Hume, J. B., Recknagel, H., Bean, C. W.,, Adams, C. E. and Mable, B. K. (2018) RADseq and mate choice assays reveal unidirectional gene flow among three lamprey ecotypes despite weak assortative mating: Insights into the formation and stability of multiple ecotypes in sympatry. Molecular Ecology, 27(22), pp. 4572-4590.

# RADseq and mate choice assays reveal unidirectional gene flow among three lamprey ecotypes despite weak assortative mating: insights into the formation and stability of multiple ecotypes in sympatry.

John B. Hume[1,2*], Hans Recknagel[2*], Colin W. Bean[2,3], Colin E. Adams[2] & Barbara K. Mable[2]

[1]*College of Agriculture & Natural Resources, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, U.S.A.*, [2]*Institute of Biodiversity, Animal Health & Comparative Medicine, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, U.K.*, [3]*Scottish Natural Heritage, Caspian House, Mariner Court, Clydebank Business Park, Clydebank G81 2NR.*

Key words: speciation; cyclostome; divergence; sympatry; RAD-sequencing.

Corresponding author: Dr John B. Hume, College of Agriculture & Natural Resources, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, U.S.A. E-mail: jhume@msu.edu; Tel.: +1 517 353-0731.

Running title: Gene flow among lamprey ecotypes.

# ABSTRACT

Adaptive divergence with gene flow often results in complex patterns of variation within taxa exhibiting substantial ecological differences among populations. One example where this may have occurred is the parallel evolution of freshwater-resident nonparasitic lampreys from anadromous parasitic ancestors. Previous studies have focused on transitions between these two phenotypic extremes but here we considered more complex evolutionary scenarios where an intermediate freshwater form that remains parasitic is found sympatrically with the other two ecotypes. Using population genomic analysis (Restriction Associated DNA sequencing) we found that a freshwater-parasitic ecotype was highly distinct from an anadromous-parasitic form ($Q_{\text{lake-P}}$ = 96.8%, $F_{\text{st}}$ = 0.154), but that a freshwater-nonparasitic form was almost completely admixed in Loch Lomond, Scotland. Demographic reconstructions indicated that both freshwater populations likely derived from a common freshwater ancestor. However, while the nonparasitic ecotype has experienced high levels of introgression from the anadromous-parasitic ecotype ($Q_{\text{anad-P}}$ = 37.7%), there is no evidence of introgression into the freshwater-parasitic ecotype. Paradoxically, mate choice experiments predicted high potential for gene flow: males from all ecotypes were stimulated to spawn with freshwater-parasitic females, which released gametes in response to all ecotypes. Differentially fixed single nucleotide polymorphisms identified genes associated with growth and development, which could possibly influence the timing of metamorphosis, resulting in significant ecological differences between forms. This suggests that multiple lamprey ecotypes can persist in sympatry following shifts in adaptive peaks, due to environmental change during their repeated colonization of post-glacial regions, followed by periods of extensive gene flow among such diverging populations.

# INTRODUCTION

Colonization of novel environments is predicted to promote intraspecific phenotypic variation as a consequence of intense competition for limiting resources (van Valen 1965). Viewed across longer periods of evolutionary time, this process manifests as a continuum of increasing differentiation; ranging from resource polymorphisms within panmictic populations, to complete and irreversible reproductive isolation between species (Hendry *et al*. 2009). Thus the separation of one lineage of organisms from another (speciation) can be thought of most simply as a consequence of adaptation to a different environment; which is made easier when coupled with barriers to movement between them. Adaptive genetic divergence, where different alleles prove advantageous for groups living in different environments (ecological speciation), is predicted to reduce gene flow in those areas of the genome associated with well-adapted phenotypes (Schluter 2009). Reproductive isolation can, therefore, stem from ecological speciation in any geographic context where two populations are diverging (allopatric, parapatric, sympatric), even in the face of gene flow (Smadja & Butlin 2011). However, ecological speciation requires that: *i*) populations persist after shifts in adaptive peaks (e.g. due to environmental change or colonization of new environments; Pavey *et al*. 2010; Rogers *et al*. 2012), and *ii*) genetic divergence is associated with pre- or post-zygotic reproductive isolation (Rundle & Nosil 2005; Nosil 2012).

Northern freshwater fishes make particularly good models for understanding this continuum from ecotype formation through speciation. The advances and retreats of glaciers during the Quaternary period (15,000 – 10,000 years ago) had a profound impact on phylogeographic distributions (Petit *et al*. 2003; Rowe *et al*. 2004), with post-glacial range expansion varying greatly across the northern hemisphere (Hays *et al*. 1976; Schluter 1996). This would have created infrequent opportunity for secondary contact between populations inhabiting discrete and partly isolated post-glacial lakes during this period (Taylor & McPhail 2000; Lyons 2003). Parallel colonization of newly available lakes and streams led to the evolution of distinct gene pools within species for multiple fish groups (Rogers *et al*. 2013). As a consequence, several post-glacial fishes are characterized by substantial levels of polymorphism driven by ecological opportunity. Such polymorphisms exhibit both adaptive divergence (e.g. physiological differences, changes in maturation timing) and population persistence in the face of exposure to environmental change (e.g. dietary shifts) (Taylor 1999; Schluter 2000). For example, whitefish (*Coregonus* spp.) are highly polymorphic, with phenotypes readily identifiable by differences in head shape, gill raker number and body shape (Svärdson 1979; Kahilainen & Østbye 2006). In particular, gill raker number is a heritable trait influencing trophic ecology and plays a central role in whitefish diversification within lakes (Rogers & Bernatchez 2007). Furthermore, Arctic charr (*Salvelinus alpinus*) is arguably the most variable of vertebrates (Klemetsen 2013), with a large number of descriptions of discrete trophic specialists characterized by morphological differences functionally associated with alternative feeding strategies (e.g. Hooker *et al*. 2016), but constrained by ecological opportunity (Recknagel *et al*. 2017). Populations at different stages of the speciation process are useful for exploring mechanisms and constraints associated with phenotypic diversity and the evolution and maintenance of polymorphisms. Thus, the adaptation of fishes to newly available environments in recently glaciated freshwater systems, the rapid evolution of reproductive isolation between ecotypes, and their persistence in the face of gene flow make them excellent systems with which to investigate ecological speciation (Schluter 1996).

One conspicuous, yet understudied example of differentiation in response to ecological opportunity has occurred among lampreys (Petromyzontiformes), where several genera contain species pairs or sister species complexes, where a marine-migrating ancestral species has given rise to a single, or multiple, freshwater-resident species (Zanandrea 1959; Vladykov & Kott 1979). These are often phenotypically and genetically similar to their anadromous counterparts (Hardisty & Potter 1971), and may therefore represent diverging ecotypes of single species (Yamazaki & Goto 2000; Salewski 2003; Docker 2009). Following a protracted larval stage, all lampreys undergo metamorphosis and either: *i*) migrate to feed on other fish as ectoparasites in freshwater/marine environments prior to maturation or, *ii*) bypass the juvenile trophic phase and instead undertake sexual maturation soon after metamorphosis (Docker 2009). Lamprey ecotypes thus exhibit a disparity in body size at maturity as a consequence of post-metamorphic feeding (or the lack thereof) but most return to the upper reaches of streams to spawn (Moser *et al.* 2015; Johnson *et al.* 2015). It is widely assumed that size-assortative mating has resulted in pre-zygotic reproductive isolation between lamprey ecotypes spawning in syntopic habitats (Hardisty 1963; Beamish & Neville 1992) but this has rarely been tested experimentally (but see Beamish & Neville 1992).

European river (*Lampetra fluviatilis*) and brook lampreys (*L. planeri*) are common to coastal watersheds in Europe and are representative of a typical lamprey species pair; with the former being an anadromous parasite as an adult and the latter a freshwater-resident nonparasitic species. Genetic differentiation between them is only apparent in former glacial refugia (e.g. Iberia, Mateus *et al*. 2013), with no evidence for species-specific fixed genetic differences elsewhere, possibly as a result of ongoing gene flow (Schreiber & Engelhorn 1998; Espanhol *et al.* 2007; Pereira *et al.* 2011; Bracken *et al.* 2015; Rougemont *et al.* 2015, 2016, 2017; Mateus *et al*. 2016). This has led to the persistence of two alternative hypotheses: *i*) these species are the result of very recent divergence (either a single divergence event or multiple parallel events) or *ii*) that these lampreys comprise just a single species consisting of multiple ecotypes (Espanhol *et al*. 2007). However, the situation could be even more complicated than this because some catchments include more than just the two phenotypic extremes (parasitic anadromous and nonparasitic freshwater forms): in some locations an intermediate form (usually considered an ecotype of *L. fluviatilis*) that is freshwater resident but continues to feed parasitically has been reported (Collett 1905; Tuunainen *et al*. 1980; Morris 1989; Inger *et al*. 2010; Tsimbalov *et al*. 2015).

Since all previous studies have focused on the two extremes, the relative importance of two key life-history elements fundamental to lamprey speciation remains unresolved: the transition from a migratory to a freshwater resident life cycle, and the switch from parasitic to nonparasitic feeding strategies. Whether or not this two-part transition has resulted from a single evolutionary step repeatedly among lamprey species pairs remains an unanswered but important question. Although restricted to only a few large postglacial lakes in Europe, the freshwater parasitic ecotype has frequently been hypothesized to represent a necessary transitional stage during lamprey speciation in post-glacial regions (Zanandrea 1959, 1961; Beamish 1985; Morris 1989; Salewski 2003; Hardisty 2006). Alternatively, such populations could represent an independent evolutionary step that only persists under certain environmental contexts (e.g. allopatry in deep post-glacial lakes). However, no studies to date have compared genetic and behavioural isolation between all three ecotypes where they occur sympatrically to distinguish between these hypotheses.

In the Loch Lomond catchment, Scotland, U.K., which is a large post-glacial lake basin, three lamprey ecotypes occur: anadromous parasitic (which we refer to as anad-P), freshwater-resident parasitic (lake-P) and freshwater-resident non-parasitic (non-P) (Morris 1989). Analysis of mtDNA variation (2077 bp concatenated across the *ND3*, *cyt b* and *ATPase 6/8* genes) from across this catchment area identified 7 haplotypes among 10 individuals; with a single haplotype that was shared among all three ecotypes, one haplotype that was shared between lake-P and non-P individuals and the rest unique to either anad-P or non-P individuals (Hume 2013; Table 7.2). While this could suggest that the two freshwater-resident forms shared a common ancestor more recently than the anad-P ecotype, the sampling was too restricted to draw definitive conclusions. Moreover, analysis of mtDNA precludes interpretation of hybridization and introgression, which has been found to have played an important role in the differentiation of other lamprey populations (Rougemont *et al.* 2015, 2016, 2017). In this study, we thus focused on nuclear variation using restricted-associated DNA sequencing, combined with behavioural observations of mating preferences. We use samples collected from a single stream in the Loch Lomond catchment, supporting spawning populations of all three ecotypes, to investigate the evolutionary forces resulting in the formation of ecotypes in sympatry. Our aims were to: *i*) test the extent of differentiation among the three lamprey ecotypes using a combination of population genomics and behavioural assays to test patterns of reproductive isolation; *ii*) reconstruct the evolutionary history of these ecotypes; and *iii*) identify regions of the genome associated with adaptive divergence in relation to migration and feeding transitions.

## METHODS

### Collection & maintenance of experimental animals

Between October and April three ecotypes of lampreys were captured in static, double-funnel traps (Morris & Maitland 1987) as they migrated upstream to spawning grounds in the Endrick Water, Loch Lomond, Scotland, U.K. (56°3’17·3” N, 4°27’16·2” W) (Fig. 1). Twelve individuals (six male, six female) each of: *i*) anadromous-parasitic *Lampetra fluviatilis* (hereafter ‘anad-P’); *ii*) freshwater resident-parasitic *L. fluviatilis* (hereafter ‘lake-P’); and *iii*) freshwater resident-nonparasitic *L. planeri* (hereafter 'non-P') were separated based on published phenotypic descriptions (Morris 1989; Renaud 2011). Specifically, anad-P and non-P can be easily distinguished by their discrete body sizes as adults (non-P rarely exceeds 170 mm whereas anad-P can exceed 490 mm total length (TL) (Renaud 2011). Lake-P is readily separated from the others due its black coloration compared with olive-brown colors of the other two (Morris 1989). Each ecotype was held separately in 175 L tanks using water drawn from Loch Lomond at ambient temperature in a flow-through system, and exposed to artificial light that tracked natural photoperiod. The same individuals (*n* = 36) were used in all methodological approaches outlined below.

