Arch. Hydrobiol. Spec. Issues Advanc. Limnol. 57, p. 1–10, July 2002
Biology and Management of Coregonid Fishes – 1999

Genetic structure of Polish populations of vendace (*Coregonus albula* L.) inferred from mitochondrial DNA

P. Brzuzan, J. Kozlowski and D. Fopp

with 4 figures and 4 tables

Abstract: A PCR based RFLP analysis of mtDNA was applied to test the genetic and phylogeographic relatedness of five vendace (*Coregonus albula*) populations from Poland. Seventeen composite haplotypes were detected by restriction enzyme analysis of the d-loop and of the ND3/4 (NADH dehydrogenase subunits 3 and 4), employing five endonucleases. UPGMA and AMOVA analyses revealed and confirmed the existence of two clonal groups which differed in their geographical distribution: one predominating in west/central, and one in east Poland. It has been estimated that these two lineages shared a common ancestor approximately 0.1–0.5 million years ago, during the middle to late Pleistocene. This suggests that vendace from the studied geographical region are derived from two glacial refugia and that they recolonized the area via two different routes.

Introduction

Morphological diversity and ecological plasticity of vendace populations (*Coregonus albula* L.) have been widely reported [for example, SVÄRDSON (1979)]; surprisingly, few studies have documented the genetic structure of this species. Moreover, there is a lack of genetic data regarding vendace populations inhabiting Polish lakes. This study uses a hierarchical framework for resolving the intraspecific phylogeny of vendace inhabiting northern Poland in relation to the Pleistocene glaciation, which has shaped genetic structure of the Coregoninae in a variety of ways (BERNATCHEZ et al. 1989, BERNATCHEZ & DODSON 1991).

In this study we have applied a PCR-based RFLP analysis of mtDNA to reveal the genetic and phylogeographic structure of the vendace populations at different geographical scales: within lakes, among lakes within regions (northwestern *vs.* northeastern Poland), and among regions. Because mtDNA is maternally inherited, it is particularly sensitive to drift and fixation on a short evolutionary timescale, which in turn makes studies utilizing mtDNA effective for detection of population differentiation, especially when compared to methods based on nuclear DNA (HAUSER et al. 1995, NEIGEL 1996).

Address of the authors: P. Brzuzan and D. Fopp, Department of Evolutionary Ecology, WM University in Olsztyn, 10-718 Olsztyn, Poland. J. Kozlowski, Department of Fish Biology and Culture, WM University in Olsztyn, 10-718 Olsztyn, Poland.

0071–1128/02/0057–001 $ 2.50

Material and methods

Populations

A total of 153 autumn-spawning European vendace (*Coregonus albula*) were collected from five localities in both northwestern and northeastern Poland (Fig. 1). Fish were caught in 1998 with gill nets during their spawning time. The fish were frozen immediately upon capture.

Mitochondrial DNA analysis

DNA was extracted using WIZARD® DNA purification Kit (Promega). Two DNA segments of the mitochondrial genome were amplified: ND3/4 (NADH dehydrogenase; subunits 3 and 4) with the primers described by Nielsen et al. (1998), and the control region (D-loop) as described by Brzuzan (1998).

Double-stranded PCR amplifications, separately for either ND3/4 or control region, were performed in reaction volumes containing 2 units of *Thermophilus aquaticus* DNA polymerase (Promega), 5 µl reaction buffer (500 mM KCl, pH 9.0; 1% Triton X-100), 20 pmol (picomole) of each primer (MWG-BIOTECH, Ebersberg, Germany), 2.5 mM $MgCl_2$, and 500 µM dATP, dCTP, dGTP and dTTP (Promega). As a template, 2 µl of the DNA preparation was

