Title: The Genomic Consistency of the Loss of Anadromy in an Arctic Fish (*Salvelinus alpinus*)

Authors: Sarah J Salisbury[1*], Gregory R McCracken[1], Robert Perry[2], Donald Keefe[3], Kara KS Layton[4,5], Tony Kess[4], Cameron M Nugent[6], Jong S Leong[7], Ian R Bradbury[1,4,5], Ben F Koop[7,8], Moira M Ferguson[6], Daniel E Ruzzante[1]

1 Department of Biology, Dalhousie University, Halifax, NS, Canada

2 Department of Environment, Fish and Wildlife Division, Government of Yukon, Whitehorse, YT, Canada

3 Department of Fisheries, Forestry, and Agriculture, Forestry and Wildlife Branch, Government of Newfoundland and Labrador, Corner Brook, NL, Canada

4 Department of Fisheries and Oceans, Northwest Atlantic Fisheries Centre, St. John's, NL, Canada

5 Department of Ocean Sciences, Memorial University of Newfoundland, St. John's, NL, Canada

6 Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

7 Department of Biology, University of Victoria, Victoria, BC, Canada

8 Centre for Biomedical Research, University of Victoria, Victoria, BC, Canada

Corresponding author email: sarahsalisbury13@gmail.com

Co-author emails: gregory.mccracken@dal.ca, Robert.Perry@yukon.ca, donkeefe@gov.nl.ca, karakslayton@gmail.com, Tony.Kess@dfo-mpo.gc.ca, camnugent@gmail.com, jong@uvic.ca, Ian.Bradbury@dfo-mpo.gc.ca, bkoop@uvic.ca, mmfergus@uoguelph.ca, Daniel.Ruzzante@dal.ca

Current Address for KKS Layton: School of Biological Sciences, University of Aberdeen, Aberdeen, United Kingdom

Keywords: parallelism, allopatry, landlocked, anadromy, incipient speciation, SNPs

Article Type: Main Article

ORCID

Sarah J. Salisbury https://orcid.org/0000-0001-7637-7742

Gregory R. McCracken https://orcid.org/0000-0002-1701-1529

Kara K. S. Layton https://orcid.org/0000-0002-4302-3048

Tony Kess https://orcid.org/0000-0002-1079-3791

Cameron M. Nugent https://orcid.org/0000-0002-1135-2605

Jong S. Leong https://orcid.org/0000-0001-6521-9239

Ben F. Koop https://orcid.org/0000-0003-0045-5200

Ian R. Bradbury https://orcid.org/0000-0002-8152-4943

Daniel E. Ruzzante https://orcid.org/0000-0002-8536-8335

Abstract

The potentially significant genetic consequences associated with the loss of migratory capacity of diadromous fishes which have become "landlocked" in freshwater are poorly understood. Consistent selective pressures associated with freshwater residency may drive repeated differentiation both between allopatric landlocked and anadromous populations and within landlocked populations (resulting in sympatric morphs). Alternatively, the strong genetic drift anticipated in isolated landlocked populations could hinder consistent adaptation, limiting genetic parallelism. Understanding the degree of genetic parallelism underlying differentiation

has implications for both the predictability of evolution and management practices. We employed an 87k SNP array to examine the genetic characteristics of landlocked and anadromous Arctic Charr (*Salvelinus alpinus*) populations from five drainages within Labrador, Canada. One gene was detected as an outlier between sympatric, size-differentiated morphs in each of two landlocked lakes. While no single locus differentiated all replicate pairs of landlocked and anadromous populations, several SNPs, genes, and paralogs, were consistently detected as outliers in at least 70% of these pairwise comparisons. A significant C-score suggested the amount of shared outlier SNPs across all paired landlocked and anadromous populations was greater than expected by chance. Our results indicate that despite their isolation, selection due to the loss of diadromy may drive consistent genetic responses in landlocked populations.

Introduction

The loss of migratory capacity is a fundamental promoter of neutral and adaptive differentiation (Waters et al. 2020). Such a loss is frequently observed in diadromous fishes, whose "landlocking" in post-glacial lakes offers a unique opportunity to study the predictability of evolution (Elmer and Meyer 2011). These populations were formed subsequent to the last glacial maximum (< 20 000 years) when anadromous individuals became trapped in freshwater environments (typically lakes) through a variety of mechanisms including isostatic rebound and physical impoundments (Lee and Bell 1999). Once landlocked, fish maintain a freshwater resident life history, typically exchanging minimal to no gene flow with other populations (e.g., Hindar et al. 1991; Palkovacs et al. 2008; Delgado et al. 2019). This independence makes them ideal natural replicates of evolution (Lee and Bell 1999) that may be compared to assess the

consistency of their genetic differentiation in response to common selective pressures (Elmer et al. 2014; Jacobs et al. 2020; McGee et al. 2020).

The loss of anadromy has predictable selective consequences (Delgado and Ruzzante 2020). For example, landlocked and diadromous populations may reliably differ in their diets (Palkovacs et al. 2008) as well as the predators (Hendry et al. 2004), parasites (Bouillon and Dempson 1989), and fishing pressure (Hendry et al. 2004) they experience. Landlocked populations are released from the selective pressures imposed by saltwater environments resulting in predictable physiological changes in osmoregulation (Velotta et al. 2014) and swimming capacity (Velotta et al. 2018). Given these consistent environmental and phenotypic differences, one might expect that the same genetic differences repeatedly underlie this adaptation to the loss of anadromy (i.e., genetic parallelism). Parallelism has been observed in Threespine Stickleback (*Gasterosteus aculeatus*) where loci such as *Eda* (Nelson and Cresko 2018) and *Pitx* (Xie et al. 2019) are known to play a role in the repeated colonization of freshwater from the marine environment. However, the degree of genomic parallelism that underlies the allopatric differentiation of landlocked and diadromous fishes more broadly remains largely unknown (Delgado et al. 2020; Kjærner-Semb et al. 2020).

Divergent selection can also drive morph differentiation within lakes (Lee and Bell 1999; Schultz and McCormick 2012). If divergent selection is consistent within multiple lakes, this can result in repeated morph differentiation (Schluter and Nagel 1995; Schluter 1996). For example, sympatric limnetic and benthic morphs have recurrently evolved in many lacustrine stickleback (Taylor and McPhail 1999) and whitefish (Bernatchez et al. 2010) populations. A growing number of studies have employed genomic data to investigate the degree of genetic parallelism underlying lacustrine radiations (e.g., Elmer et al. 2014; Meier et al. 2018; Jacobs et al. 2020;

Härer et al. 2021; Jacobs and Elmer 2021). However, expectations of genetic parallelism underlying repeated morph differentiation in landlocked lakes are generally unknown, particularly in non-model fish.

Different "levels" of genetic parallelism could also be important for driving repeated morph differentiation (Salisbury and Ruzzante, 2022). For example, identical alleles/SNPs could contribute to the same phenotypic differentiation. Alternatively, parallelism could occur at the level of the gene, where different SNPs/mutations but within the same gene cause morph differentiation in different locations. Repeated morph differentiation could also be due to the employment of different paralogous copies of the same gene. Such genetic parallelism at the level of the paralog is largely unexplored (Nichols et al. 2008; Conte et al. 2012) but may be particularly important for salmonids given their recent whole genome duplication resulting in numerous homeologs (Macqueen and Johnston 2014).

Many mechanisms could potentially undermine this genetic parallelism. For instance, genetic drift facilitated by a lack of gene flow (Bernatchez et al. 2002), founder effects (Ramstad et al. 2004), or by reduced carrying capacities (McCracken et al. 2013) may greatly impede genetic parallelism. Genetic parallelism could also be limited if replicate morphs are subject to different local selective pressures (Campbell and Bernatchez 2004) or are phylogenetically distant causing a reduction in shared genetic variation (Conte et al. 2012). Alternatively, genetic parallelism will be reduced where multiple genetic pathways can be employed to achieve the same morph differentiation. We were therefore interested in examining the degree of genetic parallelism and the factors limiting it in polymorphic Arctic Charr (*Salvelinus alpinus*) 1) between replicate landlocked and anadromous populations and 2) between replicate sympatric morphs within landlocked lakes.

Labrador, Canada is an ideal location for such work as it contains numerous landlocked charr populations inhabiting the same drainage as anadromous populations (Anderson 1985), thus forming natural paired replicates of allopatric differentiation. This differentiation has occurred recently as Labrador was deglaciated 9000 years BP (Bryson et al. 1969; Occhietti et al. 2011). Anadromous populations are fished in Labrador as part of economically and culturally important subsistence, recreational, and commercial fisheries (Andrews and Lear 1956; Scott and Crossman 1973; Dempson et al. 2008) and have been genetically well-studied in Labrador (e.g., Bernatchez et al. 1998; Layton et al. 2020; 2021). However, comparatively little is known about landlocked populations. We previously found lower genetic diversity in landlocked than in anadromous populations using microsatellites (Salisbury et al. 2018) and mtDNA (Salisbury et al. 2019). Though neutral genetic differentiation between landlocked and anadromous charr populations has previously been assessed (Bernatchez et al. 1998; Kapralova et al. 2011; Salisbury et al. 2018), adaptive genetic differences between landlocked and anadromous populations remain uncharacterized in this species.

Genetically distinguishable sympatric Arctic Charr morphs have been previously identified in two landlocked lakes in Labrador using neutral microsatellites (Salisbury et al. 2018). Size-differentiated, genetically distinguishable ecotypes of Arctic Charr have been also observed within nearby lakes in Newfoundland (Kess et al. 2021) and northern Quebec (Power et al. 2009). However, the repeatability of the adaptive genetic differentiation associated with such sympatric morphs within landlocked lakes remains unknown. Our recent work in this region has revealed limited genetic parallelism across size-differentiated sympatric morphs (consistent with putative resident and anadromous morphs) occurring in three sea-accessible Labrador lakes (Salisbury et al. 2020). Whether the genetic mechanisms driving sympatric morph differentiation

in charr are the same in landlocked and sea-accessible lakes remains unexplored. In addition, little genetic parallelism has been observed between ecologically-differentiated sympatric morphs across Scottish and Russian landlocked populations (Jacobs et al. 2020). However, unlike Scottish and Russian landlocked charr populations which were founded by single lineages (Atlantic or Siberian, respectively; Moore et al. 2015), Labrador landlocked populations demonstrate mtDNA haplotypes consistent with Acadian, Atlantic, and Arctic glacial lineages reflecting the secondary contact of all three lineages within this region (Salisbury et al. 2019). Individual landlocked populations in Labrador could have therefore been founded by very different ancestral populations (e.g., different glacial lineages).

This potential for differential ancestry among landlocked populations as well as the genetic drift likely experienced by these isolated lakes may have reduced shared genetic variation among landlocked populations thereby thwarting genetic parallelism between both replicate landlocked and anadromous populations and between replicate sympatric morph differentiation within landlocked lakes. Alternatively, the recent establishment and close geographic proximity of these landlocked populations could have resulted in them sharing genetic variation and selective pressures contributing to genetic parallelism. Therefore, the expected degree of genetic parallelism in response to the loss of anadromy and driving incipient speciation within these Labrador landlocked lakes is unknown.

Such insight into the character and consistency of the adaptive genetic differentiation between allopatric landlocked and anadromous populations and between sympatric morphs within landlocked populations is crucial for both their management and our understanding of the predictability of evolution. We employed a newly-designed 87k SNP array (Nugent et al. 2019) to characterize the adaptive differentiation between paired landlocked and anadromous

populations from the same drainage area and between sympatric morphs within landlocked populations. Our study takes advantage of natural replicates of landlocked populations across five drainages in Labrador, Canada to assess for genetic parallelism at the level of the SNP, gene, and paralog. We hypothesize that 1) consistent selective pressures due to the loss of anadromy has resulted in parallel genetic differentiation of landlocked populations, and 2) consistent divergent selection within landlocked lakes has resulted in parallel genetic differences between sympatric lacustrine charr morphs.

Methods

Sampling

Tissue samples (gill/fin) (N = 342, Table S1) were collected between 2010 and 2017 from landlocked and anadromous Arctic Charr populations from five drainages in Labrador; these were, from north to south: Saglek Fjord, Hebron Fjord, Okak Region, Anaktalik River, and Voisey Bay (Fig.1). Landlocked populations (code ending with "-L") were sampled using variable sized standardized nylon monofilament gillnets (1.27–8.89 cm diagonal) while anadromous populations (code ending with "-A") were electrofished. Landlocked specimens were weighed (g), measured for fork length (mm) and assessed for sex and maturity. All samples were immediately stored in 95% ethanol or RNAlater.

For comparative purposes, in a subset of our analyses we also included N = 178 individuals reported in Salisbury et al. (2020) from three sea-accessible lakes (according to Anderson 1985, but see Van der Velden et al. 2013) named Ramah (R), Brooklyn (B), and Esker North (E), each of which was found to contain sympatric big (putative anadromous) and small (putative resident) morphs (Table S2). These populations were sampled, extracted, and

genotyped identically to the anadromous and landlocked populations that are the focus of this study.

Extraction, Sequencing, Genotyping and Quality Control

DNA was extracted using either a glassmilk protocol (modified from Elphinstone et al. 2003), a Phenol Chloroform protocol (modified from Sambrook and Russell 2006), or a Qiagen DNeasy 96 Blood and Tissue extraction kit (Qiagen). Samples were quantified using QuantIT PicoGreen (Life Technologies) on the LightCycler 480 II (Roche) and normalized using epMotion 5075 liquid handling robot (Eppendorf) to a volume of 40 µL and a median concentration of 12 ng/µL.

DNA samples were sent to the Clinical Genomic Centre of Mount Sinai Hospital (Toronto, Canada) for sequencing using an 87k Affymetrix Axiom Array (Nugent et al. 2019). We employed the "best practices workflow" for a diploid organism in Axiom Analysis Suite (Version 4.0.1.9) (supporting information). After applying Axiom Analysis Suite QC we retained a total of N = 307 individuals (Table S1) for further analyses.

A minor allele frequency (MAF) filter of 0.01 was applied using PLINK (Version 1.9; Chang et al. 2015) when investigating structure within each landlocked population, between each pair of landlocked/anadromous populations, and across all landlocked and anadromous populations. PGDSpider (Version 2.1.1.5)(Lischer and Excoffier 2012) was used to convert between PLINK and Genepop files and the R package (R Core Team 2013) *genepopedit* (Stanley et al. 2017) was used to order and arrange Genepop files for downstream analyses.

Genetic differentiation within landlocked populations

Before we could compare landlocked with anadromous populations it was first necessary to identify any sympatric morphs present within landlocked lakes to prevent genetic sub-structuring within lakes biasing subsequent comparisons between landlocked and anadromous populations. Therefore, landlocked populations were assessed for K = 1-5 with ADMIXTURE (Version 1.3; Alexander et al. 2009) using 10 cross-validations. Saglek Fjord landlocked samples (WP132 and WP133) were analyzed together due to their close proximity and previously noted genetic similarity (Salisbury et al. 2018). Voisey Bay landlocked samples (SLU-L and GB-L) were analyzed together due to their low sample sizes. For landlocked populations where the best (lowest average cross-validation error) K-value >1, individuals were assigned to genetic groups based on ADMIXTURE Q-values. Genetic sub-structuring was confirmed using 1) the R package *PCAdapt* (Version 4.1.0; Luu et al. 2017) testing K = 1-20 (K = 1-10 where the number of samples <20) with the default Mahalanobis distance, and 2) the *snmf* function in the R package *LEA* (Frichot and François 2015) testing K = 1-5 using 10 repetitions. Before assessing the effects of genetic group assignment and maturity on fork length (mm) we conducted Levene tests of the equality of variances. Where the variance of fork length differed among groups, we conducted Welch's ANOVAs and Games-Howell posthoc tests; otherwise we conducted ANOVAs and posthoc pairwise t-tests using a Bonferroni correction (α=0.05).

