passive dispersal was thought to constrain its regional genetic differentiation. By contrast, the present analysis has revealed deep allozyme and cytochrome oxidase I mitochondrial DNA divergence between populations from North America and Europe. Moreover, North American populations are separated into four allopatric assemblages, whose distribution suggests their derivation from different Pleistocene refugia. These lineages show higher haplotype diversity and deeper sequence divergence than those of any fish from temperate North America. Its distinctive life history traits have evidently sheltered lineages of *Sida* from extinction, contributing to a remarkably comprehensive and high resolution phylogeographic record.

Keywords: allozymes, Cladocera, dispersal, glacial refugia, mtDNA, phylogeography

Received 13 February 2000; revision received 18 July 2000; accepted 7 September 2000

Introduction

The biota of North America was profoundly affected by the recurrent glacial advances and retreats during the Pleistocene. Life in formerly glaciated regions consists of immigrant taxa, derived from either Beringia or ice-free areas in the temperate zone, while life in the south is comprised of assemblages exposed to range adjustments as species tracked shifting climate regimes (Hewitt 1996, 2000). Until recently, efforts to reconstruct histories of habitat occupancy and rates of range extension were conjectural, based on the inspection of modern distributions and the identification of morphological discontinuities among populations (Dadswell 1974; Hocutt & Wiley 1986; Stemberger 1995). However, phylogeographic studies are adding new rigor to the field by enabling a critical evaluation of refugial origins and histories of habitat tenancy (Avise *et al*. 1987, 1998). Such investigations are providing novel insights concerning the interactions of ecology, life history, and abiotic factors in shaping the evolution and distribution of extant populations on both local and continental scales (Bermingham & Moritz 1998; Avise 2000).

The life of inland waters has been the target for a considerable amount of phylogeographic study (Bermingham & Avise 1986; Bernatchez & Wilson 1998). Most work has focused on fishes — active dispersers capable of migration against currents within watershed boundaries. These phylogeographic investigations have provided some generalizations, revealing, for example, that most north temperate species include a number of refugial lineages (Bernatchez & Dodson 1991; Wilson & Hebert 1998). These studies have also shown that the lineages of fishes which are largely restricted to glaciated regions are young, suggesting the vigorous pruning of diversity during the Pleistocene. By contrast, species in the unglaciated regions of North America include much older lineages (Billington & Hebert 1991; Bernatchez & Wilson 1998).

Although work on fishes has now been comprehensive enough to provide a good sense of their phylogeographic

Correspondence: Andrea J. Cox. Fax: (519) 766 1656 E-mail: acox@uoguelph.ca

patterning, it is unlikely that these results are representative of freshwater life at large. The impact of dispersal strategy on phylogeographic structure merits particular attention. The most striking axis of variation in this regard relates to the reliance of some taxa, such as fishes, on active dispersal, while others, such as the plankton, employ passive dispersal. It has long been assumed that regional genetic variation would be much more limited in the latter groups because of their ability to cross watershed boundaries (Darwin 1859; Mayr 1963). Moreover, their production of diapausing stages which remain viable for hundreds of years (Weider *et al*. 1997) would seem to provide additional security against extinction, perhaps making these organisms less prone to local lineage loss. The present study had the primary goal of advancing knowledge of phylogeographic structure in passively dispersed life. First, it tests the expectation that regional genetic divergence is limited in such taxa. Second, it examines whether the patterning of genetic diversity in these organisms shows any correspondence that observed in lineages employing active dispersal.

Sida crystallina, the target for the present analysis, is a cyclically parthenogenetic cladoceran crustacean found in the littoral zone of lakes and ponds throughout the Holarctic. Its broad distribution has evidently been achieved through the dispersal of diapausing eggs which often adhere to substrates such as macrophytes (Fryer 1996), and are both freeze-tolerant and digestion resistant, providing an avenue for dispersal via migratory waterfowl (Korovchinsky & Boikova 1996). Eurasian and North American populations of *Sida* have been classified as distinct subspecies because of subtle morphological differences (Korovchinsky 1992), suggesting that intercontinental gene flow is limited. There is, however, no knowledge of the patterning or extent of gene flow among populations on each continent or their history of habitat tenancy. Because *S. crystallina* is eurythermal, as evidenced by its occurrence from the subtropics to low arctic of North America (Korovchinsky 1992), populations may have persisted in ice-free areas, despite shifting thermal regimes during the Pleistocene. Its prevalence in regions which were both ice-covered during the Pleistocene and distant from glacial refugia also indicates that many modern populations owe their origin to recent range extension.

The present paper examines the extent of genetic divergence between European and North American populations to ascertain the recency of contact between *Sida* populations from these areas. Also, a comprehensive analysis of genetic population structure in North American lineages is undertaken to evaluate their origins and the extent of gene flow across this continent. Finally, the patterning and extent of geographical variation in this species is compared with that in both freshwater fishes and life reliant on passive dispersal.

Materials and methods

Specimen collection

Populations of *Sida crystallina* s.l. were collected from 68 habitats in North America and three sites in Europe between June 1995 and September 1997 (Appendix I; Fig. 1). Based on their geographical origin, North American populations of *Sida* were assigned to one of seven freshwater biogeographic provinces according to Burr & Mayden (1992): Cascadia (CD), Great Lakes (GL), Hudson Bay (HB), Mississippi (MI), North Appalachian (NA), South-eastern (SE), and Yukon-Mackenzie (YM). Derived from surveys of fish assemblages, the boundaries of these provinces provide an ideal framework against which to examine biogeographic patterning in other freshwater life, as they reflect the historical structure of drainage systems on the continent (Hocutt & Wiley 1986). North American populations were assigned a two digit code (01–68) followed by a two letter code indicating their biogeographic province of origin, while European collections were identified as E1–E3, followed by a two letter country code. Specific details for each habitat locality including latitude and longitude are provided in Appendix I.

Samples were obtained using a 250-μm mesh net towed through littoral zone macrophytes, sorted in the field, and either flash-frozen in liquid nitrogen or preserved in 90% ethanol for subsequent molecular analyses. Representatives of each population were later examined morphologically for the diagnostic features which ordinarily differentiate Eurasian and North American forms (Korovchinsky 1992). Due to variation in sample size and preservation method, not all populations were used in both DNA and allozyme analyses (see Appendix I).

Allozyme electrophoresis

Levels of allozyme variation were examined in 54 populations. Wherever possible, 20–40 individuals were examined for variation at seven commonly polymorphic and well-resolved loci using standard cellulose acetate electrophoresis (Hebert & Beaton 1993). Enzymes screened included arginine phosphokinase (APK, EC 2.7.3.3), supernatant aspartate amino transferase (sAAT, EC 2.6.1.1), fumarate hydratase (FUM, EC 4.2.1.2), glucose-6-phosphate isomerase (GPI, EC 5.3.1.8), supernatant malate dehydrogenase (MDH, EC 1.1.1.37), mannose-6-phosphate isomerase (MPI, EC 5.3.1.8), and phosphoglucomutase (PGM, EC 5.4.2.2). Two individuals from Beech Lake (49 GL) were included as reference standards in each gel run. The dominant allele at each locus in the standard population was assigned a relative mobility (R_f) value of 1.0; the mobilities of all other alleles were scored with respect to this standard. Where sample sizes were sufficient, side-by-side

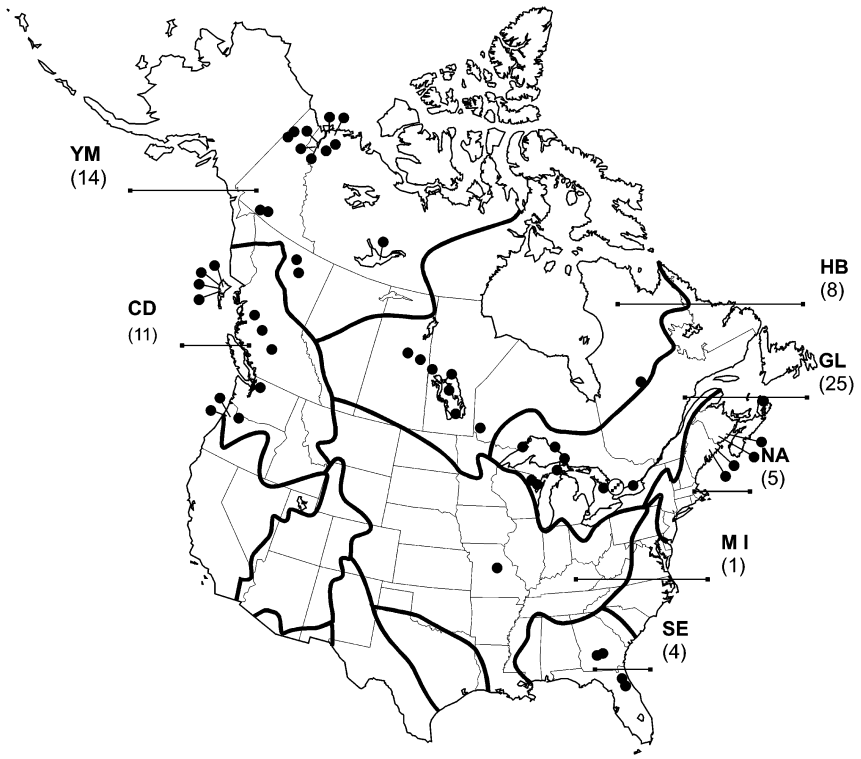

Fig. 1 Map of North America showing sampling locations for *Sida crystallina*. Biogeographic provinces (from Burr & Mayden 1992) sampled in this study are abbreviated as follows: Cascadia (CD), Great Lakes (GL), Hudson Bay (HB), Mississippi (MI), North Appalachian (NA), South-eastern (SE), and Yukon-Mackenzie (YM). Numbers in parentheses indicate the number of populations obtained from each region. The large, diagonally striped circle in the GL region overlies 20 habitats.

comparisons were made to confirm the presence of putative novel alleles.

