Short Communication

Mitochondrial DNA fails to reveal genetic structure in sea-lampreys along European shores

V.C. Almada [a], A.M. Pereira [a], J.I. Robalo [a,*], J.P. Fonseca [a],
A. Levy [a], C. Maia [a,c], A. Valente [a,b]

[a] *Unidade de Investigação em Eco-Etologia, Instituto Superior de Psicologia Aplicada, Rua Jardim do Tabaco 44, 1149-041 Lisboa, Portugal*
[b] *Departamento de Zoologia e Antropologia, Faculdade de Ciências, Universidade do Porto, Praça Gomes Teixeira, 4099-002 Porto, Portugal*
[c] *Planeta Vivo—Centro de Investigação Ambiental, Travessa da Cedofeita, 62 4050-448 Porto, Portugal*

1. Introduction

The sea-lamprey, *Petromyzon marinus* Linnaeus, 1758 occurs on both sides of the North-Atlantic. While in North America it became a serious problem for many aquatic eco-systems after it invaded the great lakes (Bryan et al., 2005), in Western Europe its conservation is raising increasing concerns (Maitland, 2003). Being an adromous, the sea-lamprey migrates into rivers and must reach gravel streams where spawning takes place. The larvae are carried down-stream and must find suitable muddy/sandy substrata where they live, buried in the sediment for several years, fil-ter feeding on particles carried by the moving water. After the larval stage, the young lampreys, which acquire fully functional eyes and become active swimmers, migrate to the sea, where they live at least two years biting and suck-ing blood and other tissues from fish (e.g. Hardisty and Potter, 1971). With such a life-cycle, sea-lampreys are vul-nerable to a variety of disturbances: they must survive the crossing of estuaries, typically among the most polluted parts of rivers; they must find their way upstream, often a problem when dams and other kinds of river obstructions have been built; they must find water with the right flow and gravel with the right size for the adults to spawn and the larvae to hatch; and, growing larvae must find clean productive grounds where water velocity and sediment structure are appropriate for their growth. In addition, in some European countries, there is a very profitable fishery of sea-lampreys, placing an additional burden on the pop-ulations (Kelly and King, 2001).

Despite its economical value and high conservation prior-ity, little information is available on the structure of sea-lam-prey populations that spawn in Western Europe. Rodriguez-Muñoz et al. (2004), using a fragment of 511 bp of the con-trol region of the sea-lamprey mitochondrial genome, found that individuals spawning in the rivers Minho and Cela, separated by about 500 km, showed similar patterns of hap-lotype distribution. They also noted that no haplotypes were shared with samples from North America. In addition, their data suggested that the Iberian populations were much less diverse genetically (three haplotypes in 45 specimens) than the Western Atlantic populations (11 haplotypes in 71 spec-imens). These findings led Rodriguez-Muñoz et al. (2004) to raise the possibility that the sea-lamprey underwent a severe bottleneck in Southwest Europe. Bryan et al. (2005), using microsatellites, also found that in a Portuguese River (Mondego), the genetic diversity was very low when compared with that found in North-western Atlantic rivers, supporting the finding of Rodriguez-Muñoz et al. (2004). They suggested that a serious reduction of effective popula-tion size may have occurred in Europe as a result of the fisheries that exploit sea-lamprey in several countries.

Both microsatellites (Bryan et al., 2005) and mtDNA con-trol region (Waldman et al., 2006) revealed a notorious lack of differentiation among the North American rivers draining into the Atlantic. This lack of differentiation is not caused by insufficient variability in the markers used, as demon-strated by the differentiation among populations of lakes where the sea lamprey is possibly native (Bryan et al., 2005; Waldman et al., 2006). More likely, the lack of structure in Atlantic rivers results from the lack of homing in this species (Bergstedt and Seelye, 1995). Instead of homing, the sea lampreys are attracted by unknown cues coming from the

---

* Corresponding author. Fax: +351 21 8860954.
E-mail address: jrobalo@ispa.pt (J.I. Robalo).

rivers and by pheromones secreted by amocoetes (e.g. Bjerse-lius et al., 2000) which means that after a local extinction the re-colonization of a river may be difficult.