## 1. Differentiation of Ecotypes

### ***Genetic comparison of ecotypes***

#### *DNA sequencing and bioinformatic processing*

A small piece of tissue was removed from the second dorsal fin of each individual and stored in 95% ethanol at -20°C until required. DNA was extracted using Qiagen DNeasy Blood & Tissue Kits (QIAGEN Sample & Assay Technologies, Copenhagen, Denmark) following the manufacturer's instructions.

Extracted samples (*n* = 36; at least 5 μg per sample) were sent to the GenePool (now Edinburgh Genomics, University of Edinburgh) where a paired-end restriction-site associated DNA (RAD) library was prepared using *SbfI* enzyme digestion (Baird *et al.* 2008). The library was sequenced on an Illumina HiSeq v3 on a single sequencing lane, with a target read length of 100 bp. A total of *c.* 305 million reads were generated and analysed using STACKS (v. 0.99994, Catchen *et al.* 2011). Reads were first assigned to each of the sequenced individuals, identified by a specific barcode at the beginning of each read and trimmed to a final length of 93 bp. All individual reads were then merged into stacks of loci, by allowing identical reads and those with up to two variant positions to be merged into a single locus. Since there is not a closely related reference genome (closest relative sequenced is *Petromyzon marinus*, ~16 million years divergence [Kuraku *et al*. 2006]), loci were assembled *de novo*. A catalogue of loci was built based on all individuals, and each individual's loci were then matched to this catalogue. Each locus was assigned to a specific ID and contained information on which, and how many, individuals contained that locus, what the respective coverage was, and whether it contained single nucleotide polymorphisms (SNPs). Average coverage per locus and individual was 53.4x (median: 48.4x; standard deviation: 23.0x). Loci with coverage higher than 8x, present in all three ecotypes, and in at least 50% of the individuals of each ecotype were extracted for further genomic analyses. A single individual ('B09 Anad') of the anad-P ecotype was excluded from all further genetic analyses because of limited data (median coverage 5x).

### *Differentiation of ecotypes*

A total of 7678 loci were analysed using STRUCTURE (v. 2.3.4, Pritchard *et al.* 2000). For polymorphic loci containing more than one SNP, only the first was included to minimize the number of linked markers. Polymorphic loci that contained a SNP differentially fixed between ecotypes and outlier loci (see *Detecting fixed SNPs and outlier loci*) were excluded from further analysis, since these loci might be under selection and would not represent neutral genetic differentiation between ecotypes. Population differentiation was assessed using a model of admixture and correlated allele frequencies. A total of 500,000 Markov-chain Monte Carlo (MCMC) repetitions were performed after discarding a burn-in of 100,000. We tested for up to five different genetic clusters ($K$ = 1-5) in four consecutive independent runs. To compare likelihoods between different numbers of genetic clustering, $\Delta K$ was estimated (Evanno *et al*. 2005). Using the most likely number of clusters identified, admixture of each individual was assessed using the $Q$-value in STRUCTURE, which indicates the proportion of each individual's genome assigned to each of the pre-defined genetic clusters. In addition, a co-ancestry matrix and population structure were estimated using FineRADStruture (Malinsky et al. 2018), using default parameters. Degrees of admixture between individuals and genetic clusters were compared to assess introgression and hybridization at the individual level and ecotype/genetic clusters.

Additionally, a Principal Coordinate Analysis (PCoA) was performed on the same set of polymorphic loci with the programme GenAlEx (v. 6.5., Peakall & Smouse 2012), with interpolation of missing genotypes. We estimated the genetic variance explained by differentiation among ecotypes using the AMOVA $F_{\mathrm{st}}$ implemented in GenAlEx and using 1000 permutations. The proportion of heterozygous loci ($H_{\mathrm{obs}}$) was calculated for each individual from the three ecotypes separately, and then compared between ecotypes. On average, 16,541 loci (including monomorphic sites) were extracted per individual (9,177 - 19,440 loci) to derive these estimates.

## ***Behavioural comparison of ecotypes***

### *Preparation of experimental animals*

Following fin-clipping for genetic analysis all lamprey were returned to holding tanks, where they were examined periodically to assess the progress of sexual maturation. Sexually mature female lampreys were identified when they became swollen with eggs that were visible through a patch of translucent skin near the cloacal opening and had developed a post-cloacal fin-fold.

Sexually mature male lampreys were identified by an obvious genital papilla that extended several millimeters from the cloacal opening. All individuals were fully ripe (ovulating and spermiating) prior to undertaking experiments.

Each of the 36 individuals from the three ecotypes (six males, six females per ecotype) used for genetic analyses were included in mate choice tests. Lampreys were anaesthetized using a benzocaine solution, measured to the nearest 1 mm $L_T$ and had a Visible Implant Alpha tag inserted beneath the skin on the dorsal surface of the branchial chamber to allow for easy identification during/after the trial period. Prior to inclusion in tests, lampreys were held in 10 L tanks as same-sex, same-ecotype pairs to prevent spawning. These tanks were maintained under the same conditions as the holding tanks.

*Experimental design*

Behavioural differences between ecotypes were examined by employing a multi-male mate choice design, commonly used for testing for the presence of pre-zygotic barriers to gene flow in closely related taxa (e.g. Kozak *et al.* 2013). An artificial stream measuring 5.72 m in length was used to simulate natural conditions as closely as possible during each trial. The base of the stream was covered to a depth of approximately 4 cm by gravel (0.5 – 2 cm diameter) collected locally. The stream was partitioned by fine mesh (1 mm) screens, creating six discrete experimental arenas measuring 91 cm long by 58 cm wide. Water from Loch Lomond was pumped through the stream at velocities of 5 – 20 cm $s^{-1}$, and temperatures ranged 8.5 – 11.5°C, which are typical of the conditions experienced by these lamprey in the wild (Maitland 1980). Artificial, low-light levels on a natural photoperiod were maintained throughout to aid observations.

Mate choice trials comprised a single focal female and three males (one representing each ecotype). The members of each of these male triads remained constant throughout. Thus, each of the 18 females was exposed to the same six male triads, resulting in a total of 108 trials. For each trial, groups of the four experimental animals were placed in separate stream sections and allowed to acclimate for five minutes before observations began. Each trial lasted six hours, during which time all reproductive activity was recorded (direct observations). To prevent exhaustion of the female's egg stock, they were removed after one trial and allowed to rest for a minimum of six hours. Males were rested after two consecutive trials. Only females from the same ecotype were tested in successive trials, to reduce any possible residual effects from pheromones, or other chemical stimuli. Following each trial, the stream section was examined for eggs, which were collected using a siphon. The gravel was then scoured to remove any traces of nests, and that section allowed to dry for at least 12 hours. After all six females from an ecotype were tested with the six male triads, the stream water was drained for a period of 24 hours, and the gravel scoured before refilling with fresh drawn water. The experimental period lasted 15 days, during which no experimental animal died.

Lamprey spawning comprises several discrete behaviours (Johnson *et al.* 2015). Typical spawning occurs within nests created in shallow gravel beds and begins when females attach to an object immediately upstream of the nest. A male then glides along the female's body, attaches to the female's head and wraps his tail around her trunk, forming a tight loop that acts to express the ova from her body. Both partners vibrate rapidly as gametes are released and fertilization takes place externally within the nest. The complete sequence of behaviours leading to potential production of progeny thus begins with a male attaching to a female and culminating in the release of gametes by both sexes. However, male attachment does not always follow a gliding movement along the female's body, a female may not respond to a male’s attachment, or she may reject a male by detaching from the nest. Therefore, we considered “successful spawning” as a combined measure of female and male motivation for mating and “frequency of mating attempts” to quantify stimulation of male reproductive behaviours in the presence of particular females.

### *Statistical analysis*

A generalized linear mixed modeling (GLMM) approach was used to determine whether successful spawning (female gamete release) or spawning attempts (stimulation of male reproductive behaviours) varied in relation to female or male ecotype, or differences in size between potential mating partners. Spawning success and whether or not a spawning attempt was made by a given male when presented with a given female (classified as binary response variables) were considered as separate response variables and analysed using the binomial family using the *glmer* function, in the *lmer* package of the statistical programming environment R (R Core Development Team, 2017). Given that ecotypes differed in size, we specifically tested whether interactions between female ecotype*male ecotype, female ecotype*size difference or male ecotype*size difference contributed significantly to the observed variation.  Female ecotype, male ecotype, size difference between potential partners and the three interactions were treated as fixed effects. Since females were tested independently and so could exhibit a choice but males were tested in trios and always had only one potential female to mate with, female ID and male trio were considered as random effects, to account for repeated measures. Model selection was performed using likelihood ratio tests, with fixed variables not significantly altering the likelihood removed from the final model. All random effects were included in the final model, to account for any variance they contributed.

The frequency of spawning attempts by males was also analysed as an integer trait and analysed using the *glmer.nb* function in R, using a negative binomial as the underlying distribution. We also tested whether the amount of admixture displayed by individuals contributed to male mating attempts and successful spawning (both treated as binary responses), by considering "own Q" values (i.e. the proportion of their genome attributed to the dominant genotype found in their own ecotype) from the STRUCTURE analysis. We included the random effects of female_ID and male_trio and considered the fixed effects of female own_Q, male own_Q and their interactions. We also used a contingency chi-squared analysis to determine whether there was a deviation from random mating when differences in the relative frequency of activities in each of the ecotypes were accounted for.

## 2. Estimating evolutionary history and admixture between ecotypes

We assessed the likelihood of twelve alternative scenarios of lamprey demographic history by testing all plausible evolutionary relationships between ecotypes and time points, direction and extents of gene flow (Fig. 2). We used FASTSIMCOAL2 (Excoffier *et al.* 2013) to compare all alternative scenarios of past history for ecotype evolution. The dataset included 7665 biallelic loci (i.e. excluding loci with more than two variants). We performed 25 runs of 100,000 simulations for each of the twelve evolutionary scenarios. Subsequently, we selected the run with the highest likelihood score for each scenario and used the parameters estimated as priors for an additional 1 million simulations to obtain final maximum likelihoods for each scenario. We compared these likelihoods and calculated the AIC to select the best model. To assess how robust the results obtained from FASTSIMCOAL2 analyses were, we performed additional simulations using the Approximate Bayesian Computation software DIYABC (Cornuet *et al.* 2014). Scenarios seven to twelve included gene flow not resulting in the formation of a new lineage, an option currently not implemented in DIYABC and therefore not included here. For the first six scenarios, Hudson's algorithm was applied to the simulated SNP dataset (Fig. 2). We generated 5 million datasets that were compared to the empirical data to assess how likely each of these scenarios was. This was done by computing posterior probabilities (PP) for each model using the direct estimate implemented in the programme (number of times simulated data of a model is closest to the observed data) and a logistic regression estimate (regression of proportion of the model supported and distance of the simulated data to the observed data) including 95% confidence intervals. Estimates from the DIYABC simulations were compared to the FASTSIMCOAL2 likelihoods (Table S1). In addition, we performed a TREEMIX search. First, migration events were set to 0 (no admixture) and model residuals compared. Migration was then increased until model residuals were equivalent between ecotypes (Pickrell & Pritchard 2012). Three-population *f3* statistics (Reich *et al.* 2009) were estimated to assess whether lamprey ecotypes were significantly admixed. Negative values in *f3* and in the *Z*-score indicate that a lineage has experienced admixture events.

## 3. Predicting candidate genes associated with adaptive divergence

To identify genomic regions that may have facilitated adaptive differentiation, candidate SNPs fixed between ecotypes were detected by extracting loci that were differentially fixed in at least two ecotypes (for example 'A' in one, 'G' in another and 'A/T' in the third). Included loci had to be present in at least two ecotypes and in at least four individuals per ecotype, resulting in a final set of 7726 SNPs. In a second approach, a global outlier analysis was performed using the programme BayeScan (v. 2.1, Foll & Gaggiotti 2008), with default conditions using loci with a minor allele frequency (MAF) of 0.05. Additionally, outlier analyses between migration strategy (anadromy *vs*. residency) and feeding type (parasitic *vs*. nonparasitic) were performed. All loci that had a probability higher than $P = 0.5$ were treated as putative outliers.