Analysis of allozyme data and calculations of descriptive and hierarchical population statistics were carried out using Genetic Data Analysis (GDA) software (Lewis & Zaykin 1999). Genotypic frequencies were determined for each population, and polymorphic loci (0.99 criterion) were examined for deviations ($\alpha = 0.05$) from Hardy–Weinberg (HW) equilibrium using Fisher's exact test, followed by sequential Bonferonni corrections (Rice 1989). Nei's genetic distance (Nei 1978) was used to estimate a UPGMA phenogram to examine relationships among populations. Bootstrapping was subsequently performed over loci in order to test the significance of population clusters within the phenogram using the program Genetic Distance and Phylogenetic Analysis (DISPAN) (Ota 1993).

The extent of allele frequency divergence among populations was evaluated in GDA using a three level hierarchical analysis of Wright's $F_{ST.}$ The hierarchy was constructed such that divergence was measured between individuals in single populations (F_{IS}), between single populations within a biogeographic province (F_{SP}), and between biogeographic provinces across the entire geographical range (F_{PT}). Confidence intervals (C.I. 95%) were estimated for each level in the hierarchy by bootstrapping across loci (5000 replicates).

Mitochondrial DNA analysis

Total DNA was isolated from a single individual from each of 32 populations, and for 2–3 individuals from another 15 populations, using modified proteinase K methods described by Schwenk (1996). Individual animals were homogenized in 50 µL of H3 extraction buffer (10 mM Tris-HCl pH 8.3, 0.05 M potassium chloride, 0.005% Tween-20 and 0.005% NP-40; Replitherm Reaction Buffer, Biozym) and 20 µg of proteinase K, and incubated overnight in a 50 °C oven. Following denaturation of proteinase K *via* incubation for 12 min in a 94 °C waterbath, extracts were stored at −20 °C. The polymerase chain reaction (PCR) (Saiki *et al*. 1988) was subsequently utilized to amplify a 680-bp fragment of the mitochondrial (mt) cytochrome *c* oxidase subunit I (COI) gene with the primer pair LCOI490 and HCO2918 (Folmer *et al*. 1994). Each 50-µL PCR reaction consisted of 5 µL of genomic DNA template, 9 mM Tris-HCl (pH 8.3), 45 mM KCl, 2.2 mM $MgCl_2$, 0.26 µM of each dNTP (C,G,A,T), 0.36 µM of each primer, and 1.0 U of *Taq* DNA polymerase (Qiagen Inc). The thermal regime for amplification consisted of one cycle of 1 min at 94 °C; 40 cycles of 1 min at 94 °C, 1.5 min at 45 °C, and 1.5 min at 72 °C; followed by one cycle of 5 min at 72 °C. The product was gel purified using the Qiaex II kit (Qiagen Inc), subjected to dye terminator cycle-sequencing reactions (25 cycles, 55 °C annealing) with

primer LCOI490 and AmpliTaq® DNA polymerase FS (Perkin Elmer), and sequenced on an ABI 377 automated sequencer (Applied Biosystems).

Electropherograms were aligned with the SeqApp 1.9a sequence editor (Gilbert 1992), and nucleotide compositions were calculated in MEGA 1.02 (Kumar *et al*. 1993). Preliminary phylogenetic analyses utilizing the sidid *Diaphanosoma* as an outgroup revealed that European *Sida* are a divergent sister group to all North American populations. Accordingly, European populations of *Sida* were used as an outgroup for all subsequent analyses. As comparisons between sequences for the multiple individuals in 15 populations revealed that, in all cases, these individuals showed very limited sequence divergence (mean divergence = 0.1%; maximum divergence = 0.5%) relative to divergences among populations (4–7%), all subsequent phenetic and cladistic analyses were carried out using one individual per population. A distance matrix of pairwise nucleotide sequence divergence was calculated using the Kimura 2-parameter model (Kimura 1980), and used to estimate a neighbour-joining (NJ) phenogram with confidence limits determined in MEGA (1000 bootstrap pseudoreplicates). Maximum parsimony (MP) heuristic searches were conducted in PAUP* 4.0b2 (Swofford 1998), with steepest descent and tree bisection reconnection options invoked. Transitions and transversions were weighted equally as differential weightings did not affect the topology of the phylogeny. Confidence in the cladistic analyses was also assessed in PAUP*, both a priori, by estimation of the g_1 skewness statistic from 100 000 random tree length distributions (Hillis & Huelsenbeck 1992), and a posteriori, by bootstrap analysis with 1000 pseudoreplicates.

Estimated rates of sequence divergence for COI range from 1.4 to 2.6% per million years (Knowlton *et al*. 1993; Knowlton & Weight 1998; Schubart *et al*. 1998). In addition to the inherent problems in all rate estimates (Hillis & Moritz 1990; Martin & Palumbi 1993), no currently available calibrations for COI are based on the precise fragment amplified in the present study. Therefore, we chose an intermediate clock rate of 2% per million years for COI divergence time estimates, which is comparable to the arthropod mtDNA clock described by Brower (1994), as well as previous work by Brown *et al*. (1979).

Results

Species identification

All individuals from North America were unambiguously assigned to the subspecies *Sida crystallina americana*, while European populations were identified as *S. c. crystallina*. Males were only detected in collections from the Yukon-Mackenzie biogeographic province, and in one late autumn collection from the Great Lakes province.

Nuclear genetic variation

Allozyme variation was detected at five of the seven loci (Appendix II). The percentage of polymorphic loci in North American populations ranged from 0 to 71% with a mean of 31% (Table 1). An average of 1.4 alleles per locus were detected, while individual heterozygosities ranged from 0 to 22%, with seven populations monomorphic for all loci. Mean heterozygosities were highest in the Great Lakes biogeographic province, moderate in Hudson Bay, Yukon-Mackenzie, and Cascade regions, and low in the North Appalachian and South-eastern provinces (Table 1). The single European population examined for allozyme variation showed 3% heterozygosity and 14% polymorphic loci, with fixed allelic differences at two loci (*sAAT* and *PGM*) between it and all North American populations (Appendix II). Most populations were in HW equilibrium (Table 1); only a single significant deviation, linked to heterozygote deficit, was revealed when the data were corrected for multiple comparisons with sequential Bonferonni tests ($k = 54$, $P < 0.001$) (Rice 1989).

A UPGMA dendrogram of Nei's genetic distance revealed marked allozyme divergence between North American and European ($D \cong 0.70$) populations. Most of the North American populations showed limited genetic divergence ($D < 0.1$), with similar allelic arrays in populations ranging from the Yukon-Mackenzie, Hudson Bay, North Appalachian, and South-eastern biogeographic provinces (Fig. 2). A few populations, such as those in Cascadia and some from the Hudson's Bay region, showed greater divergence, but bootstrap support (not shown) for these groups of populations was weak (<45%).