Overall population structure

To investigate the relationships among all populations and confirm that landlocked populations were most genetically similar to those anadromous populations within the same drainage we performed a *PCAdapt* analyses (testing K=1-30). We included all populations as

well as the sympatric small and big morphs identified in three lakes (Ramah, Brooklyn, Esker North) in Labrador by Salisbury et al. (2020) for a total of N = 485 samples. Weir and Cockerham (1984) F_{ST}s were also estimated between all populations using the package *hierfstat* (Goudet 2005). To assess if regions of high linkage disequilibrium (LD) were potentially influencing the genetic structure of all populations, we used an LD pruning method as suggested by Lotterhos (2019). Specifically, using only those SNPs mapped to one of the 39 charr linkage groups, we applied a MAF filter of 0.01 before using the –indep-pairwise function in PLINK to scan the genome in windows of 50 SNPs, shifting 5 SNPs at a time, with an r^2 threshold of 0.5.

Genetic differentiation between landlocked and anadromous populations

Each sympatric morph within a single landlocked population was subsequently separately compared with the anadromous population within the same drainage. The genetic structure between paired landlocked and anadromous populations was assessed using 1) *PCAdapt* and 2) the *snmf* function in the R package *LEA* using identical parameters to the tests within landlocked lakes as outlined above.

Outlier Detection

Outlier SNPs were detected using two methods. First, PCAdapt using default parameters and a MAF of 0.01 was used to detect outlier SNPs based on their correlation with the first PC axis after p-values were corrected using the False Discovery Rate (FDR; Storey and Tibshirani 2003) with the R package *qvalue* (Version 2.14.1; Storey 2015). Second, using Weir and Cockerham (1984) F_{ST}s estimated from PLINK, SNPs with an F_{ST} > 3 SD above the mean F_{ST} were considered outliers. We compared SNPs to the *Salvelinus* genome (NCBI,

https://www.ncbi.nlm.nih.gov/genome/86400) using BEDOPS (Neph et al. 2012) and identified the closest coding sequence (CDS) within 5000 bp of each SNP. By design, many SNPs on the chip were located within the CDS of genes, however, given the observed linkage disequilibrium decay between paired comparisons (Fig.S1, S2), associating SNPs with the closest CDS within 5000 bp is likely a conservative, yet reasonable approach.

Allelic frequencies of outlier SNPs were visualized using the R package *ComplexHeatMap* (Gu et al. 2016). Allelic trends for outlier SNPs were defined as parallel if the direction of the difference in major allele frequency (positive or negative) was the same across all the paired comparisons for which the SNP was detected as an outlier.

Gene Ontology (GO) enrichment analyses for Biological Processes (BP) were conducted using the R package *TopGO* employing the protein GO annotation file generated for charr (Christensen et al. 2018). GO term significance was assessed with a Fisher's exact test using the "weight01" algorithm, p-values were corrected using FDR ($\alpha = 0.05$).

Assessing Genetic Parallelism at the Level of the SNP, Gene, and Paralog

For landlocked populations with multiple sympatric morphs, the outliers detected by comparing each morph separately with a common downstream anadromous population were pooled over all morph comparisons. This reflects the fact that each sympatric morph does not represent an independent replicate of landlocking given the potential for gene flow between sympatric morphs. We calculated C-scores (Yeaman et al. 2018) using the R package *dgconstraint* to investigate if the amount of shared outlier SNPs detected across replicate pairs of 1) landlocked vs. anadromous populations, and 2) sympatric morph pairs within landlocked lakes was statistically significant. We applied a binary assignment to outliers and non-outliers (1 or 0,

respectively) and employed a hypergeometric approach when comparing two paired comparisons but a chi-squared approach using 100,000 permutations when comparing more than two paired comparisons (see supporting information for more details).

We were also interested in investigating the importance of paralogs to both local adaptation and parallel morph differentiation. Given our use of a SNP chip we were unable to identify paralogs by direct sequence comparison. Instead, we took the same conservative approach to identifying paralogs as employed by Salisbury et al. (2020), identifying outlier SNPs annotated with identical protein names but different protein codes ("XP_") as putative paralogs. While this does not allow for the comprehensive identification of paralogs within our data (e.g., our approach prevents detection of paralogs with non-identical names and does not allow assessment of the timing of duplication) and relies upon the accuracy of the reference genome, it nevertheless provides some insight into whether paralogous copies of the same gene may contribute to local adaptation or incipient divergence. Parallelism at the level of paralog was inferred when different paralogous copies of a given gene differentiated different replicates.

Results

Genetic differentiation within landlocked populations

Population structure analyses revealed evidence of genetic sub-structuring (consistent with sympatric morphs) within only two of the seven landlocked locations sampled in this study. After filtering using a MAF of 0.01 (resulting in 6404 – 16702 SNPs per landlocked lake, Table S3), ADMIXTURE analyses supported within-lake genetic sub-structuring (where the best K > 1) in only WP-L and LO-L. In each of these two locations, K = 2 genetic groups were detected (Fig.2a,d, Table S4). This was confirmed with PCAdapt and *snmf* structure analyses (Fig.S3 –

S6). The two genetic groups detected within WP-L were apparent within each of the two neighboring lakes associated with this location (WP132-L and WP133-L), suggesting recent gene flow between the lakes despite intervening falls (Anderson 1985). Six WP-L individuals had intermediate Q-values (0.4 < Q-value < 0.6) suggesting they were hybrids between the two "purebred" genetic groups. These samples were removed before conducting all outlier detection analyses to more easily detect signatures of divergent selection. Pairwise mean F_{ST} values between these ADMIXTURE-defined groups was 0.12 within WP-L (calculated excluding putative hybrids) and 0.12 within LO-L. There was no evidence of sympatric differentiation within any of the other landlocked lakes (see supporting information, Table S4, Fig.S7, S8).

Size differences between sympatric genetic groups

The genetic groups detected within each of WP-L and LO-L differed by fork length. In WP-L, significant differences in length were detected (Levene test $F_{(5,49)} = 8.76$, p-value < 0.001; one-way Welch's ANOVA, $F_{(5,7.37)} = 22.44$, p-value < 0.001; based on N = 55 individuals after removing 3 individuals with unknown maturity status) and posthoc Games-Howell tests revealed significant differences in length between genetic groups (Fig.2b, Table 1; for differences by sex see Fig.S9). The ADMIXTURE-assigned hybrids were larger than either of the "purebred" big or small morphs ($\bar{x}_{hybrids} = 385$ mm ($N = 6$); $\bar{x}_{purebred\ big} = 329$ mm (N = 24); $\bar{x}_{purebred\ small} = 145$ mm (N = 28); (Fig.2b, Table 1). In LO-L (N = 29) there were also significant differences in length [Levene test $F_{(3,25)} = 4.75$, p-value < 0.01; one-way Welch's ANOVA, $F_{(3,4.57)} = 68.60$, p-value < 0.001], with posthoc Games-Howell tests again suggesting differences between genetic groups (Fig.2e, Table 1; for differences by sex see Fig.S10)). Within both WP-L and LO-L, each genetic group comprised mature and immature individuals of both

sexes (Table 1; see Table S5 for information from other landlocked locations). We refer to each genetic group within each of WP-L and LO-L as "small" and "big" morphs hereafter.

Overall population structure

Having identified two landlocked lakes with sympatric morphs we were interested in assessing the overall genetic structure of landlocked and anadromous populations in our study system as well as the genetic relationships of these big and small morphs with those big and small morphs detected in Ramah, Brooklyn, and Esker North from Salisbury et al. (2020). After applying a MAF of 0.01 across all samples (N = 485), we retained N = 21201 SNPs. PCAdapt results were nearly identical after applying LD pruning; we report only the results using the full set of 21201 SNPs (but see Fig.S11). The first two PCs separated all anadromous populations from all landlocked populations (Fig.3). Landlocked populations demonstrated strong genetic distinctiveness, in contrast to the anadromous populations. Interestingly, big morphs from Ramah grouped with anadromous populations unlike big and small morphs from Brooklyn and Esker North which grouped separately from anadromous populations, like those big and small morphs in each of WP-L and LO-L. Weighted pairwise F_{ST} estimates confirmed these results and generally demonstrated that landlocked populations were most genetically similar to the anadromous population within their drainage (Fig.S12). We next compared landlocked and anadromous populations within drainages to minimize detection of genetic differences potentially due to population structure and local adaptive differences across drainages.

Genetic differentiation between landlocked and anadromous populations

Consistent with our overall PCA, paired landlocked and anadromous populations demonstrated significant genetic differences. Between 17321-22540 SNPs passed filtering for each landlocked versus anadromous population comparison (Table 2, Table S6). Mean F_{ST} values between paired landlocked and anadromous populations from the same drainage ranged from 0.10 – 0.26 (Table 2). Significant genetic distances between paired landlocked and anadromous populations were confirmed by PCAdapt and *snmf* population structure analyses (Fig.S13-S16). As expected given this strong genetic differentiation, a large number of outlier SNPs were detected between paired landlocked and anadromous populations (between 370-2296 SNPs; Fig.4, Table 2, File S1). Outlier SNPs were detected across the genome, in 35 or more linkage groups in all comparisons (Table 2; see Table S7 for SNPs detected by each method and Table S8 for polymorphic SNPs overlapping across comparisons).

Genetic parallelism across landlocked vs. anadromous comparisons

A total of 6357 SNPs were detected as outliers in at least one of the seven landlocked vs. anadromous population comparisons. None were detected in all seven comparisons, one SNP was detected as an outlier in six comparisons, 29 were detected in five comparisons, 76 in four, 356 in three, 1269 in two, and 4626 in only one comparison. This degree of overlap in outlier SNPs across all landlocked vs. anadromous population comparisons was statistically significant ($\text{C-score}_{\text{chisq}} = 26.34$, $\text{p-value}_{\text{chisq}} < 10^{-5}$).

We limit our discussion to only those 30 outlier SNPs that were detected in five or more of the seven landlocked vs. anadromous population comparisons. Of these 30 SNPs, 28 showed evidence of parallel allelic trends (Fig.5) and these 30 SNPs corresponded to 21 genes (Table 3).

An additional two genes contained outlier SNPs in five or more comparisons, but the specific outlier SNP differed between comparisons. Therefore, a total of 23 outlier genes were detected in five or more landlocked vs. anadromous populations. No significant BP GO terms were enriched among these 23 genes after adjusting p-values using FDR (Table S9).

Parallelism at the level of the paralog was detected across replicate landlocked vs. anadromous comparisons. For N = 16 genes at least five of the seven landlocked vs. anadromous comparisons demonstrated outlier SNPs in different paralogous copies of the same gene. However, none were detected in all seven landlocked vs. anadromous comparisons (Table S10, see File S2 for all paralogs detected in multiple landlocked vs. anadromous comparisons).

Genetic parallelism of sympatric divergence within landlocked lakes

A total of N = 108 outlier SNPs across 22 linkage groups were detected between sympatric morphs in WP-L, while N = 400 outlier SNPs across 37 linkage groups were detected between sympatric morphs in LO-L (Fig.2c, f, File S3). The number of SNPs detected by each method is reported by lake (Table S11). While 6183 SNPs were polymorphic in both WP-L and LO-L, only a single outlier SNP (AX-181980220) was detected in common between morphs in both of WP-L and LO-L (Table S12). It was located within the VPS10 domain-containing receptor SorCS2 gene and demonstrated parallel allelic trends (Fig.S17, Table S13). This gene is associated with neural development (Rezgaoui et al. 2001), was also identified as an outlier between sympatric big and small morphs in Esker North (Salisbury et al. 2020) and genetically differentiates fluvial "coaster" and adfluvial, resident Brook Trout (*Salvelinus fontinalis*) (Elias et al. 2018) as well as resident "black" Kokanee Salmon and river-spawning Sockeye Salmon (*Oncorhynchus nerka*) (Veale and Russello 2017). We speculate that this gene may be important

in contributing to incipient speciation in salmonids. However, the overlap of a single outlier SNP between sympatric morphs was not statistically significant ($C\text{-score}_{hyper} = 0.38$, p-value = 0.50) and no paralogous copies of the same gene were found to differentiate sympatric morphs across replicate landlocked lakes.

Discussion

Our results demonstrate a small but consistent degree of genetic parallelism underlying geographically paired landlocked and anadromous populations. Less evidence was found, however, for parallel genetic differences between genetically distinguishable, size-differentiated, sympatric morphs observed in the two landlocked locations. Drift, local adaptation, or differential colonization histories may have contributed to this overall lack of genetic parallelism. However, the functionally relevant genes found to consistently differentiate landlocked vs. anadromous populations suggest the potential for a common genetic basis for the loss of anadromy.

Limited Overall Genetic Parallelism

An overall limited amount of genetic parallelism was found in this study between paired landlocked and anadromous populations. No single locus consistently differentiated all paired comparisons, conflicting with recent observations that migratory life history in Rainbow Trout (*Oncorhynchus mykiss*) (Arostegui et al. 2019; Pearse et al. 2019) and in Japanese Grenadier Anchovy (*Coilia nasus*) (Zong et al. 2020) are dictated by consistent inverted genomic regions. Threespine Stickleback also demonstrate strong parallel genetic differentiation at loci like *Eda* and *Pitx* (Nelson and Cresko 2018; Xie et al. 2019). However, even for these loci of large effect,

parallelism is not always perfect (DeFaveri et al. 2011; Weinstein et al. 2019) and other loci across the genome may show little evidence of parallelism (Hale et al. 2013; Liu et al. 2018). Indeed, despite high genome-wide genetic differentiation between replicate landlocked populations of *O. mykiss* and a downstream anadromous population, few loci demonstrated parallelism across replicate landlocked populations apart from inversions on Omy05 and Omy20 (Campbell et al. 2021).

Similarly, only a single outlier SNP differed between small and big morphs in both landlocked locations, and this overlap was not greater than expected by chance based on C-Score analyses. Interestingly, pappalysin-2, which was detected as an outlier between small and big morphs in all locations studied by Salisbury et al. (2020) and has been associated with growth in mice and humans (Conover et al. 2011; Dauber et al. 2016) was not detected as an outlier in either of WP-L or LO-L. Limited amounts of genetic parallelism have also been found to underlie intralacustrine radiations in other populations of charr (Jacobs et al. 2020) and cichlids (Elmer et al. 2014).

Several mechanisms could be responsible for this overall lack of genetic parallelism underlying morph differentiation. Non-parallel genetic pathways may have been employed to achieve similar adaptive differentiation (Campbell and Bernatchez 2004) potentially facilitated by the abundance of genomic material provided by the whole genome duplication at the base of Salmonidae (Campbell et al. 2021). Alternatively, inconsistent environmental conditions across populations could have resulted in local adaptation and reduced genetic parallelism (Campbell and Bernatchez 2004). Given the colonization of Labrador by three glacial lineages (Salisbury et al. 2019), landlocked populations could have also been founded by different glacial lineages, thereby reducing shared genetic variation and the probability of genetic parallelism. However,

though some landlocked populations differ by mtDNA haplotypes, different haplotypes could have been fixed in each population due to drift if all landlocked populations were founded by a common admixed population. Given the isolation of landlocked populations, drift could have contributed to the non-parallel fixation of not only mtDNA haplotypes but also nDNA. Though alternative sources of non-parallelism require further study, we speculate that drift has primarily contributed to a lack of genetic parallelism in this system.

Genetic Parallelism Across Landlocked vs. Anadromous Comparisons

Despite the isolation of landlocked populations, a few key loci seem to consistently differ between landlocked and anadromous populations. The parallel allelic trends demonstrated by 28 of the 30 outliers SNPs detected in at least five pairs of landlocked lakes and anadromous populations strongly suggests that these loci are responding to consistent directional selection. Our significant C-score analysis also suggests parallelism at the level of the SNP.

