

RESEARCH ARTICLE

European Lampreys: New Insights on Postglacial Colonization, Gene Flow and Speciation

Catarina Sofia Mateus^{1,2*}, Pedro Raposo Almeida^{1,3}, Natacha Mesquita², Bernardo Ruivo Quintella^{1,4}, Maria Judite Alves²

1 MARE – Centro de Ciências do Mar e do Ambiente, Universidade de Évora, Évora, Portugal, **2** Museu Nacional de História Natural e da Ciência & Centro de Ecologia, Evolução e Alterações Ambientais, Universidade de Lisboa, Lisboa, Portugal, **3** Departamento de Biologia, Escola de Ciências e Tecnologia, Universidade Évora, Évora, Portugal, **4** Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

* csmateus@fc.ul.pt



OPEN ACCESS

Citation: Mateus CS, Almeida PR, Mesquita N, Quintella BR, Alves MJ (2016) European Lampreys: New Insights on Postglacial Colonization, Gene Flow and Speciation. PLoS ONE 11(2): e0148107. doi:10.1371/journal.pone.0148107

Editor: Sergios-Orestis Kolokotronis, Fordham University, UNITED STATES

Received: February 5, 2015

Accepted: January 13, 2016

Published: February 12, 2016

Copyright: © 2016 Mateus et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Financial support was provided by Fundação para a Ciência e a Tecnologia (FCT), FEDER, COMPETE, and QREN through project funding (PTDC/BIA-BDE/71826/2006), the pluriannual funding program to the Center of Oceanography (PEst-OE/MAR/UI0199/2011) and by National Funds through FCT - Foundation for Science and Technology under the Project UID/MAR/04292/2013. CSM and NM were supported by FCT grants SFRH/BD/44029/2008 and SFRH/BPD/14557/2003, respectively. The funders had no role in study design,

Abstract

Ice ages are known to be the most dominant palaeoclimatic feature occurring on Earth, producing severe climatic oscillations and consequently shaping the distribution and the population structure of several species. Lampreys constitute excellent models to study the colonization of freshwater systems, as they commonly appear in pairs of closely related species of anadromous *versus* freshwater resident adults, thus having the ability to colonize new habitats, through the anadromous species, and establish freshwater resident derivatives. We used 10 microsatellite loci to investigate the spatial structure, patterns of gene flow and migration routes of *Lampetra* populations in Europe. We sampled 11 populations including the migratory *L. fluviatilis* and four resident species, *L. planeri*, *L. alavariensis*, *L. auremensis* and *L. lusitanica*, the last three endemic to the Iberian Peninsula. In this southern glacial refugium almost all sampled populations represent a distinct genetic cluster, showing high levels of allopatric differentiation, reflecting long periods of isolation. As result of their more recent common ancestor, populations from northern Europe are less divergent among them, they are represented by fewer genetic clusters, and there is evidence of strong recent gene flow among populations. These previously glaciated areas from northern Europe may have been colonized from lampreys expanding out of the Iberian refugia. The pair *L. fluviatilis*/*L. planeri* is apparently at different stages of speciation in different locations, showing evidences of high reproductive isolation in the southern refugium, and low differentiation in the north.

Introduction

The Quaternary climatic oscillations and geographic restrictions imposed by the impassable glaciated areas are thought to have had major effects on the evolution and dispersal of various

data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

species (e.g. [1,2]). It is now clear that most fauna and flora presently distributed across Europe were isolated in southern refugia during the glacials, many in the Mediterranean peninsulas of Iberia, Italy and the Balkans [3]. After the glacials, and as the climate warmed rapidly, founder populations at the northern limits of the southern refugia expanded northwards, into the new available habitats, leading to a reduction from southern to northern Europe in the extent of the number of species, subspecific differentiation and allelic variation [3].

Recently deglaciated regions were relatively inaccessible to freshwater fishes, as they do not normally disperse among river basins; they were, however, easily reached by anadromous fishes (migratory species that reproduce in freshwater before migrating to the sea where they grow to the adult stage), which could access isolated basins *via* sea. These fish breed in fresh water, having ample opportunities to colonize these unexploited systems and establish freshwater isolates [4]. In some genera of lampreys, closely related species show divergent life histories: parasitic and anadromous vs. non-parasitic and freshwater resident; these species are called “paired species”, and the freshwater resident (brook) species have apparently evolved from a form similar to that of an extant anadromous one [5,6]. In some cases, more than one freshwater resident species has derived from a single anadromous species; these are called “satellite species” [7]. For this reason, lampreys constitute excellent systems to study the postglacial colonization processes and emergence of freshwater derivatives from the founder anadromous forms.

The anadromous European river lamprey (*Lampetra fluviatilis*) and the resident European brook lamprey (*Lampetra planeri*) are considered paired species. They only occur in European watersheds, extending from southern Norway to the western Mediterranean and the Iberian Peninsula in the south. The three brook lampreys *Lampetra alavariensis*, *Lampetra auremensis* and *Lampetra lusitanica*, which are endemic to western Iberian Peninsula (Portugal) [8], are apparently derived from an extinct anadromous form, more ancestral than the solely extant anadromous form occurring in Iberia, *L. fluviatilis* [9]. In this region, while *L. planeri* is found in several river basins, *L. alavariensis*, *L. auremensis* and *L. lusitanica* are confined to one or two basins, and the anadromous *L. fluviatilis* currently occurs in the Tagus river basin only [8,10]. The current distribution of the extant Iberian lamprey lineages is largely allopatric and the genetic divergence between them is consistent with extended periods of isolation during survival in separate glacial refugia throughout the ice ages [9,11]. The three brook lampreys *L. alavariensis*, *L. auremensis* and *L. lusitanica* are well supported monophyletic groups, divergent from the present-day *L. fluviatilis*. However, *L. planeri* share haplotypes with the parasitic form, implying that their emergence was more recent [9,11]. The taxonomy of *L. fluviatilis* and *L. planeri* has thus been considered problematic, as studies using different markers have revealed lack of differentiation between the species (e.g. [11–13]), leaving open two possible scenarios: 1) a very recent divergence event or 2) a single species with phenotypic plasticity. The recent study of Mateus *et al.* [14], using genome-wide sequencing in sympatric populations of these species in the Iberian Peninsula, represented an important step forward in this long-standing question, as it successfully identified fixed allelic differences between the two forms in this region. These findings imply that these species are result of a recent divergence event (scenario 1), with populations in different phases of speciation across their ranges.

In this study, we investigate postglacial colonization of European fresh waters by analyzing patterns of genetic differentiation in a group with alternative life histories. We analyzed 10 polymorphic microsatellite loci, coupled with two mitochondrial genes, in the anadromous *L. fluviatilis* and in four resident species, *L. planeri*, *L. alavariensis*, *L. auremensis* and *L. lusitanica*. We tested the hypotheses that the Iberia is a source for the northward recolonization by the migratory species, the possible existence of a speciation gradient in Europe, and test for contemporary gene flow among and within species.

Materials and Methods

Ethics Statement

This study was carried out in strict accordance with the recommendations present in the Guide for the Care and Use of Laboratory Animals of the European Union—in Portugal represented by the Decree-Law n°129/92, Portaria n°1005/92. By the time the experimental work took place, the University’s ethics committee only dealt with research involving humans. Nevertheless, since two authors have an official license for animal experimentation (Category C from FELASA), issued by the Veterinary National Authority (DGV), Portuguese Ministry of Agriculture and Sea, the experiments involving living animals were performed in accordance to international rules regarding animal welfare. Collection of samples in Portugal was carried out without euthanizing the specimens, using the collecting permits provided by the Institute for Nature Conservation and Forestry, I.P. (ICNF, IP). In Belgium and Germany sampling authorizations are provided by Gouvernement Wallon—*Travaux publics, Agriculture, Ruralité, Nature, Forêt et Patrimoine*, and the German Federal Agency for Nature Conservation, respectively. In Finland, the samples have been collected from commercial lamprey catches, so those animals were caught for human consumption, and no authorization was needed.

Sampling and DNA extraction

Lampreys were collected by electro fishing, and after being anaesthetized by immersion in 2-phenoxyethanol (0.3 ml L⁻¹), a piece of tissue was removed from the dorsal fin. After recovery individuals were released near the capturing sites. Sampled species were the European river lamprey *L. fluviatilis*, the European brook lamprey *L. planeri*, and the three recently described Iberian brook lampreys *L. alavariensis*, *L. auremensis* and *L. lusitanica* [8]. Ten sites were sampled, with only one species being present in each site, with the exception of the Sorraia River in the Tagus Basin (Iberian Peninsula), where *L. fluviatilis* and *L. planeri* are found in sympatry (Table 1). In the tables and across the manuscript, acronyms are labelled so that patterns are more readily understood: acronyms followed by *m* refer to the migratory species (*L. fluviatilis*), and by *r* to resident species (the brook lampreys *L. planeri*, *L. alavariensis*, *L. auremensis* and *L. lusitanica*). When treated together, populations from Belgium, Germany and Finland are

Table 1. Locations listed north to south, acronyms and sample sizes (n) of *Lampetra* samples included in the study.

Country	Basin	River	Acronym	n	Species
Finland	Lestijoki	Lestijoki	LEST _m	29	<i>L. fluv</i>
Germany	Warnow	Beke	BEKE _r	30	<i>L. plan</i>
Germany	Elbe	Schaale	ELBE _m	40	<i>L. fluv</i>
Belgium	Meuse	Warche	WARC _r	35	<i>L. plan</i>
Portugal	Esmoriz	Esmoriz	ESM _r	33	<i>L. alavar</i>
Portugal	Lis	Lis	LIS _r	33	<i>L. plan</i>
Portugal	Ribeiras do Oeste	Ribeira de S. Pedro	OES _r	31	<i>L. plan</i>
Portugal	Tagus	Nabão	NAB _r	35	<i>L. aurem</i>
Portugal	Tagus	Sorraia ¹	SPL _r	52	<i>L. plan</i>
Portugal	Tagus	Sorraia ¹	SFL _m	46	<i>L. fluv</i>
Portugal	Sado	Marateca	SADO _r	51	<i>L. lusit</i>

Species: *L. plan*, *L. planeri*; *L. fluv*, *L. fluviatilis*; *L. alavar*, *L. alavariensis*; *L. aurem*, *L. auremensis*; *L. lusit*, *L. lusitanica*

Acronyms followed by *m* refer to migratory species (*L. fluviatilis*), and by *r* to resident species (*L. planeri*, *L. alavariensis*, *L. auremensis* and *L. lusitanica*).

