A mosaic of genomic architectures underpins parasitism loss in a jawless vertebrate

Arne Jacobs[1,†], Nolwenn Decanter[2], Ole K. Tørresen[3], Benedicte Garmann-Aarhus[3,4], Maria Capstick[1], Quentin Rougemont[5], Frédéric Guillaume[6], Romane Normand[7], Julien Tremblay[8,9], Jean-Pierre Destouches[8,9], Anne-Laure Besnard[2], Ahmed Souissi[10], Gilles Lassalle[2], Solenn Stoeckel[2], Eric J. Petit[2], Siv NK Hoff[3], Daniel J. Park[11,12], Bernard J. Pope[12], Sissel Jentoft[3], Leif Asbjørn Vøllestad[3], Kjetill S. Jakobsen[3] & Guillaume Evanno[2,9]

1 School of Biodiversity, One Health & Veterinary Medicine, University of Glasgow, Glasgow, UK

2 DECOD, INRAE, Institut Agro, IFREMER, Rennes, France

3 Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Norway

4 Natural History Museum, University of Oslo, Norway

5 Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Société Evolution, Gif sur Yvette, France

6 Department of Organismal and Evolutionary Biology, University of Helsinki, Helsinki, Finland

7 Department of Biology, School of Science, RMIT University, Bundoora, VIC 3083, Australia

8 Unité Expérimentale d'Ecologie et d'Ecotoxicologie Aquatique - U3E, Rennes, France

9 Management of Diadromous Fish in Their Environment, OFB, INRAE, Institut Agro, UPPA, Rennes, France

10 Laboratory of Biodiversity, Parasitology and Ecology of Aquatic Ecosystems, Biology department, University of Tunis El Manar, Tunis, Tunisia

11 Department of Biochemistry and Pharmacology, Melbourne Bioinformatics, The University of Melbourne, Parkville, Australia

12 Melbourne Bioinformatics, The University of Melbourne, Parkville, Australia

†Corresponding author: arne.jacobs@glasgow.ac.uk

Abstract

Lampreys are the only ancestrally parasitic vertebrate lineage, yet parasitism has been repeatedly lost alongside a suite of life-history changes, such as loss of migration and juvenile feeding and accelerated maturation. Combining whole-genome resequencing, haplotype-resolved assemblies, hybrid-zone genotyping, multi-tissue transcriptomics, and sperm phenotyping, we map this life-history syndrome in European *Lampetra* to six chromosomes spanning a mosaic of genomic architectures: a ~20 Mb low-recombination region on chromosome 1 lacking chromosomal rearrangements within *Lampetra* but involving inter-specific rearrangements across deep lamprey lineages; a translocated inversion with ecotype-dependent sperm-velocity effects; and ecotype-divergent deletions overlapping genes crucial for nervous system (*CNTNAP2*) and reproductive development (*FSHR*). However, this genomic basis is not shared with a convergent sister lineage, pointing to independent routes to a recurring life-history transition in lampreys.

Introduction

Life-history variation within species has often been linked to chromosomal rearrangements that differ in orientation between divergent populations and suppress recombination in heterozygotes. Examples include large chromosomal inversions impacting local adaptation in seaweed flies (*1*) and Atlantic silverides (*2*), life-history in monkeyflowers (*3*), or migratory strategies in cod (*4*, *5*) and rainbow trout (*6*), or reproductive strategies in ruff (*7*). While most studies have focused on simple structural variants, such as inversions, more complex intraspecific rearrangements can also facilitate adaptation under gene flow, such as inverted-translocations in Timema stick insects (*8*), or chromosomal fusions and fissions in butterflies (*9*), threespined sticklebacks (*10*), and codfishes (*11*). Yet, recombination might be reduced without polymorphic rearrangements between life-history forms, e.g. through historical interspecific chromosomal fusions (*10*, *12*) or recombination deserts on sex-chromosomes (*13*). The underlying genomic mechanisms are not well understood, though, and empirical studies linking low recombination regions to life-history evolution in the absence of polymorphic structural variation have been largely missing - despite the potential importance for adaptation.

Parasitic feeding is ubiquitous across the tree of life, yet comparatively rare in vertebrates, with famous examples including vampire bats (*14*), vampire finches (*15*) and lampreys (*16*). In contrast to other vertebrates, the ancestor of all extant lampreys was likely parasitic (*16*, *17*). However, lampreys have repeatedly lost their parasitic lifestyle in many independent lineages, involving drastic changes in development and life-history between closely related, genetically distinct species (*16–19*). Although they are phenotypically indistinguishable as larvae, they are very distinct as adults. In contrast to parasitic lampreys, non-parasitic species show a suite of paedomorphic changes, including loss of the juvenile feeding phase (non-trophic), intestinal degeneration during metamorphosis, residency in natal streams, and accelerated sexual maturation (Fig. 1A) (*16*, *20*). This makes non-parasitic lampreys to our knowledge the only vertebrates not feeding as juveniles and/or adults, and creates a coordinated life-history syndrome (referred to as parasitic vs non-parasitic). Lampreys have fascinated biologists for over a century (*21*) and they become increasingly important for comparative genomic studies to elucidate the evolution of key vertebrate features (*22*, *23*). Yet, the genomic basis of this fascinating life-history syndrome within this early-diverging jawless vertebrate lineage remains unresolved.

In this study, we take advantage of the recent loss of parasitism in the European lamprey species complex (*Lampetra* spp.) to resolve the underlying genomic basis by integrating comparative genomics of multiple haplotype-resolved genome assemblies, population genomics of parasitic and non-parasitic individuals from across Europe, targeted genotyping in a narrow hybrid zone, and transcriptomic analyses of multiple tissues. We show that life-history variation is associated with a multitude of major effect loci concentrated in a few genomic regions containing crucial genes for development, behaviour and metabolism. Interestingly, we find a major parasitism-associated region with low recombination that is maintained in the absence of genomic rearrangements between ecotypes despite ongoing gene flow, suggesting a crucial role of ancient interspecific rearrangements in intraspecific evolution.

Results

Rapid evolution of European lamprey life-history diversity under ongoing gene flow.

European lamprey life-history forms are taxonomically divided into the ancestral anadromous-parasitic river lamprey (*Lampetra fluviatilis*), lake-migratory parasitic river lamprey (*L. fluviatilis*), and the resident non-parasitic brook lamprey (*L. planeri*), (referred to as 'ecotypes' hereafter) (*16*, *19*, *24*). These ecotypes have likely evolved recently (<1 MYA) (*17*) and frequently hybridise under secondary contact (*24*, *25*) due to incomplete reproductive isolation (*24*, *26*, *27*). We investigated the evolutionary relationship of ecotypes from eight rivers across Europe (different combinations of ecotypes) using whole genome sequencing data for 189 individuals (mean coverage depth of 8.4x; range: 2.7x to 24.3x) (Fig. 1A,B table S1). Genome-wide phylogenetic relationships based on neutral SNPs (see methods for SNP filtering) suggested a recent and rapid divergence of ecotypes within and across rivers, as individuals largely cluster by ecotype and river but are only weakly divergent, as highlighted by a star-shaped phylogeny with high uncertainty in phylogenetic clustering (Fig.1C, fig. S1). This pattern is similar for all SNPs across the genome, with slightly stronger clustering by river and ecotype (fig. S1). The lack of ecotype clustering across rivers and uncertain phylogenetic relationship is consistent with the rapid colonisation of rivers and repeated divergence from a shared common ancestor, which is supported by a lack of clustering by ecotype or river for mitochondrial genomes (fig. S2). An exception is the lake-parasitic Endrick ecotype, which was highly divergent from other parasitic populations, suggesting an independent origin of this lake-specialised ecotype (Fig. 1C,D, fig. S3-S6). Yet, the non-parasitic Endrick ecotype showed strong signals of introgression from the sympatric lake-parasitic ecotype (Fig. 1D-F, fig. S5-S6) (*24*). Furthermore, the longer branch lengths (fig. S1 and S6) and stronger divergence (Fig.1D; fig. S4) of non-parasitic populations without sympatric parasitic individuals (e.g. daNP, rbNP, gloNP, names in Fig. 1B), compared to those in sympatry with parasitic lampreys, suggests strong isolation and genetic drift without recurrent genetic influx from the anadromous forms (*28*). Closer comparisons of sympatric ecotypes showed strong historical and ongoing admixture in sympatry (Fig. 1D-F, fig. S3-S6), with non-parasitic ecotypes displaying >25% anadromous-parasitic and lake-parasitic ancestry in the Oir and Endrick, respectively (Fig. 1E; fig. S5). Parasitic forms showed weaker proportions of non-parasitic ancestry (Fig.1E, fig. S5), in line with size-related mating behaviour between ecotypes (*24*, *29*).

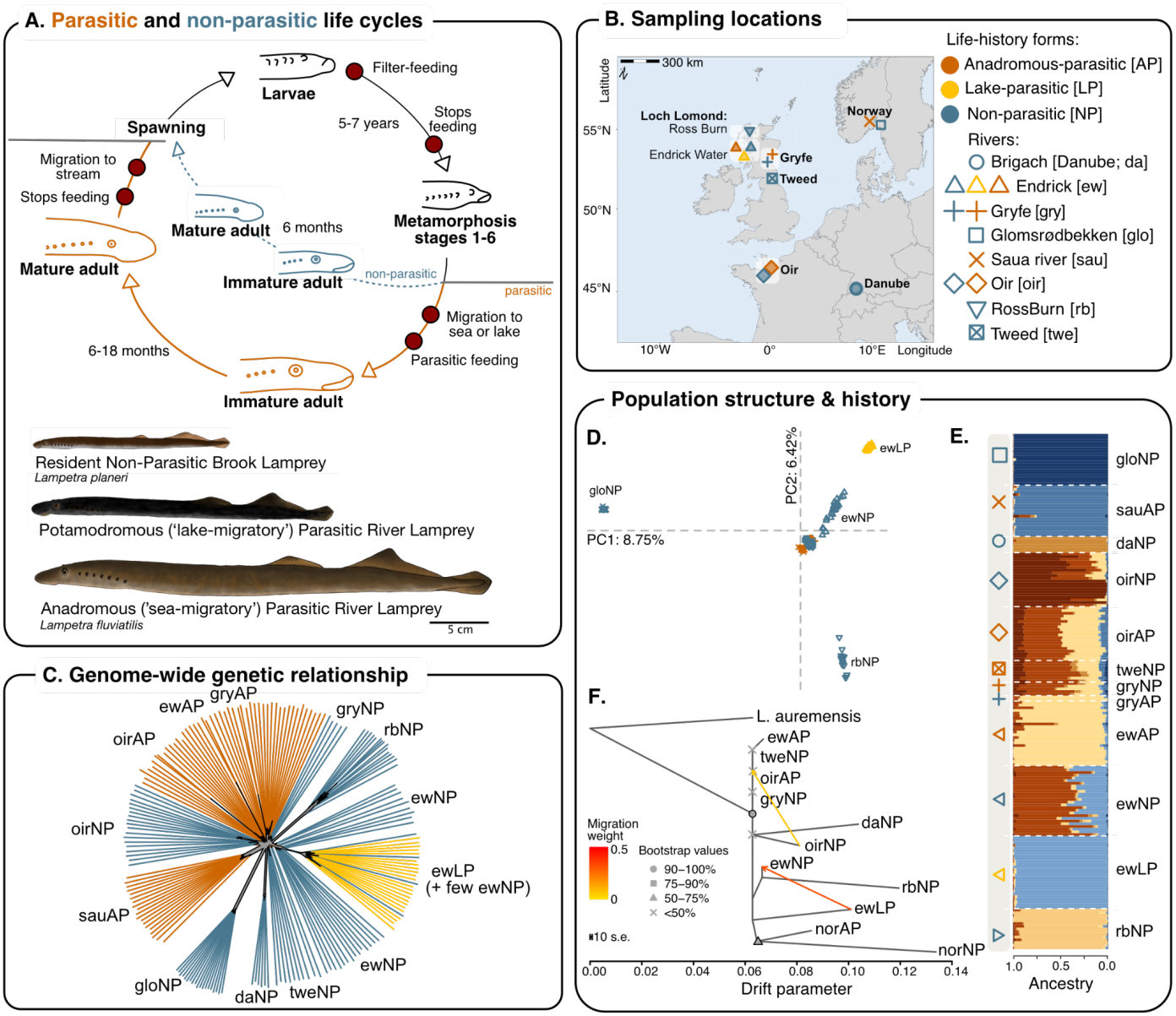

Figure 1: Life-history variation and population structure of European lampreys.

(**A**) Schematic representation of the life cycle of parasitic (orange) and non-parasitic (blue) European lampreys, highlighting how life-histories diverge between ecotypes at the end of metamorphosis. The drawings below are representative examples of the three ecotypes from the river Endrick, UK, with taxonomic classification (by A Jacobs).

(**B**) Map of Europe showing the sampling locations of individuals used for whole-genome sequencing. The exact locations are given in Table S1. Life-history forms are highlighted by colour, whereas different populations are distinguished by shape.

(**C**) NeighbourNet network based on genetic distances calculated for neutral SNPs (thinned, non-divergent SNPs outside high-LD regions) showing the phylogenetic relationship between ecotypes across populations. The grey area shows the phylogenetic outline to highlight close relationships, with the exact phylogenetic network topology shown in fig. S1. A network for thinned non-divergent SNPs is shown in fig. S1. Branches are coloured by ecotype and populations are named as shown in panel (B).

(**D**) Principal components analysis based on thinned, non-divergent SNPs for all ecotypes and populations. Additional PCAs and eigenvalue distributions for all SNPs and thinned, non-divergent SNPs are shown in fig. S3. PCAs and ancestry plots for the Loch Lomond populations and Oir population are given in fig. S5.

(**E**) Genome-wide ancestry estimated based on thinned, non-divergent SNPs with *PCAngsd* (K=8). Individuals are ordered by ecotype (coded by colour) and populations (shape) (see panel (B) for detail). Estimates for other levels of K are shown in fig. S4.

(**F**) Genome-wide maximum-likelihood tree with bootstrap support and the two strongest gene flow events as estimated with *TreeMix* based on neutral SNPs. Populations are named following the legend in (B). *TreeMix* trees with different numbers of migration events and residual correlations are provided in fig. S6.

Major-effect loci control life-history divergence

To identify the genomic basis of life-history, we performed genome-wide association analysis across all parasitic (n=91) and all non-parasitic (n=98) individuals. The presence of genetic admixture between ecotypes, replication across watersheds, and the ability to control for differences in migration strategy through parasitic ecotypes with different migration tactics (lake-parasitic vs anadromous-parasitic; Fig.1B) makes this a powerful approach. We identified seven genomic regions on six chromosomes with significant life-history associations ($p < 10^{-8}$) (Fig. 2A) (these binary-trait quantitative trait loci are hereafter referred to as 'parasitism-QTL'). The strongest associations (p=1.42e-32) were in a narrow ~100kb region on chromosome (chr) 30, yet most significant SNPs (7550 out of 7812 significant SNPs) mapped to a ~20Mb region on chr1, structured into seven individual peaks (Fig. 2A,B; fig. S7). Furthermore, we identified a broad ~5Mb parasitism-QTL on chr56 and three weaker parasitism-QTL on chr6, chr8 and chr25. The parasitism-QTL on chr1 and chr30 were also significantly associated with life-history across all three admixed sympatric ecotypes in the Endrick (fig. S8), supporting their consistent role in this life-history shift.