Outlier loci and those that were differentially fixed between ecotypes were individually checked against the Ensembl (v. 71, Flicek *et al.* 2014) sea lamprey (*P. marinus*) genomic database to identify whether they were present in transcribed or non-transcribed genomic locations. If they were part of a transcribed gene, the gene name was recorded and its function and process was checked on the National Center for Biotechnology information (NCBI) gene database (http://www.ncbi.nlm.nih.gov/gene/). Genomic location and gene ontology were similarly checked for the SNPs differentially fixed between ecotypes.

Catalogued loci were identified from the first read and then assembled into contigs using both the first and the second reads from the paired-end sequencing. A BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed to match those loci to the genome of the distantly related sea lamprey to identify their genomic location. At present, the sea lamprey genome consists of more than 2500 unlinked scaffolds, making it difficult to gain a general overview of the distribution of identified loci across the genome. However, we used the linkage groups identified in a previous RAD analysis of the sea lamprey (Smith *et al.* 2014) to label each scaffold that we found with a RAD marker and its position on the respective linkage group. If several markers were matched to a single scaffold, that scaffold was tabulated with the marker with the lowest position. To visualize if genomic regions showed evidence of selection in the global analysis and the ecotype contrasts, RAD loci were indexed with positional information and their $F_{st}$ values were plotted along the genetic map.

# RESULTS

## 1. Differentiation Among Ecotypes

### *Genetic comparison of ecotypes*

Three genetic clusters were strongly supported by the STRUCTURE analysis ($\Delta K = 118$, *vs*. < 53 in $K = 2$, 4 and 5; Figs. 3, S1, S2). Although these clusters were broadly consistent with the three ecotypes (anad-P, lake-P and non-P), levels of admixture varied significantly between them (Fig. 3). The highest degree of genetic distinctiveness was exhibited by the lake-P ecotype ($Q_{\text{lake-P}} = 96.8\%$), which showed low levels of admixture with the other two ecotypes ($Q_{\text{anad-P}} = 1.7\%$, $Q_{\text{non-P}} = 1.5\%$). In comparison, the non-P ecotype was almost completely admixed. Genotypes within the non-P ecotype were also assigned at relatively high frequency to both anad-P ($Q_{\text{anad-P}} = 37.7\%$) and lake-P ($Q_{\text{lake-P}} = 15.2\%$) STRUCTURE clusters. However, for anad-P individuals, one of the females and one male also showed high levels of admixture with the non-P type ($Q_{\text{anad-P}} = 64\%$, 51%; $Q_{\text{non-P}} = 35\%$, 41%), while all other individuals showed a minimum of 91% match to their own ecotype (Tables S2, S3).

Three individuals assigned to the non-P ecotype ('B43 Resi', 'B44 Resi' and 'B45 Resi') were genetically assigned to the lake-P cluster (Fig. 4, Fig. S4). In addition, no individual morphologically assigned to the non-P ecotype based on these prior assumptions was found to have a non-admixed genome (highest individual value, $Q_{\text{non-P}} = 54.1\%$). Finally, all individuals morphologically assigned to the anad-P ecotype were found to fall within the anad-P genetic cluster, and levels of admixture with the other two ecotypes were low (average $Q_{\text{lake-P}} = 2.78\%$, average $Q_{\text{non-P}} = 8.74\%$). Estimates of heterozygosity corroborated these observed patterns, with the non-P ecotype exhibiting significantly higher values compared with the other two ecotypes (Tukey's HSD, $P < 0.0001$, Fig. S2).

Consistent with the STRUCTURE analysis, the PCoA returned three separate clusters, consistent with the three ecotypes (Fig. 4*b*). The anad-P and lake-P ecotypes were clearly differentiated on the axis that explained the majority of genetic variation (PC1, 33.5%). Although the non-P ecotype was not clearly differentiated on PC1, having an intermediate position between the other two, the second axis clearly separated the non-P ecotype from the others (PC2, 17.9%). However, in accordance with STRUCTURE analysis, the same three individuals (all males) previously mentioned assigned to the non-P ecotype clustered with individuals belonging to the lake-P ecotype. Notably, all three of these individuals spawned with females of all three ecotypes but showed a much higher number of attempted matings with lake-P females than non-P or anad-P (Table S2). Euclidean distances derived from the PCoA were smallest between the non-P and lake-P ecotype, and greatest between the lake-P and anad-P ecotypes.

Pairwise population differentiation estimates ($F_{st}$) indicated that the lake-P ecotype was highly differentiated from the anad-P ecotype ($F_{st} = 0.154$), whereas the non-P ecotype was less differentiated from both anad-P ($F_{st} = 0.062$) and lake-P ecotypes ($F$st = 0.049). These data are consistent with the population genetic analysis that revealed high levels of admixture in the non-P ecotype.

### *Behavioural comparison of ecotypes*

Average body length of anad-P individuals was 296 mm (range = 258-327 mm), for lake-P, 221 mm (range = 193-283 mm), and for non-P 146 mm (range = 129-164 mm), with males being smaller than females for the lake-P (214 mm *vs* 228 mm) and anad-P (294 mm *vs* 299 mm) but not non-P (both 145 mm on average) individuals tested, confirming that ecotype was not independent of size difference. However, mean body size differences did not preclude between-type mating attempts. Overall, spawning activity by males was recorded on 966 occasions from 72 trials (*n* = 108 trials total; Tables 1, S2, S3); all males tested attempted to spawn with females of more than one ecotype and all attempted to spawn with both anad-P and lake-P females (see Totals under Male Mating Attempts in Table 1). In contrast, few males showed reproductive responses to nonparasitic females (5/6 non-P males and a single lake-P male; Table S3). Three of the males phenotypically classified as non-P that spawned the most frequently ('B43', 'B44' and 'B45') were genetically assigned as lake-P in both the PCA and admixture analyses ($Q_{\text{lake-P}}$ = 94%, 88%, and 90%, respectively); they all showed a strong preference for lake-P females but also attempted to mate with the other two ecotypes (Table S4).

There was more substantial variation among females in their propensity to release gametes following initiation of reproductive behaviour by males but there was little evidence of assortative mating by ecotype (Table S4). Release of gametes by females occurred in 32 pairings: eight anad-P females (stimulated by four anad-P and four lake-P males); 23 lake-P (eight non-P, 11 lake-P and four non-P males); and only one non-P (who was stimulated only by one of her own conspecifics).

The qualitative patterns were supported by the GLMM analyses, which took into account the non-independence of individual females and male trios and attempted to disentangle the effects of size difference and ecotypes. When male mating attempts were treated as a binary response variable, the best-fitting model based on likelihood ratio tests included female ecotype, male ecotype and their interaction, but did not include an individual effect of size difference (Table S5). The significant interaction was driven by an increase between lake-P females and non-P males (estimate = 1.4373; *P* = 0.0446) compared to the intercept. Similar conclusions were reached when considering spawning frequency under a negative binomial distribution but there was no effect of male ecotype*size difference (df = 2, $\chi^2$ = 3.711, *P* = 0.1637; AIC = 1130.3) and a much larger difference between the effects of female ecotype*size difference (df = 2, $\chi^2$ = 13.624, *P* = 0. 0011, AIC = 1120.4) and female ecotype*male ecotype (df = 4; $\chi^2$ = 20.386, *P* = 0. 0004, AIC = 1117.6).

For "successful" spawning in terms of female gamete release, since only a single non-P female spawned, statistical comparisons were made only between anad-P and lake-P females. No significant interactions were found but the final model included the individual effects of female ecotype (df = 1, $\chi^2$ =5.3177, *P* = 0.021, AIC = 146.82), male ecotype (df = 2, $\chi^2$ = 11.609, *P* = 0.003, AIC = 150.63) and size difference (df = 1, $\chi^2$ = 5.8003, *P* = 0.016, AIC = 146.34). More lake-P females spawned than anad-P (estimate = 3.636; *P* = 0.00355); both lake-P (estimate = -2.602; *P* = 0.01359) and non-P (estimate = -6.209; *P* = 0.00144) males induced females to spawn less than anad-P males. Increasing size difference decreased female gamete release (estimate = -0.073; *P* = 0.0717).

Considering admixture of interacting mating partners on male gamete release (considered as a binomial variable), the GLMM indicated no significant interaction between female_own Q and male_own Q or male_own Q on its own but males attempted to mate significantly more with less admixed females (i.e., with a higher own_Q; LR compared to intercept model: df = 1, $\chi^2$ = 23.963, *P* = $9.8 \times 10^{-7}$; estimate = 5.847; *P* = $4.15 \times 10^{-9}$). This effect was not apparent when excluding nonparasitic females, who were the most admixed. For female gamete release, there was a significant interaction between male and female admixture (df = 1, $\chi^2$ = 8.3915, *P* = 0.00377). Considering only parasitic females, there was no significant interaction but both male (LR: df = 1, $\chi^2$ = 7.0274, *P* = 0.008; estimate = 1.9884, *P*

= 0.01564) and female (LR: df = 1, $\chi^2$ = 7.3425, $P$ = 0.0067; estimate = 51.3658, $P$ = 0.01374) admixture decreased successful mating in terms of female gamete release. Only a single anad-P female (B03) showed extensive admixture ($Q_{\text{anad-P}}$ = 64%; $Q_{\text{non-P}}$ = 35%) and this individual failed to spawn with any male. None of the other anad-P females showed q values lower than 0.91 and the lake-P females were highly "pure" (q = 0.95-1). Similarly, only a single anad-P male (B05) was extensively admixed ($Q_{\text{anad-P}}$ = 51%; $Q_{\text{non-P}}$ = 41%); he only induced a single female (lake-P) to release gametes, although he attempted to mate with females from all ecotypes.

## 2. Estimating evolutionary history and admixture between ecotypes

Although there was some inconsistency among the best-fitting models found using different approaches to predict the past history of ecotype evolution, all analyses supported the conclusion that a strictly bifurcating scenario was unlikely. There was less consensus on whether both parasitic forms contributed to introgression into non-P, or only one form.

Coalescent simulations using FASTSIMCOAL2 showed that all strictly bifurcating tree-like scenarios, which do not include introgression between ecotypes (i.e. scenarios *i* – *iii*), were unlikely. Of the simulations allowing admixture, scenario *v* was most likely, suggesting that the anad-P split first from the ancestor of the two freshwater resident ecotypes, with a subsequent split between the lake-P and non-P ecotypes and introgression from anad-P into non-P. The second best model (ΔAIC = 8.4 compared to scenario *v*, Table S2) was scenario *vii*, which is similar but also includes secondary contact and gene flow from lake-P into non-P after divergence. All other tested models were considerably less likely (ΔAIC > 10, Table S2). For the DIYABC analyses, posterior probability estimates supported scenario *iv* as the most likely demographic history supported by the data (direct approach: PP = 0.703, CI = 0.576-0.829; logistic approach: PP = 0.682, CI = 0.465-0.885; see Fig. 2; Table S1). This suggests that the non-P ecotype arose after splitting from the anad-P ecotype, followed by substantial introgression with the lake-P ecotype, sometime after it split from the common ancestor with anad-P. In agreement with the coalescent analysis, all strict isolation scenarios were unlikely. However, scenario *i*, implying a common ancestor for the two freshwater resident ecotypes, showed the second highest PP in the logistic approach (PP = 0.318, CI = 0.114-0.535). In the direct approach, the second best model was scenario *vi* suggesting that non-P originated *via* hybridization between anad-P and lake-P (PP = 0.180, CI = 0.074-0.287). Substantial admixture was further corroborated by the TREEMIX analysis, which suggested unidirectional gene flow from the anad-P to the non-P ecotype, also supporting scenario *v*. Three-population *f3* statistics and the *Z*-score also suggested significant levels of introgression from anad-P and lake-P into non-P (*f3* = -0.0032; *Z*-score = -5.39), but no significant signals of admixture in the other two ecotypes (*f3* > 0.0191; *Z*-score > 19.82).