¹Location where the paired *L. fluviatilis* and *L. planeri* occur in sympatry.

doi:10.1371/journal.pone.0148107.t001

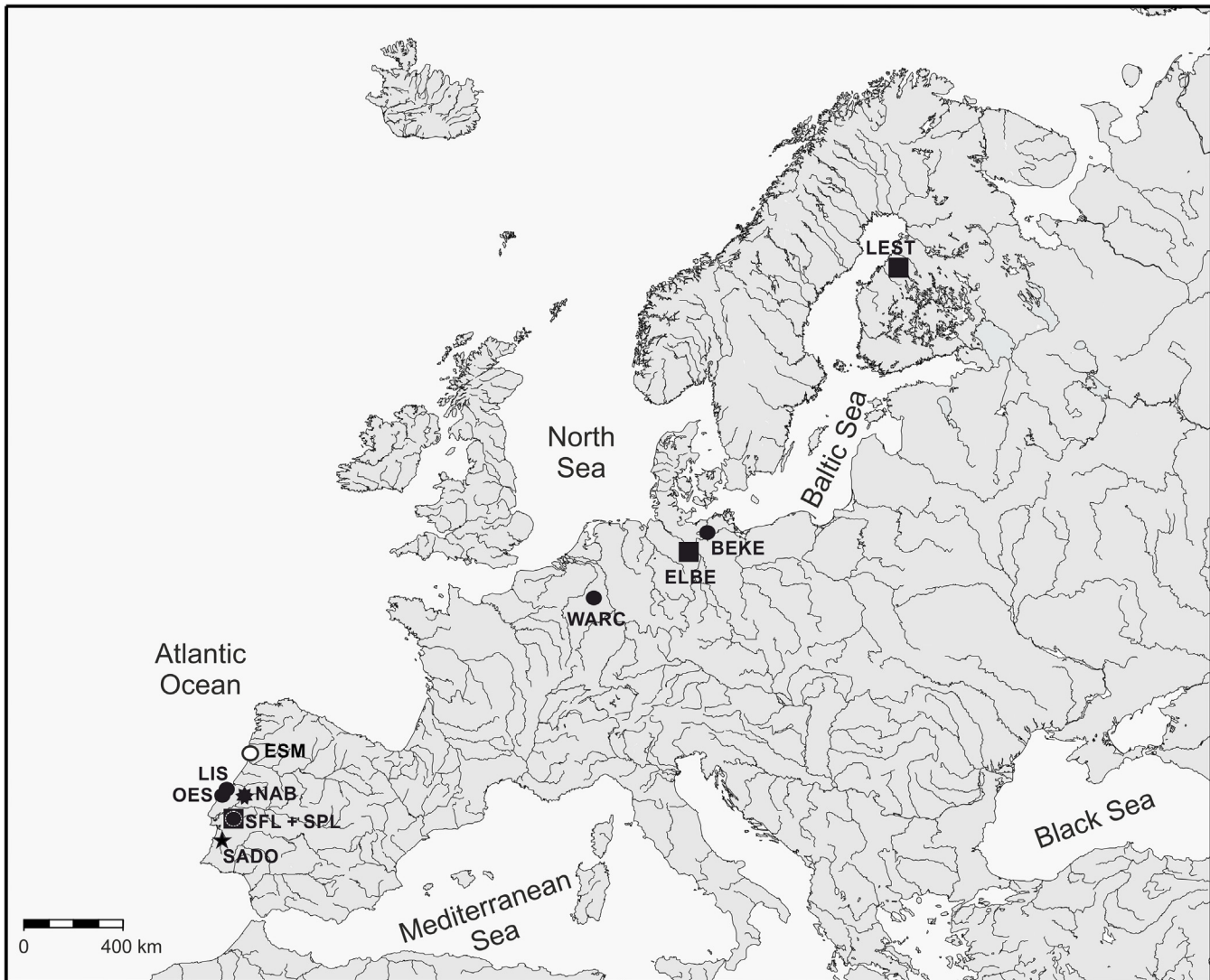


Fig 1. Sampling sites of *Lampetra* populations in Europe. Species: ■, *L. fluviatilis*; ●, *L. planeri*; ○, *L. alavariensis*; ★, *L. lusitanica*; ◆, *L. auremensis*. See [Table 1](#) for details about species and sampling sites.

doi:10.1371/journal.pone.0148107.g001

hereinafter referred as “northern populations” and populations from the Iberian Peninsula as “southern populations”. All rivers sampled in the Iberian Peninsula drain to the Atlantic Ocean, rivers Warche (Meuse basin) and Schaale (Elbe basin) drain to the North Sea, whereas rivers Beke (Warnow basin) and Lestijoki drain to the Baltic Sea ([Fig 1](#)).

Total genomic DNA was extracted following a standard phenol-chloroform protocol [[15](#)] and stored at -20°C. DNA concentration was measured using a Thermo Scientific NanoDrop™ 1000 Spectrophotometer and standardized to 50 ng μl⁻¹ per sample.

Microsatellite analysis

A total of 415 specimens from 10 sites were used in the analysis, with sample sizes ranging from 29 to 52 ([Table 1](#)). Initially, 49 microsatellite primer sets developed for other lamprey species (*Petromyzon marinus*: [[16,17](#)]; *Lethenteron* sp. N: [[18](#)]; *Ichthyomyzon unicuspis* and *Ichthyomyzon fossor*: [[19](#)]; *Lampetra richardsoni*: [[20](#)]) were screened using the described protocols and

were further optimized to the target species. Ten primer sets produced unambiguously determined bands and were polymorphic: Iun 2, Iun 5, Iun 7, Iun 10 and Iun 14 [19]; Lspn 010–2, Lspn 019c, Lspn 044 and Lspn 094 [18]; and Pma μ 5 [16]. These 10 loci were used to genotype 415 individuals; all others were rejected. The reverse primers were 5'-labelled with 6-FAM, NED, PET or VIC (Applied Biosystems[®]) fluorescent dyes. Primer sets were grouped into three multiplex reactions (S1 Table), and polymerase chain reactions (PCR) were set up in 12 μ L volumes containing 2 μ L of 50 ng μ L⁻¹ genomic DNA, 1.0 to 3.0 mM MgCl₂, 0.2 mM dNTP mix, 0.5 μ M for each primer, 1 unit of DreamTaq[™] DNA Polymerase (Fermentas) and 1 \times DreamTaq[™] Buffer. PCR conditions were as follows: initial denaturation at 94°C for 1 min, followed by 25 cycles of 30 sec at 94°C, annealing for 30 sec at temperatures ranging from 55 to 60°C and 30 sec at 72°C, and a final extension of 7 min at 72°C. A number of sets of difficult amplification were completed using a Multiplex PCR Kit (Qiagen[®]) with 5 μ L Qiagen Multiplex PCR master Mix, 3 μ L RNase-free water, 1 μ L Primer Mix (2 μ M each primer) and 1 μ L of 50 ng μ L⁻¹ of genomic DNA, using the following protocol: initial activation step at 95°C for 15 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 57°C for 90 sec and extension at 72°C for 60 sec, and a final extension of 30 min at 60°C. The PCR reactions were conducted on a Bio-Rad[®] thermal cycler.

Samples were genotyped in an ABI PRISM[®] 310 Genetic Analyzer and fragments were sized with GeneScan[™]-500 LIZ[™] Size Standard. Allele sizes were determined using the software GeneMapper[®] 3.7 (Applied Biosystems[®]).

Microsatellite loci were first tested for null alleles, stuttering and large allele dropout using the software MICRO-CHECKER 2.2.3 [21]. Each microsatellite locus was tested for Hardy–Weinberg equilibrium and departures from linkage equilibrium were assessed for all pairs of loci in each population with 10⁴ permutations, as implemented in ARLEQUIN 3.11 [22]. Genetic diversity was measured as the mean allelic richness (AR), observed heterozygosity (Ho), unbiased expected heterozygosity (He, *sensu* Nei 1978) and mean number of alleles across loci (MNA), inferred using GENETIX 4.05.2 [23], with the exception of allelic richness, which was calculated and corrected for sample dimension by rarefaction using HP-Rare [24].

The genetic differentiation among samples was assessed through pairwise F_{ST} using the Weir & Cockerham's estimator [25], and significance was assessed with 10⁴ permutations, as implemented in GENETIX. The distribution of genetic variation was assessed among and within the 11 samples, the sympatric *L. fluviatilis* and *L. planeri*, and the genetic clusters attained with population structure analysis, through analysis of molecular variance (AMOVA) [26]. These analyses were performed in ARLEQUIN, using the allelic frequencies as the genetic distance and 10⁴ permutations.

The Bayesian model-based clustering approach implemented in STRUCTURE 2.2 [27] was used to assemble individuals into groups (genetic clusters). Runs were performed under the admixture model, with correlated allelic frequencies, and with a number of groups (K) set between 1 and 12. For each K, 10 simulations were performed with a burn-in period of 10⁵, followed by 5 million Markov steps. Using the same parameters, two additional structure analyses were performed, one including solely the eight samples of *L. planeri* and *L. fluviatilis* (K between 1 and 9), and the other including the four samples from the North (K between 1 and 5). These allow detection of further structure in these populations, if present, that otherwise would be hidden due to the high differentiation among the five species, and between the northern and southern samples. The optimal K, and clustering achieved, was inferred using the protocol defined by Evanno *et al.* [28] as implemented in STRUCTURE HARVESTER 0.6.93 [29], and taking into account the biological meaning of the clusters. The software DISTRUCT 1.1 [30] was used for the graphical display of population clusters.

The software GeneClass2 2.0.h [31] was used to test the assignment of individuals to their sampling sites, including a likelihood-based method in which individuals are assigned to the

locality in which the individual's genotype is most likely to occur. The Bayesian statistical approach of Rannala & Mountain [32] was implemented. GeneClass2 2.0.h was also used to determine whether our samples might contain individuals that were first generation (F_0) immigrants from unsampled populations (the so-called 'ghost populations', [33]). We used the Bayesian assignment procedure of Rannala & Mountain [32], and the Paetkau *et al.* method [34] to compute probabilities from 10,000 simulated genotypes.

Patterns of differentiation were visualized by principal coordinates analysis (PCoA), a multivariate technique that allows to find and plot the major patterns within a multivariate dataset, like multiple loci and multiple samples. This analysis was computed using GenAlEx 6.5 [35,36].

The software NewHybrids 1.1 [37] was used for the detection and classification of putative hybrids between sympatric populations of *L. fluviatilis* and *L. planeri* from Portugal. NewHybrids uses a Bayesian approach to identify different categories of hybrid individuals through the computation of the posterior probability that individuals fall into different hybrid (F_1 , F_2 and backcrosses) or pure parental categories. It uses the allele frequencies of multilocus genotypes and a Markov Chain Monte Carlo procedure. Simulations were performed with a burn-in period of 10^5 , followed by a sampling period of 10^5 Markov steps. A threshold of posterior probability > 50% was set up to classify an individual as belonging to a certain category.

Estimates of recent migration rates (m) between migratory populations were inferred using a Bayesian assignment test-based method in the program BAYESASS 3.0.1 [38]. We also performed migration rate analysis between all populations, whose results are presented in [S2 Table](#). BAYESASS estimates migration rates over the last two generations using a Markov chain Monte Carlo procedure and does not assume that populations are in migration-drift or Hardy–Weinberg equilibrium. Because BAYESASS focuses on contemporary migration rates, estimates are unaffected by the colonization processes. A total of 10^7 MCMC iterations (discarding the first 10^6 iterations as burn-in) were performed, and samples were collected every 2000 iterations. The convergence and stability of the MCMC algorithm was checked by visual inspection of plotted posterior parameter estimates using the software Tracer 1.6 [39] ([S1 Fig](#)). Delta values for migration rate, allele frequencies, and inbreeding values coefficients were set at 0.20, 0.40 and 0.60, respectively.