Of the outlier gene/proteins detected in five or more comparisons, several have putative functions relevant to the loss of anadromy. Myomesin-2 is associated with cardiac and fast-twitch muscle function (Schoenauer et al. 2008), traits expected to differentiate anadromous and non-anadromous populations (Delgado and Ruzzante 2020). In addition, myomesin-1 is significantly upregulated during Atlantic Salmon smoltification (Seear et al. 2010). The outlier inactive dipeptidyl peptidase 10 is associated with neuronal potassium regulation (Jerng et al. 2004) and genetically differentiates resident and diadromous populations of *Galaxias maculatus* (Delgado et al. 2019).

Interestingly, 7 of the 30 outlier SNPs detected in at least five landlocked vs. anadromous population comparisons were also detected as outliers between sympatric small and big morphs

in either Ramah, Brooklyn, or Esker North lakes (Salisbury et al. 2020) (Fig.S18, Table S14). Five of these SNPs were in a ~240 kb region of AC17 and were detected as outliers between sympatric small and big morphs in Ramah Lake. For all five SNPs, the most prevalent allele in anadromous populations was also so for the Ramah big morph (Fig.S18). This is notable as the big, putative anadromous morph, in Ramah was more genetically similar to the anadromous populations studied here than those big morphs from Brooklyn and Esker North (Fig.3). This region in AC17 (containing the gene inactive dipeptidyl peptidase 10 discussed above) might therefore be key to the loss of anadromy both in sympatry (between resident and anadromous morphs) and in allopatry (between landlocked and anadromous populations).

Multiple genes also demonstrated evidence for parallelism at the paralog level in five or more landlocked vs. anadromous population comparisons. Most putative paralogs were located on distinct mapped linkage groups though some were unmapped and may not represent distinct paralogous copies. The presence of some paralogs on homeologous linkage groups suggests the potential importance of the recent whole genome duplication in salmonids (Macqueen and Johnston 2014) on contemporary adaptive differentiation. Furthermore, several of the genes demonstrating parallelism at the paralog level were associated with functionally relevant functions. One was chloride channel protein 2, which is associated with osmoregulation and is differentially expressed in salinity-tolerant and salinity-sensitive populations of Sacramento Splittail (*Pogonichthys macrolepidotus*) (Jeffries et al. 2019; Mundy et al. 2020). Another, pro-neuregulin-3, membrane-bound isoform is associated with neural development in mice (Zhang et al. 1997; Anton et al. 2004) and genetically differentiates migratory and non-migratory Brown Trout (*Salmo trutta*) (Lemopoulos et al. 2018). Interestingly, paralogous copies of this gene were also found to genetically differentiate sympatric small and big morphs in each of Ramah,

Brooklyn, and Esker North (Salisbury et al. 2020). Our results therefore suggest that parallelism at the level of the paralog could play a key role in morph differentiation in charr and other species. Though we had limited ability to detect paralogs using the SNP data employed in this study our results argue for further investigation of this overlooked level of parallelism.

Genetic Structure of Sympatric Morphs

The ecologies of the small and big morphs in WP-L and LO-L remain ambiguous. Genetically distinguishable morphs within WP-L had previously been described as cryptic based on 11 microsatellites, as no size-difference had been detected (Salisbury et al. 2018). This was however, likely due to a failure to remove putative hybrid individuals prior to size comparisons (see Fig.S19). Size-differentiated charr morphs have been observed within lakes in Alaska (May-Mcnally et al. 2015), Europe (Westgaard et al. 2004) and Canada (Kess et al. 2021). Indeed, the size differences between morphs in LO-L and WP-L is similar to that between genetically distinguishable small, littoral morph and the large benthic morphs within Lake Aigneau in northern Quebec, Canada (Power et al. 2009). Small and large morphs are also present within Charr Lake in Hebron Fjord, Labrador (Bouillon and Dempson 1989). Though we suspect these size-differentiated morphs arose in sympatry within these lakes, we cannot rule out the possibility of recent allopatry with subsequent secondary contact. For example, the putative hybrids identified in WP-L could reflect an earlier stage of incipient speciation prior to complete reproductive isolation or could alternatively be due to a collapse of morph differentiation. Sympatric morphs were not however, founded by different glacial lineages as both morphs from WP-L had Atlantic lineage mtDNA whereas both morphs from LO-L had Arctic lineage mtDNA

(Salisbury et al. 2019; Table S5). Further investigation of these morphs' ecologies as well as the environmental factors driving/maintaining this genetic differentiation is therefore needed.

In addition, of those sea-accessible populations with sympatric big (putative anadromous) and small (putative resident) charr investigated in Salisbury et al. (2020) only the big morphs from Ramah, but not those of Brooklyn or Esker North, were genetically similar to the anadromous populations. Though Anderson (1985) suggests both Esker North and Brooklyn are sea-accessible, given their remoteness there is some uncertainty about this status particularly Esker North which was considered landlocked by Van der Velden et al. (2013). It is possible then that big morphs from Esker North and Brooklyn could be non-anadromous, but as stated in Salisbury et al. (2020) it would be useful to verify the migratory phenotype of these morphs using telemetry and stable isotopes. Regardless, the genetic distinctiveness of Brooklyn and Esker North charr likely contributed to the limited genetic parallelism observed between sympatric small and big morphs across Ramah, Brooklyn, and Esker North (Salisbury et al. 2020).

Conclusions

Despite the isolation of landlocked populations, our results demonstrate that their genetic diversity was sufficient to allow for both incipient speciation as well as their consistent, potentially adaptive genetic differentiation from anadromous populations. While the former result has previously been observed (e.g., Guðbrandsson et al. 2019; Jacobs et al. 2020; Østbye et al. 2020), the latter has rarely been assessed using a replicated pairwise design of natural populations as studied here. Our experimental design allowed us to uncover a number of candidate genes and paralogs demonstrating genetic parallelism across drainages, many of which

were associated with ecologically relevant functions. Furthermore, some of the genes that consistently genetically differentiated landlocked and anadromous populations had also previously been found to differentiate sympatric putative resident and putative anadromous Arctic Charr in other Labrador populations. Some of these genes had also been associated with migratory and non-migratory life histories in other fish species. Our results propose the intriguing possibility that the underpinnings of migration in both Arctic Charr and other fishes may share a genetic commonality.

Acknowledgements

Thanks go to our editor Dr. Lau and associate editor Dr. Lotterhos as well as two anonymous reviewers whose suggestions greatly improved this work. We greatly appreciate S. Avery, J. Callahan, S. Duffy, S. Hann, L. Pike, R. Solomon, A. Walsh, for their indispensable help with fieldwork. We thank Parks Canada for allowing us access to the Torngat Mountains National Park and the Nunatsiavut government for allowing us to collect samples from their lands. We thank A. Belay at Mount Sinai Hospital for her help with sequencing, A. Mesmer for help with genotyping, and S. Lehnert for insightful data analysis suggestions. We also thank the Institute for Biodiversity, Ecosystem Science, and Sustainability of the Department of Environment and Conservation of the Government of Labrador and Newfoundland for funding for this project; NSERC for the Strategic Grant STPGP 430198 and Discovery Grant awarded to DER, for the CGS-D awarded to SJS; the Killam Trust for the Level 2 Izaak awarded to SJS; and the Government of Nova Scotia for the Graduate Scholarship awarded to SJS.

Statement of Authorship

SJS and DER conceptualized the study. SJS, DER, RP, DK, IRB developed and conducted sampling design. SJS and GRM conducted lab work. JSL, CMN, BFK, and MMF contributed to the data collection (sequencing and genotyping). DER, GRM, IRB, KKSL, TK contributed to data analysis, visualization, and interpretation. SJS analyzed data and led the writing under the supervision of DER. All authors contributed to editing the paper.

Data Accessibility Statement

Genepop files of SNP data as well as metadata are available in Dryad Repository: https://doi.org/10.5061/dryad.cz8w9gj42 (Salisbury et al. 2022). Genepop files of SNP data for Ramah, Brooklyn, and Eskar North locations has been previously published in Dryad Repository: https://doi.org/10.5061/dryad.cz8w9gj1f.

Literature Cited

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655–1664.

Anderson, T. 1985. Rivers of Labrador. Canadian Special Publication of Fisheries and Aquatic Sciences 81, Ottawa, Canada.

Andrews, C. W., and E. Lear. 1956. The Biology of Arctic Char (*Salvelinus alpinus* L.) in Northern Labrador. Journal of the Fisheries Board of Canada 13:843–860.

Anton, E. S., H. T. Ghashghaei, J. L. Weber, C. McCann, T. M. Fischer, I. D. Cheung, M. Gassmann, et al. 2004. Receptor tyrosine kinase ErbB4 modulates neuroblast migration and placement in the adult forebrain. Nature Neuroscience 7:1319–1328.

Arostegui, M. C., T. P. Quinn, L. W. Seeb, J. E. Seeb, and G. J. McKinney. 2019. Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative freshwater ecotypes in rainbow trout. Molecular Ecology 28:1412–1427.

Bernatchez, L., J. B. Dempson, and S. Martin. 1998. Microsatellite gene diversity analysis in anadromous arctic char, *Salvelinus alpinus*, from Labrador, Canada. Canadian Journal of Fisheries and Aquatic Sciences 55:1264–1272.

Bernatchez, L., S. Renaut, A. R. Whiteley, N. Derome, J. Jeukens, L. Landry, G. Lu, et al. 2010. On the origin of species: Insights from the ecological genomics of lake whitefish. Philosophical Transactions of the Royal Society B: Biological Sciences 365:1783–1800.

Bernatchez, L., J. G. Rhydderch, and F. W. Kircheis. 2002. Microsatellite Gene Diversity Analysis in Landlocked Arctic Char from Maine. Transactions of the American Fisheries Society 131:1106–1118.

Bouillon, D. R., and J. B. Dempson. 1989. Metazoan parasite infections in landlocked and anadromous Arctic charr (*Salvelinus alpinus* Linnaeus), and their use as indicators of movement to sea in young anadromous charr 67:2478–2485.

Bryson, R. A., W. M. Wendland, J. D. Ives, and J. T. Andrews. 1969. Radiocarbon Isochrones on the Disintegration of the Laurentide Ice Sheet. Arctic and Alpine Research 1:1–13.

Campbell, D., and L. Bernatchez. 2004. Generic Scan Using AFLP Markers as a Means to Assess the Role of Directional Selection in the Divergence of Sympatric Whitefish Ecotypes. Molecular Biology and Evolution 21:945–956.

Campbell, M. A., E. C. Anderson, J. C. Garza, and D. E. Pearse. 2021. Polygenic basis and the role of genome duplication in adaptation to similar selective environments. Journal of Heredity. esab049.

Chang, C. C., C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience 4:1–16.

Christensen, K. A., E. B. Rondeau, D. R. Minkley, J. S. Leong, C. M. Nugent, R. G. Danzmann, M. M. Ferguson, et al. 2018. The arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. PLoS ONE 13:1–30.

Conover, C. A., H. B. Boldt, L. K. Bale, K. B. Clifton, J. A. Grell, J. R. Mader, E. J. Mason, et al. 2011. Pregnancy-associated plasma protein-A2 (PAPP-A2): tissue expression and biological consequences of gene knockout in mice. Endocrinology 152:2837–2844.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B 279:5039–5047.

Dauber, A., M. T. Muñoz-Calvo, V. Barrios, H. M. Domené, S. Kloverpris, C. Serra-Juhé, V. Desikan, et al. 2016. Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability. EMBO molecular medicine 8:363–374.

DeFaveri, J., T. Shikano, Y. Shimada, A. Goto, and J. Merilä. 2011. Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). Evolution: International Journal of Organic Evolution 65:1800–1807.

Delgado, M. L., A. Manosalva, M. A. Urbina, E. Habit, O. Link, and D. E. Ruzzante. 2020. Genomic basis of the loss of diadromy in *Galaxias maculatus*: insights from reciprocal transplant experiments. Molecular Ecology 29:4857–4870.

Delgado, M. L., K. Górski, E. Habit, and D. E. Ruzzante. 2019. The effects of diadromy and its loss on genomic divergence: The case of amphidromous *Galaxias maculatus* populations. Molecular Ecology 28:5217–5231.

Delgado, M. L., and D. E. Ruzzante. 2020. Investigating diadromy in fishes and its loss in an-omics era. Iscience 101837. https://doi.org/10.1016/j.isci.2020.101837.

Dempson, B. J., M. Shears, G. Furey, and M. Bloom. 2008. Resilience and stability of north Labrador Arctic charr, *Salvelinus alpinus*, subject to exploitation and environmental variability. Environmental Biology of Fishes 83:57–67.

Elias, A., R. McLaughlin, R. Mackereth, C. Wilson, and K. M. Nichols. 2018. Population structure and genomic variation of ecological life history diversity in wild-caught Lake Superior brook trout, *Salvelinus fontinalis*. Journal of Great Lakes Research 44:1373–1382.

Elmer, K. R., S. Fan, H. Kusche, M. Luise Spreitzer, A. F. Kautt, P. Franchini, and A. Meyer. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nature Communications 5:1–8.

Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: Insights from parallelism and convergence. Trends in Ecology and Evolution 26:298–306.

Elphinstone, M. S., G. N. Hinten, M. J. Anderson, and C. J. Nock. 2003. An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. Molecular Ecology Notes 3:317–320.

Frichot, E., and O. François. 2015. LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution 6:925–929.

Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes 5:184–186.

Gu, Z., R. Eils, and M. Schlesner. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849.

Guðbrandsson, J., K. H. Kapralova, S. R. Franzdóttir, Þ. M. Bergsveinsdóttir, V. Hafstað, Z. O. Jónsson, S. S. Snorrason, and A. Pálsson. 2019. Extensive genetic differentiation between recently evolved sympatric Arctic charr morphs. Ecology and evolution 9: 10964-10983.

Hale, M. C., F. P. Thrower, E. A. Berntson, M. R. Miller, and K. M. Nichols. 2013. Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, oncorhynchus mykiss. G3: Genes, Genomes, Genetics 3:1273–1285.

Härer, A., D. I. Bolnick, and D. J. Rennison. 2021. The genomic signature of ecological divergence along the benthic-limnetic axis in allopatric and sympatric threespine stickleback. Molecular Ecology 30:451–463.

Hendry, A. P., T. Bohlin, B. Jonnson, and O. K. Berg. 2004. To sea or not to sea? Anadromy versus non-anadromy in salmonids. Pages 92-125 *in* A. P. Hendry, S. B. Stearns, ed. Evolution illuminated: salmon and their relatives. Oxford University Press, Oxford.

Hindar, K., B. Jonsson, N. Ryman, and G. Stâhl. 1991. Genetic relationships among landlocked, resident, and anadromous brown trout, Salmo Trutta L. Heredity 66:83–91.

Jacobs, A., M. Carruthers, A. Yurchenko, N. V. Gordeeva, S. S. Alekseyev, O. Hooker, J. S. Leong, et al. 2020. Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. PLoS Genetics 16:1–34.

Jacobs, A., and K. R. Elmer. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. Molecular Ecology. 30:4955-4969.

Jeffries, K. M., R. E. Connon, C. E. Verhille, T. F. Dabruzzi, M. T. Britton, B. P. Durbin-Johnson, and N. A. Fangue. 2019. Divergent transcriptomic signatures in response to salinity exposure in two populations of an estuarine fish. Evolutionary Applications 12:1212–1226.

Jerng, H. H., Y. Qian, and P. J. Pfaffinger. 2004. Modulation of Kv4.2 channel expression and gating by dipeptidyl peptidase 10 (DPP10). Biophysical Journal 87:2380–2396.

Kapralova, K. H., M. B. Morrissey, B. K. Kristjánsson, G. Á. Lafsdóttir, S. S. Snorrason, and M. M. Ferguson. 2011. Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. Heredity 106:472–487.

Kess, T., J. B. Dempson, S. J. Lehnert, K. K. S. Layton, A. Einfeldt, P. Bentzen, S. J. Salisbury, et al. 2021. Genomic basis of deep-water adaptation in Arctic Charr (*Salvelinus alpinus*) morphs. Molecular Ecology, 30:4415– 4432.