### 3. Predicting candidate genes associated with adaptive divergence

Consistent with the STRUCTURE analyses, the majority of differentially fixed SNPs (*n* = 10, from 16 SNPs) were identified between anad-P and lake-P ecotypes (Table S6). None of the substitutions in these differentially fixed SNPs were non-synonymous. A global outlier analysis detected 60 out of 5286 loci that had a probability > 50% of being outliers based on comparison between the three ecotypes (Fig. 5). Out of those 60 outliers, four mapped to the lamprey genome and were located in a gene (Table S7). Two of these were involved in cytoskeleton binding and transport, while the other two were involved in RNA binding or apoptosis and viral infection. An outlier analysis contrasting anadromous (anad-P) and resident individuals (lake-P and non-P) detected a total of 13 loci (total $n = 2509$), including eleven of the outlier loci detected in the global analysis plus an additional two outliers. Four of these outliers mapped to three different genes, with two outliers mapping to a different position in the same gene (Table S7). There was one synonymous and one non-synonymous substitution in this gene (CD 109). The three genes are reported to be involved in protein maturation and lipidation, apoptosis and central nervous system development. A third analysis, comparing parasitic (anad-P and lake-P) and non-parasitic (non-P) individuals, identified 22 outliers from a total of 3263 loci, of which four were shared with the outliers detected in the global analysis (Fig. 5). Four of these mapped to a gene, including one that is part of the dystrophin protein family (RAD-ID: 89575). Dystrophin is part of a protein complex connecting the cytoskeleton to the extracellular matrix (Pasternak *et al*. 1995). The other three genes are involved in transmembrane transport, embryo implantation, extracellular matrix organization and DNA replication.

When contrasting $F_{st}$ values between anadromous (anad-P) and resident individuals (lake-P and non-P), there was no apparent pattern along the linkage groups and outliers were randomly distributed (Fig. 5). In contrast, there were several loci with higher $F_{st}$ values on linkage group six (LG6) when contrasting parasitic (anad-P and lake-P) and non-parasitic individuals (non-P). A t-test revealed significantly higher $F_{st}$ values in LG6 compared with the rest of the genome ($t = 5.47$, $P < 0.001$).

## DISCUSSION

Our study has shown that three highly differentiated ecotypes of lampreys occurring sympatrically in the Loch Lomond catchment exhibit substantial variability in genetic and behavioural differentiation that does not conform to current taxonomic schemes. Similar conclusions have been drawn previously, for the species considered in this study and for others (Yamazaki & Goto 1997, 2000; Kucheryavyy *et al.* 2007, 2016; Artamonova *et al.* 2011; Yamazaki *et al.* 2011; Spice *et al.* 2014; Makhrov & Popov 2015; Vatandoust *et al.* 2015; Rougemont *et al.* 2015, 2016; Mateus *et al.* 2016). However, the inclusion of a freshwater-feeding parasitic ecotype (lake-P) from a post-glacial lake basin in this current study enables new insights into the evolutionary history, possible process of divergence, and taxonomic distinctiveness of alternative life history strategies in lampreys that has not been possible previously. Specifically, we were able to consider genomic differentiation associated with two distinct evolutionary transitions: from anadromy to freshwater-residency and from parasitic to nonparasitic life history strategies.

Because freshwater-parasitic lamprey ecotypes are putatively intermediate in life history strategy between marine-parasitic and freshwater nonparasitic, it has been frequently postulated that they represent a necessary transitional stage during the colonization of post-glacial regions (Zanandrea 1959, 1961; Beamish 1985; Morris 1989; Salewski 2003; Hardisty 2006). Our data are somewhat consistent with this hypothesis, as demographic reconstructions suggests a common ancestry for both freshwater ecotypes. However, the phenotype of that ancestor – and whether or not it was parasitic – remains uncertain. Consistent with studies conducted in other regions (Rougemont *et al*. 2015, 2017), we found that the nonparasitic freshwater ecotype was genetically introgressed with the other two forms. However, intriguingly, the freshwater parasitic ecotype in our study showed the highest degree of reproductive isolation, despite appearing highly attractive to other ecotypes during spawning and being intermediate in life-history and morphology. In nature, these freshwater parasitic ecotypes of typically anadromous parasitic species occur only rarely (e.g. Maitland 1980;

Tsimbalov *et al.* 2015), which could suggest a fitness cost relative to the more commonly found non-parasitic freshwater forms. Alternatively, populations such as the lake-P ecotype of Loch Lomond could represent a life history transition that only evolves under certain geographic contexts (e.g. allopatry in large lake basins containing abundant foraging opportunity during glacial advance-retreat, Collett 1905; Tuunainen *et al*. 1980; Morris 1989; Inger *et al*. 2010; Tsimbalov *et al*. 2015). We suggest that the formation of lamprey ecotypes in Loch Lomond, and the European river and brook lamprey "species pair" more generally, are dependent on an interplay between ecological opportunity and introgression.

### ***Differentiation and introgression of lamprey ecotypes***

Genetic variability, based on RAD sequencing, is not consistent with currently recognized taxonomic units. The lake-P ecotype, currently considered *L. fluviatilis*, was found to be more highly genetically distinct from the anad-P ecotype (also considered *L. fluviatilis*) than the non-P ecotype (considered *L. planeri*), which was highly admixed with the other two. Demographic analyses ruled out that the two parasitic ecotypes (anad-P and lake-P) share a more recent common ancestor than with the non-P ecotype (Fig. 3, Table S1). Intriguingly, three males that would have been classified as non-P based on their morphology showed a Q value of less than 11% for their own ecotype but greater than 89% for the lake-P ecotype. This could suggest that the nonparasitic phenotype is controlled by a relatively small portion of the genome, which was supported by our finding of a single genomic region strongly associated with the transition to nonfeeding. There was no evidence for admixture in the lake-P individuals but a single male and a single female that were morphologically anad-P showed extensive levels of admixture with the non-P ecotype. Additionally, the lake-P form appears to contain a subset of mtDNA haplotypes found in the non-P population from Loch Lomond (Table S8; data from Hume

2013). Only two mtDNA haplotypes were found in the lake-P individuals sampled: one that was found in all Scottish populations sampled and one (which showed a single bp mutation from the common haplotype) that was shared with non-P but was not found among the anad-P individuals sampled. Although based on only a small sample size, both anad-P and non-P from the same catchment showed more mtDNA variation than the lake-P ecotype, suggesting that the latter could have experienced a more substantial population bottleneck in the relatively recent past or less historical introgression relative to the other two ecotypes.

Across Europe, anad-P (*L. fluviatilis*) and non-P (*L. planeri*) lampreys appear to have resulted from multiple and parallel speciation events (Mateus *et al.* 2016; Rougemont *et al.* 2016). Based on mtDNA and nuclear gene markers, no species-specific fixed differences have been found (Schreiber & Engelhorn 1998; Espanhol *et al.* 2007; Pereira *et al.* 2011) and neither do they exhibit consistent microsatellite allele frequency differences across their range (Bracken *et al.* 2015; Rougemont *et al.* 2015; Mateus *et al.* 2016). A detailed comparison among sympatric and parapatric populations of the two "species" in France using both microsatellite genotypes (Rougement *et al.* 2015) and genome-wide, single nucleotide polymorphisms based on RAD sequencing (Rougemont *et al.* 2016) provides little evidence for a restriction of gene flow between them. Secondary contact and introgression between lamprey ecotypes and demographic models with asymmetric gene flow from the anad-P into the non-P were also evident from these locations (Rougemont *et al*. 2017), consistent with our data from Loch Lomond. Hybridization and introgression are increasingly recognized as evolutionary forces creating diversity and possibly promoting the evolution of differentiated ecotypes (e.g. charr, Garduño-Paz *et al*. 2012; Jacobs *et al*. 2017; crater lake cichlids, Martin *et al*. 2015; whitefish, Rougeux *et al*. 2017). For example, secondary contact and introgression between evolutionary lineages of post-glacial fishes seems to frequently result in ecological divergence and the formation of new ecotypes (Jacobs *et al*. 2017; Rougeux *et al*. 2017). Overall, phylogeography indicates that gene flow between anad-P (*L. fluviatilis*) and non-P ecotypes (*L. planeri*) is opportunistic, and introgression likely varies in accordance with contemporary and historic barriers (physical or environmental) separating these migratory and resident ecotypes (Bracken *et al.* 2015; Mateus *et al*. 2016).

Given this high potential for gene flow and sensitivity to ecological disturbance, most ecotypes formed in sympatry are thus unlikely to attain the status of "good biological species" (Mallet 2008; Hendry *et al*. 2009; Weissing *et al*. 2011; Lowry 2012; Østman *et al*. 2014; Rudman & Schluter 2016). However, revealing those circumstances that do lead to reproductive isolation is critical to understanding speciation. In our study, despite clear genetic differentiation of the lake-P ecotype from the other two, there was little evidence of pre-zygotic isolation among them, due particularly to the high attractiveness of lake-P females to all males and their willingness to mate with any of the three ecotypes tested. Thus, there does indeed appear to be the potential for gene flow among ecotypes if they were to spawn in syntopic habitats. The inclusion of an intermediate sized lake-P ecotype enables more fine-scale testing of the prediction that size-assortative mating is the primary driver of reproductive isolation between diverging lamprey ecotypes (Beamish & Neville 1992). We found evidence for an interaction between body size difference and female ecotype, which explained nearly as much variation as the interaction between male and female ecotypes for male mating behaviours – but, there was no separate effect of body size when the latter interaction was accounted for. In our assays, the interaction was driven largely by reduced mating between small non-P and large anad-P individuals; anad-P females did not prefer their own conspecifics over lake-P males and anad-P males attempted to mate with females from all ecotypes. Thus, while size is certainly a factor in lamprey mate choice, the relative size differences between ecotypes does not preclude the potential for gene flow.

### ***Factors influencing ecological speciation***

The patterns of introgression among lamprey ecotypes in this study invokes three non-mutually exclusive hypotheses generally explaining ecological speciation in sympatry: *i*) temporal or spatial isolation restricts gene flow between ecotypes under current natural conditions; *ii*) frequency-dependent encounter rates influence mate choice between ecotypes; and *iii*) hybrids between ecotypes suffer high fitness costs as a consequence of intrinsic or extrinsic post-zygotic barriers. Our results are consistent with a combination of these factors influencing the parallel differentiation of ecotypes in lampreys, which could be more broadly representative of post-glacial expansion of freshwater fishes.

Firstly, the importance of geographical isolation to the acquisition of adaptive differences between populations via divergent selection has been recognized since Darwin's time (Mallet 2008). The role of physical separation of lamprey populations in facilitating genetic divergence has been highlighted previously, where migratory ecotypes become separated from stream-resident forms as they pursue foraging opportunities in large river systems, lakes, or marine environments (e.g. due to geological or climate shifts, construction of in-stream barriers, Zanandrea 1959; Hardisty & Potter 1971; Kan & Bond 1981; Yamazaki *et al*. 2011; Bracken *et al*. 2015). While some intriguing patterns of parallel speciation and adaptation have been revealed previously by genetic analysis, suggesting some degree of isolation between lamprey ecotypes (Mateus *et al.* 2013, 2016; Bracken *et al.* 2015; Rougemont *et al.* 2015; 2016), there is strong evidence that lampreys can continue to exhibit multiple ecotypes yet exist in true sympatry (Beamish 1987; Yamazaki *et al*. 1998; Kucheryavyy *et al*. 2007; this study). Post-glacial fishes, particularly Arctic charr, exemplify this "problem" (Klemetson 2010), with evidence of ongoing gene flow between charr ecotypes apparent in some regions (e.g. Salisbury et al. 2017). However, site fidelity and associated assortative mating among ecotypes of charr is required to maintain discrete forms in sympatry (Jonsson & Jonsson 2001). Observations in nature have provided evidence against spatial or temporal segregation of many lamprey ecotypes on spawning grounds (Morman 1979; Manion & Hanson 1980; Kucheryavyy *et al*. 2007; Cochran *et al*. 2008; Lasne *et al*. 2010). Furthermore, sneak male mating tactics among ecotypes and between species have also been documented (Cochran 2008; Hume *et al*. 2013a). Thus, strict reproductive isolation is an unsatisfactory explanation for the maintenance of discrete lamprey ecotypes across entire geographic ranges. However, recent work suggests that in fact more variable and larger ecosystems can support more variable populations or multiple ecotypes as a consequence of greater ecological opportunity (Mahler *et al*. 2013; Præbel *et al*. 2013; Wellborn & Langerhans 2014; Recknagel *et al*. 2017). This suggests that the Loch Lomond system, a large and deep lake itself, is an important but uncommon ecological opportunity driving the formation of the two freshwater parasitic lamprey ecotypes in Europe.