Demographic signatures of recent bottlenecks were tested using the heterozygosity excess method implemented in BOTTLENECK 1.2.02 [40] under three different mutational models: infinite allele model (IAM), stepwise mutation model (SMM) and two-phase model (TPM). Significant deviations from mutational-drift equilibrium were tested using the Wilcoxon sign rank test with 10^5 simulations, and the distribution of allele frequency classes was examined for a deviation from the normal L-shaped distribution [41]. Past reductions in population size were also evaluated using the M ratio ($M = k/r$) statistic test as implemented in M_P_VAL [42], where k is the number of alleles present at a given microsatellite locus and r is the overall range in allele size. In recently reduced populations M is expected to be smaller than in populations at equilibrium, since the loss in any allele will contribute to a reduction in k , whereas only a loss of the smallest or largest alleles will contribute to a reduction in r , and thus k is expected to decrease more quickly than r . Reductions in population size were considered significant if less than 5% of the replicates are below the observed M value. Following Garza & Williamson (2001) [42], we used the default settings for the two-phase mutation model (TPM) $p_s = 0.9$, $\Delta_g = 3.5$ and three values of θ ($\theta = 4$, $\theta = 10$ and $\theta = 20$), where $\theta = 4Ne\mu$ (Ne = effective population size; μ = mutation rate), p_s is the proportion of one-step mutations, and Δ_g is the average size of non one-step mutations.

Populations LIS and OES were not included in the demographic analysis because they present only one polymorphic locus.

MtDNA sequence analysis

For a better understanding of the historical processes that may underlie present population structure, we performed a phylogenetic analysis of mtDNA sequences of the populations included in this study, and additional populations from across the species distributional range.

A total of 44 samples from northern Europe (LEST_m, BEKE_r, ELBE_m, WARC_r) were amplified and sequenced for both cytochrome *b* and ATPase 6/8 mitochondrial genes, following the protocol in Mateus *et al.* [9]. For the remaining populations included in the microsatellite analysis, the corresponding haplotypes were obtained from the GenBank nucleotide data base [9,11], as well as additional haplotypes which represent other populations from across the distributional range of *L. planeri* and *L. fluviatilis*. MtDNA sequencing of 44 individuals from LEST_m, BEKE_r, ELBE_m and WARC_r, revealed a total of 14 new haplotypes (H83 to H96), which combined with the 52 haplotypes from GenBank [9,11] resulted in a dataset of 66 haplotypes for phylogenetic analyses. New nucleotide sequences are available at the GenBank database under the accession numbers KT275288- KT275301.

DNA sequences were manually aligned and edited using Sequencher V4.8 (Gene Codes Corp., Ann Arbor, MI, USA). One sequence each of the species *L. alavariensis*, *L. auremensis* and *L. lusitanica* was included and used as outgroup (haplotypes H26, H48 and H53, respectively; [9,11]).

The phylogenetic relationships among mtDNA haplotypes were reconstructed using PAUP* [43] by the neighbour-joining algorithm [44], according to the selected model and using the parameter settings as estimated with MODELTEST software [45]. Bootstrap support values in neighbour-joining analysis were computed with fast stepwise addition using 1000 pseudo-replicates. For maximum parsimony, bootstrap values were computed with full heuristic search using 500 replicates, each consisting of 100 TBR searches of random addition sequence.

Results

Genetic diversity and differentiation

For microsatellites, signs of null alleles were detected with MICRO-CHECKER for a few situations across loci and populations: Pma μ 5 in SPL_r and NAB_r; Iun 10 in SPL_r and BEKE_r; Iun 5 in ELBE_m and BEKE_r; and Iun 14 in SPL_r and BEKE_r. Significant null alleles' signature is related with heterozygote deficit and therefore with deviations from Hardy-Weinberg equilibrium, as seen in S2 Table.

The summary statistics of the genetic diversity indices for each locus and sample are provided in S2 Table. The total number of alleles per locus across populations varied from two, at the loci Iun7 and Lspn010-2, to 13 at the locus Iun14. Twelve private alleles were found: NAB_r, SPL_r and SADO_r showed three unique alleles each, BEKE two, and SFL one (S2 Table). The mean number of alleles (MNA) across loci ranged from 1.1 (LIS_r and OES_r) to 3.8 (ELBE_m), mean allelic richness (AR) from 1.08 (LIS_r) to 2.62 (ELBE_m), and expected heterozygosity (He) from 0.0239 (LIS_r) to 0.4417 (NAB_r) (S2 Table).

A considerable level of genetic differentiation among samples was observed (average $F_{ST} = 0.498$, $P < 0.001$) with pairwise F_{ST} values ranging from 0.0114 (ELBE_m-LEST_m) to 0.8915 (OES_r-ESM_r), all being significant ($P < 0.001$ for all pairs, with the exception of ELBE_m-LEST_m where $P < 0.05$) (Table 2).

AMOVA analysis among the 11 samples indicated that 48.34% of genetic variance occurred among samples ($P < 0.001$); AMOVA between the sympatric paired *L. fluviatilis*/*L. planeri* from the Tagus basin revealed that 31.64% of the variance was significantly ($P < 0.001$)

Table 2. Pairwise estimates of genetic differentiation (F_{ST}) among sites (above diagonal) and corresponding P values (below diagonal). For populations' acronyms, please check [Table 1](#).

	LEST _m	BEKE _r	ELBE _m	WARC _r	ESM _r	LIS _r	OES _r	NAB _r	SPL _r	SFL _m	SADO _r
LEST _m	-	0.0519	0.0114	0.2849	0.5035	0.6642	0.6736	0.3820	0.2602	0.1463	0.4557
BEKE _r	***	-	0.0530	0.3963	0.5301	0.6959	0.7104	0.3989	0.2919	0.2378	0.5243
ELBE _m	0.04	***	-	0.2486	0.4644	0.6269	0.6402	0.3712	0.2330	0.1100	0.4127
WARC _r	***	***	***	-	0.5692	0.7047	0.6974	0.4464	0.3408	0.3704	0.6056
ESM _r	***	***	***	***	-	0.8877	0.8915	0.5472	0.3697	0.5131	0.7529
LIS _r	***	***	***	***	***	-	0.6166	0.5820	0.5656	0.6897	0.8396
OES _r	***	***	***	***	***	***	-	0.5839	0.5887	0.6996	0.8291
NAB _r	***	***	***	***	***	***	***	-	0.3909	0.4423	0.6273
SPL _r	***	***	***	***	***	***	***	***	-	0.3167	0.4931
SFL _m	***	***	***	***	***	***	***	***	***	-	0.3989
SADO _r	***	***	***	***	***	***	***	***	***	***	-

*** $P < 0.001$.

doi:10.1371/journal.pone.0148107.t002

explained among species; and AMOVA among the eight genetic groups attained with STRUCTURE (see below and [Fig 2A](#)) indicated that the majority of variance occurs among groups (46.79%) and within samples (48.89%), whereas variance among samples within groups is low (4.32%) ($P < 0.001$ for all the three levels).

Population structure and admixture

The STRUCTURE analysis using the microsatellite data revealed that the 11 samples are grouped in eight distinct genetic clusters: 1) LEST_m+BEKE_r+ELBE_m, 2) WARC_r, 3) ESM_r, 4) LIS_r+OES_r, 5) NAB_r, 6) SPL_r, 7) SFL_m and 8) SADO_r ([Fig 2A](#) and [Table 3](#)). The first group exhibits strong evidence of admixture with the resident *L. planeri* from Meuse (WARC_r, group 2) and *L. fluviatilis* from Tagus (SFL_m, group 7), with a greater number of admixed individuals in the anadromous populations (LEST_m and ELBE_m) ([Fig 2](#)). Most individuals of group 2 (WARC_r) are distinct and constitute a distinct genetic cluster with high proportion of membership (0.949; [Table 3](#)). This is also the case in other groups comprised of resident species, namely, ESM_r, LIS_r+OES_r, NAB_r and SADO_r (groups 3, 4, 5, and 8, respectively; [Fig 2A](#) and [Table 3](#)). The sympatric paired *L. planeri* and *L. fluviatilis* from Tagus (SPL_r and SFL_m) constitute two distinct genetic clusters (6 and 7, respectively) that present a few admixed individuals between them. SPL_r also shows some evidence of admixture with ESM_r ([Fig 2A](#) and [Table 3](#)). When STRUCTURE was run only for the eight samples of *L. planeri* and *L. fluviatilis* ([Fig 2B](#), $K = 5$), and for the four northern samples ([Fig 2C](#), $K = 2$), no additional genetic clusters were achieved, indicating that there is no hidden structure caused by the high differentiation of the Iberian endemics, and that the genetic cluster that groups the northern populations (LEST_m+BEKE_r+ELBE_m) is well supported.

Individual assignment tests were applied to further investigate the genetic distinctiveness of the populations. In four populations of resident species (WARC_r, ESM_r, NAB_r and SADO_r) 100% of the individuals were assigned to their correct source population ([Table 4](#)), which is in agreement with the STRUCTURE analysis. Samples from northern Europe were the ones with more individuals assigned to other populations; *L. fluviatilis* from Lestijoki and Elbe (LEST_m and ELBE_m, respectively) and the resident *L. planeri* from Warnow (BEKE_r) had individuals assigned among the three populations, in agreement with STRUCTURE, in which the three populations form a distinct genetic cluster (see above). The sympatric *L. fluviatilis* and *L.*