In this paper, we used a fragment of 624 bp of mtDNA of the sea-lamprey control region, from specimens collected in rivers ranging from the Rhine to the Guadiana (Southern Portugal) and encompassing a substantial fraction of the Western European shore. This fragment includes the region sequenced by Rodriguez-Muñoz et al. (2004). Our aim was to test for possible structure when fish coming from different rivers were compared.

2. Methods

2.1. Taxon sampling

DNA sequences were obtained from 273 individuals (GenBank Accession Nos. from EF565470 to EF565742; for collection site locations, see Fig. 1 and Table 3 in supplementary material). Amocoetes were collected by electrofishing, while tissues of adults were acquired from local fisherman that operates at river mouths.

2.2. DNA procedures

Total genomic DNA was extracted from fin clips or pieces of muscle preserved in ethanol by an SDS/proteinase-k based protocol (Sambrook et al., 1989). A total of 624 bp of the control region (d-loop) was amplified using the primers LampFor 5′-ACACCCAGAAACA GCAACAAA-3′ and LampRev 5′-GCTGGTTTACAAGACCAGTGC-3′ designed for this study. The amplification process was conducted as follows: 4 min at 94 °C, 30 cycles of [94 °C (1 min), 55 °C (1 min) and 72 °C (1 min)], 10 min at 72 °C. Sequencing reactions were performed by Macrogen Inc.
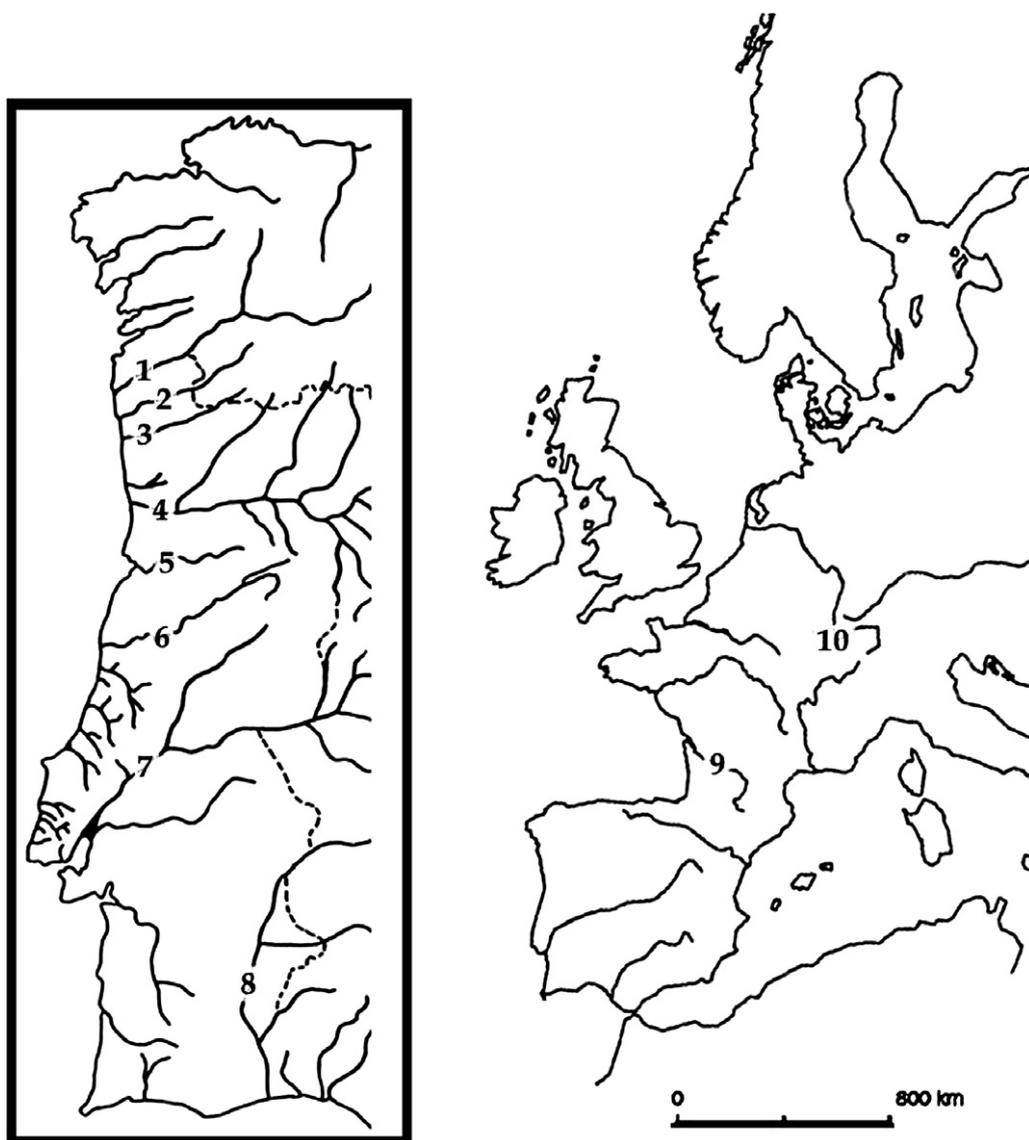