Kjærner-Semb, E., R. B. Edvardsen, F. Ayllon, P. Vogelsang, T. Furmanek, C. J. Rubin, A. E. Veselov, et al. 2020. Comparison of anadromous and landlocked Atlantic salmon genomes reveals signatures of parallel and relaxed selection across the Northern Hemisphere. Evolutionary Applications 1–16.

Layton, K. K. S., B. Dempson, P. V. R. Snelgrove, S. J. Duffy, A. M. Messmer, I. G. Paterson, N. W. Jeffery, et al. 2020. Resolving fine-scale population structure and fishery exploitation using sequenced microsatellites in a northern fish. Evolutionary Applications 13:1055–1068.

Layton, K.K.S., Snelgrove, P.V.R., Dempson, J.B., Kess, T., Lehnert, S.J., Bentzen, P., Duffy, S.J., et al. 2021. Genomic evidence of past and future climate-linked loss in a migratory Arctic fish. Nature Climate Change 1758-6798.

Lee, C. E., and M. A. Bell. 1999. Causes and consequences of recent freshwater invasions by saltwater animals. Trends in Ecology and Evolution 14:284–288.

Lemopoulos, A., S. Uusi-Heikkilä, A. Huusko, A. Vasemägi, and A. Vainikka. 2018. Comparison of migratory and resident populations of brown trout reveals candidate genes for migration tendency. Genome Biology and Evolution 10:1493–1503.

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

Liu, S., A.-L. Ferchaud, P. Grønkjær, R. Nygaard, and M. M. Hansen. 2018. Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. Molecular ecology 27:4725–4743.

Lotterhos, K. E. 2019. The effect of neutral recombination variation on genome scans for selection. G3: Genes, Genomes, Genetics 9:1851–1867.

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

Macqueen, D. J., and I. A. Johnston. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proceedings of the Royal Society B 281:20132881.

May-Mcnally, S. L., T. P. Quinn, P. J. Woods, and E. B. Taylor. 2015. Evidence for genetic distinction among sympatric ecotypes of Arctic char (*Salvelinus alpinus*) in south-western Alaskan lakes. Ecology of Freshwater Fish 24:562–574.

McCracken, G. R., R. Perry, D. Keefe, and D. E. Ruzzante. 2013. Hierarchical population structure and genetic diversity of lake trout (*Salvelinus namaycush*) in a dendritic system in Northern Labrador. Freshwater Biology 58:1903–1917.

McGee, M. D., S. R. Borstein, J. I. Meier, D. A. Marques, S. Mwaiko, A. Taabu, M. A. Kishe, et al. 2020. The ecological and genomic basis of explosive adaptive radiation. Nature 586:75–79.

Meier, J. I., D. A. Marques, C. E. Wagner, L. Excoffier, and O. Seehausen. 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. Molecular Biology and Evolution 35:1489–1506.

Moore, J. S., R. Bajno, J. D. Reist, and E. B. Taylor. 2015. Post-glacial recolonization of the North American Arctic by Arctic char (*Salvelinus alpinus*): genetic evidence of multiple northern refugia and hybridization between glacial lineages. Journal of Biogeography 42:2089–2100.

Mundy, P. C., K. M. Jeffries, N. A. Fangue, and R. E. Connon. 2020. Differential regulation of select osmoregulatory genes and Na+/K+-ATPase paralogs may contribute to population differences in salinity tolerance in a semi-anadromous fish. Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology 240:110584.

Nelson, T. C., and W. A. Cresko. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evolution Letters 2:9–21.

Neph, S., M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, et al. 2012. BEDOPS: High-performance genomic feature operations. Bioinformatics 28:1919–1920.

Nichols, K. M., A. F. Edo, P. A. Wheeler, and G. H. Thorgaard. 2008. The genetic basis of smoltification-related traits in *Oncorhynchus mykiss*. Genetics 179:1559–1575.

Nugent, C. M., J. S. Leong, K. A. Christensen, E. B. Rondeau, M. K. Brachmann, A. A. Easton, C. L. Ouellet-Fagg, et al. 2019. Design and characterization of an 87k SNP genotyping array for Arctic charr (*Salvelinus alpinus*). PLoS ONE 14:1–17.

Occhietti, S., M. Parent, P. Lajeunesse, F. Robert, and É. Govare. 2011. Late Pleistocene-Early Holocene Decay of the Laurentide Ice Sheet in Québec-Labrador. Pages 601-630 *in* J. Ehlers, P. L. Gibbard, and P. D. Hughes, ed. Developments in Quaternary Science. Vol. 15. Elsevier.

Østbye, K., M. Hagen Hassve, A. M. Peris Tamayo, M. Hagenlund, T. Vogler, and K. Præbel. 2020. "And if you gaze long into an abyss, the abyss gazes also into thee": four morphs of Arctic charr adapting to a depth gradient in Lake Tinnsjøen. Evolutionary Applications 13:1240–1261.

Palkovacs, E. P., K. B. Dion, D. M. Post, and A. Caccone. 2008. Independent evolutionary origins of landlocked alewife populations and rapid parallel evolution of phenotypic traits. Molecular Ecology 17:582–597.

Pearse, D. E., N. J. Barson, T. Nome, G. Gao, M. A. Campbell, A. Abadía-Cardoso, E. C. Anderson, et al. 2019. Sex-dependent dominance maintains migration supergene in rainbow trout. Nature Ecology and Evolution 3:1731–1742.

Power, M., G. Power, J. D. Reist, and R. Bajno. 2009. Ecological and genetic differentiation among the Arctic charr of lake Aigneau, northern Québec. Ecology of Freshwater Fish 18:445–460.

Ramstad, K. M., C. A. Woody, G. K. Sage, and F. W. Allendorf. 2004. Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. Molecular Ecology 13:277–290.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rezgaoui, M., G. Hermey, I. B. Riedel, W. Hampe, H. C. Schaller, and I. Hermans-Borgmeyer. 2001. Identification of SorCS2, a novel member of the VPS10 domain containing receptor family, prominently expressed in the developing mouse brain. Mechanisms of Development 100:335–338.

Salisbury, S. J., C. Booker, G. R. McCracken, T. Knight, D. Keefe, R. Perry, and D. E. Ruzzante. 2018. Genetic divergence among and within arctic char (*Salvelinus alpinus*) populations inhabiting landlocked and sea-accessible sites in labrador, Canada. Canadian Journal of Fisheries and Aquatic Sciences 75:1256–1269.

Salisbury, S. J., G. R. McCracken, D. Keefe, R. Perry, and D. E. Ruzzante. 2019. Extensive secondary contact among three glacial lineages of Arctic Char (*Salvelinus alpinus*) in Labrador and Newfoundland. Ecology and Evolution 9:2031–2045.

Salisbury, S. J., G. R. McCracken, R. Perry, D. Keefe, K. K. S. Layton, T. Kess, C. M. Nugent, et al. 2020. Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). Molecular Ecology 29:4280–4294.

Salisbury S. J. and D. E. Ruzzante. 2022. Genetic Causes and Consequences of Sympatric Morph Divergence in Salmonidae: A Search for Mechanisms. Annual Reviews in Animal Biosciences 10: https://doi.org/10.1146/annurev-animal-051021-080709.

Salisbury, S. J., G. R. McCracken, R. Perry, D. Keefe, K. K. S. Layton, T. Kess, C. M. Nugent, et al. 2022. Data from: The Genomic Consistency of the Loss of Anadromy in an Arctic Fish (*Salvelinus alpinus*). American Naturalist, Dryad Digital Repository, https://doi.org/10.5061/dryad.cz8w9gj1f.

Sambrook, J., and D. W. Russell. 2006. Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harbor Protocols 1:pdb-prot4455.

Schluter, D. 1996. Ecological speciation in postglacial fishes. Philosophical Transactions of the Royal Society B: Biological Sciences 351:807–814.

Schluter, D., and L. M. Nagel. 1995. Parallel Speciation by Natural Selection. The American Naturalist 146:292–301.

Schoenauer, R., S. Lange, A. Hirschy, E. Ehler, J. C. Perriard, and I. Agarkova. 2008. Myomesin 3, a Novel Structural Component of the M-band in Striated Muscle. Journal of Molecular Biology 376:338–351.

Schultz, E. T., and S. D. McCormick. 2012. Euryhalinity in An Evolutionary Context. Fish Physiology 32:477–533.

Scott, W. B., and E. J. Crossman. 1973. Freshwater Fishes of Canada; Bulletin 184.

Seear, P. J., S. N. Carmichael, R. Talbot, J. B. Taggart, J. E. Bron, and G. E. Sweeney. 2010. Differential gene expression during smoltification of Atlantic salmon (*Salmo salar* L.): a first large-scale microarray study. Marine Biotechnology 12:126–140.

Stanley, R. R. E., N. W. Jeffery, B. F. Wringe, C. DiBacco, and I. R. Bradbury. 2017. genepopedit: a simple and flexible tool for manipulating multilocus molecular data in R. Molecular Ecology Resources 17:12–18.

Storey, J.D., A. J. Bass, A. Dabney, and D. Robinson. 2015. qvalue: Q-value estimation for false discovery rate control. R package version 2.10.0. http://github.com/jdstorey/qvalue

Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the USA 100:9440–9445.

Taylor, E. B., and J. D. McPhail. 1999. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. Biological Journal of the Linnean Society 66:271–291.

Van der Velden, S., M. S. Evans, J. B. Dempson, D. C. G. Muir, and M. Power. 2013. Comparative analysis of total mercury concentrations in anadromous and non-anadromous Arctic charr (*Salvelinus alpinus*) from eastern Canada. Science of the total environment 447: 438-449.

Veale, A. J., and M. A. Russello. 2017. Genomic changes associated with reproductive and migratory ecotypes in sockeye salmon (*Oncorhynchus nerka*). Genome Biology and Evolution 9:2921–2939.

Velotta, J. P., S. D. McCormick, A. W. Jones, and E. T. Schultz. 2018. Reduced swimming performance repeatedly evolves on loss of migration in landlocked populations of alewife. Physiological and Biochemical Zoology 91:814–825.

Velotta, J. P., S. D. McCormick, R. J. O'Neill, and E. T. Schultz. 2014. Relaxed selection causes microevolution of seawater osmoregulation and gene expression in landlocked Alewives. Oecologia 175:1081–1092.

Waters, J. M., B. C. Emerson, P. Arribas, and G. A. McCulloch. 2020. Dispersal Reduction: Causes, Genomic Mechanisms, and Evolutionary Consequences. Trends in Ecology and Evolution 35:512–522.

Weinstein, S. Y., F. P. Thrower, K. M. Nichols, and M. C. Hale. 2019. A large-scale chromosomal inversion is not associated with life history development in rainbow trout from Southeast Alaska. PLoS ONE 14:e0223018.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F- statistics for the analysis of population structure. Evolution 38: 1358-1370.

Westgaard, J. I., A. Klemetsen, and R. Knudsen. 2004. Genetic differences between two sympatric morphs of Arctic charr confirmed by microsatellite DNA. Journal of Fish Biology 65:1185–1191.

Xie, K. T., G. Wang, A. C. Thompson, J. I. Wucherpfennig, T. E. Reimchen, A. D. C. Maccoll, D. Schluter, et al. 2019. DNA fragility in the parallel evolution ofpelvic reduction in stickleback fish. Science 363:81–84.

Yeaman, S., A. C. Gerstein, K. A. Hodgins, and M. C. Whitlock. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. PLoS genetics 14:e1007717.

Zhang, D., M. X. Sliwkowski, M. Mark, G. Frantz, R. Akita, Y. Sun, K. Hillan, et al. 1997. Neuregulin-3 (NRG3): A novel neural tissue-enriched protein that binds and activates ErbB4. Proceedings of the National Academy of Sciences of the USA 94:9562–9567.

Zong, S., Y. Li, and J. Liu. 2020. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. Molecular Biology and Evolution msaa290.

Tables

Table 1 Number of small and big morph Arctic Charr samples detected within landlocked locations WP-L and LO-L. Arctic and Atlantic lineage haplotypes from Salisbury et al. (2019). Note that the sum of all immature/mature males/females may not equal N, as some samples had unknown maturity status.

Location	Morph	N	Mean (Median) Length (mm)	Immature Males	Immature Females	Mature Males	Mature Females	Atlantic Lineage	Arctic Lineage
WP-L	small	28	145 (136)	6	4	7	9	15	0
	big	24	329 (352)	8	8	3	4	14	0
	hybrid	6	385 (372)	1	1	3	1	4	0
LO-L	small	21	145 (125)	2	4	8	7	0	16
	big	8	328 (303)	3	3	1	1	0	7

Table 2 All landlocked populations and anadromous population comparisons. Note that in those landlocked populations with multiple sympatric morphs, each was compared independently to the downstream anadromous population to avoid within-lake population structure biasing outlier detection. Outliers detected using each paired comparison with a sympatric morph were then pooled for a single landlocked lake as each sympatric morph does not represent an independent replicate of landlocking given the potential for gene flow between sympatric morphs since landlocking.

Landlocked vs. Anadromous Population Comparison	Drainage	Landlocked Population Compared	Anadromous Population Compared	SNPs	Mean Pairwise F_{ST}	Outlier SNPs	Pooled Outlier SNPs over all sympatric morph comparisons	Linkage groups with outlier SNPs
1	Saglek	small morph WP-L (N = 28)	SWA-A (N = 30)	20393	0.242348	1616	1994	39
	Saglek	big morph WP-L (N = 24)	SWA-A (N = 30)	20361	0.256677	1366		39
2	Hebron	HEB-L (N = 30)	IKA-A (N = 25)	19613	0.191969	548		39
3	Okak	BS-L (N = 29)	K05-A (N = 29)	20334	0.12467	370		35
4	Okak	small morph LO-L (N = 21)	K05-A (N = 29)	22540	0.100295	465	2454	38
	Okak	big morph LO-L (N = 8)	K05-A (N = 29)	22185	0.130439	2250		39
5	Anaktalik	KNU-L (N = 16)	ANA-A (N = 30)	19994	0.187325	2296		39
6	Voisey	SLU-L (N = 10)	REI-A (N = 9)	17385	0.155127	546		37
7	Voisey	GB-L (N = 12)	REI-A (N = 9)	17321	0.17172	479		38

Table 3 Genes containing one or more outlier loci in five or more of seven landlocked vs. anadromous population comparisons (1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs. K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A). The method by which each SNP was identified as an outlier is denoted for each landlocked vs. anadromous population comparison (P – PCAdapt, F- F_{ST}).