F-statistics indicated substantial differentiation at all levels of the hierarchical analyses (Table 2), with greater divergence among populations within single biogeographic provinces ($F_{SP} = 0.48$) than between provinces ($F_{PT} = 0.31$). This high F_{SP} value reflects local populations fixed for alternate alleles (MPI), and the presence of otherwise rare alleles (GPI, PGM) at high frequencies in a few populations.

mtDNA sequence variation

The 614 bp alignment of 47 COI sequences was unambiguous as no gaps were present. Forty North American and three European haplotypes (GenBank accession numbers AF277849–AF277891) were detected with 169 variable sites, 157 of which were phylogenetically informative. NJ analyses indicated the presence of four phylogeographic assemblages: Atlantic, Pacific, Mississippian, and

Table 1 Allozyme diversity based on a survey of seven loci in North American populations of *Sida crystallina* s.l. Data are grouped by biogeographic province: Cascadia (CD); Great Lakes (GL); Hudson Bay (HB); Northern Appalachian (NA); South-eastern (SE); Yukon-Mackenzie (YM). *n*, mean number of individuals analysed per locus; P, percentage polymorphic loci (0.99 criterion); A_p, mean number of alleles per polymorphic locus; H_O, observed heterozygosity; H_E, expected heterozygosity

Region	*n*	P	Ap	H_O	H_E	Region	*n*	P	Ap	H_O	H_E
CASCADIA						HUDSON BAY					
02 CD	20.9	29	2.00	0.08	0.12	26 HB	38.9	29	2.50	0.06	0.08
04 CD	29.9	43	2.67	0.20	0.21	27 HB	25.0	14	2.00	0.08	0.07
05 CD	9.9	14	2.00	0.06	0.05	28 HB	20.0	14	2.00	0.03	0.03
06 CD	19.0	43	2.00	0.09	0.09	29 HB	23.0	29	2.00	0.09	0.07
07 CD	19.9	29	2.50	0.12	0.12	30 HB	20.0	43	2.00	0.06	0.08
08 CD	20.0	43	2.00	0.11	0.12	31 HB	19.9	29	2.50	0.07	0.07
09 CD	20.0	0	—	0.00	0.00	32 HB	39.6	43	2.00	0.11	0.12
10 CD	19.4	43	2.00	0.09	0.10	33 HB	19.0	14	2.00	0.02	0.02
mean	**19.9**	**30.5**	**2.2**	**0.09**	**0.10**	**mean**	**25.7**	**26.9**	**2.1**	**0.06**	**0.07**
GREAT LAKES						NORTH APPALACHIAN					
36 GL	20.0	29	2.00	0.03	0.03	60 NA	19.7	29	2.00	0.05	0.05
39 GL	22.0	43	2.33	0.22	0.19	61 NA	15.0	0	—	0.00	0.00
40 GL	35.9	57	2.25	0.16	0.16	64 NA	54.0	0	—	0.01	0.01
41 GL	26.0	29	2.00	0.12	0.11	**mean**	**29.6**	**9.7**	**2.0**	**0.02**	**0.02**
42 GL	41.3	57	2.00	0.15	0.18	SOUTH-EASTERN					
43 GL	31.0	57	2.25	0.13	0.15	65 SE	22.0	0	—	0.00	0.00
44 GL	37.0	57	2.25	0.20	0.20	66 SE	20.0	0	—	0.00	0.00
45 GL	43.7	43	2.67	0.15	0.16	67 SE	15.0	14	2.00	0.07	0.05
46 GL	38.9	57	2.75	0.13	0.17	68 SE	22.0	0	—	0.00	0.00
47 GL	27.6	43	2.33	0.10	0.12	**mean**	**19.8**	**3.5**	**2.0**	**0.02**	**0.01**
48 GL	43.0	43	2.00	0.10	0.11	YUKON-MACKENZIE					
49 GL	36.9	43	2.33	0.12	0.13	12 YM	19.9	43	2.00	0.07	0.07
50 GL	18.9	43	2.00	0.09	0.10	13 YM*	39.3	71	2.60	0.07	0.10
51 GL	16.0	57	2.25	0.10	0.14	14 YM	28.0	14	2.00	0.04	0.04
52 GL	43.0	71	2.20	0.12	0.12	15 YM	10.0	14	2.00	0.07	0.07
53 GL	30.9	43	2.33	0.18	0.17	16 YM	20.0	0	—	0.00	0.00
54 GL	41.0	43	2.00	0.10	0.11	17 YM	40.0	14	2.00	0.04	0.06
55 GL	19.0	43	2.67	0.18	0.19	18 YM	20.0	14	2.00	0.07	0.07
56 GL	43.9	43	2.00	0.12	0.14	19 YM	30.0	43	2.00	0.06	0.06
57 GL	43.4	43	2.00	0.09	0.11	20 YM	20.0	14	2.00	0.04	0.03
mean	**33.0**	**47.2**	**2.2**	**0.13**	**0.14**	22 YM	29.0	0	—	0.00	0.00
						24 YM	28.9	14	3.00	0.07	0.08
						mean	**25.9**	**21.9**	**2.2**	**0.05**	**0.05**

*deviation from Hardy–Weinberg expectation after sequential Bonferonni tests (k = 54, $P < 0.001$) at the PGM locus.

Beringian, with strong bootstrap support for all four clusters (Fig. 3).

Due to the large number of sequences, it was not possible to complete a search for parsimonious cladograms of all taxa. Therefore, MP was carried out on three (MIS, PAC, ATL) to six (BER) divergent representatives from within each of the major haplotype clusters identified by NJ analyses, as well as one haplotype that did not fall into a major cluster (03 CD), and three haplotypes from Europe, for a total of 19 taxa with 147 parsimony informative sites. MP heuristic searches yielded 36 equally parsimonious trees of length 243 (C.I. = 0.77, retention index (R.I.) = 0.87). Both strong phylogenetic signal ($g_1 = -2.16$; $g_{crit} = -0.09$; $P < 0.01$) and high bootstrap values support the presence of four phylogenetically divergent clades of haplotypes in North America, which correspond to the four major haplotype clusters identified in the NJ analysis. A strict consensus of the 36 trees is shown in Fig. 4. The Atlantic clade was the sister group to the other three clades (Beringian, Mississippi, Pacific), which formed an unresolved trichotomy. Additional MP analyses conducted on 19 different taxa gave congruent results (data not shown), indicating that cladistic support for the major phylogeographic groups is unaffected by taxon choice.

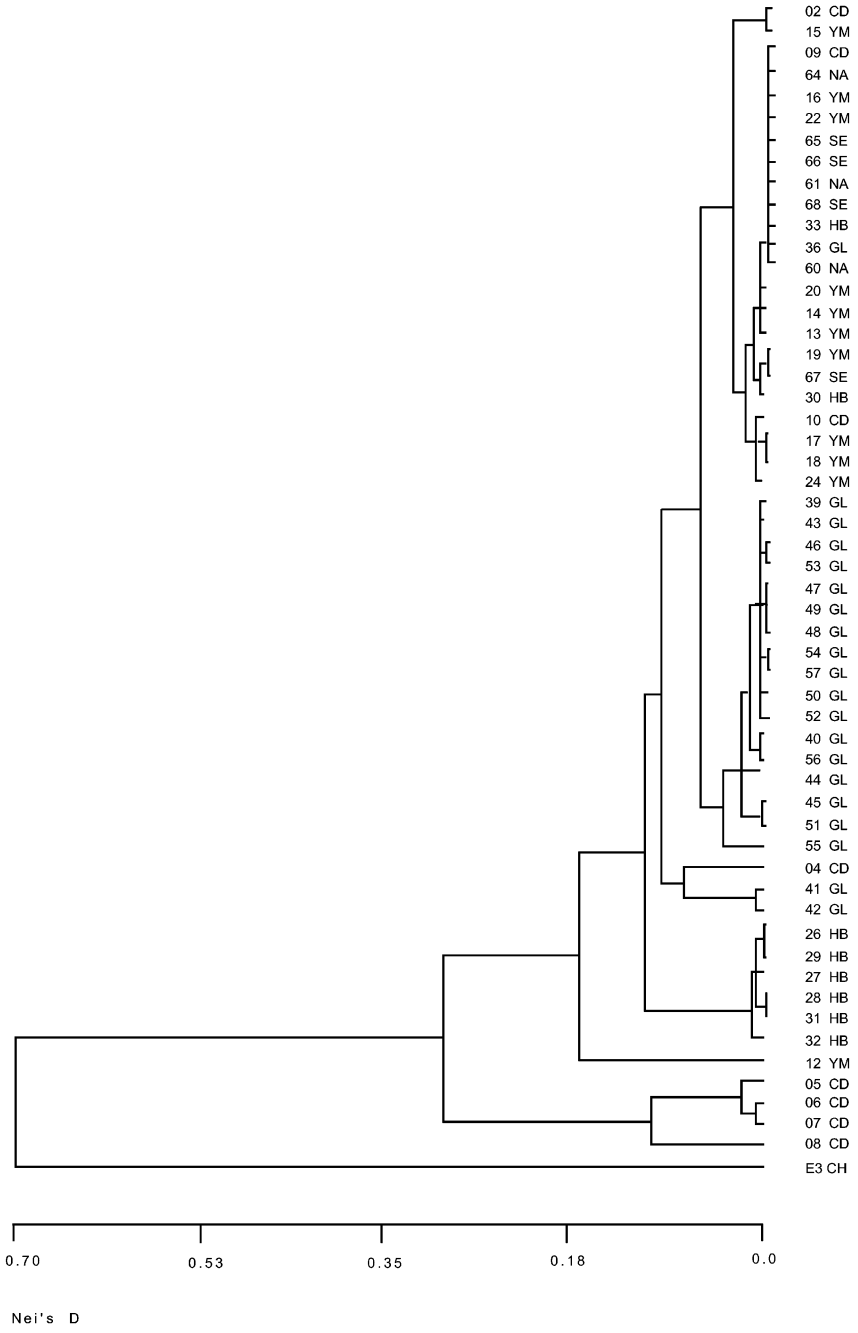

Fig. 2 UPGMA dendrogram of genetic distances (Nei 1978) among one European and 54 North American populations of *Sida crystallina*. Abbreviations for population locations are in Appendix I. Estimates of D are based on gene frequencies at five polymorphic allozyme loci. No clusters were supported by bootstrap values greater than 45%.