Secondly, sexually reproducing organisms can benefit from selecting between multiple potential mates; but being too choosy can incur high fitness costs in terms of potential loss of reproductive opportunity, energetic costs and high mortality (Andersson 1994). One important factor influencing mate choice is encounter rate, where low rates increase the risks of not mating and often results in reduced choosiness (Jennions & Petrie 1997). Based on sampling the Loch Lomond catchment during the lamprey spawning migration, the anad-P ecotype is rare, the non-P ecotype the commonest, and the lake-P ecotype somewhat intermediate (Morris

& Maitland 1987; Morris 1989; Adams *et al*. 2008; Hume 2011, 2013). These relative frequencies are consistent with the relative rarity of freshwater parasitic populations of lampreys compared to non-feeding forms in northern lakes in general (e.g. Maitland 1980; Tsimbalov *et al.* 2015). This suggests that continuing to feed in freshwater represents a strategy that only evolves in response to particular ecological opportunity (Schluter 2000; Siwertsson *et al*. 2010; Lowry 2012; Recknagel *et al*. 2017); principally, in river basins with large lakes and abundant foraging opportunity  (Collett 1905; Tuunainen *et al*. 1980; Morris 1989; Inger *et al*. 2010; Tsimbalov *et al*. 2015). There could also be a fitness disadvantage relative to nonparasitic forms in having to exploit new food resources (i.e. freshwater fishes) in a finite environment. If we assume that weak assortative mating occurs naturally in this system (i.e. not an artifact of behavioural assays in this study) then the likelihood of anad-P and lake-P ecotypes mating with the more common non-P ecotype seems high. Thus, if mate choice operates by lampreys only choosing the "correct" ecotype when encounter rates with them are frequent, then logically spawning with the other two should occur when encounter rates with their own kind are infrequent. This is consistent with Rougemont *et al*. (2017) who found higher effective population sizes for anad-P compared to non-P forms, as well as asymmetric introgression between ecotypes. Hybridization has been associated with a scarcity of conspecifics (Mayr 1963, 1970; Avise & Saunders 1984; Leaniz & Verspoor 1989; McGowan & Davidson 1992; Wirtz 1999; Randler 2002; Quilodrán *et al*. 2014). However, the influence of encounter rate between ecotypes on mate choice has rarely been experimentally tested (Willis *et al*. 2011). Our results are consistent with relative frequency of the three ecotypes during the spawning season explaining the patterns of introgression observed.

Thirdly, intrinsic post-zygotic barriers such as hybrid inviability, or extrinsic barriers such as maladaptation of hybrids to parental environments, may also contribute to reproductive isolation between ecotypes (Mallet 2008; Hendry *et al*. 2009; Lowry 2012). Loss of genetic compatibility during adaptive divergence is a principal cause of reproductive isolation (Coyne & Orr 2004). When alleles with beneficial effects in one population are recombined in a hybrid genome with alleles originating at different loci in a second population, hybrid inviability can occur (Burke & Arnold 2001; Presgraves *et al*. 2003). The lack of introgression in the lake-P lamprey ecotype and low admixture in the anad-P individuals in this study could be the result of high mortality of hybrids that adopt a migratory-parasitic life history strategy. Metamorphosis in lampreys is a highly complex, protracted event requiring drastic changes in physiology and anatomy (Manzon *et al*. 2015), and conceivably is a source of substantial mortality of parasitic lampreys. Although there is evidence for high survivorship of hybrid embryos between lamprey ecotypes in the laboratory (Enequist 1937; Piavis *et al*. 1970; Beamish & Neville 1992; Hume *et al*. 2013b; Rougemont *et al*. 2015), the viability of adults, as well as F1 or F2 hybrids remains untested. The low frequency of hybrids between lampreys in the wild has been suggested to be a consequence of selection against those hybrids (Rougemont *et al*. 2017). It is predicted that where incipient speciation is occurring selection against hybrids should also have an ecological component (Schluter 1996); where hybrids experience reduced foraging efficiency (e.g. Hatfield & Schluter 1999) and/or suffer higher rates of mortality (e.g. Nilsson et al., 2017). Therefore, offspring resulting from inter-ecotype matings between lampreys in Loch Lomond may suffer higher fitness costs when they adopt migratory-parasitic strategies, compared to the relatively sedentary and cryptic life cycle of nonparasitic lampreys.

### *Adaptive divergence of lamprey ecotypes*

Despite evidence for gene flow between parasitic and nonparasitic ecotypes, there is still substantial neutral and adaptive genetic differentiation between them. Neutral differentiation may reflect the limited dispersal capability of lampreys confined to freshwater systems (e.g. Taylor *et al.* 2012), as anadromous lampreys consistently exhibit high levels of gene flow among populations across a variety of spatial scales (Spice *et al.* 2012; Hess *et al.* 2013; Yamazaki *et al.* 2014; Bracken *et al.* 2015). The identities of transcribed genes that are differentially fixed, or are genomic outliers, in anadromous-parasitic *vs* resident-parasitic and resident-nonparasitic ecotypes are potentially revealing in regards to adaptive differences in life history strategy between them.

In our study, we found candidate genes associated with different important physiological functions that could differ in relation to a transition between feeding environments. For example, the gene *cd109* may influence growth rate, as it interacts selectively with transforming growth factor-β, a multifunctional peptide that controls proliferation, differentiation and other functions in many cell types (e.g. Hockla *et al.* 2010). Another important factor in the transition between anadromous and resident life history strategies is the immune response (Mateus *et al.* 2013). Communities of pathogens in marine and freshwater environments differ substantially, therefore resident and anadromous lampreys will experience different selection pressures for pathogen resistance. We identified the gene *nckap-1* as a genomic outlier between anadromous and both resident ecotypes. This gene has been previously identified as crucial in resistance against a bacterial pathogen in channel catfish *Ictalurus punctatus* (Zhou *et al.* 2017). This emphasizes the power of outlier analyses, even using genome sampling methods like RAD sequencing in revealing potential genomic regions involved in functionally important traits for adapting to new environments. We do not imply that any of the SNPs identified here are causal mutations for adaptive phenotypes, but encourage that future studies investigate genes and genomic regions associated with adaptive divergence identified here using more fine-scale genomic data.

Outlier analyses also revealed potential differences related to the transition to a nonparasitic life cycle. A single region of the genome (linkage group 6) revealed significant differences between parasitic ecotypes and the nonparasitic form and these may be consistent with directional selection in lamprey feeding mode despite high levels of gene flow (e.g. Nosil & Crespi 2004). The linkage of genes in so-called genomic islands may facilitate the establishment of reproductive isolation between ecotypes and eventually lead to speciation (Feder & Nosil 2010). The *reck* gene is associated with linkage group 6 and is implicated in the organization of extracellular matrix and cell migration. Knockdown experiments of this gene in zebrafish result in impaired vascular integrity and lack of dorsal root ganglia formations, suggesting that *reck* is important during embryonic development (Prendergast *et al.* 2012). This gene is also suggested to have a crucial role during embryonic development in other vertebrates (de Oliveira *et al.* 2010; Yamamoto *et al.* 2012). Two additional genes, *scn4aa* and an orthologue of *rev3l*, were detected as genomic outliers. The former gene codes for a sodium-channel protein and is involved in development of the heart, while the latter is a DNA polymerase also known to function in embryonic development. In mice, gene mutations of *rev31* resulted in smaller neonates with retarded development (Wittschieben *et al.* 2000). Although speculative at present, genes such as these may be vital in lamprey embryological development, and alterations of their expression could impact metamorphosis and potentially be involved in the transition from a parasitic to nonparasitic life history strategy.

### *Conclusions*

Our data suggest that freshwater-resident feeding forms of the typically anadromous *Lampetra fluviatilis* evolve and persist only in large post-glacial lake basins throughout Europe. In Loch Lomond, this ecotype (lake-P) is genetically distinct from the anadromous form (anad-P), whereas a nonparasitic ecotype (non-P here; often referred to as *L. planeri*) was introgressed with the other two. Although we were unable to fully resolve the question of whether lake-P ecotypes are a necessary transitional step in the evolution of nonparasitism among lampreys more generally, we found both lake-P and non-P ecotypes in Loch Lomond do share a common freshwater ancestor. This suggests ecological opportunity, followed by variation in the probability of introgression, strongly influences the extent of divergence between sympatric lamprey ecotypes.

## ACKNOWLEDGEMENTS

We thank A. Adam for help with DNA extractions, Edinburgh Genomics for sequencing and bioinformatic advice, and J. Buckley for assistance during bioinformatic analyses and interpretation. We also thank three anonymous reviewers for their extensive and helpful suggestions that substantially improved this manuscript.

## DATA ACCESSIBILITY

Data can be accessed using BioProject ID PRJNA488599.

# REFERENCES

Adams, C. E., Bissett, N., Newton, J. & Maitland, P. S. (2008). Alternative migration and host parasitism strategies and their long term stability in river lampreys from the River Endrick, Scotland. *Journal of Fish Biology*, 72, 2456-2466.

Andersson, M. (1994). Sexual selection. Princeton University Press, 624 pp.

Artamonova, V. S., Kucheryavyy, A. V. & Pavlov, D. S. (2011). Nucleotide sequences of the mitochondrial cytochrome oxidase subunit I (COI) gene of lamprey classified with *Lethenteron camtschaticum* and the *Lethenteron reissneri* complex show no species-level differences. *Doklady Biological Sciences* 437, 113-118.

Avise, J. C. & Saunders, N. C. (1984). Hybridization and introgression among species of sunfish (Lepomis): analyzed by mitochondrial DNA and allozyme markers. *Genetics* 108(1), 237-255.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A…Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3 e3376.

Beamish, R. J. & Neville, C-E. M. (1992). The importance of size as an isolating mechanism in lampreys. *Copeia* 1992, 191-196.

Beamish, R. J. (1985). Freshwater parasitic lamprey on Vancouver Island and a theory of the evolution of the freshwater parasitic and nonparasitic life history types. *In:* Foreman, R.E., Gorbman, A., Dodd, J.M. & Olsson, R, eds.) Evolutionary Biology of Primitive Fish. Plenum Publishing Corporation pp. 123-140.

Beamish, R. J. (1987). Evidence that parasitic and nonparasitic life history types are produced by one population of lamprey. *Canadian Journal of Fisheries and Aquatic Sciences* 44, 1779-1782.

Bracken, F. S. A., Hoelzel, A. R., Hume, J. B. & Lucas, M. C. (2015). Contrasting population genetic structure among freshwater-resident and anadromous lampreys: the role of demographic history, differential dispersal and anthropogenic barriers to movement. *Molecular Ecology* 24, 1188-1204.

Burke, J. M. & Arnold, M. L. (2001). Genetics and the fitness of hybrids. *Annual Reviews in Genetics* 35, 31-52.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H. & De Koning, D-J. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* 1, 171–182.

Cochran, P. A., Bloom, D. D., Wagner, R. J. (2008). Alternative reproductive behaviours in lampreys and their significance. *Journal of Freshwater Ecology* 23, 437-446.

Collett, R. (1905). Meddelelser om Norges fiske i Aarene 1884–1901 Cristiania: I Kommission hos Jacob Dybwad, pp. 1–165.

Cornuet, J-M, Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R…Estoup, A. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187–1189.

Coyne, J.A. & Orr, H.A. (2004). Speciation. Sinauer, 480 pp.

de Oliveira Demarchi., Zambuzzi, W. F., Paiva, K. B. S., Silva-Valenzuela, M. G., Nunes, F. D., Figeuira, R. C. S., Sasahara, R. M…Granjeiro, J. M. (2010). Development of secondary palate requires strict regulation of ECM remodeling: sequential distribution of RECK, MMP-2, MMP-3, and MMP-9. *Cell and Tissue Research*, 340, 61-69.

Docker, M. F. (2009). A review of the evolution of nonparasitism in lampreys and an update of the paired species concept: Biology, Management, and Conservation of Lampreys in North America pp. 71–114. American Fisheries Society Symposium 72, Bethesda, Maryland.

Enequist, P. (1937). Das Bachneunauge als ökologische Modification des Flussneunauge. Üeber die Flussneunaugen und Bachneunaugen Schwedens. Ark. Zool. 29, 1-22.

Espanhol, R.. Almeida, P. R., Alves, M. J. (2007). Evolutionary history of lamprey paired species *Lampetra fluviatilis* (L.) and *Lampetra planeri* (Bloch) as inferred from mitochondrial DNA variation. *Molecular Ecology*, 16, 1909-1924.