						Method of Outlier Detection by Landlocked vs. Anadromous Comparison								
Protein Name	**Linkage Group**	**Protein Code**	**SNP Code**	**Position (Mbp)**	**Position Relative to CDS (kbp)**	**small morph WP -L vs. SWA-A**	**big morph WP-L vs. SWA-A**	**HEB-L vs. IKA-A**	**BS-L vs. K05-A**	**small morph LO-L vs. K05-A**	**big morph LO-L vs. K05-A**	**KNU-L vs. ANA-A**	**SLU-L vs. REI-A**	**GB-L vs. REI-A**
1-acyl-sn-glycerol-3-phosphate acyltransferase gamma	AC02	XP_023860472.1	AX-181934436	17.1	0		P	P			P,F	P	P,F	P,F
neurexin-3a-like	AC04q.2	XP_023842075.1	AX-181947935	28.8	0								P,F	
			AX-181915470	28.9	0	P		P		F	P	P		
			AX-181937420	29.2	0	P	P							
LOW QUALITY PROTEIN: protocadherin-11 X-linked-like	AC08	XP_023847824.1	AX-181937960	0.8	0	P	P				P	P	P	P,F
extended synaptotagmin-1	AC11	XP_023852472.2	AX-181939957	30.7	0			P	P,F	P,F	P,F	P	P,F	
E3 ubiquitin-protein ligase DTX3L isoform X1	AC11	XP_023852474.1	AX-181933794	30.7	0			P	P,F	P,F	P,F	P	P,F	
			AX-181933793	30.7	0			P	P,F	P,F	P,F	P	P,F	
			AX-181941389	30.7	-0.9			P	P,F	P,F	P,F	P	P,F	
EEF1A lysine methyltransferase 3 isoform X2	AC11	XP_023852598.1	AX-181922045	32.8	0.6		P		F		P,F	P	P,F	
uncharacterized protein LOC111970338	AC11	XP_023852785.1	AX-181945827	38.6	-4.5	P			F		P	P	P	
partner of Y14 and mago B	AC17	XP_023861958.1	AX-181916308	22.6	0.1				P,F	P,F	P,F	P	P,F	P,F
			AX-181916309	22.6	0				P,F	P,F	P,F	P	P,F	P,F
PAN2-PAN3 deadenylation complex catalytic subunit PAN2 isoform X1	AC17	XP_023861482.1	AX-181952107	22.7	-2.6				P,F	P,F	P,F	P,F	P	P,F
nuclear envelope integral membrane protein 1-like isoform X1	AC17	XP_023862374.1	AX-181967022	22.7	0				P,F	P,F	P,F	P	P,F	P,F

LOW QUALITY PROTEIN: serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C-like	AC17	XP_023862177.1	AX-181980622	22.9	2.7				P,F	F	P,F	P	P,F	P,F
			AX-181983398	22.9	2.7				P,F	F	P,F	P	P,F	P,F
inactive dipeptidyl peptidase 10	AC17	XP_023860785.1	AX-182162437	22.9	0				P,F	P,F	P,F	P	P,F	P,F
			AX-181987181	23.0	0				P,F	P,F	P,F	P		P,F
E3 ubiquitin-protein ligase BRE1B isoform X2	AC18	XP_023862569.1	AX-181935230	13.0	0			P,F	P,F	P,F		P		P
			AX-181935231	13.0	0			P,F	P,F	P,F		P		
parafibromin	AC19	XP_023865178.1	AX-181969955	34.1	0	P	P	P,F	P	F		P		
calpain-9	AC21	XP_023869744.1	AX-181973095	1.4	0	P	P	P,F	P,F	P,F	P	P		
uncharacterized protein LOC111982472	AC21	XP_023869810.1	AX-181936535	1.5	0	P	P	P,F	P,F	P,F	P	P		
myomesin-2	AC21	XP_024003594.1	AX-181973093	1.5	0	P	P	P		P,F	P,F	P		P
			AX-181936531	1.6	-1.8	P	P	P		P,F	P,F	P		P
LOW QUALITY PROTEIN: lengsin	AC21	XP_023869550.1	AX-181936530	1.6	0.8	P	P	P		P,F	P,F	P		P
peroxisome assembly protein 12	AC23	XP_023824703.1	AX-181933438	19.3	0	P	P				P	P	P,F	P,F
exonuclease V-like isoform X2	AC23	XP_023823204.1	AX-181933434	19.7	0	P	P				P	P	P,F	P,F
collagen alpha-1(XXVI) chain-like	AC23	XP_023823891.1	AX-182165632	20.0	-3.2	P	P				P	P	P,F	P,F
ubl carboxyl-terminal hydrolase 18-like	AC24	XP_023825191.1	AX-181937172	9.6	-0.3	P	P			P,F	P	P	P,F	P
			AX-181944480	9.6	-1.9	P	P			P,F				
DET1- and DDB1-associated protein 1	AC32	XP_023833368.1	AX-181924007	10.3	-1.2		P				P	P		
			AX-181924006	10.3	-1.2			P						P

Figure Legends

Fig.1 Sampling locations in Labrador, Canada in five drainages: Saglek Fjord, Hebron Fjord, Okak Region, Anaktalik River, Voisey Bay. Within each drainage, orange squares indicate landlocked Arctic Charr (Salvelinus alpinus) populations, purple circles indicate anadromous Arctic Charr populations. Landlocked and anadromous population codes are coloured by drainage. Green triangles indicate additional putatively sea-accessible lakes identified as containing sympatric small and large morphs by Salisbury et al. (2020) (R-Ramah, B-Brooklyn, E-Esker North). Map generated using data from CanVec (Government of Canada).

Fig.2 ADMIXTURE plots of K = 2 for a) WP-L and d) LO-L. Orange bars indicate small morph individuals, blue bars indicate big morph individuals. Boxplots demonstrating length of fish by maturity (immature (I), mature (M)) and morph type (small (S), big (B), hybrid (H)) in b) WP-L (N = 55) and e) LO-L (N = 29). Asterisks indicate significant posthoc Games-Howell tests (* = $p < 0.05$, ** = $p < 0.01$, *** $p < 0.001$). Manhattan plots demonstrating FST values of outlier loci detected between small and big morphs (excluding hybrids) in c) WP-L and f) LO-L. Red lines indicate 3 standard deviations above the mean FST and detected outliers are highlighted.

Fig.3 PCA based on N = 21201 SNPs of landlocked (-L; denoted by squares) and anadromous (-A; denoted by circles) populations as well as small and big morphs from three putatively sea-accessible lakes (Ramah-R, Brooklyn-B, Esker North-E; denoted by triangles) detected by Salisbury et al. (2020). Note that small and big morphs were also detected in landlocked populations WP-L, and LO-L (see Fig.2).

Fig.4 Manhattan plots demonstrating FST values of outlier loci detected between landlocked vs. anadromous populations. Red lines indicate 3 standard deviations above the mean FST. Note that mean pairwise FST was calculated separately for each of the two genetic subgroups (corresponding to small and big morphs) within each of LO and WP. However, outlier SNPs detected between either morph and the corresponding anadromous population were pooled when identifying SNPs detected in multiple landlocked vs. anadromous comparisons. Therefore, we identified outlier SNPs detected in 2-4 (black points) and in five or more (red points) of seven landlocked vs. anadromous populations: 1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs. K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A

Fig.5 Heatmap of allele frequencies for those loci detected as outliers for five or more of seven paired landlocked and anadromous populations: 1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs. K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A. The names of SNPs which show parallel allelic trends across the locations in which a SNP was detected as an outlier are highlighted in red.