To further test if parasitism-QTL are consistently involved in life-history evolution across independent watersheds, we performed topology-weighting analyses across sympatric life-history pairs, which test the degree of monophyly of the three possible phylogenies across genomic regions (*30*, *31*). These tests suggested that haplotypes around parasitism-QTL were largely shared across populations (Fig 2C). Parasitism-QTL showed a higher weighting for the life-history topology (T_{LH}), which clustered individuals by life-history across rivers, compared to the geographic topology (T_{Geo}), which clustered individuals by river of origin (Kruskal-Wallis test for T_{LH} weights: W=145644457, p-value < 2.2e-16) or the control tree (T_C), despite T_{Geo} being the most common topology across genome, likely due to ongoing gene flow (Fig. 2C,D, fig. S9). A skew toward the T_{LH} topology across all parasitism-QTL is unlikely under incomplete lineage sorting (ILS) (*TWISST'N'TERN*: D_{LR}=-0.0177; P=0.015, fig. S10), as all topologies are equally likely under ILS, suggesting the maintenance of parasitism-associated regions through selection (*31*). Peaks in T_{LH} weighting around parasitism-QTL on chr1 and chr30 suggested that ecotypes share similar haplotypes in these regions (Fig. 2E), with the lack of fixation for the T_{LH} topology (T_{LH} weight < 1; Fig. 2D; fig. S9) indicating the presence of genetic variation in parasitism-QTL across rivers, e.g. through genetic drift and gene flow.

The repeated loss of parasitism via similar phenotypic changes across independent lamprey taxa (*16*), might indicate a shared genomic basis (*32*). However, comparisons to parasitic and non-parasitic ecotypes from a North American sister lineage to the European lamprey (*Occidentis ayresii*) (*33*), which diverged ~7.5MYA (*17*), indicated that genomic regions homologous to parasitism-QTL in European *Lampetra* were not differentiated between *O. ayresii* ecotypes (fig. S11). This suggests that parasitism was lost through alternative genomic routes (*17*, *33*). Yet, similar to European lamprey, *O. ayresii* ecotypes seemed to be differentiated by a mosaic of genomic architectures, including putative structural variants (fig. S11 B,C), indicating comparable genomic architectures.

Parasitism-QTL are associated with major developmental and metabolic pathways

The diversity of genes under parasitism-QTL (Fig. 2A, Supplementary file1) and associated gene ontology terms (Fig. 2F, fig. S12, Supplementary file 1) reflect the myriad phenotypic changes associated with parasitism loss in lampreys, including the loss of juvenile migration and feeding, developmental variation during metamorphosis, earlier sexual maturation, and drastic changes in metabolism in non-parasitic lampreys (*16*, *34*). For example, the strongest peak on chr30 contained a glycoprotein hormone gene (*GPHB5*) and follicle-stimulating hormone-receptor gene (*FSHR*) (*35*), which are crucial for sexual development and reproduction, known key traits distinguishing parasitic and non-parasitic lampreys (*16*, *36*). Furthermore, the peak on chr1 contained the *INHBB* gene, which is part of the TGF-β superfamily and encodes a highly-conserved protein involved in reproductive development and fertility, e.g. by regulating follicle-stimulating hormone secretion (*37*). The dual effect on the FSH pathway might indicate a potential functional interaction between divergent genes on chr1 and chr30. Many genes on chr1 play key roles in major signalling pathways involved in development (e.g. Hedgehog; Transforming growth factor (TGF)-β; WNT; Insulin-like Growth Factor (IGF)) (*38*–*41*). This included thyroid-hormone responsive genes (e.g. *IHH, GLI2, GALNT12, WNT1*), with thyroid hormone being the master regulator of metamorphosis (*42*); the developmental timepoint at which phenotypic differences between ecotypes arise (*16*, *20*, *34*). Parasitism-QTL further overlapped numerous metabolic genes involved in fatty-acid metabolism and glycolysis on chr1 (e.g. *HACL1, INSIG2, OSBPL6, PHYH, PDK2, PFKP*), chr8 (*CPT1A),* chr25 (*PPP5C*), and chr56 (*CPT2*, *IRS1*) (Supplementary file 1). This suggests that ecotypes likely differ in metabolism (*43*, *44*), which aligns with differences in trophic ecology; non-parasitic lamprey do not feed as juveniles, while the parasitic lamprey actively migrate, parasitically feed and rapidly grow (*16*, *36*). Additionally, we identified genes with known roles in feeding behaviour/appetite on chr1 and chr56 (*INSIG2, GHSR*, *GPR83).* Lastly, many parasitism-QTL genes were associated with brain/neural development and function (e.g. *AMPH, EN2, CNTNAP2, NLGN3, DLG2*), pointing toward differences in brain development between ecotypes, with the brain drastically restructuring during lamprey metamorphosis, reflecting ontogenetic shifts in behaviour and ecology (*45*).

The functional role of genes under parasitism-QTL was further investigated through differential expression analyses in several tissues (i.e. brain, liver, gonad and gill) between spawning-ripe anadromous-parasitic and non-parasitic individuals from the Oir (fig. S13). Of the 177 genes under parasitism-QTL, 47 (50 transcripts) were differentially expressed in at least one tissue (ranging from 2 in the gonad to 34 in the liver; fig. S13), including key metabolic genes in the liver, such as *CPT1A* and *CPT2*, which play a role in fatty-acid metabolism, and *IRS1*, an insulin receptor. It is important to note that we only assayed gene expression in adults, and other genes under parasitism-QTL might be either differentially spliced (*46*, *47*) or differentially regulated in other life stages (*48*).

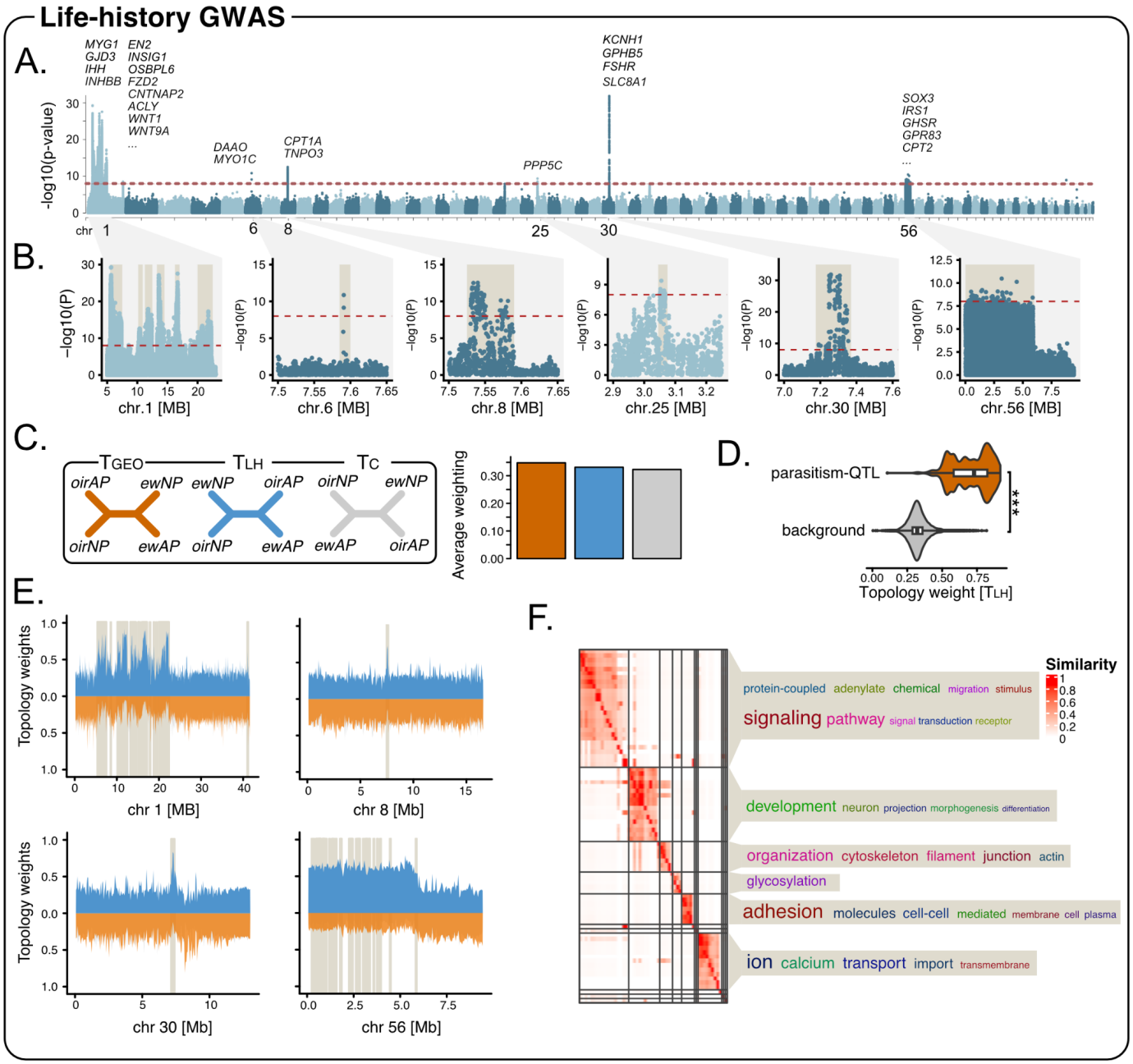

Figure 2: Genomic signals of parasitism-loss in European lampreys.

(**A**) Manhattan plot showing genetic associations for life history (parasitic vs non-parasitic) estimated using a linear mixed model in *GEMMA* across all populations. The red dashed line shows the significance threshold of $p = 10^{-8}$. We show some key candidate genes with parasitism-QTL above individual peaks (not all shown for chr1 and chr56). The full list of candidate genes, their full names, and locations are given in Supplementary file 1, with functional annotations in Supplementary file 1.

(**B**) Candidate regions on chromosomes with parasitism associated genomic regions ("Parasitism-QTL"), zoomed in from the plot above (A).

(**C**) Four-taxon topologies for the anadromous-parasitic (AP) and non-parasitic (NP) ecotypes from the Endrick (ew) and Oir that were tested using *TWISST*. T_{GEO} shows the geographic topology, clustering individuals by location of origin; T_{LH} shows the life-history topology, clustering individuals by ecotype; T_C shows the control topology without any clustering by geography and ecotype. The barplot on the right shows their average weighting by topology (proportion across all trees) across the genome.

(**D**) Comparison of the life-history topology (T_{LH}) weights for 50kb genomic windows overlapping parasitism-QTL and the genomic background (genomic regions without parasitism-QTL) across the genomes. The genomic background distribution was obtained by the random resampling with replacement of x loci for 1,000 times (x = number of windows overlapping parasitism-QTL). Asterisks highlight the significance level for differences in topology weight distributions estimated using a Kruskal-Wallis test (W=145644457, p-value < 2.2e-16).

(**E**) Distribution of topology weights across the four chromosomes (chr 1, 8, 30, 56) with top parasitism-QTL (highlighted by brown rectangles). To aid visualisation, the T_{LH} (blue) distribution is upwards, whereas the distribution for T_{GEO} (orange) is plotted downwards.

(**F**) Semantic similarity clustering of overrepresented biological process gene ontology (GO) terms for genes overlapping parasitism-QTL, with similarities shown using a white to red scale. Descriptions of GO terms are shown on the right, with the size of words describing overrepresented terms shown relative to their frequency. Full GO term overrepresentation results are shown in Supplementary file 1 and fig. S12.

Interspecific genomic rearrangements are associated with recombination cold spots

To elucidate how parasitism-QTL are maintained despite extensive gene flow, we tested if they colocalised with recombination-suppressing chromosomal rearrangements between ecotypes (on chromosomes 1, 25, 30, 56). The presence of three distinct genomic clusters in PCAs for parasitism-QTL on chr56 indicated a large rearrangement, and a smaller putative rearrangement around the parasitism-QTL at the end of chr1 (Fig. 3A, fig. S14-15). Although some of the other regions display three broad clusters in region-specific PCAs, these still show evidence of recombination (e.g. intermediate individuals between clusters) and are likely associated with low-recombination regions but not recombination-suppressing SVs (*1*). Comparisons of five genome assemblies for both ecotypes, including new haplotype-resolved, chromosome-scale genomes for the parasitic and non-parasitic ecotypes from Scandinavia (*49*) and France (Oir, provided in this study) (table s4), and genomes for two evolutionarily distant species [sea lamprey (*Petromyzon marinus*) and Korean lamprey (*Lethenteron reissneri*)], that diverged between 5 and 25MYA, (*17*, *50*, *51*), confirmed the presence of chromosomal rearrangements between ecotypes on chr56, but did not provide evidence for large rearrangements on chr1 or the other chromosomes (Fig. 3B, fig. S16-17). While most of the genome showed high collinearity across *Lampetra* genomes (fig.S19), we detected a ~3Mb inverted-translocation (‘*trans-inv56*’) between ecotypes on chr56 (Fig. 3B, fig. S17). The same region was also translocated across more distantly related species, indicating high structural turnover (Fig. 3B). *Trans-inv56* showed high linkage disequilibrium (fig. S18) and low recombination rates in the non-parasitic Oir ecotype (Fig. 3C, fig. S19-S20, fig. S21-22 for linkage maps). However, population-level recombination rates were only slightly decreased in the non-parasitic Endrick ecotype (Fig. 3C, fig. S19), potentially because this ecotype harbours all three *trans-inv56* karyotypes.

Despite the lack of evidence for large genomic rearrangements between ecotypes around the other parasitism-QTL (fig. S16-S17) recombination rates were generally lower around parasitism-QTL compared to the genomic background (fig. S20), and these regions also showed increased Fst and LD (Fig. 3C). Compared to most of the genome, the low-recombination region on chr1 showed substantial interspecific genomic rearrangements across lamprey lineages, including fusions, translocations and inversions, but no rearrangements between parasitic and non-parasitic European lamprey ecotypes (Fig.3B, fig. S17). We hypothesise that these ancient interspecific rearrangements have potentially led to long-term recombination suppression on chr1, e.g. by altering the chromatin landscape (*12*, *52*), contributing to the maintenance of co-adapted alleles under gene flow (*10*, *13*, *53*).

While we did not detect chromosomal rearrangements between ecotypes around most parasitism-QTL, except chr56, we found evidence for divergent copy-number variants (CNVs) between sympatric ecotypes based on normalised read depth (Fig. 3D; fig. S23), which were most likely deletions in non-parasitic lamprey compared to the parasitic reference. Three divergent CNVs (CNV-based QST > mean SNP-based Fst)

between ecotypes were located within parasitism-QTL on chr1 and chr30 (Fig. 3D). Two of these CNV partially overlapped annotated genes, *CNTNAP2* and *AMPH* on chr1, which have known roles in brain development and neural function, and *FSHR* on chr30, alluding to a putative functional role of CNVs in ecotype divergence.

