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

Catarina Sofia Mateus1,2*, Pedro Raposo Almeida1,3, Natacha Mesquita2, Bernardo Ruivo Quintella1,4, 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 de Évora, Évora, Portugal, 4 Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal

* csmateus@fc.ul.pt

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
Postglacial Colonization, Gene Flow and Speciation in Lampreys

The authors have declared that no competing interests exist.

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 nº129/92, Portaria nº1005/92. 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>
<thead>
<tr>
<th>Country</th>
<th>Basin</th>
<th>River</th>
<th>Acronym</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>Lestijoki</td>
<td>Lestijoki</td>
<td>LESTₘ</td>
<td><em>L. fluviatilis</em></td>
</tr>
<tr>
<td>Germany</td>
<td>Warnow</td>
<td>Beke</td>
<td>BEKEₘ</td>
<td><em>L. planeri</em></td>
</tr>
<tr>
<td>Germany</td>
<td>Elbe</td>
<td>Schaaale</td>
<td>ELBEₘ</td>
<td><em>L. fluviatilis</em></td>
</tr>
<tr>
<td>Belgium</td>
<td>Meuse</td>
<td>Warche</td>
<td>WARCₘ</td>
<td><em>L. planeri</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Esmoriz</td>
<td>Esmoriz</td>
<td>ESMₘ</td>
<td><em>L. alavariensis</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Lis</td>
<td>Lis</td>
<td>LISₘ</td>
<td><em>L. planeri</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Ribeiras do Oeste</td>
<td>Ribeira de S. Pedro</td>
<td>OESₘ</td>
<td><em>L. auremensis</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Tagus</td>
<td>Nabo</td>
<td>NABₘ</td>
<td><em>L. auremensis</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Tagus</td>
<td>Sorraia¹</td>
<td>SPLₘ</td>
<td><em>L. planeri</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Tagus</td>
<td>Sorraia¹</td>
<td>SFLₘ</td>
<td><em>L. fluviatilis</em></td>
</tr>
<tr>
<td>Portugal</td>
<td>Sado</td>
<td>Marateca</td>
<td>SADOₘ</td>
<td><em>L. lusitanica</em></td>
</tr>
</tbody>
</table>

Species: *L. planeri*, *L. planeri*, *L. fluviatilis*, *L. alavariensis*, *L. auremensis*, *L. auremensis*, *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
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 Pmaμ 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 μL volumes containing 2 μL of 50 ng μL⁻¹ genomic DNA, 1.0 to 3.0 mM MgCl₂, 0.2 mM dNTP mix, 0.5 μM for each primer, 1 unit of DreamTaq™ DNA Polymerase (Fermentas) and 1× 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 μl Qiagen Multiplex PCR master Mix, 3 μl RNase-free water, 1 μl Primer Mix (2 μM each primer) and 1 μl of 50 ng μL⁻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 accessed through pairwise FST using the Weir & Cockerham’s estimator [25], and significance was assessed with 10⁴ permutations, as implemented in GENETIX. The distribution of genetic variation was accessed 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 2.0.1h [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 (F₀) 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 (F₁, F₂ 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 10⁶, followed by a sampling period of 10⁵ 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 10⁷ MCMC iterations (discarding the first 10⁶ 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 10⁵ 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*ₕ = 0.9, *Δ*ₖ = 3.5 and three values of θ (*θ* = 4, *θ* = 10 and *θ* = 20), where *θ* = 4Neμ (Ne = effective population size; μ = mutation rate), *p*ₕ is the proportion of one-step mutations, and *Δ*ₖ 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 (LESTm, BEKEr, ELBEm, WARCr) 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 LESTm, BEKEr, ELBEm and WARCr, 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: Pmaμ 5 in SPLr, and NABr; Iun 10 in SPLr, and BEKEr; Iun 5 in ELBEm and BEKEr; and Iun 14 in SPLr and BEKEr. 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 Lsp2010-2, to 13 at the locus Iun14. Twelve private alleles were found: NABr, SPLr and SADOr showed three unique alleles each, BEKE two, and SFL one (S2 Table). The mean number of alleles (MNA) across loci ranged from 1.1 (LISr and OESr) to 3.8 (ELBEm), mean allelic richness (AR) from 1.08 (LISr) to 2.62 (ELBEm), and expected heterozygosity (He) from 0.0239 (LISr) to 0.4417 (NABr) (S2 Table).