Evanno, G., Regnaut, S., Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9, e1003905.

Feder, J. L. & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution; International Journal of Organic Evolution*, 64, 1729–1747.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S…Searle, S. M. (2014). Ensemble 2014. *Nucleic Acids Research*, 42(database issue), 749-755.

Foll, M. & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180, 977-993.

Garduño-Paz, M. V., Adams, C. E., Verspoor, E., Knox, D. & Harrod, C. (2012). Convergent evolutionary processes driven by foraging opportunity in two sympatric morph pairs of Arctic charr with contrasting post-glacial origins. *Biological Journal of the Linnean Society* 106(4), 794-806.

Hardisty, M. W. (1963). Fecundity and speciation in lampreys. *Evolution* 17(1), 17-22.

Hardisty, M. W. (2006). Lampreys: Life Without Jaws. Forrest.

Hardisty, M.W., Potter, I.C. (1971). Paired species: The Biology of Lampreys, pp. 249–278. Academic Press, London.

Hatfield, T. & Schluter, D. (1999). Ecological speciation in sticklebacks: environmental-dependent hybrid fitness. *Evolution* 53(3), 866-873.

Hays, J. D., Imbrie, J. & Shackleton, N. J. (1976). Variations in the earth's orbit: pacemaker of the ice ages. *Science*, 194, 1121-1132.

Hendry, A. P., Bolnick, D. I., Berner, D. & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology* 75(8), 2000-2036.

Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F. & Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22, 2898-2916.

Hockla, A., Radisky, D. C. & Radisky, E. S. (2010). Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109. *Breast Cancer Research and Treatment*, 124, 27-38.

Hooker, O. E., Barry, J., van Leeuwen, T. E., Lyle, A., Newton, P., Cunningham, P. & Adams, C. E. (2016). Morphological, ecological and behavioural differentiation of sympatric profundal and pelagic Arctic charr (*Salvelinus alpinus*) in Loch Dughaill Scotland. *Hydrobiologia* 783(1), 209-221.

Hume, J. B. (2011). Adult lamprey survey of the Endrick Water SSSI and SAC 2009-2010. *Scottish Natural Heritage Commissioned Report No. 480*.

Hume, J. B., Adams, C. E., Mable, B. & Bean, C. W. (2013*a*). Sneak male mating tactics between lampreys (Petromyzontiformes) exhibiting alternative life-history strategies. *Journal of Fish Biology*, 82, 1093-1100.

Hume, J. B., Adams, C. E., Mable, B. & Bean, C. W. (2013*b*). Post-zygotic hybrid viability in sympatric species pairs: a case study from European lampreys. *Biological Journal of the Linnaean Society*, 108, 378-383.

Hume, J.B. (2013). The evolutionary ecology of lampreys (Petromyzontiformes). PhD thesis, University of Glasgow.

Inger, R., McDonald, R. A., Rogowski, D., Jackson, A. L., Parnell, A., Preston, S. J…Bearhop, S. (2010). Do non-native invasive fish support elevated lamprey populations? *Journal of Applied Ecology*, 47, 121-129.

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N., Alekseyev, S., Hooker, O.,...& Elmer, K. R. (2018). Convergence in form and function overcomes non-parallel evolutionary histories in Arctic charr. *bioRxiv*, 265272.

Jennions, M.D. & Petrie, M. (1997). Variation in mate choice and mating preferences: a review of causes and consequences. *Biological Reviews of the Cambridge Philosophical Society* 72(2), 283-327.

Johnson, N. S., Buchinger, T. & Li, W. (2015). Reproductive ecology of lampreys. *In:* (Docker, M.F. ed.) Lampreys: Biology, Conservation and Control, Vol 1. Springer, Dordrecht pp. 265-304.

Jonsson, B. & Jonsson, N. (2001). Polymorphism and speciation in Arctic charr. *Journal of Fish Biology* 58(3), 605-638.

Kahilainen, K. & Østbye, K. (2006). Morphological differentiation and resource polymorphism in three sympatric whitefish *Coregonus lavaretus* (L.) forms in a subarctic lake. *Journal of Fish Biology* 68(1), 63-79.

Kan, T. T. & Bond, C. E. (1981). Notes on the biology of the Miller Lake lamprey *Lampetra* (*Entosphenus*) *minima*. *Northwest Science* 55(1), 70-77.

Klemetsen, A. (2010). The charr problem revisited: exceptional phenotypic plasticity promotes ecological speciation in postglacial lakes. *Freshwater Reviews* 3, 49-74.

Klemetsen, A. (2013). The most variable vertebrate on earth. *Journal of Ichthyology* 53, 781-791.

Kozak, G. M., Head, M. L., Lackey, A. C. R., Boughman, J. W. (2013). Sequential mate choice and sexual isolation in threespine stickleback species. *Journal of Evolutionary Biology*, 26, 130-140.

Kucheryavyy, A. V., Saavaitova, K. A., Pavlov, D. S. Gruzdeva, M. A., Kuzishchin, K. V. & Stanford, J. A. (2007). Variations of life history strategy of the Arctic lamprey *Lethenteron camtschaticum* from the Utkholok River (Western Kamchatka). *Journal of Ichthyology,* 47(1), 37-52.

Kucheryavyy, A. V., Tsimbalov, I. A., Kostin, V. V., Nazariv, D. Yu., Zvezdin, A. O. & Pavlov. D. S. (2016). Polymorphism of spawners of the resident form of European river lamprey *Lampetra fluviatilis* (Petromyzontidae). *Journal of Ichthyology*, 56(5), 744-751.

Kuraku, S., & Kuratani, S. (2006). Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zoological science*, *23*(12), 1053-1064.

Lasne, E., Sabatié, M-R., Evanno, G. (2010). Communal spawning of brook and river lampreys (*Lampetra planeri* and *L. fluviatilis*) is common in the Oir River (France). *Ecology of Freshwater Fish*, 19, 323-325.

Leaniz, C. G. & Verspoor, E. (1989). Natural hybridization between Atlantic salmon, Salmo salar, and brown trout, Salmo trutta, in Northern Spain. Journal of Fish Biology 34(1), 41-46.

Lowry, D. B. (2012). Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society* 106(2), 241-257.

Lyons, S. (2003). A quantitative assessment of the range shifts of Pleistocene mammals. *Journal of Mammology* 84, 385-402.

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, *341*(6143), 292-295.

Maitland, P. S. (1980). Review of the ecology of lampreys in northern Europe. *Canadian Journal of Fisheries and Aquatic Science,* 37, 1944-1952.

Maitland, P. S., Morris, K. H., East, K., Schoonoord, M. P., van der Wal, B. & Potter, I. C. (1984). The estuarine biology of the river lamprey, *Lampetra fluviatilis*, in the Firth of Forth, Scotland, with particular reference to size composition and feeding. *Journal of the Zoological Society, London,* 203, 211-225.

Makhrov, A. A. & Popov, I. Yu. (2015). Life forms of lampreys (Petromyzontidae) as a manifestation of intraspecific diversity of ontogenesis. *Russian Journal of Developmental Biology*, 46(4), 196-207.

Mallet, J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 363, 2971-2986.

Manion, P. J. & Hanson, L. H. (1980). Spawning behaviour and fecundity of lampreys from the upper three Great Lakes. *Canadian Journal of Fisheries and Aquatic Science,* 37, 1635-1640.

Manzon, R. G., Youson, J. H. & Holmes, J. A. (2015). Lamprey metamorphosis. *In*: (Docker, M.F. ed.) Lampreys: Biology, Conservation and Control, Vol 1. Springer, Dordrecht pp. 139-214.

Martin, S.H., Dasmahapatara, K.K., Nadeau, N.J. et al. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* 23, 1817-1828.

Mateus, C. S., Almeida, P. R., Mesquita, N., Quintella, B. R. & Alves, M. J. (2016) European lampreys: new insights on postglacial colonization. Gene flow and speciation. *PLoS ONE*, 11, e0148107.

Mateus, C. S., Strange, M., Berner, D., Roesti, M., Quintella, B. R., Alves, J. A.,…Salzburger, W. (2013). Strong genome-wide divergence between sympatric European river and brook lampreys. *Current Biology*, 23: R649-50.

Mayr, E. (1963). Animal species and evolution. Belknap Press of Harvard University Press, 797pp.

Mayr, E. (1970). Populations, species, and evolution: an abridgement of animal species and evolution. Belknap Press of Harvard University Press, 453pp.

McGowan, C. & Davidson, W.S. (1992). Artificial hybridization of Newfoundland brown trout and Atlantic salmon: hatchability, survival and growth to first feeding. Aquaculture 106, 117-125.

Morman, R. H. (1979). Distribution and ecology of lampreys in the Lower Peninsula of Michigan, 1957-75. Great Lakes Fishery Commission Technical Report 33.

Morris, K. H. & Maitland, P. S. (1987). A trap for catching adult lampreys (Petromyzonidae) in running water. *Journal of Fish Biology*, 31, 513-516.

Morris, K. H. (1989). A multivariate morphometric and meristic description of a population of freshwater-feeding river lampreys, *Lampetra fluviatilis* (L.), from Loch Lomond, Scotland. *Zoological Journal of the Linnaean Society,* 96, 357-371.

Moser, M. L., Almeida, P. R., Kemp, P. S. & Sorensen, P. W. (2015). Lamprey spawning migration. *In:* (Docker, M.F. ed.) Lampreys; Biology, Conservation and Control. Springer Fish & Fisheries Series pp. 215-264.

Nilsson, P.A., Hulthén, K., Chapman, B.B. et al. (2017). Species integrity enhanced by a predation cost to hybrids in the wild. *Biology Letters* 13(7), 20170208.

Nosil, P. & Crespi, B. J. (2004). Does gene flow constrain trait divergence or vice-versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution*, 58, 101-112.

Nosil, P. (2012). Ecological speciation. Oxford University Press, 304pp.

Østbye, K., Harrod, C., Gregersen, F., Klepaker, T., Schulz, M., Schluter, D. & Vøllestad, L. A. (2016) The temporal window of ecological adaptation in postglacial lakes: a comparison of head morphology, trophic position and habitat use in Norwegian threespine stickleback populations. *BMC Evolutionary Biology*, 16, 1–16.

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O. & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, 8, p.e1002752.

Pasternak, C., Wong, S. & Elson, E.L. (1995). Mechanical function of dystrophin in muscle cells. *The Journal of Cell Biology* 128(3), 355-361.

Pavey, S.A., Colin, H., Nosil, P. & Rogers, S.M. (2010). The role of gene expression in ecological speciation. *Annals of the New York Academy of Science* 1206, 110-129.

Peakall, R. & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28, 2537-2539.

Pereira, A. M., Almada, V. C. & Doadrio, I. (2011). Genetic relationships of brook lamprey of the genus *Lampetra* in a Pyrenean stream in Spain. *Ichthyological Research* 58: 278-282.

Petit, R., Aquinagalde, I., de Beaulieu, J.L. et al. (2003). Glacial refugia: hotspots but not melting pots of biodiversity. *Science* 300(5625), 1563-1565.

Piavis, G. W., Howell, J. H. & Smith, A.J. (1970). Experimental hybridization among five species of lampreys from the Great Lakes. *Copeia***,** 1970, 29-37.

Pickrell, J. K. & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8, e1002967.

Præbel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., ... & Amundsen, P. A. (2013). Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and evolution*, *3*(15), 4970-4986.

Prendergast, A., Linbo, T. H., Swarts, T., Ungos, J. M., McGraw, H. F., Krispin, S…Raible, D. W. (2012). The metalloproteinase inhibitor Reck is essential for zebrafish DRG development. *Development*, 139, 1141-1152.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2010). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.

Quilodrán, C. Q., Currat, M. & Montoya-Burgos, J. I. (2014). A general model of distant hybridization reveals the conditions for extinction in Atlantic salmon and brown trout. PLoS ONE 9(7), e101736.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Randler, C. (2002). Hybridization, mixed pairing and female choice. Animal Behaviour 63, 103–119.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461, 489–494.

Renaud, C. B. (2011). Lampreys of the world: an annotated and illustrated catalogue of lamprey species known to date. FAO Species Catalogue for Fisheries Purposes No. 5. Rome: FAO.