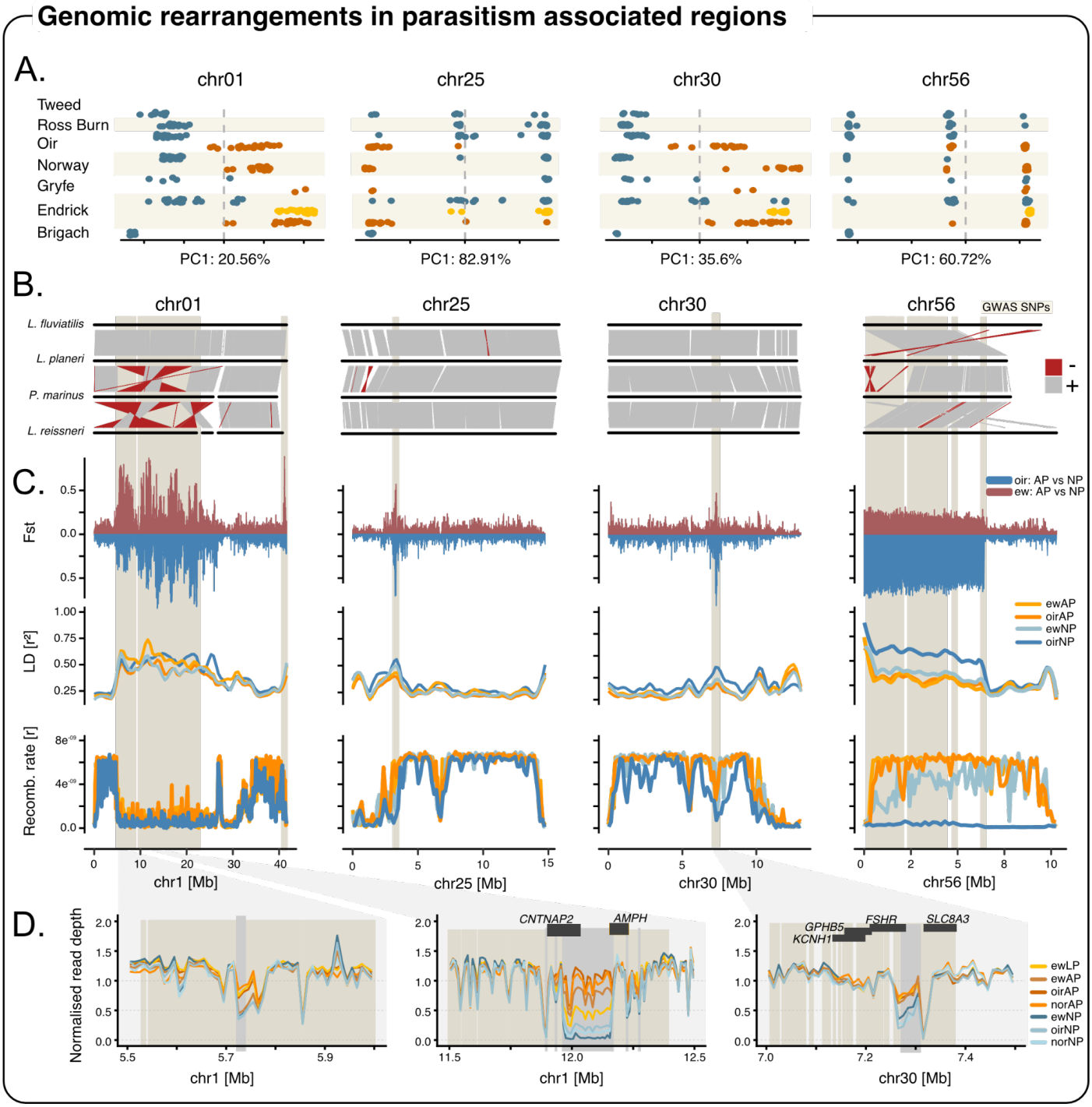

Figure 3: Genomic rearrangements around parasitism-QTL.

(**A**) Genomic principal component scores (PC1) showing the genetic variation for all SNPs within parasitism-QTL by population for chrs. 1, 25, 30, 56. Eigenvalues (in %) of PC1 for each genomic region are shown below each plot. Each dot represents an individual coloured by ecotype and ordered by population (n=189, n= 2 - 27 by ecotype and population, see table S1). Anadromous-parasitic = orange, lake-parasitic = yellow, non-parasitic = blue. The plot for chr1 is across all peaks between 5Mb and 22Mb. PCA plots (PC1 vs PC2) for each parasitism-QTL are shown in fig. S14-S15.

(**B**) Genomic synteny between the anadromous-parasitic Endrick ecotype (kcLamFluv1.1 assembly), non-parasitic Oir ecotype (*L. planeri* - haplotype 2), sea lamprey (*P. marinus* - kcPetMar1) and Korean lamprey (*L. reissneri* - ASM1570882v1), estimated with *ntSynt*. Shown are chromosomes (chrs. 1, 25, 30, 56) with major parasitism-QTL

(highlighted in beige), ordered by the extent of synteny. Grey bars connect syntenic, collinear regions, whereas red-bars show inverted regions with reversed strands. Synteny analyses across the entire genome and additional assemblies are shown in fig. S16 and S17.

(**C**) Top: Genetic differentiation (Fst) between anadromous-parasitic and non-parasitic ecotypes in the Endrick (ewAP, ewNP) and Oir (oirAP, oirNP) across the four chromosomes containing the main parasitism-QTL. Fst between Endrick ecotypes is plotted upwards and for Oir ecotypes downwards to aid visualisation (both on the same scale from 0 to 1). Middle: Linkage disequilibrium (LD) values (r^2) for replicated ecotypes from the Oir and Endrick summarized in 10kb windows across each chromosome and plotted using locally estimated scatterplot smoothing with a 0.1 span. Lines are coloured as shown in the legend. Bottom: Recombination rates for anadromous-parasitic and non-parasitic ecotypes from Endrick and Oir across candidate chromosomes, with lines coloured the same way as LD values.

(**D**) Evidence for copy number variation within parasitism-QTL. Left plot: Normalised read depth indicates reduced read depth in non-parasitic compared to parasitic ecotypes within a narrow region on chr1 as highlighted by a grey bar. This is supported by increased divergence in read depth between ecotypes across populations as measured using read-depth based Qst estimates (fig.S29). Middle plot: Variation in normalised read depth indicates the presence of a large deletion in the non-parasitic ecotype in the Endrick (ew), Oir (oir) and Norwegian (nor) rivers at around 12Mb on chr1 (grey bar). The deletion is located between two parasitism-QTL peaks, but without any significant SNPs within the deletion region as no sites were called in non-parasitic individuals in this region. Genes (*CNTNAP2 and AMPH*) partially overlapping this putative deletion are shown as dark grey rectangles at the top. Right plot: The parasitism-QTL on chr30 shows signs of reduced read depth in non-parasitic ecotypes, which could indicate a heterozygous deletion in this ecotype. The putative deletion overlaps the follicle-stimulating hormone receptor gene (*FSHR*).

Maintenance of life-history variation through selection in parasitic ecotypes

We hypothesised that selection would act more strongly on parasitic ecotypes to maintain co-adapted trait combinations related to feeding and migration. In contrast, trait loss, as observed in non-parasitic ecotypes (e.g. skipping the juvenile migration and feeding phase), has regularly been linked to relaxed selection (*16*, *54*, *55*). Parasitism-QTL showed higher levels of genetic differentiation compared to the genomic background (Fig. 4 A,B, fig. S24), with genetic differentiation being moderately correlated across ecotype pairs in the Oir and Endrick (Pearson correlation of Fst: r = 0.37, t = 163.28, df = 175083, p-value < 2.2e-16). However, not all parasitism-QTL were strongly differentiated in all ecotype pairs, e.g. with *trans-inv56* only showing weak differentiation between ecotypes in the Endrick compared to strong differentiation in the Oir (Fig. 3C,4A, fig. S25), which is likely explained by the presence of all three karyotypes in the non-parasitic Endrick ecotype. In contrast, parasitic individuals were either homozygous for the “parasitic” karyotype (higher frequency in parasitic ecotype; $AP^{trans\text{-}inv56}/AP^{trans\text{-}inv56}$) or heterozygous ($NP^{trans\text{-}inv56}/AP^{trans\text{-}inv56}$), but never homozygous for the “non-parasitic” karyotype ($NP^{trans\text{-}inv56}/NP^{trans\text{-}inv56}$), suggesting potential selection against the non-parasitic karyotype in parasitic lampreys. Thus, we argue that *trans-inv56* might not determine ecotype but rather impacts one or multiple ecotype-linked phenotypes (Fig. 3A, fig. S14). Overall, genetic differentiation within parasitism-QTL was higher between parasitic and non-parasitic ecotypes than between sympatric parasitic ecotypes (i.e. anadromous- vs lake-parasitic in the Endrick) (Fig. 4B), supporting their general role in parasitism-loss.

In line with our above predictions on selection, we found that parasitism-QTL on chr1, chr8, chr30 and chr56 showed stronger signatures of positive selection (higher population branch excess, PBE) in parasitic ecotypes compared to non-parasitic ecotypes in the Oir and Endrick (Fig. 4C, fig. S26-S27). In contrast, the parasitism-QTL on chr25, containing the *PPP5C* gene, which is linked to fat metabolism and glucose balance

(*56*), showed stronger signatures of selection in non-parasitic lamprey, suggesting the potential alteration of metabolic traits with the loss of parasitism.

Lastly, we made use of the extensive introgression from the lake-parasitic into the non-parasitic Endrick ecotype to test for signatures of selection against admixture around parasitism-QTL (Fig. 1D-F), which we argue would manifest in the (near) fixation of life-history topologies (T_{LH}) around parasitism-QTL despite genome-wide admixture. Topology weighting analyses across all three Endrick ecotypes and an adjacent non-admixed, non-parasitic population (Ross Burn; Fig. 4D) showed that a topology clustering the lake-parasitic and non-parasitic Endrick ecotypes (Admixed topology: T_{ADM}) was most common across the genome (Fig. 4E), which is expected due to their strong introgression (*24*). However, parasitism-QTL on chr1 and chr30 were enriched for windows with fixed or nearly fixed life-history topologies (T_{LH}) (Fig. 4E,F, fig. S28), as expected for selection against admixture in those regions (*31*).

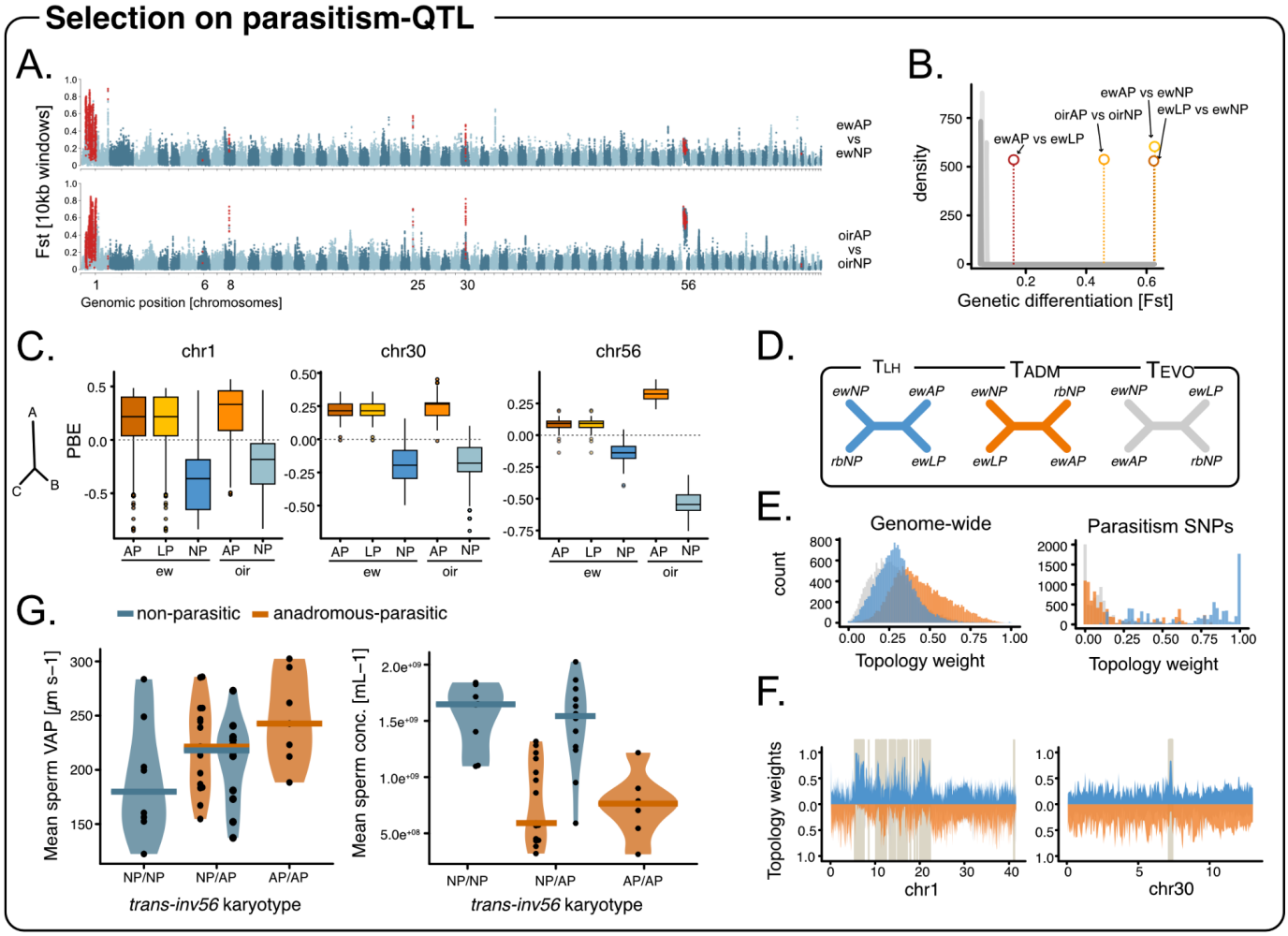

Figure 4: Selection on parasitism loci

(**A**) Landscape of genetic differentiation (Fst) in 10kb sliding windows (5kb steps) between sympatric life history forms in the Endrick (Top: Anadromous-parasitic vs non-parasitic) and the Oir (Bottom: Anadromous-parasitic vs non-parasitic). Genomic windows overlapping parasitism-QTL are highlighted in red.

(**B**) Higher genetic differentiation of genomic windows overlapping parasitism-QTL compared to the genomic background. Comparisons are shown between all ecotypes in the Endrick (ew) and the Oir ecotypes. Grey distributions show the resampled distribution of mean Fst values for n windows per comparison (n = number of parasitism-QTL overlapping genomic windows).

(**C**) Higher Population Branch Excess (PBE) in parasitic compared to non-parasitic ecotypes for genomic-windows overlapping parasitism-QTL on chr1, chr30 and chr56, which show some of the strongest signals of ecotype association and/or genetic differentiation. The drawing on the left shows the comparison, with the focal population A displaying a longer branch length compared to outgroups C and B, indicative of positive selection. PBE scores for all parasitism-QTL are shown in fig. S27.
(**D**) Topologies for the three Endrick ecotypes and the non-parasitic ecotype from a neighbouring stream (Ross burn; rbNP) used in *TWISST*. The T_{LH} topology shows the phylogenetic relationship expected for clustering by life-history, the T_{ADM} topology shows the relationship expected based on the genetic admixture between lake-parasitic and non-parasitic Endrick ecotypes, and the T_{EVO} topology shows the phylogenetic relationship between ewNP and ewAP expected based on their evolutionary relationship.
(**E**) Distribution of topology weights for each topology (in D) across the entire genome (left) and only genomic windows overlapping parasitism-QTL (right). Distributions are coloured according to their topology.
(**F**) Genome scan of topology weights in 50kb windows across chr1 and chr30, with topology weights for the life-history topology (T_{LH}; see D) shown in blue above 0 and the admixed (orange) and alternative (grey) topologies below 0. Parasitism-QTL are shown in black and broader parasitism-QTL regions highlighted as brown boxes in the plot.
(**G**) Differences in mean sperm average path velocity (VAP) and mean sperm concentration between parasitic and non-parasitic individuals from the Oir with different karyotypes for *trans-inv56*. NP/NP = homozygous for the karyotype with higher frequency in non-parasitic populations; AP/AP = homozygous for parasitic karyotype; NP/AP = heterozygous. Violin plots are coloured by ecotype (see legend), with individual datapoints in black (n=42) and coloured bars showing the mean for each group (karyotype by ecotype).