Fig. 1. European collection sites for specimens included in this study. Labels for collecting sites in the map: (1) Minho, (2) Lima, (3) Cavado, (4) Douro, (5) Vouga, (6) Mondego, (7) Tejo, (8) Guadiana, (9) Garonne, (10) Rhine.

2.3. Data analysis

Sequences were aligned with ClustalX (Thompson et al., 1997). Phylogenetic analysis was performed with PAUP 4.0 (Swofford, 2003) using maximum parsimony (MP). Bootstrap analysis was used to assess the relative robustness of branches of the MP tree (1000 replicates) (Felsenstein, 1985). *Lampetra fluviatilis* (Linneaeus, 1758) was used as outgroup in all analyses (GenBank Accession No. Y18683). Relationships among haplotypes were analyzed with a parsimony network estimated by the software TCS Version 1.18 (Clement et al., 2000).

ARLEQUIN software package Version 3.01 (Excoffier, 2005) was used to estimate genetic diversity indices, to access population differentiation, to perform neutrality tests and analysis of molecular variance (AMOVA; Excoffier et al., 1992). An analysis of the mismatch distribution (Rogers, 1995) was performed to test sudden expansion and spatial expansion models. Mean numbers of inter-population pairwise differences were corrected by subtracting the average within population differences for the populations of each pair. To allow the use of already published data, when comparing genetic diversity between North America and Europe, sequences were truncated and aligned with homologous sequences from two localities in Spain (rivers Minho and Cela) (Rodriguez-Muñoz et al., 2004) and nine rivers from the North-western Atlantic (Waldman et al., 2006). The haplotype diversity indices of North American and European drainages were compared with a Mann–Whitney U test with STATISTICA 6.0 (StatSoft, 2003). In all other calculations, the non-truncated Versions of our sequences were used.

In the absence of a calibrated molecular clock for the control region of the sea lamprey, we tentatively used a value derived from the recent literature on fish. Several studies found that the mitochondrial genome of fishes as a whole typically evolve at a divergence rate around 2% per million years. We adopted this value, although aware that most likely it is a conservative one, as the control region tends to evolve at a particularly fast rate when compared with other regions of the mitochondrial DNA. Effective population size (N_e) was inferred given the estimate of diversity index ($\theta = 2N_e\mu$).

Using empirically derived nucleotide frequencies and a single-rate model of molecular evolution, we simulated DNA matrices of identical dimensions as our empirical data, upon 1000 simulated coalescent genealogies for values of N_e from 10^4 to 10^6, using Mesquite (Maddison and Maddison, 2006), and assuming a generation time of nine years (Hardisty and Potter, 1971). In the coalescent simulations, population size was assumed to be constant through time. A scaling factor of 10^8 was selected through preliminary runs that provided a mutation rate similar to 2% per million years. From each set of simulations, we obtained the distribution of average pairwise differences (π), Watterson's nucleotide diversity (θ), and the number of segregating sites (S), using SITES (Hey and Wakeley, 1997).