Figure 1

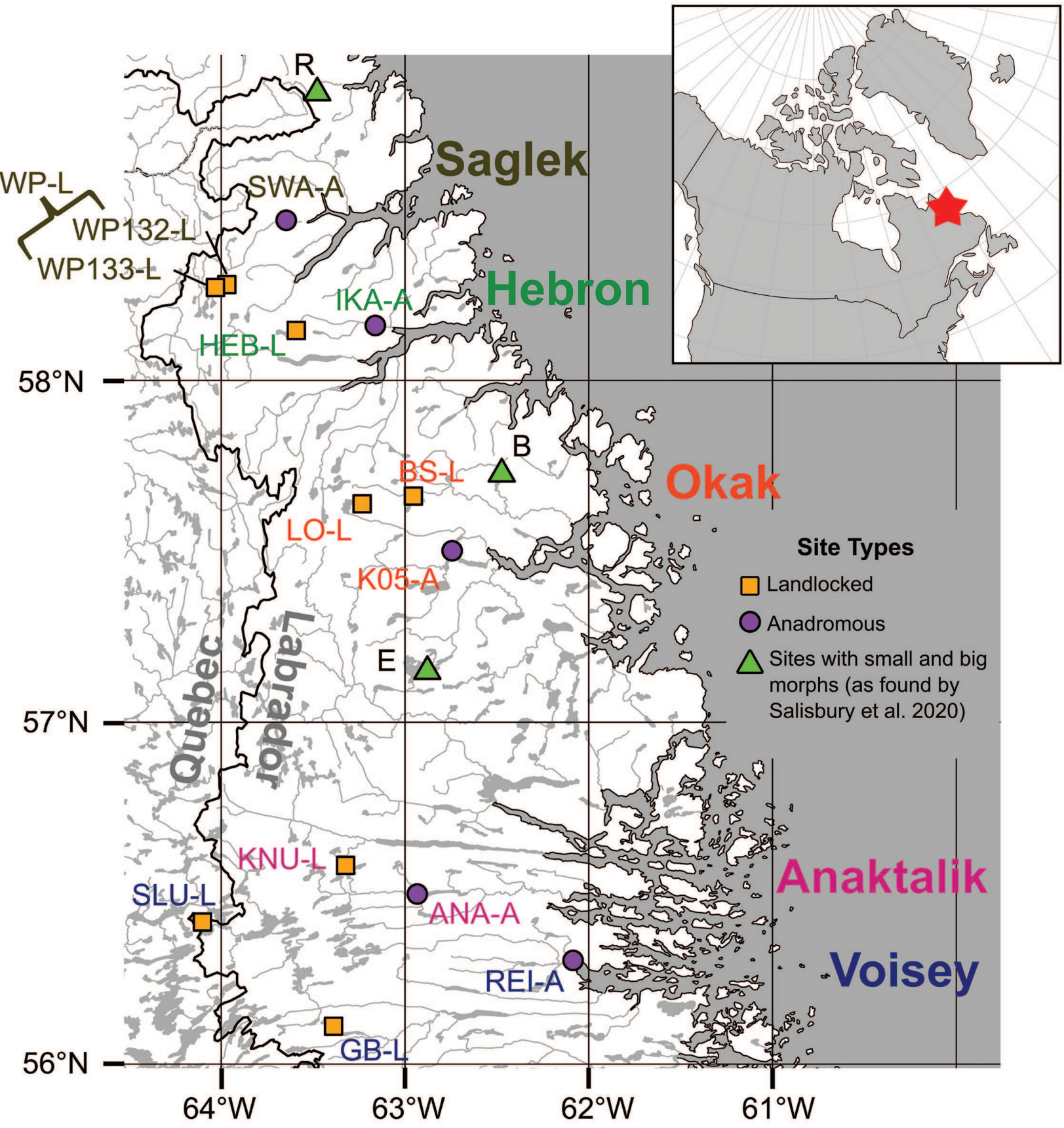


Figure 2

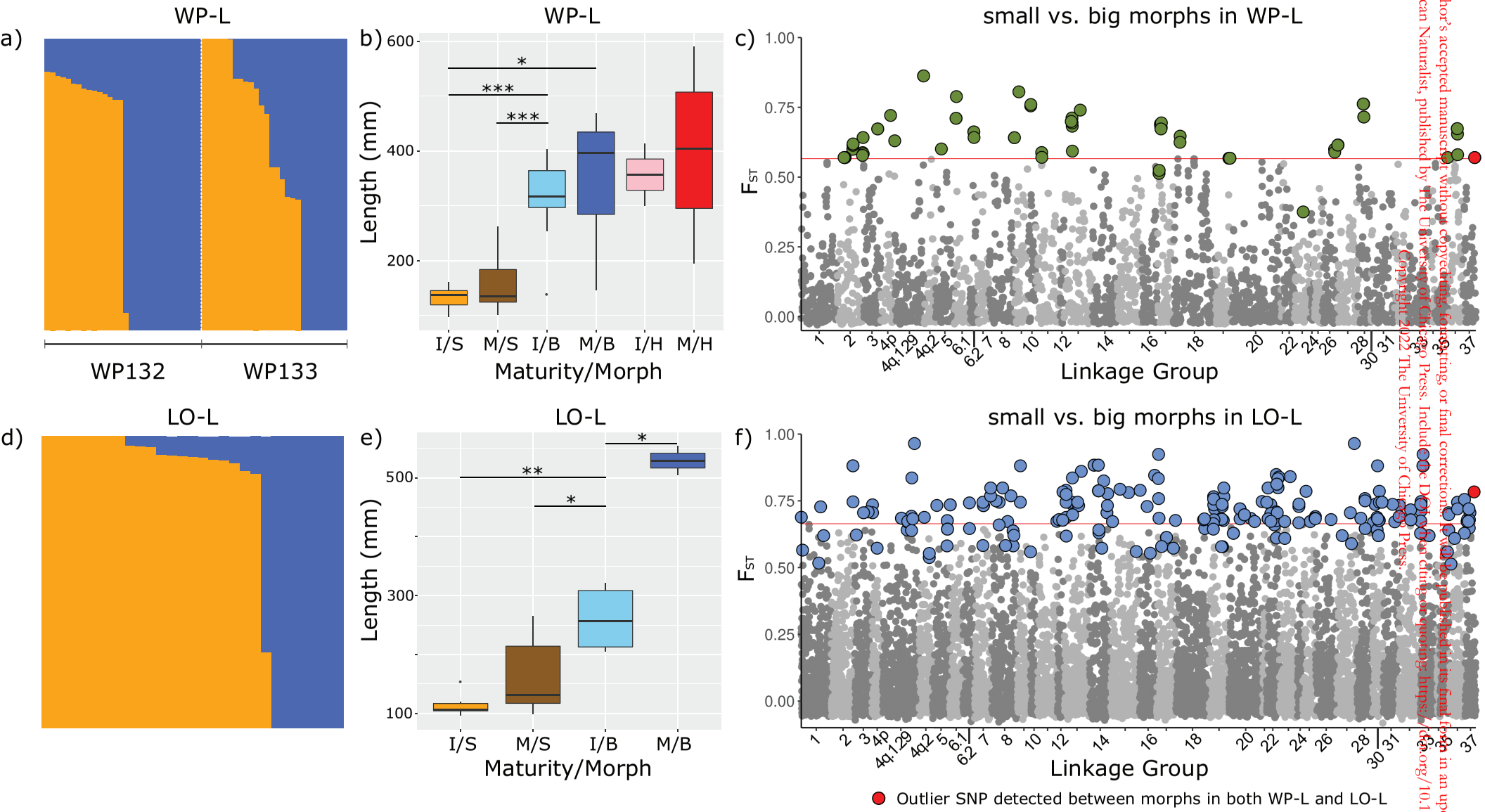


Figure 3

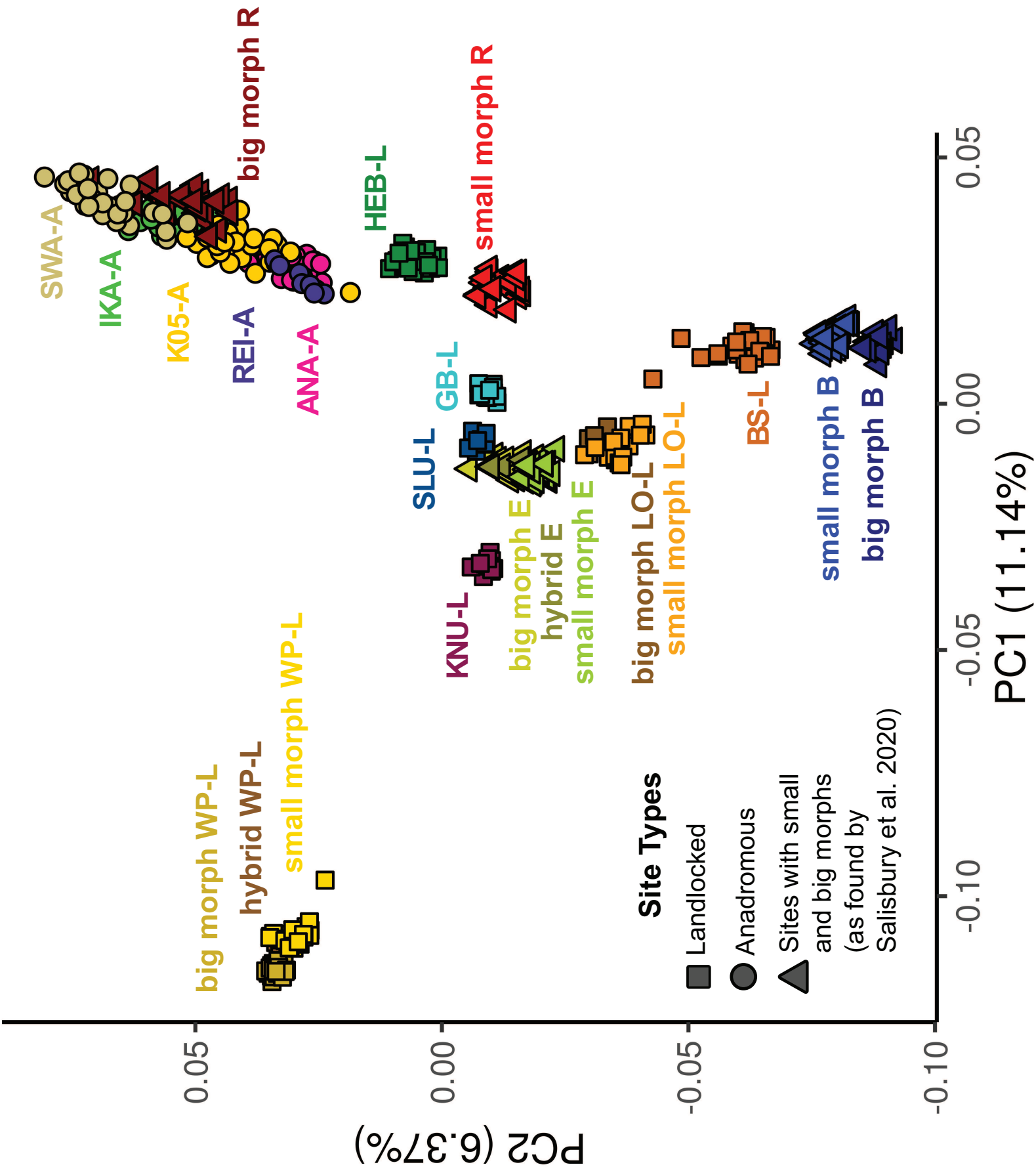


Figure 4

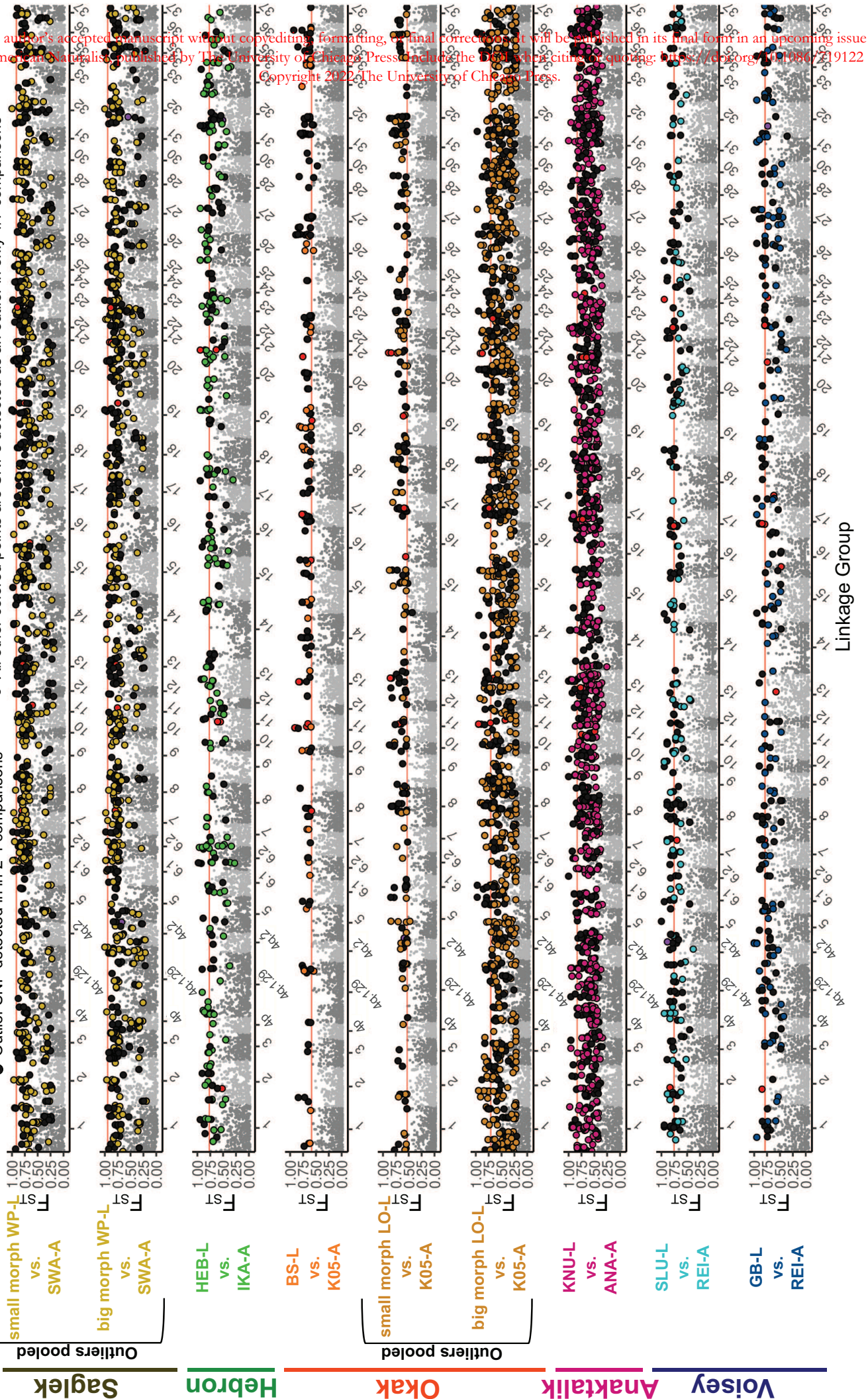
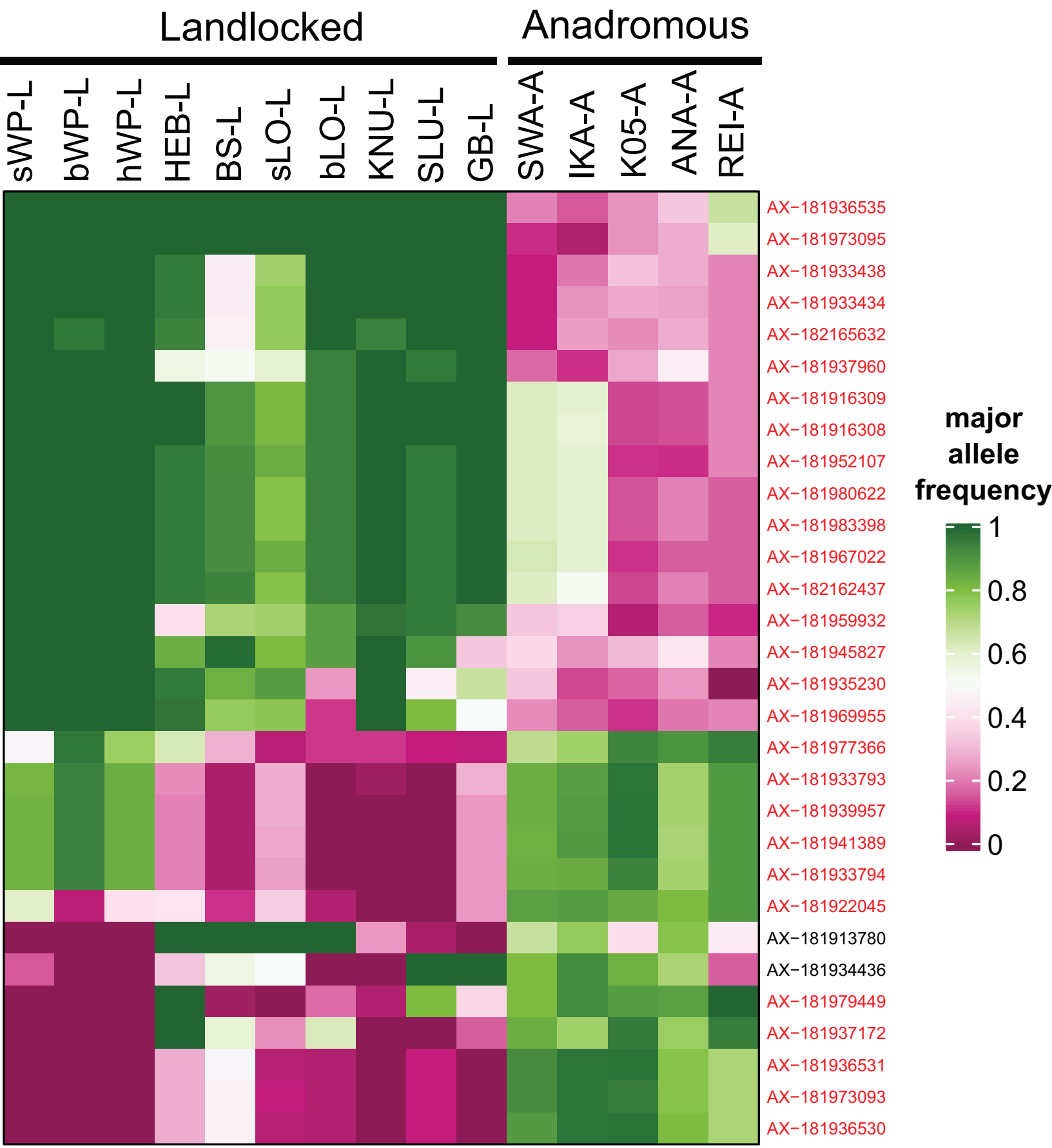


Figure 5



Supporting Information

Title: The Genomic Consistency of the Loss of Anadromy in an Arctic Fish (*Salvelinus alpinus*)

Authors: Sarah J Salisbury[1*], Gregory R McCracken[1], Robert Perry[2], Donald Keefe[3], Kara KS Layton[4,5], Tony Kess[4], Cameron M Nugent[6], Jong S Leong[7], Ian R Bradbury[1,4,5], Ben F Koop[7,8], Moira M Ferguson[6], Daniel E Ruzzante[1]

1 Department of Biology, Dalhousie University, Halifax, NS, Canada

2 Department of Environment, Fish and Wildlife Division, Government of Yukon, Whitehorse, YT, Canada

3 Department of Fisheries, Forestry, and Agriculture, Forestry and Wildlife Branch, Government of Newfoundland and Labrador, Corner Brook, NL, Canada

4 Department of Fisheries and Oceans, Northwest Atlantic Fisheries Centre, St. John's, NL, Canada

5 Department of Ocean Sciences, Memorial University of Newfoundland, St. John's, NL, Canada

6 Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

7 Department of Biology, University of Victoria, Victoria, BC, Canada

8 Centre for Biomedical Research, University of Victoria, Victoria, BC, Canada

Current Address for KKS Layton: School of Biological Sciences, University of Aberdeen, Aberdeen, United Kingdom

Corresponding author email: sarahsalisbury13@gmail.com

Journal: The American Naturalist

ORCID

Sarah J. Salisbury https://orcid.org/0000-0001-7637-7742

Gregory R. McCracken https://orcid.org/0000-0002-1701-1529

Kara K. S. Layton https://orcid.org/0000-0002-4302-3048

Tony Kess https://orcid.org/0000-0002-1079-3791

Cameron M. Nugent https://orcid.org/0000-0002-1135-2605

Jong S. Leong https://orcid.org/0000-0001-6521-9239

Ben F. Koop https://orcid.org/0000-0003-0045-5200

Ian R. Bradbury https://orcid.org/0000-0002-8152-4943

Daniel E. Ruzzante https://orcid.org/0000-0002-8536-8335

Supporting Information: Genotyping Details

When using Axiom Analysis Suite to genotype samples we used default sample quality control thresholds: dish quality control ≥ 0.82, quality control call rate ≥ 0.97, and average call rate of passing samples on a given plate ≥ 0.985. We regenerated SNP Metrics using the "Run PS Supplemental" option as recommended (Axiom Analysis Suite User manual version 3.1) for complex genomes to screen out putative paralogous sequence variants given the potential that some regions of the charr genome may remain un-diploidized after the salmonid whole genome duplication. Those SNPs categorized as "PolyHighResolution", "NoMinorHom" and "MonoHighResolution" were used in analyses. Samples from 2010-2015 (batch 1) were extracted in a different lab and sequenced at a separate time from the 2017 samples (batch 2) and were therefore analyzed as separate "batches" in accordance with Axiom Analysis Suite User manual (version 3.1). Four samples from the 2010-2015 batch were sequenced and genotyped two times in order to allow for the screening out of those SNPs which were not identically genotyped within individuals, however both replicates passed quality control measures in only three of these samples. Replicate genotypes of a single individual were combined for those SNPs where one of the two replicates was missing a genotype. We removed 321 SNPs from the analysis which demonstrated inconsistent genotypes across two technical replicates in any of the three samples for which both technical replicates passed quality control measures. Most landlocked populations were analysed in one batch, while the second batch was comprised of mostly anadromous populations. When two different batches were compared, only those SNPs passing AxiomAnalysisSuite QC in both batches were retained. Additionally, when conducting pairwise comparisons between landlocked and anadromous populations SNPs with a frequency of one particular allele > 0.95 in one batch but < 0.05 in another were removed in order to exclude those SNPs that were genotyped inconsistently across the two batches used in this study. See Table S3 for the SNPs retained after each filtering step for paired sympatric morph comparisons; Table S6 for the SNPs retained after each filtering step for landlocked vs. anadromous comparisons.

Supporting Information: C-Score Analyses

The total number of SNPs used in each C-score calculation included only those SNPs which passed filtering (either monomorphic or polymorphic) for all locations considered in the analyses. Therefore, where locations sourced from different batches were compared, as in the landlocked vs. anadromous paired comparisons, SNPs which were genotyped in one batch but not the other (as well as those SNPs that were inconsistently genotyped across batches such that the frequency of one particular allele > 0.95 in one batch but < 0.05 in another) were removed from the analysis prior to C-Score calculations.

Supporting Information: More ADMIXTURE Results for Populations other than WP-L, LO-L

While the CV was marginally lower for K = 2 than K=1 ($\Delta = 0.00236$, Table S4) for KNU-L, this was driven by only two individuals (Fig.S7). K = 2 was supported within the Voisey Bay drainage (Table S4) but this corresponded to genetic differences between the two landlocked locations GB-L and SLU-L (Fig.S8). Therefore, only WP-L and LO-L showed evidence of genetic sub-structuring.

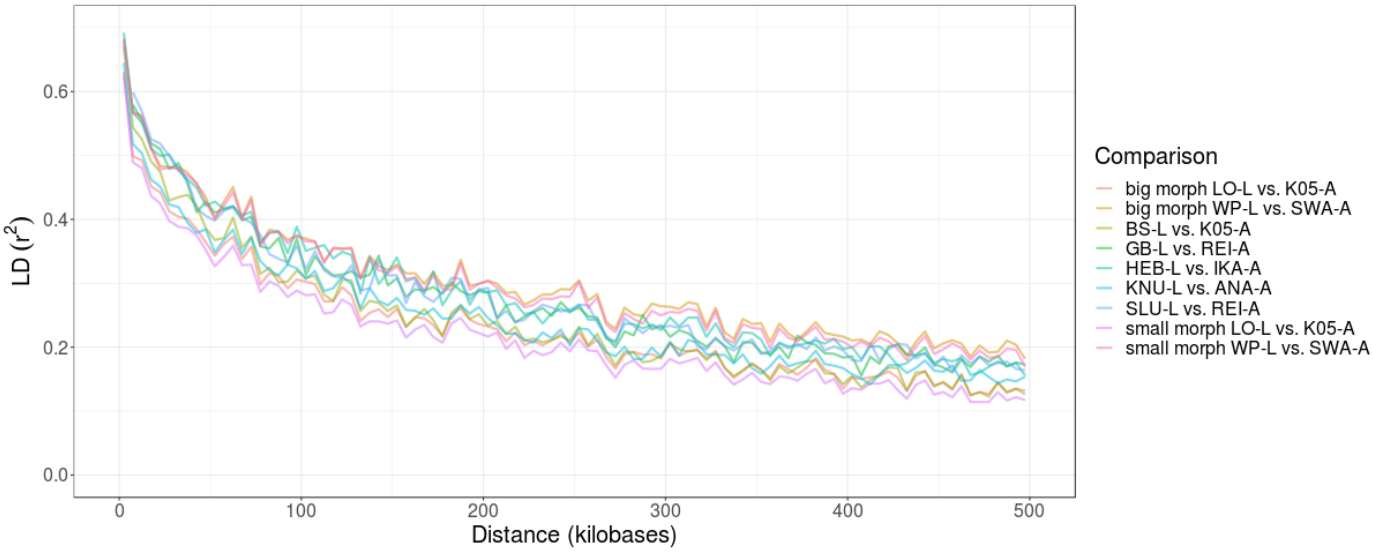


Fig.S1 Linkage disequilibrium decay in the first 500 kbp averaged over 5000 bp windows for all landlocked vs. anadromous population comparisons. Note the rapid decay in LD within the first 10000 bp.