Locus	F_{IS}	F_{SP}	F_{PT}	F_{IT}
Mpi	0.05	0.52	0.28	0.55
Pgm	0.13	0.43	0.32	0.50
sAat	0.10	0.18	0.09	0.27
Mdh	0.16	0.52	0.11	0.60
Gpi	−0.03	0.64	0.46	0.63
Overall	0.08	0.49	0.31	0.53
C.I.	0.01–0.13	0.35–0.59	0.17–0.40	0.42–0.60

Table 2 *F*-statistics for three-level hierarchical analysis of five polymorphic allozyme loci among 54 North American *Sida crystallina* populations. F_{IS}, inbreeding coefficient for individuals within populations; F_{SP}, coancestry coefficient among populations within a region; F_{PT}, coancestry coefficient for different biogeographic provinces. C.I. = upper and lower bounds of 95% confidence intervals obtained by bootstrapping over loci (5000 replicates)

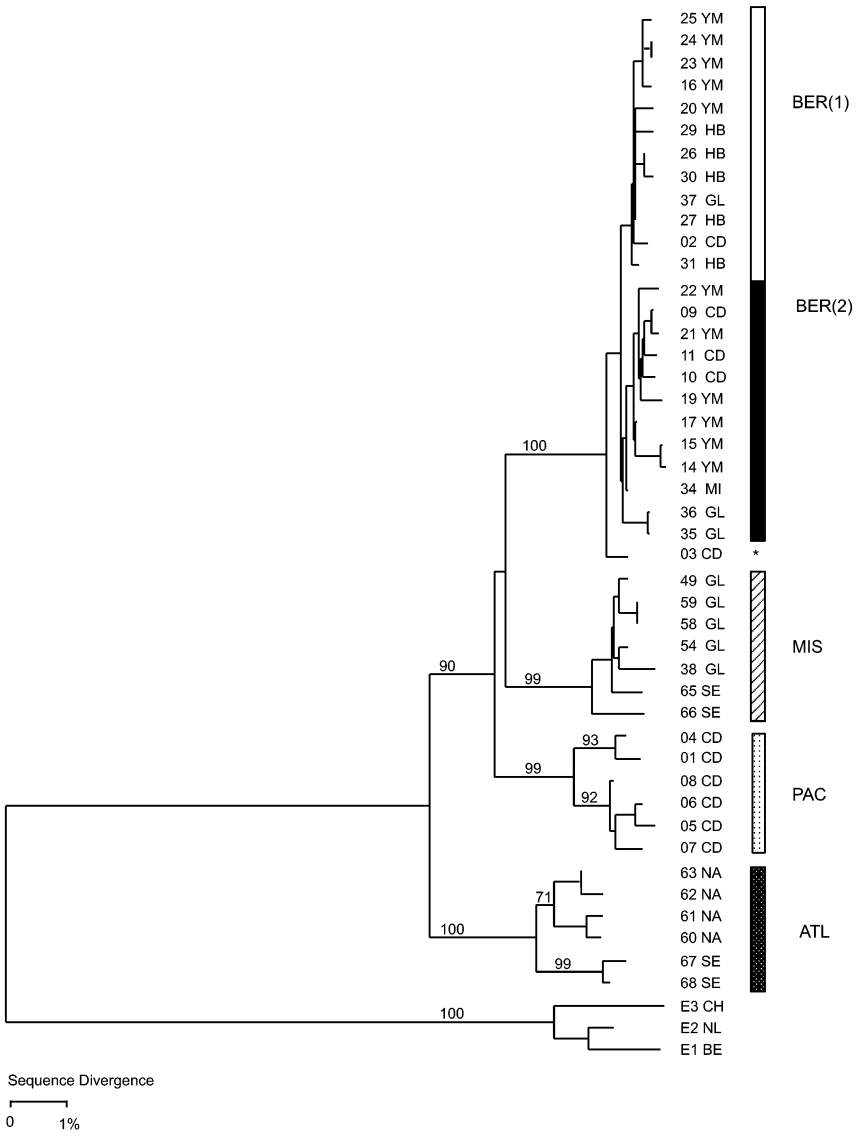

Fig. 3 Neighbour-joining (NJ) tree based on sequence variation in mitochondrial COI sequences (614 aligned sites) for a single isolate from each of 44 North American populations of *Sida crystallina americana*. Abbreviations for population locations are listed in Appendix I. The tree was rooted with three sequences obtained from European *S. c. crystallina*. Bootstrap confidence limits (1000 replicates) greater than 70% appear above the branches. Shaded bars correspond to the four clades mapped in Fig. 5: Beringia (BER), Mississippian (MIS), Pacific (PAC) and Atlantic (ATL). Two phylogeographic groups (BER1, BER2) are shown for the Beringian clade. Haplotype 03OR (*) was not assigned to a geographical cluster. The NJ tree was estimated in PAUP* 4.0b2 (Swofford 1998) using a distance matrix generated with Kimura 2-parameter model (Kimura 1980).

Divergence estimates and refugial origins

The four mitochondrial clades showed largely allopatric distributions (Fig. 5), each occurring in one of the regions known to have comprised former glacial refugia (reviewed in Pielou 1991). The Mississippian clade was dominant in the Great Lakes and Mississippian biogeographic provinces, but exhibited little sequence divergence over this broad range. By contrast, each of the other three clades was fragmented into genetically distinct subgroups with allopatric distributions. Haplotypes in the Pacific clade were restricted to populations from the Cascadian biogeographic province, but were subdivisable into a southern mainland group, and a group restricted to the Queen Charlotte Islands. The Atlantic clade also included two groups, one restricted to populations from the Southeastern biogeographic province, and a second assemblage which included populations from sites in the North Appalachian biogeographic province. The Beringian clade was additionally subdivided into two phylogeographic groups (Fig. 5), although these lineages formed an unresolved polytomy in cladistic analysis. Populations with BER1 haplotypes occupied the Yukon-Mackenzie and Hudson Bay biogeographic provinces, occurring predominantly in the Northwest Territories (NWT) and prairies. By contrast, haplotypes of the other Beringian group (BER2) were located mainly in the western Yukon-Mackenzie and mainland Cascadian biogeographic provinces, with a few populations scattered in the Great Lakes and Mississippian regions (see Fig. 5).

The average per cent nucleotide sequence divergence between the least (Mississippian and Beringian) and most

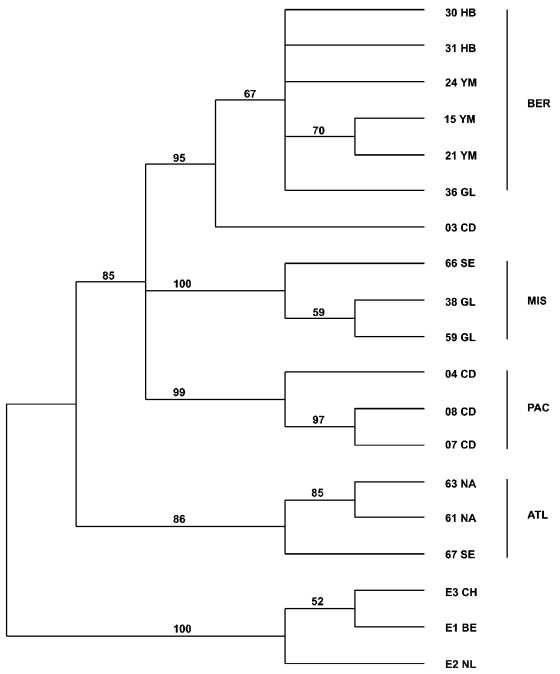

Fig. 4 Strict consensus of 36 most parsimonious trees (length 243, C.I. = 0.77, retention index (R.I.) = 0.87) for 19 *Sida crystallina* haplotypes, based on 147 phylogenetically informative characters of the COI gene. Trees were obtained using an heuristic search in PAUP* 4.0b2 (Swofford 1998), with steepest descent and tree bisection reconnection options invoked. Transitions and transversions were weighted equally; additional analyses with differential weightings did not affect the topology of the phylogeny. Clades are identified by lines, bootstrap values (1000 replicates) are shown on the branches.

(Atlantic and Pacific) divergent clades in North America was 4.8% and 7.4%, respectively, while European populations showed 23.4–23.8% divergence from the four North American clades. Assuming 2% sequence divergence per million years, North American and European *S. crystallina* have not been in contact for more than 10 million years, while the four clades within North America last shared common ancestors between 2.4 and 3.7 million years ago (Ma).