A considerable level of genetic differentiation among samples was observed (average FST = 0.498, P<0.001) with pairwise FST values ranging from 0.0114 (ELBEm–LESTm) to 0.8915 (OESr–ESMr), all being significant (P<0.001 for all pairs, with the exception of ELBEm–LESTm 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)
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 $L$. planeri...
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, and OES, 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
ELBE<sub>m</sub> was assigned to BEKE<sub>r</sub> and another to WARC<sub>r</sub>; one individual of WARC<sub>r</sub> was assigned to ELBE<sub>m</sub>; one individual of SPL<sub>r</sub> was assigned to SFL<sub>m</sub> and one individual of SFL<sub>m</sub> was assigned to LEST<sub>m</sub>.

The principal coordinates analysis (PCoA), revealed the existence of mainly six distinct clusters, NAB<sub>r</sub>, SADO<sub>r</sub>, ESM<sub>r</sub>, LIS<sub>r</sub>+OES<sub>r</sub>, SPL<sub>r</sub>+WARC<sub>r</sub>, BEKE<sub>r</sub> + LEST<sub>m</sub> + ELBE<sub>m</sub> + SFL<sub>m</sub> (Fig 3). These results are congruent with STRUCTURE, with the exception that it groups SPL<sub>r</sub> with WARC<sub>r</sub> and SFL<sub>m</sub> with the northern cluster, while in STRUCTURE SPL<sub>r</sub>, WARC<sub>r</sub>, SFL<sub>m</sub> and the northern populations BEKE<sub>r</sub>, LEST<sub>m</sub> and ELBE<sub>m</sub> constitute four distinct groups.

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.

<table>
<thead>
<tr>
<th>Population</th>
<th>Inferred clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>LEST&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0.734</td>
</tr>
<tr>
<td>BEKE&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.917</td>
</tr>
<tr>
<td>ELBE&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0.552</td>
</tr>
<tr>
<td>WARC&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.009</td>
</tr>
<tr>
<td>ESM&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.003</td>
</tr>
<tr>
<td>LIS&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.003</td>
</tr>
<tr>
<td>OES&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.002</td>
</tr>
<tr>
<td>NAB&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.008</td>
</tr>
<tr>
<td>SPL&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.016</td>
</tr>
<tr>
<td>SFL&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0.035</td>
</tr>
<tr>
<td>SADO&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.007</td>
</tr>
</tbody>
</table>

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

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.

<table>
<thead>
<tr>
<th>Assigned population</th>
<th>LEST&lt;sub&gt;m&lt;/sub&gt;</th>
<th>BEKE&lt;sub&gt;r&lt;/sub&gt;</th>
<th>ELBE&lt;sub&gt;m&lt;/sub&gt;</th>
<th>WARC&lt;sub&gt;r&lt;/sub&gt;</th>
<th>ESM&lt;sub&gt;r&lt;/sub&gt;</th>
<th>LIS&lt;sub&gt;r&lt;/sub&gt;</th>
<th>OES&lt;sub&gt;r&lt;/sub&gt;</th>
<th>NAB&lt;sub&gt;r&lt;/sub&gt;</th>
<th>SPL&lt;sub&gt;r&lt;/sub&gt;</th>
<th>SFL&lt;sub&gt;m&lt;/sub&gt;</th>
<th>SADO&lt;sub&gt;r&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEST&lt;sub&gt;m&lt;/sub&gt;</td>
<td>80</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BEKE&lt;sub&gt;r&lt;/sub&gt;</td>
<td>3</td>
<td>94</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ELBE&lt;sub&gt;m&lt;/sub&gt;</td>
<td>10</td>
<td>15</td>
<td>68</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WARC&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ESM&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LIS&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>73</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OES&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NAB&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SPL&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>SFL&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>96</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SADO&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

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.
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 (SPLr and SFLm, 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.