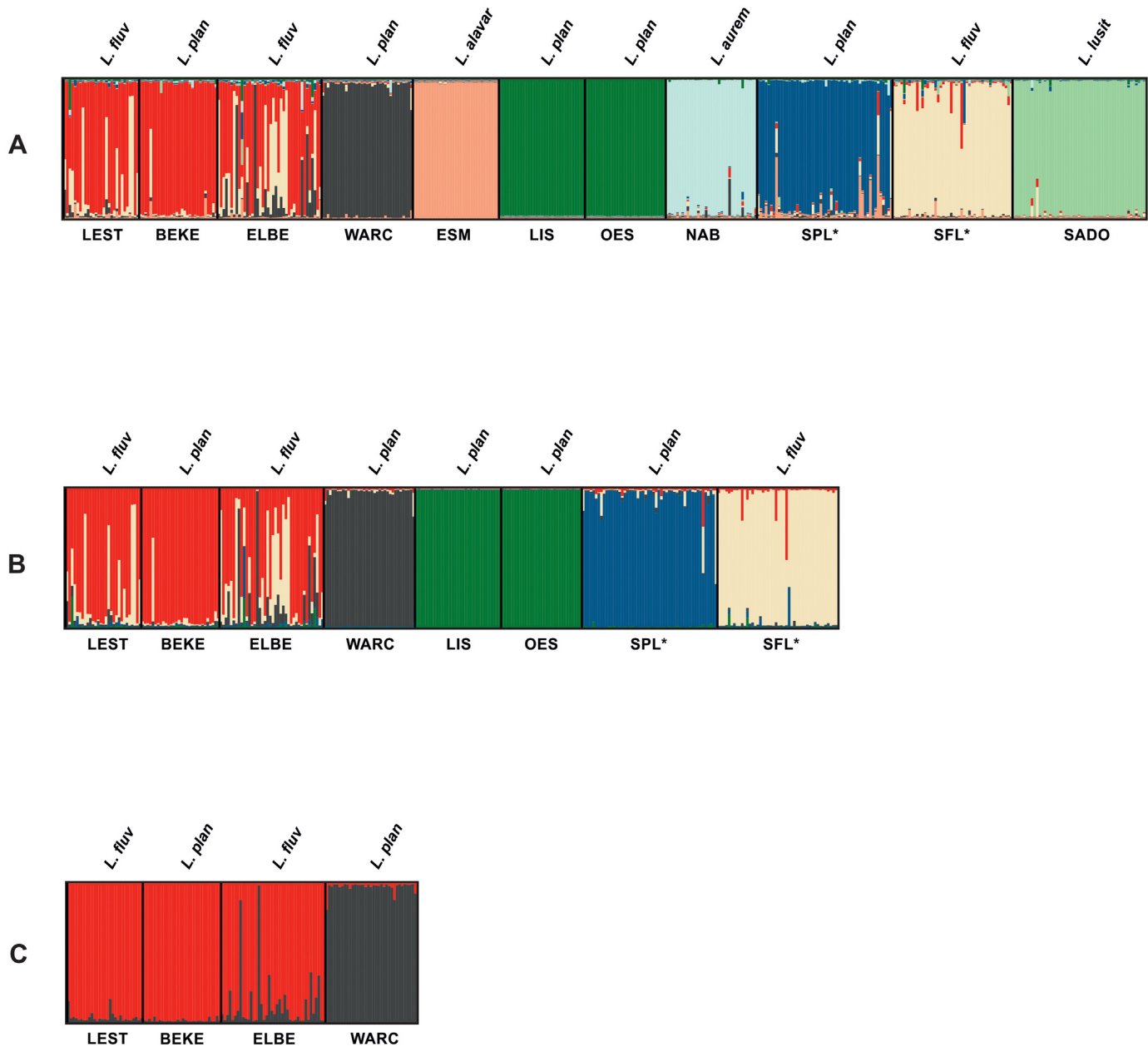


Fig 2. Most likely population structure, computed under the admixture model with correlated allelic frequencies in STRUCTURE, considering: A) all the 11 samples, K = 8; B) the populations of *L. planeri* and *L. fluviatilis*, K = 5; and C) the four northern populations, K = 2. Each individual is represented by a vertical bar. Sampled locations below plot and corresponding *Lampetra* species above are in accordance with Fig 1 and Table 1.

doi:10.1371/journal.pone.0148107.g002

planeri (SFL_m and SPL_r) had almost all individuals assigned correctly (96% and 98%), and small percentages (4% and 2%) assigned between them. LIS_r and OES_r showed 73% and 97%, respectively, of correctly assigned individuals, and the remaining were assigned also among each other (Table 4). This last result is consistent with the STRUCTURE and PCoA analyses, which revealed a close genetic relation between these two populations (Figs 2 and 3). Using a critical *P* of 0.01 for first-generation (*F*₀) immigrant analysis, a total of seven individuals had a probability below the threshold value, but all were assigned to a sampled location: one individual of LEST_m was assigned to BEKE_r; one of BEKE_r was assigned to ELBE_m; one individual of

Table 3. STRUCTURE analysis for the 11 samples. Proportion of membership of each pre-defined population in each of the eight genetic clusters. For populations' acronyms, please check [Table 1](#).

Population	Inferred clusters							
	1	2	3	4	5	6	7	8
LEST _m	0.734	0.028	0.015	0.013	0.007	0.010	0.189	0.005
BEKE _r	0.917	0.008	0.017	0.006	0.008	0.006	0.032	0.007
ELBE _m	0.552	0.135	0.016	0.012	0.014	0.017	0.242	0.011
WARC _r	0.009	0.949	0.009	0.003	0.006	0.005	0.014	0.004
ESM _r	0.003	0.003	0.977	0.003	0.003	0.004	0.005	0.003
LIS _r	0.003	0.002	0.003	0.981	0.003	0.003	0.003	0.002
OES _r	0.002	0.002	0.002	0.983	0.003	0.002	0.002	0.003
NAB _r	0.008	0.018	0.009	0.007	0.943	0.006	0.005	0.003
SPL _r	0.016	0.012	0.052	0.006	0.009	0.860	0.038	0.006
SFL _m	0.035	0.009	0.016	0.007	0.004	0.016	0.902	0.012
SADO _r	0.007	0.004	0.005	0.006	0.004	0.005	0.013	0.957

Clusters: 1, LEST_m+BEKE_r+ELBE_m; 2, WARC_r; 3, ESM_r; 4, LIS_r+OES_r; 5, NAB_r; 6, SPL_r; 7, SFL_m; 8, SADO_r

Values in bold represent the proportion of membership in the dominant genetic cluster.

doi:10.1371/journal.pone.0148107.t003

ELBE_m was assigned to BEKE_r and another to WARC_r; one individual of WARC_r was assigned to ELBE_m; one individual of SPL_r was assigned to SFL_m; and one individual of SFL_m was assigned to LEST_m.

The principal coordinates analysis (PCoA), revealed the existence of mainly six distinct clusters, NAB_r, SADO_r, ESM_r, LIS_r+OES_r, SPL_r+WARC_r, BEKE_r+LEST_m+ELBE_m+SFL_m ([Fig 3](#)). These results are congruent with STRUCTURE, with the exception that it groups SPL_r with WARC_r, and SFL_m with the northern cluster, while in STRUCTURE SPL_r, WARC_r, SFL_m and the northern populations BEKE_r, LEST_m and ELBE_m constitute four distinct groups.

Table 4. Assignment tests performed with GeneClass2. Values represent the percentage of individuals from each studied sample assigned to each of the sampled populations based on the Bayesian method. For populations' acronyms, please check [Table 1](#).

	Assigned population										
	LEST _m	BEKE _r	ELBE _m	WARC _r	ESM _r	LIS _r	OES _r	NAB _r	SPL _r	SFL _m	SADO _r
LEST _m	80	3	17	0	0	0	0	0	0	0	0
BEKE _r	3	94	3	0	0	0	0	0	0	0	0
ELBE _m	10	15	68	5	0	0	0	0	0	2	0
WARC _r	0	0	0	100	0	0	0	0	0	0	0
ESM _r	0	0	0	0	100	0	0	0	0	0	0
LIS _r	0	0	0	0	0	73	27	0	0	0	0
OES _r	0	0	0	0	0	3	97	0	0	0	0
NAB _r	0	0	0	0	0	0	0	100	0	0	0
SPL _r	0	0	0	0	0	0	0	0	98	2	0
SFL _m	0	0	0	0	0	0	0	0	4	96	0
SADO _r	0	0	0	0	0	0	0	0	0	0	100

Each row contains the samples from one sampled locality and the columns indicate the localities to which the samples were assigned (i.e., in which their genotypes had the highest likelihood of occurring).

Values along the diagonal (in bold) represent the proportion of individuals assigned to the population in which they were sampled.

doi:10.1371/journal.pone.0148107.t004

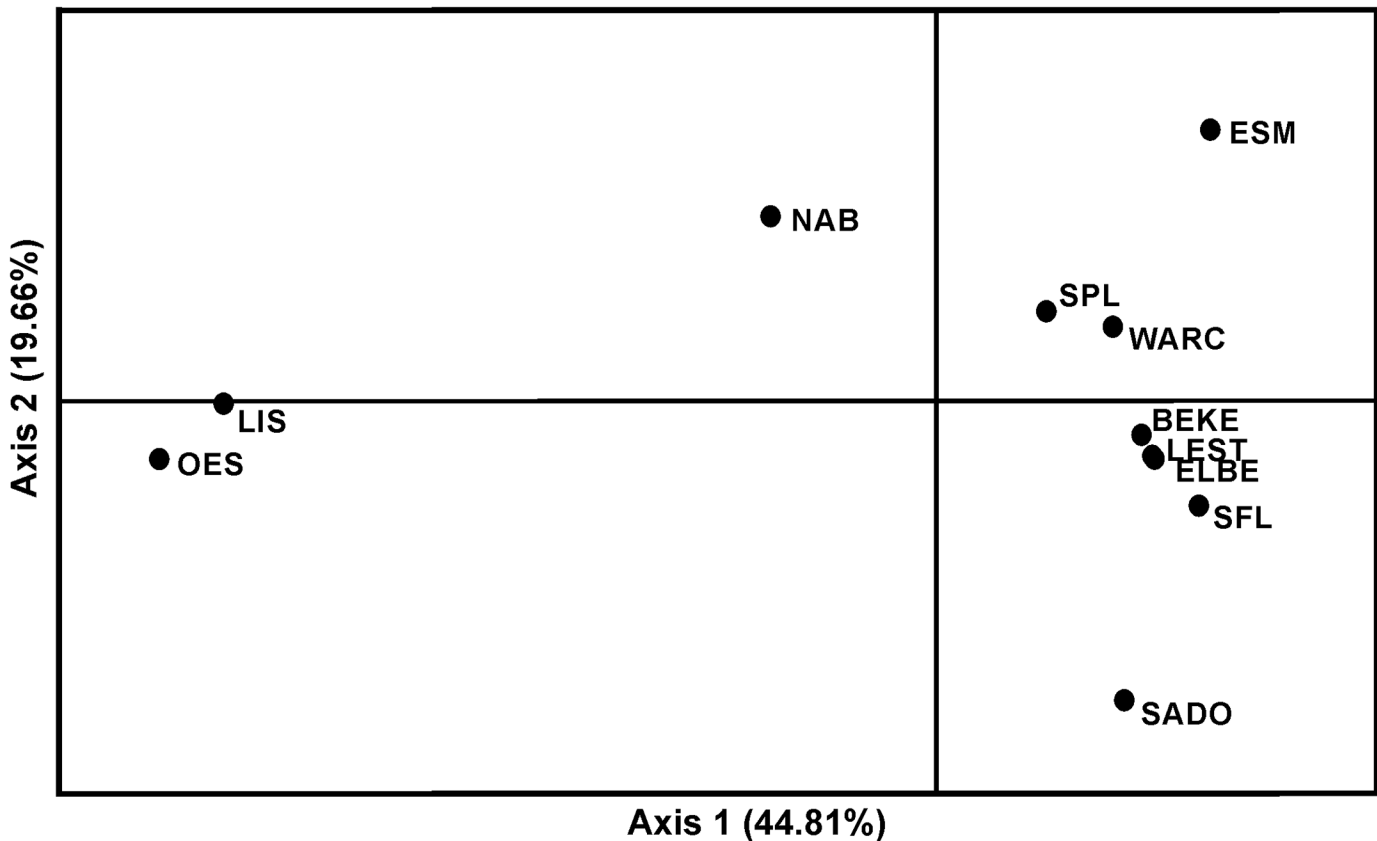


Fig 3. Principal coordinates analysis plot (PCoA) computed by GenAlEx. The percentage of variation explained by each axis is shown. Samples' acronyms as in Fig 1 and Table 1.

doi:10.1371/journal.pone.0148107.g003

The neighbour-joining phylogenetic tree attained from mtDNA sequences is characterized by low structuring, with many basal polytomies (Fig 4). Only three well supported lineages (>80 bootstrap support) were found: two including exclusively haplotypes from the Iberian refugium and a third containing haplotypes from southern France (H10 and H11). The remaining haplotypes representing samples from the Iberian Peninsula and Europe do not show clear genetic structure.