Discussion

This paper supports the prior taxonomic discrimination of *Sida crystallina* lineages from Europe and North America; allozyme and mtDNA divergences were far greater between populations from the different continents than across North America. In fact, this intercontinental divergence was greater than that between many other congeneric species of cladocerans (e.g. Taylor *et al*. 1998), indicating that these taxa merit recognition as distinct species. Regardless of their final taxonomic placement, the 23% sequence divergence at COI suggests that the colonization event(s) which led to the Holarctic distribution of *Sida* occurred before the Pliocene. Our failure to detect the haplotypes and allozyme alleles diagnostic of European lineages in North America suggests that long distance dispersal between these continents is so uncommon that it has not, at least in North America, impacted subsequent evolutionary trajectories.

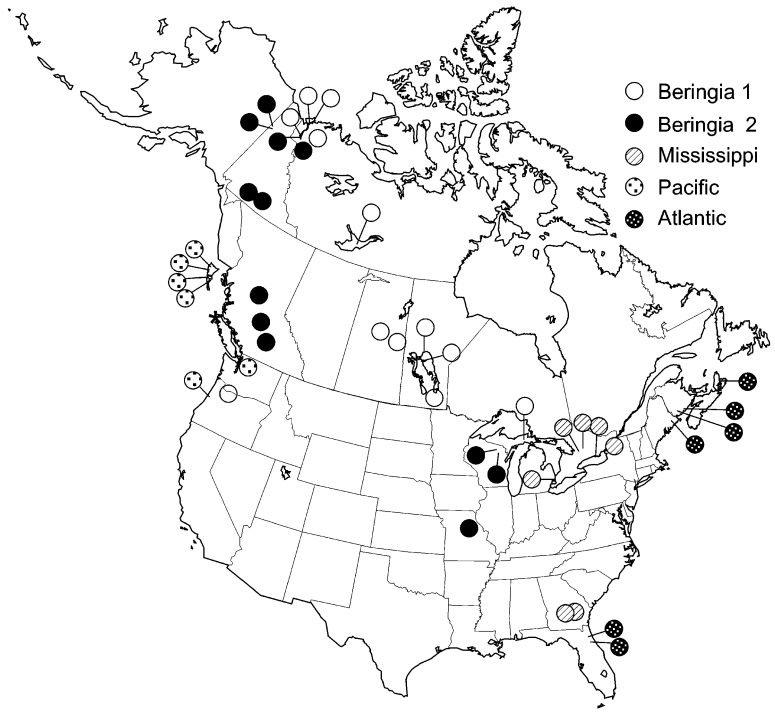

Fig. 5 Geographic distribution of the four major clades of *Sida crystallina americana* within North America. The Beringian clade is broken into two phylogeographic groups (BER1, BER2). Shaded circles and the single asterix correspond to haplotypes identified within the clusters of the NJ tree in Fig. 3.

Allozyme analyses suggested that North American populations now assigned to *S. crystallina* are a single species. There was, for example, no evidence of the HW disturbances which characterize populations where sibling species occur in sympatry (Witt & Hebert 2000). Also, allelic arrays were shared by populations across the continent, and the regional patterning of gene frequencies was modest. Populations in proximity did exhibit marked gene frequency shifts, but this is typical of populations founded from a small number of colonists capable of parthenogenesis (Boileau *et al*. 1992).

The analysis of sequence variation at COI revealed high diversity with 40 haplotypes detected among the 44 North American populations. Moreover, cladistic analysis indicated that these haplotypes were members of four clades showing marked (4–7%) sequence divergence. These clades had largely allopatric distributions, which suggested their derivation from source populations in the Mississippian, Atlantic, Beringian and Pacific refugia. Based on a clock calibration of 2% per million years (Myr) these lineages diversified 2.4–3.7 Ma, consistent with their isolation just prior to, or at the onset of the Pleistocene (Webb & Bartlein 1992). In addition to isolation imposed by advancing ice sheets during the Pleistocene, southern refugial groups were undoubtedly further separated by other physiographic features. For example, the clear divergence observed between Atlantic lineages of *Sida* and the neighbouring Mississippian refugial group to the west likely reflects isolation induced by the Appalachian Mountains. Similarly, the Western Cordillera presented a barrier to gene flow between populations on the Pacific coast and those in the prairies.

The depth of divergence between refugial groups of *Sida* and their largely allopatric distributions indicate that these populations exhibit a Category I phylogeographic pattern (Avise 2000), which is characterized by the presence of spatially isolated haplogroups showing marked genetic divergence. Closer inspection of taxa which exhibit such patterns often reveals further geographical structure within phylogroups, which may reflect isolation by distance or restricted contemporary gene flow (Avise 2000). In fact, all of the phylogroups in *Sida*, except the Mississippian lineage, showed further geographical structure. The substantial COI divergence between populations of *Sida* from the Queen Charlotte Islands and their mainland counterparts on the Pacific Coast suggests the fragmentation of this refugium into two isolates, a conclusion reinforced by evidence of allozymic divergence in these populations. *Sida* populations along the Atlantic coast were also separable into two subgroups, one from the South-eastern biogeographic province, and the other from the North Appalachian biogeographic province.

Despite their shallow sequence divergence, Beringian isolates were similarly separated into two phylogroups, which merit particular attention from a colonization perspective. The distribution of the BER1 haplotypes suggested their dispersal from Beringia through the arc of proglacial lakes which extended into modern Lake Winnipeg (Dyke & Prest 1987). The absence of the second lineage (BER2) from these habitats suggests that these two haplogroups were not panmictic within Beringia during the last glaciation, but experienced restricted gene flow and subsequently utilized separate dispersal corridors. The generally western distribution of the BER2 haplotypes suggests their derivation from a Nahanni refuge (Lindsey & McPhail 1986; Dyke & Prest 1987) with dispersal occurring southwards along the Pacific coast, in a similar fashion to that reported in lake trout (Wilson & Herbert 1998). The presence of the BER2 lineage to the south-west of the Great Lakes region likely represents its subsequent eastward dispersal, but further sampling in the prairies is required to test this speculation.

Collectively, these haplotype distributions indicate that post-Wisconsinan corridors enabled only a unidirectional passage for *Sida* — allowing the southward movement of Beringian lineages, but not the northerly migration of Mississippian or Pacific refugial groups. In addition, the presence of phylogeographic structure over smaller spatial scales *within* major refugial groups indicates that, despite the ability for overland dispersal in *Sida*, such events are either too infrequent to overcome divergence due to lineage sorting and random drift among isolated populations, or that pre-existing populations are so large that they resist penetration by secondary colonists.

The phylogeographic groups observed in North American *Sida* are remarkable in two ways. First, the various lineages show a higher degree of sequence divergence than reported in fishes (Bernatchez & Wilson 1998). Second, the boundaries between lineages are very well defined, with little evidence of the admixture which would be expected if dispersal events were nondirectional, as might be predicted due to their ability to disperse overland. The general coincidence between the boundaries of *Sida* lineages, and biogeographical provinces in fishes, suggest that much dispersal in *Sida* has been accomplished through the same interconnections which facilitated gene flow in fish. However, the occurrence of rare, long-distance dispersal events is suggested by the presence of disjunct haplotypes such as the Mississippian refugial lineages in the South-eastern biogeographic province, and a Beringian lineage in one habitat within the Cascade biogeographic province.

Few studies have examined the patterning of genetic diversity in other freshwater organisms reliant on passive dispersal. Work on the genus *Daphnia* has largely focused on polar taxa which show substantial evidence of lineage admixture and long distance dispersal (Weider & Hobaek 1997; Colbourne *et al*. 1998; Weider *et al*. 1999a,b). Studies on a single temperate zone assemblage, the *D. laevis* complex,

revealed a pattern similar to that in *S. crystallina* with marked genetic divergence between allopatric lineages that showed little evidence of secondary contact (Taylor *et al*. 1998). Moreover, the Appalachians and Western Cordillera played the same important role in defining phylogeographic boundaries as they did in *Sida*. This contrasting pattern of secondary contact between temperate and polar taxa suggests that in areas where sustained occupancy has been possible, lineages have long histories of tenure and have pre-empted the diffusion of secondary colonists. By contrast, in areas remote from refugia, the prevalence of vacant habitats has provided greater opportunities for lineage mixture.

Comparative phylogeography

There is now sufficient information to gain a preliminary sense of the response of freshwater organisms with different dispersal strategies to the cyclic advance and retreat of ice sheets during the Pleistocene. The recurrent formation of specific refugia, as a result of either persistent physiographic barriers, such as mountain ranges, or the regeneration of habitats linked to sea level lowering, provided a basis for the sustained differentiation of lineages (Hewitt 1996). However, the potential for the admixture of lineages during each interglacial set the stage for the erosion of incipient divergence among different refugial stocks (Bernatchez & Wilson 1998). Also, despite the repeated construction of specific refugia and the successful retreat of populations to these sites, the risk of local extinction was likely high as a joint consequence of the long duration of each glacial advance and the limited number of habitats in each refugium (Hewitt 2000).