PLOS ONE | DOI:10.1371/journal.pone.0148107 February 12, 2016 12 / 22
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 ($m/0.02$) 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.

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

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 $F_1$).

*Individual identified as hybrid.

D\oi: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.

<table>
<thead>
<tr>
<th>Migration from</th>
<th>Migration into</th>
<th>LEST&lt;sub&gt;m&lt;/sub&gt;</th>
<th>BEKE&lt;sub&gt;r&lt;/sub&gt;</th>
<th>ELBE&lt;sub&gt;m&lt;/sub&gt;</th>
<th>WARC&lt;sub&gt;r&lt;/sub&gt;</th>
<th>ESM&lt;sub&gt;r&lt;/sub&gt;</th>
<th>LIS&lt;sub&gt;r&lt;/sub&gt;</th>
<th>OES&lt;sub&gt;r&lt;/sub&gt;</th>
<th>NAB&lt;sub&gt;r&lt;/sub&gt;</th>
<th>SPL&lt;sub&gt;r&lt;/sub&gt;</th>
<th>SFL&lt;sub&gt;r&lt;/sub&gt;</th>
<th>SADO&lt;sub&gt;r&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEST&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0.9019 (0.0274)</td>
<td>0.2554 (0.0216)</td>
<td>0.2428 (0.0205)</td>
<td>0.0074 (0.0071)</td>
<td>0.0079 (0.0075)</td>
<td>0.0071 (0.0070)</td>
<td>0.0073 (0.0071)</td>
<td>0.0071 (0.0068)</td>
<td>0.0067 (0.0063)</td>
<td>0.0141 (0.0112)</td>
<td>0.0053 (0.0053)</td>
<td></td>
</tr>
<tr>
<td>BEKE&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0083 (0.0082)</td>
<td>0.6753 (0.0084)</td>
<td>0.0064 (0.0063)</td>
<td>0.0072 (0.0070)</td>
<td>0.0076 (0.0075)</td>
<td>0.0073 (0.0071)</td>
<td>0.0073 (0.0071)</td>
<td>0.0072 (0.0070)</td>
<td>0.0057 (0.0056)</td>
<td>0.0059 (0.0058)</td>
<td>0.0054 (0.0054)</td>
<td></td>
</tr>
<tr>
<td>ELBE&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0.0085 (0.0082)</td>
<td>0.0076 (0.0074)</td>
<td>0.6995 (0.0136)</td>
<td>0.0071 (0.0070)</td>
<td>0.0075 (0.0074)</td>
<td>0.0072 (0.0070)</td>
<td>0.0076 (0.0073)</td>
<td>0.0073 (0.0070)</td>
<td>0.0061 (0.0059)</td>
<td>0.0060 (0.0058)</td>
<td>0.0053 (0.0052)</td>
<td></td>
</tr>
<tr>
<td>WARC&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0082 (0.0079)</td>
<td>0.0077 (0.0075)</td>
<td>0.0063 (0.0062)</td>
<td>0.9261 (0.0210)</td>
<td>0.0074 (0.0071)</td>
<td>0.0073 (0.0069)</td>
<td>0.0076 (0.0077)</td>
<td>0.0076 (0.0094)</td>
<td>0.0057 (0.0057)</td>
<td>0.0060 (0.0058)</td>
<td>0.0054 (0.0053)</td>
<td></td>
</tr>
<tr>
<td>ESM&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0085 (0.0083)</td>
<td>0.0077 (0.0075)</td>
<td>0.0061 (0.0060)</td>
<td>0.0076 (0.0075)</td>
<td>0.9244 (0.0209)</td>
<td>0.0071 (0.0069)</td>
<td>0.0075 (0.0074)</td>
<td>0.0076 (0.0094)</td>
<td>0.0115 (0.0131)</td>
<td>0.0058 (0.0058)</td>
<td>0.0054 (0.0053)</td>
<td></td>
</tr>
<tr>
<td>LIS&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0084 (0.0082)</td>
<td>0.0077 (0.0073)</td>
<td>0.0062 (0.0059)</td>
<td>0.0071 (0.0070)</td>
<td>0.0074 (0.0072)</td>
<td>0.6745 (0.0207)</td>
<td>0.2579 (0.0210)</td>
<td>0.0071 (0.0071)</td>
<td>0.0053 (0.0052)</td>
<td>0.0058 (0.0057)</td>
<td>0.0054 (0.0053)</td>
<td></td>
</tr>
<tr>
<td>OES&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0083 (0.0082)</td>