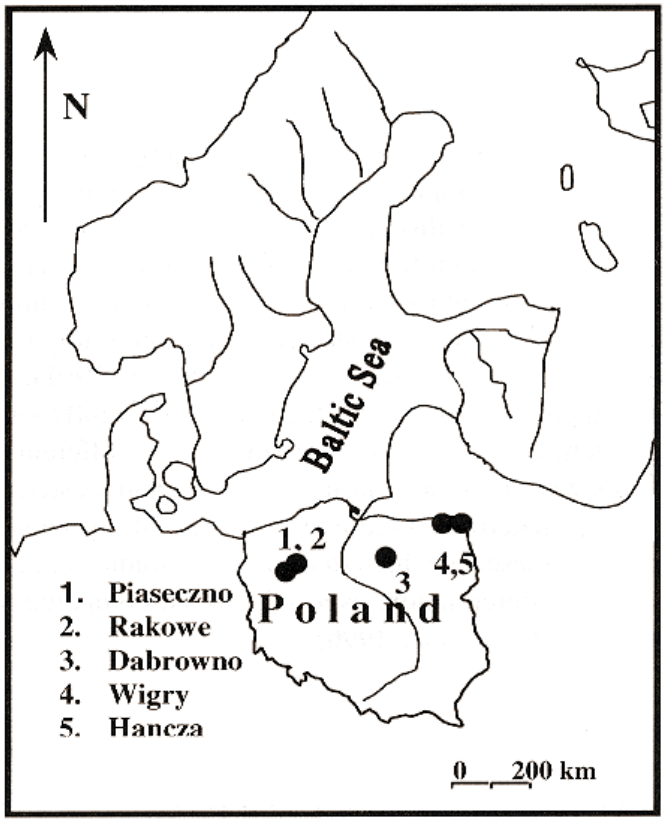

Fig. 1. Sampling sites (closed circles) for vendace (*Coregonus albula*) in Poland.

added to the PCR mix. DNA was amplified in an Perkin-Elmer (2400) thermal cycler beginning with preliminary denaturation at 95 °C for 1 min. The amplification cycle consisted of 92 °C for 1 min, 50 °C for 1 min, and 72 °C for 1.5 min. Reactions were cycled 30 times.

The amplified segments of DNA were screened for restriction site polymorphism using five restriction endonucleases: *Alu* I, *Cfo* I, *Hae* III, *Msp* I, and *Rsa* I. DNA digests were subjected to agarose gel electrophoresis to separate fragments according to their molecular weight. Electrophoresis was performed using horizontal slab gels containing 1.0 or 1.5 % agarose at 80 V for 1 hr. Ethidium bromide (0.1 µg/ml) was added to the gel and mtDNA bands were visualised for photography under UV light.

Data analysis

The size of restriction fragments was estimated by a comparison to a hundred-base-pair ladder (Promega). Fragments larger than approximately 100 bp were recorded (Table 1). Restriction fragment patterns were coded for presence or absence of sites in a binary matrix which was used for both distance (*d*-values) and character-based analyses. The genetic relationship among

Table 1. Approximate size in base pairs (bp) of restriction fragments for European vendace d-loop and ND3/4 digested with four endonucleases: *Alu* I, *Hae* III, *Msp* I, and *Rsa* I.

D-loop

Restriction endonuclease	*Alu* I		*Msp* I		*Rsa I*	
Restriction Morph	A	B	A	B	A	B
Fragment sizes (bp)	600		550	550	700	700
	550	550	500		400	250
		300	250	250	230	230
		300		230		150
				230		

ND3/4

Restriction endonuclease	*Hae* III					*Msp* I		*Rsa* I			
Restriction morph	A	B	C	D	E	A	B	A	B	C	D
Fragment sizes (bp)	520	520		520	600		550		1500	1500	
	500			500	520	500	500				900
	400	400	400	400	400	350		800			600
	320	320	320	320	320	330	330	700		500	
	300	300	300	300	300	250	250	400	400		400
	300	300	300	150	200	250	250	250	250	250	250
		300	300	150		200	200	100	100		100
		200	300			200	100				
			220			100					
			200								

Only fragments larger than one hundred base pairs are included.

haplotypes was analyzed by calculating the mean number of nucleotide subsitutions between all pairs of haplotypes (Nei & Tajima 1981) using the program D in REAP (McElroy et al. 1992). The resulting distance matrix was clustered by UPGMA method (Sneath & Sokal 1973) using the program NEIGHBOR from PHYLIP 3.5c package (Felsenstein 1993). The mtDNA haplotypes were organized into a Wagner network using the program MIX from PHYLIP 3.5c (Felsenstein 1993). Levels of inter- and intrapopulation genetic diversity were quantified by indices of haplotype diversity (h, Nei & Tajima 1981), nucleotide diversity within (π; Nei 1987) and nucleotide divergence (d_{xy}, Nei & Tajima 1983) among populations using the program DA in REAP. A calibration of 1% sequence divergence per million years (Smith 1992) was chosen to calculate divergence times since a common ancestor. Pairwise tests for homogeneity of haplotype frequencies between samples were performed using a Monte Carlo simulation (Roff & Bentzen 1989) with 1,000 randomizations using the MONTE option in REAP. Levels of significance were adjusted according to the sequential Bonferroni method as described by Rice (1989).

