Convergence in form and function overcomes non-parallel evolutionary histories in Arctic charr

Arne Jacobs[1], Madeleine Carruthers[1], Andrey Yurchenko[1], Natalia V. Gordeeva[2], Sergei S. Alekseyev[3], Oliver Hooker[1,4,†], Jong S. Leong[5], Eric B. Rondeau[5], Ben F. Koop[5], Colin E. Adams[1,4] & Kathryn R. Elmer[1*]

[1]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, G12 8QQ, Glasgow, UK

[2]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 199333 Russia

[3]Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, 119991 Russia

[4] Scottish Centre for Ecology and the Natural Environment, University of Glasgow, Rowardennan, Loch Lomond, Glasgow G63 0AW, UK

[5]Biology/Centre for Biomedical Research, University of Victoria, British Columbia, Canada

[†]current address: PR statistics, 6 Hope Park Crescent, Edinburgh, EH8 9NA, UK

*Correspondence to: kathryn.elmer@glasgow.ac.uk, Tel: +44 141 330 6617

Abstract:

Understanding the extent to which evolution is predictable is fundamental to advancing research on biodiversity origins. We studied the evolution of ecotypes in Arctic charr in 18 freshwater lakes and two evolutionary lineages. We find significant phenotypic parallelism across replicates, with the degree of parallelism most pronounced in adaptive traits associated with foraging. Genome-wide analyses of the underlying population variation revealed a strong influence of historical demography and gene flow. Some divergent genomic regions were significantly shared across ecotype replicates, though there was a high proportion of population-specific variation. Functional molecular bases to ecotypes in gene expression and biological pathways were extensively and significantly shared across ecotype pairs. These results demonstrate a spectrum of predictability, from low parallelism at the nucleotide level to high parallelism in regulatory molecular bases and functional adaptive aspects of morphology, and advance our understanding of parallel evolution through divergent evolutionary routes.

Whether the evolution of biodiversity is predictable remains one of the greatest controversies in evolutionary biology (*1,2*). Numerous species in nature have repeatedly evolved similar phenotypes in response to similar environmental challenges, strongly suggesting repeatability and predictability amongst evolutionary trajectories (*3-8*) and highlighting the role of natural selection in evolution (*9,10*). Despite this, extensive variation in the magnitude and orientation of evolutionary trajectories exists among some classic 'parallel' populations, suggesting some historical contingency and contemporary stochasticity (*6, 11-12*). Stochastic factors such as differences in the local environment, gene flow, and selection, or contingencies such as genomic background, demographic history, and the genetic architecture of adaptive traits, can lead to departure from phenotypic parallelism and unpredictable patterns at the genomic level (*6,8,13,14*). In order to improve our ability to predict evolutionary outcomes and thus how biodiversity emerges, it is critical to understand the repeatability of evolutionary routes leading to parallel outcomes, particularly where these occur under different evolutionary and genomic backgrounds.

The replicated ecological and morphological post-glacial diversification of fishes into distinct trophic types in freshwater lakes provides an ideal natural experiment for testing these questions (*15-17*). In the Holarctic salmonid species Arctic charr (*Salvelinus alpinus*), sympatric ecotype pairs that differ in several heritable phenotypic traits, such as body shape, body size, trophic morphology, and life history, are abundantly replicated (*16-20*). Common ecotypes distinguished by diet and foraging tactics include: Planktivores feed in the pelagic, Benthivores feed in the littoral zone, and Piscivores feed in open water (*16-19*). Here we studied the extent of parallelism in phenotype, evolutionary history, genomic basis, and transcriptomic expression in repeated divergences of Arctic charr in environmental context from 18 lakes within and across two evolutionary lineages (Atlantic and Siberian). Arctic charr provides an excellent system for studying the interplay of contingency and predictability in parallel evolution within and across a range of evolutionary hierarchies.

To test for parallelism in ecologically relevant phenotypes, we assessed morphology within and across lineages based on traits from 1529 individuals from 17 lake populations and 13 replicate ecotype pairs (fig. S1, table S1). Based on linear measures and body shape (fig. S2), we found differences in the extent of phenotypic differentiation and shape across populations and lineages (Fig. 1A and 1C; figs. S4 to S8; tables S2 and S5). The parallelism in morphology was stronger in Siberian ($\eta^2_{\text{Ecotype, Siberian}} = 0.282$) compared to Atlantic populations ($\eta^2_{\text{Ecotype, Atlantic}} = 0.00163$) (table S3, figs. S5 and S6). The unique evolutionary history of populations explained the highest proportion of variance in both lineages, suggesting a strong influence of evolutionary contingency (table S3). Trajectory analysis of body shape in the four sympatric ecotype pairs from the Atlantic lineage (N = 347) revealed variable levels of parallelism in overall body shape morphology, with some pairs exhibiting parallelism and others not (Fig. 1B, table S7; see *21*). However, the main axis of phenotypic variation (PC1) for all fish was associated with head and eye size (Fig. 1B), and those morphological components were highly parallel across all ecotype pairs (table S6: $\eta^2_{\text{Ecotype, Atlantic, PC1}} = 0.49$, $\eta^2_{\text{Lake, Atlantic, PC1}} = 0.39$; $\eta^2_{\text{Ecotype x Lake (ExL), Atlantic, PC1}} = 0.26$). Indeed, individual traits that are functionally related to foraging showed the highest levels of parallelism, with the strongest parallelism across benthivorous-planktivorous ecotype pairs being in eye diameter, with the parallel ecotype effect explaining most of the phenotypic variation (Fig. 1D: $\eta^2_{\text{ED,bn-pl}} = 0.323$, $p < 0.001$) (table S4) even across Atlantic-Siberian lineages. Across planktivorous-piscivorous ecotype pairs, head depth showed the highest degree of parallelism (Fig. 1D: $\eta^2_{\text{HDE,pl-pisc}} = 0.212$, $p < 0.001$) (table S4). All other assessed linear head

traits, except maxillary length in piscivorous ecotypes, were significantly parallel across lakes though also strongly influenced by unique evolutionary histories of populations ($\eta^2_{Ecotype}<\eta^2_{Lake}$) (table S4). Thus, despite a considerable level of non-parallelism in morphology within and across lineages, there is significant parallelism amongst components of fish morphology that have functional significance for foraging efficiency (head shape, eye diameter) (*22*) across replicates.

To determine if parallel ecotypes have evolved via similar population genetic processes and under contemporary environmental influences, we analyzed a set of up to 12,215 single-nucleotide polymorphisms (SNPs) in 648 individuals from 18 populations. The degree and pattern of genetic differentiation (Fst) differed extensively within and among lakes, forming a continuum (fig. S9, table S8) that was not strictly associated with the geography of populations, and suggested important differences in evolutionary history and reproductive isolation among populations (*22*). For example, variation in genetic differentiation among ecotypes within lakes was positively correlated with the mean degree of phenotypic differentiation (fig. S10, table S9: Pearson's $r_{outlierSNPs} = 0.695$, P = 0.038; Pearson's $r_{neutralSNPs} = 0.648$, P = 0.062), consistent with isolation-by-adaptation (*12*,*23*). We predicted that some of this variability in genetic and phenotypic divergence may be due to local environmental differences (9), as has been found to be important in charr *(24)*. Indeed, the level of genome-wide genetic diversity across lakes was mainly explained by ecosystem size (Fig. 1B; Pearson's r = - 0.75, P< 0.001), suggesting an important role of ecological opportunity (*24*). Further, the magnitude of divergence in phenotypic traits (Pst) between sympatric morphs was positively correlated with ecosystem size (fig. S10, table S9; Pearsons correlation, Pst_{mean}: r=0.61, P=0.078; Pst_{HL}: r = 0.82, P = 0.0098). These patterns strongly suggest a key role of local environmental opportunity in the extent of the phenotypic divergence in Arctic charr.

To understand if replicate populations in these environments have evolved by parallel evolutionary routes, we assessed demographic and evolutionary genetic histories using genome-wide SNPs and mitochondrial haplotypes. Population genetic ancestry and phylogenetic patterns (Fig. 2; figs. S9, and S11 to S15) revealed strong variation in the degree of differentiation and admixture among ecotypes and populations. Most populations form genetically distinct clusters (figs. S12 to S14) but varying levels of shared ancestry across populations and ecotypes indicate histories of differential secondary gene flow or ancestral population structuring (Figs. 2D and 2E, figs. S13 and S14). In two cases populations formed polyphyletic genetic clusters, i.e. included both sympatric ecotypes and a nearby generalist lake population (Dug-Uai and Kir3-Kir4), which suggested strong secondary gene flow or shared ancestry of similar ecotypes (Fig. 2A to 2C). Discordance in population structure between mitochondrial and nuclear DNA further supports a complex history of divergence and admixture, with near complete haplotype sharing across ecotypes (fig. S16). Formal analysis of introgression and secondary gene flow using *f3*- and *f4*-statistics found evidence for extensive secondary gene flow among most populations in both lineages (Fig. 2B and C; tables S10 and S11), and evidence for introgressive hybridization in the planktivorous ecotype in Dughaill and the small-piscivorous ecotype in Kalarskii Davatchan (Fig. 2B and 3C, table S11). Coalescence simulations implemented in *fastsimcoal2* (*25)* suggested that most ecotype pairs likely evolved via secondary contact that admixed ancestral lineages. Considerable variation exists in inferred divergence times (910 – 16,690 generations) and admixture proportions (0.003 – 1.0) across lakes (figs. S17 and S18; table S12). Models suggest that during the last glacial maximum (LGM) ancestral populations in Siberia were strongly isolated (fig. S18) while in the Atlantic lineage weak gene flow was detected during the LGM (fig. S18), adding to the evolutionary contingency across lineages. In two

Scottish populations with low contemporary genetic differentiation between ecotypes (Fst = 0.002 and 0.001; Fig. 2D, table S8), we found evidence for sympatric divergence of ecotypes after the LGM (isolation-with-migration model; $T_{DIV,\ Awe}$ = 1245 generations, $T_{DIV,\ naS}$ = 4795) (fig. S18). However, in most cases, models including recent introgression are better supported than those without (figs. S17 and S18; table S12), suggesting a pervasive effect of secondary contact and introgression in sympatric ecotypes, similar to other polymorphic postglacial fish species (*19,26*). Subsequent to complex histories, genetic and ecological data suggest most ecotypes likely diverged in sympatry (*27*), facilitated by genetic variation that arose in allopatry across ancestral populations. Thus, findings from multiple approaches suggest that despite replicated parallel evolution of phenotype within and across lineages, there is a high level of variation across populations and lineages in evolutionary genetic and demographic histories.

To test the extent to which evolution overcomes complex differing genomic histories resulting in parallelism, we compared genome-wide patterns of differentiation across ecotype pairs. We found one large genomic region of 4.2 Mb (Contig985 on LG1) that exhibited extreme differentiation (top 5% Fst outlier-SNPs) in five out of six benthivorous-planktivorous ecotype pairs and one piscivorous-planktivorous pair (fig. S19), strongly suggesting that this region plays a potentially important role in benthivorous-planktivorous divergence. However, in most cases replicate ecotype pairs and lineages shared fewer outlier genomic regions and loci under selection than expected by chance, when comparing particular outlier SNPs or their surrounding regions (contigs) (Fig. 3A and 3B, figs. S19 and S22). Patterns of genome-wide differentiation and diversity were highly variable across populations, strongly influenced by the underlying evolutionary history (figs. S19 and S21, table S8); populations for which we had inferred histories of secondary contact had few outliers and symmetrical values of ZFst relative to $\Delta\pi$, while populations with ecotypes that diverged recently under gene flow without ancestral secondary contact (by IM) showed reduced π in one compared to other ecotype in differentiated loci (fig. S21), typical of response to selection. Thus, despite evidence for parallel divergence in a small number of genomic regions across two or more ecotypes, most genomic response to selection is unique and contingent on demographic history.

To identify parallelism in the functional genomic basis underlying these ecotypes, we compared genome-wide association results between lineages. Loci associated with ecotype were widely distributed across the genome and not shared across lineages (fig. S22). Although candidate loci from association analyses were non-parallel across the evolutionary lineages, we found that the allele frequencies were to some extent parallel across replicate ecotypes within lineages (fig. S23; table S13; Scotland: $\eta^2_{ecotype,\ PC4} = 0.70$, $\eta^2_{ExL,\ PC4} = 0.30$; Siberia: $\eta^2_{ecotype,\ PC1} = 0.943$, $\eta^2_{ExL,\ PC1} = 0.508$). Similar to the phenotypic analysis, GWAS candidate loci showed a generally higher level of parallelism within the Siberian lineage. These findings demonstrate that the genetic basis of ecotype differs across lineages in the set of loci we interrogated.

Given the non-parallelism in genomic patterns but parallelism in phenotype replicates, we hypothesized that regulatory variation would be parallel among ecotypes. We analyzed genome-wide gene expression for 44 white-muscle transcriptomes between ecotypes from a subset of five lakes (table S14) and compared differential expression, co-expression modules, and biological pathways. Similar to population genetic patterns, we found a continuum of differentiation in gene expression across ecotype pairs (fig. S24). Contrary to the genomic analysis, differentially expressed genes (DEGs) were shared between parallel ecotype pairs more often than expected by chance (Fig. 3D; table S14) and modules of co-expressed genes (fig. S25; table S16) were correlated with ecotype across populations and lineages. Analyses of biological pathways found

three characterized genes that were consistently differentially expressed in five out of seven ecotype pairs: *ABCC8*, *NDPK,* and *ALDOA* (Fig. 3C, table S15 and S18). These genes are biologically relevant because of their association with growth rate, body size, and metabolism (*28-30*), phenotypes that differ significantly between ecotypes (*16-20*), including in our morphology analyses (fig. S4). Other biologically relevant DEGs were shared between up to four pairs (19 genes, table S18), such as hemoglobins *HBA-4* and *HBB-1* (fig. 26) which might facilitate different swimming performance of ecotypes, and *RERG*, a growth inhibitor (Fig. 3E). We found *RERG* expression is correlated to the mean body size of ecotypes across all our populations, an important trait that differentiates ecotypes in many populations (Fig. 3E; fig. S4). DEGs were enriched for 15 functional biological pathways and some were shared between multiple populations (table S19), such as oxidative phosphorylation (Awe, Tay, Kam bn-pisc; dre00190, FDR < 0.1), ribosome biogenesis (Awe, Tok; dre03008, FDR < 0.1), and ECM-receptor interaction (Awe, Tok; dre04512, FDR < 0.1). An additional seven pathways (in four modules) were significantly enriched across co-expression modules (table S20), of which two were shared with DE gene sets: ECM-receptor interaction, dre04512 and glycerolipid metabolism, dre00561 (tables S17 and S19). Genes that we had found to be associated with ecotype divergence or to be under selection at the genome level, were also involved in some these pathways, such as the ECM-receptor interaction or ABC transporter pathways (tables S20 and S21) in the Siberian or Atlantic populations, suggesting that gene expression differences have at least partially an underlying genetic basis. Genes in highly differentiated regions (Contig 985), and regions under selection and associated with ecotype in both lineages (Contig 1512), were also involved in differentially and co-expressed pathways (e.g. fructose and mannose metabolism, glycloysis, phosphatidylinositol signaling system; Contig 985), or significantly enriched for a co-expressed pathway (ErbB signaling pathway (dre04012); Contig1512), that play important roles in growth, development and performance (see Supplementary text). These demonstrate that there is a significant level of parallelism in the functional genomic underpinnings of replicate ecotypes in charr.

Our findings demonstrate parallel phenotypic changes in several traits associated with foraging, morphology, and ecological divergence in Arctic charr over a large number of replicate but variable environments. This parallelism has evolved within and across two evolutionary lineages, spanning time scales from hundreds to tens of thousands of years. We also demonstrate a pervasive role of population genetic structure, genomic architecture, and demographic history that strongly influence non-parallelism at the genome level. We find that gene expression in an ecologically relevant tissue, white muscle, was significantly parallel across ecotypes and that candidate genes and functional pathways encompass many aspects of phenotype, such as growth and metabolism, giving a distinctive insight to the functional basis of replicate ecotypes of fishes in their natural environment and suggesting a mechanism by which eco-morphological parallelism is achieved. Thus, our findings highlight a predictability of evolution 'replaying the tape of life' (*2*), overcoming constraints of stochasticity and contingency of different evolutionary and genomic backgrounds, in one of the most expansive natural experiments in replicated evolution (*9*).

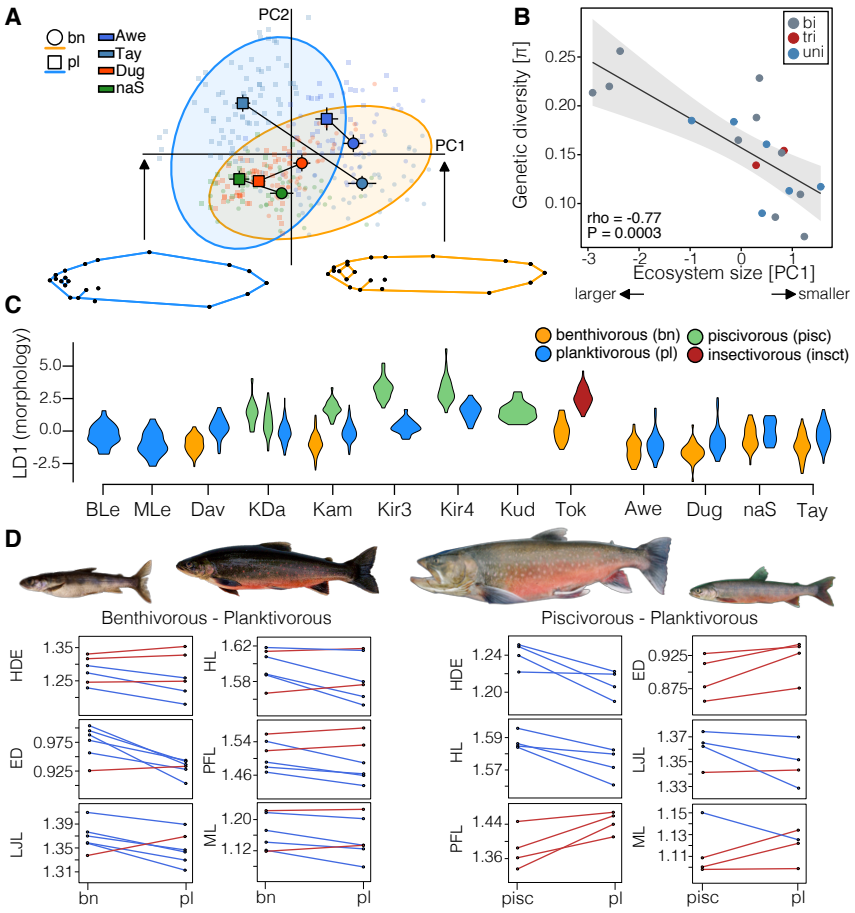


Fig. 1. Phenotypic divergence, parallelism and genetic diversity. (A) Divergence in body shape between sympatric ecotypes in the Atlantic lineage (N = 347). Phenotypic trajectories show the phenotypic parallelism along the first principal component (table S6), although overall divergence is highly parallel (table S7). Wireframes show the phenotypic extremes along PC1, highlighting the differences between benthivorous (orange) and planktivorous (blue) ecotypes, particularly in eye and head size relative to body size. Ellipses show 95% confidence intervals by ecotype. (B) Genetic diversity is positively correlated to ecosystem size, independent on the number of ecotypes within a population [Bimodal (bi) with two ecotypes, trimodal (tri) with three ecotypes and unimodal (uni) with only one or no ecotypes]. Grey areas around the regression line show 95% confidence intervals. (C) Divergence in morphology based on linear discriminant analysis (LD1) of six linear morphological traits (N = 1529). (D) Trajectories for individual phenotypic traits in benthivorous-planktivorous and piscivorous-planktivorous ecotype pairs (red or blue in either direction), with exemplar images of ecotypes above.

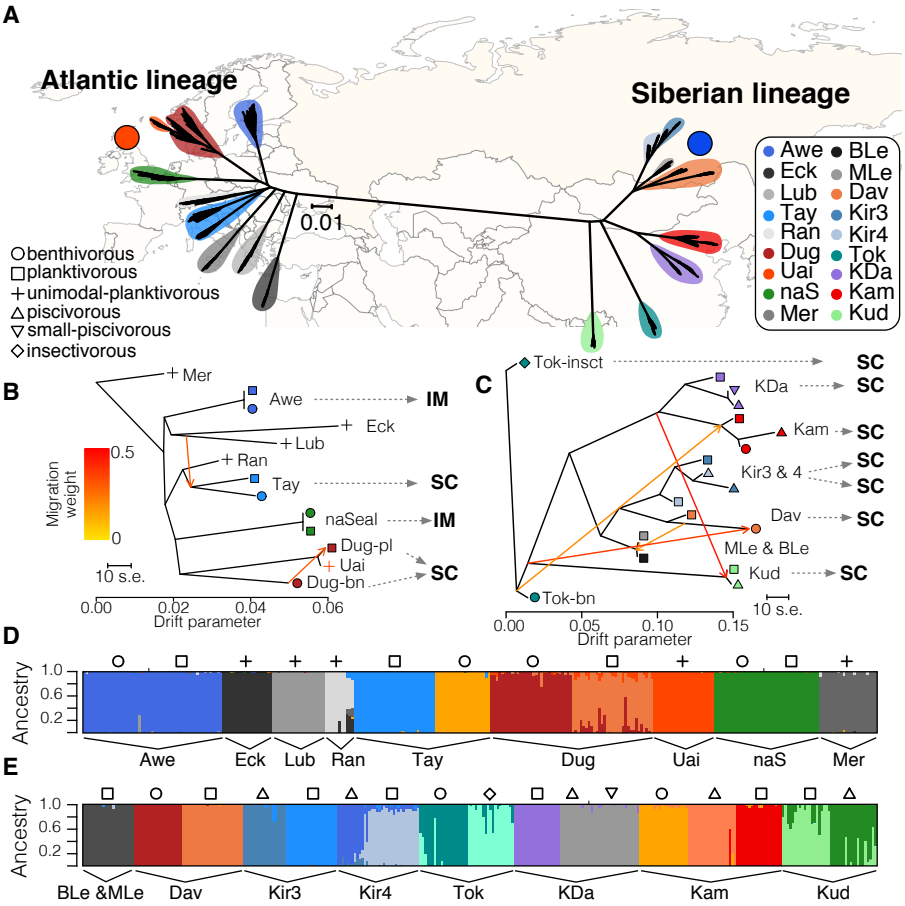


Fig. 2. Population structure and evolutionary history. (A) Neighbor-joining tree based on 12,215 SNPs illustrating the relationship between populations and lineages (N = 648). A version including bootstrap support values is available in fig. S15. (B, C) The most likely evolutionary history based on demographic models and genetic divergence patterns based on maximum-likelihood trees for (B) Atlantic and (C) Siberian population and ecotype allele frequencies. The strongest migration events across populations are shown as arrows colored by weight. (D, E) Genetic ancestry plots by lineage, K = 11 (cv = 0.317) in the Atlantic lineage (D), and for K=16 (cv = 0.184) in Siberia (E).

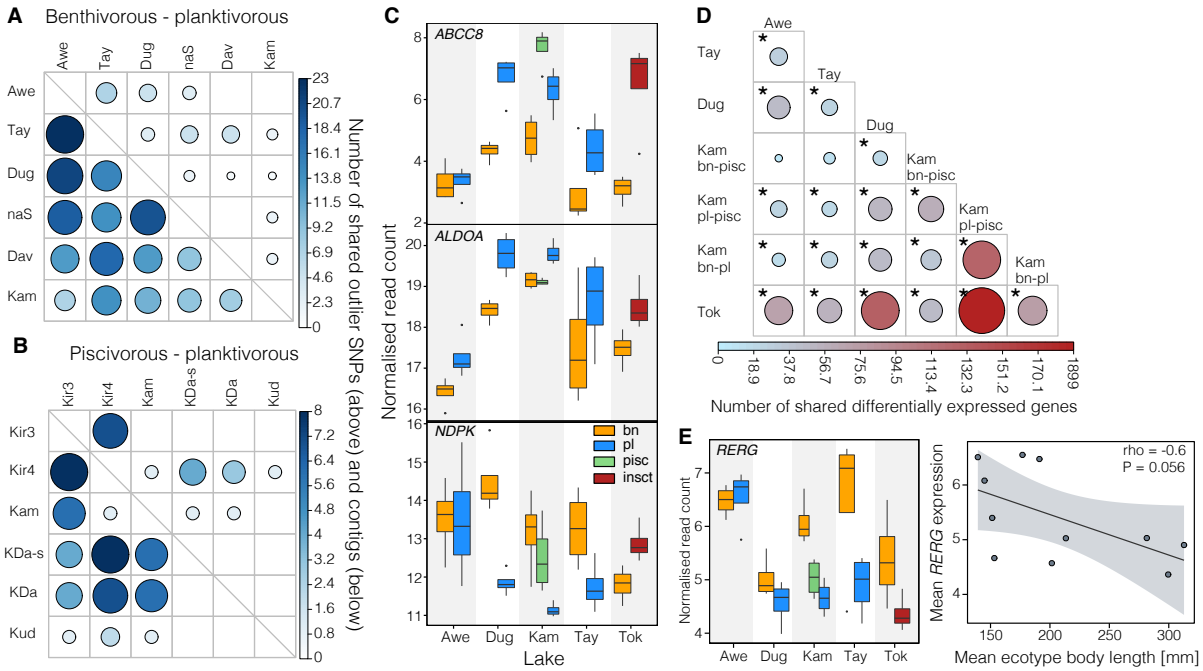


Fig. 3. Quantification of parallel and non-parallel genetic and transcriptomic divergence across ecotype pairs. (A-B) The majority of divergent SNPs (above 95th quantile of empirical Fst distribution) were unique to a single ecotype pair and not significantly shared more often than expected by chance (based on permutation test of observed versus expected number of shared outlier SNPs, $P < 0.05$). (C) Differential gene expression across all populations for three genes significantly shared. (D) Genes that are differentially expressed across ecotype pairs (extent of sharing weighted by color). Significant comparisons are highlighted with an asterisk. (E) Expression patterns of *RERG*, a growth inhibitor by ecotype that correlates with mean body size across ecotypes.

Materials and Methods

Study populations, ecotype classification, sample collection and processing

Arctic charr were sampled from nine Scottish lakes (Atlantic lineage) and from nine Transbaikalian lakes (Siberian lineage) (*31*), between 1995 and 2015 using standard gill nets (fig. S1). The sampled lake populations (we refer to all Arctic charr individuals within a lake as population) contained different numbers and combinations of trophic specialists (we refer trophic specialists as ecotypes), and detailed sample sizes for each analysis, locations and ecotype compositions are given in table S1.

Individuals were assigned to ecotype based on sampling location, spawning time, stomach content, parasites, and eco-morphology or body size *(17,19,20,27,32–35)*.

We distinguished and classified ecotypes based on primary diet (trophic) (see *(16,17,19,20,32–35)* for detailed overviews on trophic and phenotypic divergence). Although proportions of main food sources (benthic, plankton, insects and fish) differ across lakes and some populations are more opportunistic than others, we distinguished four main ecotypes in this study, based on established classifications (i.e. (*16*) for an overview) and to allow comparability across lakes and lineages: the planktivorous ecotype that mainly feed on plankton throughout the year; the benthivorous ecotype that also consumes a substantial proportion of benthic diet, particularly during autumn and winter; the piscivorous ecotype that mainly feeds on other fish; and the insectivorous ecotype for which terrestrial insects make up a large proportion of their diet. Although many ecotypes supplement their diet with plankton during certain times of the year when it is abundant, for example during the spring/summer algae bloom, ecotypes strongly differentiate based on diet (trophic niche) during resource-limited periods of the year and this is how we have classified them here. Many of these ecotypes also differ in body size, development and growth rate *(17,19,33,35,36)*.

We sampled a total of nine populations from the Atlantic lineage (table S1). Four of the nine Scottish populations (na Sealga, Tay, Awe and Dughaill) contained two ecotypes, benthivorous and planktivorous (*19,20,32*). We further sampled four populations (Eck, Lubnaig, Uaine, Merkland) containing only one non-divergent (generalist)-planktivorous populations (unimodal-planktivorous), (*37*), as outgroups for the phenotypic and genomic analysis. All Scottish populations except Loch na Sealga have been extensively studied ecotypes (*19,20,32,37*), but ecological and morphological data for the Loch na Sealga population are presented as part of this study (see below). Although Loch Rannoch harbours three different ecotypes (*38*), samples used in this study include representatives of at least two ecotypes and were used as genetic outgroup samples.

All Siberian populations used in this study were multimodal, harbouring different combinations of benthivorous, planktivorous, piscivorous, and insectivorous ecotypes (*17,27,33–36,39*). However, not all ecotypes were available for each population, either because they are extinct or very rare. For Bol'shoe Leprindo and Maloe Leprindo only the planktivorous ecotype was available and no phenotypic data were available for the small planktivorous ecotype from Kudushkit.

After collection a photograph of the left side of each fish was taken (Scottish samples) or individual fish were preserved in formaldehyde (Siberian samples) for subsequent phenotypic analysis. White muscle tissue (from underneath the dorsal fin and above the lateral line) and/or fin clips were taken for subsequent genetic and transcriptomic analysis and stored in absolute ethanol or RNAlater at -20ºC.

Analysis of linear morphological traits

A total of 545 individuals from the Atlantic lineage (44 to 119 individuals per population, mean ± SD = 68.13 ± 26.24) and 984 individuals from the Siberian lineage (20 to 335 individuals per population, mean ± SD = 109.33 ± 105.89) were used for eco-morphological analysis based on linear traits (table S1). Seven linear morphological measurements and fork length were taken for each individual from photographs using ImageJ v. 1.50i (*40*) for the Atlantic populations or directly from formaldehyde fixated fish for Siberian samples (fig. S2A; *(17)*): FL – fork length, HDO – head depth at operculum, HDE – head depth at eye, HL – head length, ED – eye diameter, ML – maxilla length, LJL – lower jaw length,

PFL – pectoral fin length. All measurements, except HDO, were comparable between Siberian and Atlantic samples. Therefore, HDO was not used in combined analyses of both lineages. Linear traits were chosen based on previous studies on eco-morphological divergence in salmonid fishes and their potential functional importance (*e.g. (17,35,41,42)*.

As these linear traits are correlated to body length, all measurements were scaled to mean fork length independently for Siberian and Atlantic populations using the allometric formula as described in (*41*):

$$\log^{10} Y_i = \log^{10} M_i + b * (\log^{10} L_m - \log^{10} L_i)$$

where Y_i is the corrected trait-value, M_i is the measured trait value, b is the slope (regression coefficient) of the regression of the trait value against fork length (L_i) within each lake and ecotype, and L_m is the mean fork length of all fish within a lineage. The slope was calculated using population and ecotype as covariates to take variation in mean fork length into account. For all subsequent statistical test, these size-adjusted measurements were used.