Trans-inv56 is associated with ecotype divergence in sperm velocity

To better understand patterns of admixture and the role of *trans-inv56*, we genotyped 1072 individuals (945 adults and 127 larvae of unknown ecotype) from a hybrid zone in the Oir at 131 SNPs, targeting SNPs within parasitism-QTL on chr1 (n=9 SNPs; fig. S29A), *trans-inv56* (n=21 SNPs, fig. S29A) and non-divergent SNPs outside parasitism-QTL ('neutral': n=101 SNPs). This confirmed high levels of genome-wide admixture between ecotypes (fig. S29A), although proportions of admixed individuals with ancestry proportions (Q) above or below 25% and 75%, respectively, varied between life-stages (fig. S29). Of parasitic and non-parasitic adults, 48.9% and 58.9% were classified as admixed, respectively (adult mean: 53.9%), whereas 79.5% of larvae were admixed (Adult vs larvae admixture proportion: X^2 = 28.8, df = 1, p = 8e-08). A decrease in admixture with age could point toward potential viability selection against hybrids at later life stages (*57*).

Individual-based simulations of hybridization between parasitic and non-parasitic lampreys (fig. S30) suggest that without selection or hybrid breakdown, the non-parasitic ecotype nearly disappears, likely due to the higher fecundity of the anadromous-parasitic ecotype. Thus, similar to previous models (*58*), our simulations indicate the need for selective advantages and / or selection against hybrids to explain the persistence of the non-parasitic ecotype in sympatry.

Despite the potential selection against hybrids, we detected an excess of *trans-inv56* heterozygotes ($NP^{trans\text{-}inv56}/AP^{trans\text{-}inv56}$) in both ecotypes, which was stronger in the non-parasitic ecotype (fig. S31). This could potentially indicate overdominance (i.e. heterozygote advantage) in non-parasitic lamprey. In line with this overdominance hypothesis, we found that non-parasitic heterozygotes ($NP^{trans\text{-}inv56}/AP^{trans\text{-}inv56}$) have faster but not more sperm compared to homozygotes ($NP^{trans\text{-}inv56}/NP^{trans\text{-}inv56}$) (LMM: beta = 27.72, $t_{(38)}$ = 2.50, p = 0.017; Fig. 4G; fig. S32), a likely fitness advantage in communal spawners and for smaller sneaker males (*59*). In contrast, heterozygous parasitic males ($NP^{trans\text{-}inv56}/AP^{trans\text{-}inv56}$) had slower sperm compared to homozygous males ($AP^{trans\text{-}inv56}/AP^{trans\text{-}inv56}$) (Fig. 4G), supporting a scenario of selection against the non-

parasitic karyotype in the parasitic ecotype. A role of *trans-inv56* in sperm velocity is further supported by the presence of genes linked to sperm traits inside the inversion (e.g. *CUL4B*, *MFF*, CPT2, *IRS1*; table S2), providing key candidate genes underlying variation in sperm velocity between lamprey ecotypes, and targets for future functional studies.

Discussion

Our study resolves the genomic basis of a striking life-history transition in an early-diverging vertebrate lineage; the repeated loss of parasitism in lampreys. Rather than the single supergene-style architecture often inferred for analogous transitions in other taxa, such as large inversions controlling life-history variation in rainbow trout (*6*), ruff (*7*), or cod (*5*), or the highly polygenic basis of migration-history in brown trout (*60*), we find that this life-history syndrome in lampreys maps to a small number of regions with distinct genomic architectures. Below we consider, in turn, what these regions implicate biologically, how their architectures are maintained, and what their non-shared identity in a North American sister lineage reveals about the repeatability of this transition.

The genes implicated by these regions cluster around three major axes that mirror the components of the syndrome itself. The peak on chr30, including a putative deletion in non-parasitic lampreys, overlaps key genes in the gonadotropin pathway (*FSHR*, *GPHB5*), a central regulator of vertebrate sexual maturation (*61*). This is consistent with the accelerated post-metamorphic maturation that distinguishes non-parasitic from parasitic lampreys (*16*). Direct functional evidence in fish establishes FSHR as a key component of maturation transition, with *fshr*-knockout in Atlantic salmon preventing testicular maturation (*62*), and *fshr* loss-of-function in medaka causing folliculogenesis arrest (*63*). In humans, FSHR variants are associated with timing of pubertal onset and reproductive length (*64*, *65*). Notably, the dominant maturation-timing locus in salmonids (*vgll3*) (*66*, *67*) acts upstream of the gonadotropin axis, whereas in European lamprey variation in sexual maturation seems to be orchestrated through changes in the receptor itself. Furthermore, genes under parasitism-QTL on chr1 and within the chr56-inversion implicate neural development and synaptic function (e.g. *CNTNAP2*, *KCNH1*, *AMPH*, *EN2*, *GSHR*, *WNT1*), pathways plausibly associated with the transition from a migratory and predatory behaviour in parasitic lampreys to a brief, stream-resident juvenile phase in non-parasitic lampreys (*16*). Metabolic genes under parasitism-QTL on e.g. chr1, chr8 and chr56, that were also differentially expressed in livers (e.g. *CPT1A*, *CPT2*, *IRS1*, *INSIG1*) further implicate fatty-acid β-oxidation and insulin signalling, consistent with the metabolic transition away from juvenile feeding in non-parasitic lampreys toward reliance on stored reserves through metamorphosis and spawning (*16*, *43*, *44*). This convergence between gene function and phenotypic axis points toward specific developmental and metabolic pathways that must be co-modified for this life-history transition to occur. However, the underlying mechanistic changes regulating this life-history shift remain unresolved.

Yet, how is genetic divergence between lamprey ecotypes maintained in the face of gene flow? Most parasitism-QTL on chromosome 1 fall within a ~20 Mb region of suppressed recombination that does not seem to be caused by large-scale intra-specific rearrangements. Yet, the same region carries inter-specific rearrangements. One parsimonious interpretation is that these ancestral rearrangements established a

recombination cold spot through altered chromatin architecture, which has subsequently captured and held together ecotype-divergent variation. This is consistent with recent work on ancient rearrangement-derived recombination cold spots in mammals (*12*, *13*) and the suggested impact of ancient chromosomal fusions on adaptation in other systems (*10*, *52*, *53*, *68*). However, the link between low-recombination regions due to ancient rearrangements and adaptive divergence has been relatively unexplored so far, and direct functional tests are needed to establish mechanistic links between these regions and phenotypic change. Altogether, these observations suggest that ancient, fixed rearrangements may act as a substrate for the maintenance of adaptive variation under gene flow in a manner that is mechanistically distinct from, but functionally analogous, to the well-described inversion supergenes. In contrast, a large translocated-inversion on chr56 (trans-inv56) seems to contribute to life-history variation by affecting variation in sperm velocity, an ecotype-related trait (*59*) that is critical for sperm competition in externally fertilising species (*59*, *69*, *70*). Interestingly, the genomic basis of sperm trait evolution has rarely been dissected within or between species in the wild (*71*, *72*), and links to chromosomal inversions are rarer still (*72*). The persistence of karyotype differences between ecotypes, despite the excess of heterozygotes, might indicate the presence of evolutionary trade-offs, a balanced-polymorphism mechanism well-established for inversion supergenes (*73–75*). Together with the chr1, trans-inv56 illustrates that different forms of recombination-suppressed architecture can coexist within a single life-history transition.

Lastly, our results for the North American sister lineage *Occidentis ayresii*, which has independently evolved a similar parasitic vs non-parasitic ecotype pair, indicates that the underlying loci are largely non-shared across lineages. Yet *O. ayresii* parasitism loss appears itself to involve a mosaic of broad and narrow regions of genetic differentiation, suggesting that the architectural mosaic is conserved across the genus while the specific loci are not. This pattern contrasts strikingly with canonical examples of repeatable parallel evolution at the locus level - *Eda* and *Pitx1* in three-spine sticklebacks (*76–78*) *or optix* in *Heliconius (79)* - and argues that life-history transitions of this type can recur through architecturally similar but locus-distinct genomic routes. Our comparisons across lineages further suggest that rearrangements on chr1 likely arose within the last 10 million years (*17*) and potentially following the split of *Lampetra* and *Occidentis* (divergence ~8MYA) (*17*), consistent with this lineage-specific genomic architectures. We note, however, that the *Occidentis* comparison rests on low-coverage sequencing (~0.5×) and that high-coverage re-sequencing will be required to definitively rule out shared loci of small effect. Future comparative and functional genomic studies - especially across more distantly related parasitic / non-parasitic species pairs and across the full lamprey phylogeny - will reveal whether mosaic architectures of this kind are a common feature of similar life-history transitions and provide insights into the mechanistic basis of parasitism loss and associated developmental changes in lampreys.

Acknowledgement

Funding: AJ was supported by a NERC Independent Research Fellowship (NE/W008963/1), a Leverhulme Trust Early Career Fellowship (ECF-2020-509), a Lord Kelvin Adam Smith Fellowship from the University of Glasgow, a Carnegie Trust Incentive Grant to Kathryn Elmer and Arne Jacobs (Co-I AJ), and funding from the Glasgow Natural History Society. This project was further supported by the Research Council Norway (to KSJ).

Samples: We want to thank Hannele Honkanen, John Hume, Peter Koene, Matt Newton and Flora Rendell-Bhatti for help with the collection of Scottish samples. The samples from Oir used in this study were provided by the Biological Resource Centre Colisa (DOI: Biological Resource Centre Colisa) part of BRC4Env (DOI: https://doi.org/10.15454/TRBJTB), of the Research Infrastructure AgroBRC-RARe. Norwegian samples were collected by Ole-Håkon Heier and Eivind Sandvik Schartum. German samples were provided by Alexander Brinker, LAZBW.

Sequencing: We also want to thank Sophie Michon-Coudouel and Romain Causse-Vedrines (EcogenO Platform, University of Rennes) for the generation of the HiPlex data as well as Inês Trancoso for high molecular weight DNA extraction. The sequencing of Norwegian samples was performed by the Norwegian Sequencing Center (NSC).

Furthermore, we want to thank Kathryn Elmer and Thomas Boehm for help with sample collection, funding acquisition and data generation, Catherine Labbé for sperm speed analyses and the Evolution & Diversity Theme and Shared Interest Group at the University of Glasgow as well as Claire Mérot for stimulating discussions.

Author contributions:

Conceptualisation: AJ, GE; Sampling: AJ, GE, JT, JPD, QR, LAV, ND; Data generation: AJ, MC, ALB, SS, EJP, GL, DJP, BJP, GE, AS, QR, ND, RN, KSJ, SJ, SNKH, BGA; Analysis: AJ, ND, BGA, OKT, FG; Funding: AJ, GE, KSJ; Writing initial draft: AJ, GE; Writing final draft: All authors.

Conflicts of Interest

The authors declare no conflicts of interest

Data and code availability

All data will be made available upon acceptance of the manuscript. The raw sequencing files will be accessible on ENA under the BioProject XXXXX. Additional data will be available on xxxx.

Methods

Sampling

European lamprey used in this study were collected using a range of methods in different waterways across Europe (Table S1). Sampling strategies are described below.

UK: Juvenile and adult individuals were collected in the River Endrick, Gryfe and Ross Burn in Scotland (UK) using electrofishing, smolt traps or specialised lamprey traps ('Maitland traps') during the upstream migration (Autumn) between 2019 and 2022 (*80*). Ecotypes were identified based on size and colouration (Fig. 1A) (*24*, *81*), euthanised using an overdose of MS-222 and tissue samples collected for DNA extraction. Larvae were sampled in the River Tweed using electrofishing as described in (*82*) and whole-specimens frozen. These were determined to be non-parasitic, as sample sites were not reachable by anadromous-parasitic lamprey. Sampling was performed under the University of Glasgow ethical approval, with sampling permits from the Marine Directorate of Scotland.

France/Oir: Parasitic and non-parasitic individuals were collected in sympatry in the Oir, a tributary of the Sélune, France, at spawning time as described in (*25*, *27*). Individuals were anesthetized with benzocaine to collect a fin clip, which was preserved in 95% EtOH.

Norway: Non-parasitic individuals were sampled using electrofishing in Glomsrødbekken, Østfold county, and parasitic individuals were sampled using electrofishing in a tributary to the Saua River, Lake Nordsjø, Telemark county. The parasitic population presently have no or only limited access to the sea following construction of different dams and locks.

Brigach, Danube: Larval samples were collected by the LAZBW in the Brigach, a tributary of the Danube in Germany, using electrofishing and whole larvae preserved in ethanol. These were determined to be non-parasitic, as anadromous-parasitic lampreys don't exist in this system.

Whole-genome resequencing

DNA was extracted from fin clips using either the Qiagen Blood & Tissue Kit or using a bead-based extraction protocol (dx.doi.org/10.17504/protocols.io.b46bqzan). Whole genome sequencing libraries for all individuals, except the Norwegian populations, were prepared using NEBnext Ultra II FS DNA library prep with 100ng of DNA as input and using half-reactions. Pooled libraries were sequenced on NOVASEQ X or NOVASEQ X Plus 25B lanes. Whole genome sequencing libraries for Norwegian samples were prepared at the Norwegian Sequencing Centre from DNA extracted using the Qiagen DNeasy Blood and Tissue Kit (from 25 mg muscle tissue) sequenced on an Illumina NovaSeq 6000 lane. We furthermore downloaded Illumina whole genome sequencing data for a single *Lampetra auremensis* individual from Portugal (SRR24224961), a sister species to *Lampetra fluviatilis/planeri*.

WGS processing and SNP calling

Raw sequencing data for all individuals were processed and trimmed using FastP with the following settings: *-l 30 --trim_poly_g --cut_right --detect_adapter_for_pe (83)*. We aligned processed reads to the Darwin Tree of Life *Lampetra fluviatilis* reference genome (kcLamFluv1.1; GCA_964198595.1 on NCBI) using *bwa-mem* (*84*), merged replicated samples using the merge command in *sambamba* v.0.8.2 (*85*), removed reads with mapping quality below 30 and removed duplicated reads with *sambamba* using *--remove-duplicates*, clipped overlapping reads using the *bamUtils clipOverlap* command (https://github.com/statgen/bamUtil), and indexed the bam files using *sambamba index*. Finally, we performed indel realignment per sample using *GATK3.8* with the *RealignerTargetCreator* and *IndelRealigner* commands. We estimated the depth of coverage for each bam file using *mosdepth* v.0.3.11 with the -n --fast-mode --by 500 --mapq 30 flags (*86*).