An analysis of molecular variance (AMOVA, Excoffier 1995) was used to test the hierarchical partitioning of genetic variation among populations, and for calculation of Φ-statistics: Φ_{ST}, Φ_{CT}, and Φ_{SC} (Excoffier et al. 1992). For the partitioning of variance components into different hierarchical levels, vendace populations were organized into two groups according to their relative position on the population tree. The analyses were performed with the ARLEQUIN 1.1 software (Schneider et al. 1997). The relationship between geographical and genetic distance was investigated by comparing distances measured from lake to lake, with genetic distance computed as Φ_{ST} (Slatkin 1993). The relationship was tested by use of the program Mantel from the GENEPOP 3.1 package (Raymond & Rousset 1995).

Results

Restriction sites for *Cfo* I enzyme were conserved among the five populations studied in both d-loop and ND3/4 DNA segment. Digestion patterns of the d-loop fragment revealed two size variants (1.2 and 1.3 kb) and restriction site differences for three endonucleases: *Alu* I, *Msp* I and *Rsa* I (Table 1). The size of the amplified ND3/4 fragment was 2.3 kb. In this DNA segment, restriction morphs were revealed by using *Hae* III, *Msp* I and *Rsa* I (Table 1). In all cases, differences between morphs could be explained by the gain or loss of one restriction site.

Seventeen composite haplotypes were identified (Table 2). The Wagner parsimony analysis generated a total of 100 equally parsimonious networks requiring a minimum of 18 site changes. However, the analysis was unable to resolve phylogenetic relationships between all pairs of haplotypes, owing to homoplastic recognition sites detected by *Alu* I, *Hae* III and *Rsa* I (data not shown). Genetic relatedness among the observed haplotypes was resolved on the UPGMA tree which identified two lineages, labeled A and B (Fig. 2). Each lineage contained a dominant haplotype with mutational derivatives. The dominant haplotype of clade A was haplotype 1, whereas haplotype 13 predominated in lineage B. Estimated pairwise sequence divergences among all haplotypes ranged from 0.35 to 2.67%, with a mean value of 1.02%. The estimated sequence divergence (mean ± SE) between clades A and B was 1.10 ± 0.64%.

The number of haplotypes per population averaged 7.2, ranging from 3 to 16, and several 'private' mtDNA genotypes were observed (Table 2). The table shows that haplotype 13 was

Table 2. Distribution of vendace mtDNA haplotypes, haplotype diversity (± standard error), and percent nucleotide diversity among lakes. Haplotypes are denoted by capital letters in the following order: d-loop- *Alu* I, *Msp* I, *Rsa* I; ND3/4- *Hae* III, *Msp* I and *Rsa* I.

	Lake					
Haplotype	Piaseczno Duze	Rakowe	Dabrowno	Wigry	Hancza	TOTAL
1 AAAAAA		7	5	6	16	34
2 BAAAAA			3	1	4	8
3 AAAABA		1			1	2
4 AAABBA		1			1	2
5 BAAABA			4		1	5
6 BAAACA				3	1	4
7 BAADBA	1					1
8 BAABAA					1	1
9 AAABAA					1	1
10 AAADAA			1	2	2	5
11 BAACBA	4	1	10	1	6	22
12 BAABCA					1	1
13 BAABBA	23	19	17	1	3	63
14 BAAEDB					1	1
15 BBAABA					1	1
16 BABBBA					1	1
17 AAACAA					1	1
TOTAL	28	29	40	14	42	153
Haplotype diversity	0.32 (± 0.10)	0.53 (± 0.09)	0.74 (± 0.05)	0.79 (± 0.09)	0.83 (± 0.05)	0.64 (± 0.01)
Nucleotide diversity	0.14	0.48	0.55	0.77	0.81	0.55

the most common and appeared in all populations studied. Levels of haplotype diversity among vendace populations ranged from 0.32 (± 0.10) in Lake Piaseczno Duze to 0.83 (± 0.05) in Lake Hancza, and the overall haplotype diversity was 0.64 (± 0.01) (Table 2). Nucleotide diversity values (π) were high (0.14–0.81%) most likely due to sympatry of fish from two haplotype groups, A and B (Fig. 2; Table 2).