Kruskal-Wallis tests with a *post hoc* Dunn-test were performed on corrected linear measures for each trait to test for significant divergence between sympatric ecotypes. Further, principal component analyses (PCA) were used to uncover the major axes of phenotypic variation in the Atlantic and Siberian lineage using *prcomp* in R. To determine the effectiveness of distinguishing ecotypes within and across populations based on all linear measurements (excluding HDO) combined and for analysing the degree of morphological differentiation among ecotypes within and across lakes, we performed a canonical variate analysis (CVA) using the *Morpho* R-package with 1,000 permutations for both lineages combined. P-values were adjusted for multiple testing using a Benjamini-Hochberg correction. We plotted the derived linear discriminant scores using *ggplot2*.

Analysis of phenotypic parallelism based on linear traits

In order to determine the contribution of parallel and non-parallel evolutionary effects on the overall morphological divergence of ecotypes within and across populations, we used the MANCOVA (multivariate analysis of co-variance) approach outlined in (*43*). This analysis was performed for each lineage separately and both lineages combined using all linear traits. Unimodal populations and the single benthivorous-insectivorous ecotype pair (Tokko) were excluded for the analyses of parallelism. MANCOVAs were used to determine the effect sizes of shared (parallel) divergence between ecotypes across populations (Ecotype effect), the unique evolutionary history of each population (Lake effect) and the differences in magnitude and direction of divergence (non-parallel effect) in each population (indicated by the Ecotype x Lake ('ExL') interaction effect). We estimated Wilk's partial η^2 using the R-package *BaylorEdPsych* to determine the explanatory ability of each variable compared to the unexplained total variance. We further used linear models to estimate effect sizes on a trait-by-trait basis for benthivorous – planktivorous and for planktivorous – piscivorous ecotype pairs, and estimated Wilk's partial η^2 using the *BaylorEdPsych* R-package.

Geometric morphometric analysis of body shape in bimodal Atlantic lineage populations

In addition to the analysis of linear traits, we used geometric morphometric analysis to examine the degree of differentiation and parallelism in body shape. This analysis was done for the four bimodal Atlantic populations (Awe, Tay, na Sealga and Dughaill N = 347), as suitable high-quality photographs were not available for the other populations. For this subset of individuals, 23 homologous landmarks (fig. S2B) on each image using tpsDIG2 (*44*) to investigate body shape. Landmarks were chosen based on previous studies and based on their previously established functional importance in foraging and locomotion *(19-21,41,42)*. Shape differences due to bending of fish were removed using the unbend function in tpsutil (*Rohlf 2004*) based on landmark 1, 8, 22, 23 and 18. The resulting superimposed shape files were used for geometric morphometric analysis in *geomorph (45)*. We performed a general Procrustes fit to remove variation due to size and orientation of individuals.

First, we analysed the full dataset with all four lakes together. We tested for homogeneity of allometric curves using the *procD.allometry* function. As allometric curves were not parallel we did not compare mean shapes but rather differences in the slope of allometric curves across ecotypes and lakes, using a linear model with the logarithmic centroid size as a covariate and the ecotype and lake interaction as grouping. The model was implemented using the *advanced.procD.lm* function in *geomorph*. We further performed a principal component analysis to explore the body shape variation within and across lakes and reveal the major axes of variation. Eigenvectors were plotted using *ggplot2*.

Similar to the analysis of linear traits we used individual ANOVAs for the first five principal component axes to assess the effect of shared parallel divergence (Ecotype), unique evolutionary history of each population (Lake effect), and the non-parallel aspects of divergence across lakes (Ecotype x Lake interaction). We estimated η^2 using the *BaylorEdPsych* R-package. Furthermore, we used a phenotypic trajectory analysis (PTA) (*46*) as a complementary analysis for determining the level of parallelism in body shape across the four Scottish populations. The PTA was performed based on Procrustes distances using the *geomorph* R-package.

Ecological analysis of Loch na Sealga ecotypes

We hypothesized the presence of distinct ecotypes, as individuals of comparable age within Loch na Sealga differ strongly in fork length (fig. S4), which has been associated with ecological divergence in a range of other Arctic charr populations (e.g. *(17,19)*) and other salmonid species (e.g. Bernatchez 2010). Arctic charr were caught in Loch na Sealga during spawning time using benthic and pelagic gill nets. To determine the extent of ecological segregation among putative ecotypes in Loch na Sealga, we analysed parasite infection intensity and stable isotope signatures.

First, we analysed the infection intensity of individuals by *Diphyllobothrium spp*. larvae, a parasitic cestode following (*20*). The infection by this cestode is used as an indicator for a predominantly planktivorous diet, as this parasite is trophically transmitted through feeding on planktonic copepods, the first intermediate host for the *Diphyllobothrium spp.* parasite (*47*). Infection intensities between putatively planktivorous and benthivorous individuals were analysed using a Wilcoxon rank sum test in R.

Second, we performed a carbon and nitrogen stable isotope analysis on white muscle tissue to analyse the feeding behaviour of fish during the summer feeding period. The stable isotope analysis was performed as described in (*20*). Differences in stable isotope signatures between putatively planktivorous and benthivorous individuals were analysed using an ANOVA in R.

Chromosome-level assembly of the draft genome

We created a draft chromosome-level assembly of the Arctic charr draft genome using a publically-available genotyping-by-sequencing based high-density Arctic charr linkage map (*48*) and synteny between the Atlantic salmon and Arctic charr genomes. Ordered linkage map markers from the male and female linkage map were aligned to the draft genome using *bowtie2* and the --very-sensitive setting. Markers with a mapping quality below 20 were subsequently removed. Furthermore, we analysed synteny between the Arctic charr draft genome contigs and the Atlantic salmon reference genome (*(49), ICSASG_v2)* using the software *Chromosomer* (*50*). Finally, we used mapped linkage map markers as well as the synteny results from Chromosomer as input for *All-Maps* (*51*) to create a chromosome-level assembly. The number of chromosomes were assigned based on the female linkage map. Furthermore, we used a weight of 2 for the female and male linkage map and a weight of 1 for the synteny results. We excluded the male linkage group 3 for the final assembly as it caused large misassembles.

Annotation of the draft genome

To annotate this complex draft genome, we used a method realized in the *GeMoMa* 1.4.2 software (*52*). *GeMoMa* uses the homology-based prediction of genes with dynamic programming and intron position conservation supported by RNA-seq data to precisely predict genes in evolutionary related genomes. Intron positions were identified based on the STAR-aligned (*53*) bam files from 10 individuals (see below). Coding sequences were extracted from the NCBI RefSeq annotation of the salmon genome (*49*)

and blasted (tblastn, *(54)*) against the charr genome in parallel with GNU PARALLEL (*55*). Next, genes were predicted using the *CLI GeMoMa* module with maximum allowed intron length of 70 kbp, a maximum number of five transcripts per gene and at least three split reads to infer an intron border position based on RNA-seq data. The gene predictions were further filtered using the *GeMoMa* Annotation Filter (GAF) with the following settings: relative score filter of 2, allowing non-complete gene predictions (due to the fragmented nature of the genome), common border filter of 0.2, and with all other settings set to default. The quality of the gene prediction was estimated using *BUSCO* v.1.22 (56) using the gene set mode (--m OGS) and the Actinopterygii database (actinopterygii_odb9) with 4584 mostly single-copy genes.

To find orthologs between Arctic charr and zebrafish (*Danio rerio*) genes (*57*) we used *Orthofinder* (*58*) and the longest isoforms of genes extracted from both species.

DNA extraction, library preparation and sequencing

DNA was extracted from fin clips and muscle tissue using the NucleoSpin Tissue kit (Macherey-Nagel), following the manufacturers recommendations. DNA quality and quantity were assessed using agarose gel electrophoresis and the Qubit Fluorometer with the dsDNA BR Assay (Life Technologies). ddRADseq libraries were prepared using a modified version of the Recknagel *et al*. (*24*) ddRADseq protocol for Illumina sequencing platforms. The modified protocol used a different set of adapters designed for Illumina platforms with combinatorial barcodes. Paired-end 75-bp sequencing was performed on the Illumina NextSeq500 platform at Glasgow Polyomics (University of Glasgow) at coverage of 3-4M reads per individual.

Amplification, sequencing and analysis of the mitochondrial ND1 gene

The mitochondrial ND1 gene was amplified for 107 individuals (between 2 and 11 individuals per population) using the primer-pair B1NDF/B1NDF (59) and the PCR conditions as described in (*59*). The PCR product was cleaned and Sanger sequenced in both directions at *DNA Sequencing and Services* (MRC I PPU). Contigs were assembled from forward and reverse reads using *Sequencher* v. 5.4 (*http://www.genecodes.com*) after removing low quality reads and trimming read ends. Reads for all individuals were aligned using *Muscle* (*60*) in *MEGA* v. 7 (*61*) and trimmed to a common length. A TCS haplotype network was built in *POPART* (*62*).

Bioinformatic processing of ddRADseq data

Raw sequence data quality was assessed using *FastQC v.0.11.3* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). The *process_radtags* pipeline in *Stacks* version 1.46 (*63*) was used for demultiplexing raw sequence data based on unique barcodes, quality filtering and read-trimming to 70bp of all libraries. Cleaned and trimmed reads were aligned to the Arctic charr draft genome with *bwa mem* v. 0.7.15 (*64*) using a seed length of 25bp and the –M option. Reads with a mapping quality below 20 were removed using *samtools* v. 1.6 (*65*). We further used *Stacks* v. 1.46 and the *ref_map.pl* pipeline for building of RAD-loci and SNP calling. The *populations* module was used to export genotype calls in VCF format for further filtering in *vcftools* v. 0.1.15 (66). We created three different datasets; a global dataset for the Atlantic and Siberian lineages combined, and a separate dataset for each lineage. SNPs were retained when the following criteria were fulfilled: (i) present in at least 66% of all individuals within a population and 2/3 of all populations, (ii) global minor allele frequency (MAF) ≥ 0.05 or ≥ 0.01 for the global dataset, (iii) heterozygosity ≤ 0.5 (iv) in Hardy-Weinberg equilibrium ($P > 0.05$) in at least 2/3 of all populations and (v) with a minimum coverage of 6x. Filtering and data conversion of data into the different formats was performed using *Stacks*, *vcftools*, *PLINK* version 1.90 *(www.cog-genomics.org/plink/1.9/; (67))* and *PGDspider* v. 2.11.2 (*68*). Datasets for creating site frequency spectra were filtered in a similar way, except no MAF cut-off was used in order to retain informative low frequency sites and a maximum of 10% missing data within each dataset (e.g. each ecotype pair) were allowed. For all analysis, only one SNP per locus was retained to reduce the effect of linkage.

Summary statistics and analysis of population structure

We calculated nucleotide diversity for each ecotype by population and all populations using *vcftools*. To assess population structuring across and within each lineage, and within lakes, we used a range of different approaches using the global SNP datasets and lineage-specific datasets. First, we used principal component analyses in *adegenet* (*69,70*) to assess major axes of genetic variation for the global dataset and for each lineage specific dataset. Second, *Admixture* v.1.3 (*71*) was used with a ten-fold cross-validation to detect the most likely numbers of clusters and genetic ancestry proportions within lineages. *Genodive* v.2.0b27 (72) was used to estimate pairwise genetic differentiation (Fst; *(73)*) between ecotypes within and across lakes using the global dataset, and significance was assessed using 10,000 permutations. We also assessed the relationship among populations within and across lineages using a neighbour-joining network as well as neighbour-joining tree based on uncorrected P distances using *SplitsTree4 v.4.14.4* (*74*). Bootstrap support values for the neighbour joining tree were calculated using 100 bootstrap replicates.

Furthermore, to assess genetic co-ancestry among individuals within and across lakes, we used haplotype-based population inference approach implemented in *fineRADstructure* v.0.1 (*75*). We used the same filtering criteria as for the SNP dataset and performed the analyses using default settings. In brief, we calculated the co-ancestry matrix using *RADpainter* as implemented in the *fineRADstructure* package and subsequently inferred populations using a burn-in of 100,000 and 100,000 MCMC iterations. Finally, the results were visualized using the provided R-scripts (http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html).

Analysis and visualization of introgression and differential admixture

We first used *Treemix* version 1.13 (76) to explore and visualize secondary gene flow within each lineage. We build maximum likelihood trees only for non-admixed individuals (admixture threshold of 0.25 as inferred with *Admixture*) to reduce the effect of contemporary admixture on the inference of secondary gene flow. We fitted up to six and ten migration edges for the Atlantic and Siberian populations, respectively, and chose the strongest and most explanatory migration events based on the maximized variance explained, the maximum likelihood of the tree and the significance of each migration edge. To quantitatively test for secondary gene flow in the tree we used *f4-statistics* implemented in the *fourpop* function of *Treemix*. Furthermore, to formally test for differential admixture or introgressive hybridization in a three-taxon comparison (Target taxon C; Reference taxon A, Reference taxon B) we used *f3-statistics* as implemented in the *threepop* function in *Treemix*. Significant negative deviations from zero indicate the presence of differential admixture, e.g. in the form of introgressive hybridization, of two populations (e.g. A and B) in a third population (target population C).

Inference of evolutionary histories

To distinguish between alternative evolutionary scenarios leading the ecotype diversity within lakes, we used coalescence simulations implemented in *fastimcoal2* v.2.5.2.3 (*25*) and information contained in the multidimensional site frequency spectraum (SFS). 2-population and 3-population SFSs were created using *δaδi* v.1.6.3 (*77*) for each lake population and parapatric outgroups if appropriate. Populations were down-sampled by around 30% to reduce the effect of missing data (maximum 10% missing data in used SNP datasets). Since no trinucleotide substitution matrix is available for salmonids and no outgroup species was sequenced for ancestral state reconstruction and correction, the minor folded site frequency spectrum was used. To determine absolute values for divergence times and other inferred parameters, we corrected the number of monomorphic sites in the SFS as described in (*78*). Adjusting the number of monomorphic sites to the number of variant sites used to build the SFS is necessary to determine absolute parameter values, as only one SNP per RAD locus was retained for the analysis, skewing the ratio of monomorphic to variant sites (*78*). Furthermore, a mutation rate of $1x10^{-8}$ was used as no accurate mutation rate for Arctic charr or salmonids is available (*26*).

For all sympatric ecotype pairs, seven pairwise demographic models describing different historical divergence scenarios were tested (fig. S17): Strict isolation (SI), Ancient migration (AM), Isolation-with-migration (IM), Secondary contact (SC), Secondary Contact with introgression (AdmSC), Isolation-with-migration with a historical change in migration rate (IMchange) and an IM-model with a historical introgression event sympatric species (IMint) were tested. In lakes with three sympatric ecotypes (Kamkanda and Kalarskii Davatchan), or strong admixture across lakes (Loch Dughaill and Loch Uaine), models describing different combinations of strict isolation, isolation-with-migration, secondary contact, introgression and hybrid speciation were tested for all three ecotypes/populations together (fig S16). These models tested in general two different evolutionary histories. The isolation-with-migration models test a history of divergence under constant gene flow (with differing rates), whereas secondary contact models mainly test the occurrence of historical secondary contact between different distinct lineages or populations (e.g. distinct gene pools or glacial refugial populations) prior to ecotype divergence, and not necessarily multiple invasions.

We ran a total of 30 iterations for each model and lake, and selected the most likely model based on the AIC *(25)*. Each run consisted of 40 rounds of parameter estimation with 100,000 coalescent simulations. Point estimates of inferred parameters were taken from the most likely model and averaged over the top five runs.

Outlier analysis, positive selection, and genome-wide patterns of differentiation

To determine outlier loci between sympatric ecotypes, we screened the genome for loci showing genetic differentiation (Weir-Cockerham Fst calculated using *vcftools*) above the 95th quantile of the Fst distribution (outlier loci). We estimated the number and proportion of pairwise shared outlier SNPs across benthivorous-planktivorous and piscivorous-planktivorous ecotype pairs, and identified potential loci under parallel selection. We also identified contigs containing outlier SNPs (but not the same SNPs across contigs) that are shared among ecotype pairs. In order to determine if more outlier SNPs or contigs are shared among two ecotype pairs than expected by chance, we resampled N number of SNPs (N = number of outlier SNPs for each ecotype pair) 10,000 times with replacement from the full dataset and determined the mean number of shared SNPs from that distribution. We used a Chi-square test to determine if more outlier SNPs/contigs are shared than expected. We also plotted the standardized Fst (ZFst) against delta nucleotide diversity ($\Delta\pi = \pi_{ecotype2} - \pi_{ecotype1}$) between sympatric ecotypes, to determine if outlier loci show reduced genetic diversity in one or both ecotypes.

We used *pcadapt* 3.0.4 (*79*) to identify SNPs potentially under positive selection in each lineage-specific dataset. As the PCAdapt approach is based on detecting loci that separate along different principal components, we used the first 10 and 14 principal component axes for the Atlantic and Siberian populations, respectively, as these separate the different lakes and sympatric ecotypes within lakes. We considered loci with *q-values* of 0.01 or below (false discovery rate of below 1%) to be candidates for being under positive selection. P-values were transformed to q-values using the *qvalues* R-package.

Genome-wide association analysis

To detect loci significantly associated with ecotype, we used two different approaches, a univariate linear mixed model (LMM) and a Bayesian sparse linear mixed model (BSLMM) implemented in *GEMMA* (*80*). BSLMM also calculates a set of hyperparameters describing the genetic architecture underlying trait (ecotype) variation. For these association analyses, ecotypes were coded numerically, with planktivorous as 0, benthivorous as 1 and piscivorous as 2. We excluded unimodal-planktivorous outgroup populations from Scotland and the ecotypes from Tokko from this analysis, and performed the association analysis separate for each lineage. As these approaches do not allow missing data in the dataset, we imputed missing data using the LD-kNNi method (*81*) implemented in Tassel5 (82). This method has been specifically designed and tested for RADseq data and is based on linkage to the k-nearest neighbours. We imputed missing genotypes based on the 10 closest, in terms of linkage and not genomic position, genotypes using the default settings. To test the imputation accuracy, we calculated the Pearson correlations between allele frequencies before and after imputation for the full dataset and the subset with

the highest proportion of missing data. Genotype and phenotype data were formatted in *Plink* binary ped file format.

As the LMM and BSLMM approaches correct for the presence of population structure using a relatedness matrix, we calculated a centred relatedness matrix for each lineage using *GEMMA*. First, we used a linear mixed model to test for association between genotype and ecotype using the Wald test and report p-values. We used a conservative cut-off of $\alpha = 0.0001$ and $-\log10(\alpha) = 4$ to determine statistical significance of associations (*83*). Second, we used a Bayesian sparse linear mixed model to determine associations of genotypes with ecotype and estimate the genetic architecture underlying ecotype divergence. In contrast to single-SNP association mapping approaches, such as the linear mixed model, BSLMM is a polygenic model that simultaneously assesses the association of multiple SNPs with trait variation and performs well for a wide range of genetic architectures underlying phenotypic variation, ranging from polygenic architectures to 'sparse' architectures with a few large effect loci underlying phenotypic variation (*80*). Furthermore, the BSLMM model also estimates a range of hyper-parameters describing the genetic architecture underlying phenotypic variation. We ran ten independent chains for the standard linear BSLMM, each with 25 million generations for the Atlantic dataset and 50 million generations for the Siberian lineage, excluding the first 5 and 25 million generations as burn-in for the Atlantic and Siberian dataset, respectively. We assessed convergence of the runs based on the mixing of the hyper-parameters. We estimated the posterior inclusion probability (PIP) for each SNP, the proportion of generations a SNP has an effect on phenotype, by averaging across the ten independent chains. We used a stringent cut-off of PIP > 0.1 to identify candidate SNPs associated with ecotype. Furthermore, we summarized hyper-parameters describing the genetic architecture underling ecotype divergence within each lineage by calculating the mean and 95% confidence intervals for the parameter posterior distributions across all independent chains.

In order to assess if allele frequencies of candidate loci are similar across parallel ecotypes within lineages, we first performed principal component analyses using BSLMM candidate loci for all lakes within each lineage. Subsequently, we performed two-way ANOVAs on individual scores for each principal component separately (PC1-5 in Atlantic and PC1-9 in Siberian lineage), with ecotype, lake and ecotype-lake interaction as variables. Similar to the phenotypic analysis, we then estimated the effect size (η^2) of each variable on the genetic variation along each significant principal component using the *BaylorEdPsych* R-package.

Annotation and functional characterization of Fst outlier loci and GWAS candidate loci

We identified all annotated genes containing ecotype associated loci (LMM and BSLMM) and those under selection (PCAdapt) and the closest gene within 10 kb of those SNPs. We used all genes with zebrafish orthologues (one-to-one or one-to-two in zebrafish) for an Overrepresentation Enrichment Analyses (ORA) in *Webgestalt* (*84*), using the zebrafish genome annotation as background. Identified KEGG pathways, both significantly enriched and none-enriched, were then compared across ecotype pairs to identify parallel pathways. Pathways were identified as significantly enriched with a false-discovery rate (FDR) below 0.1.

Correlation of genetic divergence, phenotypic divergence and ecosystem size

Identifying the drivers of phenotypic and genetic divergence is important for understanding the processes that lead to deviations from parallelism or even parallelism itself. We tested for correlations between the degree of phenotypic differentiation (Pst), genetic divergence (mean neutral Fst and mean outlier Fst) and ecosystem size (PC1), as a proxy for ecological opportunity (*24*), for all ecotype pairs combined and for benthivorous – planktivorous and piscivorous – planktivorous pairs separately (genetic divergence vs. phenotypic divergence, genetic divergence vs. ecosystem size, phenotypic divergence vs. ecosystem size) using Pearson's correlation (*cor.test* function in R). We used the first principal component from a PCA performed using *prcomp* based on maximum lake depth and surface area, estimated using Google Earth Pro, as a proxy for ecosystem size (*24*). Pst values between all sympatric ecotype pairs were calculated for each trait and the mean was calculated across trait-specific Pst-values to estimate the proportion of

phenotypic variance among ecotypes relative to the total variance (*4*) using the *Pstat* R-package. Pst values are unit-less and directly analogous to Fst (*4*). We corrected for multiple testing using a Benjamini-Hochberg correction implemented in the R-package *p-adjust*. However, due to the relatively small number of data points (N = 14) and the large number of pairwise comparisons, we were primarily interested in the strength of correlation rather than p-value significance, treating Spearman correlations (rho) over 0.5 as relevant for hypothesis generation and further analyses (*4*).

RNA extraction, sequencing and alignment

RNA was extracted from white muscle tissue from 44 individuals (N = 4 per ecotype per lake) from five different lakes (Awe, Tay, Dug, Kam, Tok), representing all possible ecotypes (benthivorous, planktivorous, insectivorous, piscivorous), using PureLink RNA Mini Kits (Life Technologies, Carlsbad, CA). Extractions were mostly carried out following the manufacturer's instructions, with an additional homogenization step using a FastPrep-24 (MP Biomedicals), before RNA extraction (*99*). RNA was quantified using the RNA HS Assay kit for the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA) and quality was assessed using the 2200 Tapestation using the High Sensitivity RNA ScreenTape (Agilent, Santa Clara, CA). Extracted RNA was of high quality, yielding A260/280 ratios between 1.9 and 2.1 and RIN (RNA Integrity Number) values above 8.3.

RNA-seq libraries were prepared and sequenced at Glasgow Polyomics (University of Glasgow) for Awe, Tay, Dug and Kam and at BGI (Shenzhen, China) for Tok. Individual cDNA libraries were prepared for each individual using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, San Diego, CA) in combination with a Poly-A selection step. Libraries were sequenced on the illumina NextSeq 500 system (Illumina, San Diego, CA) using 75 bp paired-end sequencing with a sequencing depth of 25-30 million reads per library.

Raw reads were processed using *Scythe v0.9944 BETA* (https://github.com/vsbuffalo/scythe/) and *Trimmomatic v0.36* (*85*) to remove any adapters, ambiguous bases and low-quality sequences prior to alignment. Leading and trailing bases with a Phred quality score below 20 were removed and a sliding window approach (4 bp window size) was used to trim reads at positions with Phred scores below 20. A minimum read length of 50 bp was allowed after trimming. We used *FastQC v0.11.2* to assess read-quality before and after processing. Processing removed 2% of all reads, resulting in a dataset of 1.81 billion cleaned reads. The resulting reads were aligned against the Arctic charr draft genome using STAR *v*2.5.2b, with default parameters.

Differential gene expression analysis

To identify differentially expressed genes, reads were counted for each gene based on the longest isoform annotations using the HTSEQ-count python package (*86*) with the unstranded (--stranded=no), CDS–based (--type=CDS, --idattr=Parent) settings. The read count table was used for the differential gene expression analysis between sympatric ecotypes using *DESeq2* package (*87*) on a per lake basis. Only genes with at least 20 read counts per lake were used.

Network analysis

To further examine the functional bases of trophic divergence Arctic charr, we used a Weighted Gene Co-Expression Network Analysis (WGCNA) to identify modules of co-expressed genes (*88*). Network analyses were only performed on benthivorous–planktivorous ecotype pairs (from Awe, Tay, Dug and Kam, n=32, eight per lake, four per ecotype within lake), given that sample sizes for insectivorous and piscivorous ecotypes were too low (n = 4 per ecotype) to robustly identify co-expressed network modules for these ecotypes. Prior to further analyses, count data were transformed to a $\log_2$ scale using the *rlog* function in DESeq2.

To reduce stochastic background noise in our expression data resulting from lake-specific effects, we used the R package *variancePartition v*3.6 (*89*) to identify a subset of genes with expression variation attributed to ecotype. *VariancePartition* applies a linear mixed model (LMM) to partition the expression variance for each gene into the percentage of expression variation explained by 'lake' and 'ecotype'

effects. All remaining expression variation not attributable to these factors was termed residual variation. Batch effects were removed from expression data using an initial LMM, with 'batch' as a random effect, as per the *varianceParition* pipeline (*89*). The residuals from the initial model were then used as the response variable for a subsequent LMM, with 'lake' and 'ecotype' specified as explanatory variables; 'lake' and 'ecotype' represent categorical variables and were thus modelled as random effects. *VariancePartition* assesses each gene individually and the results for all genes are aggregated afterwards (*89*). All genes for which 'ecotype' explained more than 10% of the total expression variation across individuals were used for network construction.

A single network was constructed for benthivorous and planktivorous individuals from the set 1,512 ecotype-associated genes. Following the standard procedure of WGCNA, pair-wise Pearson's correlations were calculated for all genes and raised to a power 12 to generate the adjacency matrix. A topological overlap matrix (TOM) was then calculated from the adjacency matrix to measure network interconnectedness (the strength of a co-expression relationship between any two genes with respect to all other genes in the network; *(90)*). Genes with highly similar co-expression relationships were clustered using average linkage hierarchical clustering of the topological overlap dissimilarity measure (1-TOM). Finally, network modules were defined using the dynamic treecut algorithm, with a minimum module size of 25 genes and a cut height of 0.992 (*91*). The module eigengene distance threshold was set to 0.25 to merge highly similar modules. Modules of co-expressed genes were labelled by color (assigned arbitrarily by WGCNA).

To determine the significance of module-trait relationships, Spearman's correlations were calculated between module eigengenes (defined as the first principal component of the expression profile for a given module) and trait measurements (lake and ecotype). Module -trait relationships were considered significant when p-values were <0.05 after Benjamini-Hochberg correction. The direction of correlation (i.e. positive or negative) determined whether modules were associated with expression in benthivorous or planktivorous ecotypes. Trait values for ecotype were specified in binary format (planktivorous = 0 and benthivorous = 1), therefore positive correlation represent up-regulation in benthivorous ecotypes (over planktivorous), and negative correlations represent up-regulation in planktivorous ecotypes (over benthivorous).

Functional characterization of differentially expressed genes

To find genetic pathways associated with ecotype divergence in Arctic charr, we performed functional enrichment analyses on differentially expressed genes (all pair-wise comparisons of ecotype, across all five lakes) and co-expressed gene modules (benthivorous – planktivorous ecotype comparisons, form Awe, Tay, Dug and Kam), using the web interface of the *WebGestalt* tool (*84*). For differentially expressed gene sets we used Gene Set Enrichment Analyses (GSEA; *(92)*) to identify enriched between different ecotypes. Before the analysis, the DESeq2 raw p-values for each lake-ecotype comparison were modified to create a ranked list of genes using the following formula: -log10(p-value)*log2FoldChange. To identify enriched pathways within co-expression modules we performed an Overrepresentation Analyses (ORA) for each module individually.

For the GSEA and ORA analyses, we used only charr genes which had 1:1 or 2:1 orthologs in the zebrafish genome. The analysis was performed with the following settings: minimum number of genes for a category = 5, maximum number of genes for a category = 500, number of permutations = 1000, number of categories with leading-edge genes = 20, KEGG pathways (*93*), organism = *Danio rerio*.

Gene sharing between comparisons.

To calculate an expected number of the shared DEGs between the comparisons we used a permutation-based approach, similar to the outlier comparison (see), with 10,000 permutations. We randomly sampled n genes (n = the number of DEGs in a comparison) 10,000 times from each dataset with replacement and calculated the expected number of shared DEGs as the mean number of shared resampled genes in each comparison. We calculated proportion-based p-values based on the number of observed shared DEGs, expected DEGs and the total number of genes for each pairwise comparison using R function *prop.test*.

Supplementary Results

Ecological divergence in Loch na Sealga

Arctic charr from Loch na Sealga that were classified as benthivorous and planktivorous based on stomach content were strongly differentiated in the intensity of *Diphyllobothrium spp*. infections (fig. S3B; Wilcoxon rank sum test: $P < 0.001$), indicating strong segregation in trophic niche *(47)*. Furthermore, the stable isotope analyses revealed that planktivorous and benthivorous ecotypes displayed divergent nitrogen signatures (fig. S4A; $F_{1, 62} = 20.65$, $P < 0.001$). Planktivorous and benthivorous also differed in the length of the lower jaw (fig. S4; table S2; $Pst_{LJL} = 0.783$) and fork length (fig. S4; table S2; $Pst_{SL} = 0.967$). These results indicate the presence of ecologically and morphologically distinct ecotypes, although this divergence is relatively subtle in most aspects.

Continuum of eco-morphological divergence across lakes

We used seven linear morphological traits and fork length to analyze the extent of morphological divergence between ecotypes within and across lakes and lineages. Allometric effects cannot be fully removed and body size will in some cases have an effect on the downstream analysis, where trait variation and body size variation are strongly correlated. However, the fact that the allometric correction fully removed the effect of body size for most traits, such as head depth at eye (Spearman test: rho = -0.002, P = 0.94), head length (Spearman test: rho =-0.05, P = 0.052), and lower jaw length (Spearman test: rho = -0.039, P = 0.1251), and that other traits only display very weak although significant ($P < 0.05$) correlations with individual fork length, i.e. eye diameter (rho = -0.29) and maxilla length (rho = -0.10), suggests that the effect of body size on downstream analyses are negligible (*35*).

We found that the main axis of morphological variation in the Atlantic lineage primarily separated the different bimodal populations (figs. S5 and S6). Sympatric (benthivorous and planktivorous) ecotypes split along PC2 in Awe, Tay and na Sealga but mainly along PC1 in Dughaill. In the Siberian lineage, however, the main axis of morphological variation (PC1) differentiated benthivorous and planktivorous ecotypes, followed by the split of planktivorous and piscivorous ecotypes along PC2 (figs. S5 and S6). Furthermore, the insectivorous ecotype in Tokko separated from the benthivorous and planktivorous ecotypes along PC1, but also split from all other ecotypes along PC3 (figs. S5 and S6).