The mitochondrial genomes of coldwater fish species record a history of extinction events and lineage contact throughout the Pleistocene in three ways. First, it appears that modern populations often derive from just one or two refugia (Bernatchez & Wilson 1998), indicating that many isolates perished. Second, the number of haplotypes in each refugial lineage is ordinarily low, indicating the loss of diversity, as expected if the population size in each refugium was small (Hewitt 1996). Third, the extent of sequence divergence among modern haplotypes from different refugia is small, suggesting that frequent secondary contact achieved via active dispersal through postglacial meltwaters, and subsequent genetic drift, led to the loss of diversity which originated earlier in the Pleistocene (Bernatchez & Wilson 1998; Wilson & Hebert 1998). Hence, refugial lineages in coldwater fish species rarely show more than 1% sequence divergence, while those in more southerly species show more than four times this level (Bermingham & Avise 1986; Billington & Hebert 1991; Avise 2000).

The mitochondrial genome of *Sida* records a different history of response to the repeated glacial cycles than that observed in freshwater fishes; one which signals a much reduced sensitivity to extinction. First, there is evidence from this study that its lineages persisted in all of the major refugia available to freshwater life in temperate North America. Second, haplotype diversity within refugial groups of *Sida* is high relative to that observed within refugial groups of teleosts from the same regions. Third, even the lowest amount of sequence divergence between phylogroups is greater than that observed in freshwater fish subject to the same glaciation events, suggesting that phylogroups of *Sida* successfully survived recurrent cycles of ice advance and retreat. Considered jointly, these results indicate that refugial lineages of *Sida* were little impacted by either extinction events or secondary contact throughout the Pleistocene. Their resistance to extinction is likely a joint consequence of the production of diapausing eggs which can survive unfavourable conditions for centuries, and their parthenogenetic mode of reproduction which allows the refounding of populations from a single individual (Hebert 1987; Korovchinsky & Boikova 1996). The potentially large size of populations in each habitat undoubtedly provided further protection from stochastic lineage extinction. These results are important from a phylogeographic perspective because they suggest that further work on organisms sharing these life history attributes will provide an unusually comprehensive phylogeographic record in comparison with studies on fishes, where much of the historical record has been erased through lineage loss. Differential exposure to extinction may also explain an otherwise puzzling biogeographic pattern — the absence of fish species endemic to polar freshwaters, but the presence of numerous invertebrates restricted to these settings.

Acknowledgements

We are particularly grateful to Jonathan Witt for stimulating discussions on the phylogeography of aquatic organisms. We thank S. Adamowicz, T. Crease, J. Garza, R. Gregory and two anonymous reviewers for helpful comments on earlier drafts of the manuscript. R. Dooh, J. Havel, S. Potter, J. Rhydderch, C. Rowe, A. Sands, Y. Um, and J. Witt assisted with North American collections, while H. Ketelaars, T. Little, and K. Schwenk provided samples from Europe. We thank A. Holliss for assistance with DNA sequencing, and I. Smith for technical help with figures. Funding for this research was provided by a Natural Sciences and Engineering Research Council grant to P.D.N.H., by an Ontario Graduate Scholarship in Science and Technology, and a Northern Studies Training Grant to A.J.C. Logistical support for collections in the arctic were provided by the Polar Continental Shelf Project.

References

Avise JC (2000) *Phylogeography. The History and Formation of Species*. Harvard University Press, London, UK.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London Series B*, **265**, 1707–1712.

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

Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, **45**, 1016–1035.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, **7**, 431–452.

Billington N, Hebert PDN (1991) Mitochondrial DNA diversity in fishes and its implications for introductions. *Canadian Journal of Fisheries and Aquatic Sciences*, **48** (Suppl.1), 80–94.

Boileau MG, Hebert PDN, Schwartz SS (1992) Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *Journal of Evolutionary Biology*, **5**, 25–40.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the USA*, **91**, 6491–6495.

Brown JM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA*, **76**, 1967–1971.

Burr BM, Mayden RL (1992) Phylogenetics and North American freshwater fishes. In: *Systematics, Historical Ecology, and North American Freshwater Fishes* (ed. Mayden RL), pp. 18–75. Standford University Press, Stanford, California.

Colbourne JK, Crease TJ, Weider LJ *et al.* (1998) Phylogenetics and evolution of a circumarctic species complex (Cladocera: *Daphnia pulex*). *Biological Journal of the Linnean Society*, **65**, 347–365.

Dadswell MJ (1974) Distribution, ecology and postglacial dispersion of certain crustaceans and fishes in eastern North America. *National Museum of Canada Publications in Zoology*, **11**.

Darwin C (1859) *On the Origin of Species by Means of Natural Selection*. John Murray, London.

Dyke AS, Prest VK (1987) Paleogeography of northern North America, 18 000–5000 years ago. *Geological Survey of Canada, Map 1703A, Scale,* 1:12, 500 000.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fryer G (1996) Diapause, a potent force in the evolution of freshwater crustaceans. *Hydrobiologia*, **320**, 1–14.

Gilbert D (1992) *SeqApp, version 1.9a. A multiple sequence editor for Macintosh computers*. Published electronically on the Internet at fly.bio.indiana.edu.

Hebert PDN (1987) Genotypic characteristics of the Cladocera. *Hydrobiologia*, **145**, 183–193.

Hebert PDN, Beaton MJ (1993) *Methodologies for Allozyme Analysis using Cellulose Acetate Electrophoresis. A Practical Handbook*. Helena Laboratories, Beaumont, TX.

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

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

Hillis DM, Huelsenbeck JP (1992) Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity*, **83**, 189–195.

Hillis DM, Moritz C (1990) An overview of application of molecular systematics. In: *Molecular Systematics* (eds Hillis DM, Moritz C), pp. 502–515. Sineaur Associates, Sunderland.

Hocutt CH, Wiley EO (1986) *The Zoogeography of North American Freshwater Fishes*. John Wiley and Sons, New York.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Knowlton N, Weight LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London, Series B*, **256**, 2257–2263.

Knowlton N, Weigt LA, Solorzano LA, Mills DK, Bermingham E (1993) Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *Science*, **260**, 1629–1632.

Korovchinsky NM (1992) *Sididae and Holopediidae: (Crustacea: Daphniiformes)*. SPB Academic Publishing, The Hague, The Netherlands.

Korovchinsky NM, Boikova OS (1996) The resting eggs of the Ctenopoda (Crustacea: Branchiopoda): a review. *Hydrobiologia*, **320**, 131–140.

Kumar S, Tamura K, Nei M (1993) *MEGA: Molecular Evolutionary Genetics Analysis, Version 1.02*. Distributed by the authors. Pennsylvania State University, University Park, PA.

Lewis PO, Zaykin D (1999) *Genetic Data Analysis: a computer program for the analysis of allelic data, Version 1.0 (d13)*. Distributed by the authors on the internet from the GDA homepage, http://alleyn.eeb.uconn.edu/gda/.

Lindsey CC, Mcphail JD (1986) Zoogeography of fishes of the Yukon and Mackenzie basins. In: *The Zoogeography of North American Freshwater Fishes* (eds Hocutt CH, Wiley EO), pp. 639–674. John Wiley and Sons, New York.

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the USA*, **90**, 4087–4091.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Ota T (1993) *DISPAN: Genetic Distance and Phylogenetic Analysis*. Distributed by the Institute of Molecular Evolutionary Genetics: http://www.bio.psu.edu/People/Faculty/Nei/Lab/Programs.html Pennsylvania State University, University Park, PA.

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

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Saiki RK, Gelfand DH, Stoffel S (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487–491.

Schubart CD, Diesel R, Hedges SB (1998) Rapid evolution to terrestrial life in Jamaican crabs. *Nature*, **393**, 363–365.

Schwenk K (1996) *Evolutionary genetics of* Daphnia *species complex: hybridism in syntopy*. Publication 2231. Netherlands Institute of Ecology, Centre for Limnology, Nieuwersluis, The Netherlands.

Stemberger RS (1995) Pleistocene refuge areas and postglacial dispersal of copepods of the northeastern United States. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 2197–2210.

Swofford DL (1998) *PAUP*: Phylogenetic Analysis Using Parsimony, Version 4.0b2*. Sinauer associates, Sunderland MA.

Taylor DJ, Finston TL, Hebert PDN (1998) Biogeography of a widespread freshwater crustacean: pseudocongruence and cryptic endemism in the North American *Daphnia laevis* complex. *Evolution*, **52**, 1648–1670.

Webb T, Bartlein PJ (1992) Global changes during the last 3 million years: climatic controls and biotic responses. *Annual Review of Ecology and Systematics*, **23**, 141–173.

Weider LJ, Hobaek A (1997) Postglacial dispersal, glacial refugia, and clonal structure in Russian/Siberian populations of the arctic *Daphnia pulex* complex. *Heredity*, **78**, 363–372.