Significant differences in haplotype frequencies were found for the majority of comparisons between populations (Table 3); however, only the Hancza and Wigry populations were significantly different from all other populations. These two populations showed the largest average indices of differentiation to the other populations, with pairwise ϕ_{ST} values ranging from 0.078 to 0.452 (Table 3). An UPGMA phenogram clustering all studied populations (Fig. 3) showed that Wigry and Hancza composed a distinct group from the three other populations. The clustering pattern was largely caused by the differential distribution of A- and B-haplotypes among populations; the similarity of Wigry and Hancza populations and their distinction from the other three populations was due to the near absence of haplotype 13 of clade B and the dominance of haplotype 1 of clade A in these two populations. Nucleotide divergence values among populations between the two groups ranged from 0.122 to 0.513% (mean=0.251%) and were significantly higher than those obtained from within group comparisons: 0.004% in a group Wigry/Hancza, and 0.012–0.073% (mean=0.045) in the Piaseczno/Rakowe/Dabrowno group.

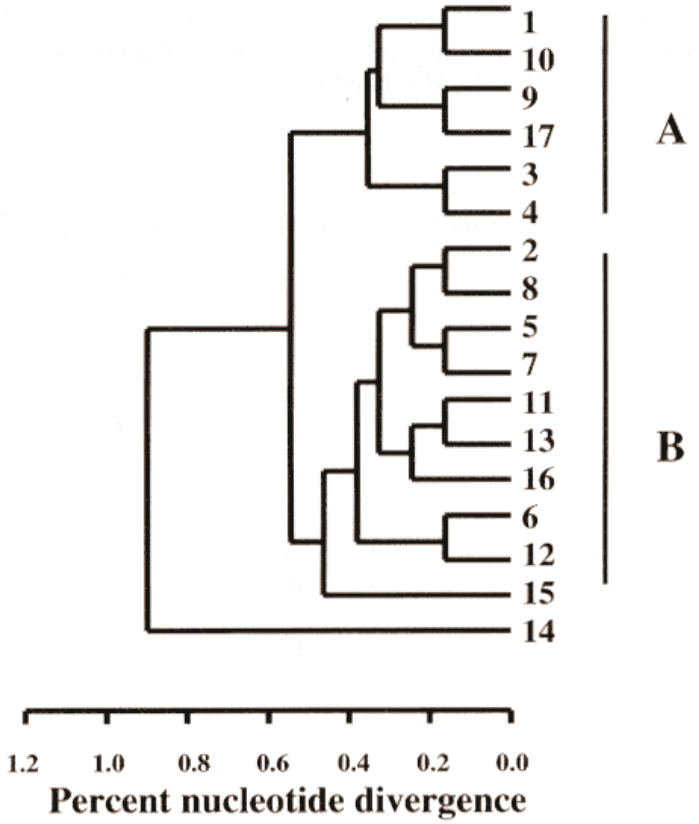

Fig. 2. UPGMA phenogram summarizing the genetic relationship among vendace haplotypes observed in the study. The tree was based on the average number of nucleotide substitution per site between haplotypes (*d* -values; Nei & Tajima 1981), and constructed by using the program NEIGHBOR from the PHYLIP 3.5c program package (Felsenstein 1993). MtDNA lineages are labelled A and B.

Most of the genetic variance was distributed within populations (Table 4). The among-populations-within-region level contributed least to the total variance (4.7%), and was mainly due to differences among the three western populations. In contrast, a notable proportion (18.7%) was distributed among regions. The relationship between genetic and geographical distance (Fig. 4) was not significant (Mantel test: P=0.053).

Discussion

Observed levels of haplotype diversity among vendace populations are similar to those reported for other northern freshwater species. Overall haplotype diversity of populations in Table 2 (h=0.64) was comparable to values for lake whitefish, *Coregonus clupeaformis* (h=0.73, 0.52) (Bernatchez & Dodson 1990, 1991) and walleye, *Stizostedion vitreum* (h=0.69) (Stepien & Faber 1998). These values were considerably lower, however, than those for European whitefish, *Coregonus lavaretus* (h=0.90) (Bernatchez et al. 1989).