3. Results

For the d-loop 63 sites were variable and six were parsimony informative. Eighteen haplotypes were found in the 273 samples studied. MP analysis resulted in six trees, with 67 steps (not shown). The MP tree lacked any noticeable structure representing a typical star-like phylogeny. There were no geographical patterns except for some very short branches which included only samples from one or a few locations.

This absence of pattern is clearly visible in the parsimony network obtained with TCS (Fig. 2). All populations are dominated by an ancestral haplotype (PMVG8) which occurs in the majority of fish samples from each population. Indeed, 78% of samples correspond to this haplotype. A total of 18 distinct haplotypes were found, with no haplotype separated by more than three mutational steps from the ancestral one. The second most common haplotype (PMVG6) was separated from the ancestral one by a single mutational step. It occurred in 11% of samples and was present in several Portuguese rivers and in the Rhine. Thus, 89% of samples correspond to these two haplotypes, for which no geographical structure is observed. Although the results of AMOVA were significant, the within population variance was overwhelmingly larger than the among population component (Groups considered: Rhine, Garonne, Minho, Lima, Cávado, Douro, Vouga, Mondego, and Tejo; among population variance = 2.67%; within population variance = 97.33%; FST = 0.03; $p < 0.05$). The
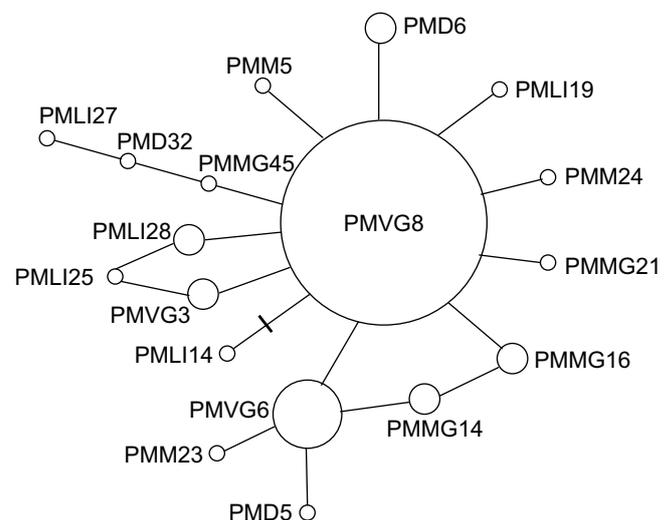

Fig. 2. TCS network representing all haplotypes obtained in this study. Labels, number of fish and distribution on each haplotype: PMVG8: 214 fish, Garonne, Cávado, Douro, Guadiana, Lima, Minho, Mondego, Rhine, Tejo and Vouga; PMVG6: 29 fish, Cávado, Douro, Lima, Mondego, Rhine, Tejo and Vouga; PMD5: 1 fish, Douro; PMD6: 4 fish, Douro; PMD32: 1 fish, Douro; PMLI14: 1 fish, Lima; PMLI19: 1 fish, Lima; PMLI25: 1 fish, Lima; PMLI27: 1 fish, Lima; PMLI28: 3 fish, Lima; PMM5: 1 fish, Minho; PMM23: 1 fish, Minho; PMM24: 1 fish, Minho; PMMG16: 5 fish, Mondego; PMMG14: 3 fish, Mondego; PMMG21: 1 fish, Mondego; PMMG45: 1 fish, Mondego; PMVG3: 4 fish, Vouga, Minho, Cávado and Lima.

genetic diversity indices, neutrality test statistics, inter-population corrected pairwise differences, and pairwise FSTs are presented in Tables 1 and 2. The data in Table 2 suggest that the results of AMOVA are, in large measure, an artefact caused by a disproportionately large sampling effort in some Portuguese rivers, which cause the detection of several rare haplotypes. In contrast, the samples from the River Garonne only contained the ancestral haplotype. Inspection of Table 1 shows that rivers as far apart as the Tagus, Minho and Rhine, did not yield significant pairwise FSTs, a finding that we take as clear evidence of lack of a global phylogeographic pattern in the data.