Putative hybrids between sympatric *L. fluviatilis* and *L. planeri*

The existence of putative hybrids between *L. planeri* and *L. fluviatilis* was investigated using sympatric populations, in this case *L. planeri* and *L. fluviatilis* from the Tagus basin, Portugal (SPL_r and SFL_m , respectively), using NewHybrids. In this analysis, each individual was assigned a posterior probability (p) of belonging to one of the six different genotype classes resulting from two generations. From the 52 samples of *L. planeri*, 49 (94%) were classified as being pure *L. planeri* using the posterior probability threshold of 0.5, 20 of which showing $p > 0.99$ and 25 showing $0.8 < p < 0.99$ (Fig 5). For this species only one individual was classified as hybrid (F_2 ; second generation hybrid) with posterior probability of 0.664, and one individual was classified as being pure *L. fluviatilis* (posterior probability = 0.537) (Table 5). All the 46 individuals of *L. fluviatilis* were identified as such (pure *L. fluviatilis*), from which 40 exhibit $p > 0.99$, and 5 showing $0.8 < p < 0.99$. No F_1 or backcross hybrids were found in any of the species (Table 5).

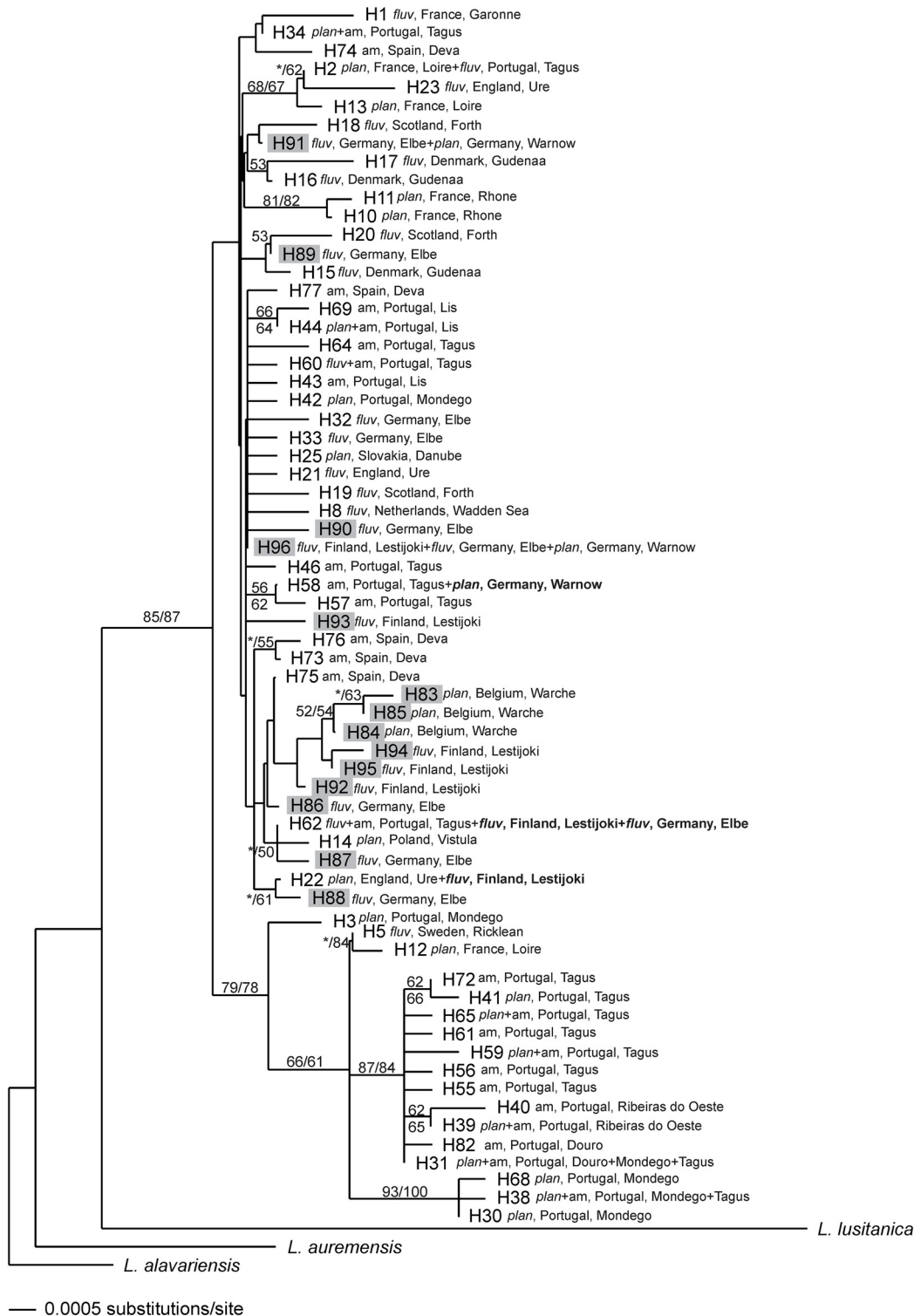


Fig 4. Neighbour-joining phylogenetic tree of 66 mitochondrial haplotypes of *Lampetra* (14 from this study (H83 to H96; grey shading) and the remaining 52 from GenBank database). For each haplotype, the species, country and river basin are indicated. Numbers are the bootstrap support values equal to or higher than 50% obtained from maximum parsimony (left or above) and neighbour-joining (right or below). The branches not recovered by one of these analyses are indicated with an asterisk. *fluv*, *L. fluviatilis*; *plan*, *L. planeri*; *am*, ammocoete.

doi:10.1371/journal.pone.0148107.g004

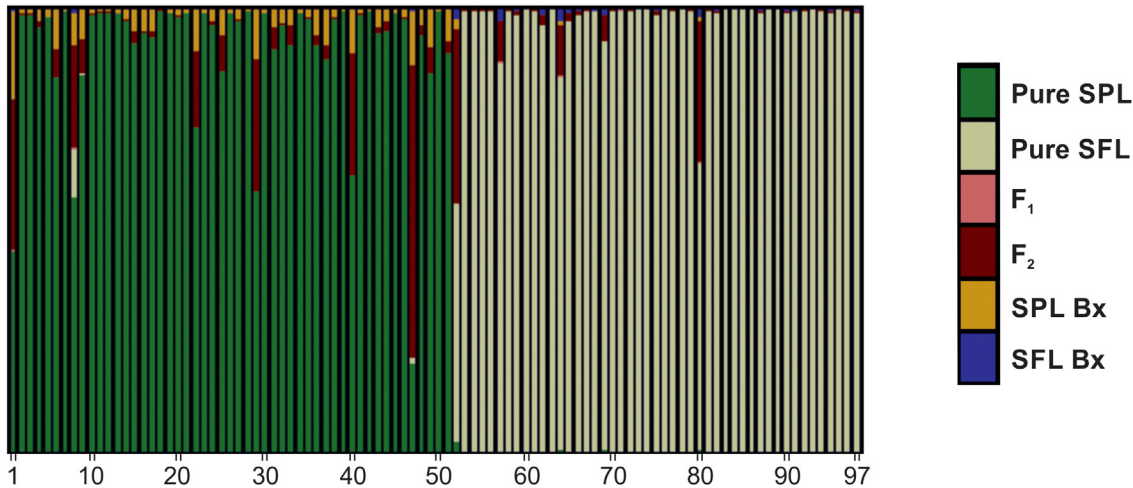


Fig 5. Estimated posterior probabilities that each individual from the sympatric populations of *L. planeri* (SPL; $n = 52$) and *L. fluviatilis* (SFL; $n = 46$) belongs to each of the six different genotype frequency categories that arise from two generations of potential interbreeding (parental species, F_1 , F_2 and backcrosses), computed by NewHybrids. Each individual is represented by a vertical bar. For the individuals identified as belonging to a certain hybrid category, posterior probability values are detailed in Table 5.

doi:10.1371/journal.pone.0148107.g005

Migration rate among populations

Recent migration rates (m) among migratory populations were estimated using BAYESASS. This analysis was performed for the three anadromous populations included in the study, i.e., *L. fluviatilis* from Tagus (SFL_m), from Elbe ($ELBE_m$) and from Lestijoki ($LEST_m$). The proportion of individuals derived from their own location was high in SFL_m ($m = 0.979$) and in $LEST_m$ ($m = 0.968$), and relatively low in $ELBE_m$ ($m = 0.705$) (Fig 6). Accordingly, a relatively high proportion of immigrants ($m = 0.286$) was detected from $LEST_m$ into $ELBE_m$. SFL_m is the most isolated population, with the highest proportion of non-immigrants ($m = 0.979$) and low migration rates ($m \leq 0.02$) in both directions (Fig 6, Table 6).

Table 5. Hybridization analysis for the sympatric *L. fluviatilis* (SFL_m) and *L. planeri* (SPL_r) from Tagus basin. Estimated posterior probabilities of belonging to one of the six genotype frequency classes (pure parental, F_1 , F_2 or backcrosses) for the individuals showing some evidence of hybridization. An individual is identified as belonging to a certain class if the posterior probability of falling into that class is above 0.5. Specimens are numbered as in Fig 5.

Species	Specimen	Pure SPL_r	Pure SFL_m	F_1	F_2	SPL_r Bx	SFL_m Bx
<i>L. planeri</i> (SPL_r) $n = 52$	1	0.452	-	-	0.342	0.200	-
	8	0.577	0.106	-	0.233	0.075	-
	22	0.733	-	-	0.171	0.096	-
	29	0.587	-	-	0.301	0.113	-
	40	0.625	-	-	0.274	0.099	-
	47*	0.198	0.013	-	0.664	0.120	-
<i>L. fluviatilis</i> (SFL_m) $n = 46$	52	0.023	0.537	-	0.394	0.023	0.022
	80	-	0.649	-	0.319	0.009	0.018

Genotype classes: Pure SPL_r , pure *L. planeri*; Pure SFL_m , pure *L. fluviatilis*; F_1 , first generation hybrid; F_2 , second generation hybrid; SPL_r Bx, *L. planeri* backcross (pure *L. planeri* mating with F_1); and SFL_m Bx, *L. fluviatilis* backcross (pure *L. fluviatilis* mating with F_1).

Bold indicates the class the individuals were classified into.