We performed population genomic analyses using genotype likelihoods in *ANGSD* v0.938 to account for genotype uncertainty (*87*, *88*). We identified variant sites across all individuals using *ANGSD* with the samtools genotype likelihood model (-GL 1), a minimum SNP value of 1e-6 (-SNP_pval 1e-6), determined major/minor alleles based on allele frequencies (-doMajorMinor 1), removed sites below and above a minimum and maximum total sequencing depth of 184 (number of individuals used to call SNPs) and 2500 (2 x total number of individuals x mean sequencing depth), respectively, across all individuals, and removed sites with missing data in more than 20% of individuals (-minInd 153), minimum base quality below 30 (-minQ 30) and a minimum individual depth of 1 (-setMinDepthInd 1). Furthermore, we removed multi-mapping reads (-uniqueOnly 1), unmapped or duplicated reads (-remove_bads 1) and those without a matching paired read (-only_proper_pairs 1). We also adjusted the mapping quality for excessive mismatches (-C 50). We only kept SNPs with a minor allele frequency of 5% (-minMaf 0.05). We then removed SNPs falling within potentially problematic genomic regions that can impact the analyses accuracy. We removed sites that fall within regions with low mappability (*89*). We estimated mappability using a kmer approach in *genmap* with 150-mers and we only kept SNPs within uniquely mapping regions (mappability of 1) (*90*). Furthermore, we removed sites with excess heterozygosity and all sites within 10kb of those sites. For that, we estimated inbreeding statistics using PCAngsd (*91*) and removed all sites (± 10kb) with F < -0.95 and P-value < 10−6 (*92*). This final, filtered SNP list was used as the basis for all further analyses across all samples.

Mitochondrial variant calling and analysis

Genotypes were called for all whole genome reads mapping to the scaffold representing the mitochondrial genome across all individuals using *FreeBayes* (*93*) with ploidy set to 1. Subsequently, we filtered the resulting dataset using *vcftools* to keep all variant sites that were genotyped in more than 90% of all individuals, with mapping and genotyping quality above 30, and with minimum coverage of 5x and a maximum coverage of 100x. We reformatted the resulting VCF into nexus format using the *vcf2phylip* script (https://github.com/edgardomortiz/vcf2phylip) and generated a median joining haplotype network using *PopArt* (*94*).

Population genomics based on genotype-likelihoods

Population structure and phylogenetics

We investigated population structure across all populations using Principal components analyses with *PCAngsd* based on all SNPs and based on a thinned SNP dataset (1 SNP every 50kb; informed by decay of linkage-disequilibrium [LD]) with highly-divergent regions on chromosome 1 and 56 removed (thinned SNP dataset). We also inferred genetic ancestry using an admixture approach in *PCAngsd* for K=2 to K=12. We evaluated the most likely K using correlations matrices of residuals in evalAdmix (*95*). We also estimated population structure for the Oir and the Clyde catchment (Endrick, Ross Burn, Gryfe), separately, using *PCAngsd*, only using sites with minor allele frequencies above 5% in these subsampled datasets. Moreover, we also estimated PCAs for parasitism-associated regions using *PCAngsd*.

Furthermore, we inferred phylogenetic relationships between individuals and populations using several ways. First, we build neighbour-joining (NJ) trees using the bioNJ command in the *ape* v.5.8-1 R-package (R version 4.3.0 (2023-04-21)) based on a genetic distance matrix (for all SNPs and the thinned SNP dataset), which was inferred using *ANGSD* (--doIBS). Second, we estimated population relationships and inferred gene flow events between populations using *TreeMix* in the *BITE* V2 R-package (*96*, *97*). We first estimated minor allele frequencies for each population ('ecotype by location') using *ANGSD*, converted them into minor allele counts by multiplying the maximum likelihood sample frequency by the SNP-specific sample size, and created a minor allele count matrix across all SNPs and populations as input for *TreeMix*. We ran *TreeMix* with the single *L. auremensis* individual as the outgroup in the *BITE* R-package for 0 to 5 migration edges. First, *BITE* builds a phylip consensus tree with 500 bootstraps, and then models migration events along this tree. We analysed SNPs in blocks of 1000 SNPs to account for linkage between SNPs. We plotted trees with *BITE* for 1 to 3 migration events, and determined the most likely tree based on the Evanno method in OptM (*98*) and residuals plots. Second,, as the phylogenetic relationship between populations likely differs across the genome due to selection, gene flow and incomplete lineage sorting, we performed topology weighting analyses in *TWISST* (*30*) across the genome for i) the anadromous-parasitic and non-parasitic ecotypes from Oir and Endrick and ii) all three ecotypes from the Endrick together with the non-parasitic Ross Burn population. Due to the use of genotype likelihoods in our analyses to account for genotype uncertainty, we implemented a custom pipeline for *TWISST* (*89*). First, we estimated the IBS matrix per 50kb window (based on LD decay, below) using single read sampling (-doIBS 1) in *ANGSD* and then generated one NJ tree per window using the *bioNJ* command in *ape*. The weightings across all subtrees were calculated using the 'complete' method in *TWISST*.

Linkage disequilibrium (LD):

We inferred LD per ecotype by population using *ngsLD v1.1.1* (*99*). We created two different LD datasets. First, to identify windows of high LD across the genome, we ran *ngsLD* on all sites per ecotype by population and estimated LD between sites that were maximum 1kb apart to reduce computational burden. For plotting, we summarised LD estimates in 10kb sliding windows (2.5kb steps) using the *WindowScanR* R-package, and

plotted the loess-smoothed values (span=0.1) across the genome using ggplot2. Second, to estimate LD decay rates, we only kept sites on chromosome 2, 5 and 10, and estimated LD up to a maximum distance of 100kb between sites for each ecotype by river using *ngsLD*. LD decay was plotted using custom scripts in R.

Genome-wide association study for life-history using GEMMA:

To map the genomic basis of life-history, we used a binary genome-wide association analysis (GWAS) in *GEMMA* (*100*). We performed the GWAS for life-history based on mean genotypes to account for genotype uncertainty. Mean genotypes were inferred as the mean of scaled genotype probabilities per SNP using custom scripts and genotype probabilities stored in beagle format (generated with *ANGSD* using the -GL 3 command). Mean genotypes in BIMBAM format were used as input for *GEMMA* to first generate a covariance matrix and subsequently test for genetic associations with life-history coded as 0 for parasitic and 1 for non-parasitic using linear mixed models with the Wald test (-lmm 1). We performed GWAS for all individuals in the dataset and only for individuals from the Endrick catchment to reduce the potential impact of population structure.

GO term enrichment analysis:

We performed the functional annotation based on annotated genes (REFseq annotation for kcLamFluv1.1 from NCBI) and gene ontology terms provided with the REFseq annotation (.gaf file). using *clusterProfiler* version 4.8.2 (*101*) in R. In *clusterProfiler*, we performed gene ontology overrepresentation analyses (ORA) for genes containing significant parasitism-associated SNPs, and gene set enrichment analyses (GSEA) with genes ranked by the highest p-value (from parasitism GWAS) within each gene that contained a SNP. We did not perform a correction for multiple testing due to the hierarchical structure of GO terms. However, we summarised the results based on semantic similarity using the *simplifyEnrichment* R-package (*102*) and plotted the top 25 GO terms for each test using *clusterProfiler*.

Genetic differentiation and selection:

We estimated genetic differentiation between ecotypes within rivers using *ANGSD* based on all filtered SNPs. We estimated the 2D site frequency spectrum for each comparison using *realSFS* and used *realSFS fst* to estimate Fst for each comparison. We summarised Fst values in 10kb sliding windows (5kb steps) using the *realSFS fst stats2* command, and plotted the results as manhattan plots using the *CMplot* R-package and ggplot2.

To estimate ecotype specific selection, we estimated the population branch excess statistic (PBE), an extension of the population branch statistic (PBS), using the approach outlined in Shpak et al. (*103*). PBS aims to identify selective sweep signatures by comparing the focal branch length of a locus within a population to two outgroup populations, and the PBE statistic extends this approach and estimates selective sweeps by testing if the PBS value at a site exceeds the expected value under neutrality. PBE for each locus (i.e. Fst for 10kb sliding window) in the focal population A was estimated as: $PBE_A = PBS_A - PBS_{A\text{-}exp} = PBS_A - [(T_{BC} \times \text{med}(PBS_A)) / \text{med}(T_{BC})]$, with $\text{med}(PBS_A)$ and $\text{med}(T_{BC})$ being the median values of their respective statistics

across all loci. $PBS_A = (T_{AB} + T_{AC} - T_{BC}) / 2$, with T being the linearised Fst between populations at this locus (T = -log(1-Fst). Fst values were mean Fst values per 10kb sliding window as estimated above, and for estimates in the Endrick system we used the three ecotypes (ewAP, ewNP, ewLP), changing the focal population to estimate population-specific PBE values. For the Oir, we used the two ecotypes and the lake-parasitic ecotype from the Endrick as an outgroup. We used these estimates to compare PBE values for genomic windows containing parasitism associated SNPs between ecotypes by river using Wilcoxon rank sum tests.

Population genomics of Occidentis ayresii

To test if the genomic basis of parasitism loss is convergent or divergent across lamprey lineages, we re-analysed low-coverage whole genome sequencing data (mean coverage depth = 0.51x, range = 0.27x to 1.26x) for sympatric parasitic Western River Lamprey and the non-parasitic Western Brook Lamprey ecotypes of *Occidentis ayresii* (formerly *Lampetra ayresii* and *L. richardsoni,* respectively) (*33*). *Occidentis ayresii* is a North American sister lineage to *Lampetra*. These lineages have diverged approximately 5-10MYA (*17*, *33*). As no highly contiguous reference genome was available for *O. ayresii*, we performed synteny-guided scaffolding of a fragmented draft genome (*33*) against the DToL *Lampetra fluviatilis* reference genome (kcLamFluv1.1) using *RagTag* (*104*) with default settings, anchoring ~882Mb (~87% of the genome) into chromosomes. We also lifted over the annotation from kcLamFluv1.1 to the scaffolded reference using *Lifton* with default settings (*105*). We identified variant sites using *ANGSD*, retaining sites using the same filtering strategy as for European lamprey (see above), changing the depth and minimum individual filters. We retained sites with a minimum mean depth above 39 and below 100, and present in at least 18 individuals. We generated a genetic distance matrix based on single read sampling (-doIBS 1) in *ANGSD* and used multidimensional scaling to confirm the population structure. Furthermore, we estimated per-SNP Fst between ecotypes with *ANGSD* and *realSFS* as described for European lamprey. Per SNP Fst estimates were plotted using CMplot. Lastly, we performed principal components analysis for the main outlier regions (based on elevated Fst) between ecotypes, with *PCAngsd* to identify putative signals of structural variation.

Structural variant analysis

Genome assemblies

For the analyses of chromosomal rearrangements, we used newly, publicly available chromosome-scale reference genome assemblies for the anadromous-parasitic European river lamprey (kcLamFluv1.1 primary assembly from the Darwin Tree of Life [NCBI] and *kcLamFluv2.1* haplotype 1 and 2 from the Norwegian BioGenome Project (*49*)) and European brook lamprey (kcLamPlan1.2 haplotype 1 and 2 from the Norwegian BioGenome Project (*49*)). These haplotype-resolved assemblies from the Norwegian BioGenome Project are earlier versions of the ones available on NCBI (kcLamFLuv2.2 h1 and kcLamPlan2.2 h2), but are comparable and just differ in scaffold naming. Furthermore, we assembled haplotype-resolved and chromosome-scale genomes for one European river lamprey and one European brook lamprey from the Oir using a combination

of PacBio Hifi and Hi-C (*49*) (see Table S3 for assembly details). For genome assemblies, fresh milt (sperm) was collected from spawning-ripe males caught in the Oir, immediately flash-frozen in liquid nitrogen, and stored at -80°C until further processing. High-molecular weight (HMW) DNA was extracted from 100µl of milt using the Nanobind DNA extraction kit for tissue, following the providers manual, and quality control performed using a Nanodrop Spectrophotometer, Agarose gel, Qubit and Agilent Tapestation with the Genomic DNA tape. HMW DNA with an average fragment length above 60kb was sent for library preparation and sequencing on one PacBio Revio SMRT cell per sample at Edinburgh Genomics. Hi-C libraries were prepared from 100µl of milt from the same individual using the PhaseGenomics Proximo Hi-C kit following the provider's manual. Final Hi-C libraries were sequenced to 100M reads per library on a NovaSeq X Plus 10B flow cell (together with other samples). We performed genome assemblies using established pipelines for European lampreys (*49*). First, we filtered PacBio Hifi reads using HiFiAdapterFilt-3.0.0 and assembled haplotype-resolved contigs from filtered PacBio Hifi reads using *hifiasm v.0.19.7-r598* (*106*) with the Hi-C option for phasing, and everything else at default settings. Second, we processed Proximo Hi-C reads following provider's recommendations, aligned reads to the contigs using *bwa mem* using the -5SP setting, and performed quality control of Hi-C alignments using the *hic_qc.py* script provided by PhaseGenomics (https://github.com/phasegenomics/hic_qc). After marking duplicates with *sambamba*, we scaffolded the haplotype-specific assemblies using *YAHS* (*107*) with a mapping quality threshold of 10 (-q 10) and default settings. We visualised scaffolded assemblies using *PretextView* (*108*). Lastly, we improved the scaffolded haplotype-resolved assemblies through an additional round of reference-guided scaffolding using *RagTag* v2.1.0 (*104*). The brook lamprey assembly was scaffolded against LamPlan1.1_h2 and the river lamprey assembly was scaffolded against the reference genome (kcLamFluv1.1).

Chromosomal rearrangement identification

We used two different approaches to identify chromosomal rearrangements between European lamprey ecotypes (see above for details on all assemblies) and across lamprey lineages. We used chromosome-scale assemblies for the parasitic *Petromyzon marinus* (UKy_Petmar_22M1.pri1.0 GCF_048934315.1) (*51*) and non-parasitic *Lethenteron reissneri* (ASM1570882v1 - GCF_015708825.1) (*50*) as outgroup species.

First, we investigated synteny and identified large-scale chromosomal rearrangements across different combinations of genome assemblies using *ntSynt* (*109*) using default settings and visualised results using *ntSynt-viz* (*110*).

Second, we identified rearrangements between *Lampetra spp.* assemblies using *SyRi* (*111*). As *SyRi* requires genomes to have the same sequence names, we scaffolded all genomes for each comparison against a common reference (kcLamFluv1.1) using *RagTag* v2.1.0 (*104*) and scaffolds renamed following the reference genome. Re-scaffolded assemblies were aligned against each other using *minimap2* with the -ax asm5 -t 8 --eqx settings (*112*), and pairwise alignments used as input for *SyRi* using default settings. SyRi results were plotted using *plotsr (113)*.

Copy-number variation:

We inferred copy number variation (CNV) between ecotypes from normalised read depth following the approach outlined in (*114*). First, we estimated read depths in 10kb windows across the genome for each individual using *mosdepth* (*86*) for reads with mapping qualities above 30. Next, we normalised read depths by dividing the estimated read depth for each window by the genome-wide median using custom R scripts, and identified putative CNVs as windows with normalised read depth 2 standard deviations above or below the mean. Furthermore, to identify genomic windows that differ in normalised read depth between sympatric ecotypes, indicative of divergence CNVs, we estimated Qst values for each window between sympatric ecotypes using a linear mixed model approach outlined in (*114*), which adapted the Qst and population variation approach outlined in (*115*). In brief, we performed linear mixed models in the lme4 R-package to extract within and among population variance based on unnormalised read depths, correcting for variation in average read depth by including the per-sample median read depth as a covariate, and population/ecotype as a random factor. The variance component attributed to the population was used as the among-population variance ($V_{A,\ among}$) and the model's residual variance component was used as the within-population variance ($V_{A,\ within}$), which were used to estimate Qst using the following formula: $Qst = V_{A,\ among} / (V_{A,\ among} + 2 * V_{A,\ within})$. We conservatively identified outlier CNVs as those above the 1% genome-wide Fst distribution following a Qst - Fst approach (*114*), and inspected the normalised read depth distribution around putative outlier CNVs overlapping parasitism-QTL.