After our sequences were truncated, 12 haplotypes remained, which included the three haplotypes reported by Rodriguez-Muñoz et al. (2004). The rank order of haplotype frequencies between the two studies is identical. All Eastern Atlantic haplotypes are absent from Western Atlantic samples from Waldman et al. (2006), and vice versa. Overall, haplotype diversity of the Eastern Atlantic (0.289) is much lower than in the Western Atlantic (0.771); river basin haplotype diversities were significantly lower in Europe than in North America (Mann–Whitney U test, $Z = 3.78$, $n = 20$, $p < 0.001$).

When all our samples are pooled together, estimates of effective population size, assuming a mutation rate of 2%/Myr, vary from 4200 and 13,055 individuals, using the Tajima (1983) and Watterson (1975)) estimates of $\theta \pm SD$, respectively $\theta_\pi \pm SD = 0.472 \pm 0.456$; $\theta_S \pm SD = 1.455 \pm 0.559$. Empirical levels of nucleotide diversity (θ and S) are consistent with simulations assuming a 2% mutation rate and a population size (N_e) near 50,000 individuals. However, the average pairwise differences (π) of our data (0.00028) fall within the lower first percentile of simulation results.

Both Tajima's D and Fu's Fs were negative and significant (Table 2), suggesting that the population underwent a recent expansion. Mismatch analysis of the data showed a good fit between expected and observed differences for the pooled data (SSD = 0.00003129; $p = 0.9582$; Harpending's Raggedness index = 0.1645527; $p = 0.5477$), also supporting the hypothesis of a sudden expansion.

Both the sudden expansion and spatial expansion models estimate a time of expansion (τ) of 0.5 mutations/generation, or about 40,000 years ago (95% CI: 0–10,000 years). This value falls well within the Pleistocene and, given our likely underestimation of mutation rate when applied to the

Table 1
Below the diagonal: population pairwise FSTs

	Tejo	Minho	Cávado	Douro	Lima	Garonne	Guadiana	Vouga	Rhine	Mondego
Tejo	—	0.0061	−0.01833	0.00534	0.01877	0.04412	0.04412	−0.00724	−0.02778	0.02637
Minho	0.01269	—	0.00653	0.01368	0.00864	0.00753	0.00753	−0.00655	−0.00896	0.01287
Cávado	−0.04721	0.01467	—	0.00885	0.01763	0.04615	0.04615	−0.00665	−0.02393	0.02916
Douro	0.00247	0.02378	0.01389	—	0.02508	0.03754	0.03754	0.00708	−0.00511	0.03108
Lima	0.017	0.01175	0.02282	0.03363*	—	0.02553	0.02553	0.00665	0.0045	0.02808
Garonne	0.24380*	0.02517	0.17143*	0.08383*	0.04150*	—	0	0.01587	0.02222	0.02147
Guadiana	−0.625	−0.93333	−0.65714	−0.78632	−0.88194	0	—	0.01587	0.02222	0.02147
Vouga	−0.0193	−0.0179	−0.0172	0.0114	0.00738	0.08362	−0.82222	—	−0.01905	0.01353
Rhine	−0.08233	−0.02899	−0.06779	−0.02734	−0.01774	0.26241	−0.77778	−0.05733	—	0.01036
Mondego	0.05318*	0.02745	0.06014*	0.05774*	0.04812*	0.04868*	−0.82825	0.02865	0.01349	—

Distance method: pairwise difference. *Represent significant differences between populations. Above the diagonal: corrected average pairwise differences.