<td>0.0077 (0.0075)</td>
<td>0.0065 (0.0063)</td>
<td>0.0070 (0.0070)</td>
<td>0.0075 (0.0077)</td>
<td>0.2609 (0.0196)</td>
<td>0.6747 (0.0078)</td>
<td>0.0074 (0.0075)</td>
<td>0.0052 (0.0052)</td>
<td>0.0059 (0.0059)</td>
<td>0.0056 (0.0055)</td>
<td></td>
</tr>
<tr>
<td>NAB&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0085 (0.0082)</td>
<td>0.0078 (0.0077)</td>
<td>0.0061 (0.0061)</td>
<td>0.0073 (0.0070)</td>
<td>0.0075 (0.0073)</td>
<td>0.0072 (0.0068)</td>
<td>0.0076 (0.0075)</td>
<td>0.9228 (0.0208)</td>
<td>0.0055 (0.0055)</td>
<td>0.0059 (0.0058)</td>
<td>0.0053 (0.0053)</td>
<td></td>
</tr>
<tr>
<td>SPL&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0081 (0.0081)</td>
<td>0.0078 (0.0075)</td>
<td>0.0063 (0.0059)</td>
<td>0.0076 (0.0076)</td>
<td>0.0076 (0.0073)</td>
<td>0.0073 (0.0070)</td>
<td>0.0073 (0.0071)</td>
<td>0.0073 (0.0072)</td>
<td>0.9272 (0.0195)</td>
<td>0.0059 (0.0057)</td>
<td>0.0053 (0.0053)</td>
<td></td>
</tr>
<tr>
<td>SFL&lt;sub&gt;m&lt;/sub&gt;</td>
<td>0.0229 (0.0189)</td>
<td>0.0076 (0.0075)</td>
<td>0.0076 (0.0072)</td>
<td>0.0083 (0.0082)</td>
<td>0.0075 (0.0073)</td>
<td>0.0071 (0.0069)</td>
<td>0.0076 (0.0074)</td>
<td>0.0074 (0.0072)</td>
<td>0.0083 (0.0074)</td>
<td>0.9330 (0.0191)</td>
<td>0.0055 (0.0055)</td>
<td></td>
</tr>
<tr>
<td>SADO&lt;sub&gt;r&lt;/sub&gt;</td>
<td>0.0084 (0.0081)</td>
<td>0.0076 (0.0075)</td>
<td>0.0063 (0.0063)</td>
<td>0.0074 (0.0074)</td>
<td>0.0076 (0.0073)</td>
<td>0.0071 (0.0071)</td>
<td>0.0075 (0.0072)</td>
<td>0.0073 (0.0072)</td>
<td>0.0054 (0.0053)</td>
<td>0.0059 (0.0057)</td>
<td>0.9462 (0.0154)</td>
<td></td>
</tr>
</tbody>
</table>

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<sub>r</sub>), 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<sub>r</sub>) 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<sub>m</sub>), 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
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 Warthe 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 Fs [58] and Tajima’s D [59] statistics revealed non-significant values (WARC, calculated in the present study, and NAB, 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, and OES, 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_r 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_r and ELBE_r) 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.

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

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 (Na) with number of private alleles in parentheses, mean allelic richness (AR), unbiased expected heterozygosity (He), observed heterozygosity (Ho), 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 MICROCHECKER 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%.

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

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 Fluviário 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. Messquita 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


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.