The canonical analysis of variance for both lineages combined showed that individuals clustered by ecotype, even across lineages, but with substantial overlap in morphospace (Fig 1C, fig. S7). The average assignment success of individuals to the respective ecotype across lakes was 70.5%. The overall degree of divergence between sympatric and allopatric ecotypes differed across comparisons (measured as mahalanobis distance), and in several cases trophically similar allopatric ecotypes could not be significantly differentiated based on morphology, for example the planktivorous ecotype from Bol'shoe Leprindo from the planktivorous ecotypes in Kamkanda and Kiryalta-3 or the piscivorous ecotype in Kudushkit from the piscivorous ecotypes in Kamkanda or Kalarskii Davatchan (fig. S8). However, all sympatric ecotypes, except in na Sealga (Mahalanobis distance = 1.295, p = 0.1), could be significantly distinguished based on overall morphology. Overall, we found that the level of morphological divergence varied strongly across lakes but that sympatric ecotype pairs were significantly diverged, at least in one or more traits.

Geometric morphometrics of body shape

In addition to linear traits, we also examined the level of divergence and parallelism in body shape using geometric morphometric analysis on the four Atlantic benthivorous – planktivorous ecotype pairs (Tay, Dughaill, Awe, na Sealga). The principal component analysis showed that all sympatric ecotypes split along PC1 (38.2%), and to varying degrees along PC2 (20.3%) (Fig. 1A). Most of the variance in body shape along PC1 and PC2 was explained by differences in head and eye size compared to the rest of the body, as well as body depth (Fig. 1A). Benthivorous ecotypes had in general larger heads and bigger eyes in relation to their body size.

To test for differences in body shape among ecotypes and lakes, we first analyzed all ecotype pairs combined and compared differences in allometric curves among sympatric ecotypes, as allometric curves were not homogenous across lakes ($F = 1.576$, $Z = 5.67$, $p = 0.001$). We found that sympatric ecotypes in Awe ($Z = 5.4$, $p = 0.003$), Tay ($Z = 5.6$, $p = 0.006$) and Dughaill ($Z = 2.2$, $p = 0.044$) significantly differed in the angle of allometric slopes but not distances, meaning that shape changes differently among these ecotypes during ontogeny (table S5). Allometric curves did not differ between ecotypes in Loch na Sealga (table S5; $Z = -0.38$, $p = 0.569$).
Using a similar approach to the analysis of linear trait, separate ANCOVAs for each principal component showed that most of the phenotypic trait variation (PC1) was explained by the parallel ecotype effect (table S6: $\eta^2_{PC1,\ Ecotype} = 0.486$), even more than the effect of evolutionary history ($\eta^2_{PC1,\ Lake} = 0.391$). However, the non-parallel 'lake term' and 'ecotype x lake interaction' term explained most of the variation along PC2 to PC5, suggesting that the head region and body depth show the strongest level of parallelism across populations (Fig. 1A). To further explore the level of parallelism in overall body shape, we used a phenotypic trajectory analysis (PTA) (*43*), which takes the full phenotypic variance for all PCs into account. The PTA revealed that Tay and na Sealga are parallel in the angle of phenotypic trajectories (table S7; direction of change; $\theta = 26.96$, $Z = 1.89$, $p = 0.06$), whereas na Sealga and Awe ($Z = -0.57$, $p = 0.72$) and na Sealga and Dughaill ($Z = 1.11$, $p = 0.15$) are parallel in the length of phenotypic trajectories (table S7; magnitude of change). This shows that some populations are more parallel in the direction of change and others more in the magnitude of change. Overall, this demonstrates that even though that most ecotype pairs are not fully parallel in body shape, that certain aspects of their shape, mainly in the head region, show significant parallelism across lakes.

Genome annotation, orthology inference and chromosome assembly.

We were able to assemble 62.6% of the draft genome (1.346 Gb) into 39 Arctic charr chromosomes (*48*) with a total of 26.7% (0.573 Gb) being oriented. 37.4 % (0.804 Gb) of all contigs could not be anchored to the linkage map.

We annotated 44102 genes with 55443 alternative isoforms using homology-based *GeMoMa* gene prediction software. The BUSCO analysis estimated that the annotation contained 88% of completely assembled BUSCOs and 3.4% of fragmented BUSCOs, while 7.6% of BUSCOs were missing. Among the 88% completely assembled BUSCOs, 62% were duplicated, which is consistent with the estimation for the salmon genome (*49, 100*).

Orthofinder returned 15485 orthogroups with a mean and median orthogroup size equal to 2.7 and 2, respectively, which can be explained by the relatively recent whole genome duplication in salmonids (*49*). We detected 8428 single-copy orthogoups.

Genetic clustering and shared ancestry

Using a combined (12,215 SNPs) and lineage-specific (Siberia: 4475 SNPs; Atlantic: 6,039 SNPs) SNP datasets, we analyzed the major axes of genetic variance, patterns of genetic clustering and shared ancestry. We used principal component analyses and individual clustering analyses to infer the major axes of genetic variation, genetic ancestry within and across populations and infer potential modes of speciation.

The combined principal component analysis revealed a main clustering of populations by lineage and then catchment within each evolutionary lineage (figs. S11 and S12). Under the scenario of a single colonization event of each lake and a subsequent intralacustrine diversification, we would expect the divergence along independent principal components within each population and that the resulting ecotypes would be equidistant to all other outgroup populations (*94,95*). However, contrary to our expectations we found that ecotypes in several intralacustrine radiations did not diverge along independent principal components but often split along non-independent axes and share varying degrees of ancestry with other populations (fig. S12). In Kiryalta-3 and Kiryalta-4, populations split along the same principal components (PC9 and PC11) and clustered by ecotype (fig. S12). Similarly, in Dughaill and Uaine, we observed that the planktivorous Dughaill ecotype is equidistant to the benthivorous

Dughaill ecotype and the unimodal population from Uaine. This was supported by significant *f3*-statistic, showing significant introgression from the Uaine population and the benthivorous Dughaill ecotype (or ancestral populations) into the planktivorous ecotype (table S10). This significant signals of selection was also detected based on allopatric comparisons (Tay-bn:Uai or Awe-pl:Uai), suggesting that introgression occurred through an ancestral lineage that is genetically similar to the benthivorous Dughaill ecotype and southern Scottish populations (i.e. Tay, Awe, Eck). This is supported by considerable sharing of mitochondrial haplotypes between those populations. In other cases, this genetic clustering is less extreme but differences in shared genetic ancestry could still be observed. For example, the planktivorous ecotype in Tay is genetically closer to the Loch Awe population than the benthivorous Tay ecotype, and the planktivorous ecotype from Davatchan is genetically intermediate between the benthivorous Davatchan ecotype and the (Bol'shoe and Maloe) Leprindo populations (figs. S11 and S12). In general, these scenarios of secondary gene flow across populations were supported by significant *f4*-statistics for almost all populations (table S11), indicating that gene flow was prevalent across populations.

However, ecotypes in Kamkanda and Kalarskii Davatchan split along independent principal components (Kam: PC10, KDa: PC8), suggesting their divergence in sympatry (*94*). The benthivorous ecotype in Kamkanda is intermediate and equidistant to the piscivorous and planktivorous ecotypes, supporting a divergence in sympatry with similar levels of gene flow between ecotypes. In Kalarskii Davatchan, the small-piscivorous ecotype is placed intermediate to the large-piscivorous and the planktivorous ecotype, but genetically more similar to the large-piscivorous ecotype, supporting the results of the *f3*-statistics (table S10), suggesting introgression into the small-piscivorous ecotype from the other sympatric ecotypes or historical populations, because signals of introgression were also significant for allopatric and more distant populations (the piscivorous Kudushkit and benthivorous Tokko ecotypes).

The analyses of genetic co-ancestry (figs. S13 and S14) revealed varying levels of shared genetic co-ancestry across populations and ecotypes, with mostly higher levels of shared co-ancestry within catchments (e.g. Dughaill and Uaine or Tay and Rannoch; fig. S13) but also in some cases across more distant populations (e.g. the planktivorous ecotype in na Sealga showed higher genetic co-ancestry with Eck than the benthivorous na Sealga ecotype). Similar patterns were observed for the Siberian populations (fig. S14), with for example Tokko sharing different levels of genetic co-ancestry with Kamkanda and Kalarskii Davatchan.

Thus, these combined analyses of genetic clustering and shared co-ancestry suggest a complex history of ancestral admixture and subsequent ecotype divergence.

Demographic models

To test whether sympatric ecotypes evolved under a scenario of secondary contact (historical contact and ancestral admixture of different lineages or gene pools) or completely in the presence of gene flow only between the contemporary sympatric ecotypes, we used a demographic modelling approach based on the folded joint allele frequency spectrum. We found that the most likely demographic history of most ecotype pairs in the Atlantic and Siberian lineage include signs of historical secondary contact and/or introgression (table S12). These models tested that ancestral populations or lineages came into secondary contact before or while colonizing newly formed postglacial lakes, presumably following the retraction of Pleistocene glaciers (*17,96*). These admixed populations most likely subsequently diverged phenotypically in sympatry due to disruptive selection.

We found that the contemporary genetic diversity in most Siberian populations most likely originated under a secondary contact scenario, in which distinct ancestral populations (two or more) came into secondary contact after the retraction of Pleistocene glaciers (fig. S18, table S12). The mean divergence time among ancestral populations was 8180 generations, which translates into around 24,450 – 81,800 years for generation times of 3 - 10 years (interval of the first maturation of different ecotypes) ((*27*), *personal observations*), and a mean timing of introgression around 861 generations (2,583 - 8,610) years ago (fig. S18). The demographic modelling approach could not fully resolve the evolutionary history of ecotypes in Kudushkit. Although the secondary contact model (SCint) in Kudushkit is more

likely than the model of isolation-with-migration and introgression (IMint) (ΔAIC = 4.47), we inferred a recent divergence time between genetic clusters of 1,580 generations (fig. S18), which corresponds into a divergence time of 9,480 – 20,540 (mean = 15,010) years (mean generation time of 9.5 years, ~6 years in the planktivorous and ~13years in the piscivorous ecotype; (*33*)).

In the Atlantic lineage, the most likely model of divergence among sympatric ecotypes was the isolation-with-migration and introgression (IMint) model, in which ancestral populations and/or ecotypes diverged under gene flow and experienced an introgression event at some point in time with a change in migration rate (fig. S18). However, inferred divergence times and therefore the interpretation of the evolutionary history differed strongly across lakes. In Dughaill and Tay, ecotypes most likely diverged prior to the last glacial maximum around 20,699 and 10,931 generations ago (62,097 and 32,793 years with a 3-year generation time), respectively, with a respective timing of secondary contact around 90 and 1176 generations ago (270 - 3,528 years). Although gene flow could be detected between the postglacial lineages during the last glacial maximum, the rate of gene flow was around one to two magnitudes lower before the secondary contact compared to contemporary gene flow (fig. S18). In contrast, ecotypes from Awe and na Sealga genetically diverged around 1,245 and 4,795 generations ago (3,735 - 14,385 years; 3-year generation time), so most likely after the last glacial maximum. Introgression in Awe and na Sealga occurred around 0-10 generations ago, meaning the model effectively converged into an isolation-with-migration model. However, it could also be caused by very recent introgression into those populations from an external population. In Awe and na Sealga, the inferred rates of gene flow prior to introgression ($m_{avg.prior,Awe}$ = 3.75e^{-3}; $m_{avg.prior,naS}$ = 3.18e^{-3}) were stronger compared to recent gene flow after introgression ($m_{mean.post,Awe}$ = 1.53e^{-9}; $m_{mean.post,naS}$ = 5.48e^{-4}).

In some cases, we cannot significantly distinguish competing models with a threshold of ΔAIC of 4, mainly because of limited information contained in the site frequency spectrum and complex evolutionary histories. Therefore, other modes of speciation or variations of the tested modes are also possible (table S12). However, in most cases the next best model is a variation of the best fitting model, e.g. secondary contact without introgression (SC) rather than with introgression (SCint).

As pairwise comparisons in systems with more than two ecotypes are prone to underestimate gene flow and misinterpret scenarios of introgression and gene flow, we tested between six to eight different models for each system with three ecotypes (Kamkanda, Kalarskii Davatchan and Dughaill-Uaine) using different combinations of the pairwise models (fig. S17). Due to the significant *f3* statistics for the small-piscivorous Kalarskii Davatchan ecotype (KDa-pisc-s) and the planktivorous Dughaill ecotype (Dug-pl), suggesting a significant history of introgressive hybridization, we also tested models describing introgression and hybrid speciation in those two systems.

In the case of Loch Dughaill and Uaine, the most likely model was the IMint model (ΔAIC to next best model = 89.32; table S12, fig. S18), similar to the other populations from the Atlantic lineage. Under this model, the oldest split between the benthivorous Dughaill ecotype and the ancestor of the planktivorous Dughaill ecotype and Uaine population occurred before the last glacial maximum, 16,690 generations ago, which is about 50,000 years with a generation time of three years. The planktivorous ecotype from Dughaill and the Loch Uaine split after the LGM around 3,954 generations ago (~12,000 years ago). We also detected a recent introgression from the benthivorous Dughaill ecotype and Uaine into the planktivorous Dughaill ecotype, with nearly complete admixture from Uaine into the planktivorous Dughaill ecotype (adm = 1.00). Only weak admixture occurred between the benthivorous and planktivorous ecotypes in Dughaill (adm = 0.003).

Both Siberian populations, Kalarskii Davatchan (KDa) and Kamkanda (Kam) most likely evolved under an evolutionary history of secondary contact scenario (SC) without significant introgression (table S12, fig. S18). This contradicts the *f3*-statistics that showed significant introgressive hybridization in KDa. However, the next best model in KDa is the IntSC model that includes introgression and is also weakly supported (ΔAIC = 5.27), so the possibility of introgressive hybridization cannot be fully excluded. In Kamkanda we cannot exclude the possibility of isolation-with-migration as the mode of divergence, as the IM model with ongoing gene flow during divergence was also highly supported compared to the SC model (ΔAIC = 0.32). In both cases, though, the oldest split most likely occurred

before the LGM (Kam: 10,716 generations; KDa: 3,783 generations). Generation times of piscivorous ecotypes in Siberian lakes are significantly higher compared to Atlantic and other Siberian ecotypes (10 years and older), translating it into divergence times of up to more than 107,160 years in Kamkanda (mean = 80,370 with a mean generation time across ecotypes of roughly 7.5 years, *(35)* and up to more than 37,830 in Kalarskii Davatchan (mean = 28,373 years with a mean generation time across ecotypes of roughly 7.5 years, *(33)*.

Genomic patterns of differentiation

We compared genomic patterns of relative differentiation (Weir & Cockerham Fst) and genetic diversity (π) across ecotype pairs to identify differences and similarities in genome-wide patterns, particularly among ecotype pairs with different evolutionary histories. Patterns of $\Delta\pi$ (π (ecotype1) - π (ecotype2)) plotted against the standardized Fst (ZFst) differed across ecotype pairs, with ecotype pairs that most likely diverged after the last glacial maximum (Awe, na Sealga) showing strong differential within-group diversity ($\Delta\pi \neq 0$) at outlier loci (ZFst > 4) compared to ecotype pairs that most likely diverged before the last glacial maximum after secondary contact (fig. S21). In Kudushkit, the pattern of genetic differentiation is intermediate and does not clearly support one mode of speciation, similar to the demographic modeling results, which suggest a recent (post-LGM) divergence (figs. S18 and S21).

Ecotype pairs that show strong support for divergence under secondary contact also show high average genome-wide differentiation with several fixed variants across the genome, ranging from a maximum of 76 fixed SNPs in Davatchan (3.33% of the genome) to a minimum of 1 fixed SNP in Kudushkit and in the planktivorous-piscivorous ecotype pair in Kamkanda (table S8). In general, the number of fixed differences is higher in the Siberian lineage (mean number of fixed SNPs = 16.4) compared to Atlantic populations (mean number of fixed SNPs = 1).

Sharing of Fst outlier loci on Contig985

In context of the outlier SNPs shared across replicates, Contig985, which is located on LG1, contained outlier loci in five out of six benthivorous-planktivorous ecotype pairs in a 4.2 Mb region, and was only significantly differentiated in one of the planktivorous-piscivorous ecotype pairs. Genes in this region are involved in a range of pathways (dre00512, dre00030, dre00052, dre00052, dre00010, dre03018, dre01230, dre04070, dre01200, dre04110) of which several are also enriched for differentially expressed genes (Fructose and mannose metabolism dre00052; Biosynthesis of amino acids dre01230 and Glycolysis / Gluconeogenesis dre00010, table S16) or modules of co-expressed genes (Phosphatidylinositol signaling system dre04070, table S19).

Correlation of phenotype, genotype and environment

To gain insights into the potential drivers of phenotypic and genetic divergence among sympatric ecotypes, we investigated correlations of within- and between-group genetic diversity (π, Fst), phenotypic divergence (Pst for each trait and mean across traits) and ecological opportunity (PC1 for ecosystem size). We found a significant correlation between the degree of genetic differentiation (calculated based on outlier loci, Fst above 95th quantile) and mean phenotypic divergence (mean Pst) (table S9; Pearson's r = 0.695, P = 0.038) and for pectoral fin length (Pst-PFL) and genetic differentiation (table S9; Pearson's r = 0.699, P = 0.038). Neutral genetic differentiation was also correlated with phenotypic divergence, but less strongly and was not statistically significant after correction for multiple testing. Variation in the mean phenotypic divergence among sympatric ecotypes may be explained by variation in ecosystem size, a proxy for ecological opportunity (*24*), although the correlation was not significant (Pearson's r = 0.609, p = 0.077; table S9). However, our dataset was not specifically tested for this kind of analysis and these results should be observed as hypothesis creation rather than strict hypothesis testing.

Signals of selection and ecotype-association

As a complementary approach to the Fst outlier scans, we also identified loci under selection in each evolutionary lineage and loci associated with ecotype. 599 (out of 6,039) and 666 (out of 4,475) SNPs

were identified to being under positive selection in the Atlantic and Siberian lineage, respectively (fig. S20). These loci were associated with the first 10 and 14 axes of genetic variation in the Atlantic and Siberian, respectively, and were under selection within and across populations. Loci under selection were widely distributed across the genome in both lineages, although some clusters could be observed (fig. S20). Only 18 loci under selection were shared across lineages (expected number of shared loci by chance: mean = 14, 95%-CI = 8 - 22). Contig985 also contained loci (but not the same SNPs) under selection in both lineages. Genes associated with loci under selection were involved in a wide range of different biological functions (table S21), but none of the GO-terms or KEGG pathways were significantly enriched. Cell differentiation (GO:0030154) is the only biological process (GO-term) that was shared among lineages, although similar functions were under selection, such as cellular and developmental processes, neural development or signalling pathways.

Furthermore, we performed genome-wide association analysis using LMM and BSLMM approaches based on lineage-specific imputed SNP datasets. Correlation between non-imputed and imputed SNP datasets (allele frequencies for SNPs with more than 20% missing data) of 0.92 – 0.97 (>0.98 for all SNPs), suggests the imputation did not introduce a bias in allele frequencies. We identified a total of 35 and 228 candidate loci that were associated with ecotype in the Atlantic and Siberian lineage, respectively (fig. S22). 8 (Atlantic) and 11 (Siberia) of those loci were identified using both methods, the univariate LMM and the BSLMM. Genes associated with ecotype in the Siberian lineage (table S20) were significantly enriched for genes involved in habenula (GO:0021986) and epithalamus (GO:0021538) development (table S18, FDR < 0.1). Genes involved in ECM-receptor interaction (dre04512) were co-localised with loci under selection and loci associated with ecotype divergence in the Siberian lineage. None of the pathways in Atlantic populations were significantly enriched (table S20) but for example the adrenergic signaling in cardiomyocytes (dre04261) KEGG pathway was under selection in Siberian populations (table S20 and S21).

Two and 51 of the candidate loci associated with ecotype were also detected to be under selection in the Atlantic or Siberian population, respectively, and even though none of the candidate loci were directly shared across lineages, we found that Contig1512 contained loci associated with ecotype and under selection in both lineages. Genes located on this contig were significantly enriched for functions related to the ErbB signalising pathway (dre04012). ErbB receptor tyrosine kinases link the binding of extracellular growth factor ligands to intracellular pathways, regulating processes such as cell proliferation, differentiation or apoptosis (*97*), and these pathways are suggested to play an important role in skeletogenesis (*98*).

Genetic architecture underlying ecotype divergence

Hyper-parameters inferred using the BLSMM were used to investigate the genetic architecture underlying ecotype evolution in the Atlantic and Siberian lineages (table S22). We found that a large proportion of phenotypic variation is explained by the genotype (PVE), ranging from 0.9676 (95%-CI: 0.8617-0.9999) in the Atlantic lineage to 0.9992 (95%-CI: 0.9967 - 0.9999) in the Siberian lineage. However, in the Atlantic lineage only 63.1% of PVE was explained by 51 sparse (major-) effect loci (n.gamma), whereas in Siberia nearly all of the PVE (PGE = 96.8%) was explained by loci with sparse effect (n.gamma = 91). Overall, large proportions of ecotype divergence were explained by genotype in both lineages.

RNA-seq data processing and alignment.

79.17 % ± 0.747 (mean ± SEM) of RNA-seq reads could be uniquely mapped to the Artic charr draft genome. The majority of the unmapped reads could not be mapped because they were too short after the trimming step. In total (with multiple mapping) between 15 and 37 million reads per individual were mapped.

We detected a wide range in the extent of differential gene expression between ecotypes, from complete separation in the PCA (i.e. Dughaill in fig. S24) up to relatively strong overlap among ecotypes (i.e. in Awe; fig. S24). In Kamkanda, we find a similar relationship between ecotypes based on gene expression compared to SNP data, with the benthivorous and piscivorous ecotypes being more similar to

each other compared to the planktivorous ecotype, which splits of along PC1 from both other ecotypes. Ecotype pairs from different populations also showed very high overlap in differentially expressed genes (DEGs) and nearly all pairwise comparisons were significant (table S14). The only two non-significant pairwise comparisons were across non-parallel ecotypes pairs, comparing the benthivorous – piscivorous ecotype pair from Kamkanda to the benthivorous – planktivorous ecotypes from Tay and Awe. We further found a range of differentially expressed genes that were shared between three (122 DEGs), four (19 DEGs) and five (4 DEGs) out of seven ecotype pairs (table S15 and S18). Out of the 19 DEGs that were differentially expressed in four ecotype pairs, seven transcripts belonged to different haemoglobin alpha and beta subunits (fig. S26). These *HBA* and *HBB* subunits were all expressed in the same direction between ecotypes within populations but not always in the same direction across ecotypes. Other shared genes include heat shock proteins *HSP16*.1, a growth inhibitor (as-related and estrogen-regulated growth inhibitor; Fig. 3E) or mid1-interacting protein 1-B-like (table S18). Four genes were shared across five ecotype pairs, of which three were characterized (Fig. 3C, table S18). The fourth transcript belonged to an uncharacterized gene (table S18). In general, these genes showed comparable expression patterns across populations.

Furthermore, we detected 13 modules containing co-expressed genes that were also correlated with ecotype (benthivorous or planktivorous) across populations (fig. S25, table S16). Genes within four modules were enriched for a range of different functional pathways (KEGG; table S19) related to growth and development (TGF-beta signalling pathway, ECM-receptor interaction) or metabolism (Glycerolipid metabolism).

References

1. S.C. Morris, Life's solution: inevitable humans in a lonely universe (Cambridge University Press, 2003).

2. S.J. Gould, Wonderful life: the Burgess Shale and the nature of history (WW Norton & Company, 1990).

3. K.R. Elmer, A. Meyer, Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution* **26**, 298–306 (2011).

4. R. Kaeuffer, C. Peichel, D. Bolnick, A. Hendry, Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution* **66**, 402–18 (2012).

5. L. Mahler, T. Ingram, L. Revell, J. Losos, Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. *Science* **341**, 292–295 (2013).

6. K. R. Elmer, S. Fan, H. Kusche, M. Spreitzer, A. Kautt, P. Franchini, A. Meyer, Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat Commun* **5**, 5168 (2014).

7. Perrier C. Perrier, V. Bourret, M. Kent, L. Bernatchez, Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon. *Mol Ecol* **22**, 5577–93 (2013).

8. J. Kowalko, N. Rohner, T. Linden, S. Rompani, W. Warren, R. Borowsky, C. Tabin, W. Jeffery, and M. Yoshizawa, Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of Astyanax mexicanus. *Proc National Acad Sci* **110**, 16933–16938 (2013).

9. D. Schluter, The ecology of adaptive radiation (OUP Oxford, 2000).

10. J.A. Endler, Natural selection in the wild (No. 21. Princeton University Press, 1986).

11. K. B. Oke, G. Rolshausen, C. LeBlond, A. P. Hendry, How Parallel Is Parallel Evolution? A Comparative Analysis in Fishes. *Am Nat* **190**, 1–16 (2017).

12. Y. E. Stuart T. Veen, J. Weber, D. Hanson, M. Ravinet, B. Lohman, C. Thompson, T. Tasneem, A. Doggett, R. Izen, N. Ahmed, R. Barrett, A. Hendry, C. Peichel, and D. Bolnick, Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat Ecol Evol* **1**, 0158 (2017).

13. G. Conte, M. Arnegard, C. Peichel, D. Schluter, The probability of genetic parallelism and convergence in natural populations. *Proceedings. Biological sciences* **279**, 5039–47 (2012).

14. G. Conte, M. Arnegard, J. Best, Y. Chan, F. Jones, D. Kingsley, D. Schluter, and C. Peichel, Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. *Genetics* **201**, 1189–200 (2015).

15. D. Schluter, A. Rambaut, Ecological Speciation in Postglacial Fishes [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **351**, 807–814 (1996).

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

17. S.S. Alekseyev, V.P. Samusenok, A.N. Matveev, M.Y. Pichugin, Diversification, sympatric speciation, and trophic polymorphism of Arctic charr, Salvelinus alpinus complex, in Transbaikalia. *Environmental Biology of Fishes* **22**, 97–114 (2002).

18. C.E. Adams, F.A. Huntingford, Inherited differences in head allometry in polymorphic Arctic charr from Loch Rannoch, Scotland. *Journal of Fish Biology* **60**, 515–520 (2002).

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

20. O. Hooker, J. Barry, T. E. Van Leeuwen, A. Lyle, J. Newton, P. Cunningham, C. E. Adams, Morphological, ecological and behavioural differentiation of sympatric profundal and pelagic Arctic charr (Salvelinus alpinus) in Loch Dughaill Scotland. *Hydrobiologia* **783**, 209–221 (2015).

21. C. E. Adams, F. A. Huntingford, The functional significance of inherited differences in feeding morphology in a sympatric polymorphic population of Arctic charr. *Evol Ecol* **16**, 15–25 (2002).

22. R. Riesch, M. Muschick, D. Lindtke, R. Villoutreix, A. Comeault, T. Farkas, K. Lucek, E. Hellen, V. Soria-Carrasco, S. Dennis, C. de Carvalho, R. Safran, C. Sandoval, J. Feder, R. Gries, B. Crespi, G. Gries, Z. Gompert, and P. Nosil, Transitions between phases of genomic differentiation during stick-insect speciation. *Nat Ecol Evol* **1**, 0082 (2017).

23. P. Nosil, S. Egan, D. Funk, Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**, 316–336 (2008).

24. H. Recknagel, O. Hooker, C. E. Adams, K. R. Elmer, Ecosystem size predicts eco-morphological variability in a postglacial diversification. *Ecology and Evolution* **7**, 5560–5570 (2017).

25. L. Excoffier, I. Dupanloup, E. Huerta-Sánchez, V. Sousa, M. Foll, Robust Demographic Inference from Genomic and SNP Data. *PloS Genetics* **9**, e1003905 (2013).

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

27. N. V. Gordeeva, S. S. Alekseyev, A. N. Matveev, V. P. Samusenok, Parallel evolutionary divergence in Arctic char Salvelinus alpinus complex from Transbaikalia: variation in differentiation degree and segregation of genetic diversity among sympatric forms. *Can J Fish Aquat Sci* **72**, 96–115 (2015).

28. Y. Morbey, P. Couture, P. Busby, B. Shuter, Physiological correlates of seasonal growth patterns in lake trout Salvelinus namaycush. *Journal of Fish Biology* **77**, 2298–2314 (2010).

29. A. Palstra, M. Rovira, D. Rizo-Roca, J. Torrella, H. Spaink, and J. Planas, Swimming-induced exercise promotes hypertrophy and vascularization of fast skeletal muscle fibres and activation of myogenic and angiogenic transcriptional programs in adult zebrafish. *BMC Genomics* **15**, 1–20 (2014).

30. S. Flanagan, V. Dũng, J. Houghton, E. Franco, C. Ngoc, A. Damhuis, F. Ashcroft, L. Harries, and S. Ellard, An ABCC8 Nonsense Mutation Causing Neonatal Diabetes Through Altered Transcript Expression. *J Clin Res Pediatric Endocrinol* **9**, 260–264 (2017).

31. P.C. Brunner, M.R. Douglas, A. Osinov, C.C. Wilson, L. Bernatchez, Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution* **55**, 573–586 (2001).

32. C.E. Adams D. Fraser; I. McCarthy; S. Shields; S. Waldron; G. Alexander, Stable isotope analysis reveals ecological segregation in a bimodal size polymorphism in Arctic charr from Loch Tay, Scotland. *Journal of Fish Biology* **62**, 474–481 (2003).

33. S.S. Alekseyev, M. Mina, E. Smirina, A. Sokolov, Late ontogeny growth acceleration and size form transformations in Transbaikalian Arctic charr, *Salvelinus alpinus* complex: evidence from fin ray cross section growth layers. *Environmental Biology of Fishes* **86**, 487–505 (2009).

34. S.S. Alekseyev, N.V. Gordeeva, V.P. Samusenok, A.N. Matveev, R.S. Andreev, A.L. Yur'ev, E.M. Smirina, Extant and extinct forms of arctic charr *Salvelinus alpinus* (L.) complex from the Leprindo lake system (Transbaikalia): Differentiation in life history, morphology, and genetics. *Journal of Ichthyology* **53**, 792–803 (2013).

35. S.S. Alekseyev, N.V. Gordeeva, A.N. Matveev, V.P. Samusenok, A.I. Vokin, A.L. Yur'ev, Three sympatric forms of Arctic charr *Salvelinus alpinus* complex (Salmoniformes, Salmonidae) from Lake Kamkanda, Northern Transbaikalia. *Journal of Ichthyology* **54**, 384–408 (2014).

36. Yu. M. Pichugin, Development of skeleton elements in juvenile dwarf and small sympatric forms of *Salvelinus alpinus* complex from Lake Davatchan (Transbaikalia). *Journal of Ichthyology* **49**, 741-758 (2009).