Recombination rate estimation

To estimate recombination rate variation across the lamprey genome for specific ecotypes, we used different complementary approaches.

First, we used allele frequency data and machine learning approaches to estimate population-scale recombination rates for each ecotype in the Oir and Endrick population from whole-genome sequencing data using *ReLERNN (116)*. We estimated allele frequencies per ecotype and population using *ANGSD* for all filtered SNPs, and ran *ReLERNN* using default settings with allele frequencies, which has been shown to be accurate for reconstructing recombination rate landscapes (*117*).

Second, we reconstructed linkage maps for a hybrid family based on RADseq data to investigate recombination patterns. To construct our linkage map, we used the RAD sequencing data generated for a single cross of parent caught in the Oir (*25*). We first demultiplexed the raw data using the program process_radtags of the *Stacks* software v2.4 (*118*, *119*). *process_radtags* was used to check reads overall quality, trimmed read to 85bp and removed reads containing Illumina adapters. The reads of each individual were aligned to the HiFi NP draft genome (earlier version) using BWA-mem (*84*), and the output transformed to bam format with samtools (*120*). We then ran *gstacks* (Stacks v2.4) in reference-based mode and removed the PCR duplicates via gstacks. We removed unpaired reads and a minimum PHRED-scaled mapping quality of 25 was set to consider a read. We also applied a maximum soft clipping level of 5% of the read length to reduce sequencing errors. The *populations* program (*Stacks*) was used to export the results in VCF format for further filtering and only keep informative individuals for linkage map construction: the 2 parents

(anadromous-parasitic female and non-parasitic male from Oir population) and their 90 offspring. The SNPs at positions 60 were removed due to a sequencing problem occurring at this position (*25*). We filtered SNPs using *vcftools* (*121*) to remove low quality genotypes and SNPs. We kept genotypes with a (mean) coverage ranging from 5x to 100x. We only included bi-allelic sites with less than 60% missing data, and a minor allele count of one. Subsequently, we filtered out two individuals with a mean depth below 10 and more than 25% missing data. We checked for mendelian inheritance errors with *bcftools*, leading to the removal of one individual. We then repeated the SNP filtering as before but only kept SNPs with less than 30% missing data. We also sought and removed singletons with the --singletons option of *vcftools*.

We used *LepMap3* (*122*) to construct the linkage maps. Two different pipelines were tested to construct the map: with or without *Filtering2*. When using filtering, *Filtering2* was run with the following parameters: removeNonInformative=1 and dataTolerance=0.0001. Small dataTolerance like this is advised in single-family crosses. Then, for the SeparateChromosomes2 module we computed different linkage maps testing different lodLimit ranging from 6 to 12 and removed the LG computed with less than 20 SNPs with sizeLimit=20. When not using *Filtering2*, we ran *SeparateChromosomes2* the same way as before but with the parameter distorsion Lod=1 as advised for single family data. We expected approximately 82 chromosomes in *L. planeri* and *L. fluviatilis (123)*. We chose to continue the analysis with the pipeline with filtering because without, the first LGs contained too many markers and very few were left in the others. The selected map was passed through the module *JoinSingles2All* with parameters lodLimit=6 and lodDifference=2.

Subsequently, we aligned RADtags containing SNPs to the kcLamFluv1.1 reference genome using *bwa mem* and ordered SNPs used for linkage mapping along the reference genome. We removed RADtags that did not uniquely mapped and filtered out SNPs, for the female and male map separately, that did not show a collinear relationship between physical and genetic distance, unless ten or more SNPs showed that pattern (e.g. inversions).

Individual-based simulations of hybridization

We aimed at comparing the observed proportions of admixture with simulated proportions computed according to various scenarios of mating and hybrid breakdown. We used the forward-in-time individual-based simulation software *NEMO* v.2.3.55 (*124*). We aimed at simulating a life cycle as close as possible as the one of lampreys.

We implemented two different models. First, we modelled assortative mating (i.e. within ecotype) based on a binary indicator phenotype (AP (1) or NP (0)) and tested different levels of random mating (i.e. possibly between ecotypes): 10%, 30% or 50%. Second, we implemented a model of hybrid incompatibility where hybrid survival depended on their pedigree, with overdominance expressed in the F1 hybrids and underdominance (hybrid breakdown) expressed in the F2, backcrosses and later hybrid crosses. Individual fitness (offspring survival) was determined by ten pairs of Dobzhansky Müller Incompatibilities (DMI) loci, with free recombination. NP and AP parental types were initialised with opposite alleles (0/0 in AP, 1/1 in NP), facilitating the distinction between F1 hybrids (all 0/1 or 1/0 genotypes at all loci) from F2 and later generation

crosses (mixed 0/1 and 1/0 genotypes). Genotype fitness were set such that survival rates were 0.82 for pure-type offspring, 1 for F1's, and a minimum of 0.11 for 100% incompatible heterozygotes in F2's.e. Given the different body size of ecotypes, the fecundity of parasitic lamprey (AP) was multiplied by 3 or 5 during breeding. The phenotype of the offspring (parasitic lamprey or non-parasitic lamprey) was inherited by ancestry percentage and generations were non-overlapping. We simulated three populations: one sink population mimicking a hybrid zone receiving a unidirectional migration from two other populations containing either non-parasitic or parasitic individuals (fig. S29A). At the first generation, the sink population contained only non-parasitic individuals. We recorded the individual genotypes at the DMI loci and their phenotypes in the sink population after 200 generations for further analyses. The populations' carrying capacities were set at 1,000. We performed 50 replicates of each combination of parameters (dispersal, fecundity and assortative mating, fig. S29B). The classification of hybrids with the *NEMO* output was performed with a custom script.

Hi-Plex genotyping:

To investigate the population structure and role of chr1 and chr56 in ecotype divergence, we performed targeted genotyping of over 1000 lampreys from a hybrid zone in the Oir using a Hi-Plex amplicon sequencing approach.

Target SNP selection

Hi-Plex is an amplicon sequencing technique allowing co-amplification of all SNPs in a single multiplex PCR prior to Illumina sequencing. Here we developed a set of 148 SNPs allowing the distinction of *L. planeri* and *L. fluviatilis* and including both loci with a high divergence between ecotypes, and randomly selected SNPs. We used RAD sequencing data (609 individuals distributed across 17 populations) from (*25*) for target SNP selection. Only individuals from French populations with a genome-wide Fst < 0.1 were retained for downstream analyses (n = 121 individuals). These populations experience high gene flow and are therefore more appropriate for investigating reproductive isolation than allopatric populations with low gene flow in which highly differentiated markers may reflect population-specific drift. To call SNPs, we first demultiplexed the raw data using *process_radtags* in *STACKS* v2.4 (*119*), and used process_radtags to check read quality, trim reads to 85bp and removed reads containing Illumina adapters. Reads were aligned to a *L. planeri* draft genome assembly (based on PacBio Hifi data; see (*125*) for details) using *BWA mem*, and transformed to a sorted .bam file with SAMtools. We ran the *gstacks* module in reference-based mode to call SNPs, after using it to remove PCR duplicates, unpaired reads and removed reads with a minimum PHRED-scaled mapping quality of 25. We also applied a maximum soft clipping level of 5% of the read length to reduce sequencing errors. We exported a VCF file for further processing using the *populations* module of *STACKS*. SNPs at position 60 were removed due to sequencing problems occurring at this position (*25*).

After data visualization, we retained genotypes with a mean coverage between 5× and 100× and a minimum minor allele count of 5. Only biallelic sites with a maximum of 50% missing genotypes were kept. Eight individuals with more than 40% missing data were removed. Sites with more than 20% missing data

in either *L. planeri* or *L. fluviatilis* individuals were excluded. Finally, we retained biallelic sites with a minimum minor allele count of 5 that were genotyped in at least 85% of individuals. The filtered VCF file was converted into a GENEPOP file using *PGDspider* v2.1.1.5 (*126*). Fst was computed for both GENEPOP files using the function diffCalc from the package *diveRsity* (*127*). In total, 190 outlier loci with Fst > 0.3 between ecotypes were identified. To improve genome-wide representativeness, we added putatively neutral SNPs (hereafter "neutral SNPs"), as outlier SNPs were primarily located on chromosomes 1 and 56. Neutral SNPs were identified using the populations program (Stacks) with individuals from the Oir river. A minimum of 2% of individuals were required to present the locus, and a minimum minor allele count of 3 was required to process loci. The dataset was then split by ecotype, and genotypes with a mean coverage between 5× and 100× and a minimum minor allele count of 5 were retained. Only biallelic sites with a maximum of 50% missing data were included. Individuals with more than 60% missing data were removed. Using *VCFtools*, loci with more than 15% missing data, loci deviating from Hardy–Weinberg equilibrium, and loci with a minor allele frequency below 0.1 were excluded. The populations program was then used to filter out loci with heterozygosity below 0.5. Markers present in both *L. planeri* and *L. fluviatilis* individuals were retained.

Hi-Plex primer design

To design primers for the HiPlex assay, we extracted a 145bp sequence around each selected SNP from *L. planeri* draft genome contigs. These sequences were aligned against the *L. planeri* genome using BLASTn and only matches with an E-value < 1×10^{-20} were retained. A custom script parsed BLAST output and retained only sequences with a unique match. The original FASTA file was filtered accordingly, yielding 138 sequences for outlier SNPs and 9,105 sequences for neutral SNPs. Repeats were screened using Rad4SnpsRepbaseFilter.py, resulting in the removal of 9 outlier SNP sequences and 1,037 neutral SNP sequences. A custom script was then used to retain only non-overlapping SNPs, followed by filtering of non-sequenceable sequences. This resulted in 63 outlier SNP sequences and 5,120 neutral SNP sequences. Neutral SNP sequences were randomly subsampled to retain one sequence per contig, resulting in 585 neutral SNP sequences. The final set of 648 sequences was then used for primer design using the Hi-Plex protocol (*128*). Designed primers were blasted against the *L. planeri* draft genome contigs to remove repeated loci and nonspecific amplifications.

Hi-Plex genotyping

A total of 1,291 individuals (536 non-parasitic *L. planeri*, 480 parasitic L. *fluviatilis* and 275 *Lampetra* larvae of unknown ecotype) were sampled in the Oir River between 2010 to 2021 by electrofishing. DNA was extracted from fin clips using the NucleoSpin® 96 Tissue kit (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. All individuals were genotyped for 148 SNPs using a modified version of the Hi-Plex protocol (*128–130*). Dual indexing was applied to each sample to allow demultiplexing. Reads were aligned to the *L. planeri* draft genome using *BWA mem*, and allele counts were obtained with *SAMtools* (*120*). For each SNP in every individual, genotypes were inferred from the observed allele counts obtained through sequencing by calculating genotype likelihoods under a multinomial probability model. The genotype exhibiting the highest posterior probability was assigned to the corresponding SNP–individual combination

only if this probability exceeded a threshold of 70%; otherwise, the genotype was treated as missing. We only retained SNPs that were genotyped in more than 75% of individuals and retained individuals with less than 25% missing data. We re-aligned RADtags around target SNPs to the kcLamFluv1.1 reference genome used in this study using *BWA mem*. We removed 'neutral SNPs' that were located on chromosomes with parasitism-QTL, even if they were not located directly within these, to remove potentially impacts on neutral population structure through linkage. This left a total of 131 SNPs, with 9 SNPs within parasitism-QTL on chr1, 21 SNPs within *trans-inv56* and 101 non-divergent SNPs outside parasitism-QTL ('neutral SNPs').

Analyses of Hi-Plex genotyping data:

To investigate the population structure of lampreys in the Oir hybrid zone, we performed principal components analyses and admixture analyses for the three different SNP subsets, SNPs on chr1, trans-inv56 and neutral SNPs. First, we performed PCA for each SNP set in *adegenet* and plotted them with *ggplot2*. For SNPs within *trans-inv56* we determined the karyotype of each individual using a k-means clustering approach based on PC1 and PC2, and estimated the karyotype frequencies for each ecotype and life stage based on these estimated karyotypes. Furthermore, to test for an excess of *trans-inv56* heterozygotes within each ecotype, we tested for deviations from Hardy-Weinberg Equilibrium within each ecotype using the *HWChisq* function in *HardyWeinberg* R-package. Furthermore, we estimated genetic ancestry proportions for each SNP dataset using *Admixture* (*131*) with a K of 2. Lastly, we used linear mixed models with *lmer* in the *lme4* R-package to test for correlations of genetic variation for each SNP subset (PC1 for chr1 and neutral SNPs, karyotype for chr56) with sperm concentration and sperm speed (VAP), which were measured for a subset of genotyped individuals (n=42) in a previous study (*59*). In brief, adult *L. fluviatilis* and *L. planeri* were caught electrofishing in the Oir, kept at the field station and anaesthetized with benzocaine before fin clipping and stripping. Fin clips were used for genotyping using the HiPlex protocol as described above. Sperm traits were measured within 5 to 30min post stripping. Sperm velocity was measured using a Computer-assisted Sperm Analysis system (*OpenCASA*) and sperm number was counted using a Thoma cell counting chamber on three subsamples per male (see (*59*) for details). We included ecotype as a random effect in the model.

RNAseq analysis

To identify differentially expressed genes between ecotypes across tissues, we performed RNA-seq on liver, gonad, gill and brain of spawning-ripe anadromous-parasitic and non-parasitic lamprey caught in the Oir. Eight parasitic *L. fluviatilis* adults (four males and four females) and eight non-parasitic *L. planeri* (four males and four females) were caught by electrofishing in the Oir River in April 2018, anesthetized, and then killed with a lethal dosage of benzocaine. We collected their liver, brain, gonad and gill (total = 32 samples) for the adults and immediately froze them in liquid nitrogen. Total RNA was extracted from brains and gonads with the NucleoSpin® RNA Plus XS extraction kit following manufacturer's instructions (Macherey-Nagel), and other organs were extracted using TRIzol Reagent and chloroform phase separation. Individual Illumina Stranded TruSeq RNA libraries were 150bp paired-end RNA sequenced on a HiSeq3000 Illumina sequencer to an average of $3x10^7$ reads per sample.

Raw RNAseq data were aligned to the kcLamFluv1.1 reference genome using a 2-pass mapping approach in *STAR* (*132*). We created read count tables against transcript annotation features for the REFseq annotation using *featurecounts* (*133*), only using uniquely mapping reads. For the differential gene expression analyses we removed female gonad samples, leaving 40 samples. We performed gene expression analyses across all individuals and tissues using *DESeq2* (*134*), retaining all genes with a minimum of 10 counts per million (~1 cpm per individual) in at least 3 individuals, and the following model: ~ sex + group; with group being a combination of ecotype and tissue. PCA was performed based on vst-transformed data in *DESeq2*. We tested for differential expression between ecotypes by tissue, identifying genes as differentially expressed with an FDR < 0.05. We identified differentially expressed genes that overlapped with genes under parasitism-QTL using *bedtoolsr* intersect (*135*).

References:

1. C. Mérot, E. L. Berdan, H. Cayuela, H. Djambazian, A.-L. Ferchaud, M. Laporte, E. Normandeau, J. Ragoussis, M. Wellenreuther, L. Bernatchez, Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly. *Mol. Biol. Evol.* **38**, 3953–3971 (2021).

2. M. Akopyan, A. Jacobs, J. A. Rick, A. P. Wilder, Z. A. Baumann, D. Conover, H. Baumann, N. O. Therkildsen, Multiple chromosomal inversions modulate continuous local adaptation along a steep thermal cline. *Science* **391**, 1015–1021 (2026).