Weider LJ, Hobaek A, Colbourne JK *et al.* (1999a) Holarctic phylogeography of an asexual species complex I. mtDNA variation in arctic *Daphnia*. *Evolution*, **53**, 777–792.

Weider LJ, Hobaek A, Hebert PDN, Crease TJ (1999b) Holarctic phylogeography of an asexual species complex–II. Allozyme variation and clonal structure in arctic *Daphnia*. *Molecular Ecology*, **8**, 1–13.

Weider LJ, Lampert W, Wessels M, Colbourne JK, Limburg P (1997) Long-term genetic shifts in a microcrustacean egg bank associated with anthroprogenic changes in the Lake Constance ecosystem. *Proceedings of the Royal Society of London, Series B*, **254**, 1613–1618.

Wilson CC, Hebert PDN (1998) Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1010–1024.

Witt JDS, Hebert PDN (2000) Cryptic species diversity and evolution in the amphipod genus *Hyallela* within central glaciated North America: a molecular phylogenetic approach. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 687–698.

Andrea Cox recently completed her M. Sc. in Paul Hebert's laboratory. Her thesis work probed the importance of life history traits in determining the impact of Pleistocene glaciations on the genetic structure of freshwater crustaceans in North America.

Appendix I

Locations and abbreviations for North American and European habitats with *Sida crystallina* s.l. Populations are designated by both number and biogeographic province of origin (Burr & Mayden 1992): CD (Cascadia); GL (Great Lakes); HB (Hudson Bay); MI (Mississippi); NA (North Appalachian); SE (South-eastern); YM (Yukon-Mackenzie). Country abbreviations are: CA (Canada); US (United States); BE (Belgium); NL (Netherlands); CH (Germany). Type of data collected from each population are indicated by: D, mtDNA only; A, allozymes only; *mtDNA and allozyme data

Code	Sample Site	Province / Territory / State	Country	Lat (N)	Long (W)	Data type
01 CD	Saunders Lake	Oregon	US	43.53	124.22	D
02 CD	Fern Ridge Lake	Oregon	US	44.05	123.29	*
03 CD	Unnamed Pond	Oregon	US	44.39	121.80	D
04 CD	Cain Lake	Washington	US	48.65	122.33	*
05 CD	Mosquito Lake	British Columbia	CA	53.07	132.07	*
06 CD	Skidegate Lake	British Columbia	CA	53.11	131.98	*
07 CD	Pure Lake	British Columbia	CA	53.87	132.08	*
08 CD	Mayer Lake	British Columbia	CA	53.64	132.05	*
09 CD	Cluculz Pond	British Columbia	CA	53.91	123.64	*
10 CD	Fraser Lake	British Columbia	CA	54.06	124.85	*
11 CD	Babine Lake	British Columbia	CA	54.85	126.18	D
12 YM	Boya Lake	British Columbia	CA	59.37	129.10	A
13 YM	Blue Lake	British Columbia	CA	59.83	129.13	A
14 YM	Twin Lakes	Yukon	CA	61.70	135.93	*
15 YM	Klusha Pond	Yukon	CA	61.59	135.84	*
16 YM	Ferry Pond	Northwest Territories	CA	67.33	134.92	*
17 YM	Small Frog Lake	Northwest Territories	CA	67.38	134.15	*
18 YM	Three Forks Marsh	Northwest Territories	CA	67.42	133.92	A
19 YM	Arctic Red River 2	Northwest Territories	CA	67.69	132.07	*
20 YM	Crossley Lakes 15	Northwest Territories	CA	67.86	131.48	*
21 YM	Old Crow 16	Yukon	CA	67.97	139.50	D
22 YM	Old Crow 20	Yukon	CA	68.19	139.50	*
23 YM	Eskimo Lakes 5	Northwest Territories	CA	68.56	133.75	D
24 YM	Eskimo Lakes 3	Northwest Territories	CA	68.64	133.54	*
25 YM	Great Slave Lake	Northwest Territories	CA	62.40	114.32	D
26 HB	Lac La Ronge	Saskatchewan	CA	55.09	105.31	*
27 HB	Amisk Lake	Saskatchewan	CA	54.69	102.08	*
28 HB	Reed Lake	Manitoba	CA	54.59	100.50	A
29 HB	Lake Wekusko	Manitoba	CA	54.64	99.25	*
30 HB	Lake William	Manitoba	CA	53.85	99.25	*
31 HB	Lake Manitoba	Manitoba	CA	50.73	98.14	*
32 HB	Lake of the Woods	Ontario	CA	49.58	94.53	A
33 HB	Quebec Pond	Quebec	CA	48.08	69.93	A
34 MI	Springfield Lake	Missouri	US	37.20	93.28	D
35 GL	Frank Lake	Wisconsin	US	46.07	89.55	D
36 GL	Plum Lake	Wisconsin	US	46.00	89.50	*
37 GL	Soldier Lake	Michigan	US	46.33	84.88	D
38 GL	Lake St. Clair	Ontario	CA	42.47	82.67	D
39 GL	Sparrow Lake	Ontario	CA	44.80	79.40	A
40 GL	Lake Couchiching	Ontario	CA	44.67	79.38	A
41 GL	Prospect Lake	Ontario	CA	44.99	79.13	A
42 GL	Echo Lake	Ontario	CA	45.18	79.07	A
43 GL	Raven Lake	Ontario	CA	45.20	78.85	A
44 GL	Paint Lake	Ontario	CA	45.22	78.95	A
45 GL	Muskoka Lake	Ontario	CA	45.00	79.42	A
46 GL	Lake Rosseau	Ontario	CA	45.17	79.58	A
47 GL	Maple Lake	Ontario	CA	45.12	78.67	A
48 GL	Pine Lake	Ontario	CA	45.12	78.58	A
49 GL	Beech Lake	Ontario	CA	45.08	78.72	*
50 GL	Grass Lake	Ontario	CA	45.03	78.55	A
51 GL	Little Boshkong Lake	Ontario	CA	45.03	78.72	A

Appendix I *Continued*

Code	Sample Site	Province / Territory / State	Country	Lat (N)	Long (W)	Data type
52 GL	Chemong Lake	Ontario	CA	44.40	78.38	A
53 GL	Lake Esson	Ontario	CA	45.02	78.27	A
54 GL	Silent Lake	Ontario	CA	44.95	78.13	*
55 GL	Coburg River	Ontario	CA	43.98	78.17	A
56 GL	Lake St. George	Ontario	CA	44.70	79.37	A
57 GL	Paudash Lake	Ontario	CA	44.97	78.05	A
58 GL	Millhaven Dock	Ontario	CA	44.18	76.75	D
59 GL	1000 Islands	Ontario	CA	44.23	76.42	D
60 NA	Cape Breton Lake	Nova Scotia	CA	45.77	60.60	*
61 NA	Second Lake	New Brunswick	CA	45.40	65.82	*
62 NA	Magaguadavic Lake	New Brunswick	CA	45.72	67.20	D
63 NA	Patrick Lake	Maine	US	44.88	67.38	D
64 NA	Gardener Lake	Maine	US	44.75	67.33	A
65 SE	Berry Reservoir	Georgia	US	34.17	85.17	*
66 SE	Alatoona Lake	Georgia	US	34.17	84.73	*
67 SE	Santa Fe Lake	Florida	US	29.74	82.08	*
68 SE	Dorr Lake	Florida	US	29.00	81.63	*
E1 BE	River Meuse	Hastiere	BE	51.16	4.82	D
E2 NL	Lake Maarsseveen	Utrecht	NL	52.14	5.09	D
E3 CH	Schöehsee	Plön	CH	54.02	10.05	*

Appendix II

Summary of allele frequencies in North American and European populations of *Sida crystallina* s.l. at five polymorphic allozyme loci. Alleles are labelled according to their mobility relative to the designated standard. Population codes are presented in Appendix I