37. C.E. Adams, D. Fraser, A.J. Wilson, G. Alexander, M.M. Ferguson, S. Skulason, Patterns of phenotypic and genetic variability show hidden diversity in Scottish Arctic charr. *Ecology of Freshwater Fish* **16**, 78–86 (2007).

38. C. Bryce, A. Fraser, R. Knudsen, R. Greer, C.E. Adams, Divergent functional traits in three sympatric Arctic charr Salvelinus alpinus morphs are not coupled with the age of the lineage divergence. *Hydrobiologia* **783**, 177–189 (2016).

39. D. Pavlov, A. Osinov, Reproduction and development in the dwarf form of Arctic charr *Salvelinus alpinus* from Lake Davatchan, Transbaikalia. *Journal of Ichthyology* **48**, 96–113 (2008).

40. C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, **9**, 671–675 (2012).
41. A. Siwertsson, R. Knudsen, C.E. Adams, K. Præbel, P.A. Amundsen, Parallel and non-parallel morphological divergence among foraging specialists in European whitefish (*Coregonus lavaretus*). *Ecology and Evolution* **3**, 1590–602 (2013).

42. S. Skoglund, A. Siwertsson, P.A. Amundsen, R. Knudsen, Morphological divergence between three Arctic charr morphs – the significance of the deep-water environment. *Ecology and Evolution* **5**, 3114–3129 (2015).

43. B.R. Langerhans, T.J. DeWitt, Shared and unique features of evolutionary diversification. *The*

American Naturalist **164**, 335–49 (2004).

44. F. Rohlf, TpsDig, Department of Ecology and Evolution, State University of New York, Stony Brook, NY (2004).

45. D.C. Adams, E. Otárola-Castillo, geomorph: an r package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* **4**, 393–399 (2013).

46. D.C. Adams, M. Collyer, A general framework for the analysis of phenotypic trajectories in evolutionary studies. *Evolution* **63**, 1143–1154 (2009).

47. R. Knudsen, A. Klemetsen, F. Staldvik, Parasites as indicators of individual feeding specialization in Arctic charr during winter in northern Norway. *Journal of Fish Biology* **48**, 1256–1265 (1996).

48. C.M. Nugent, A.A. Easton, J.D. Norman, M.M. Ferguson, R.G. Danzmann, A SNP Based Linkage Map of the Arctic Charr (*Salvelinus alpinus*) Genome Provides Insights into the Diploidization Process After Whole Genome Duplication. *G3-Genes Genomes Genetics* **7**, 543–556 (2016).

49. S. Lien, B.F. Koop, S.R. Sandve, J.R Miller, M.P. Kent, T. Nome, T.R. Hvidsten, J.S. Leong, D.R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R.A. Hermansen, K. von Schalburg, E.B. Rondeau, A. Genova, J.K.A. Samy, J. Vik, M.D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. Våge, P. de Jong, T. Moen, M. Baranski, Y. Palti, D.R. Smith, J.A. Yorke, A.J. Nederbragt, A. Tooming-Klunderud, K.S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D.A. Liberles, R. Vidal, P. Iturra, S.J.M. Jones, I. Jonassen, A. Maass, S.W. Omholt, W.S. Davidson, The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–205 (2016).

50. G. Tamazian, P. Dobrynin, K. Krasheninnikova, A. Komissarov, K.P. Koepfli, S.J. O'Brien, Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. *GigaScience* **5**, 38 (2016).

51. H. Tang, X. Zhang, C. Miao, J. Zhang, R. Ming, J. Schnable, P. Schnable, E. Lyons, and J. Lu, ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biology* **16**, 3 (2015).

52. J. Keilwagen, M. Wenk, J. Erickson, M. Schattat, J. Grau, and F. Hartung, Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research* **44**, e89–e89 (2016).

53. A. Dobin, C. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29.1,** 15-21 (2013).

54. S. Altschul, T. Madden, A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402 (1997).

55. O. Tange, Gnu parallel-the command-line power tool. *The USENIX Magazine* **36.1**, 42-47 (2011).

56. F.A. Simão, R. Waterhouse, P. Ioannidis, E. Kriventseva, E. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
57. K. Howe, M. Clark, C. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. Collins, S. Humphray, K. McLaren, L. Matthews, The zebrafish reference genome sequence and its relationship to the human genome. *Nature 496* (2013).

58. D. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157 (2015).

59. T. Schenekar, E. Lerceteau-Köhler, S. Weiss, Fine-scale phylogeographic contact zone in Austrian brown trout Salmo trutta reveals multiple waves of post-glacial colonization and a pre-dominance of natural versus anthropogenic admixture. *Conservation Genetics* **15**, 561–572 (2014).

60. R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).

61. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870–1874 (2016).

62. J.W. Leigh, D. Bryant, popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110–1116 (2015).

63. J. Catchen, A. Amores, P. Hohenlohe, W. Cresko, and J. Postlethwait, Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics* **1**, 171–182 (2011).

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

65. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

66. P. Danecek, A. Auton, G. Abecasis, C. Albers, E. Banks, M. DePristo, R. Handsaker, G. Lunter, G. Marth, and S. Sherry, The variant call format and VCFtools. *Bioinformatics* **27** (2011).

67. C. Chang, C. Chow, L. Tellier, S. Vattikuti, S. Purcell, and J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 1–16 (2015).

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

69. T. Jombart, adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

70. T. Jombart, I. Ahmed, adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27** (2011).

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

72. P.G. Meirmans, P.H. Tienderen, genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792–794 (2004).

73. B. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).

74. D.H. Huson, D. Bryant, Estimating phylogenetic trees and networks using SplitsTree 4. Manuscript in

preparation, software available from www. splitstree. org (2005).

75. M. Malinsky, E. Trucchi, D. Lawson, D. Falush, RADpainter and fineRADstructure: population inference from RADseq data. *BioRxiv* (2018) doi**:** https://doi.org/10.1101/057711.

76. J.K. Pickrell, J.K. Pritchard, Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* **8**, e1002967 (2012).

77. R. Gutenkunst, R. Hernandez, S. Williamson, C. Bustamante, Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics* **5**, e1000695 (2009).

78. A. Kautt, G. Machado-Schiaffino, A. Meyer, Multispecies Outcomes of Sympatric Speciation after Admixture with the Source Population in Two Radiations of Nicaraguan Crater Lake Cichlids. *PLoS Genetics* **12**, e1006157 (2016).

79. K. Luu, E. Bazin, M.G. Blum, pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* **17**, 67–77 (2017).

80. X. Zhou, P. Carbonetto, M. Stephens, Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLoS Genetics* **9**, e1003264 (2013).

81. D. Money, K. Gardner, Z. Migicovsky, H. Schwaninger, G.-Y. Zhong, S. Myles, LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. *G3: Genes, Genomes, Genetics* **5**, 2383–2390 (2015).

82. P.J. Bradbury, Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).

83. A. Brelsford, D.P.L. Toews, D.E. Irwin, Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proc. R. Soc. B* ***284***, 20171106 (2017).

84. J. Wang, S. Vasaikar, Z. Shi, M. Greer, and B. Zhang, WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Research* **45**, W130–W137 (2017).

85. A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, (2014).

86. S. Anders, P. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, (2015).

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

88. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 1–13 (2008).

89. G.E. Hoffman, E.E. Schadt, variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).

90. A.M. Yip, S. Horvath, Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* **8**, 1–14 (2007).

91. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24** (2007).

92. A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, (2005).

93. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* **27** (1999).

94. G. McVean, A Genealogical Interpretation of Principal Components Analysis. *PLoS Genetics* **5**, e1000686 (2009).

95. C.H. Martin, J.S. Cutler, J.P. Friel, C. Touokong, G. Coop, P.C. Wainwright, Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution* **69**, 1406–1422 (2015).

96. K. Lambeck, Glacial rebound of the British Isles—II. A high-resolution, high-precision model. *Geophysical Journal International* **115**, 960–990 (1993).

97. R.N. Jorissen, F. Walker, N. Pouliot, T. Garrett, C.W. Ward, A.W. Burgess, Epidermal growth factor receptor: mechanisms of activation and signalling. *Experimental Cell Research* **284**, 31–53 (2003).

98. E. Ahi, Signalling pathways in trophic skeletal development and morphogenesis: Insights from studies on teleost fish. *Developmental Biology* **420**, 11–31 (2016).

99. H.M. Gunter, S. Fan, F. Xiong, P. Franchini, C. Fruciano, A. Meyer, Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. *Molecular Ecology* **22**, 4516-4531.

100. M. Carruthers, A. Yurchenko, J. Augley, C. Adams, P. Herzyk, and K.R. Elmer, De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species, *Bmc Genomics* **19**, 32 (2018).

Acknowledgements: N.V. Alekseyeva, R.S. Andreev, A.A. Kalinina, V.S. Khlystov, I.G. Khoro, I.B. Knizhin (deceased), A. Loshakov, A.N. Matveev, A.G. Osinov, M. Yu Pichugin, V.K. Pomazkin, V.V. Pulyarov, V.P. Samusenok, D.V. Schepotkin, F.N. Shkil', A.A. Sokolov (deceased), S.D. Sviridov, A.I. Vokin, A.L. Yur'ev, I.I Yur'ev, provided assistance with sampling, D.A. Akzhigitova, M. Yu Pichugin, A.A. Sokolov provided assistance with morphometric measures. We thank A. Adam for technical support in the lab, P.C Robinson for generating some sequence data, and Glasgow Polyomics for NGS. **Funding**: this work was supported by funding from the European Union's INTERREG IVA Programme (Project 2859 "IBIS") managed by the Special EU Programmes Body (CEA), BBSRC-Westbio DTP studentship (to MC with CEA and KRE) (BB/J013854/1), Wellcome Trust-Glasgow Polyomics ISSF Catalyst (KRE) (Wellcome Trust [097821/Z/11/Z]), Marie Curie CIG 321999 "GEN ECOL ADAPT" (KRE), Carnegie Turst Research Incetive Grant (70287) (KRE), Natural Sciences and Engineering Council of Canada Strategic Grant (NSERC-Strategic) (BFK), and Russian Foundation for Basic Research (project no. 18-04-00092) (SSA and NVG). **Authors contributions:** KRE, CEA conceived the project; NVG, SSA, CEA, KRE designed experiments; NVG, SSA, MC, AJ, OH, CEA collected / contributed samples; BFK, JSL, EBR generated the draft genome scaffolds; AJ, MC generated data; AJ analyzed genomic and morphology data; MC, AY analyzed transcriptomic data; AJ, AY anchored and annotated the genome; KRE, CEA, AY supervised the project; AJ, KRE drafted the manuscript; all authors contributed to and approved the final manuscript. **Data and material availability:** Sequence data are available from [with acceptance].

Supplementary Figures and Tables:

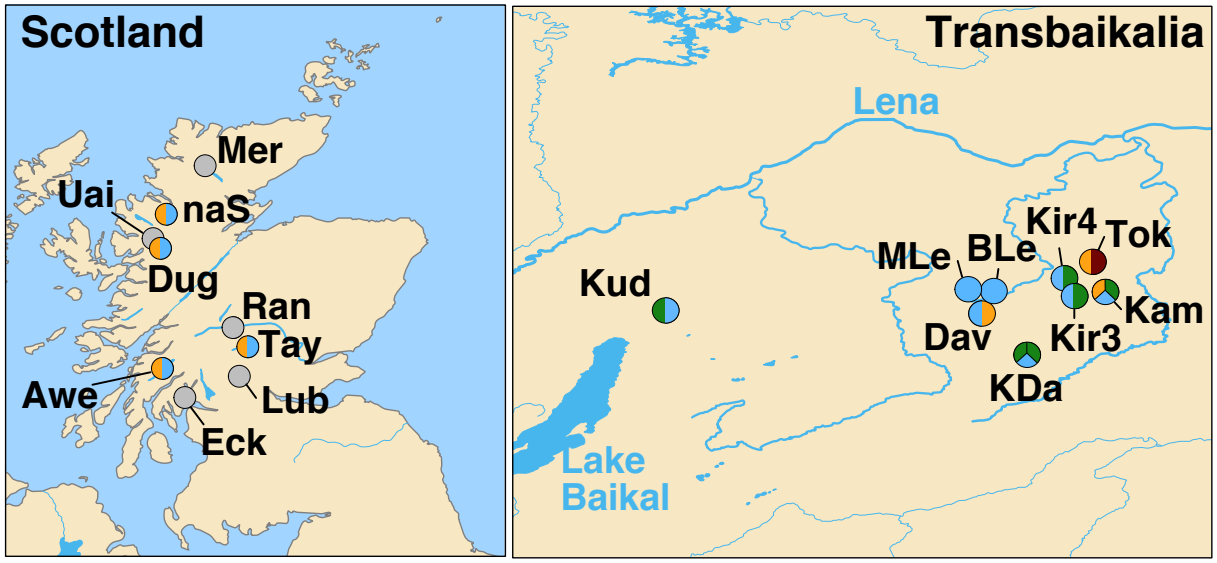


Fig. S1. Map of sampling sites in the Atlantic lineage (Scotland) and the Siberian lineage (Transbaikalia). Exact coordinates, population abbreviations and sample sizes are given in table S1. The following colour code was used: orange – benthivorous, blue – planktivorous, green – piscivorous, dark red – insectivorous, grey – unimodal-planktivorous/unassigned.

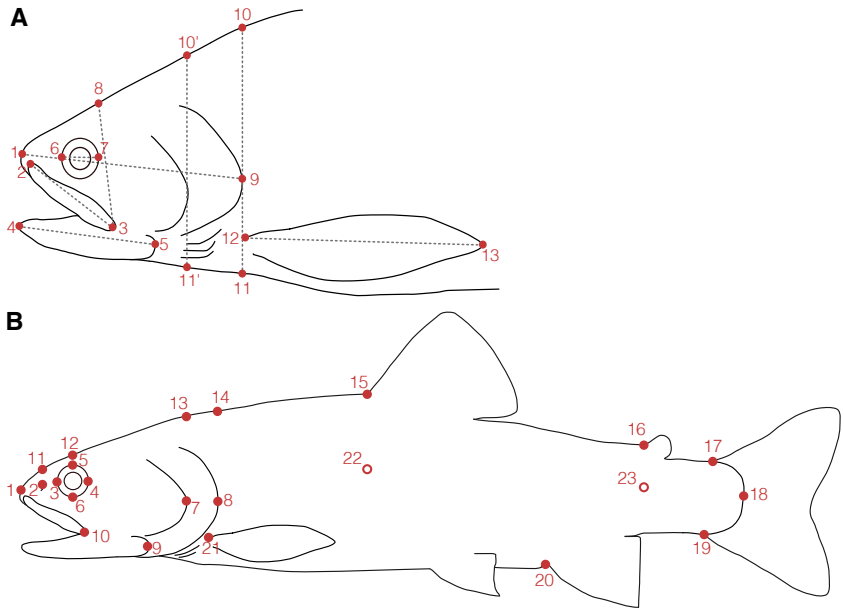


Fig. S2. Illustration of linear measurements and landmark positions for geometric morphometrics. (**A**) Linear measurements taken for phenotypic analysis: 1-9 Head length (HL), 2-3 Maxilla length (ML), 4-5 Lower jaw length (LJL), 6-7 Eye diameter (ED), 3-8 Head depth at eye (HDE), 10-11 Head depth at operculum (HDO - Atlantic lineage), 10'-11' Head depth at pre-operculum (HDO - Siberian population), 12-13 Pectoral fin length (PFL). (**B**) Landmarks used for geometric morphometric analysis: 1 – Tip of the snout, 2 – Nostril, 3-6 – Corners of the eye (opposite of each other), 7 – posterior end of the pre-operculum, 8 – posterior end of the operculum, 9 – posterior end of the lower jaw, 10 – posterior end of the maxilla, 11 – point on top of snout directly above the nostril, 12 – point on top of head directly above the eye, 13 and 14 – directly above posterior ends of pre-operculum and operculum, 15 – anterior base of dorsal fin, 16 – anterior base of adipose fin, 17 – dorsal anterior base of caudal fin, 18 – posterior end of lateral line at base of caudal fin, 19 – ventral anterior base of caudal fin, 20 – anterior base of anal fin, 21 – dorsal anterior base of pectoral fin. Landmark 22 and 23 are helper points to remove bending during the analysis and are placed on the lateral line directly below landmark 15 and 16.

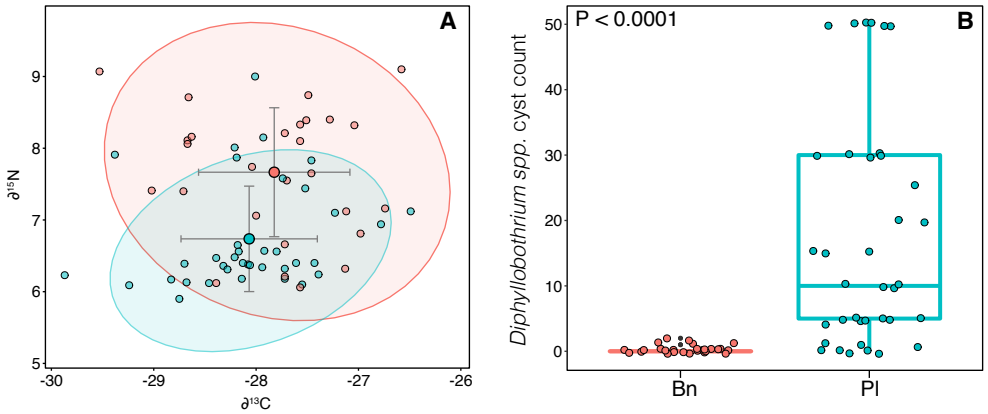


Fig. S3. Ecological divergence between ecotypes in Loch na Sealga (Benthivorous ecotypes in red; planktivorous ecotypes in green). (**A**) Ecotypes separate based on nitrogen stable isotope signatures but not based on carbon. (**B**) Ecotypes show contrasting intensities of *Diphyllobothrium spp*. infection.

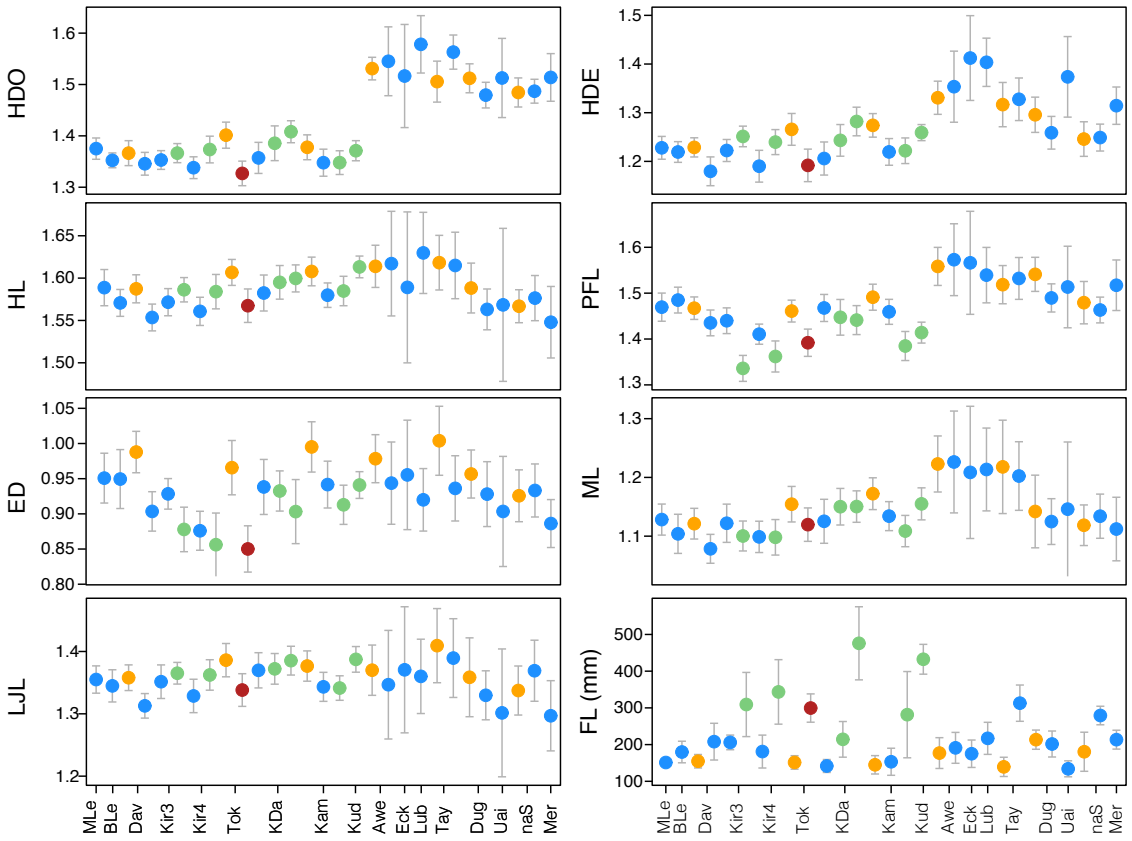


Fig. S4. Divergence of linear traits. Shown are means and standard errors (± SEM) for each linear trait for each population and ecotype. HDO =Head depth at operculum (Atlantic lineage), Head depth at pre-operculum (Siberian population), HL = head length, ED = Eye diameter, LJL = Lower jaw length, HDE = Head depth at eye, PFL = Pectoral fin length; ML=Maxilla length; FL = fork length. The colour code follows fig. S1.

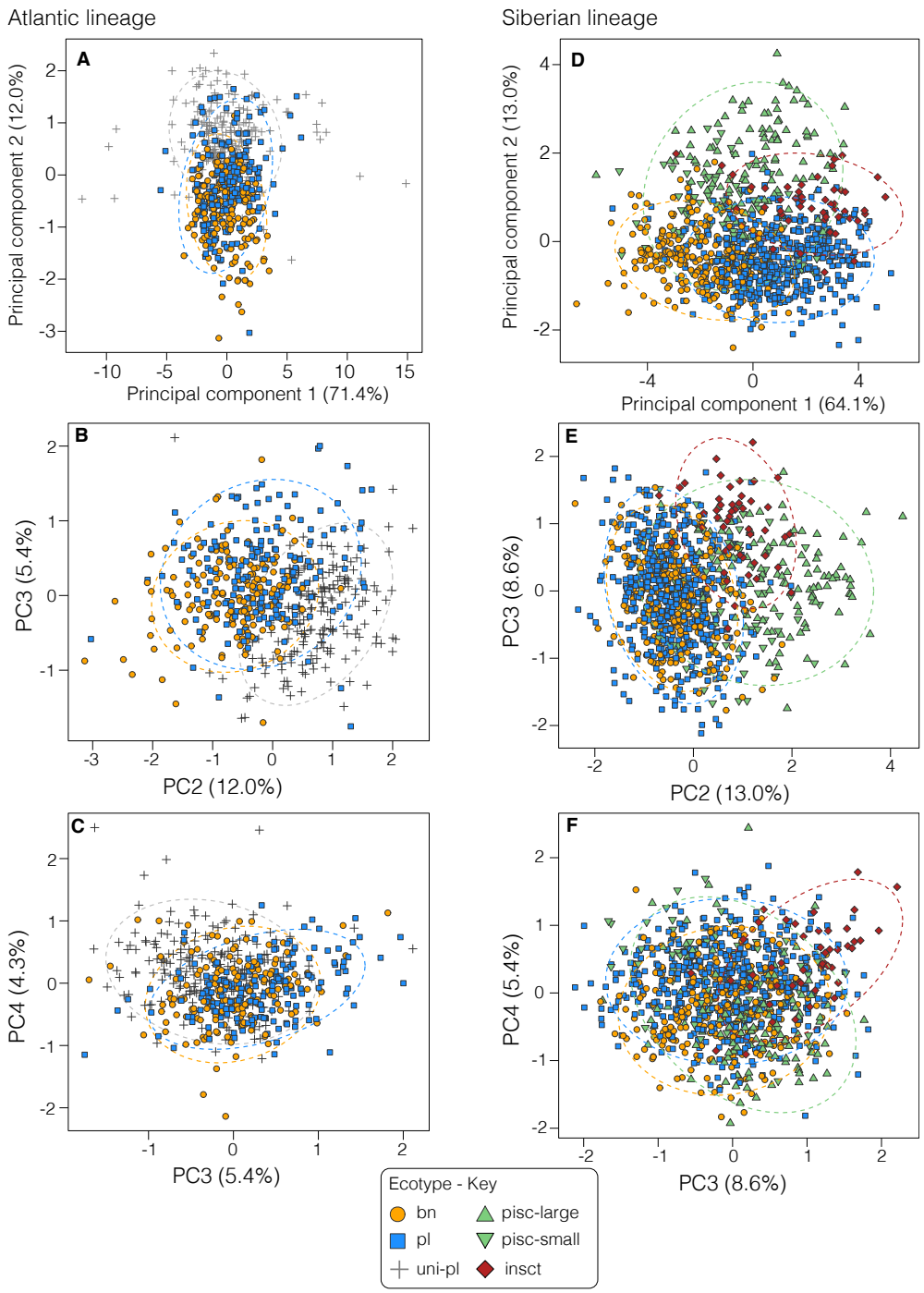


Fig. S5. Principal component plots based on linear traits by lineage. PCA plots with each dot representing a single individual for the first five eigenvectors in the Atlantic (**A-C**) and Siberian (**D-F**) lineage. Ellipses represent 95%-confidence intervals for each ecotype.

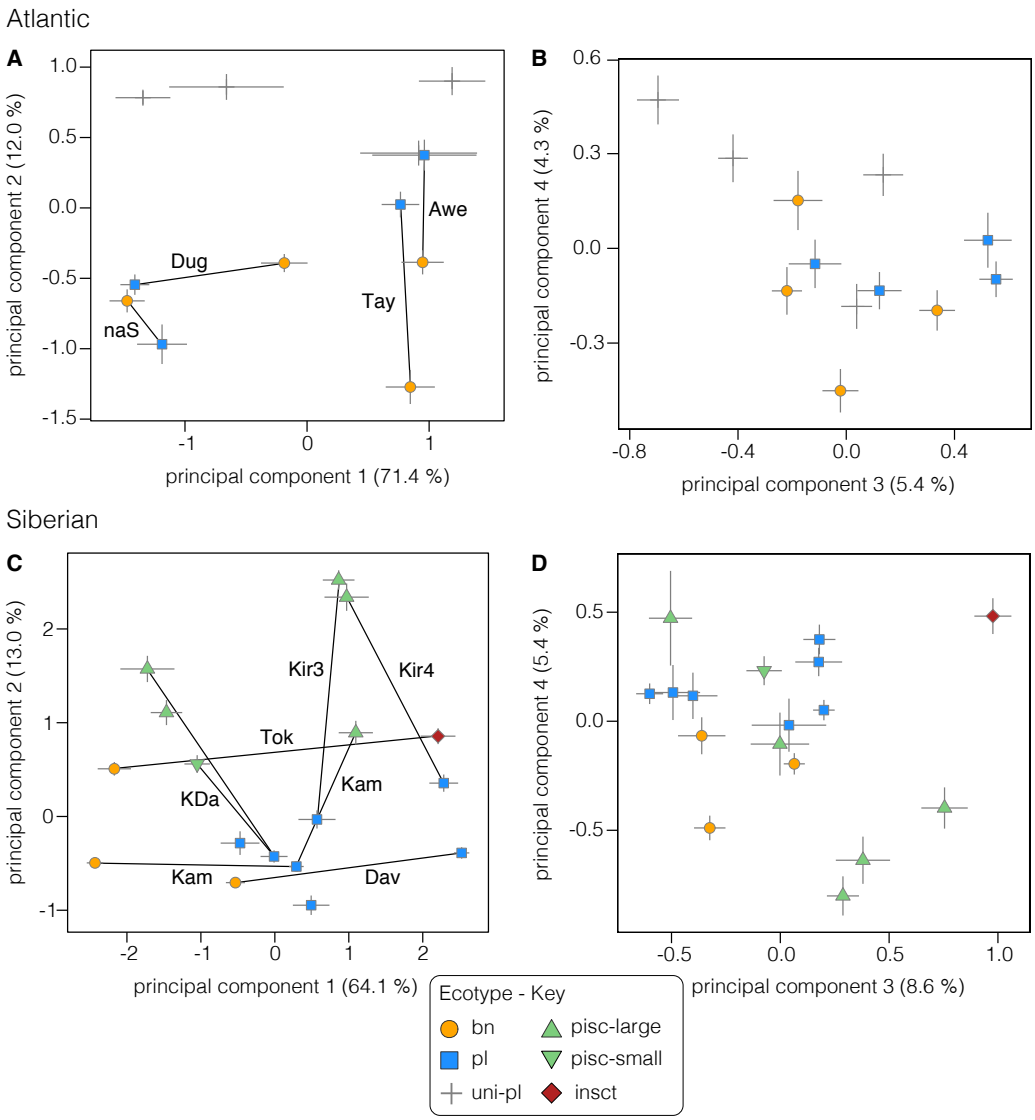


Fig. S6. Principal component plots based on linear traits. Individual dots show the centroid ± SEM for each ecotype in the Atlantic (**A,B**) and the Siberian (**C,D**) lineage. Phenotypic trajectories across the first two principal component axes are shown for each lineage (**A,C**).

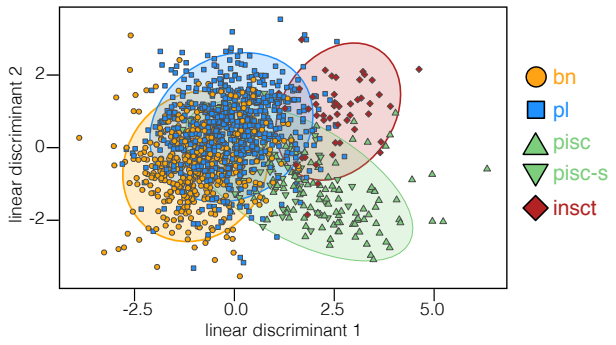


Fig. S7. Linear discriminant plot showing the two main axes of phenotypic (six linear traits) variation across both lineages. Individuals were clustered by ecotype for this analysis. Ellipses show 95%-confidence intervals.