Table 2
Number of individuals sampled (N), number of haplotypes, number of polymorphic sites, gene diversity, mean number of pairwise differences, nucleotide diversity, Tajima's D, Tajima's Dp, Fu Fs and Fu Fsp

	N	Number of haplotypes	Number of polymorphic sites	Gene diversity	Mean number of pairwise differences	Nucleotide diversity	Tajima's D	Tajima's D p	Fu Fs	Fu Fs p
Tejo	17	2	1	0.3824 ± 0.1132	0.382353 ± 0.379406	0.000614 ± 0.000682	0.00000	1.000	0.83402	0.526
Minho	31	6	4	0.3527 ± 0.1088	0.436559 ± 0.402509	0.000701 ± 0.000719	−1.25289	0.068	−4.38603	<10⁻⁶
Cávado	26	3	2	0.4277 ± 0.0949	0.446154 ± 0.409976	0.000716 ± 0.000733	0.00000	1.000	−0.24975	0.331
Douro	37	5	4	0.4880 ± 0.0908	0.627628 ± 0.503761	0.001007 ± 0.000899	−1.14569	0.097	−1.65309	0.093
Lima	37	8	8	0.5360 ± 0.0945	0.813814 ± 0.598244	0.001306 ± 0.001067	−2.00635	0.002	−4.45570	0.002
Garonne	26	1	0	0.0000 ± 0.0000	0.000000 ± 0.000000	0.000000 ± 0.000000	0.00000	1.000	n.a.	n.a
Guadiana	1	1	0	1.0000 ± 0.0000	0.000000 ± 0.000000	0.000000 ± 0.000000	0.00000	1.000	n.a.	n.a.
Vouga	28	3	2	0.3148 ± 0.1024	0.325397 ± 0.338302	0.000522 ± 0.000605	−1.00000	1.000	−0.78960	0.180
Rhine	10	2	1	0.3556 ± 0.1591	0.355556 ± 0.375279	0.000571 ± 0.000681	−1.00000	1.000	0.41670	0.392
Mondego	60	6	4	0.3548 ± 0.0771	0.457062 ± 0.408810	0.000732 ± 0.000727	−0.63978	0.282	−3.37914	0.011
Total	273	18	13	0.3745 ± 0.0363	0.471908 ± 0.412565	0.000756 ± 0.000731	−1.93024	0.002	−21.93892	<10⁻⁶

Values are presented for each population and for all populations pooled together (Western Europe).

control region, probably after the end of the last glaciation. The two models vary however in their post-expansion population size estimates: 29,500 individuals after sudden expansion versus 431,700 following a spatial expansion (sudden expansion model $\theta_0 = 0.00$, $\theta_1 = 3.314$, $\tau = 0.516$; spatial expansion model $\theta = 0.001$, $M = 48.489$, $\tau = 0.482$). In both cases, pre-expansion populations are estimated to include only a few thousand individuals.

4. Discussion

The present results confirm and extend the pattern described by Rodriguez-Muñoz et al. (2004) to an area of Western Europe ranging from the North Sea to Portugal.

The available evidence strongly supports the conclusion that in Europe sea-lampreys form a single population, which must be recovering from a recent bottleneck. In addition, the very low genetic diversity of European sea-lamprey when compared with the Atlantic coast of North America was clearly confirmed.

The low diversity and haplotype mismatch distribution suggest that the Eastern Atlantic population may be expanding following a recent bottleneck. Our analysis suggest its effective population size may be as low as 50,000. The low average pairwise difference observed in our data relative to that expected in a sample drawn from a constant sized population of 50,000 individuals, suggests that the population is probably growing and the time of the bottleneck may be overestimated.

In the last glacial maximum and probably repeatedly during the Pleistocene, many of the areas now occupied by the sea-lamprey in Europe were glaciated (Climap, 1981), and although the waters of North-west Africa were probably of suitable temperature, few rivers were available in the region for the lampreys to spawn and grow. So, it is likely that Iberia was the centre of a much smaller refugial area where the species could survive. Unlike North America, Europe has very few rivers flowing to the South, which makes latitudinal shifts in distribution of freshwater fish much more difficult. On the other hand, it is possible that the aridity and relatively high temperature of Mediterranean rivers limited the use of the Mediterranean area as a refugium for this species.

The lack of differentiation found in European rivers is consistent with the lack of homing repeatedly found for this species (Bergstedt and Seelye, 1995) and its ability to travel many hundreds or thousands of kilometres. Fishes with some degree of homing (e.g. salmonids, Neville et al., 2006) tend to display some degree of differentiation among the populations that spawn in different rivers.