*Individual identified as hybrid.

doi:10.1371/journal.pone.0148107.t005

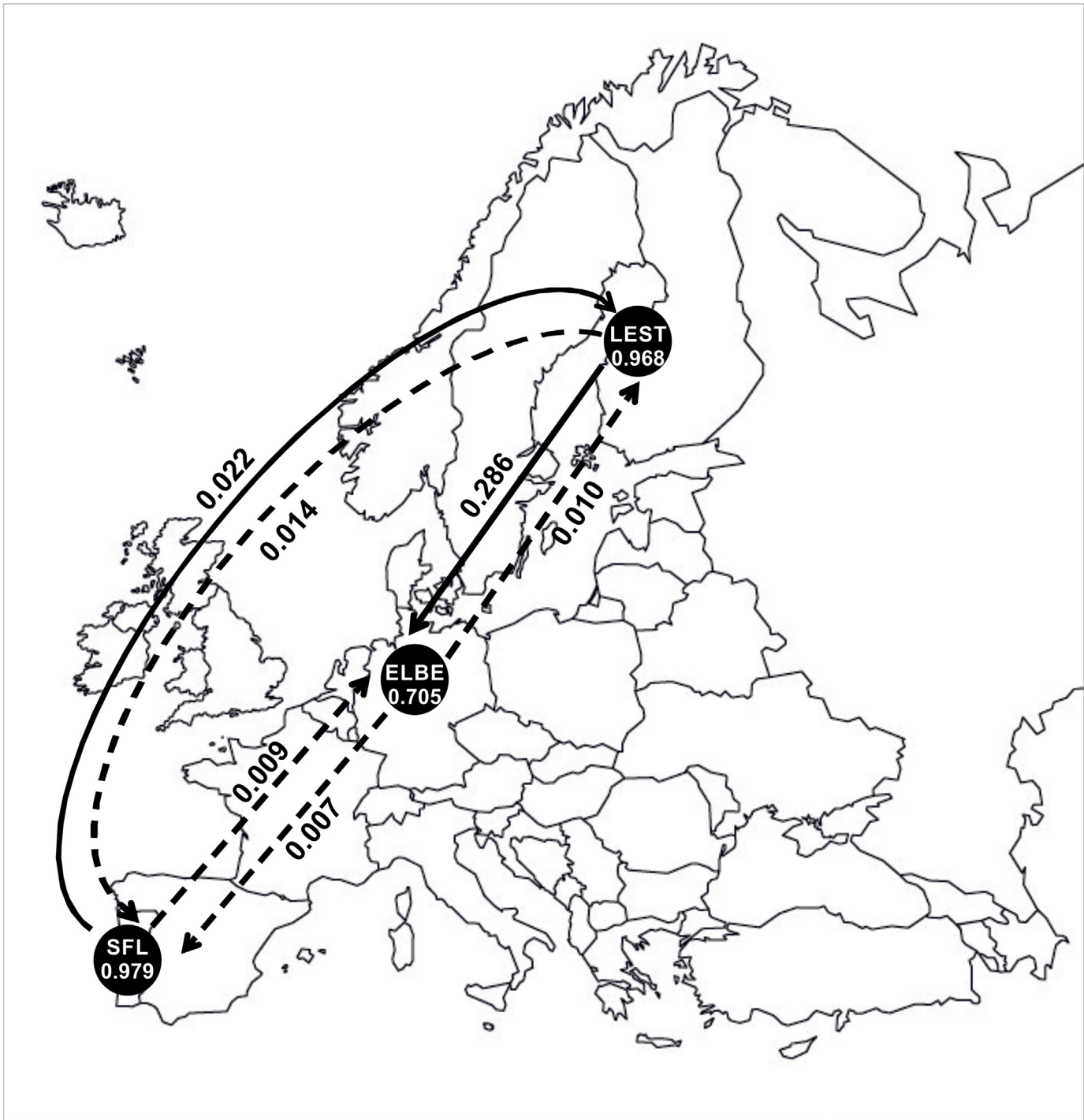


Fig 6. Recent migration rates (m) between migratory populations estimated using BAYESASS. Within circles, acronyms represent samples as in Fig 1 and Table 1, and numbers denote the proportion of non-immigrants within populations. Arrows indicate direction of gene flow among populations and respective m value. Dashed arrows represent values of m lower than 0.02.

doi:10.1371/journal.pone.0148107.g006

Table 6. Bayesian estimates of recent migration rates (*m*) among all populations using the program BAYESASS. Values shown are the mean migration rate into each population and their respective standard deviation of the marginal posterior distribution in parentheses. Values along the diagonal (in bold) represent the proportion of non-immigrants within populations. Values of *m* higher than 0.02 are presented in italics.

Migration from	Migration into										
	LEST _m	BEKE _r	ELBE _m	WARC _r	ESM _r	LIS _r	OES _r	NAB _r	SPL _r	SFL _m	SADO _r
LEST _m	0.9019 (0.0274)	0.2554 (0.0216)	0.2428 (0.0205)	0.0074 (0.0071)	0.0079 (0.0075)	0.0071 (0.0070)	0.0073 (0.0071)	0.0071 (0.0068)	0.0067 (0.0063)	0.0141 (0.0112)	0.0053 (0.0053)
BEKE _r	0.0083 (0.0082)	0.6753 (0.0084)	0.0064 (0.0063)	0.0072 (0.0070)	0.0076 (0.0075)	0.0073 (0.0071)	0.0073 (0.0071)	0.0072 (0.0070)	0.0057 (0.0056)	0.0059 (0.0058)	0.0054 (0.0054)
ELBE _m	0.0085 (0.0082)	0.0076 (0.0074)	0.6995 (0.0136)	0.0071 (0.0070)	0.0075 (0.0074)	0.0072 (0.0070)	0.0076 (0.0073)	0.0073 (0.0070)	0.0061 (0.0059)	0.0060 (0.0058)	0.0053 (0.0052)
WARC _r	0.0082 (0.0079)	0.0077 (0.0075)	0.0063 (0.0062)	0.9261 (0.0210)	0.0074 (0.0071)	0.0073 (0.0069)	0.0076 (0.0077)	0.0115 (0.0094)	0.0057 (0.0057)	0.0060 (0.0058)	0.0054 (0.0053)
ESM _r	0.0085 (0.0083)	0.0077 (0.0075)	0.0061 (0.0060)	0.0076 (0.0075)	0.9244 (0.0209)	0.0071 (0.0069)	0.0075 (0.0074)	0.0076 (0.0074)	0.0187 (0.0131)	0.0058 (0.0058)	0.0054 (0.0052)
LIS _r	0.0084 (0.0082)	0.0077 (0.0073)	0.0062 (0.0059)	0.0071 (0.0070)	0.0074 (0.0072)	0.6745 (0.0076)	0.2579 (0.0210)	0.0071 (0.0071)	0.0053 (0.0052)	0.0058 (0.0057)	0.0054 (0.0053)
OES _r	0.0083 (0.0082)	0.0077 (0.0075)	0.0065 (0.0063)	0.0070 (0.0070)	0.0075 (0.0077)	0.2609 (0.0196)	0.6747 (0.0078)	0.0074 (0.0075)	0.0052 (0.0052)	0.0059 (0.0059)	0.0056 (0.0055)
NAB _r	0.0085 (0.0082)	0.0078 (0.0077)	0.0061 (0.0061)	0.0073 (0.0070)	0.0075 (0.0073)	0.0072 (0.0068)	0.0076 (0.0075)	0.9228 (0.0208)	0.0055 (0.0055)	0.0059 (0.0058)	0.0053 (0.0053)
SPL _r	0.0081 (0.0081)	0.0078 (0.0075)	0.0063 (0.0059)	0.0076 (0.0076)	0.0076 (0.0073)	0.0073 (0.0070)	0.0073 (0.0071)	0.0073 (0.0072)	0.9272 (0.0195)	0.0059 (0.0057)	0.0053 (0.0053)
SFL _m	0.0229 (0.0189)	0.0076 (0.0075)	0.0076 (0.0072)	0.0083 (0.0082)	0.0075 (0.0073)	0.0071 (0.0069)	0.0076 (0.0074)	0.0074 (0.0072)	0.0083 (0.0074)	0.9330 (0.0191)	0.0055 (0.0055)
SADO _r	0.0084 (0.0081)	0.0076 (0.0075)	0.0063 (0.0063)	0.0074 (0.0074)	0.0076 (0.0073)	0.0071 (0.0071)	0.0075 (0.0073)	0.0073 (0.0072)	0.0054 (0.0053)	0.0059 (0.0057)	0.9462 (0.0154)

doi:10.1371/journal.pone.0148107.t006

Demographic history

Bottleneck analysis revealed consistent signs for recent contraction of population size only in the population from Meuse (WARC_r), which showed significant ($P < 0.05$) heterozygote excess according to the three mutational models tested, and a shift in the distribution of allele frequency classes from the expected L-shaped distribution (Table 7). The population from Nabão (NAB_r) also presents significant ($P < 0.01$) heterozygote excess according to IAM, but no deviation from the expected L-shaped distribution (Table 7). Using the *M* ratio statistic test, we found strong evidence of past reduction in population size for the migratory *L. fluviatilis* from Tagus (SFL_m), as the *M* ratio was significantly smaller than the equilibrium expectation ($P < 0.05$) for all the prebottleneck θ values considered. None of the remaining populations presented signs of reduction in population size (Table 7).

Discussion

Genetic diversity, population structure and postglacial dispersal

The colonization processes that took place after the glacial periods, when populations from the southern Mediterranean peninsulas expanded north across Europe, shaped, together with recent processes, the biodiversity of current taxa. Southern populations isolated in refugia and sub-refugia accumulated variation through the ice ages, and the founders that rapidly moved northward during interglacials only represented a subsample of the southern diversity [46]. This pattern is clearly observed for the genus *Lampetra* in this study. This genus presents a

Table 7. Demographic analysis. *P* values for one-tailed heterozygote excess (bold indicates significant *P* values), deviation of allele frequency classes from a normal L-shaped distribution and *M* ratio tests. For populations' acronyms, please check Table 1.

	IAM	TPM	SMM	L-shape	<i>M</i> ratio value
LEST _m	0.326	0.820	0.993	no deviation	0.72
BEKE _r	0.578	0.963	0.994	no deviation	0.75
ELBE _m	0.248	0.590	0.936	no deviation	0.71
WARC _r	0.014	0.020	0.037	deviation	0.79
ESM _r	0.156	0.563	0.906	no deviation	0.83
NAB _r	0.007	0.064	0.082	no deviation	0.75
SPL _r	0.097	0.216	0.784	no deviation	0.78
SFL _m	0.326	0.674	0.976	no deviation	0.66 ¹
SADO _r	0.422	0.422	0.578	no deviation	0.82

IAM, infinite allele model; TPM, two-phase model; SMM, stepwise mutation model.

The populations LIS_r and OES_r were not included in the analysis as they only have one polymorphic locus (Lspn 094) (S2 Table).

¹Location with *M* ratio value significantly smaller than the equilibrium expectation (*P* < 0.05) for all the prebottleneck θ values considered (4, 10 and 20).

doi:10.1371/journal.pone.0148107.t007

higher number of species and higher genetic diversity in the Iberian Peninsula compared with those of central and northern Europe, with the majority of private alleles being found in southern samples. This pattern of reduced richness in northern regions has already been observed by Bracken *et al.* [47] for the same species inhabiting the British Isles, and by Boguski *et al.* [48] for North American *Lampetra*, where haplotypic richness is greatest in regions south of the Columbia River. Similarly, Goodman *et al.* [49] and Spice *et al.* [50] found that in the anadromous Pacific lamprey (*Entosphenus tridentatus*), haplotypic richness increases from north to south.

This study is another evidence of the Iberian Peninsula as a glacial refugium, as proposed by Taberlet *et al.* [1], and posteriorly supported by several case-studies, namely in fish. For instance, a number of studies on endemic Iberian cyprinid species support this scenario, like for *Barbus sclateri* [51], *Squalius aradensis* [52] and *Pseudochondrostoma polylepis* [53]. Likewise, the study of Consuegra *et al.* [54] with Atlantic salmon demonstrates the presence of this species in the Iberian glacial refugium during the last 40 000 years and also points to the Iberian Peninsula as the likely source of the most common haplotype within the Atlantic lineage in Europe.