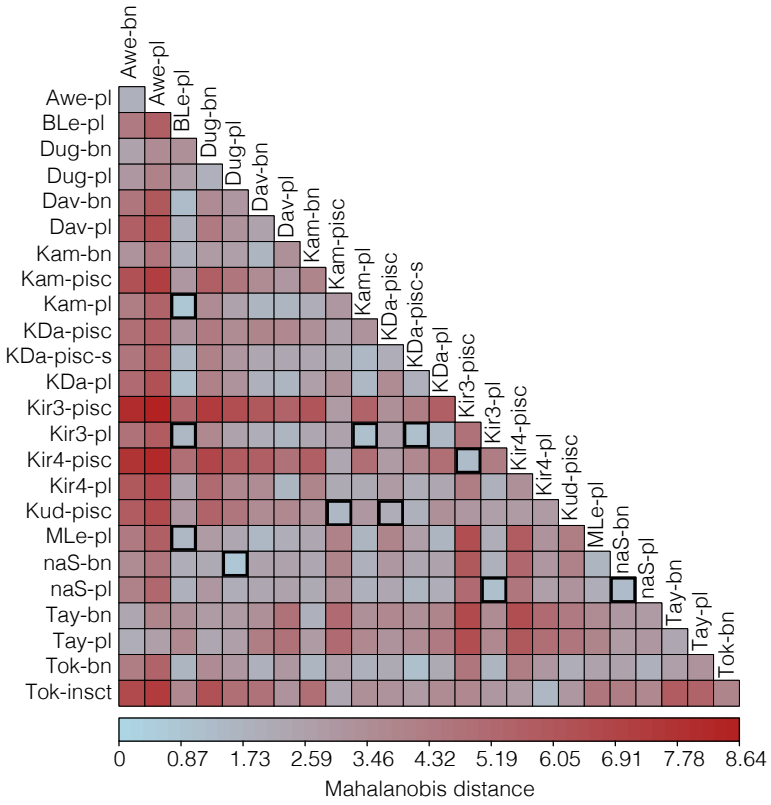


Fig. S8. Variation in phenotypic divergence among ecotypes. Mahalanobis distances were calculated using six linear traits (without HDO). Ecotypes pairs that did not significantly differ in overall phenotype (P > 0.05) are highlighted with black boxes.

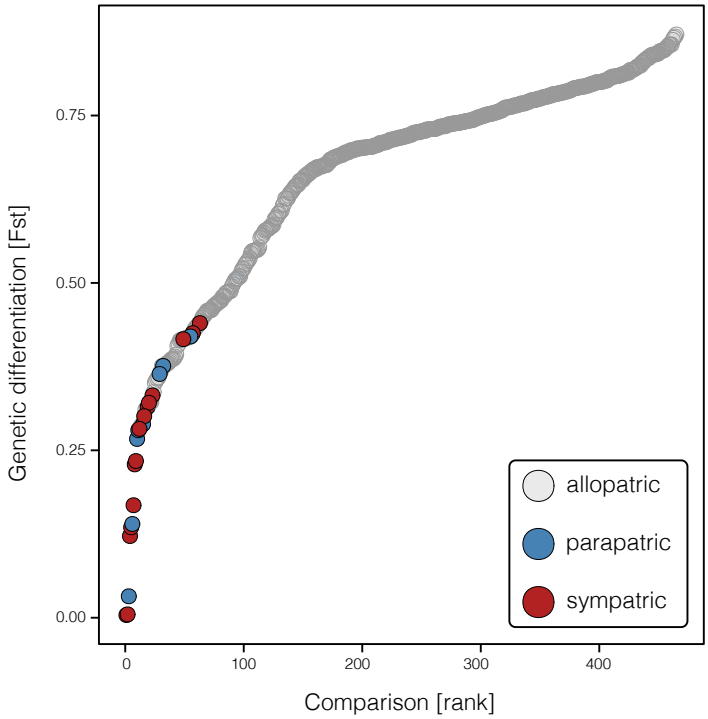


Fig. S9. Continuum of genetic differentiation across sympatric ecotype pairs (sympatric), connected populations (parapatric) or isolated populations (allopatric) within and across lineages. It should be noted that some sympatric ecotype pairs show higher levels of genetic divergence than parapatric or allopatric populations.

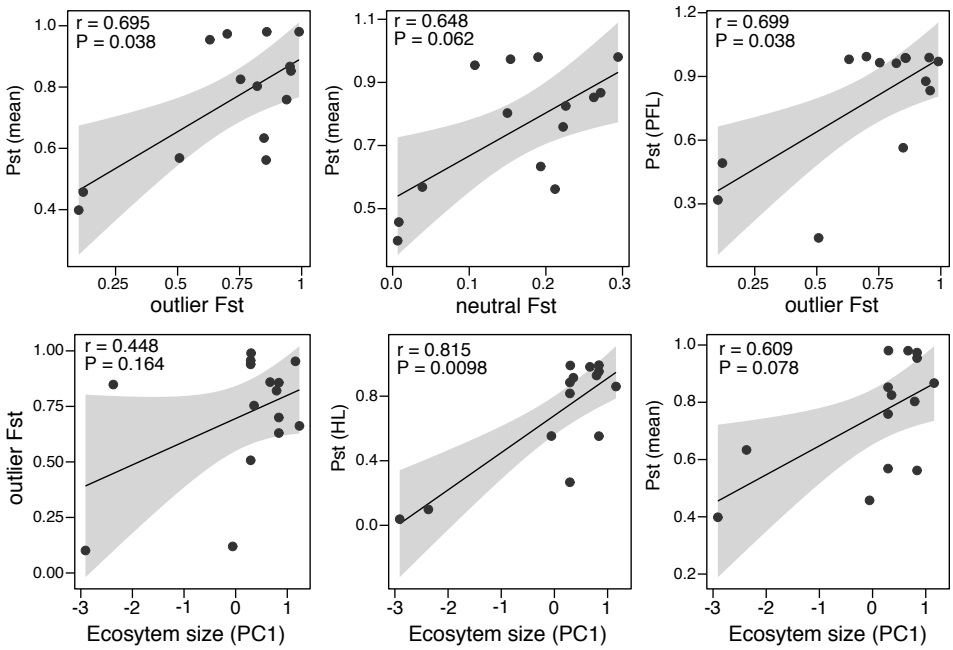


Fig. S10. Correlations between phenotypic differentiation (Pst), genetic differentiation (Fst) and ecosystem size (PC1). These correlations show correlations (although not all significant) between phenotypic differentiation and genetic differentiation, and between phenotypic

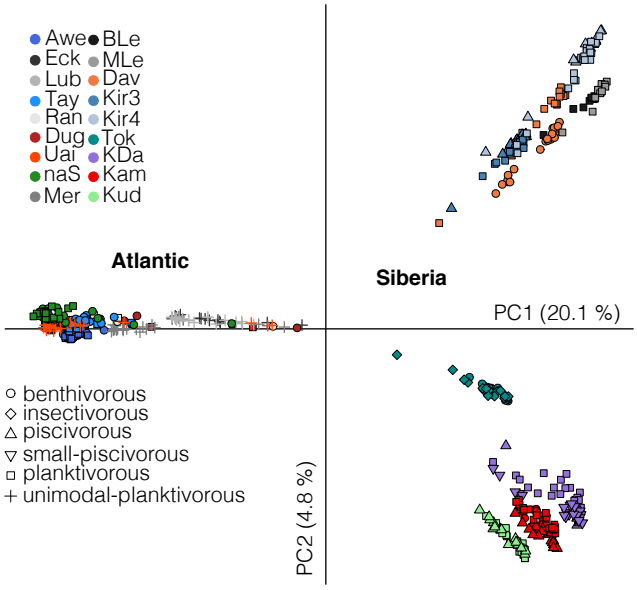


Fig. S11. Principal component plot based 12,215 SNPs for the Atlantic and Siberian lineage (N = 648). Populations split between lineages along PC1, and the Siberian lineage separates by catchment along PC2. Ecotypes are illustrated by shape and populations by color.

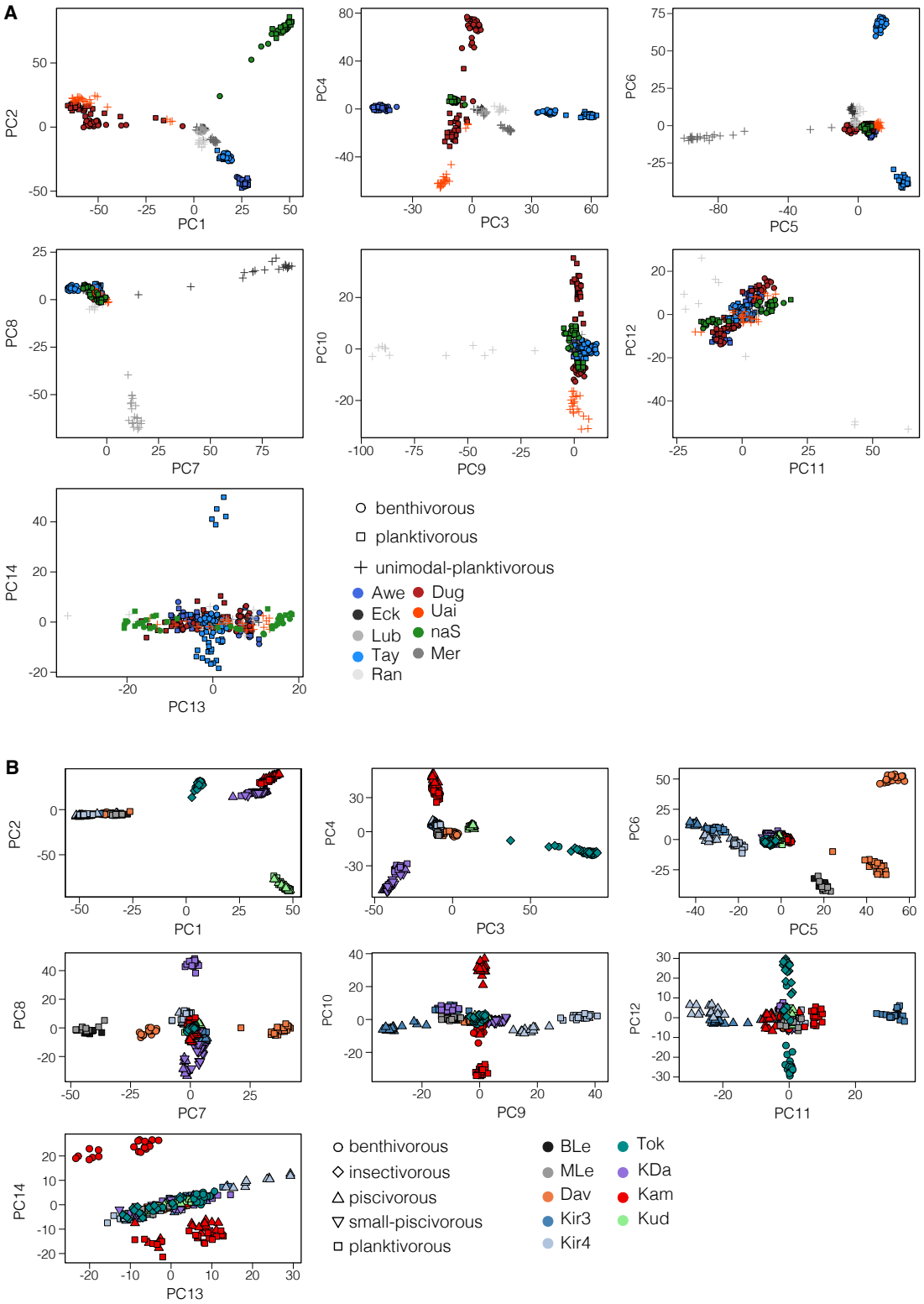


Fig. S12. Major axes of genetic variation. Principal component plots displaying the major axes of genetic variation in the Atlantic (**A**) and Siberian (**B**) lineage. Note that ecotype pairs usually split along different principal component axes but that some populations and ecotypes

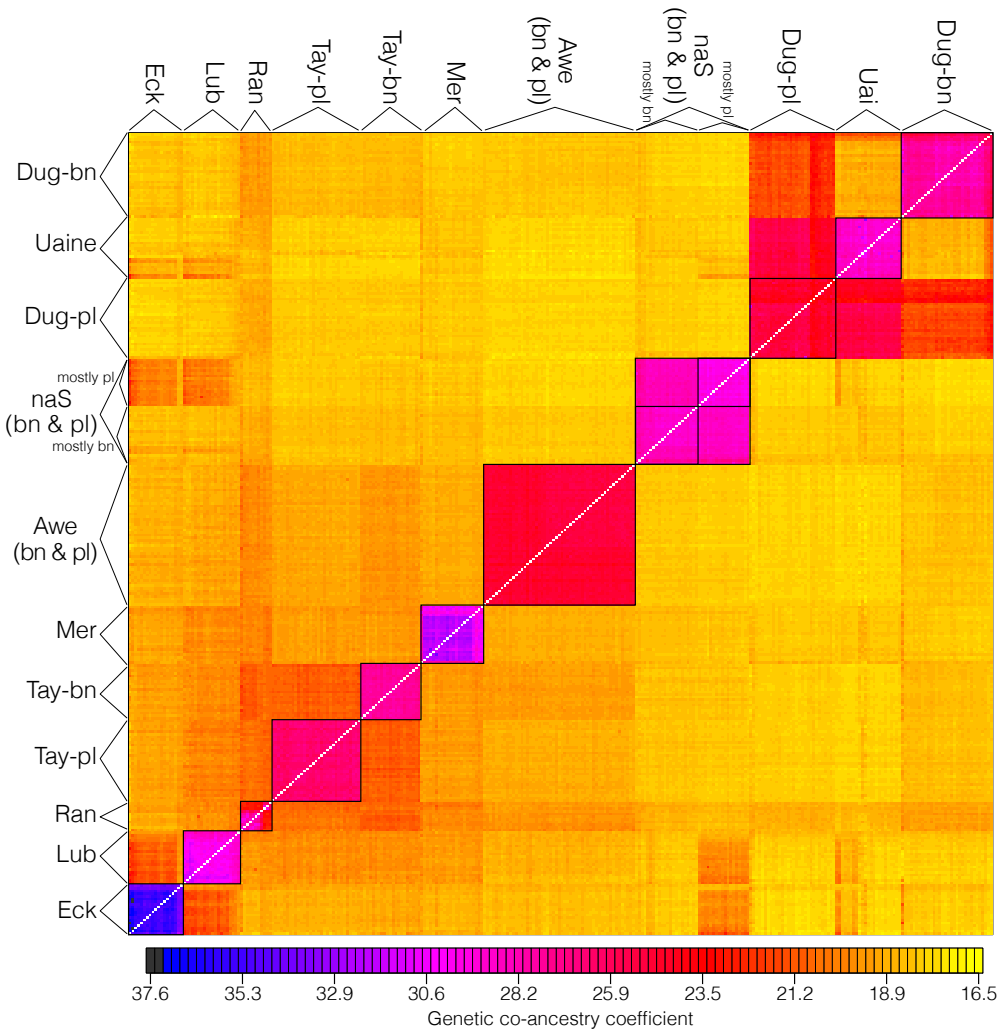


Fig. S13. Genetic co-ancestry across ecotypes and populations in the Atlantic lineage based on haplotype information. Individuals were clustered using *fineRADstructure* based on genetic co-ancestry. Black boxes highlight populations and ecotypes within populations. The hgiher genetic co-ancestry between the planktivorous Dughaill ecotype (Dug-pl) and the Uaine population compared to the sympatric benthivorous ecotype is clearly visible.

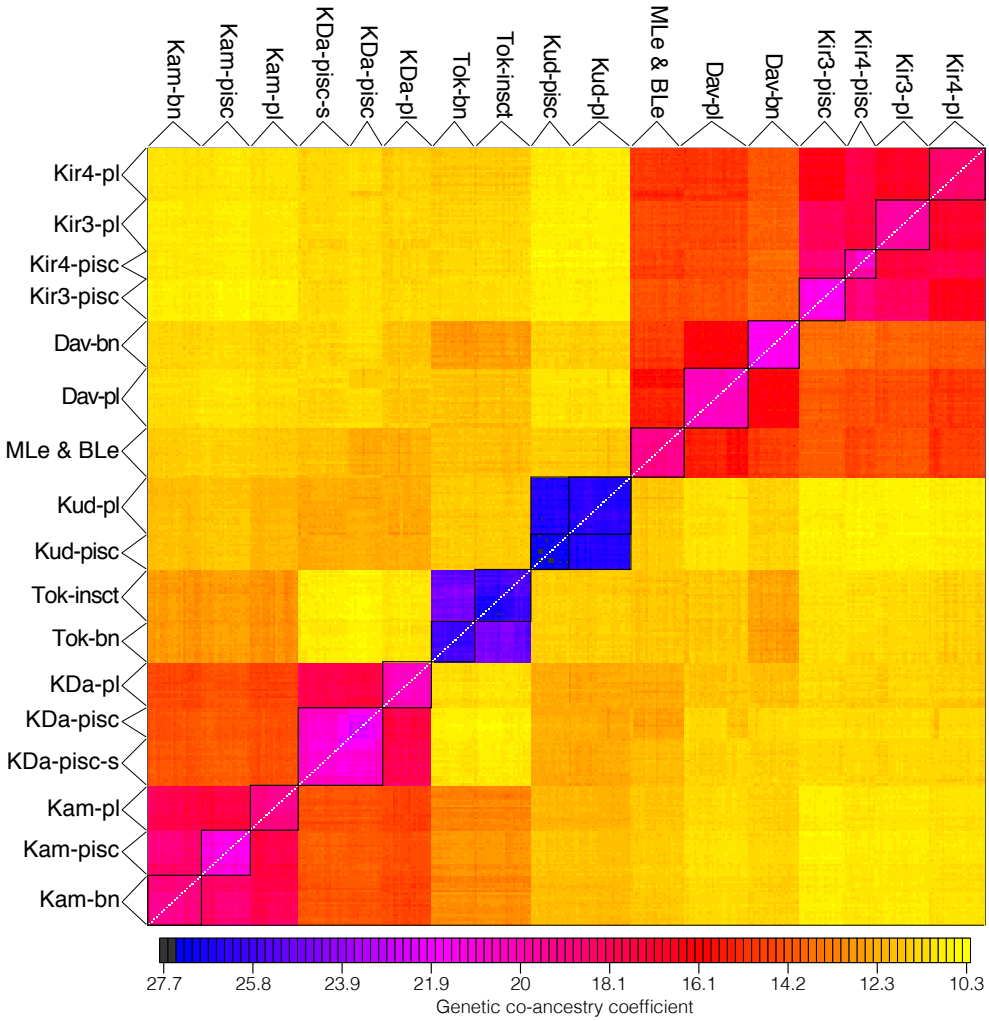


Fig. S14. Genetic co-ancestry across ecotypes and populations in the Atlantic lineage based on haplotype information. Individuals were clustered using *fineRADstructure* based on genetic co-ancestry. Populations strongly cluster by catchment (e.g. BLe, MLe, Dav, Kir3 and Kir4). Kudushkit and Tokko show extremely high genetic co-ancestry of individuals within populations and relatively low levels of shared ancestry with other populations.

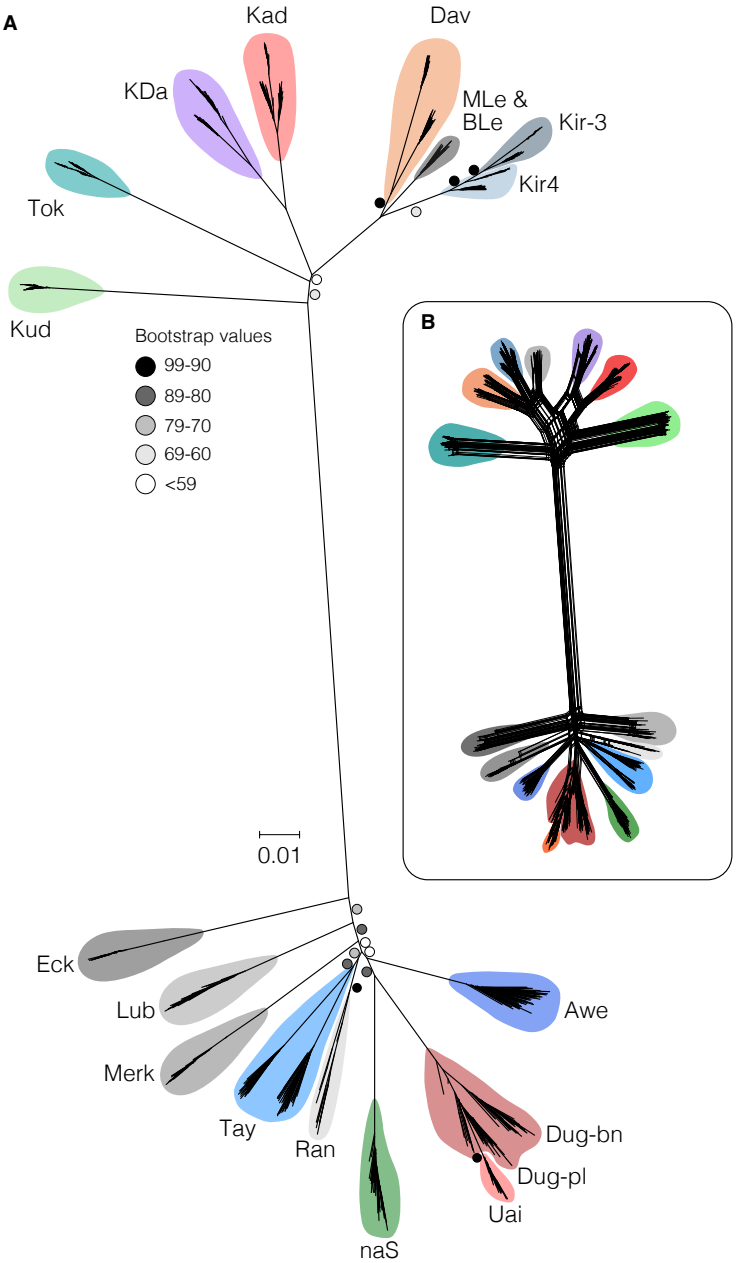


Fig. S15. Neighbour-joining tree and network highlighting the relationships between populations and lineages. (**A**) Neighbour-joining tree highlighting with bootstrap support values (100 bootstrap replicates) based on 12,215 SNPs (N = 648). (**B**) Neighbour-joining tree highlighting the strong uncertainty in the relationships between populations.

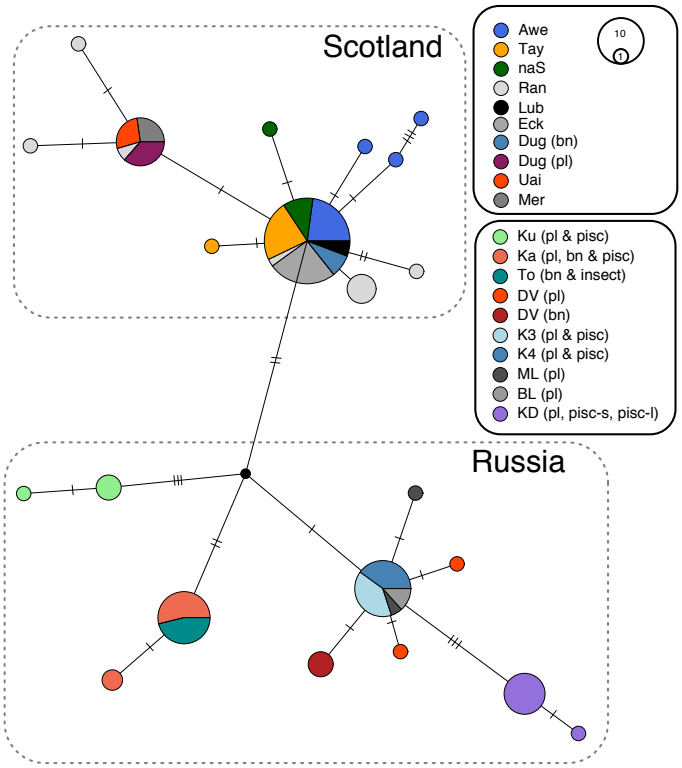
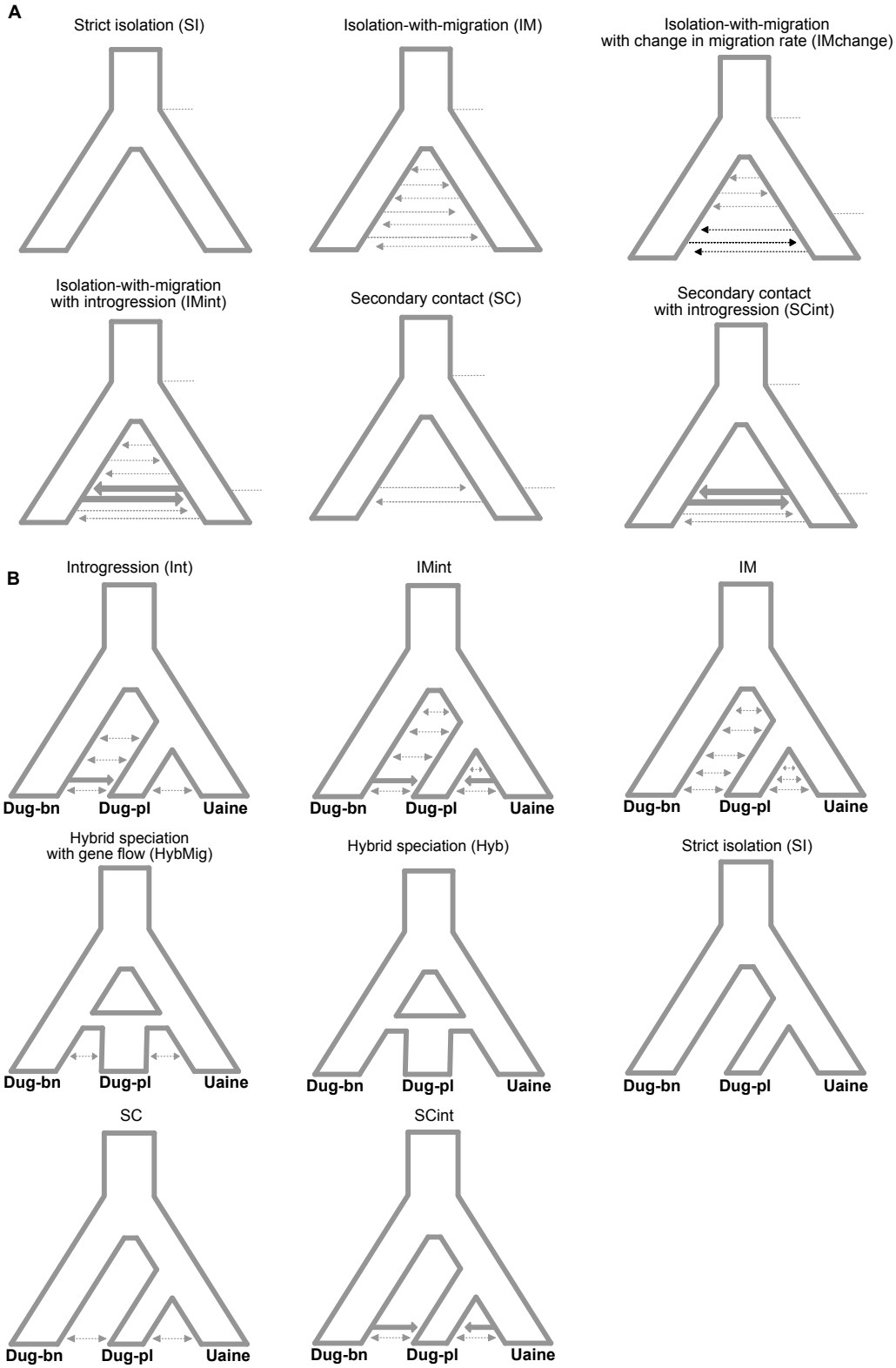


Fig. S16. Haplotype network based on the mitochondrial ND1 gene (N = 107). Most sympatric ecotypes share the same haplotypes. Exceptions are Loch Dughaill and Davatchan. In these cases, ecotypes were coloured differently. The size of the circles corresponds to the number of individuals sharing a haplotype.


A
Strict isolation (SI)
Isolation-with-migration (IM)
Isolation-with-migration
with change in migration rate (IMchange)
Isolation-with-migration
with introgression (IMint)
Secondary contact (SC)
Secondary contact
with introgression (SCint)
B
Introgression (Int)
IMint
IM
Dug-bn
Dug-pl
Uaine
Hybrid speciation
with gene flow (HybMig)
Hybrid speciation (Hyb)
Strict isolation (SI)
SC
SCint

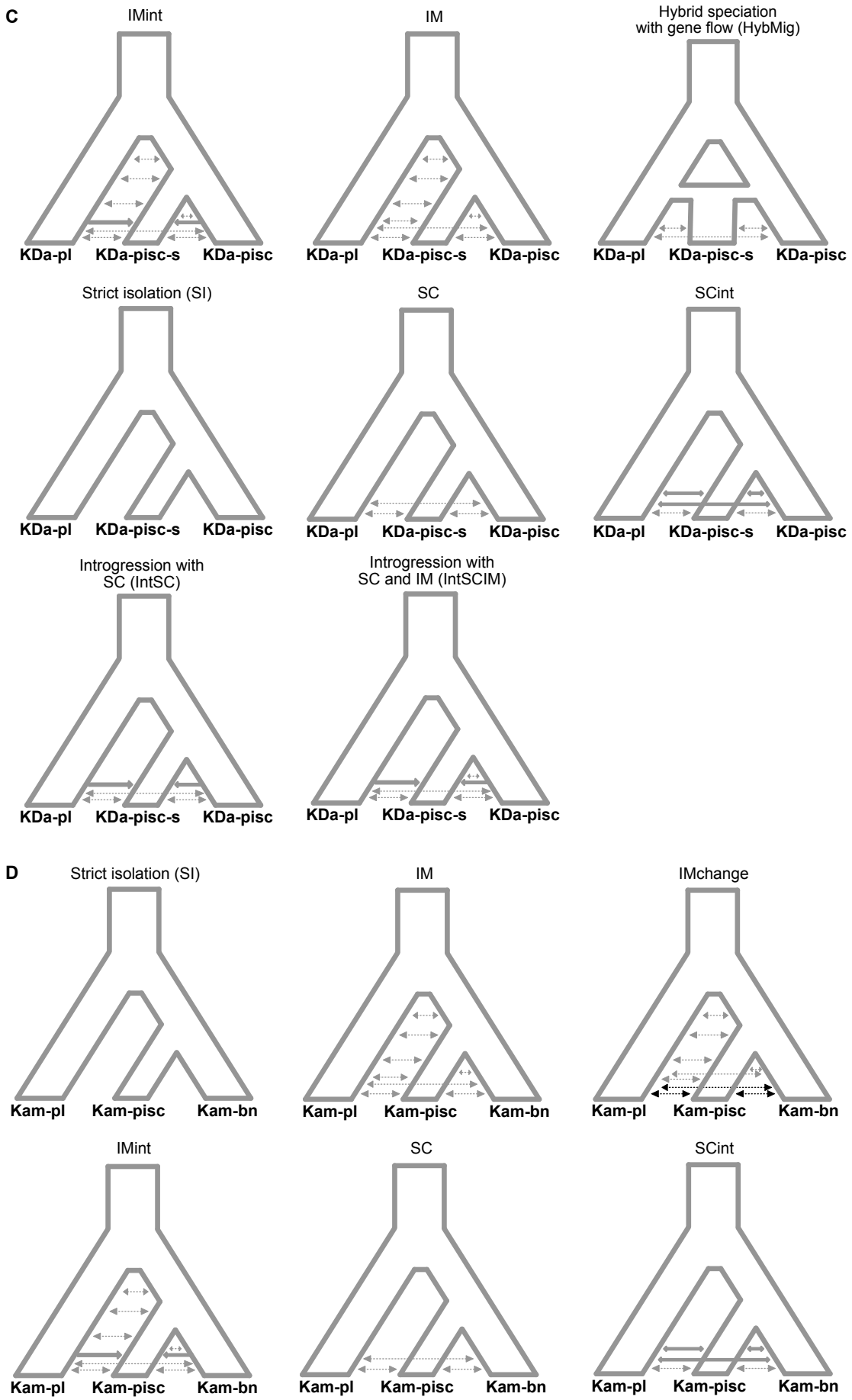

C
IMint
KDa-pl
KDa-pisc-s
KDa-pisc
IM
KDa-pl
KDa-pisc-s
KDa-pisc
Hybrid speciation
with gene flow (HybMig)
KDa-pl
KDa-pisc-s
KDa-pisc
Strict isolation (SI)
KDa-pl
KDa-pisc-s
KDa-pisc
SC
KDa-pl
KDa-pisc-s
KDa-pisc
SCint
KDa-pl
KDa-pisc-s
KDa-pisc
Introgression with
SC (IntSC)
KDa-pl
KDa-pisc-s
KDa-pisc
Introgression with
SC and IM (IntSCIM)
KDa-pl
KDa-pisc-s
KDa-pisc
D
Strict isolation (SI)
Kam-pl
Kam-pisc
Kam-bn
IM
Kam-pl
Kam-pisc
Kam-bn
IMchange
Kam-pl
Kam-pisc
Kam-bn
IMint
Kam-pl
Kam-pisc
Kam-bn
SC
Kam-pl
Kam-pisc
Kam-bn
SCint
Kam-pl
Kam-pisc
Kam-bn

Fig. S17. Overview of all tested demographic models in *fastsimcoal2*. (**A**) Illustrations of two-population models tested. (**B**) Three-population models tested for the evolutionary history of Dughaill and Uaine. (**C**) All three-population models tested for Kalarskii-Davatchan and (**D**) Kamkanda. We inferred most parameters, such as divergence times, timing of secondary contact and introgression, strength of introgression and gene flow and effective population sizes.