High number of haplotypes observed in most vendace populations suggests that they either arose from large numbers of founder individuals or have maintained large population sizes

Table 3. Pairwise tests on homogeneity of mtDNA haplotype distribution between populations (above the diagonal), and pairwise tests on the extent of genetic differentiation between populations (below the diagonal). Above the diagonal are χ^2 values and their probabilities (P) from Monte Carlo tests (Roff & Bentzen 1989, McElroy et al. 1992). Below the diagonal are pairwise φ_{ST} values from AMOVA tests and the probability (P) of obtaining a random number greater than the value.

Lake (N)	Piaseczno (28)	Rakowe (29)	Dabrowno (40)	Wigry (14)	Hancza (42)
Piaseczno	–	χ^2=11.18	χ^2=15.85	χ^2=34.09	χ^2=48.94
	–	P=0.009*	P=0.002**	P<0.001**	P<0.001**
Rakowe	φ_{ST}=0.078	–	χ^2=16.47	χ^2=21.68	χ^2=33.26
	P=0.024*	–	P=0.008*	P<0.001**	P<0.001**
Dabrowno	φ_{ST}=0.128	φ_{ST}=0.066	–	χ^2=22.77	χ^2=28.81
	P=0.005*	P=0.026*	–	P<0.001**	P<0.001**
Wigry	φ_{ST}=0.452	φ_{ST}=0.240	φ_{ST}=0.140	–	χ^2=10.56
	P=0.000**	P=0.001**	P=0.003**	–	P=0.869
Hancza	φ_{ST}=0.355	φ_{ST}=0.197	φ_{ST} =0.100	φ_{ST}=–0.011	–
	P=0.000**	P=0.000**	P=0.000**	P=0.665	–

* denotes a significant difference at $P < 0.05$; ** denotes a significant difference at $P<0.005$ for multiple *post hoc* tests, using a sequential Bonferroni method (Rice 1989).

over time (Avise et al. 1984). This observation is concordant with relatively high levels of allozyme diversity observed in vendace populations (Vuorinen et al. 1981). As a short-living planktivorous fish, vendace population sizes have probably been large in all but smallest lakes. Haplotype distributions are unlikely to be a consequence of past stocking programs, because only two of the five lakes (Dabrowno and Wigry) were known to have been stocked, and the stocking material for both lakes was derived from native spawners.

The mtDNA analysis identified two distinct lineages among the Polish vendace populations. Because the mitochondrial molecular clock for salmonids has been estimated at 1% divergence per million years (Smith 1992), this suggested that the two groups shared a common ancestor 0.1–0.5 million years ago, during the mid- to late Pleistocene. Although the molecular data should be treated with caution, it seems likely that divergence between the

Table 4. Genetic variance components and Φ-statistics for vendace populations estimated by AMOVA (Excoffier et al. 1992).

Variance component	Variance	% total	Φ-statistics
Among regions	0.080	18.7	Φ_{CT}=0.187 P=0.096
Among populations within regions	0.020	4.7	Φ_{SC}=0.058 P=0.003
Within populations	0.327	76.6	Φ_{ST}=0.234 P<0.001

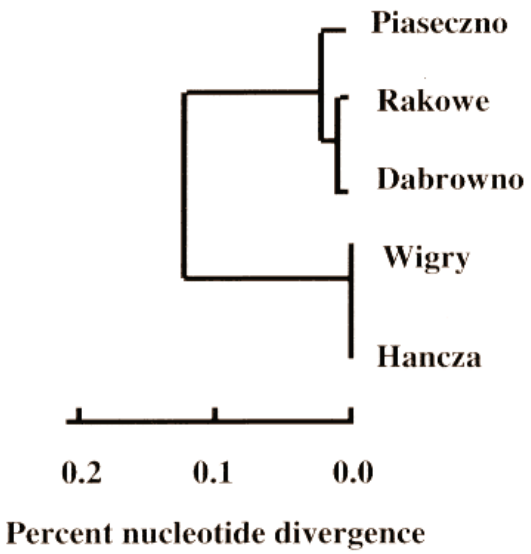

Fig. 3. UPGMA phenogram clustering vendace populations according to the distance matrix resulting from maximum likelihood estimation of the net average number of nucleotide substitutions per site between populations (d_{XY}, nucleotide divergence; NEI & TAJIMA 1983).