STRUCTURE analysis for the 11 analyzed samples revealed the existence of eight genetic clusters: whereas the southern populations were grouped in several genetic clusters reflecting their high levels of differentiation, northern populations were grouped in the same genetic cluster as result of their more recent common ancestor. The exception was the resident *L. planeri* from Warche in Belgium, which constitutes a distinct genetic cluster with high proportion of membership. This population seems to be facing a genetic bottleneck (see Table 6), which explains the relatively low number of alleles found in each microsatellite locus (maximum of three, see S2 Table). This may reflect the isolation of this population, which is located very upstream in the river basin, with several important obstacles downstream (M. Ovidio pers. comm.). Glaciations have been considered an important factor in brook lamprey evolution, both in Europe and North America [48,55,56]. These long periods favoured the abandonment of anadromous habits due to blocking of migratory routes. A cryptic Belgian refugium is another hypothesis for this differentiation, as several cryptic northern refugia have been hypothesized, one of which in the Belgian Ardennes. These northern refugia would have been in areas of sheltered topography that provided suitable stable microclimates [57]. We tested,

using mtDNA, whether the population from Warche represent an independently evolved population that survived during glaciations in the Belgian Ardennes refugium (see Fig 4). The results attained were not conclusive because although the three private haplotypes from Warche (H83 to H85) group in a clade, this branch shows low support.

Demographic analyses revealed consistent signs for recent contraction of population size in the population from Meuse (WARC_r). Nabão (NAB_r) population presented significant heterozygote excess according to the infinite allele model (IAM). These results, however, were not corroborated by mtDNA data, where Fu's F_s [58] and Tajima's D [59] statistics revealed non-significant values (WARC_r calculated in the present study, and NAB_r from Mateus *et al.* [9]), unlike what would be expected from a recent population bottleneck. This might be explained by the different signals attained from both markers, i.e., the more ancestral signal retrieved from mtDNA *versus* the recent history retrieved from microsatellite loci. This would mean that populations historically not showing evidence for contraction in population size (as attained from mtDNA data), may have faced a recent genetic bottleneck, as revealed by microsatellite data. Grouping of Iberian populations by STRUCTURE reflects in general their specific status, with the exception of LIS_r and OES_r that were grouped together, but not grouped with the other population of *L. planeri* from Portugal (SPL_r). Those two populations present very low levels of genetic diversity, having one single polymorphic locus and no private alleles. In general, the alleles present in those populations are rather common, most of the times having a frequency of more than 50% in other populations, namely SPL_r. The differentiation of those populations from SPL_r seems, therefore, to reflect this lack of diversity, which statistically makes them unique. Those populations may be facing a genetic bottleneck, but this analysis could not be performed due to the existence of a single polymorphic locus.

This study revealed that anadromous populations from central and northern Europe have high proportions of membership from the population of *L. fluviatilis* from the Iberian Peninsula (the SFL_m genetic cluster). Because contemporary gene flow from south to north and vice-versa is apparently happening in very small proportions (see Fig 6), this signal is likely due to ancestral polymorphism, as a result of the colonization process from a southern *L. fluviatilis*-type, rather than ongoing gene flow. The hypothesis of ancestral polymorphism is also supported by the mtDNA analysis, as northern samples of *L. fluviatilis* (LEST_m and ELBE_m) share haplotype 62 with *L. fluviatilis* from the Tagus watershed (Iberian Peninsula) (see Fig 4), and is in tune with the scenario of recent dispersal and founding of the northern populations ([9,11]; discussed above). For a better understanding of the contemporary patterns of gene flow in more recently established northern populations, recent migration routes among and within a wider range of northern *L. fluviatilis* and *L. planeri* populations should be further investigated.

Gene flow among species and populations

The repeated emergence of resident forms from ancestral migratory ones in different locations and times is a phenomenon very well known in lampreys [5–7,60], and should promote varying degrees of reproductive isolation between founder and derived species.

Previous data on mtDNA suggested that the three Iberian brook lampreys *L. alavariensis*, *L. auremensis* and *L. lusitanica* are well supported species, while the brook *L. planeri* shares haplotypes with the parasitic form, which is evidence that its emergence was more recent [9,11]. Present microsatellite data is concordant with the mtDNA data, as it revealed no signal of ongoing gene flow between the Iberian endemics with any of the samples included in this study. In the paired *L. planeri* and *L. fluviatilis* the scenario is different, as levels of genetic differentiation vary among populations. Sympatric *L. fluviatilis* and *L. planeri* from the Tagus basin present significant genetic differentiation and almost no signal of hybridization. These

results are in agreement with the previous work using restriction site-associated DNA (RAD) sequencing [14] that suggests that these populations are two distinct taxa that diverged recently. Results attained for northern populations of *L. fluviatilis* and *L. planeri*, however, indicate that these species are grouped in a single cluster (STRUCTURE analysis) and may be experiencing, or have experienced until recently, gene flow. This scenario is most likely explained by the postglacial colonization of northern habitats by a southern *L. fluviatilis*-type and, consequently, later emergence of northern populations, as explained below. Differentiation between resident and anadromous populations may, therefore, be an ongoing process in many locations, where speciation is still underway. This scenario was also found for the same species in France, where variable levels of divergence were found among sympatric population pairs [61]. This may also be the case in other lamprey paired species; for instance, no significant genetic differences were found between sympatric populations of the paired silver (*Ichthyomyzon unicuspis*) and northern brook (*Ichthyomyzon fossor*) lampreys from the Great Lakes, suggesting the existence of ongoing gene flow between them at least in this region [62].

BAYESASS revealed high recent gene flow between the migratory northern populations ELBE_m and LEST_m, which is corroborated by the assignment tests (see Table 4). In contrast, the migratory *L. fluviatilis* from the Iberian Peninsula has a signature similar to a resident species, showing almost absence of ongoing gene flow with northern populations and high degree of isolation and differentiation. This is consistent with the findings of [63], who suggested that lampreys with smaller body size may show limited dispersal and greater genetic differentiation. The genetic isolation of *L. fluviatilis* from the Iberian Peninsula, together with the relatively small size of individuals may reveal reduced levels of mobility during the parasitic adult phase, probably associated with their permanence in the large Tagus estuary (ca. 300 km²) and adjacent coastal area. *Lampetra fluviatilis* migrants have been separated on the basis of size into “typical” and “praecox” forms, whose mean lengths are approximately 30 and 24 cm, and mean weights about 53 and 22 g, respectively [64]. The size difference between the typical and praecox forms is thought to be due to differences in the time spent feeding in the sea, the last reducing their marine feeding phase by at least 1 year [64]. The population from the Tagus river basin resembles these smaller praecox forms; specimens analysed by the authors had in average 26 cm total length and 33 g weight, and one of the individuals was as small as 20 cm of total length and 19 g weight. These low values contrast, for instance, with those registered by Kemp *et al.* [65] for this species in north-east England, 80.7 g and 36.3 cm. The southern population from Tagus river basin shows strong evidence of past reduction in population size (this study), and the low number of individuals caught in the last years is representative of the rareness of this population [10], which is apparently isolated from the remaining European populations.

Conclusions

The use of microsatellites, coupled with information of mtDNA data, has proved effective in unravelling the patterns of colonization in European *Lampetra*, placing the Iberian Peninsula as a major source of postglacial colonization for this genus. We suggest the existence of a speciation continuum between *L. fluviatilis* and *L. planeri*, i.e., different populations of this pair likely represent differentiation at different stages, as result of the different timing of colonization and refugial persistence. This hypothesis should be further investigated with the addition of more samples, namely of sympatric populations.

Gene flow between migratory populations is high in northern regions, but low between northern populations and the southern Iberian population. We hypothesise that the migratory *L. fluviatilis* from the Iberian Peninsula has limited dispersion movements, resembling the “praecox” form described for this species by Abou-Seedo and Potter (1979).

Supporting Information

S1 Fig. Analysis of convergence of the MCMC algorithm for the estimates from BAYESASS (Bayesian estimates of recent migration rates) using the software Tracer 1.6. Plots represent the log probability along iterations, where the burn-in iterations are indicated in light grey and sample iterations in black; and the Bayesian posterior density of the parameter estimates for a) the three migratory populations and b) all the 11 populations. As should be expected in a case of convergence, the log probability oscillates around a plateau, and the oscillations are quite regular, i.e., there are no persistent lows or highs (valleys or hills) in the plot.

(DOCX)

S1 Table. List of the 10 polymorphic primer sets used, allelic range (bp), fluorescent label and multiplex panel.

(DOCX)

S2 Table. Measures of genetic diversity assayed at ten microsatellite DNA loci for each sampled location. Sample acronyms correspond to locations as in [Fig 1](#) and [Table 1](#). Number of alleles per locus (N_a) with number of private alleles in parentheses, mean allelic richness (AR), unbiased expected heterozygosity (H_e), observed heterozygosity (H_o), significance of departure from Hardy–Weinberg Equilibrium (HWE), mean number of alleles across loci (MNA) and number of polymorphic loci in each location (P). Grey shading indicates loci where MICRO--CHECKER detected signs of null alleles and relation with deviations from Hardy–Weinberg equilibrium. NS, non-significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; n , sample size; †, private allele with frequency $> 50\%$.

(DOCX)

S3 Table. Data of the ten microsatellite loci assayed for each of the 415 samples.

(PDF)

Acknowledgments

The authors especially thank Alpo Huhmarniemi and Jaakko Erkinaro (Finnish Game and Fisheries Research Institute, Finland), Michael Ovidio (University of Liège, Belgium) and Martin Krappe (Society for Nature Conservation and Landscape Ecology, Kratzeburg, Germany) for the donation of samples. We are also grateful to Hugo Gante and Vítor Sousa for valuable suggestions on data analysis, to Maria Drago for assistance in early phases of the labwork and to Carlos Alexandre and Tadeu Pereira for assistance in the fieldwork. We thank the collecting permits provided by Instituto da Conservação da Natureza e das Florestas (ICNF), and Fluvial de Mora for logistic support. Financial support was provided by Fundação para a Ciência e a Tecnologia (FCT), FEDER, COMPETE, and QREN through project funding (PTDC/BIA-BDE/71826/2006) and), through the pluriannual funding program to the Center of Oceanography (PEst-OE/MAR/UI0199/2011.) and by National Funds through FCT - Foundation for Science and Technology under the Project UID/MAR/04292/2013. C.S. Mateus and N. Mesquita were supported by FCT grants SFRH/BD/44029/2008 and SFRH/BPD/14557/2003, respectively.

Author Contributions

Conceived and designed the experiments: CSM MJA PRA BRQ NM. Performed the experiments: CSM. Analyzed the data: CSM. Contributed reagents/materials/analysis tools: CSM MJA PRA BRQ NM. Wrote the paper: CSM MJA PRA BRQ NM. Responsible for sampling: CSM MJA PRA BRQ.