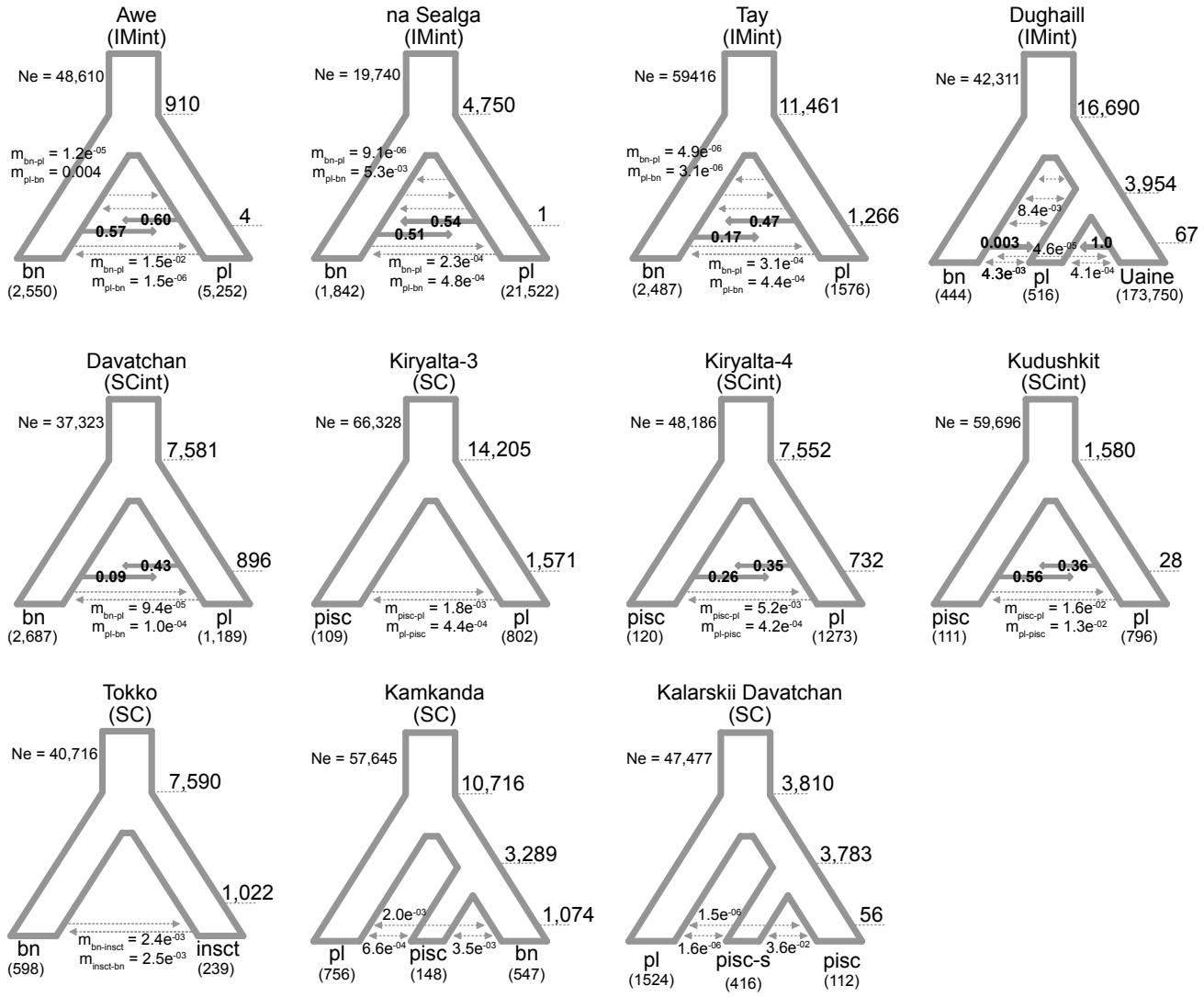


Fig. S18. Illustrations of the most likely evolutionary history for each population and inferred parameters using *fastsimcoal2*. All parameters are point estimates that were averaged across the five runs with the highest likelihood.

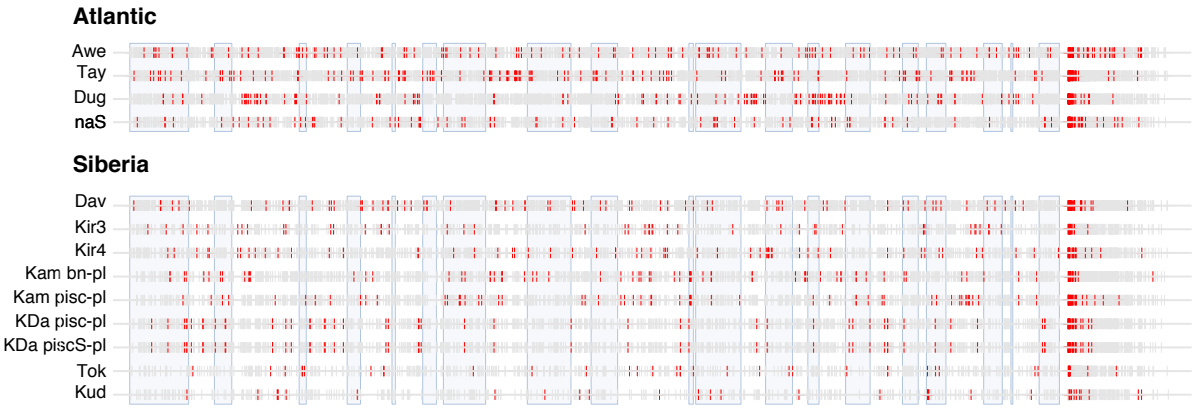


Fig. S19. Distribution of Fst outlier loci across the Arctic charr genome. Loci with Fst values above the 95th-quantile are highlighted in red (Fst outlier loci). Ordered chromosomes are alternatingly marked with grey boxes. Unplaced contigs are placed at the end of the genome.

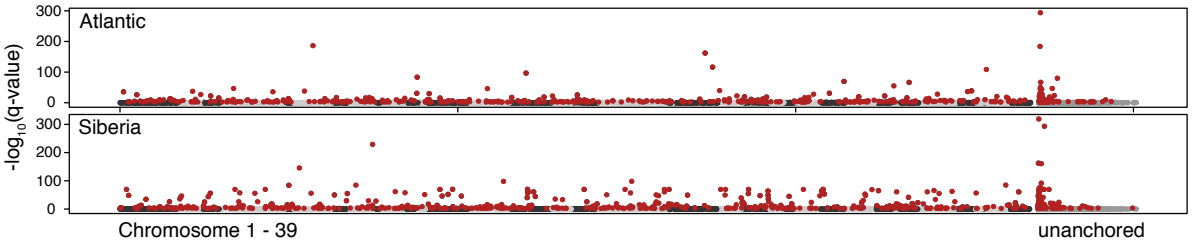


Fig. S20. Manhattan plot showing loci potentially under selection within each lineage. The log-transformed q-value for each SNP (inferred using PCAdapt) was plotted along the genome. Loci showing signs of selection (q-value < 0.01) are highlighted in red.

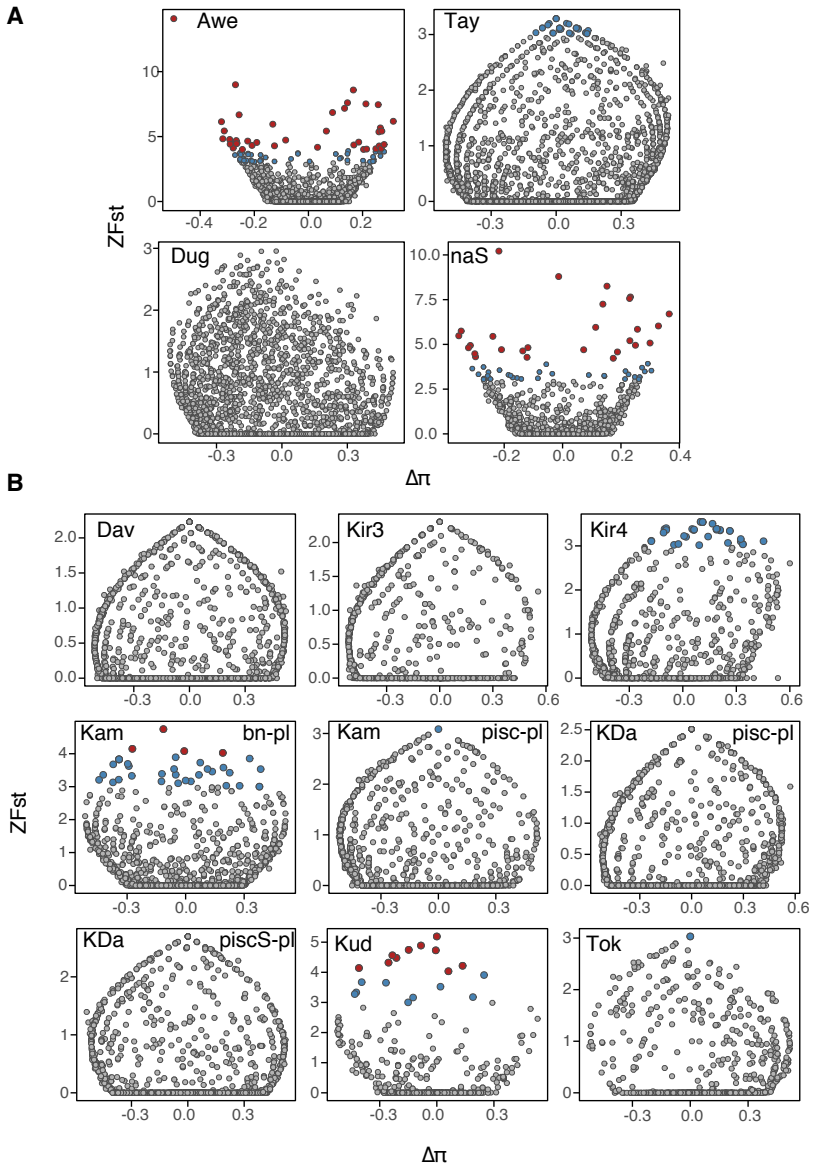


Fig. S21. Genome-wide patterns of differentiation between sympatric ecotypes in the Atlantic (**A**) and Siberian (**B**) lineage. The standardised Fst (ZFst) is plotted against the delta nucleotide-diversity (Δπ; bn – pl and pisc – pl). If the Δπ deviates from zero, it shows genetic diversity at this locus is reduced in one of the ecotypes. Loci with ZFst values above 4 were assumed to be significantly differentiated (red dots) and loci with ZFst above 3 (blue dots) are also reported. Two main patterns are clearly visible: differentiated loci show reduced diversity in one of the ecotypes (i.e. Awe and naS) or differentiated loci show similar levels of genetic diversity in both ecotypes (i.e. Tay, Dug, Dav or Kir3). Furthermore, populations displaying the first pattern of divergence tend to contain stronger outlier loci. These populations are also candidates for incipient speciation under ongoing gene flow. Two ecotype pairs (Kud and Kam bn-pl) show intermediate patterns of differentiation.

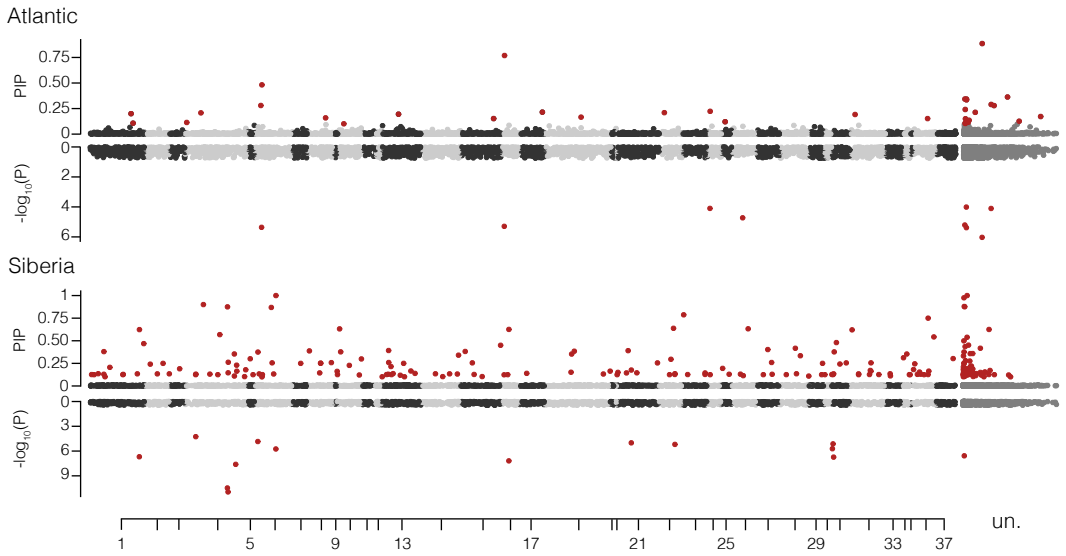


Fig. S22. Manhattan plots showing the results for the genome-wide association with ecotype. The top panel for the Atlantic and Siberian lineage show results from the Bayesian sparse linear mixed model (BSLMM) and variants with a posterior inclusion probability (PIP) above 0.1 are highlighted in red. These variants are assumed to have a significant effect on phenotype. The lower inverted panel shows the results for the univariate linear mixed effects model and variants signficiantly associated with ecotype ($-\log_{10}(P) > 4$) are highlighted in red. None of these variants is shared across lineages.

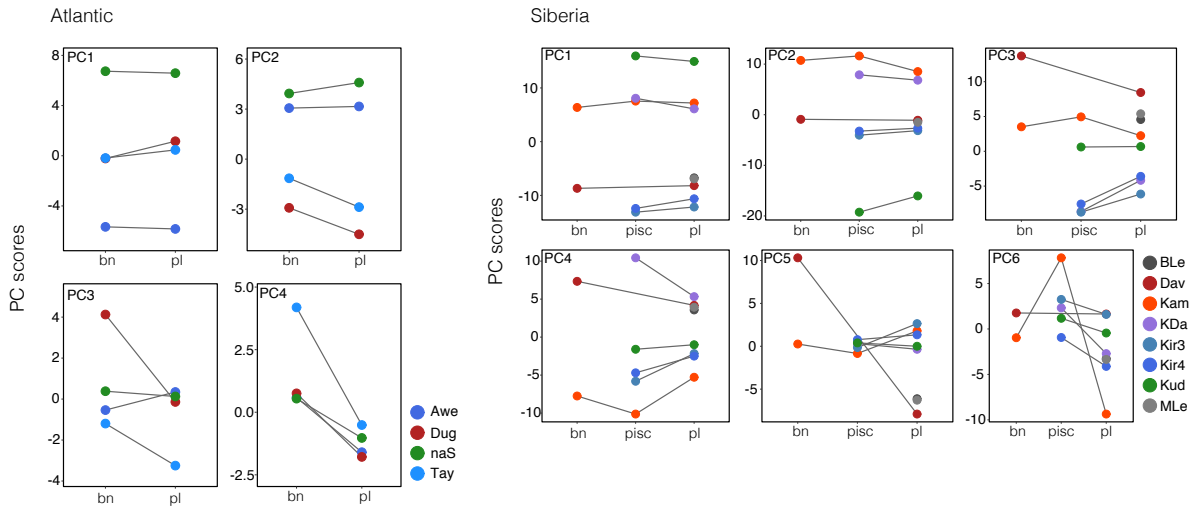


Fig. S23. Allele frequency parallelism of GWAS candidate loci across populations within lineages. Principal component analyses were performed using variants associated with ecotype in the BSLMM analysis (PIP > 0.1) for each lineage. Shown are the trajectories connecting ecotypes within populations along the first four and six principal component (PC) axes for the Atlantic and Siberian lineage, respectively. PC4 shows the highest level of parallelism in the Atlantic linages, with the parallel ecotype effect explaining most of the genetic variation. In the Siberian lineage, PC1 to PC4 explain large proportions of the genetic variation, more than the non-parallel ecotype x lake interaction, but never more than the unique evolutionary history of each population. Results of the statistical analysis are given in table S13.

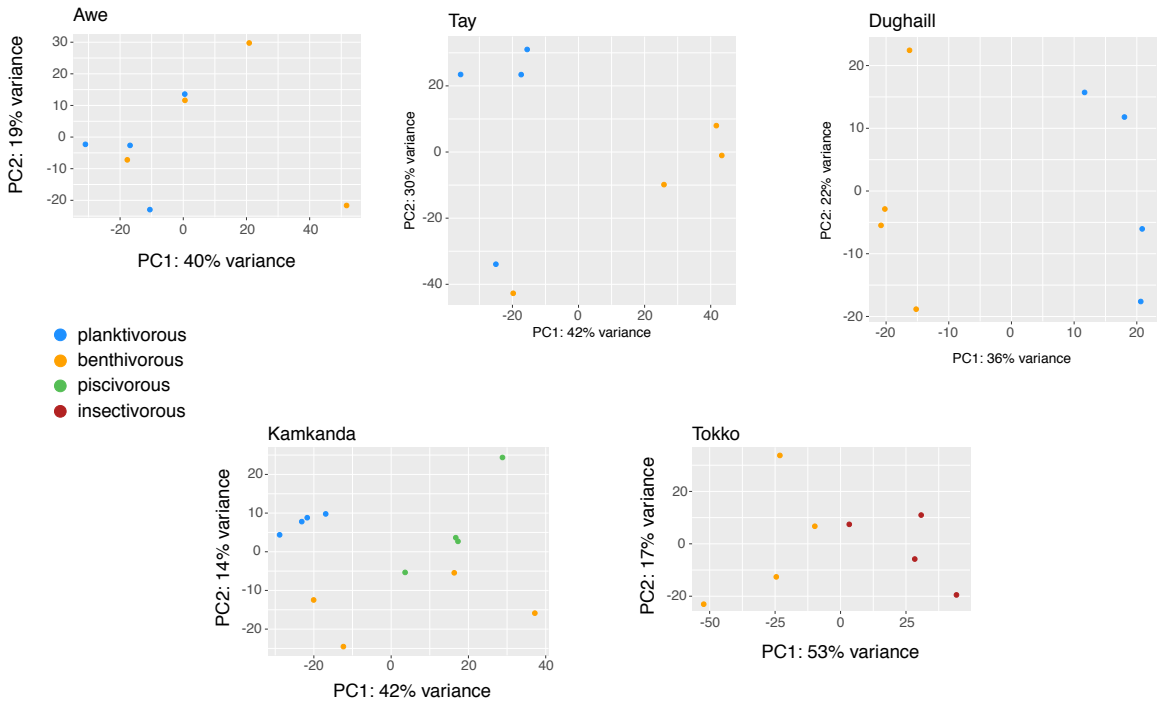


Fig. S24. Variation in gene expression between ecotypes within populations. Principal component analysis was performed based on the gene expression profiles (normalised read count) of the top 500 differentially expressed genes between ecotypes.

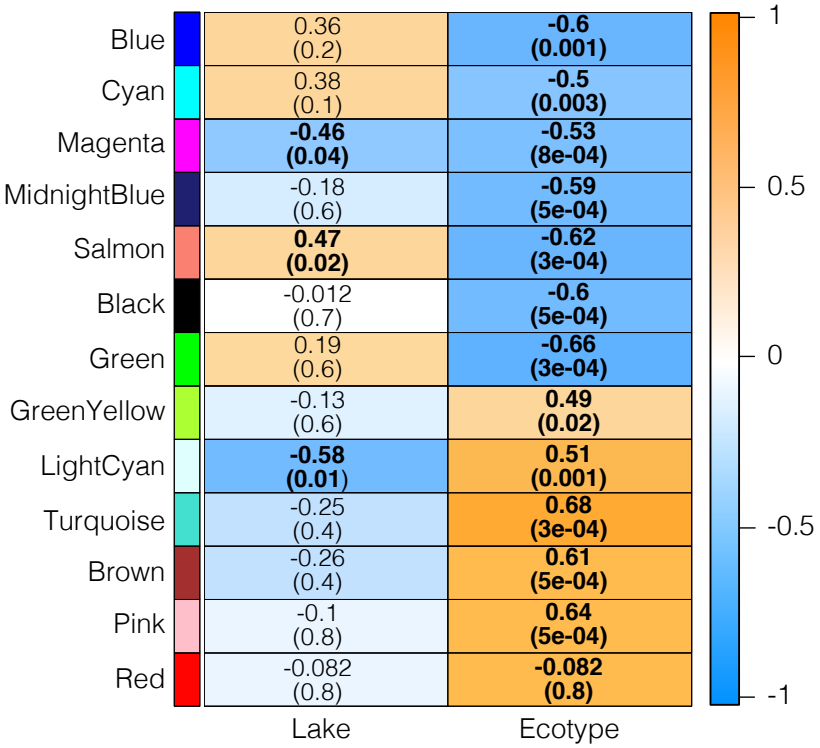


Fig. S25. Correlation between expression of WGCNA module eigengenes and lake (population of origin) and ecotype. Each row represents a module of co-expressed genes (identified by colour). Spearman's correlation coefficients for the module expression – variable correlation are given in each cell. The corresponding corrected p-values are given in parenthesis. Significant correlations are highlighted in bold. Cells are coloured based on their correlation, with orange cells being correlated with up regulation of gene expression in benthivorous ecotypes and blue being associated with up-regulation in planktivorous ecotypes.

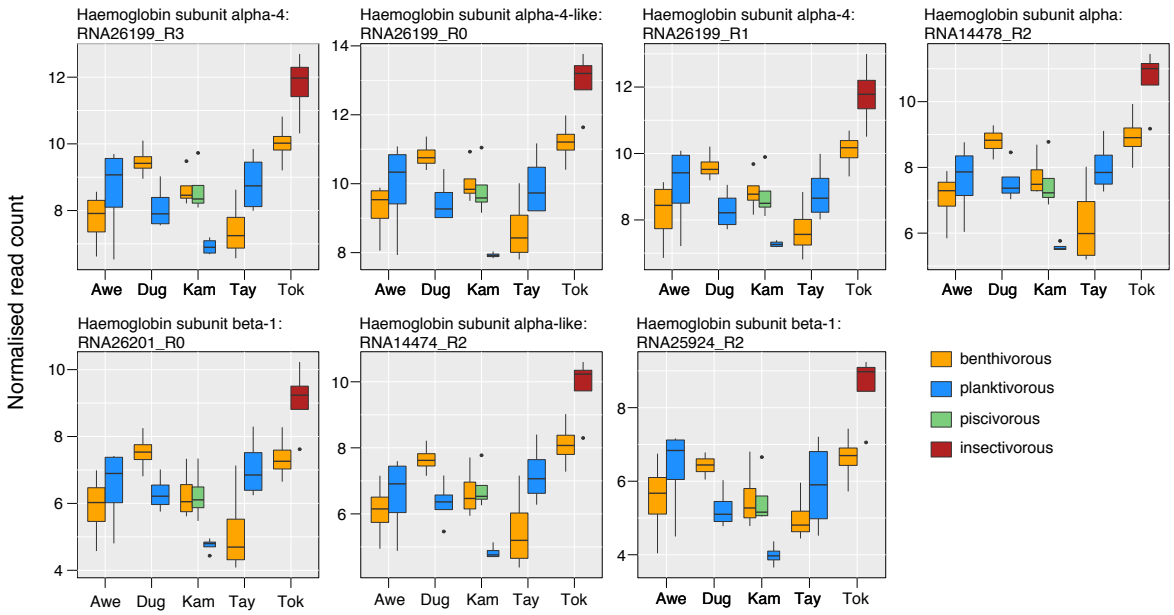


Fig. S26. Expression profiles of haemoglobin subunits (*HBA* and *HBB*) across ecotypes and populations. Haemoglobin subunits are significantly differentially expressed in up to four ecotype pairs but the direction of gene expression change is not always parallel across lakes.

Table S1. Sampling sites and sample sizes for populations used for phenotypic and genomic analysis of Arctic charr from the Atlantic and Siberian lineage.

Atlantic

Lake	Abb[1]	Catchment	Lat. (N)	Lon. (W)	N phenotyped individuals			N genotyped individuals		
					Bn	Pl	Uni	Bn	Pl	Uni
Tay	Tay	Tay	56˚30’	004˚10’	43	76	-	21	31	-
Rannoch	Ran	Tay	56˚41’	004˚16’	-	-	-	-	-	11
Awe	Awe	Awe	56˚20’	005˚05’	34	36	-	25	28	-
Eck	Eck	Cowal	56˚01’	004˚58’	-	-	56	-	-	19
Lubnaig	Lub	Forth	56˚17’	004˚17’	-	-	44	-	-	20
Dughaill	Dug	Ewe	57˚28’	005˚20’	44	52	-	31	31	-
Uaine	Uai	Ewe	57˚31’	005˚23’	-	-	50	-	-	23
naSealga	naS	-	57˚79’	004˚30’	41	19	-	19	21	-
Merkland	Mer	Shin	58˚14’	004˚44’	-	-	50	-	-	22

Siberian

Lake	Abb[1]	Catchment	Lat. (N)	Lon. (E)	N phenotyped individuals				N genotyped individuals			
					Bn	Pl	Pisc	Insct	Bn	Pl	Pisc	Insct
Kudushkit	Kud	Chaya	56˚33’	110˚37’	-	-	20	-	-	20	19	-
Davatchan	Dav	Chara, Olekma	56˚27’	117˚33’	76	114	-	-	20	25	-	-
Kalarskii Davatchan	KDa	Kalar,Vitim	56˚16.5’	116˚38.5’	-	83	15/ 81[2]	-	-	19	12/ 21[2]	-
Kiryalta-3	Kir3	Chara, Olekma	57˚08.5’	119˚27’	-	23	30	-	-	20	19	-
Kiryalta-4	Kir4	Chara, Olekma	57˚06.5’	119˚28’	-	42	25	-	5	21	13	-
Tokko	Tok	Tokko, Chara, Olekma	57˚11’	119˚41’	42	-	-	47	18	20	-	19
Kamkanda	Kam	Khani, Olekma	57˚05.5’	199˚48.5’	146	163	26	-	20	19	20	-
Bol’shoe Leprindo	BLe	Chara, Olekma	56˚37’	117˚31’	25	-	-	-	10	-	-	-
Maloe Leprindo	MLe	Chara, Olekma	56˚36’	117˚22’	26	-	-	-	11	-	-	-

Note: [1]Abbrevations used for these populations in the manuscript. [2] There are two piscivorous ecotypes in Kalarskii Davatchan, one large piscivorous (first number) and a small piscivorous (last number) ecotype. Ecotype abbreviations: Bn – benthivorous, PL – planktivorous, Uni – unimodal-planktivorous populations or without phenotypic information (Rannoch), Pisc – piscivorous, Insct – insectivorous.

Table S2. Divergence of each measured linear trait between ecotype pairs. Shown are p-values and significant comparisons are highlighted in bold.

Eco1	Eco2	HDO	HDE	HL	ED	PFL	LJL	ML	FL
Awe-bn	Awe-pl	0.5152	0.3326	0.7133	**0.0104**	0.7682	0.1855	0.6917	0.2643
Dug-bn	Dug-pl	**< 0.001**	**0.0058**	**0.0033**	**0.0056**	**< 0.001**	**0.0200**	0.1907	0.2947
Dug-bn	Uai	0.7094	**< 0.001**	**0.0444**	**< 0.001**	**0.0047**	**< 0.001**	0.6983	**< 0.001**
Dug-pl	Uai	**0.0014**	**< 0.001**	0.4148	**0.0237**	**0.0401**	0.1619	**0.0611**	**< 0.001**
naS-bn	naS-pl	0.9770	0.9849	0.4123	0.4781	0.2906	**0.0469**	0.4862	**< 0.001**
Tay-bn	Tay-pl	**< 0.001**	0.3331	0.7313	**< 0.001**	0.2232	0.2004	0.1536	**< 0.001**
DV-bn	DV-pl	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Ka-bn	Ka-pisc	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
Ka-bn	Ka-pl	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**0.17**
Ka-pisc	Ka-pl	0.8921	0.9145	0.3209	**0.0073**	**< 0.001**	0.8581	**0.0019**	**< 0.001**
KD-pisc-l	KD-pisc-s	**0.0272**	**0.0057**	0.5286	0.0823	0.5750	0.2029	0.9184	**< 0.001**
KD-pisc-l	KD-pl	**< 0.001**	**< 0.001**	**0.0092**	**0.0321**	**0.0201**	0.1033	**0.0286**	**< 0.001**
KD-pisc-s	KD-pl	**< 0.001**	**< 0.001**	**< 0.001**	0.5179	**0.0015**	0.5492	**< 0.001**	**< 0.001**
Kir3-pisc	Kir3-pl	0.0855	**0.0060**	**0.0233**	**0.0011**	**< 0.001**	0.1039	**0.0337**	**0.0377**
Kir3-pisc	Kir4-pisc	0.5460	0.2866	0.6634	0.5589	0.6227	0.5463	0.9378	0.6697
Kir3-pl	Kir4-pisc	**0.0235**	0.1136	0.0846	**< 0.001**	**< 0.001**	0.3360	**0.0504**	**0.0145**
Kir3-pisc	Kir4-pl	**< 0.001**	**< 0.001**	**< 0.001**	0.8332	**0.0122**	**< 0.001**	0.9799	**< 0.001**
Kir3-pl	Kir4-pl	0.0695	**0.0081**	0.1492	**< 0.001**	**0.0124**	**0.0110**	**0.0233**	**0.0250**
Kir4-pisc	Kir4-pl	**< 0.001**	**< 0.001**	**< 0.001**	0.6503	0.0697	**< 0.001**	0.9205	**< 0.001**
BL-pl	ML-pl	**0.0039**	0.3835	**0.0126**	0.8353	0.1886	0.3579	**0.0472**	**0.0249**
To-bn	To-insct	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**

Note: Abbreviations for lakes are shown in table one. HDO – Head depth at the operculum, HDE – Head depth at eye, HL – Head length, ED – Eye diameter, PFL – pectoral fin length, LJL – Lower jaw length, ML – Maxilla length, FL – Fork length. Reported p-values were corrected for multiple testing.