3. D. B. Lowry, J. H. Willis, A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8** (2010).

4. P. R. Berg, B. Star, C. Pampoulie, M. Sodeland, J. M. I. Barth, H. Knutsen, K. S. Jakobsen, S. Jentoft, Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Sci. Rep.* **6**, 23246 (2016).

5. M. Matschiner, J. M. I. Barth, O. K. Tørresen, B. Star, H. T. Baalsrud, M. S. O. Brieuc, C. Pampoulie, I. Bradbury, K. S. Jakobsen, S. Jentoft, Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution*, 1–13 (2022).

6. D. E. Pearse, N. J. Barson, T. Nome, G. Gao, M. A. Campbell, A. Abadía-Cardoso, E. C. Anderson, D. E. Rundio, T. H. Williams, K. A. Naish, T. Moen, S. Liu, M. Kent, M. Moser, D. R. Minkley, E. B. Rondeau, M. S. O. Brieuc, S. R. Sandve, M. R. Miller, L. Cedillo, K. Baruch, A. G. Hernandez, G. Ben-Zvi, D. Shem-Tov, O. Barad, K. Kuzishchin, J. C. Garza, S. T. Lindley, B. F. Koop, G. H. Thorgaard, Y. Palti, S. Lien, Sex-dependent dominance maintains migration supergene in rainbow trout. *Nature Ecology & Evolution*, doi: 10.1038/s41559-019-1044-6 (2019).

7. S. Lamichhaney, G. Fan, F. Widemo, U. Gunnarsson, D. S. Thalmann, M. P. Hoeppner, S. Kerje, U. Gustafson, C. Shi, H. Zhang, W. Chen, X. Liang, L. Huang, J. Wang, E. Liang, Q. Wu, S. M.-Y. Lee, X. Xu, J. Höglund, X. Liu, L. Andersson, Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nat. Genet.* **48**, 84–88 (2016).

8. Z. Gompert, J. L. Feder, T. L. Parchman, N. P. Planidin, F. J. H. Whiting, P. Nosil, Adaptation repeatedly uses complex structural genomic variation. *Science* **388**, eadp3745 (2025).

9. H. Augustijnen, L. Bätscher, M. Cesanek, T. Chkhartishvili, V. Dincă, G. Iankoshvili, K. Ogawa, R. Vila, S. Klopfstein, J. M. de Vos, K. Lucek, A macroevolutionary role for chromosomal fusion and fission in *Erebia* butterflies. *Sci Adv* **10**, eadl0989 (2024).

10. Z. Liu, M. Roesti, D. Marques, M. Hiltbrunner, V. Saladin, C. L. Peichel, Chromosomal Fusions Facilitate Adaptation to Divergent Environments in Threespine Stickleback. *Mol. Biol. Evol.* **39** (2022).

11. S. N. K. Hoff, M. Maurstad, O. K. Tørresen, P. R. Berg, K. Præbel, K. S. Jakobsen, S. Jentoft, Chromosomal fusions and large-scale inversions are key features for adaptation in Arctic codfish species, *bioRxiv* (2024)p. 2024.06.28.599280.

12. C. Marín-García, L. Álvarez-González, L. Marín-Gual, S. Casillas, J. Picón, K. Yam, M. M. Garcias-Ramis, C. Vara, J. Ventura, A. Ruiz-Herrera, Multiple genomic landscapes of recombination and genomic divergence in wild populations of house mice-the role of chromosomal fusions and Prdm9. *Mol. Biol. Evol.* **41**, msae063 (2024).

13. N. M. Foley, R. G. Rasulis, Z. Wani, M. N. Mendoza Cerna, H. V. Figueiró, K. P. Koepfli, T. Raudsepp, W. J. Murphy, An ancient recombination desert is a speciation supergene in placental mammals. *Nature*, 1–9 (2025).

14. M. L. Zepeda Mendoza, Z. Xiong, M. Escalera-Zamudio, A. K. Runge, J. Thézé, D. Streicker, H. K. Frank, E. Loza-Rubio, S. Liu, O. A. Ryder, J. A. Samaniego Castruita, A. Katzourakis, G. Pacheco, B. Taboada, U. Löber, O. G. Pybus, Y. Li, E. Rojas-Anaya, K. Bohmann, A. Carmona Baez, C. F. Arias, S. Liu, A. D.

Greenwood, M. F. Bertelsen, N. E. White, M. Bunce, G. Zhang, T. Sicheritz-Pontén, M. P. T. Gilbert, Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat Ecol Evol* **2**, 659–668 (2018).

15. A. J. Michel, L. M. Ward, S. K. Goffredi, K. S. Dawson, D. T. Baldassarre, A. Brenner, K. M. Gotanda, J. E. McCormack, S. W. Mullin, A. O'Neill, G. S. Tender, J. A. C. Uy, K. Yu, V. J. Orphan, J. A. Chaves, The gut of the finch: uniqueness of the gut microbiome of the Galápagos vampire finch. *Microbiome* **6**, 167 (2018).

16. M. F. Docker, I. C. Potter, "Life History Evolution in Lampreys: Alternative Migratory and Feeding Types" in *Lampreys: Biology, Conservation and Control: Volume 2*, M. F. Docker, Ed. (Springer Netherlands, Dordrecht, 2019), pp. 287–409.

17. L. C. Hughes, D. D. Bloom, K. R. Piller, N. Lang, R. L. Mayden, Phylogenomic resolution of lampreys reveals the recent evolution of an ancient vertebrate lineage. *Proc. Biol. Sci.* **292** (2025).

18. C. D. Brownstein, T. J. Near, Phylogenetics and the Cenozoic radiation of lampreys. *Curr. Biol.*, doi: 10.1016/j.cub.2022.12.018 (2022).

19. S. J. Giuseppe Zanandrea, Speciation among Lampreys. *Nature* **184**, 380–380 (1959).

20. R. W. Hilliard, D. J. Bird, I. C. Potter, Metamorphic changes in the intestine of three species of lampreys. *J. Morphol.* **176**, 181–196 (1983).

21. M. Schultze, Die Entwickelungs-geschichte von Petromyzon planeri. (1856).

22. S. A. Green, B. R. Uy, M. E. Bronner, Ancient evolutionary origin of vertebrate enteric neurons from trunk-derived neural crest. *Nature* **544**, 88–91 (2017).

23. F. Lamanna, F. Hervas-Sotomayor, A. P. Oel, D. Jandzik, D. Sobrido-Cameán, G. N. Santos-Durán, M. L. Martik, J. Stundl, S. A. Green, T. Brüning, K. Mößinger, J. Schmidt, C. Schneider, M. Sepp, F. Murat, J. J. Smith, M. E. Bronner, M. C. Rodicio, A. Barreiro-Iglesias, D. M. Medeiros, D. Arendt, H. Kaessmann, A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat Ecol Evol* **7**, 1714–1728 (2023).

24. J. B. Hume, H. Recknagel, C. W. Bean, C. E. Adams, B. K. Mable, 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. *Mol. Ecol.* **27**, 4572–4590 (2018).

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

26. J. B. Hume, C. E. Adams, B. Mable, C. Bean, Post-zygotic hybrid viability in sympatric species pairs: a case study from European lampreys. *Biol. J. Linn. Soc. Lond.* **108**, 378–383 (2013).

27. Q. Rougemont, A. Gaigher, E. Lasne, J. Côte, M. Coke, A.-L. Besnard, S. Launey, G. Evanno, Low reproductive isolation and highly variable levels of gene flow reveal limited progress towards speciation between European river and brook lampreys. *J. Evol. Biol.* **28**, 2248–2263 (2015).

28. Q. Rougemont, V. Dolo, A. Oger, A.-L. Besnard, D. Huteau, M.-A. Coutellec, C. Perrier, S. Launey, G. Evanno, Riverscape genetics in brook lamprey: genetic diversity is less influenced by river fragmentation than by gene flow with the anadromous ecotype. *Heredity* , doi: 10.1038/s41437-020-00367-9 (2020).

29. L. Daupagne, C. Furusawa, H. Mieda, O. Kishida, E. Lasne, É. Tentelier, I. Koizumi, Form-assortative mating behaviors of individuals from parasitic and non-parasitic populations of Arctic lamprey (*Lethenteron camtschaticum*). *Behav. Ecol.*, arae073 (2024).

30. S. H. Martin, S. M. Van Belleghem, Exploring Evolutionary Relationships Across the Genome Using

Topology Weighting. *Genetics* **206**, 429–438 (2017).

31. S. Stankowski, Z. B. Zagrodzka, M. D. Garlovsky, A. Pal, D. Shipilina, D. G. Castillo, H. Lifchitz, A. Le Moan, E. Leder, J. Reeve, K. Johannesson, A. M. Westram, R. K. Butlin, The genetic basis of a recent transition to live-bearing in marine snails. *Science* **383**, 114–119 (2024).

32. M. Bohutínská, C. L. Peichel, Divergence time shapes gene reuse during repeated adaptation. *Trends Ecol. Evol.*, doi: 10.1016/j.tree.2023.11.007 (2023).

33. G. S. Silver, R. T. Lampman, N. Percival, N. Timoshevskaya, J. J. Smith, K. T. Bentley, J. Wade, S. R. Narum, J. E. Hess, Genetic identification of lamprey genera and anadromous ecotypes in watersheds of the Northeastern Pacific ocean. *Evol. Appl.* **18**, e70108 (2025).

34. T. M. Evans, P. Janvier, M. F. Docker, The evolution of lamprey (Petromyzontida) life history and the origin of metamorphosis. *Rev. Fish Biol. Fish.* **28**, 825–838 (2018).

35. M. Freamat, S. A. Sower, A sea lamprey glycoprotein hormone receptor similar with gnathostome thyrotropin hormone receptor. *J. Mol. Endocrinol.* **41**, 219–228 (2008).

36. N. Ajmani, T. Yasmin, M. F. Docker, S. V. Good, Transcriptomic analysis of gonadal development in parasitic and non-parasitic lampreys (Ichthyomyzon spp.), with a comparison of genomic resources in these non-model species. *G3* **11** (2021).

37. C. Zhao, Y. Zhai, R. Geng, K. Wu, W. Song, N. Ai, W. Ge, Genetic analysis of activin/inhibin β subunits in zebrafish development and reproduction. *PLoS Genet.* **18**, e1010523 (2022).

38. L. Wen, T. Hasebe, T. C. Miller, A. Ishizuya-Oka, Y.-B. Shi, A requirement for hedgehog signaling in thyroid hormone-induced postembryonic intestinal remodeling. *Cell Biosci.* **5**, 13 (2015).

39. Y. Chung-Davidson, J. Ren, C. Yeh, U. Bussy, B. Huerta, P. J. Davidson, S. Whyard, W. Li, TGF-β Signaling Plays a Pivotal Role During Developmental Biliary Atresia in Sea Lamprey (Petromyzon marinus). *Hepatol Commun* **4**, 219–234 (2020).

40. J. Hildahl, M. Galay-Burgos, G. Sweeney, I. E. Einarsdóttir, B. T. Björnsson, Identification of two isoforms of Atlantic halibut insulin-like growth factor-I receptor genes and quantitative gene expression during metamorphosis. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **147**, 395–401 (2007).

41. M. A. Walkiewicz, M. Stern, Increased insulin/insulin growth factor signaling advances the onset of metamorphosis in Drosophila. *PLoS One* **4**, e5072 (2009).

42. R. G. Manzon, L. A. Manzon, Lamprey metamorphosis: Thyroid hormone signaling in a basal vertebrate. *Mol. Cell. Endocrinol.* **459**, 28–42 (2017).

43. K. R. Paton, M. H. Cake, I. C. Potter, Lipid and protein catabolism contribute to aerobic metabolic responses to exhaustive exercise during the protracted spawning run of the lamprey Geotria australis. *J. Comp. Physiol. B* **190**, 35–47 (2020).

44. S. Tamrakar, B. Huerta, Y.-W. Chung-Davidson, W. Li, Liver metabolomic profiles of sea lamprey (Petromyzon marinus) are influenced by sex and maturation stages. *Metabolomics* **21**, 69 (2025).

45. C. A. Salas, K. E. Yopak, R. E. Warrington, N. S. Hart, I. C. Potter, S. P. Collin, Ontogenetic shifts in brain scaling reflect behavioral changes in the life cycle of the pouched lamprey Geotria australis. *Front. Neurosci.* **9**, 251 (2015).

46. A. Jacobs, K. R. Elmer, Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol. Ecol.* **30**, 4955–4969 (2021).

47. J.-P. Verta, A. Jacobs, The role of alternative splicing in adaptation and evolution. *Trends Ecol. Evol.*, doi: 10.1016/j.tree.2021.11.010 (2021).

48. A. Jacobs, J. P. Velotta, A. Tigano, A. P. Wilder, H. Baumann, N. O. Therkildsen, Temperature-dependent gene regulatory divergence underlies local adaptation with gene flow in the Atlantic silverside. *Evolution* **78**, 1133–1149 (2024).

49. O. K. Tørresen, B. Garmann-Aarhus, S. Nam Khang Hoff, S. Jentoft, M. Svensson, E. Schartum, A. Tooming-Klunderud, M. Skage, A. Krabberød, L. Asbjørn Vøllestad, K. S. Jakobsen, Comparison of whole-genome assemblies of European river lamprey (Lampetra fluviatilis) and brook lamprey (Lampetra planeri). *J. Hered.*, esaf036 (2025).

50. T. Zhu, Y. Li, Y. Pang, Y. Han, J. Li, Z. Wang, X. Liu, H. Li, Y. Hua, H. Jiang, H. Teng, J. Quan, Y. Liu, M. Geng, M. Li, F. Hui, J. Liu, Q. Qiu, Q. Li, Y. Ren, Chromosome-level genome assembly of Lethenteron reissneri provides insights into lamprey evolution. *Mol. Ecol. Resour.*, doi: 10.1111/1755-0998.13279 (2020).

51. N. Timoshevskaya, K. I. Eşkut, V. A. Timoshevskiy, S. M. C. Robb, C. Holt, J. E. Hess, H. J. Parker, C. F. Baker, A. K. Miller, C. Saraceno, M. Yandell, R. Krumlauf, S. R. Narum, R. T. Lampman, N. J. Gemmell, J. Mountcastle, B. Haase, J. R. Balacco, G. Formenti, S. Pelan, Y. Sims, K. Howe, O. Fedrigo, E. D. Jarvis, J. J. Smith, An improved germline genome assembly for the sea lamprey Petromyzon marinus illuminates the evolution of germline-specific chromosomes. *Cell Rep.* **42**, 112263 (2023).

52. C. Vara, A. Paytuví-Gallart, Y. Cuartero, L. Álvarez-González, L. Marín-Gual, F. Garcia, B. Florit-Sabater, L. Capilla, R. A. Sanchéz-Guillén, Z. Sarrate, R. Aiese Cigliano, W. Sanseverino, J. B. Searle, J. Ventura, M. A. Marti-Renom, F. Le Dily, A. Ruiz-Herrera, The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. *Nat. Commun.* **12**, 2981 (2021).

53. F. Cicconardi, J. J. Lewis, S. H. Martin, R. D. Reed, C. G. Danko, S. H. Montgomery, Chromosome fusion affects genetic diversity and evolutionary turnover of functional loci but consistently depends on chromosome size. *Mol. Biol. Evol.* **38**, 4449–4462 (2021).

54. E. Y. Roback, E. Ferrufino, R. L. Moran, D. Shennard, C. Mulliniks, J. Gallop, J. Weagley, J. Miller, Y. Fily, C. P. Ornelas-García, N. Rohner, J. E. Kowalko, S. E. McGaugh, Population genomics of premature termination codons in cavefish with substantial trait loss. *Mol. Biol. Evol.*, msaf012 (2025).