References

1. Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 1998; 7: 453–464. PMID: [9628000](#)
2. Lorenzini R, Lovari S. Genetic diversity and phylogeography of the European roe deer: the refuge area theory revisited. *Biological Journal of the Linnean Society* 2006; 88: 85–100. doi: [10.1111/j.1095-8312.2006.00606.x](#)
3. Hewitt GM. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 1999; 68: 87–112. doi: [10.1006/bjil.1999.0332](#)
4. Bell MA, Andrews CA. Evolutionary consequences of postglacial colonization of fresh water by primitively anadromous fishes. In: Streit B, Stadler T, Lively CM, editors. *Evolutionary Ecology of Freshwater Animals*. Basel: Birkhauser Verlag. pp. 323–363; 1997.
5. Hubbs CL. Speciation of fishes. *The American Naturalist* 1940; 74: 198–211.
6. Zanandrea G. Speciation among lampreys. *Nature* 1959; 184: 380.
7. Vladykov VD, Kott E. Satellite species among the holarctic lampreys (Petromyzonidae). *Canadian Journal of Zoology* 1979; 57: 860–867.
8. Mateus CS, Alves MJ, Quintella BR, Almeida PR. Three new cryptic species of the lamprey genus *Lampetra* Bonnaterre, 1788 (Petromyzontiformes: Petromyzontidae) from the Iberian Peninsula. *Contributions to Zoology* 2013; 82: 37–53.
9. Mateus CS, Almeida PR, Quintella BR, Alves MJ. MtDNA markers reveal the existence of allopatric evolutionary lineages in the threatened lampreys *Lampetra fluviatilis* (L.) and *Lampetra planeri* (Bloch) in the Iberian glacial refugium. *Conservation Genetics* 2011; 12: 1061–1074.
10. Mateus CS, Rodríguez-Muñoz R, Quintella BR, Alves MJ, Almeida PR. Lampreys of the Iberian Peninsula: distribution, population status and conservation. *Endangered Species Research* 2012; 16: 183–198.
11. Espanhol R, Almeida PR, Alves MJ. Evolutionary history of lamprey paired species *Lampetra fluviatilis* (L.) and *Lampetra planeri* (Bloch) as inferred from mitochondrial DNA variation. *Molecular Ecology* 2007; 16: 1909–1924. PMID: [17444901](#)
12. Schreiber A, Engelhorn R. Population genetics of a cyclostome species pair, river lamprey (*Lampetra fluviatilis* L.) and brook lamprey (*Lampetra planeri* Bloch). *Journal of Zoological Systematics and Evolutionary Research* 1998; 36: 85–99.
13. Blank M, Jürss K, Bastrop R. A mitochondrial multigene approach contributing to the systematics of the brook and river lampreys and the phylogenetic position of *Eudontomyzon mariae*. *Canadian Journal of Fisheries and Aquatic Sciences* 2008; 65: 2780–2790.
14. Mateus CS, Stange M, Berner D, Roesti M, Quintella BR, Alves MJ. Strong genome-wide divergence between sympatric European river and brook lampreys. *Current Biology* 2013; 23: R649–R650. doi: [10.1016/j.cub.2013.06.026](#) PMID: [23928080](#)
15. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press; 1989.
16. Bryan MB, Libants S V, Warrillow JA, Li W, Scribner KT. Polymorphic microsatellite markers for the landlocked sea lamprey, *Petromyzon marinus*. *Conservation Genetics* 2003; 4: 113–116.
17. Filcek KB, Gilmore SA, Scribner KT, Jones ML. Discriminating Lamprey Species Using Multilocus Microsatellite Genotypes. *North American Journal of Fisheries Management* 2005; 25: 502–509.
18. Takeshima H, Yokoyama R, Nishida M, Yamazaki Y. Isolation and characterization of microsatellite loci in the threatened brook lamprey *Lethenteron* sp. N. *Molecular Ecology Notes* 2005; 5: 812–814.
19. McFarlane CT, Docker MF. Characterization of 14 microsatellite loci in the paired lamprey species *Ichthyomyzon unicuspis* and *I. fossor* and cross amplification in four other *Ichthyomyzon* species. *Conservation Genetics Resources* 2009; 1: 377–380.
20. Luzier CW, Docker MF, Whitesel TA. Characterization of ten microsatellite loci for western brook lamprey *Lampetra richardsoni*. *Conservation Genetics Resources* 2010; 2: 71–74.
21. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 2004; 4: 535–538.
22. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 2005; 1: 47–50.
23. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France); 1996.

24. Kalinowski ST. HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 2005; 5: 187–189.
25. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984; 38: 1358–1370.
26. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; 131: 479–491. PMID: [1644282](#)
27. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000; 155: 945–959. PMID: [10835412](#)
28. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 2005; 14: 2611–2620. PMID: [15969739](#)
29. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 2012; 4: 359–361.
30. Rosenberg NA. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* 2004; 4: 137–138.
31. Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 1999; 153: 1989–2000. PMID: [10581301](#)
32. Rannala B, Mountain JL. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* 1997; 94: 9197–9201. PMID: [9256459](#)
33. Beerli P. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology* 2004; 13: 827–836. PMID: [15012758](#)
34. Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 2004; 13: 55–65. PMID: [14653788](#)
35. Peakall R, Smouse PE. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 2006; 6: 288–295.
36. Peakall R, Smouse PE. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 2012; 28: 2537–2539. PMID: [22820204](#)
37. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* 2002; 160: 1217–1229. PMID: [11901135](#)
38. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 2003; 163: 1177–1191. PMID: [12663554](#)
39. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6. 2014. Available: <http://beast.bio.ed.ac.uk/Tracer>.
40. Piry S, Luikart G, Cornuet J-M. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* 1999; 90: 502–503.
41. Luikart G, Allendorf FW, Cornuet J-M, Sherwin WB. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* 1998; 89: 238–247. PMID: [9656466](#)
42. Garza JC, Williamson EG. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 2001; 10: 305–318. PMID: [11298947](#)
43. Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland; 2002.
44. Saitou N, Nei M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular biology and evolution* 1987; 4: 406–425. PMID: [3447015](#)
45. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998; 14: 817–818. PMID: [9918953](#)
46. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 1996; 58: 247–276.
47. Bracken FSA, Hoelzel AR, Hume JB, Lucas MC. Contrasting population genetic structure among freshwater-resident and anadromous lampreys: the role of demographic history, differential dispersal and anthropogenic barriers to movement. *Molecular Ecology* 2015; 24: 1188–1204. doi: [10.1111/mec.13112](#) PMID: [25689694](#)
48. Boguski DA, Reid SB, Goodman DH, Docker MF. Genetic diversity, endemism and phylogeny of lampreys within the genus *Lampetra sensu stricto* (Petromyzontiformes: Petromyzontidae) in western North America. *Journal of Fish Biology* 2012; 81: 1891–1914. doi: [10.1111/j.1095-8649.2012.03417.x](#) PMID: [23130690](#)

49. Goodman DH, Reid SB, Docker MF, Haas GR, Kinziger AP. Mitochondrial DNA evidence for high levels of gene flow among populations of a widely distributed anadromous lamprey *Entosphenus tridentatus* (Petromyzontidae). *Journal of Fish Biology* 2008; 72: 400–417.
50. Spice EK, Goodman DH, Reid SB, Docker MF. Neither philopatric nor panmictic: microsatellite and mtDNA evidence suggests lack of natal homing but limits to dispersal in Pacific lamprey. *Molecular Ecology* 2012; 21: 2916–2930. doi: [10.1111/j.1365-294X.2012.05585.x](https://doi.org/10.1111/j.1365-294X.2012.05585.x) PMID: [22564149](https://pubmed.ncbi.nlm.nih.gov/22564149/)
51. Gante HF, Micael J, Oliva-Paterna FJ, Doadrio I, Dowling TE, Alves MJ. Diversification within glacial refugia: tempo and mode of evolution of the polytypic fish *Barbus sclateri*. *Molecular ecology* 2009; 18: 3240–3255. doi: [10.1111/j.1365-294X.2009.04264.x](https://doi.org/10.1111/j.1365-294X.2009.04264.x) PMID: [19573028](https://pubmed.ncbi.nlm.nih.gov/19573028/)
52. Mesquita N, Hänfling B, Carvalho GR, Coelho MM. Phylogeography of the cyprinid *Squalius aradensis* and implications for conservation of the endemic freshwater fauna of southern Portugal. *Molecular Ecology* 2005; 14: 1939–1954. PMID: [15910317](https://pubmed.ncbi.nlm.nih.gov/15910317/)
53. Aboim MA, Mesquita N, Drago M, Coelho MM, Alves MJ. Assessing inter-drainage connections: patterns of genetic diversity in an Iberian cyprinid fish. *Biological Journal of the Linnean Society* 2013; 109: 656–669.
54. Consuegra S, García de Leániz C, Serdio A, González Morales M, Straus LG, Knox D, et al. Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular ecology* 2002; 11: 2037–2048. PMID: [12296947](https://pubmed.ncbi.nlm.nih.gov/12296947/)
55. Hardisty MW, Potter IC. The behaviour, ecology and growth of larval lampreys. In: Hardisty MW, Potter IC, editors. *The biology of lampreys*, Vol 1. London: Academic Press. pp. 85–125; 1971.
56. Hardisty MW. General introduction to lampreys. In: Holčík J, editor. *The freshwater fishes of Europe*, Vol 1, Part I—Petromyzontiformes. Wiesbaden: Aula-Verlag. pp. 19–83; 1986.
57. Stewart JR, Lister AM. Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* 2001; 16: 608–613. doi: [10.1016/S0169-5347\(01\)02338-2](https://doi.org/10.1016/S0169-5347(01)02338-2)
58. Fu Y-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997; 147: 915–925. PMID: [9335623](https://pubmed.ncbi.nlm.nih.gov/9335623/)
59. Tajima F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 1989; 123: 585–595. PMID: [2513255](https://pubmed.ncbi.nlm.nih.gov/2513255/)
60. Hubbs CL. The life cycle and growth of lampreys. *Papers of the Michigan Academy of Science* 1925; 4: 587–603.
61. Rougemont Q, Gaigher A, Lasne E, Côte J, Coke M, Besnard A-L, et al. Low reproductive isolation and highly variable levels of gene flow reveal limited progress towards speciation between European river and brook lampreys. *Journal of evolutionary biology* 2015; 28: 2248–2263. doi: [10.1111/jeb.12750](https://doi.org/10.1111/jeb.12750) PMID: [26348652](https://pubmed.ncbi.nlm.nih.gov/26348652/)
62. Docker MF, Mandrak NE, Heath DD. Contemporary gene flow between “paired” silver (*Ichthyomyzon unicuspis*) and northern brook (*I. fossor*) lampreys: implications for conservation. *Conservation Genetics* 2012; 13: 823–835.
63. Spice EK, Goodman DH, Reid SB, Docker MF. Neither philopatric nor panmictic: microsatellite and mtDNA evidence suggests lack of natal homing but limits to dispersal in Pacific lamprey. *Molecular Ecology* 2012; 21: 2916–2930. doi: [10.1111/j.1365-294X.2012.05585.x](https://doi.org/10.1111/j.1365-294X.2012.05585.x) PMID: [22564149](https://pubmed.ncbi.nlm.nih.gov/22564149/)
64. Abou-Seedo FS, Potter IC. The estuarine phase in the spawning run of the River lamprey *Lampetra fluviatilis*. *Journal of Zoology* 1979; 188: 5–25.
65. Kemp PS, Russon IJ, Vowles AS, Lucas MC. The influence of discharge and temperature on the ability of upstream migrant adult river lamprey (*Lampetra fluviatilis*) to pass experimental overshoot and under-shot weirs. *River Research and Applications* 2010; 27: 488–498.