lineages preceded the last glaciation. It is possible that the two major mtDNA clusters A and B are characteristic of ancient populations that evolved in two separate glacial refugia and then, subsequently, recolonized the studied geographical region following the retreat of the ice sheet. Indeed, studies of other freshwater fishes in Europe and North America have all revealed a strong correlation of genetic lineage distribution and postglacial migration from different refuge areas (reviewed in e.g., BERNATCHEZ & WILSON 1998, NESBØ et al. 1999).

Further evidence that clonal groups A and B may have been derived from two different refugia is provided by their different geographic distributions. Haplotypes of cluster A are two times

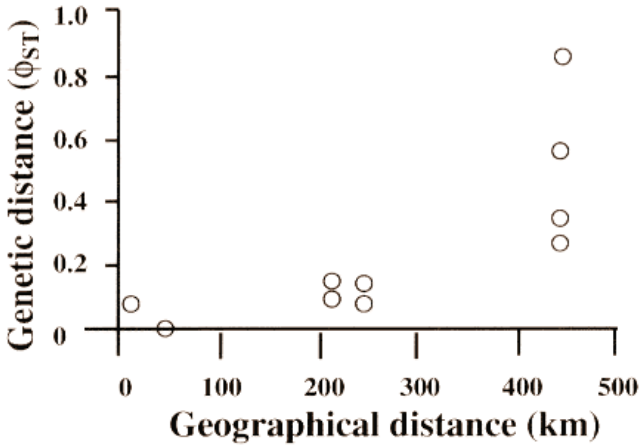

Fig. 4. Relationship between genetic (pairwise Φ_{ST} values) and geographical distances (km) between vendace populations.

more abundant in populations of eastern part of the studied region than are haplotypes of cluster B, whereas the reverse is true in central and western part of northern Poland. Similarly, the results from the analyses of molecular variance (Table 4) and UPGMA (Fig. 3) suggest considerable genetic differentiation among populations from west/central and east Poland. Moreover, we have found a positive correlation between genetic and geographical distance (Fig. 4) that may be a consequence of a clinal distribution of the two haplotype clusters over the studied area.

Bernatchez et al. (1989) proposed that recolonization of modern lakes in northern Europe by European whitefish was possible from at least three glacial refugia; the fish probably came from the east (present day Black and Caspian Seas) and some may have come from the south, following the ice. Therefore, one group of vendace (clonal group A) may have recolonized the area from east and the other (clonal group B) from the south. So far, however, the exact location of vendace refugia remains uncertain. A sampling design that includes more vendace populations over a much larger geographical range could clarify this.

The Eurasian *albula* complex consists of two Palearctic species, *C. albula* and *C. sardinella* which are regarded as allopatric (Berg 1962). It cannot be ruled out, however, that riverine and semianadromous *C. sardinella* could have taken advantage of transient connections among proglacial lakes and drainages during the Pleistocene and thus may have visited vendace refugia. Whether introgressive hybridization between the two species, followed by dispersal of introgressed individuals, occurred in the past, remains to be tested.

It is unlikely that vendace stocks originating after the withdrawal of the last glaciers can be discriminated by population-specific mtDNA markers. However, vendace populations may differ in distribution frequencies of their mtDNA lineages (Table 3). Bernatchez et al. (1989) described similar pattern for European whitefish (*Coregonus lavaretus*), a related species with an evolutionary history similar to vendace.

Acknowledgments

We thank Professor M. Luczynski (WM University in Olsztyn) for his comments on this MS, and Mr S. Ciesielski for his excellent lab assistance. We also thank two anonymous reviewers for their constructive comments. The study was financed by KBN Project No. 5P06D 043 14.

References

Avise, J.C., Neigel, J.E. & Arnold, J. (1984): Demographic influences on mitochondrial DNA survivorship in animal populations. – J. Mol. Evol. **20**: 99–105.

Berg, L.S. (1962): Freshwater fishes of the USSR and adjacent countries. – Israel Program for Scientific Transactions, Jerusalem.

Bernatchez, L. & Dodson, J.J. (1990): Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial DNA restriction analysis. – Evolution **44**: 1263–1271.