55. D. H. Drabeck, J. Wiese, E. Gilbertson, J. Arroyave, M. L. J. Stiassny, S. E. Alter, R. Borowsky, D. A. Hendrickson, D. Arcila, S. E. McGaugh, Gene loss and relaxed selection of plaat1 in vertebrates adapted to low-light environments. *Proc. Biol. Sci.* **291**, 20232847 (2024).

56. N. Grankvist, R. E. Honkanen, Å. Sjöholm, H. Ortsäter, Genetic disruption of protein phosphatase 5 in mice prevents high-fat diet feeding-induced weight gain. *FEBS Lett.* **587**, 3869–3874 (2013).

57. H. C. Justen, S. A. Blain, K. E. Delmore, The genetics of extrinsic postzygotic selection in a migratory divide between subspecies of the Swainson's thrush. *Nat. Commun.* **16**, 1–11 (2025).

58. T. M. Evans, K. E. Limburg, Parasitism offers large rewards but carries high risks: Predicting parasitic strategies under different life history conditions in lampreys. *J. Evol. Biol.* **32**, 794–805 (2019).

59. N. Decanter, R. Normand, A. Souissi, C. Labbé, E. Edeline, G. Evanno, Sperm competition experiments reveal low prezygotic postmating isolation between parasitic and nonparasitic lamprey ecotypes. *Ecol. Evol.* **13**, e9970 (2023).

60. P. A. Moran, T. J. Colgan, K. P. Phillips, J. Coughlan, P. McGinnity, T. E. Reed, Whole-Genome Resequencing Reveals Polygenic Signatures of Directional and Balancing Selection on Alternative Migratory Life Histories. *Mol. Ecol.*, e17538 (2024).

61. S. Lai, Y. Huang, S. Ma, H. Hao, A. Dai, X. Yan, J. Yang, S. Wang, Q. Ren, Y. Zhang, P. Hu, J. Li, X. Zheng, J. Brosius, C. Deng, FSHR and LHR functional compensation reveals the mechanism and treatment of Ovarian Hyperstimulation Syndrome. *Nat. Commun.*, 1–18 (2026).

62. E. Andersson, R. W. Schulz, F. Almeida, L. Kleppe, K. O. Skaftnesmo, E. Kjærner-Semb, D. Crespo, P. G.

Fjelldal, T. J. Hansen, B. Norberg, R. B. Edvardsen, A. Wargelius, Loss of fshr prevents testicular maturation in Atlantic salmon (Salmo salar L.). *Endocrinology* **165**, bqae013 (2024).

63. N. Murozumi, R. Nakashima, T. Hirai, Y. Kamei, T. Ishikawa-Fujiwara, T. Todo, T. Kitano, Loss of follicle-stimulating hormone receptor function causes masculinization and suppression of ovarian development in genetically female medaka. *Endocrinology* **155**, 3136–3145 (2014).

64. G. P. Kordes, A. S. Busch, Genetic epistasis in FSH action influences female reproductive lifespan. *Reproduction* **171**, xaag008 (2026).

65. C. P. Hagen, K. Sørensen, L. Aksglaede, A. Mouritsen, M. G. Mieritz, J. Tinggaard, C. Wohlfart-Veje, J. H. Petersen, K. M. Main, E. Rajpert-De Meyts, K. Almstrup, A. Juul, Pubertal onset in girls is strongly influenced by genetic variation affecting FSH action. *Sci. Rep.* **4**, 6412 (2014).

66. N. J. Barson, T. Aykanat, K. Hindar, M. Baranski, G. H. Bolstad, P. Fiske, C. Jacq, A. J. Jensen, S. E. Johnston, S. Karlsson, M. Kent, T. Moen, E. Niemelä, T. Nome, T. F. Næsje, P. Orell, A. Romakkaniemi, H. Sægrov, K. Urdal, J. Erkinaro, S. Lien, C. R. Primmer, Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**, 405–408 (2015).

67. E. P. Ahi, M. Sinclair-Waters, J. Moustakas-Verho, S. Jansouz, C. R. Primmer, Strong regulatory effects of vgll3 genotype on reproductive axis gene expression in juvenile male Atlantic salmon. *Gen. Comp. Endocrinol.* **325**, 114055 (2022).

68. S. N. K. Hoff, M. F. Maurstad, O. K. Tørresen, R. A. Araya, P. R. Berg, K. Præbel, K. S. Jakobsen, S. Jentoft, Rapid genome modifications including chromosomal fusions and large-scale inversions are key features in Arctic codfish species. *Genome Biol.*, 1–41 (2026).

69. S. Wigby, T. Chapman, Sperm competition. *Curr. Biol.* **14**, R100–2 (2004).

70. M. J. G. Gage, C. P. Macfarlane, S. Yeates, R. G. Ward, J. B. Searle, G. A. Parker, Spermatozoal traits and sperm competition in Atlantic salmon. *Curr. Biol.* **14**, 44–47 (2004).

71. H. S. Fisher, E. Jacobs-Palmer, J.-M. Lassance, H. E. Hoekstra, The genetic basis and fitness consequences of sperm midpiece size in deer mice. *Nat. Commun.* **7**, 13652 (2016).

72. K.-W. Kim, C. Bennison, N. Hemmings, L. Brookes, L. L. Hurley, S. C. Griffith, T. Burke, T. R. Birkhead, J. Slate, A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nat. Ecol. Evol.* **1**, 1168–1176 (2017).

73. E. L. Berdan, N. H. Barton, R. Butlin, B. Charlesworth, R. Faria, I. Fragata, K. J. Gilbert, P. Jay, M. Kapun, K. E. Lotterhos, C. Mérot, E. Durmaz Mitchell, M. Pascual, C. L. Peichel, M. Rafajlović, A. M. Westram, S. W. Schaeffer, K. Johannesson, T. Flatt, How chromosomal inversions reorient the evolutionary process. *J. Evol. Biol.*, doi: 10.1111/jeb.14242 (2023).

74. P. Nosil, V. Soria-Carrasco, R. Villoutreix, M. De-la-Mora, C. F. de Carvalho, T. Parchman, J. L. Feder, Z. Gompert, Complex evolutionary processes maintain an ancient chromosomal inversion. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2300673120 (2023).

75. C. Mérot, V. Llaurens, E. Normandeau, L. Bernatchez, M. Wellenreuther, Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nat. Commun.* **11**, 1–11 (2020).

76. M. D. Shapiro, M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, D. M. Kingsley, Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).

77. C. L. Peichel, K. S. Nereng, K. A. Ohgi, B. L. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, D. M. Kingsley, The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).

78. P. F. Colosimo, K. E. Hosemann, S. Balabhadra, G. Villarreal Jr, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, D. M. Kingsley, Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).

79. R. D. Reed, R. Papa, A. Martin, H. M. Hines, B. A. Counterman, C. Pardo-Diaz, C. D. Jiggins, N. L. Chamberlain, M. R. Kronforst, R. Chen, G. Halder, H. F. Nijhout, W. O. McMillan, optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141 (2011).

80. K. H. Morris, P. S. Maitland, A trap for catching adult lampreys (Petromyzonidae) in running water. *J. Fish Biol.* **31**, 513–516 (1987).

81. P. S. Maitland, K. H. Morris, K. East, The ecology of lampreys (Petromyzonidae) in the Loch Lomond area. *Hydrobiologia* **290**, 105–120 (1994).

82. F. Rendell-Bhatti, C. Bull, R. Cross, R. Cox, G. A. Adediran, E. Lahive, From the environment into the biomass: microplastic uptake in a protected lamprey species. *Environ. Pollut.* **323**, 121267 (2023).

83. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

84. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

85. A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, P. Prins, Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

86. B. S. Pedersen, A. R. Quinlan, Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

87. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

88. R. N. Lou, A. Jacobs, A. P. Wilder, N. O. Therkildsen, A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol. Ecol.* **30**, 5966–5993 (2021).

89. X. Wang, C.-E. T. Pedersen, G. Athanasiadis, G. Garcia-Erill, K. Hanghøj, L. D. Bertola, M. S. Rasmussen, M. Schubert, X. Liu, Z. Li, L. Lin, R. F. Balboa, E. Jørsboe, C. Nursyifa, S. Liu, V. Muwanika, C. Masembe, L. Chen, W. Wang, I. Moltke, H. R. Siegismund, A. Albrechtsen, R. Heller, Persistent gene flow suggests an absence of reproductive isolation in an African antelope speciation model. *Syst. Biol.* **73**, 979–994 (2024).

90. C. Pockrandt, M. Alzamel, C. S. Iliopoulos, K. Reinert, GenMap: ultra-fast computation of genome mappability. *Bioinformatics* **36**, 3687–3692 (2020).

91. J. Meisner, A. Albrechtsen, Testing for Hardy-Weinberg Equilibrium in Structured Populations using Genotype or Low-Depth NGS Data. *Mol. Ecol. Resour.*, doi: 10.1111/1755-0998.13019 (2019).

92. G. Garcia-Erill, C. H. F. Jørgensen, V. B. Muwanika, X. Wang, M. S. Rasmussen, Y. A. de Jong, P. Gaubert, A. Olayemi, J. Salmona, T. M. Butynski, L. D. Bertola, H. R. Siegismund, A. Albrechtsen, R. Heller, Warthog genomes resolve an evolutionary conundrum and reveal introgression of disease resistance genes. *Mol. Biol. Evol.* **39**, msac134 (2022).

93. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing, *arXiv [q-bio.GN]* (2012). http://arxiv.org/abs/1207.3907.

94. J. W. Leigh, D. Bryant, Popart: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

95. G. Garcia-Erill, A. Albrechtsen, Evaluation of model fit of inferred admixture proportions. *Mol. Ecol. Resour.* **20**, 936–949 (2020).

96. M. Milanesi, S. Capomaccio, E. Vajana, L. Bomba, J. F. Garcia, P. Ajmone-Marsan, L. Colli, BITE: an R package for biodiversity analyses, *Bioinformatics* (2017). https://www.biorxiv.org/content/10.1101/181610v1.full.pdf.

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

98. R. R. Fitak, OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc* **6**, bpab017 (2021).

99. E. A. Fox, A. E. Wright, M. Fumagalli, F. G. Vieira, ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**, 3855–3856 (2019).

100. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

101. S. Xu, E. Hu, Y. Cai, Z. Xie, X. Luo, L. Zhan, W. Tang, Q. Wang, B. Liu, R. Wang, W. Xie, T. Wu, L. Xie, G. Yu, Using clusterProfiler to characterize multiomics data. *Nat. Protoc.* **19**, 3292–3320 (2024).

102. Z. Gu, D. Hübschmann, Simplify enrichment: A bioconductor package for clustering and visualizing functional enrichment results. *Genomics Proteomics Bioinformatics*, doi: 10.1016/j.gpb.2022.04.008 (2022).

103. M. Shpak, K. N. Lawrence, J. E. Pool, The precision and power of population branch statistics in identifying the genomic signatures of local adaptation. *Genome Biol. Evol.* **17**, evaf080 (2025).

104. M. Alonge, L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, S. Aganezov, X. Wang, Z. B. Lippman, M. C. Schatz, S. Soyk, Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).

105. K.-H. Chao, J. M. Heinz, C. Hoh, A. Mao, A. Shumate, M. Pertea, S. L. Salzberg, Combining DNA and protein alignments to improve genome annotation with LiftOn. *Genome Res.* **35**, 311–325 (2025).

106. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*, doi: 10.1038/s41592-020-01056-5 (2021).

107. C. Zhou, S. A. McCarthy, R. Durbin, YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39** (2023).

108. K. Howe, W. Chow, J. Collins, S. Pelan, D.-L. Pointon, Y. Sims, J. Torrance, A. Tracey, J. Wood, Significantly improving the quality of genome assemblies through curation. *Gigascience* **10** (2021).

109. L. Coombe, P. Kazemi, J. Wong, I. Birol, R. L. Warren, ntSynt: multi-genome synteny detection using minimizer graph mappings. *BMC Biol.* **23**, 367 (2025).

110. L. Coombe, R. L. Warren, I. Birol, ntSynt-viz: Visualizing synteny patterns across multiple genomes, *bioRxiv* (2025)p. 2025.01.15.633221.

111. M. Goel, H. Sun, W.-B. Jiao, K. Schneeberger, SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).

112. H. Li, Minimap2: pairwise alignment for nucleotide sequences, *arXiv [q-bio.GN]* (2017)pp. 3094–3100.

113. M. Goel, K. Schneeberger, Plotsr: Visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022).

114. J. Wilson, V. C. Bieker, L. van Boheemen, T. Connallon, M. D. Martin, P. Battlay, K. A. Hodgins, Copy number variation contributes to parallel local adaptation in an invasive plant. *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2413587122 (2025).

115. M. C. Whitlock, Evolutionary inference from QST. *Mol. Ecol.* **17**, 1885–1896 (2008).

116. J. R. Adrion, J. G. Galloway, A. D. Kern, Predicting the landscape of recombination using deep learning. *Mol. Biol. Evol.*, doi: 10.1093/molbev/msaa038 (2020).

117. M. W. Hahn, S. R. Mishra, Estimating recombination using only the allele frequency spectrum. *Genetics*, iyaf108 (2025).

118. J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko, Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).

119. N. C. Rochette, A. G. Rivera-Colón, J. M. Catchen, Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **28**, 4737–4754 (2019).

120. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

121. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

122. P. Rastas, Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33**, 3726–3732 (2017).

123. E. S. Robinson, I. C. Potter, N. B. Atkin, The nuclear dna content of lampreys. *Experientia* **31**, 912–913 (1975).

124. F. Guillaume, J. Rougemont, Nemo: an evolutionary and population genetics programming framework. *Bioinformatics* **22**, 2556–2557 (2006).

125. N. Decanter, “Genomic divergence and speciation between the parasitic river lamprey (Lampetra fluviatilis) and non-parasitic brook lamprey (L. planeri),” thesis, Institut national d’enseignement supérieur pour l’agriculture, l’alimentation et l’environnement, Rennes (2023).

126. H. E. L. Lischer, L. Excoffier, PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012).

127. K. Keenan, P. McGinnity, T. F. Cross, W. W. Crozier, P. A. Prodöhl, diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**, 782–788 (2013).

128. F. Hammet, K. Mahmood, T. R. Green, T. Nguyen-Dumont, M. C. Southey, D. D. Buchanan, A. Lonie, K. L. Nathanson, F. J. Couch, B. J. Pope, D. J. Park, Hi-Plex2: a simple and robust approach to targeted sequencing-based genetic screening. *Biotechniques* **67**, 118–122 (2019).

129. A.-L. Besnard, D. J. Park, B. J. Pope, F. Hammet, S. Michon-Coudouel, M. Biget, S. A. Krueger-Hadfield, S. Mauger, E. J. Petit, Workflow for SNP genotyping using the Hi-Plex method. doi: 10.17504/protocols.io.8epv5jnnnl1b/v2 (2023).

130. M. H. Meek, W. A. Larson, The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Mol. Ecol. Resour.* **19**, 795–803 (2019).

131. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

132. A. Dobin, T. R. Gingeras, Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11.14.1–19 (2015).

133. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

134. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

135. M. N. Patwardhan, C. D. Wenger, E. S. Davis, D. H. Phanstiel, Bedtoolsr: An R package for genomic data analysis and manipulation. *J. Open Source Softw.* **4**, 1742 (2019).