Table S3. Phenotypic parallelism within and across lineages based on linear traits. MANCOVA results and effect sizes of evolutionary and ecological terms for each lineage and both lineages combined.

Lineage	Effect	Wilk's Lambda	df	p-value	Partial η^2	partial variance [%]
Siberian	Ecotype	0.14649	2	< 2.2e-16	0.282	28.2
	Lake	0.22067	8	< 2.2e-16	0.373	37.3
	Ecotype x Lake	0.72232	4	< 2.2e-16	0.061	6.1
Atlantic	Ecotype	0.71860	1	< 2.2e-16	0.00163	0.16
	Lake	0.34666	3	< 2.2e-16	0.351	35.1
	Ecotype x Lake	0.63203	3	< 2.2e-16	0.0109	1.09
Combined	Ecotype	0.29617	2	< 2.2e-16	0.121	12.1
	Lineage	0.39180	1	< 2.2e-16	0.195	19.5
	Lake	0.26947	11	< 2.2e-16	0.372	37.2
	Ecotype:Lineage	0.89910	1	< 2.2e-16	0.031	3.1
	Ecotype:Lake	0.73402	7	< 2.2e-16	0.031	3.1

Table S4. Trait-by-trait linear model results for benthivorous-planktivorous (bn – pl) and piscivorous-planktivorous (pisc – pl) ecotype pairs. Shown are partial η^2 values. All terms were significant and the strongest term for each phenotype is highlighted in bold.

Comparison	Trait	Ecotype	Lake	Ecotype x Lake
Bn - Pl	ED	**0.323086**	0.14578	0.091302
	PFL	**0.086307**	0.492767	0.077966
	HDE	**0.169528**	0.646455	0.150975
	LJL	**0.097425**	0.207795	0.044578
	HL	**0.122886**	0.350742	0.074271
	ML	**0.079253**	0.490287	0.042407
Pisc - Pl	ED	**0.104075**	0.32582	0.031376
	PFL	**0.35992**	0.463427	0.166505
	HDE	**0.211567**	0.048338	0.06539
	LJL	**0.102983**	0.190074	0.044351
	HL	**0.123311**	0.105932	0.021021
	ML	0.001288	0.190586	**0.103356**
	HDO	**0.194594**	0.082245	0.050367

Note: Trait abbreviations: ED – Eye diameter, PFL – Pectoral fin length, HDE – Head depth at eye, LJL – Lower jaw length, HL – Head length, ML – Maxilla length, HDO – Head depth at operculum.

Table S5. Divergence in body shape. Shown are effect sizes for differences in the angle (below diagonal) and distance (above diagonal) of allometric slopes between ecotypes, estimated using a linear model. Significant comparisons ($P < 0.05$) are highlighted in bold.

Slope angle/ distance	bn:Awe	bn:Dug	bn:naS	bn:Tay	pl:Awe	pl:Dug	pl:naS	pl:Tay
bn:Awe	-	-0.69	**2.44**	0.02	-1.07	-1.08	-0.55	-0.26
bn:Dug	**3.94**	-	0.37	-0.16	-0.56	-0.85	-0.56	-0.16
bn:naS	1.23	**3.59**	-	**3.73**	**2.44**	1.77	-1.17	**3.15**
bn:Tay	**7.58**	2.62	**2.80**	-	0.37	-0.15	-0.31	-1.28
pl:Awe	**5.41**	**3.59**	**2.78**	**7.83**	-	0.88	-0.44	-0.54
pl:Dug	0.92	**2.22**	1.06	1.83	2.56	-	-0.49	-0.30
pl:naS	0.63	0.59	-0.38	0.15	0.40	0.22	-	-0.26
pl:Tay	**5.67**	1.73	**6.21**	**5.63**	**5.12**	1.42	1.11	-

Table S6. Phenotypic parallelism in body shape. ANCOVA results for the first 5 principal component (PC) axes of body shape in Tay, Awe, Dug and naS. Shown are partial η^2 values and the strongest term, explaining most of the total variance for each PC is highlighted in bold.

PC	Ecotype	Lake	Ecotype x Lake	P-value
PC1	**0.486**	0.391	0.255	< 0.001
PC2	0.208	**0.445**	0.338	< 0.001
PC3	0.034	0.027	**0.148**	< 0.001
PC4	0.001	0.063	0.027	0.496
PC5	0.016	**0.207**	0.054	0.026

Table S7. Phenotypic trajectory analysis. Shown are differences in the distances of phenotypic trajectories across ecotype pairs (below) and angles (in degrees) between trajectories. Parallel trajectories (in distance or direction, P > 0.05) are highlighted in bold.

Angle / distance	Awe (bn-pl)	Dug (bn-pl)	naS (bn-pl)	Tay (bn-pl)
Awe (bn-pl)	0	83.19858	49.98509	36.06031
Dug (bn-pl)	0.008719617	0	57.15612	66.40108
naS (bn-pl)	**0.002360382**	**0.006359235**	0	**26.95783**
Tay (bn-pl)	0.036864199	0.028144582	0.034503817	0

Table S8. Genome scan results for sympatric ecotype pairs.

Lake	Ecotype pairs	mean Fst	outlier Fst	95th quant.	Out. loci	ZFst>2	ZFst>3	ZFst>4	Fixed SNPs	Perc. fixed
Awe	Bn-Pl	0.0113	0.1012	0.0653	175	185	86	40	0	0
Tay	Bn-Pl	0.2262	0.8483	0.7391	209	273	36	0	4	0.1
Dug	Bn-Pl	0.2524	0.7537	0.6814	178	142	0	0	0	0
naS	Bn-Pl	0.0136	0.1195	0.0795	131	146	60	27	0	0
Dav	Bn-Pl	0.3291	0.9898	0.9523	113	127	0	0	76	3.38
Kam	Bn-Pl	0.1336	0.63	0.4935	96	99	36	4	0	0
Kam	Pl-Pisc	0.2445	0.8567	0.7592	90	106	1	0	1	0.06
KDa	Pl-Pisc	0.2976	0.9569	0.8748	85	99	0	0	35	2.08
KDa	Pl-Pisc-s	0.2587	0.9393	0.8549	83	102	0	0	18	1.08
Kir-3	Pl-Pisc	0.3074	0.953	0.8845	59	51	0	0	15	1.26
Kir-4	Pl-Pisc	0.1837	0.8208	0.6731	86	112	36	0	0	0
Kud	Pl-Pisc	0.1003	0.6621	0.4677	37	40	19	10	1	0.14
Tok	Bn-Insct	0.2233	0.8593	0.7557	46	54	2	0	2	0.22

Note: 95th-quant. gives the 95th-quantile Fst values. Out. loci gives the number of loci above the 95th-quantile. Number of loci with ZFst above a certain threshold. Number of fixed SNPs and percentage of SNPs that are fixed compared to all SNPs in an ecotype pair.

Table S9. Correlation of genetic differentiation, phenotypic distance and ecosystem size.

Trait1	Trait2	Pearson's r	P-value	adj. P-value
neutralFst	Pst.HDE	0.3743279	0.1873	0.27054444
neutralFst	Pst.HL	0.5500582	0.04155	0.10803
neutralFst	Pst.ED	0.3646646	0.1999	0.27354737
neutralFst	Pst.ML	0.5880682	0.02698	0.087685
neutralFst	Pst.LJL	-0.04043237	0.8908	0.9213
neutralFst	**Pst.PFL**	**0.7287232**	**0.003113**	**0.037934**
neutralFst	Pst.FL	0.3478646	0.2229	0.27597143
neutralFst	**meanPst**	**0.6475247**	**0.01229**	**0.06235667**
outlierFst	Pst.HDE	0.4689736	0.09072	0.16355733
outlierFst	Pst.HL	0.5550213	0.03938	0.10803
outlierFst	Pst.ED	0.4039749	0.152	0.23247059
outlierFst	Pst.ML	0.5126136	0.06089	0.13312
outlierFst	Pst.LJL	-0.02910953	0.9213	0.9213
outlierFst	**Pst.PFL**	**0.6992596**	**0.005383**	**0.037934**
outlierFst	Pst.FL	0.5049314	0.06554	0.13312
outlierFst	**meanPst**	**0.6946158**	**0.005836**	**0.037934**
EcoSize	Pst.HDE	0.3515184	0.2178	0.27597143
EcoSize	**Pst.HL**	**0.8153686**	**0.0003793**	**0.0098618**
EcoSize	Pst.ED	0.04489506	0.8789	0.9213
EcoSize	Pst.ML	0.5033007	0.06656	0.13312
EcoSize	Pst.LJL	0.2339964	0.4207	0.47557391
EcoSize	Pst.PFL	0.6364766	0.01439	0.06235667
EcoSize	Pst.FL	0.4310588	0.1238	0.201175
EcoSize	**meanPst**	**0.6086634**	**0.02089**	**0.07759143**
EcoSize	neutralFst	0.2980552	0.2806	0.33161818
EcoSize	outlierFst	0.4475541	0.09436	0.16355733

Note: OutlierFst – Genetic differentiation calculated based on SNPs with Fst above 95th-quantile in the respective ecotype pair, neutralFst: calculated based on SNPs below with Fst below 95th-quantile. Mean Pst was calculated by averaging across trait-specific Pst values. EcoSize – Ecosystem size (PC1).

Table S10. Significant tests for introgressive hybridization (*f3*-statistics).

Focal	**Ref1**	**Ref2**	***f3***	**SE**	**Z-score**
Dug-pl	Dug-bn	Uai	-0.0259701	0.00110819	-23.4348
Dug-pl	Tay-bn	Uai	-0.00475052	0.00127475	-3.72663
Dug-pl	Ran	Uai	-0.00448714	0.00123844	-3.62322
Dug-pl	Tay-pl	Uai	-0.00434716	0.00126704	-3.43097
Dug-pl	Uai	Lub	-0.00435496	0.00140769	-3.0937
Dug-pl	Awe-pl	Uai	-0.0037364	0.00131955	-2.83157
Dug-pl	Eck	Uai	-0.0042912	0.00153863	-2.78897
Dug-pl	Uai	Awe-bn	-0.00371748	0.00133324	-2.7883
Dug-pl	Uai	naSeal-pl	-0.00329079	0.00140305	-2.34546
Dug-pl	naSeal-bn	Uai	-0.00295874	0.00140562	-2.10494
Dug-pl	Uai	Mer	-0.0030708	0.00146056	-2.10249
KDa-pisc-s	KDa-pisc	KDa-pl	-0.00249944	0.000420217	-5.94796
KDa-pisc-s	Kud-pisc	KDa-pisc	-0.00191366	0.000606432	-3.15561
KDa-pisc-s	KDa-pisc	Tok-bn	-0.00136393	0.000597315	-2.28343

Table S11. Significant tests for secondary gene flow (*f4*-statistics) in a four-population tree.

PopA	PopB	PopC	PopD	*f4*	SE	Z-score
Tay-pl	Tay-bn	Awe-bn	naSeal-pl	-0.0048	0.0024	-2.026
Tay-pl	Tay-bn	Awe-pl	Dug-pl	-0.0048	0.0021	-2.294
Tay-pl	Tay-bn	Awe-pl	Eck	-0.0089	0.0025	-3.579
Tay-pl	Tay-bn	Awe-pl	Mer	-0.0058	0.0023	-2.492
Tay-pl	Tay-bn	Awe-pl	naSeal-pl	-0.0053	0.0024	-2.201
Tay-pl	Tay-bn	Awe-pl	Uai	-0.0053	0.0024	-2.201
Dug-bn	Dug-pl	Awe-pl	Uai	0.0591	0.0028	20.919
Tay-pl	Tay-bn	Dug-pl	Awe-bn	0.0044	0.0021	2.076
Awe-pl	Awe-bn	Dug-pl	Lub	0.0011	0.0005	2.273
Tay-pl	Tay-bn	Eck	Awe-bn	0.0084	0.0025	3.395
Dug-bn	Dug-pl	Eck	Uai	0.0604	0.0031	19.266
Tay-pl	Tay-bn	Mer	Awe-bn	0.0053	0.0023	2.290
Dug-bn	Dug-pl	naSeal-bn	Uai	0.0570	0.0029	19.512
Tay-pl	Tay-bn	Ran	Eck	-0.0055	0.0027	-2.065
naSeal-bn	naSeal-pl	Ran	Eck	0.0011	0.0005	2.085
naSeal-bn	naSeal-pl	Ran	Lub	0.0009	0.0004	2.075
Dug-bn	Dug-pl	Ran	Uai	0.0595	0.0028	21.585
Eck	Awe-bn	Tay-bn	Lub	-0.0076	0.0030	-2.527
Eck	Awe-bn	Tay-bn	naSeal-pl	-0.0067	0.0032	-2.068
Dug-bn	Dug-pl	Tay-bn	Uai	0.0588	0.0028	20.887
naSeal-bn	naSeal-pl	Tay-pl	Ran	-0.0009	0.0004	-2.389
naSeal-bn	Awe-pl	Tay-pl	Tay-bn	0.0049	0.0024	2.070
Dug-bn	Awe-bn	Tay-pl	Tay-bn	0.0047	0.0022	2.108
Dug-bn	Awe-pl	Tay-pl	Tay-bn	0.0051	0.0022	2.321
Dug-bn	Dug-pl	Tay-pl	Uai	0.0591	0.0028	21.442
Dug-bn	Dug-pl	Uai	Awe-bn	-0.0593	0.0028	-20.894
Tay-pl	Tay-bn	Uai	Awe-bn	0.0048	0.0024	2.007
Dug-bn	Dug-pl	Uai	Lub	-0.0568	0.0028	-19.987
Dug-bn	Dug-pl	Uai	Mer	-0.0553	0.0030	-18.444
Dug-bn	Dug-pl	Uai	naSeal-pl	-0.0571	0.0029	-19.489
Dav-bn	BLe	KDa-pisc	KDa-pisc-s	-0.0027	0.0006	-4.852
Dav-bn	BLe	Kud-pisc	Kud-pl	0.0017	0.0005	3.163
Dav-bn	Dav-pl	Kir4-pl	Tok-bn	-0.0225	0.0032	-7.136
Dav-bn	Dav-pl	Kir4-pl	Tok-insct	-0.0221	0.0032	-6.895
Dav-bn	Dav-pl	MLe	Tok-bn	-0.0191	0.0031	-6.253
Dav-bn	Dav-pl	Kir4-pisc	Tok-bn	-0.0209	0.0034	-6.108
Dav-bn	Dav-pl	MLe	Tok-insct	-0.0187	0.0031	-6.070

Dav-bn	Dav-pl	Kir4-pl	Kam-bn	-0.0172	0.0029	-5.995
Dav-bn	Dav-pl	Kir3-pl	Tok-bn	-0.0202	0.0034	-5.976
Dav-bn	Dav-pl	Kir4-pisc	Tok-insct	-0.0204	0.0034	-5.932
Dav-bn	Dav-pl	BLe	Tok-insct	-0.0186	0.0031	-5.926
Dav-bn	Dav-pl	Kir3-pl	Tok-insct	-0.0197	0.0034	-5.810
Dav-bn	Dav-pl	Kir4-pl	Kud-pl	-0.0179	0.0031	-5.721
Dav-bn	Dav-pl	Kir3-pisc	Tok-bn	-0.0186	0.0035	-5.282
Dav-bn	Dav-pl	MLe	Kam-bn	-0.0138	0.0027	-5.127
Dav-bn	Dav-pl	Kir3-pisc	Tok-insct	-0.0182	0.0035	-5.117
Dav-bn	Dav-pl	Kir4-pl	KDa-pl	-0.0143	0.0028	-5.105
Dav-bn	Dav-pl	BLe	Kam-bn	-0.0138	0.0027	-5.053
Dav-bn	Dav-pl	Kir4-pisc	Kam-bn	-0.0156	0.0031	-5.042
Dav-bn	Dav-pl	Kir4-pisc	Kud-pisc	-0.0165	0.0033	-4.933
Dav-bn	Dav-pl	MLe	Kud-pl	-0.0144	0.0029	-4.925
Dav-bn	Dav-pl	Kir3-pl	Kam-bn	-0.0149	0.0030	-4.891
Dav-bn	Dav-pl	Kir4-pisc	Kud-pl	-0.0162	0.0033	-4.862
Dav-bn	Dav-pl	Kir4-pisc	Kam-pisc	-0.0157	0.0033	-4.834
Dav-bn	Dav-pl	Kir3-pl	Kud-pl	-0.0155	0.0033	-4.740
Dav-bn	Dav-pl	MLe	KDa-pl	-0.0109	0.0026	-4.199
Dav-bn	Dav-pl	Kir4-pisc	KDa-pl	-0.0126	0.0030	-4.190
Dav-bn	Dav-pl	Kir4-pl	KDa-pisc-s	-0.0119	0.0028	-4.177
Dav-bn	Dav-pl	Kir3-pisc	Kud-pisc	-0.0143	0.0034	-4.175
Dav-bn	Dav-pl	BLe	KDa-pl	-0.0108	0.0026	-4.161
Dav-bn	Dav-pl	Kir3-pisc	Kam-bn	-0.0133	0.0032	-4.156
Dav-bn	Dav-pl	Kir3-pisc	Kud-pl	-0.0139	0.0034	-4.096
Dav-bn	Dav-pl	Kir3-pl	KDa-pl	-0.0120	0.0030	-3.987
Dav-bn	Dav-pl	Kir3-pisc	Kam-pisc	-0.0135	0.0034	-3.961
Dav-bn	Dav-pl	KDa-pisc	Tok-bn	-0.0111	0.0032	-3.443
Dav-bn	Dav-pl	Kir4-pisc	KDa-pisc-s	-0.0102	0.0030	-3.377
Dav-bn	Dav-pl	Kir3-pisc	KDa-pl	-0.0104	0.0031	-3.309
Dav-bn	Dav-pl	KDa-pisc	Tok-insct	-0.0106	0.0032	-3.285
Dav-bn	Dav-pl	Kir4-pisc	KDa-pisc	-0.0098	0.0031	-3.210
Dav-bn	Dav-pl	MLe	KDa-pisc-s	-0.0085	0.0027	-3.189
Dav-bn	Dav-pl	Kir3-pl	KDa-pisc-s	-0.0096	0.0030	-3.154
Dav-bn	Dav-pl	BLe	KDa-pisc-s	-0.0085	0.0027	-3.139
Dav-bn	Dav-pl	Kir3-pl	KDa-pisc	-0.0091	0.0031	-2.990
Dav-bn	Dav-pl	KDa-pisc	Kam-bn	-0.0058	0.0021	-2.700
Dav-bn	Dav-pl	Kir3-pisc	KDa-pisc-s	-0.0080	0.0031	-2.547
Dav-bn	Dav-pl	Kir3-pisc	KDa-pisc	-0.0076	0.0032	-2.396
Dav-bn	Dav-pl	KDa-pisc	Kud-pl	-0.0064	0.0028	-2.265

Dav-bn	Dav-pl	Kam-pl	Tok-bn	-0.0059	0.0028	-2.126
Dav-bn	Dav-pl	KDa-pisc	KDa-pl	-0.0028	0.0014	-1.979
Dav-bn	Dav-pl	Kir3-pisc	Kir4-pl	0.0039	0.0019	2.043
Dav-bn	Dav-pl	Kud-pl	KDa-pisc-s	0.0060	0.0028	2.143
Dav-bn	Dav-pl	Kud-pisc	KDa-pisc-s	0.0063	0.0028	2.245
Dav-bn	Dav-pl	Kud-pisc	KDa-pisc	0.0067	0.0028	2.363
Dav-bn	Dav-pl	Kam-pl	KDa-pisc-s	0.0048	0.0020	2.366
Dav-bn	Dav-pl	Kam-pisc	KDa-pisc-s	0.0055	0.0023	2.373
Dav-bn	Dav-pl	Tok-insct	KDa-pl	0.0078	0.0033	2.393
Dav-bn	Dav-pl	Kam-pl	KDa-pisc	0.0052	0.0020	2.530
Dav-bn	Dav-pl	Kam-bn	KDa-pisc-s	0.0054	0.0021	2.535
Dav-bn	Dav-pl	Kam-pisc	KDa-pisc	0.0059	0.0023	2.545
Dav-bn	Dav-pl	Tok-bn	KDa-pl	0.0082	0.0032	2.556
Dav-bn	Dav-pl	KDa-pisc	BLe	0.0080	0.0027	2.953
Dav-bn	Dav-pl	KDa-pisc	MLe	0.0081	0.0027	2.997
Dav-bn	Dav-pl	Tok-insct	KDa-pisc-s	0.0102	0.0032	3.160
Dav-bn	Dav-pl	Tok-bn	KDa-pisc-s	0.0106	0.0032	3.321
Dav-bn	Dav-pl	KDa-pisc	Kir4-pl	0.0115	0.0029	3.982
Dav-bn	Dav-pl	Kam-pl	Kir3-pisc	0.0127	0.0032	4.001
Dav-bn	Dav-pl	Kam-pisc	Kir3-pl	0.0150	0.0032	4.691
Dav-bn	Dav-pl	Kam-pl	Kir3-pl	0.0143	0.0030	4.732
Dav-bn	Dav-pl	Kud-pisc	Kir3-pl	0.0158	0.0033	4.812
Dav-bn	Dav-pl	Kam-pisc	BLe	0.0139	0.0029	4.847
Dav-bn	Dav-pl	Kud-pl	BLe	0.0144	0.0030	4.859
Dav-bn	Dav-pl	Kam-pl	Kir4-pisc	0.0150	0.0031	4.898
Dav-bn	Dav-pl	Kam-pisc	MLe	0.0140	0.0028	4.914
Dav-bn	Dav-pl	Kud-pisc	BLe	0.0147	0.0030	4.936
Dav-bn	Dav-pl	Kam-pl	BLe	0.0132	0.0027	4.936
Dav-bn	Dav-pl	Kud-pisc	MLe	0.0148	0.0030	5.002
Dav-bn	Dav-pl	Kam-pl	MLe	0.0132	0.0026	5.058
Dav-bn	Dav-pl	Kam-pisc	Kir4-pl	0.0174	0.0030	5.741
Dav-bn	Dav-pl	Kud-pisc	Kir4-pl	0.0182	0.0031	5.779
Dav-bn	Dav-pl	Kam-pl	Kir4-pl	0.0167	0.0028	5.852
Dav-bn	Dav-pl	Tok-bn	BLe	0.0191	0.0031	6.123
Dav-bn	Kam-bn	KDa-pisc	KDa-pisc-s	-0.0014	0.0006	-2.376
Dav-bn	Kam-bn	Kir4-pisc	Kir4-pl	-0.0043	0.0020	-2.229
Dav-bn	Kam-bn	Kud-pisc	Kud-pl	0.0017	0.0008	2.209
Dav-bn	Kam-pisc	KDa-pisc	KDa-pisc-s	-0.0022	0.0006	-3.429
Dav-bn	Kam-pisc	Kud-pisc	Kud-pl	0.0020	0.0008	2.441
Dav-bn	Kam-pl	Kir4-pisc	Kir4-pl	-0.0041	0.0019	-2.147

Dav-bn	Kam-pl	Kud-pisc	Kud-pl	0.0016	0.0008	2.094
Dav-bn	KDa-pisc	Kir4-pisc	Kir4-pl	-0.0050	0.0019	-2.568
Dav-bn	KDa-pisc	Kud-pisc	Kud-pl	0.0028	0.0009	3.306
Dav-bn	KDa-pisc-s	Kir4-pisc	Kir4-pl	-0.0044	0.0019	-2.310
Dav-bn	KDa-pl	Kir4-pisc	Kir4-pl	-0.0043	0.0019	-2.248
Dav-bn	Kir3-pisc	KDa-pisc-s	KDa-pl	-0.0035	0.0018	-1.996
Dav-bn	Kud-pisc	KDa-pisc-s	KDa-pl	-0.0052	0.0023	-2.237
Dav-bn	Kud-pl	KDa-pisc	KDa-pl	-0.0061	0.0024	-2.572
Dav-bn	Kud-pl	KDa-pisc-s	KDa-pl	-0.0052	0.0023	-2.248
Dav-bn	Kud-pl	Kir4-pisc	Kir4-pl	-0.0043	0.0021	-2.083
Dav-bn	MLe	KDa-pisc	KDa-pisc-s	-0.0019	0.0005	-3.824
Dav-bn	MLe	Kud-pisc	Kud-pl	0.0010	0.0005	2.073
Kam-pisc	Dav-pl	Kud-pisc	Kud-pl	-0.0017	0.0008	-2.189
Kam-pisc	Dav-pl	KDa-pisc	KDa-pisc-s	0.0018	0.0007	2.768
Kam-pisc	Kam-bn	KDa-pisc	KDa-pisc-s	0.0008	0.0003	2.896
Kam-pisc	KDa-pisc-s	Tok-bn	Tok-insct	-0.0028	0.0012	-2.367
Kam-pisc	Kir3-pl	Kud-pisc	Kud-pl	-0.0027	0.0008	-3.445
Kam-pisc	Kir3-pl	KDa-pisc	KDa-pisc-s	0.0021	0.0006	3.339
Kam-pisc	Kir4-pl	Kud-pisc	Kud-pl	-0.0024	0.0007	-3.197
Kam-pisc	Kir4-pl	KDa-pisc	KDa-pisc-s	0.0020	0.0006	3.142
Kam-pisc	Kud-pisc	KDa-pisc	KDa-pisc-s	0.0030	0.0007	4.324
Kam-pisc	Kud-pl	KDa-pisc	KDa-pisc-s	0.0013	0.0006	1.973
Kam-pisc	Tok-bn	KDa-pisc	KDa-pisc-s	0.0025	0.0007	3.628
Kam-pisc	Tok-insct	KDa-pisc	KDa-pisc-s	0.0019	0.0007	2.818
Kam-pl	BLe	KDa-pisc	KDa-pisc-s	-0.0018	0.0006	-2.932
Kam-pl	Dav-pl	Kir4-pisc	Kir4-pl	0.0058	0.0019	2.990
Kam-pl	Kam-bn	KDa-pisc	KDa-pisc-s	-0.0006	0.0003	-2.093
Kam-pl	Kam-bn	Kir3-pisc	Kud-pisc	-0.0036	0.0017	-2.051
Kam-pl	Kam-bn	Kir3-pisc	Kud-pl	-0.0035	0.0017	-1.994
Kam-pl	Kam-pisc	KDa-pisc	KDa-pisc-s	-0.0014	0.0004	-3.627
Kam-pl	Kam-pisc	Kir4-pl	Kud-pl	-0.0042	0.0021	-1.971
Kam-pl	Kam-pisc	Kud-pisc	Dav-pl	0.0043	0.0021	2.022
Kam-pl	Kam-pisc	Kud-pisc	KDa-pisc	0.0045	0.0022	2.025
Kam-pl	Kam-pisc	Kud-pisc	Kir4-pl	0.0045	0.0021	2.166
Kam-pl	KDa-pisc	Kud-pisc	Kud-pl	0.0012	0.0005	2.211
Kam-pl	KDa-pisc-s	Tok-bn	Tok-insct	-0.0021	0.0011	-2.008
Kam-pl	Kir3-pisc	Kud-pisc	Kud-pl	-0.0024	0.0008	-3.066
Kam-pl	Kir3-pl	Kud-pisc	Kud-pl	-0.0023	0.0007	-3.154
Kam-pl	Kir4-pl	Kud-pisc	Kud-pl	-0.0020	0.0007	-2.880
Kam-pl	Kud-pisc	KDa-pisc	KDa-pisc-s	0.0017	0.0006	2.735

KDa-pisc	KDa-pisc-s	Dav-pl	BLe	-0.0023	0.0005	-4.563
KDa-pisc	KDa-pisc-s	Kir4-pl	BLe	-0.0024	0.0006	-4.348
KDa-pisc	KDa-pisc-s	Tok-bn	BLe	-0.0029	0.0007	-4.165
KDa-pisc	KDa-pisc-s	Tok-bn	Tok-insct	-0.0006	0.0002	-3.149
KDa-pisc	KDa-pisc-s	MLe	BLe	-0.0007	0.0003	-2.756
KDa-pisc	KDa-pisc-s	Kud-pl	BLe	-0.0017	0.0006	-2.750
KDa-pisc	KDa-pisc-s	Tok-bn	Kam-bn	-0.0017	0.0006	-2.634
KDa-pisc	KDa-pisc-s	MLe	Tok-insct	0.0016	0.0006	2.580
KDa-pisc	KDa-pisc-s	BLe	Tok-insct	0.0023	0.0007	3.435
KDa-pisc	KDa-pisc-s	MLe	Dav-pl	0.0015	0.0004	3.448
KDa-pisc	KDa-pisc-s	MLe	Tok-bn	0.0022	0.0006	3.449
KDa-pisc	KDa-pisc-s	MLe	Kir4-pl	0.0017	0.0005	3.499
Kir3-pisc	BLe	KDa-pisc	KDa-pisc-s	-0.0027	0.0006	-4.248
Kir3-pisc	BLe	KDa-pisc-s	KDa-pl	0.0034	0.0016	2.185
Kir3-pisc	BLe	Kud-pisc	Kud-pl	0.0025	0.0007	3.762
Kir3-pisc	Kam-bn	KDa-pisc	KDa-pisc-s	-0.0015	0.0006	-2.289
Kir3-pisc	Kam-bn	Kud-pisc	Kud-pl	0.0025	0.0007	3.281
Kir3-pisc	Kam-pisc	KDa-pisc	KDa-pisc-s	-0.0023	0.0007	-3.368
Kir3-pisc	Kam-pisc	Kud-pisc	Kud-pl	0.0027	0.0008	3.388
Kir3-pisc	KDa-pisc	Kud-pisc	Kud-pl	0.0035	0.0009	3.930
Kir3-pisc	KDa-pisc-s	Kud-pisc	Kud-pl	0.0018	0.0008	2.226
Kir3-pisc	KDa-pl	Kud-pisc	Kud-pl	0.0018	0.0008	2.373
Kir3-pisc	Kir3-pl	Dav-pl	Kud-pl	-0.0064	0.0025	-2.565
Kir3-pisc	Kir3-pl	Dav-pl	KDa-pisc-s	-0.0053	0.0023	-2.245
Kir3-pisc	Kir3-pl	Dav-pl	Kam-bn	-0.0048	0.0023	-2.096
Kir3-pisc	Kir3-pl	Dav-pl	KDa-pl	-0.0047	0.0023	-1.982
Kir3-pisc	Kir3-pl	Kud-pisc	Kir4-pl	0.0047	0.0024	1.962
Kir3-pisc	Kir3-pl	KDa-pisc	Dav-pl	0.0051	0.0024	2.169
Kir3-pisc	Kir3-pl	Kud-pisc	Dav-pl	0.0065	0.0025	2.569
Kir3-pisc	MLe	KDa-pisc	KDa-pisc-s	-0.0020	0.0006	-3.423
Kir3-pisc	MLe	KDa-pisc-s	KDa-pl	0.0035	0.0016	2.250
Kir3-pisc	MLe	Kud-pisc	Kud-pl	0.0017	0.0006	2.845
Kir3-pisc	Tok-bn	Kud-pisc	Kud-pl	0.0015	0.0007	2.103
Kir3-pisc	Tok-insct	Kud-pisc	Kud-pl	0.0018	0.0007	2.430
Kir3-pl	Kam-bn	KDa-pisc	KDa-pisc-s	-0.0014	0.0006	-2.229
Kir3-pl	MLe	KDa-pisc	KDa-pisc-s	-0.0019	0.0005	-3.405
Kir3-pl	MLe	KDa-pisc-s	KDa-pl	0.0029	0.0015	2.001
Kir4-pisc	BLe	KDa-pisc	KDa-pisc-s	-0.0019	0.0005	-3.422
Kir4-pisc	BLe	Kud-pisc	Kud-pl	0.0013	0.0006	2.272
Kir4-pisc	Kam-pisc	KDa-pisc	KDa-pisc-s	-0.0014	0.0006	-2.301