Bernatchez, L. & Dodson, J.J. (1991): Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to pleistocene glaciations. – Evolution **45**(4): 1016–1035.

Bernatchez, L., Dodson, J.J. & Boivin, S. (1989): Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. – J. Fish. Biol. **35** (Suppl. A): 233–244.

Bernatchez, L. & Wilson, C.C. (1998): Comparative phylogeography of Nearctic and Palearctic fishes. – Mol. Ecol. **7**: 431–452.

Brzuzan, P. (1998): DNA length variation and RFLP of the mitochondrial control region in two samples of whitefish, *Coregonus lavaretus*, from Lake Baikal (Russia) and Lake Maroz (Poland). – In: Eckmann, R., Appenzeller, A. & Rösch, R. (ed.): Biology and Management of Coregonid Fishes – 1996. – Arch. Hydrobiol. Spec. Issues Advanc. Limnol. **50**: 349–356.

Excoffier, L. (1995): AMOVA, Analysis of Molecular Variance, version 1.55. – University of Geneva, Switzerland.

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

Felsenstein, J. (1993): PHYLIP (Phylogeny Inference Package, version 3.5c). – Department of Genetics, SK, University of Washington, Seattle, WA.

Hauser, L., Carvalho, G.R. & Pitcher, T.J. (1995): Morphological and genetic differentiation of the African clupeid *Limnothrissa miodon* 34 years after its introduction to Lake Kivu. – J. Fish Biol. **47** (Suppl. A): 127–144.

McElroy, D., Moran, P., Bermingham, E. & Kornfield, I. (1992): REAP: an integrated environment for the manipulation and phylogenetic analysis of restriction data. – J. Hered. **83**: 157–158.

Nei, M. (1987): Molecular Evolutionary Genetics. – Columbia University Press, New York, NY.

Nei, M. & Tajima, F. (1981): DNA polymorphism detectable by restriction endonucleases. – Genetics **97**: 145–163.

Nei, M & Tajima, F. (1983): Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. – Genetics **105**: 207–217.

Neigel, J.E. (1996): Estimation of effective population size and migration parameters from genetic data. – In: Smith, T.B. & Wayne, R.K. (ed.): Molecular Genetic Approaches in Conservation. Oxford University Press, Oxford, pp. 329–346.

Nesbø, C.L., Fossheim, T., Vøllestad, A. & Jakobsen, K.S. (1999): Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. – Mol. Ecol. **8**: 1387–1404.

Nielsen, E.E., Hansen, M.M. & Mensberg, K.-L.D. (1998): Improved primer sequences for the mitochondrial ND1, ND3/4 and ND5/6 segments in salmonid fishes. Application to RFLP analysis of Atlantic salmon. – J. Fish Biol. **53**: 216–220.

Raymond, M. & Rousset, F. (1995): Population genetics software for exact test and ecumenicism. – J. Hered. **86**: 248–249.

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

Roff, D.A. & Bentzen, P. (1989): The statistical analysis of mitochondrial DNA polymorphism: χ^2 and the problem of small samples. – Mol. Biol. Evol. **6**: 539–545.

Schneider, S., Kueffer, J.M., Roessli, D. & Excoffier, L. (1997): ARLEQUIN 1.1-on line. Http:// anthropologie.unige.ch.arlequin.

Slatkin, M. (1993): Isolation by distance in equilibrium and non-equilibrium populations. – Evolution **47**: 264–279.

Smith, G.R. (1992): Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. – Syst. Biol. **41**: 41–57.

Sneath, P.H.A. & Sokal, R.R. (1973): Numerical Taxonomy. – Freeman, San Francisco.

Stepien, C.A. & Faber, J.E. (1998): Population genetic structure, phylogeography and spawning philopatry in walleye (*Stizostedion vitreum*) from mitochondrial DNA control region sequences. – Mol. Ecol. **7**: 1757–1769.

Svärdson, G. (1979): Speciation in Scandinavian *Coregonus*. – Inst. Freshwater Res, Drottingholm **57**: 1–95.

Vuorinen, J., Himberg, M.K.-J. & Lankinen, P. (1981): Genetic differentiation in *Coregonus albula* (L.) (Salmonidae) populations in Finland. – Hereditas **94**: 113–121.