		Population									
Locus		**02CD**	**04CD**	**05CD**	**06CD**	**07CD**	**08CD**	**09CD**	**10CD**	**12YM**	**13YM**	**14YM**
Mpi	**(*n*)**	**20**	**29**	**10**	**19**	**20**	**20**	**20**	**17**	**19**	**40**	**28**
0.92		0.000	0.328	0.000	0.000	0.000	0.000	0.000	0.176	0.000	0.013	0.000
1.00		0.525	0.052	0.000	0.000	0.000	0.000	1.000	0.824	0.868	0.963	0.857
1.08		0.475	0.621	1.000	1.000	1.000	1.000	0.000	0.000	0.132	0.025	0.143
Pgm	**(*n*)**	**21**	**30**	**10**	**19**	**20**	**20**	**20**	**19**	**20**	**39**	**28**
0.93		0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.075	0.051	0.000
1.00		1.000	0.633	1.000	0.974	0.775	0.125	1.000	1.000	0.925	0.756	1.000
1.03		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.103	0.000
1.05		0.000	0.350	0.000	0.026	0.225	0.875	0.000	0.000	0.000	0.090	0.000
sAat	**(*n*)**	**21**	**30**	**9**	**19**	**20**	**20**	**20**	**20**	**20**	**40**	**28**
0.71		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.025	0.000
1.00		1.000	1.000	1.000	0.974	1.000	0.950	1.000	1.000	1.000	0.975	1.000
1.09		0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mdh	**(*n*)**	**21**	**30**	**10**	**19**	**20**	**20**	**20**	**20**	**20**	**40**	**28**
0.83		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.91		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.200	0.950	0.013	0.000
1.00		1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.800	0.050	0.988	1.000
Gpi	**(*n*)**	**21**	**30**	**10**	**19**	**20**	**20**	**20**	**20**	**20**	**40**	**28**
0.55		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.80		0.000	0.383	0.200	0.474	0.625	0.625	0.000	0.025	0.000	0.063	0.000
1.00		0.786	0.617	0.000	0.000	0.050	0.000	1.000	0.975	1.000	0.938	1.000
1.16		0.214	0.000	0.800	0.526	0.325	0.375	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.04		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix II *Continued*

		Population									
Locus		**15YM**	**16YM**	**17YM**	**18YM**	**19YM**	**20YM**	**22YM**	**24YM**	**26HB**	**27HB**	**28HB**
Mpi	**(*n*)**	**10**	**20**	**40**	**20**	**30**	**20**	**29**	**29**	**39**	**25**	**20**
0.92		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.00		0.650	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.08		0.350	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pgm	**(*n*)**	**10**	**20**	**40**	**20**	**30**	**20**	**29**	**29**	**40**	**25**	**20**
0.93		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.00		1.000	1.000	1.000	1.000	0.967	1.000	1.000	1.000	1.000	1.000	1.000
1.03		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.05		0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000
sAat	**(*n*)**	**10**	**20**	**40**	**20**	**30**	**20**	**29**	**29**	**40**	**25**	**20**
0.71		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.00		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.09		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mdh	**(*n*)**	**10**	**20**	**40**	**20**	**30**	**20**	**29**	**28**	**39**	**25**	**20**
0.83		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000
0.91		0.000	0.000	0.275	0.350	0.017	0.000	0.000	0.446	0.013	0.000	0.000
1.00		1.000	1.000	0.725	0.650	0.983	1.000	1.000	0.536	0.987	1.000	1.000
Gpi	**(*n*)**	**10**	**20**	**40**	**20**	**30**	**20**	**29**	**29**	**40**	**25**	**20**
0.55		0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.088	0.000	0.000
0.80		0.000	0.000	0.000	0.000	0.200	0.000	0.000	0.000	0.638	0.560	0.900
1.00		1.000	1.000	1.000	1.000	0.800	0.875	1.000	1.000	0.275	0.440	0.100
1.16		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.04		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

		Population										
Locus		**29HB**	**30HB**	**31HB**	**32HB**	**33HB**	**36GL**	**39GL**	**40GL**	**41GL**	**42GL**	**43GL**
Mpi	**(*n*)**	**23**	**20**	**20**	**38**	**19**	**20**	**22**	**36**	**26**	**39**	**31**
0.92		0.000	0.000	0.000	0.263	0.000	0.000	0.227	0.069	0.692	0.705	0.226
1.00		1.000	0.975	1.000	0.737	1.000	1.000	0.682	0.931	0.308	0.295	0.694
1.08		0.000	0.025	0.000	0.000	0.000	0.000	0.091	0.000	0.000	0.000	0.081
Pgm	**(*n*)**	**23**	**20**	**20**	**40**	**19**	**20**	**22**	**36**	**26**	**41**	**31**
0.93		0.000	0.000	0.125	0.000	0.000	0.050	0.000	0.014	0.000	0.000	0.000
1.00		1.000	0.975	0.875	1.000	0.947	0.950	0.432	0.403	0.808	0.683	0.500
1.03		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.05		0.000	0.025	0.000	0.000	0.053	0.000	0.568	0.583	0.192	0.317	0.500
sAat	**(*n*)**	**23**	**20**	**20**	**40**	**19**	**20**	**22**	**35**	**26**	**41**	**31**
0.71		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.050	0.227	0.343	0.000	0.244	0.016
1.00		1.000	1.000	1.000	0.988	1.000	0.950	0.773	0.657	1.000	0.756	0.984
1.09		0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mdh	**(*n*)**	**23**	**20**	**20**	**40**	**19**	**20**	**22**	**36**	**26**	**42**	**31**
0.83		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.91		0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.00		0.978	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Gpi	**(*n*)**	**23**	**20**	**20**	**40**	**19**	**20**	**22**	**36**	**26**	**42**	**31**
0.55		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.80		0.696	0.400	0.875	0.688	0.000	0.000	0.000	0.000	0.000	0.012	0.016
1.00		0.304	0.600	0.100	0.313	1.000	1.000	1.000	0.972	1.000	0.988	0.984
1.16		0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.04		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Appendix II *Continued*

		Population										
Locus		**44GL**	**45GL**	**46GL**	**47GL**	**48GL**	**49GL**	**50GL**	**51GL**	**52GL**	**53GL**	**54GL**
Mpi	**(*n*)**	**37**	**44**	**39**	**27**	**43**	**37**	**19**	**16**	**43**	**30**	**41**
0.92		0.135	0.318	0.103	0.130	0.093	0.122	0.079	0.313	0.000	0.200	0.000
1.00		0.595	0.648	0.859	0.852	0.907	0.865	0.921	0.688	0.907	0.800	0.988
1.08		0.270	0.034	0.038	0.019	0.000	0.014	0.000	0.000	0.093	0.000	0.012
Pgm	**(*n*)**	**37**	**42**	**38**	**26**	**43**	**37**	**18**	**16**	**43**	**31**	**41**
0.93		0.000	0.012	0.013	0.000	0.000	0.000	0.000	0.000	0.012	0.081	0.000
1.00		0.662	0.250	0.566	0.481	0.372	0.446	0.667	0.156	0.523	0.581	0.415
1.03		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.05		0.338	0.738	0.421	0.519	0.628	0.554	0.333	0.844	0.465	0.339	0.585
sAat	**(*n*)**	**37**	**44**	**39**	**28**	**43**	**36**	**19**	**16**	**43**	**31**	**41**
0.71		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.203	0.148	0.218	0.036	0.070	0.083	0.053	0.063	0.035	0.194	0.146
1.00		0.797	0.852	0.744	0.964	0.930	0.917	0.947	0.875	0.965	0.806	0.854
1.05		0.000	0.000	0.038	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000
Mdh	**(*n*)**	**37**	**44**	**39**	**28**	**43**	**37**	**19**	**16**	**43**	**31**	**41**
0.83		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.91		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.012	0.000	0.000
1.00		1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.969	0.988	1.000	1.000
Gpi	**(*n*)**	**37**	**44**	**39**	**28**	**43**	**37**	**19**	**16**	**43**	**31**	**41**
0.55		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.80		0.014	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.035	0.000	0.000
1.00		0.986	1.000	0.987	1.000	1.000	1.000	1.000	1.000	0.965	1.000	1.000
1.16		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.04		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

		Population									
Locus		**55GL**	**56GL**	**57GL**	**60NA**	**61NA**	**64NA**	**65SE**	**66SE**	**67SE**	**68SE**	**E3CH**
Mpi	**(*n*)**	**19**	**44**	**44**	**20**	**15**	**54**	**22**	**20**	**15**	**22**	**19**
0.92		0.132	0.125	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.00		0.842	0.875	0.989	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
1.08		0.026	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
Pgm	**(*n*)**	**19**	**44**	**42**	**20**	**15**	**54**	**22**	**20**	**15**	**22**	**19**
0.93		0.053	0.000	0.000	0.050	0.000	0.009	0.000	0.000	0.000	0.000	1.000
1.00		0.500	0.250	0.476	0.950	1.000	0.991	1.000	1.000	1.000	1.000	0.000
1.03		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.05		0.447	0.750	0.524	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
sAat	**(*n*)**	**19**	**43**	**43**	**20**	**15**	**54**	**22**	**20**	**15**	**22**	**19**
0.71		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
0.89		0.605	0.244	0.128	0.125	0.000	0.009	0.000	0.000	0.000	0.000	0.000
1.00		0.395	0.756	0.872	0.875	1.000	0.991	1.000	1.000	1.000	1.000	0.000
1.09		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mdh	**(*n*)**	**19**	**44**	**44**	**18**	**15**	**54**	**22**	**20**	**15**	**22**	**19**
0.83		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.91		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.00		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Gpi	**(*n*)**	**19**	**44**	**43**	**20**	**15**	**54**	**22**	**20**	**15**	**22**	**19**
0.55		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.80		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.233	0.000	0.000
1.00		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.767	1.000	0.000
1.16		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.89		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.895
1.04		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.105