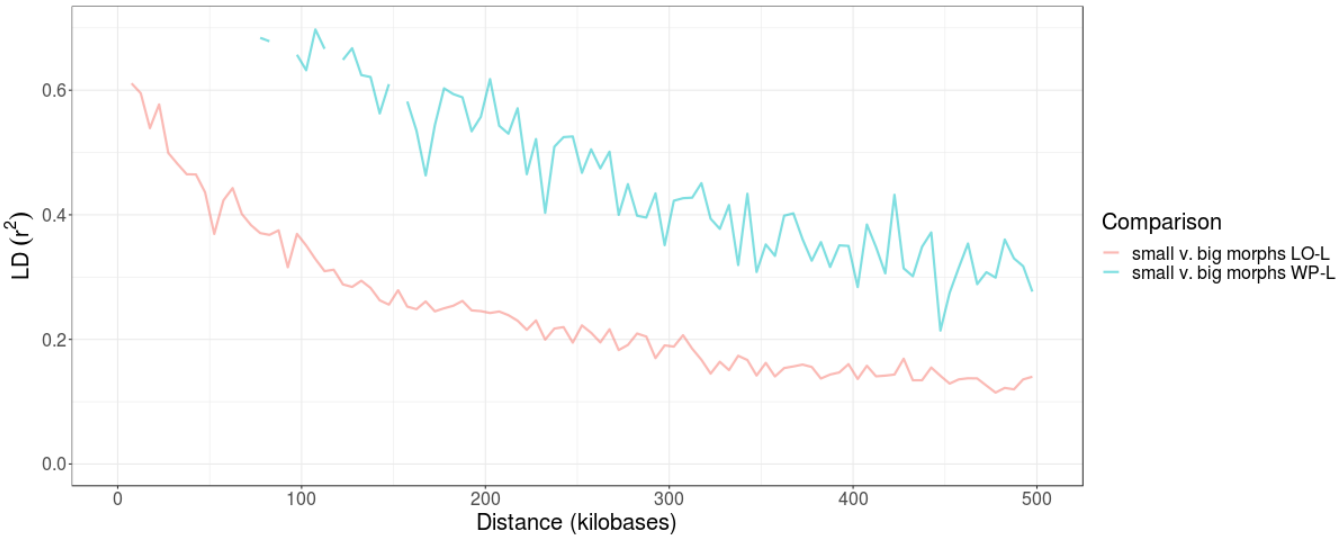


Fig.S2 Linkage disequilibrium decay in the first 500 kbp averaged over 5000 bp windows for all pairwise comparisons of sympatric small vs. big morphs.

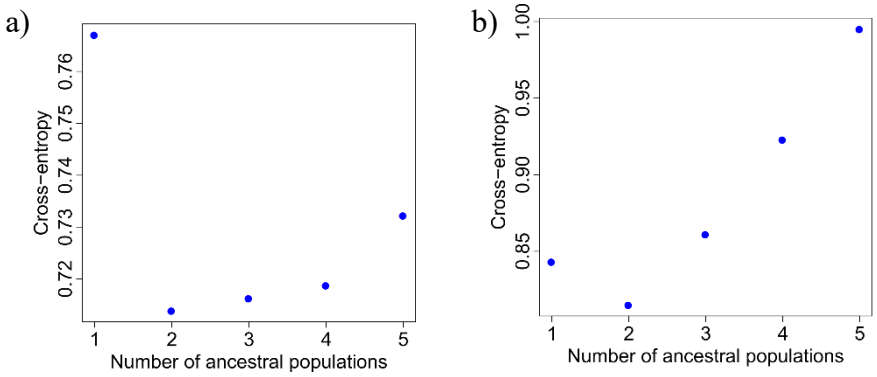


Fig.S3 SNMF cross-entropy values for K = 1-5 for a) WP-L, b) LO-L.

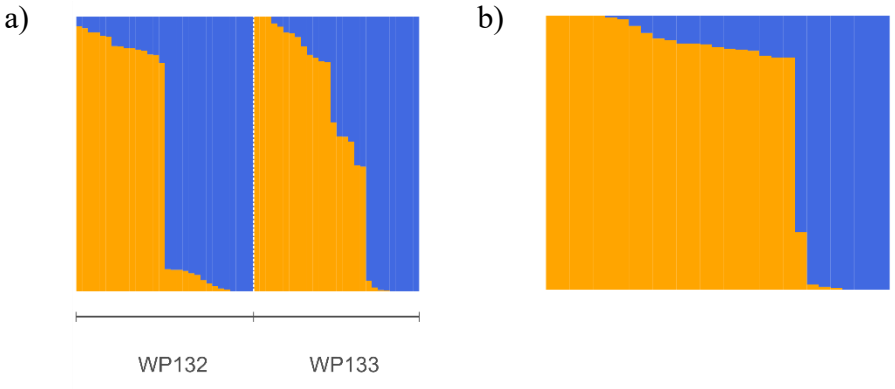


Fig. S4 SNMF plots of K = 2 for a) WP-L, b) LO-L. Orange bars indicate small morph samples, blue bars indicate big morph samples.

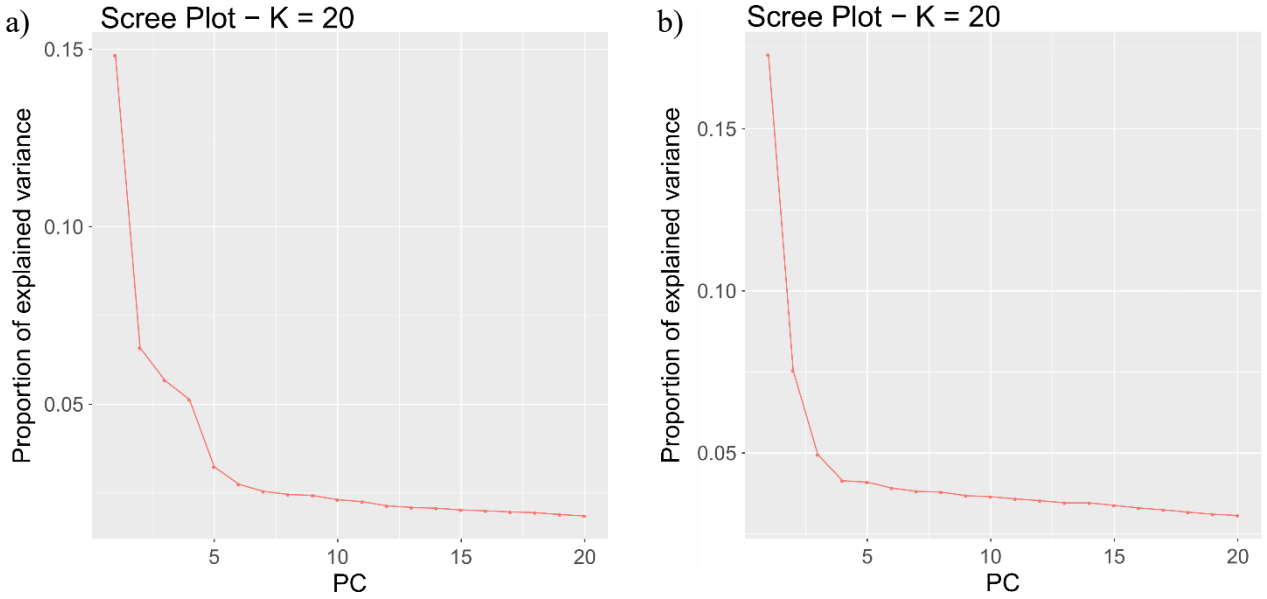


Fig.S5 Proportion of explained variance for each PC of PCAdapt population structure analysis for a) WP-L, b) LO-L.

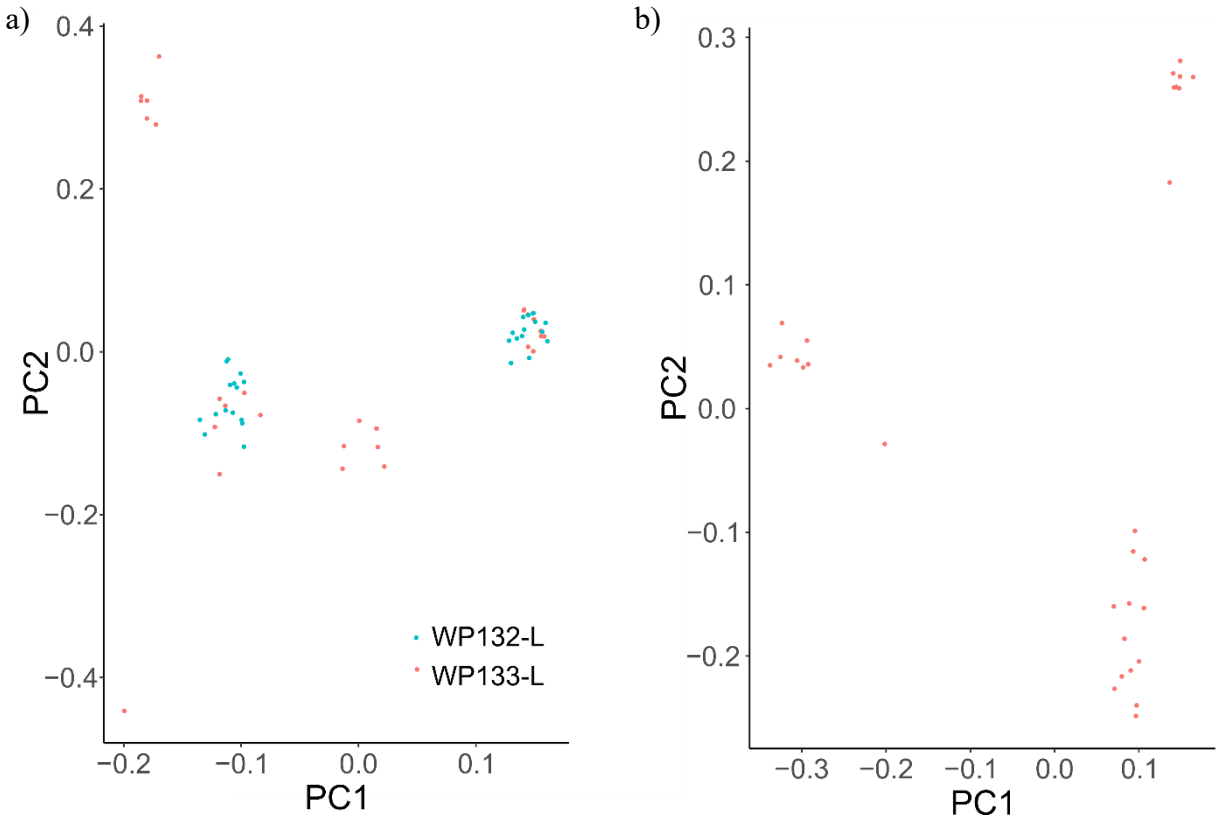


Fig.S6 Proportion of explained variance for each PC of PCAdapt population structure analysis for a) WP-L, b) LO-L.

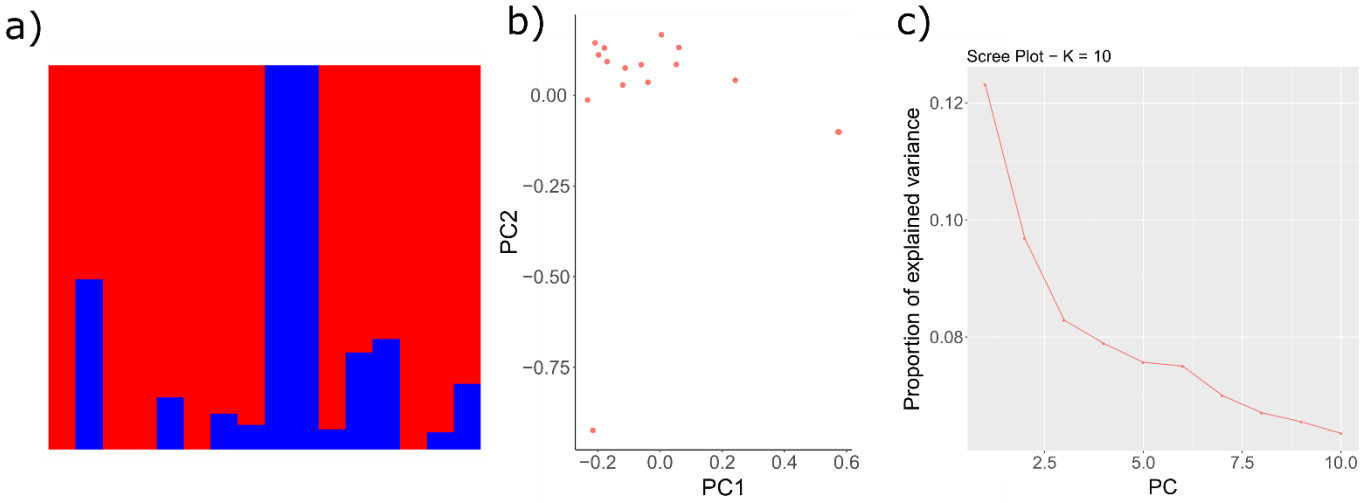


Fig.S7 Genetic structure within KNU-L based on N = 7378 SNPs. a) ADMIXTURE plot for K = 2, b) PCA, c) Scree Plot of the proportion of genetic variance explained by each PC.

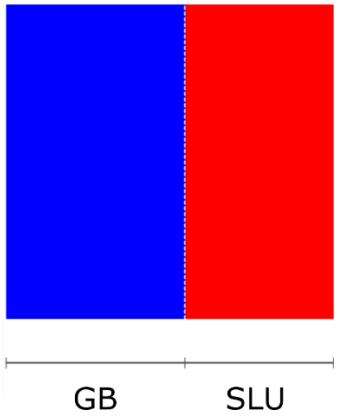


Fig.S8 ADMIXTURE plot for K = 2 within Voisey Bay landlocked lakes GB-L and SLU-L based on N = 13886 SNPs.