Rogers, S.M., Mee, J.A. & Bowles, E. (2013). The consequences of genomic architecture on ecological speciation in postglacial fishes. *Current Zoology* 59(1), 53-71.

Rogers, S.M., Tamkee, P., Summers, B., Balabahadra, S. et al. (2012). Genetic signature of adaptive peak shift in threespine stickleback. *Evolution* 66(8), 2439-2450.

Rougemont, Q., Gagnaire, P. –A., Perrier, C., Genthon, C., Besnard, A-L., Launey, S. & Evanno, G. (2017). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, 26, 142–162.

Rougemont, Q., Gaigher, A., Lasne, E., Côte, J., Coke, M., Besnard, A. –L…Evanno, G. (2015). Low reproductive isolation and highly variable levels of gene flow reveal limited progress toward speciation between European river and brook lampreys. *Journal of Evolutionary Biology*, 28(12), 2248-2263.

Rougemont, Q., Roux, C., Neuenschwander, S., Goudet, J., Launey, S. & Evanno, G. (2016). Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. *Peer Journal*, 4, e1910.

Rougeux, C., Bernatchez, L., & Gagnaire, P. A. (2017). Modeling the Multiple Facets of Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs (Coregonus clupeaformis). *Genome biology and evolution*, *9*(8), 2057-2074.

Rowe, K.C., Heske, E.J., Brown, P.W. & Paige, K.N. (2004). Surviving the ice: northern refugia and postglacial colonization. *PNAS* 101(28), 10355-10359.

Rundle, H.D. & Nosil, P. (2005). Ecological speciation. *Ecology Letters* 8(3), 336-352.

Salewski V. (2003). Satellite species in lampreys: a worldwide trend for ecological speciation in sympatry. *Journal of Fish Biology*, 63, 267-279.

Schluter, D. (1996). Along radiation along genetic lines of least resistance. *Evolution* 50(5), 1766-1774.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science* 323(5915), 737-741.

Schreiber, A. & Engelhorn, R. (1998). 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*, 36, 85-99.

Siwertsson, A., Knudsen, R., Kahilainen, K., Præbel, K., Primicerio, R., & Amundsen, P. A. (2010). Sympatric diversification as influenced by ecological opportunity and historical contingency in a young species lineage of whitefish. *Evolutionary Ecology Research*, 12, 929–947.

Smadja, B.M. & Butlin, R.K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology* 20(24), 5123-5140.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler., Jiang, N., Campbell, M. S…Li, W. (2013). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genetics*, 45, 415-421.

Spice, E.K., Goodman, D.H., Reid, S.B. & Docker, M.F. (2012). Neither philopatric nor panmictic: microsatellite and mtDNA evidence suggests lack of natal homing but limits to dispersal in Pacific lamprey. *Molecular Ecology* 21, 2916-2930.

Taylor, E. B. (1999). Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries*, 9, 299-324.

Taylor, E. B., Harris, L. N., Spice, E. K. & Docker, M. F. (2012). Microsatellite DNA analysis of parapatric lamprey (*Entosphenus* spp.) populations: implications for evolution, taxonomy, and conservation of a Canadian endemic. *Canadian Journal of Zoology*, 90, 291-303.

Taylor, E.B. & McPhail, J.D. (2000). Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society B: Biological Sciences* 267(1460), DOI: 10.1098/rspb.2000.1294.

Tsimbalov, I. A., Kucheryavyi, A. V., Veselov, A. E. & Pavlov, D. S. (2015). Description of the European river lamprey *Lampetra fluviatilis* (L., 1758) from the Lososinka River (Onega Lake Basin). *Doklady Biological Sciences*, 462, 124-127.

Tuunainen, P., Ikonen, E. & Auvinen, H. (1980). Lampreys and lamprey fisheries in Finland. *Canadian Journal of Fisheries and Aquatic Science*, 37, 1953-1959.

Vatandoust, S., Mousavi-Sabet, H., Razeghi-Mansour, M., AnvariFar, H. & Heidari, A. (2015). Morphometric variation of the endangered Caspian lamprey, *Caspiomyzon wagneri* (Pisces: Petromyzontidae), from migrating stocks of two rivers along the southern Caspian Sea. *Zoological Studies*, 54, 56.

Vladykov, V. D. & Kott, E. (1979). Satellite species among Holarctic lampreys (Petromyzontidae). *Canadian Journal of Zoology*, 57, 860-867.

Wellborn, G. A., & Langerhans, R. B. (2015). Ecological opportunity and the adaptive diversification of lineages. *Ecology and Evolution*, *5*, 176-195.

Willis, P.M., Ryan, M.J. & Rosenthal, G.G. (2011). Encounter rates with conspecific males influence female mate choice in a naturally hybridizing fish. *Behavioral Ecology*, 22(6), 1234-1240.

Wirtz, P. (1999). Mother species-father species: unidirectional hybridization in animals with female choice. Animal Behaviour 58(1), 1-12.

Wittschieben, J., Shivji, M. K. K., Lalani, E., Jacobs, M. A., Marini, F…Wood, R. D. (2000). Disruption of the developmentally regulated Rev3l gene causes embryonic lethality. *Current Biology*, 10, 1217-1220.

Yamamoto, M., Matsuzaki, T., Takahashi, R., Adachi, E., Maeda, Y., Yamaguchi, S…Noda, M. (2012). The transformation suppressor gene Reck is required for postaxial patterning in mouse forelimbs. *Biology Open*, 1, 458-466.

Yamazaki, Y. & Goto, A. (1998). Morphometric and meristic characteristics of two groups of *Lethenteron reissneri*. *Ichthyological Research*, 44, 15-25.

Yamazaki, Y. & Goto, A. (2000). Breeding season and nesting assemblages in two forms of *Lethenteron reissneri*, with reference to reproductive isolating mechanisms. *Ichthyological Research*, 47(3), 271-276.

Yamazaki, Y. & Goto, A. (2016). Molecular phylogeny and speciation of East Asian lampreys (genus *Lethenteron*) with reference to their life-history diversification. *In*: (Orlov, A. & Bemaish, R. eds.) Jawless Fishes of the World (Vol. 1). Cambridge Scholars Publishing pp. 17-57.

Yamazaki, Y., Yokoyama, R., Nagai, T. & Goto, A. (2011). Formation of fluvial non-parasitic population of *Lethenteron camtschaticum* as the first step in petromyzontid speciation. *Journal of Fish Biology*, 79, 2043-2059.

Yamazaki, Y., Yokoyama, R., Nagai, T. & Goto, A. (2014). Population structure and gene flow among anadromous Arctic lamprey (*Lethenteron camtschaticum*) populations deduced from polymorphic microsatellite loci. *Environmental Biology of Fishes*, 97, 43-52.

Zanandrea, G. S. J. (1959). Speciation among lampreys. *Nature*, 184, 380.

Zanandrea, G. S. J. (1961). Studies on European lampreys. *Evolution* 15(4), 523-534.

Zhou, T., Liu, S., Geng, X., Jin, Y., Jiang, C., Bao, L…Liu, Z. (2017). GWAS analysis of QTL for enteric septicemia of catfish and their involved genes suggest evolutionary conservation of a molecular mechanism of disease resistance. *Molecular Genetics and Genomics*, 292, 231-242.

# TABLES & FIGURES

Figure 1. Photograph of three lamprey ecotypes collected in the Loch Lomond basin and used in genetic and behavioural comparisons. (A) is nonparasitic (non-P), (B) is freshwater parasitic (lake-P), and (C) is anadromous parasitic (anad-P).

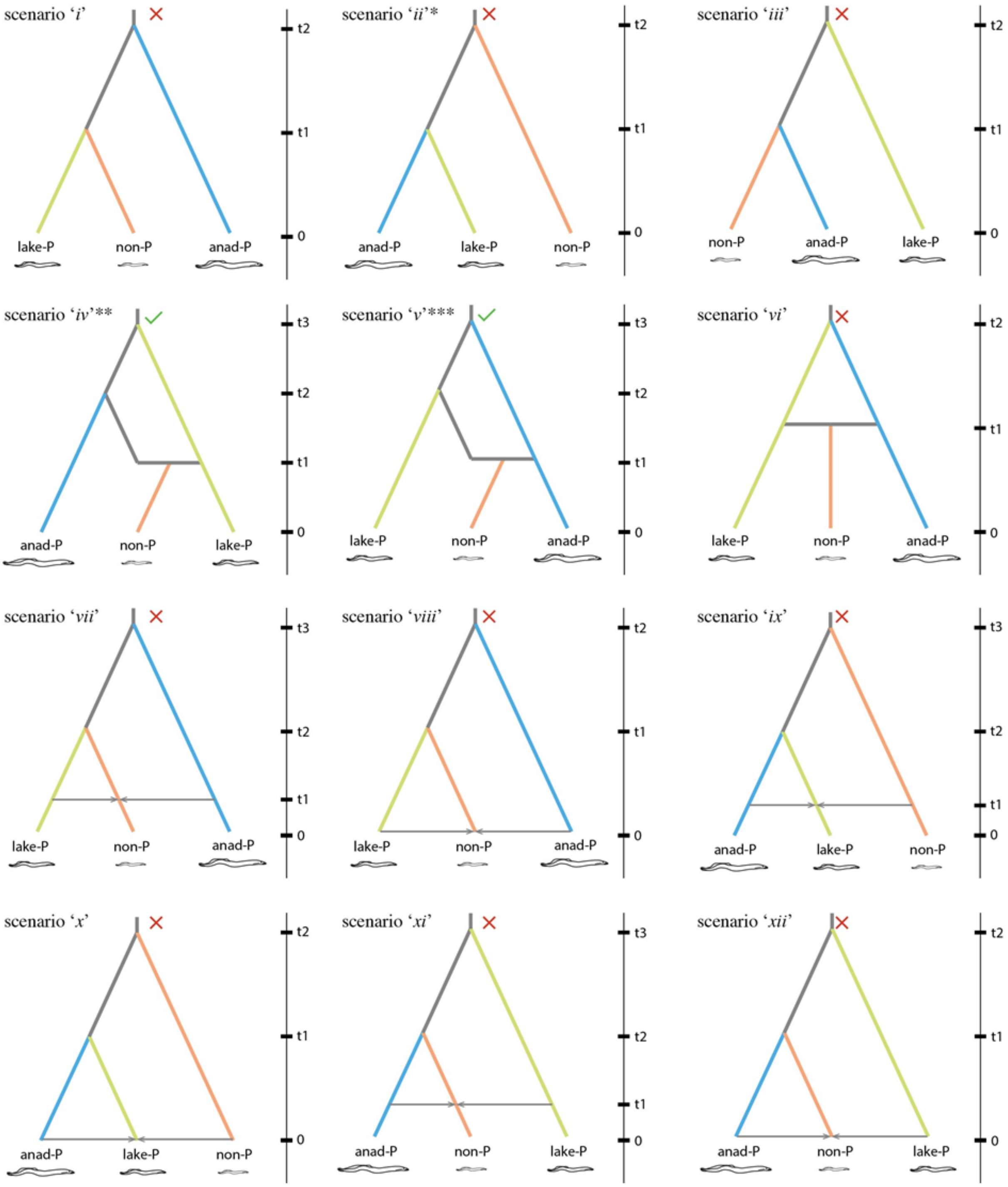


Figure 2. Hypothetical evolutionary scenarios for lamprey ecotypes tested using demographic reconstruction analyses. Scenarios *i* - *iii* are strictly bifurcating trees with no gene flow between ecotypes. Scenarios *iv* and *v* require an initial splitting event followed by a second split of the non-P ecotype and subsequent introgression into non-P from either parasitic form. Scenario *vi* suggests hybrid speciation between lake-P and anad-P, giving rise to the non-P ecotype. Scenarios *vii* - *xii* differ in the evolutionary relationships between ecotypes and include gene flow at different time points from both parasitic ecotypes into non-P. Scenarios with uneven number (*vii*, *ix*, *xi*) are evolutionary models with recent ongoing gene flow and those with even numbers (*viii*, *x*, *xii*) past gene flow from the parasitic into the non-parasitic ecotype. Scenarios with check mark were supported by analyses, and scenarios with an x-mark were not supported

in any analysis. * consistent with current taxonomy; ** supported by DIYABC analysis; *** supported by FASTSICMOAL2 and TREEMIX.