Kir4-pisc	Kam-pisc	Kud-pisc	Kud-pl	0.0015	0.0008	2.024
Kir4-pisc	KDa-pisc	Kud-pisc	Kud-pl	0.0023	0.0008	2.846
Kir4-pisc	Kir3-pisc	Kud-pisc	Kud-pl	-0.0012	0.0003	-3.946
Kir4-pisc	Kir3-pisc	KDa-pisc	KDa-pisc-s	0.0008	0.0003	2.781
Kir4-pisc	Kir3-pl	Kud-pisc	Kud-pl	-0.0012	0.0004	-3.246
Kir4-pisc	Kir3-pl	KDa-pisc	KDa-pisc-s	0.0007	0.0003	2.283
Kir4-pisc	Kir4-pl	Dav-pl	Kam-bn	-0.0060	0.0020	-3.072
Kir4-pisc	Kir4-pl	Dav-pl	KDa-pl	-0.0060	0.0020	-3.060
Kir4-pisc	Kir4-pl	Dav-pl	KDa-pisc-s	-0.0061	0.0020	-3.060
Kir4-pisc	Kir4-pl	Dav-pl	Kud-pl	-0.0060	0.0020	-2.943
Kir4-pisc	Kir4-pl	Dav-pl	BLe	-0.0038	0.0013	-2.865
Kir4-pisc	Kir4-pl	Kud-pisc	Kud-pl	-0.0008	0.0003	-2.563
Kir4-pisc	Kir4-pl	KDa-pisc	KDa-pisc-s	0.0005	0.0003	2.000
Kir4-pisc	Kir4-pl	Kud-pisc	Dav-pl	0.0052	0.0021	2.514
Kir4-pisc	Kir4-pl	MLe	Dav-pl	0.0035	0.0014	2.580
Kir4-pisc	Kir4-pl	Kam-pisc	Dav-pl	0.0054	0.0020	2.680
Kir4-pisc	Kir4-pl	KDa-pisc	Dav-pl	0.0066	0.0020	3.305
Kir4-pisc	Kud-pisc	KDa-pisc	KDa-pisc-s	0.0016	0.0006	2.572
Kir4-pisc	MLe	KDa-pisc	KDa-pisc-s	-0.0011	0.0005	-2.224
Kud-pisc	BLe	KDa-pisc	KDa-pisc-s	-0.0035	0.0007	-4.755
Kud-pisc	BLe	KDa-pisc-s	KDa-pl	0.0051	0.0022	2.320
Kud-pisc	Dav-pl	KDa-pisc	KDa-pisc-s	-0.0012	0.0006	-2.095
Kud-pisc	Kam-bn	KDa-pisc	KDa-pisc-s	-0.0022	0.0006	-3.442
Kud-pisc	Kud-pl	KDa-pisc	KDa-pisc-s	-0.0018	0.0004	-4.726
Kud-pisc	Kud-pl	KDa-pisc	Kir4-pl	-0.0032	0.0008	-3.844
Kud-pisc	Kud-pl	KDa-pisc	KDa-pl	-0.0017	0.0005	-3.321
Kud-pisc	Kud-pl	KDa-pisc	Dav-pl	-0.0025	0.0008	-3.097
Kud-pisc	Kud-pl	MLe	Kir4-pl	-0.0014	0.0005	-2.545
Kud-pisc	Kud-pl	KDa-pisc	Tok-bn	-0.0020	0.0008	-2.456
Kud-pisc	Kud-pl	KDa-pisc	MLe	-0.0018	0.0008	-2.368
Kud-pisc	Kud-pl	KDa-pisc	Tok-insct	-0.0017	0.0008	-2.108
Kud-pisc	Kud-pl	KDa-pisc	Kam-bn	-0.0011	0.0005	-2.006
Kud-pisc	Kud-pl	Kir3-pl	Tok-bn	0.0015	0.0008	1.965
Kud-pisc	Kud-pl	Kir4-pl	Tok-insct	0.0014	0.0007	2.059
Kud-pisc	Kud-pl	Kir4-pl	KDa-pl	0.0014	0.0007	2.078
Kud-pisc	Kud-pl	Kir3-pl	KDa-pisc-s	0.0018	0.0008	2.227
Kud-pisc	Kud-pl	Kir3-pl	Tok-insct	0.0018	0.0008	2.292
Kud-pisc	Kud-pl	Kir3-pl	KDa-pl	0.0018	0.0007	2.475
Kud-pisc	Kud-pl	MLe	BLe	0.0007	0.0003	2.658
Kud-pisc	Kud-pl	Dav-pl	BLe	0.0014	0.0005	2.927

Kud-pisc	Kud-pl	Kir3-pl	MLe	0.0017	0.0006	2.979
Kud-pisc	Kud-pl	Kir4-pl	Kam-bn	0.0021	0.0007	3.089
Kud-pisc	Kud-pl	Kir3-pl	Kam-bn	0.0024	0.0007	3.323
Kud-pisc	Kud-pl	Kir4-pl	BLe	0.0021	0.0006	3.494
Kud-pisc	Kud-pl	Kir3-pl	BLe	0.0024	0.0006	3.817
Kud-pisc	Kud-pl	Kir3-pl	KDa-pisc	0.0035	0.0009	3.973
Kud-pisc	MLe	KDa-pisc	KDa-pisc-s	-0.0027	0.0007	-4.041
Kud-pisc	MLe	KDa-pisc-s	KDa-pl	0.0052	0.0022	2.357
Kud-pl	BLe	KDa-pisc-s	KDa-pl	0.0051	0.0022	2.322
MLe	KDa-pisc-s	Tok-bn	Tok-insct	-0.0028	0.0013	-2.158
MLe	Kud-pl	KDa-pisc-s	KDa-pl	-0.0052	0.0022	-2.357
Tok-bn	Tok-insct	Kam-bn	KDa-pisc-s	-0.0026	0.0011	-2.263
Tok-bn	Tok-insct	BLe	KDa-pisc-s	-0.0028	0.0013	-2.152

Table S12. Most likely demographic models for each lake population. Shown are the fours best fitting models for each population/comparison and the respective ΔAIC between them.

Lake/Ecotypes	N pop.	N SNPs	Best model	2nd best (ΔAIC)	3rd best (ΔAIC)	4th best (ΔAIC)
Awe	2pop	15,891	IMint	SCint (12.4)	IMchange (25.1)	SC (50.0)
Tay	2pop	15,000	IMint	SCint (0.15)	SC (12.5)	IMchange (13.7)
naS	2pop	13,687	IMint	SCint (4.8)	IMchange (21.8)	SC (104)
Dug-bn - Dug-pl	2pop	15,096	IMint	IMchange (3.8)	SCint (68.8)	SC (69.1)
Dug-bn - Uai	2pop	11,372	IMchange	IMint (10.3)	SCint (95.5)	SC (137)
Dug-pl - Uai	2pop	7,231	IMint	IMchange (0.65)	SCint (1.9)	SC (4.1)
Dug	3pop	11,753	Imint	HybMig (89.3)	IM (129)	Int (142)
Dav	2pop	6,931	SCint	IMint (3.1)	IMchange (5.1)	SC (26.4)
Kir-3	2pop	6.493	SC	Imchange (3.8)	SCint (4.4)	IMint (6.7)
Kir-4	2pop	5,641	SCint	SC (0.65)	IMint (3.5)	IMchange (3.7)
Tok	2pop	4,285	SC	SCint (2.9)	IMchange (3.3)	IMint (5.9)
Kam-PiscPl	2pop	6,686	SC	SCint (0.4)	IMchange (2.9)	IMint (4.0)
Kam-BnPl	2pop	7,068	SC	SCint (2.3)	IMchange (5.5)	IMint (7.8)
Kam-BnPisc	2pop	6,941	SC	SCint (0.14)	IMint (4.1)	IMchange (4.9)
KDa-PiscPl	2pop	6,528	IMint	SC (4.2)	SCint (4.5)	IMchange (9.7)
KDa-PiscSPl	2pop	6,718	SC	SCint (2.7)	IMchange (3.9)	IMint (8.3)
KDa-PiscPiscS	2pop	6,700	SC	IMchange (1.5)	SCint (2.1)	IMint (7.1)
Kud	2pop	6,194	SCint	IMint (4.5)	IMchange (8.2)	SC (18.1)
Kam	3pop	6,634	SC	IM (0.33)	SCint (4.3)	IMchange (4.6)
KDa	3pop	6,310	SC	IntSC (5.27)	SCint (8.4)	Hyb (12.9)

Note: 2pop models tested the demographic history for ecotype pairs and 3pop models modelled the demographic history for three populations/ecotypes. N SNPs gives the number of SNPs used to build the folded site frequency spectrum.

Table S13. Parallelism in allele frequencies of loci associated with ecotype (BSLMM, PIP > 0.1) for each lineage. Shown are the results from individual ANCOVAs and estimated partial η^2 values for each effect term by tested principal component axis. The effect term explaining most the phenotypic variance is highlighted in bold.

Lineage	PC	Ecotype	Lake	Ecotype x Lake	P-value
Atlantic	PC1	0.087	**0.967**	0.155	< 0.001
	PC2	0.154	**0.932**	0.235	< 0.001
	PC3	0.455	**0.715**	0.509	< 0.001
	PC4	**0.704**	0.529	0.304	< 0.001
	PC5	0.009	0.103	**0.543**	0.187
Siberian	PC1	0.943	**0.997**	0.508	< 0.001
	PC2	0.918	**0.995**	0.662	< 0.001
	PC3	0.972	**0.981**	0.806	< 0.001
	PC4	0.334	**0.987**	0.846	< 0.001
	PC5	0.918	0.896	**0.944**	< 0.001
	PC6	0.896	0.832	**0.901**	< 0.001
	PC7	0.508	0.849	**0.907**	< 0.001
	PC8	0.594	**0.949**	0.607	< 0.001
	PC9	0.677	**0.903**	0.826	< 0.001

Table S14. Expected and observed number of shared differentially expressed genes (DEGs) for each pairwise comparison.

Comparison	Obs. shared DEGs	Shared genes	Exp. shared DEGs	P-value	Bonferroni P-value
Awe (bn-pl) vs. Dug (bn-pl)	47	33136	8.0	7.6E-44	1.6E-42
Awe (bn-pl) vs. Kam (bn-pisc)	5	32023	7.5	8.2E-01	1.7E+01
Awe (bn-pl) vs. Kam (bn-pl)	16	32023	7.4	7.4E-04	1.6E-02
Awe (bn-pl) vs. Kam (pl-pisc)	25	32023	7.6	1.5E-10	3.2E-09
Awe (bn-pl) vs. Tay (bn-pl)	28	33286	7.6	7.0E-14	1.5E-12
Awe (bn-pl) vs. Tok (bn-insct)	71	33537	8.5	8.1E-102	1.7E-100
Dug (bn-pl) vs. Kam (bn-pisc)	19	31474	8.2	8.1E-05	1.7E-03
Dug (bn-pl) vs. Kam (bn-pl)	47	31474	8.1	4.3E-43	9.0E-42
Dug (bn-pl) vs. Kam (pl-pisc)	50	31474	8.3	4.2E-48	8.8E-47
Dug (bn-pl) vs. Tay(bn-pl)	24	32638	8.2	2.0E-08	4.2E-07
Dug (bn-pl) vs. Tok (bn-insct)	129	32787	9.0	0.0E+00	0.0E+00
Kam (bn-pisc) vs. Kam (bn-pl)	37	33485	8.7	3.7E-22	7.8E-21
Kam (bn-pisc)vs. Kam (pl-pisc)	58	33485	8.5	3.4E-65	7.0E-64
Kam (bn-pisc)vs. Tay (bn-pl)	12	31644	8.2	9.4E-02	2.0E+00
Kam (bn-pisc)vs. Tok (bn-insct)	47	31774	8.2	2.6E-42	5.4E-41
Kam (bn-pl) vs. Kam (pl-pisc)	128	33485	8.2	0.0E+00	0.0E+00
Kam (bn-pl) vs. Tay (bn-pl)	24	31644	8.0	9.0E-09	1.9E-07
Kam (bn-pl) vs. Tok (bn-insct)	75	31774	8.3	2.8E-119	5.9E-118
Kam pl-pisc vs. Tay (bn-pl)	21	31644	9.1	4.2E-05	8.7E-04
Kam pl-pisc vs. Tok (bn-insct)	189	31774	9.1	0.0E+00	0.0E+00
Tay (bn-pl) vs. Tok (bn-insct)	54	32821	9.5	1.4E-47	2.8E-46

Table S15. Number of unique and shared differentially expressed genes.

Number of comparisons	Number of shared DEGs
5	4
4	19
3	122
2	586
1	2981

Table S16. Gene co-expression modules in network generated from 1,512 ecotype associated genes. Traits (lake and ecotype) significantly correlated (FDR < 0.05) each module are given, as well the direction of expression for ecotype: pl = up-regulated in planktivorous (over benthivorous) and bn = up-regulated in benthivorous (over planktivorous).

Module	Number genes in module	Significantly (FDR < 0.05) associated with	Ecotype associated with
Blue	221	Ecotype	pl
Cyan	42	Ecotype	pl
Magenta	73	Lake + Ecotype	pl
MidnightBlue	31	Ecotype	pl
Salmon	43	lake + Ecotype	pl
Black	98	Ecotype	pl
Green	159	Ecotype	pl
GreenYellow	43	Ecotype	bn
LightCyan	73	Lake + Ecotype	bn
Turquoise	384	Ecotype	bn
Brown	162	Ecotype	bn
Pink	73	Ecotype	bn
Red	110	Ecotype	bn

Table S17. Significantly enriched (FDR < 0.1) KEGG pathways for differentially expressed genes and associated populations.

Populations	KEGG pathway	Comparisons
3	Oxidative phosphorylation - Danio rerio (zebrafish)	Awe, Kam bn-pisc, Tay
2	Ribosome biogenesis in eukaryotes - Danio rerio (zebrafish)	Awe, Tok
2	ECM-receptor interaction - Danio rerio (zebrafish)	Awe, Tok
1	Glycolysis / Gluconeogenesis - Danio rerio (zebrafish)	Tok
1	Fructose and mannose metabolism - Danio rerio (zebrafish)	Tok
1	Valine, leucine and isoleucine degradation - Danio rerio (zebrafish)	Dug
1	Glycerolipid metabolism - Danio rerio (zebrafish)	Dug
1	Biosynthesis of amino acids - Danio rerio (zebrafish)	Tok
1	ABC transporters - Danio rerio (zebrafish)	Awe
1	Ribosome - Danio rerio (zebrafish)	Awe
1	DNA replication - Danio rerio (zebrafish)	Tok
1	Proteasome - Danio rerio (zebrafish)	Tok
1	Cell cycle - Danio rerio (zebrafish)	Tok
1	Oocyte meiosis - Danio rerio (zebrafish)	Tok
1	Progesterone-mediated oocyte maturation - Danio rerio (zebrafish)	Tok

Table S18. Differentially expressed genes that are shared between four and five ecotype pairs.

LG	Contig	Start	End	Transcript	N. eco	Name
chr2	Contig3082	449329	449805	RNA94238_R0	4	mid1-interacting protein 1-B-like
chr20b	Contig1638	1719889	1757678	RNA55399_R0	4	dehydrogenase/reductase SDR family member 11
chr20b	Contig4545	275730	276437	RNA26656_R0	4	Heat shock protein Hsp-16.1/Hsp-16.11
chr24	Contig603	1765545	1793398	RNA71652_R0	4	ras-related and estrogen-regulated growth inhibitor
chr26	Contig669	1528764	1546396	RNA66169_R1	4	cytosolic carboxypeptidase 2-like
chr28	Contig1092	180742	183781	RNA62963_R0	4	uncharacterized protein LOC106571495
chr32	Contig8	1391080	1395568	RNA58210_R0	4	carbonic anhydrase-like
Un	Contig1119	35572	39011	RNA71231_R0	4	regucalcin
Un	Contig1422	170260	172193	RNA75643_R1	4	high mobility group protein B3-like
Un	Contig4180	54066	62576	RNA69074_R0	4	kelch-like ECH-associated protein 1
Un	Contig4656	82820	83542	RNA13977_R0	4	heat shock protein Hsp-16.1/Hsp-16.11-like
Un	Contig4861	140387	141382	RNA26199_R3	4	Hemoglobin subunit alpha-4
Un	Contig4861	180208	180927	RNA26199_R0	4	hemoglobin subunit alpha-4-like
Un	Contig4861	185293	186058	RNA14478_R2	4	Hemoglobin subunit alpha
Un	Contig4861	214581	215334	RNA26201_R0	4	Hemoglobin subunit beta-1
Un	Contig4861	216743	217549	RNA14474_R2	4	hemoglobin subunit alpha-like
Un	Contig854	342514	366628	RNA60937_R0	4	collagen alpha-2(I) chain precursor
Un	Contig944	267089	267820	RNA26199_R1	4	hemoglobin subunit alpha-4
Un	Contig944	272233	273055	RNA25924_R2	4	Hemoglobin subunit beta-1
chr4	Contig2374	608188	732140	RNA41267_R0	5	ATP-binding cassette sub-family C member 8
chr18	Contig3023	278846	282347	RNA78504_R1	5	fructose-bisphosphate aldolase A
chr20a	Contig1143	437390	438376	RNA26558_R0	5	nucleoside diphosphate kinase A-like
chr31	Contig979	244747	245057	RNA59594_R1	5	uncharacterized protein LOC100194706

Note: LG – chromosome the contig is located on (Un – unanchored contig). Start and End position of the transcript on the respective contig. Transcript – Name of the transcript/gene. N. eco – Number of ecotype pairs a gene is differentially expressed in. Name – Name of the respective gene.

Table S19. Significantly enriched (FDR < 0.05) KEGG pathways in co-expressed gene modules.

Module	Up-regulated in ecotype	KEGG	KEGG pathway description	FDR
Blue	pl	dre00240	Pyrimidine metabolism - *Danio rerio* (zebrafish)	0.04
Magenta	pl	dre04510	Focal adhesion - Danio rerio (zebrafish)	0.04
Magenta	pl	dre04512	ECM-receptor interaction - Danio rerio (zebrafish)	0.04
Black	pl	dre00561	Glycerolipid metabolism - Danio rerio (zebrafish)	0.02
Black	pl	dre04350	TGF-beta signalling pathway - Danio rerio (zebrafish)	0.02
Turquoise	bn	dre04070	Phosphatidylinositol signalling system - Danio rerio (zebrafish)	0.01
Turquoise	bn	dre00562	Inositol phosphate metabolism - Danio rerio (zebrafish)	0.03

Table S20. KEGG pathways and biological processes (GOterms) associated with GWAS candidate genes by lineage.

Dataset	Biological process (GOterm) or KEGG pathway	Genes	Fold	Pvalue	FDR	Lin
GO:0021986	Habenula development	3	67.6	9.3E-06	0.03	SIB
GO:0021538	Epithalamus development	3	46.8	3.1E-05	0.05	SIB
GO:0071679	Commissural neuron axon guidance	2	81.1	2.4E-04	0.19	SIB
GO:0048048	Embryonic eye morphogenesis	3	21.0	3.8E-04	0.19	SIB
GO:0065009	Regulation of molecular function	11	3.3	4.0E-04	0.19	SIB
GO:0009966	Regulation of signal transduction	10	3.5	4.9E-04	0.19	SIB
GO:0051336	Regulation of hydrolase activity	8	4.2	5.2E-04	0.19	SIB
GO:0048667	Cell morphogenesis involved in neuron differentiation	6	5.7	6.0E-04	0.19	SIB
GO:0050790	Regulation of catalytic activity	10	3.2	8.1E-04	0.19	SIB
GO:0042476	Odontogenesis	2	45.1	8.4E-04	0.19	SIB
dre04510	Focal adhesion - Danio rerio (zebrafish)	4	3.2	3.3E-02	1.00	SIB
dre04310	Wnt signaling pathway - Danio rerio (zebrafish)	3	3.4	5.6E-02	1.00	SIB
dre04512	ECM-receptor interaction - Danio rerio (zebrafish)	2	5.0	6.1E-02	1.00	SIB
dre03060	Protein export - Danio rerio (zebrafish)	1	9.1	1.0E-01	1.00	SIB
dre04012	ErbB signaling pathway - Danio rerio (zebrafish)	2	3.6	1.1E-01	1.00	SIB
dre04140	Autophagy - Danio rerio (zebrafish)	1	6.9	1.4E-01	1.00	SIB
dre04514	Cell adhesion molecules (CAMs) - Danio rerio (zebrafish)	2	2.9	1.5E-01	1.00	SIB
dre02010	ABC transporters - Danio rerio (zebrafish)	1	5.1	1.8E-01	1.00	SIB
dre00051	Fructose and mannose metabolism - Danio rerio (zebrafish)	1	4.7	1.9E-01	1.00	SIB
dre00620	Pyruvate metabolism - Danio rerio (zebrafish)	1	4.5	2.0E-01	1.00	SIB
GO:0008063	Toll signaling pathway	1	337.9	3.0E-03	1.00	ATL
GO:0016558	Protein import into peroxisome matrix	1	337.9	3.0E-03	1.00	ATL
GO:0031057	Negative regulation of histone modification	1	337.9	3.0E-03	1.00	ATL
GO:0006625	Protein targeting to peroxisome	1	281.6	3.6E-03	1.00	ATL
GO:0035066	Positive regulation of histone acetylation	1	281.6	3.6E-03	1.00	ATL
GO:004357	Peroxisomal transport	1	281.6	3.6E-	1.00	ATL

4				03		
GO:0072662	Protein localization to peroxisome	1	281.6	3.6E-03	1.00	ATL
GO:0072663	Establishment of protein localization to peroxisome	1	281.6	3.6E-03	1.00	ATL
GO:1901985	Positive regulation of protein acetylation	1	281.6	3.6E-03	1.00	ATL
GO:2000758	Positive regulation of peptidyl-lysine acetylation	1	281.6	3.6E-03	1.00	ATL
dre04370	VEGF signaling pathway - Danio rerio (zebrafish)	2	23.1	3.0E-03	0.28	ATL
dre04620	Toll-like receptor signaling pathway - Danio rerio (zebrafish)	2	18.7	4.5E-03	0.28	ATL
dre04914	Progesterone-mediated oocyte maturation - Danio rerio (zebrafish)	2	16.9	5.5E-03	0.28	ATL
dre04933	AGE-RAGE signaling pathway in diabetic complications - Danio rerio (zebrafish)	2	14.6	7.3E-03	0.28	ATL
dre04068	FoxO signaling pathway - Danio rerio (zebrafish)	2	11.0	1.3E-02	0.39	ATL
dre04261	Adrenergic signaling in cardiomyocytes - Danio rerio (zebrafish)	2	9.0	1.9E-02	0.47	ATL
dre04010	MAPK signaling pathway - Danio rerio (zebrafish)	2	5.8	4.2E-02	0.92	ATL
dre04622	RIG-I-like receptor signaling pathway - Danio rerio (zebrafish)	1	14.9	6.6E-02	1.00	ATL
dre04146	Peroxisome - Danio rerio (zebrafish)	1	10.6	9.0E-02	1.00	ATL
dre04920	Adipocytokine signaling pathway - Danio rerio (zebrafish)	1	10.3	9.3E-02	1.00	ATL

Note: Genes – Number of candidate genes belonging to the identified pathway. Fold – Fold enrichment for each pathway, with P-value (based on compairison to total number of genes in the pathway) and false-discovery rate (FDR). Results are shown by lineage (Lin). ATL – Atlantic, SIB – Siberian.

Table S21. KEGG pathways and biological processes (GOterms) associated with genes under selection.

Dataset	Biological process (GOterm) or KEGG pathway	Genes	Fold	Pvalue	FDR	Lin
GO:0048468	Cell development	21	2.4	1.5E-04	0.2	SIB
GO:0098662	Inorganic cation transmembrane transport	12	3.5	1.6E-04	0.2	SIB
GO:0098655	Cation transmembrane transport	13	3.2	1.8E-04	0.2	SIB
GO:0098660	Inorganic ion transmembrane transport	12	3.1	4.8E-04	0.4	SIB
GO:0032879	Regulation of localization	13	2.7	9.5E-04	0.4	SIB
GO:0030154	Cell differentiation	26	1.9	1.0E-03	0.4	SIB
GO:0007409	Axonogenesis	9	3.5	1.1E-03	0.4	SIB
GO:0071804	Cellular potassium ion transport	6	5.0	1.2E-03	0.4	SIB
GO:0071805	Potassium ion transmembrane transport	6	5.0	1.2E-03	0.4	SIB
GO:0048667	Cell morphogenesis involved in neuron differentiation	9	3.3	1.7E-03	0.4	SIB
dre04512	ECM-receptor interaction - Danio rerio (zebrafish)	6	5.5	7.3E-04	0.1	SIB
dre04510	Focal adhesion - Danio rerio (zebrafish)	8	2.4	1.9E-02	1.0	SIB
dre00061	Fatty acid biosynthesis - Danio rerio (zebrafish)	2	9.2	1.9E-02	1.0	SIB
dre00562	Inositol phosphate metabolism - Danio rerio (zebrafish)	4	3.5	2.8E-02	1.0	SIB
dre03460	Fanconi anemia pathway - Danio rerio (zebrafish)	3	4.3	3.3E-02	1.0	SIB
dre00512	Mucin type O-glycan biosynthesis - Danio rerio (zebrafish)	2	5.3	5.5E-02	1.0	SIB
dre00640	Propanoate metabolism - Danio rerio (zebrafish)	2	4.6	6.9E-02	1.0	SIB
dre04115	p53 signaling pathway - Danio rerio (zebrafish)	3	3.0	7.8E-02	1.0	SIB
dre04912	GnRH signaling pathway - Danio rerio (zebrafish)	4	2.3	9.5E-02	1.0	SIB
dre04933	AGE-RAGE signaling pathway in diabetic complications - Danio rerio (zebrafish)	4	2.3	9.5E-02	1.0	SIB
GO:0009653	Anatomical structure morphogenesis	26	2.3	3.1E-05	0.1	ATL
GO:0035295	Tube development	10	4.1	1.5E-04	0.2	ATL
GO:0048869	Cellular developmental process	25	2.1	2.0E-04	0.2	ATL
GO:0048565	Digestive tract development	5	8.0	3.9E-04	0.2	ATL
GO:0097529	Myeloid leukocyte migration	4	10.3	5.7E-04	0.2	ATL
GO:0030154	Cell differentiation	23	2.1	6.1E-04	0.2	ATL
GO:0030522	Intracellular receptor signaling pathway	5	7.1	6.6E-04	0.2	ATL
GO:0048646	Anatomical structure formation involved in morphogenesis	13	2.8	7.9E-04	0.2	ATL
GO:0009888	Tissue development	18	2.3	8.2E-04	0.2	ATL
GO:0048598	Embryonic morphogenesis	12	2.9	8.7E-04	0.2	ATL
dre04910	Insulin signaling pathway - Danio rerio (zebrafish)	6	2.6	2.6E-02	1.0	ATL
dre04920	Adipocytokine signaling pathway - Danio rerio (zebrafish)	4	3.4	3.0E-02	1.0	ATL
dre04150	mTOR signaling pathway - Danio rerio (zebrafish)	6	2.4	3.6E-02	1.0	ATL
dre04330	Notch signaling pathway - Danio rerio (zebrafish)	3	4.0	3.9E-02	1.0	ATL
dre04068	FoxO signaling pathway - Danio rerio (zebrafish)	5	2.3	7.1E-02	1.0	ATL
dre04310	Wnt signaling pathway - Danio rerio (zebrafish)	5	2.2	7.8E-02	1.0	ATL
dre04141	Protein processing in endoplasmic reticulum -	5	2.1	8.5E-02	1.0	ATL

	Danio rerio (zebrafish)					
dre03022	Basal transcription factors - Danio rerio (zebrafish)	2	3.9	9.5E-02	1.0	ATL
dre04916	Melanogenesis - Danio rerio (zebrafish)	4	2.2	1.1E-01	1.0	ATL
dre00920	Sulfur metabolism - Danio rerio (zebrafish)	1	7.7	1.2E-01	1.0	ATL

Note: Genes – Number of genes under selection within each pathway. Fold – Fold enrichment for each pathway, with P-value (based on compairison to total number of genes in the pathway) and false-discovery rate (FDR). Results are shown by lineage (Lin). ATL – Atlantic, SIB – Siberian.

Table S22. Hyperparameters describing the genetic architecture underlying ecotype divergence in each lineage inferred using the BSLMM in *GEMMA*. Given are the mean and 95%-confidence interval (CI) for each hyperparameter.

Lineage	Hyperparameter	Mean	2.50%-CI	97.50%-CI
Atlantic	h	0.9712	0.8683	1.0000
	PVE	0.9676	0.8617	0.9999
	rho	0.7456	0.3515	0.9785
	PGE	0.6312	0.2140	0.9763
	pi	0.0083	0.0010	0.0222
	n.gamma	51	7	133
Siberian	h	0.9992	0.9962	1.0000
	PVE	0.9992	0.9967	1.0000
	rho	0.9707	0.9430	0.9909
	PGE	0.9679	0.9129	0.9931
	pi	0.0205	0.0136	0.0283
	n.gamma	91	67	118

Note: Parameters were estimated using a 6,039 SNPs in the Atlantic and 4,475 SNPs in the Siberian lineage. PVE – proportion of phenotypic variance explained by genotype; PGE – proportion of genetic variance explained by sparse (major) effects; n_gamma – the expected number of sparse (major) effect loci; h – approximate parameter used for the prior specification of the expected PVE; rho – used for prior specification of PGE.