Although our results strongly suggest that the sea-lampreys of Western Europe form a single large population, it is premature to draw definitive conclusions on this issue. Data from the British Isles and Scandinavia are much needed to get a complete picture of the phylogeography of the sea-lamprey in Europe. International cooperation covering as completely as possible the full range of the species, together with additional genetic markers is needed.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online Version, at doi:10.1016/j.ympev. 2007.05.024.

References

Bergstedt, R.A., Seelye, J.G., 1995. Evidence for lack of homing by sea lampreys. Transactions of the American Fisheries Society 124, 235–239.

Bjerselius, R., Li, W.M., Teeter, J.H., Seelye, J.G., Johnsen, P.B., Maniak, P.J., Grant, G.C., Polkinghorne, C.N., Sorensen, P.W., 2000. Direct behavioral evidence that unique bile acids released by larval sea lamprey (*Petromyzon marinus*) function as a migratory pheromone. Canadian Journal of Fisheries and Aquatic Sciences 57 (3), 557–569.

Bryan, M.B., Zalinski, D., Filcek, K.B., Libants, S., Li, W., Scribner, K.T., 2005. Patterns of invasion and colonization of the sea lamprey (*Petromyzon marinus*) in North America as revealed by microsatellite genotypes. Molecular Ecology 14, 3757–3773.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9, 1657–1659.

Climap Project Members, 1981. Seasonal reconstruction for the earth's surface at the last glacial maximum, GSA Map & Chart Serv., MC-36, Boulder, Geological Society of America, Colorado.

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

Excoffier, L.G.L., Schneider, S., 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47–50.

Felsenstein, J., 1985. Confidence-limits on phylogenies—an approach using the bootstrap. Evolution 39, 783–791.

Hardisty, M.W., Potter, I.C., 1971. The general biology of adult lampreys. In: Hardisty, M.W., Potter, I.C. (Eds.), The Biology of Lampreys, vol. 1. Academic Press, New York, pp. 127–206.

Hey, J., Wakeley, J., 1997. A coalescent estimator of the population recombination rate. Genetics 145, 833–846.

Kelly, F.L., King, J.J., 2001. A review of the ecology and distribution of three lamprey species, *Lampetra fluviatilis* (L.), *Lampetra planeri* (Bloch) and *Petromyzon marinus* (L.): a context for conservation and biodiversity considerations in Ireland. Biology and Environment: Proceedings of the Royal Irish Academy 101B, 165–185.

Maddison, W.P., Maddison, D.R., 2006. Mesquite: a modular system for evolutionary analysis. Version 1.12. Available from: <http://mesquite-project.org/>.

Maitland, W.P., 2003. Ecology of the brook, river and sea lamprey. Conserving Natura 2000, Rivers Ecology Series No. 5 English Nature, Petersborough.

Neville, H.M., Isaak, D.J., Dunham, B., Thurow, F., Rieman, B.E., 2006. Fine-scale natal homing and localized movement as shaped by sex and spawning habitat in Chinook salmon: insights from special autocorrelations analysis of individual genotypes. Molecular Ecology 15, 4589–4602.

Rodriguez-Muñoz, R., Waldman, J.R., Grunwald, C., Roy, N.K., Wirgin, I., 2004. Absence of shared mitochondrial DNA haplotypes between sea lamprey from North American and Spanish rivers. Journal of Fish Biology 64, 783–787.

Rogers, A.R., 1995. Genetic evidence for a Pleistocene population explosion. Evolution 49, 608–615.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

StatSoft, I., 2003. STATISTICA (data analysis software system), Version 6. Available from: <www.statsoft.com/>.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4b.10, Sinauer Associates, Sunderland, MA.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 24, 4876–4882.

Waldman, J.R., Grunwald, C., Wirgin, I., 2006. Evaluation of the native status of sea lampreys in lake Champlain based on mitochondrial DNA sequencing analysis. Transactions of the American Fisheries Society 135, 1076–1085.

Watterson, G., 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7, 256–276.