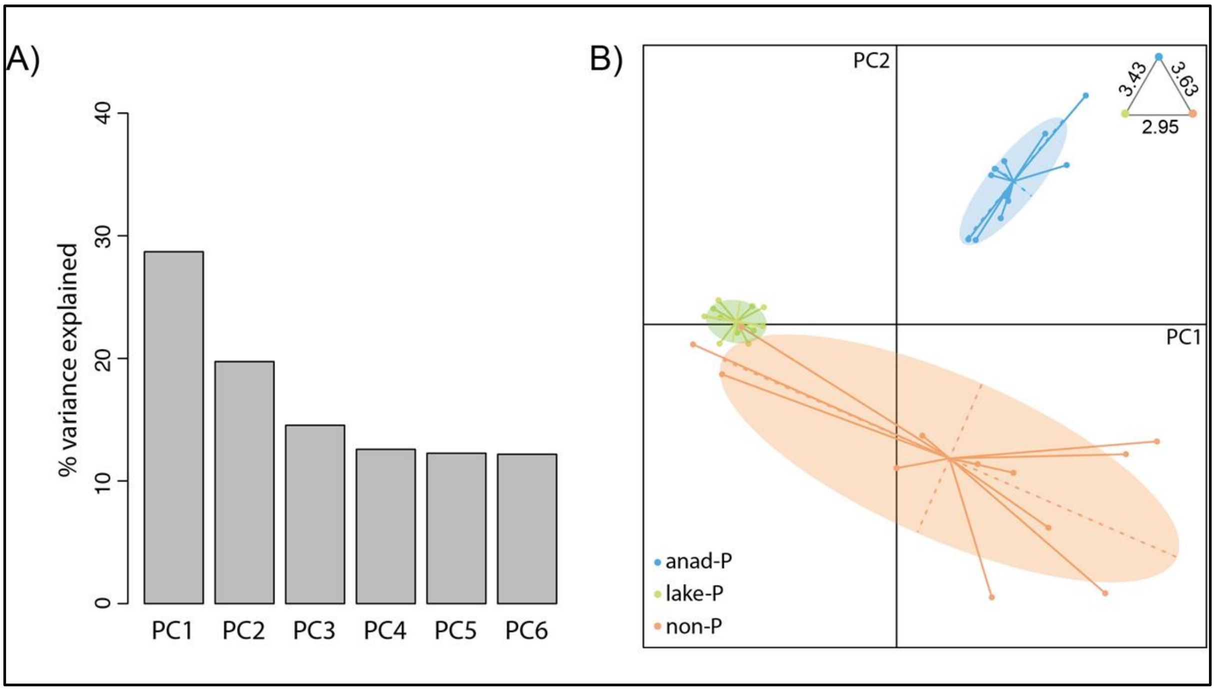


Figure 3. Multivariate analyses illustrating genetic differentiation between lamprey ecotypes. Analyses were performed with the same 35 individuals. Panel A shows relative percentages of variance explained by the first six principal components. The first two principal components accounting for 48.44% of the total variance are shown in Panel B (anad-P = blue, lake-P = green, non-P = orange), with the scores for each individual of the different ecotypes. Ellipses drawn are confidence intervals including ⅔ of the samples within each ecotype. Average Euclidean distance between ecotypes are shown in the upper right corner of PCA plots.

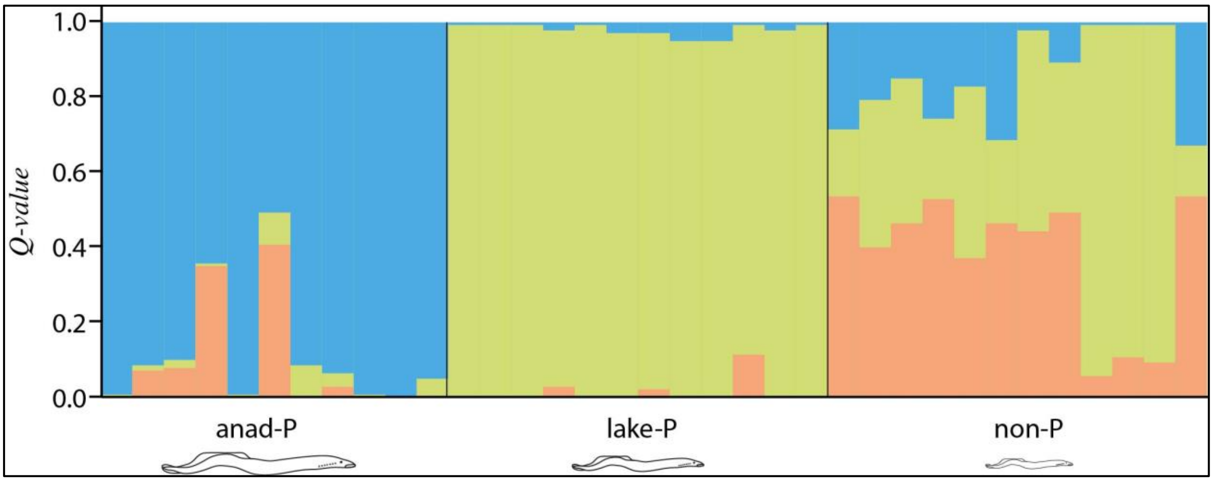
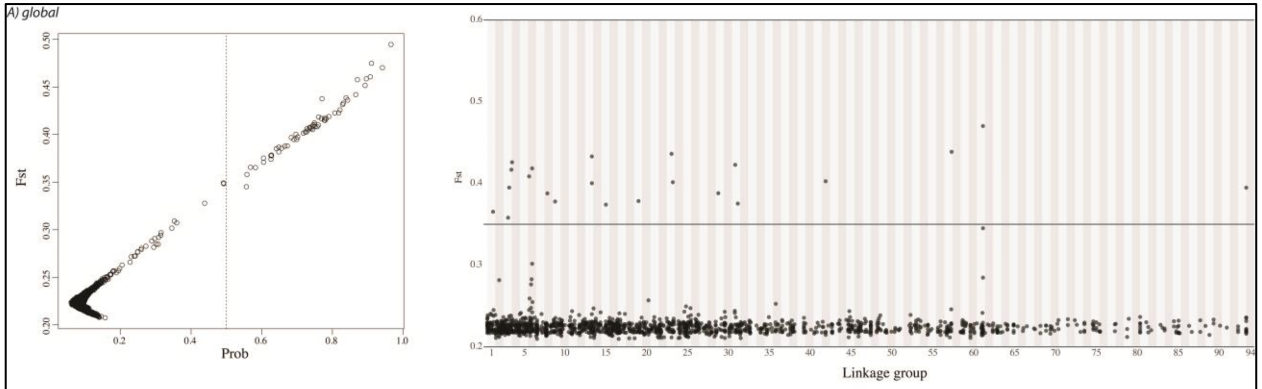
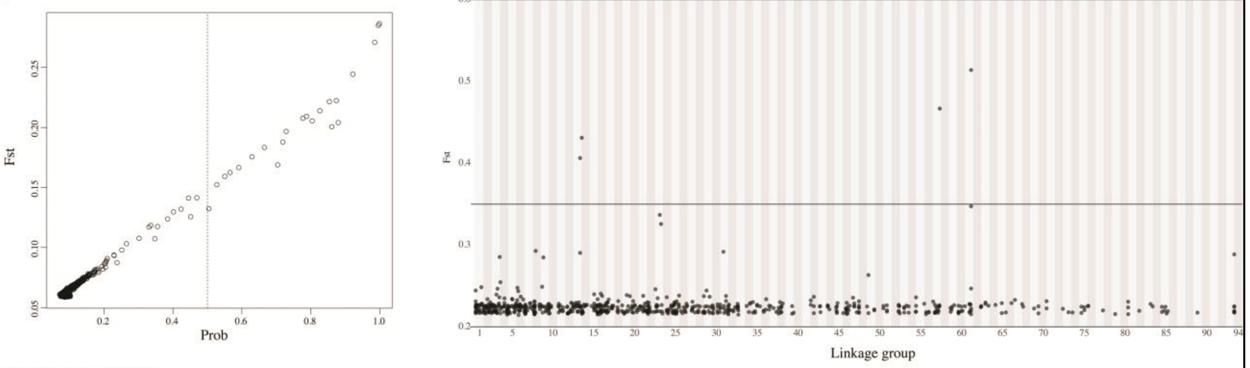


Figure 4. Output from a STRUCTURE analysis using $K = 3$. See Figure S3 for comparisons with $K = 4$ and 5. Each bar represents an individual and the relative proportion of its genotype belonging to each of the three genetic clusters. The analysis was performed excluding differentially fixed SNPs and outlier loci and includes a total of 7, 678 loci.

*A) global*



*B) resident vs. anadromous*



*C) parasitic vs. non-parasitic*

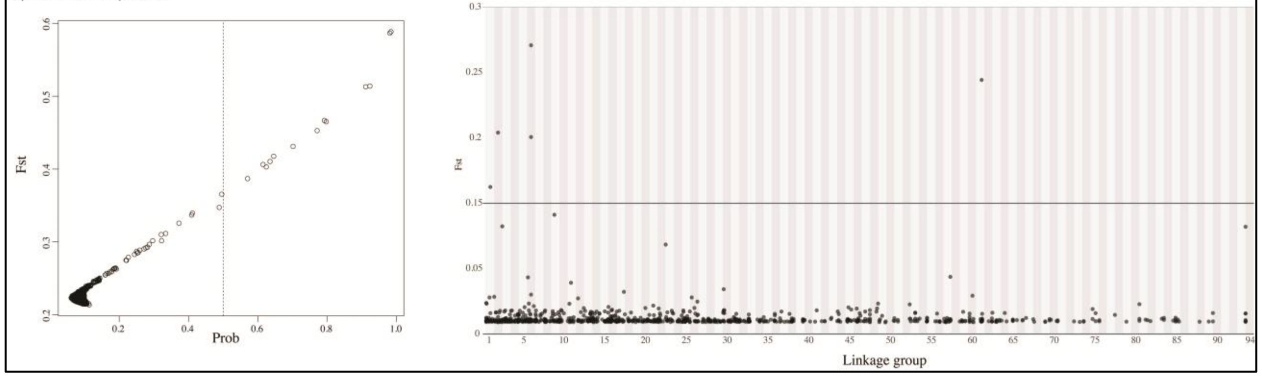


Figure 5. Plot of *F*st outliers and genome scan of all identified loci that matched to the indexed sea lamprey (*Petromyzon marinus*) genome. On the left, *F*st values obtained from BAYESCAN outlier analyses from (A) global across all ecotypes; (B) anadromous *vs*. resident individuals; and (C) parasitic *vs*. nonparasitic individuals are displayed. Loci that fall right of the vertical lines have a probability greater than 50% of being an outlier. On the right, the position of each locus along the sea lamprey linkage groups (x-axis; values normalized to a fraction of 1 for each linkage group) and the respective *F*st value (y-axis) are shown. Horizontal lines represent the probability threshold of 50% of being an outlier.

Table 1. Proportion of mating events occurring within and between ecotypes. Male mating attempts occurred in 142/324 potential pairings, whereas females only released gametes in 32. Shown are the proportion of pairings where males attempted to mate at least once (male mating attempts), the frequency of male reproductive behaviours (male spawning frequency), and the pairings where females were stimulated to release gametes (female gamete release). Anad-P = anadromous, parasitic; Lake-P = lake-resident parasitic; Non-P = lake-resident nonparasitic. The "Pro. Males" columns show the proportion of males of each ecotype that attempted to mate, spawned, or stimulated female gamete release.

| | **Male Mating Attempts** | | | | **Male Spawning Frequency** | | | | **Female Gamete Release** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Female Ecotype** | Anad-P | Lake-P | Non-P | Pro. Males | Anad-P | Lake-P | Non-P | Pro. Males | Anad-P | Lake-P | Non-P | Pro. Males |
| **Anad-P** | 0.51 | 0.53 | 0.30 | 0.45 | 0.41 | 0.42 | 0.09 | 0.31 | 0.27 | 0.33 | 0.00 | 0.25 |
| **Lake-P** | 0.49 | 0.45 | 0.55 | 0.49 | 0.59 | 0.56 | 0.85 | 0.67 | 0.73 | 0.67 | 0.80 | 0.72 |
| **Non-P** | 0.00 | 0.02 | 0.16 | 0.06 | 0.00 | 0.02 | 0.06 | 0.03 | 0.00 | 0.00 | 0.20 | 0.03 |