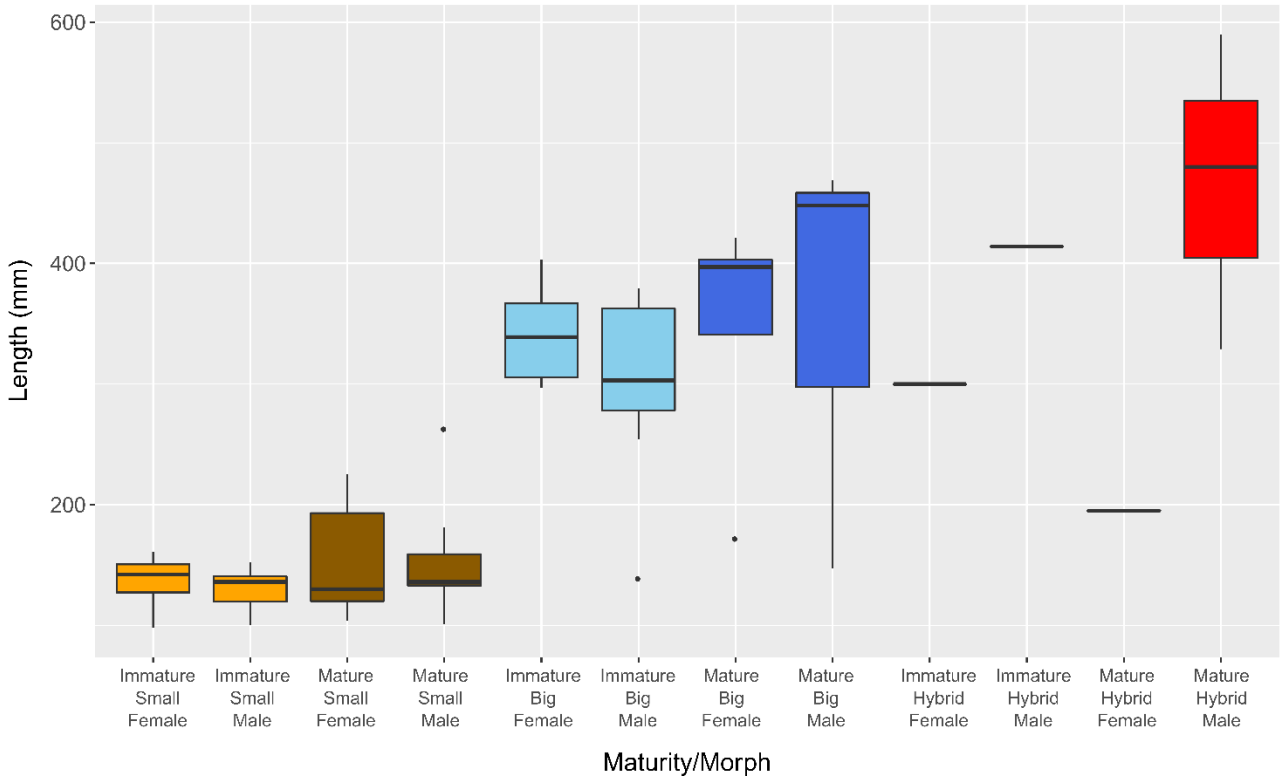


Fig.S9 Boxplots demonstrating length of fish by maturity, morph type, and sex for N=55 samples from WP-L.

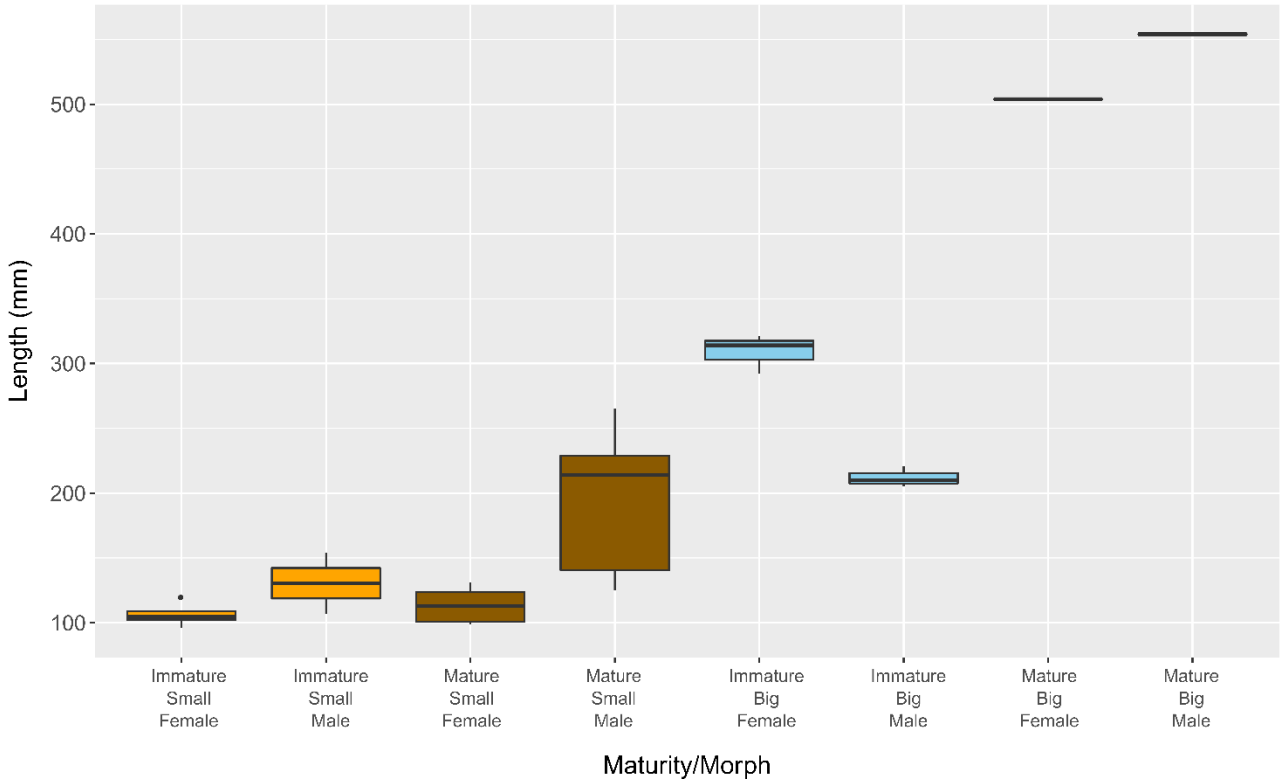


Fig.S10 Boxplots demonstrating length of fish by maturity, morph type, and sex for N=29 samples from LO-L.

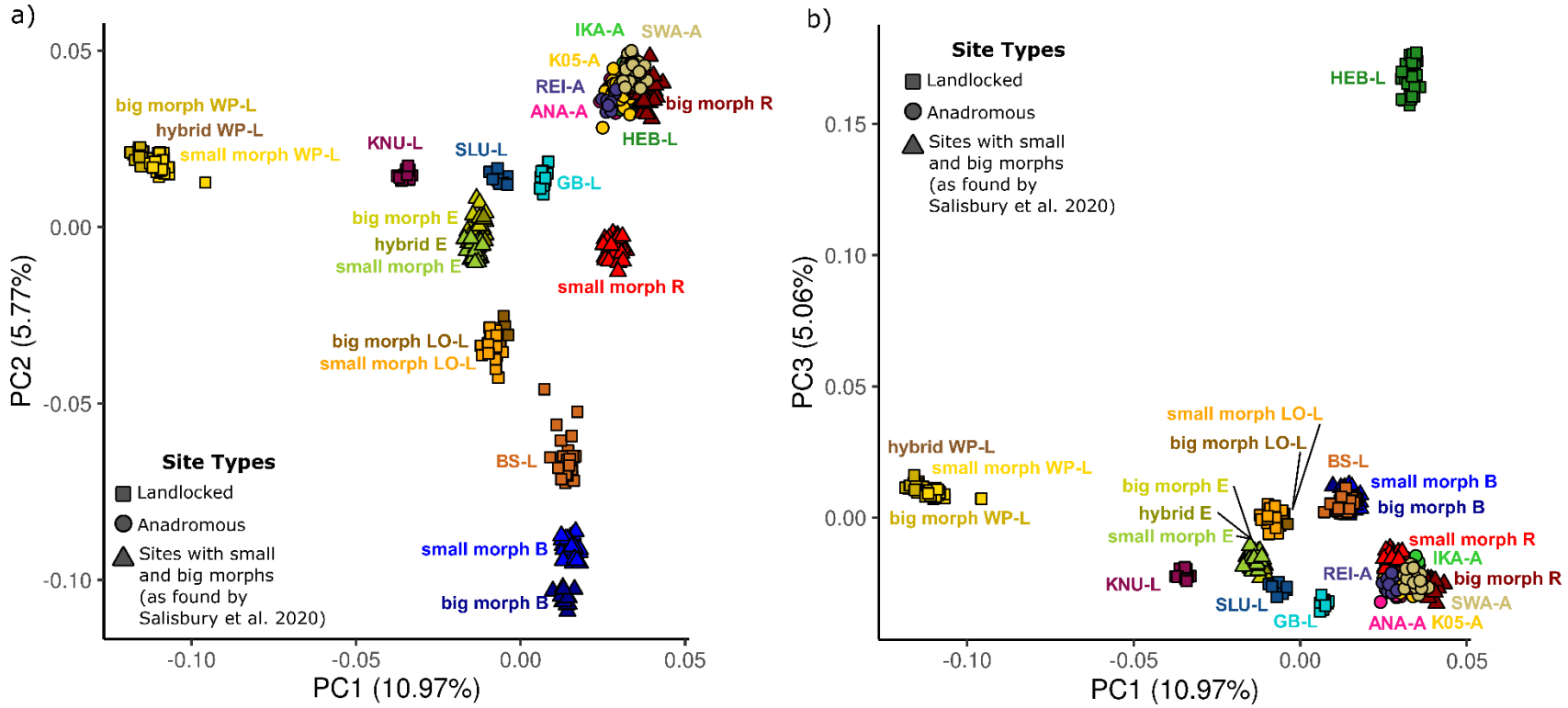


Fig.S11 PCA based on N = 8523 SNPs of landlocked (-L; denoted by squares) and anadromous (-A; denoted by circles) populations as well as small and big morphs from three putatively sea-accessible lakes (Ramah-R, Brooklyn-B, Esker North-E; denoted by triangles) detected by Salisbury et al. (2020). a) PC1 vs. PC2, b) PC1 vs. PC3. Note that unlike in Fig.3, HEB-L groups with anadromous populations (-A) in PC1 vs. PC2. However, HEB-L is clearly distinct from these anadromous populations by PC3.

	Saglek				Hebron		Okak				Anaktalik		Voisey			Ramah		Brooklyn		Esker North		
	sWP-L	**bWP-L**	**hWP-L**	**SWA-A**	**HEB-L**	**IKA-A**	**BS-L**	**sLO-L**	**bLO-L**	**K05-A**	**KNU-L**	**ANA-A**	**SLU-L**	**GB-L**	**REI-A**	**sR**	**bR**	**sB**	**bB**	**sE**	**bE**	**hE**
sWP-L (N=28)	0.00	0.16	0.11	0.36	0.57	**0.37**	0.42	0.38	0.53	**0.38**	0.58	**0.39**	0.54	0.57	**0.48**	0.41	0.37	0.44	0.59	0.36	0.40	0.45
bWP-L (N=24)	0.16	0.00	0.17	0.39	0.60	**0.41**	0.45	0.42	0.60	**0.41**	0.64	**0.42**	0.60	0.62	**0.53**	0.44	0.40	0.47	0.63	0.40	0.44	0.53
hWP-L (N=6)	0.11	0.17	0.00	0.29	0.53	**0.30**	0.37	0.32	0.49	**0.31**	0.59	**0.33**	0.50	0.54	**0.40**	0.35	0.30	0.40	0.56	0.32	0.34	0.39
SWA-A (N=30)	0.36	0.39	0.29	0.00	**0.26**	0.03	**0.16**	**0.15**	**0.21**	0.05	**0.32**	0.07	**0.22**	**0.24**	0.08	0.13	0.04	0.19	0.27	0.16	0.16	0.14
HEB-L (N=30)	0.57	0.60	0.53	**0.26**	0.00	0.26	0.33	0.33	0.43	**0.26**	0.52	**0.27**	0.43	0.45	**0.32**	0.30	0.26	0.35	0.46	0.33	0.35	0.36
IKA-A (N=25)	**0.37**	**0.41**	**0.30**	0.03	0.26	0.00	0.16	0.15	0.21	0.04	**0.32**	0.06	**0.21**	**0.23**	0.07	0.12	0.05	0.18	0.27	0.15	0.16	0.13
BS-L (N=29)	0.42	0.45	0.37	**0.16**	0.33	0.16	0.00	0.13	0.23	**0.17**	0.38	**0.18**	0.29	0.31	**0.20**	0.15	0.15	0.12	0.25	0.19	0.21	0.19
sLO-L (N=21)	0.38	0.42	0.32	**0.15**	0.33	0.15	0.13	0.00	0.18	**0.15**	0.36	**0.16**	0.26	0.29	**0.18**	0.16	0.15	0.16	0.29	0.13	0.16	0.13
bLO-L (N=8)	0.53	0.60	0.49	**0.21**	0.43	0.21	0.23	0.18	0.00	**0.22**	0.51	**0.23**	0.39	0.41	**0.27**	0.25	0.21	0.28	0.40	0.26	0.26	0.28
K05-A (N=29)	**0.38**	**0.41**	**0.31**	0.05	**0.26**	0.04	**0.17**	**0.15**	**0.22**	0.00	0.32	0.03	**0.20**	**0.21**	0.04	0.13	0.06	0.18	0.27	0.15	0.16	0.13
KNU-L (N=16)	0.58	0.64	0.59	**0.32**	0.52	**0.32**	0.38	0.36	0.51	0.32	0.00	**0.32**	0.44	0.47	**0.40**	0.36	0.32	0.39	0.55	0.33	0.36	0.41
ANA-A (N=30)	**0.39**	**0.42**	**0.33**	0.07	**0.27**	0.06	**0.18**	**0.16**	**0.23**	0.03	**0.32**	0.00	0.19	0.20	0.04	0.13	0.08	0.19	0.28	0.14	0.15	0.14
SLU-L (N=10)	0.54	0.60	0.50	**0.22**	0.43	**0.21**	0.29	0.26	0.39	**0.20**	0.44	0.19	0.00	0.27	**0.23**	0.25	0.22	0.30	0.43	0.23	0.25	0.26
GB-L (N=12)	0.57	0.62	0.54	**0.24**	0.45	**0.23**	0.31	0.29	0.41	**0.21**	0.47	0.20	0.27	0.00	**0.25**	0.26	0.24	0.31	0.45	0.26	0.28	0.30
REI-A (N=9)	**0.48**	**0.53**	**0.40**	0.08	**0.32**	0.07	**0.20**	**0.18**	**0.27**	0.04	**0.40**	0.04	**0.23**	**0.25**	0.00	0.14	0.09	0.21	0.34	0.16	0.18	0.15
sR (N=32)	0.41	0.44	0.35	0.13	0.30	0.12	0.15	0.16	0.25	0.13	0.36	0.13	0.25	0.26	0.14	0.00	0.11	0.17	0.28	0.17	0.19	0.16
bR (N=28)	0.37	0.40	0.30	0.04	0.26	0.05	0.15	0.15	0.21	0.06	0.32	0.08	0.22	0.24	0.09	0.11	0.00	0.18	0.26	0.16	0.17	0.14
sB (N=42)	0.44	0.47	0.40	0.19	0.35	0.18	0.12	0.16	0.28	0.18	0.39	0.19	0.30	0.31	0.21	0.17	0.18	0.00	0.20	0.20	0.23	0.22
bB (N=16)	0.59	0.63	0.56	0.27	0.46	0.27	0.25	0.29	0.40	0.27	0.55	0.28	0.43	0.45	0.34	0.28	0.26	0.20	0.00	0.32	0.33	0.36
sE (N=33)	0.36	0.40	0.32	0.16	0.33	0.15	0.19	0.13	0.26	0.15	0.33	0.14	0.23	0.26	0.16	0.17	0.16	0.20	0.32	0.00	0.08	0.05
bE (N=21)	0.40	0.44	0.34	0.16	0.35	0.16	0.21	0.16	0.26	0.16	0.36	0.15	0.25	0.28	0.18	0.19	0.17	0.23	0.33	0.08	0.00	0.03
hE (N=6)	0.45	0.53	0.39	0.14	0.36	0.13	0.19	0.13	0.28	0.13	0.41	0.14	0.26	0.30	0.15	0.16	0.14	0.22	0.36	0.05	0.03	0.00

F_{ST} 0.8 0.6 0.4 0.2 0

Fig.S12 Weighted pairwise Weir and Cockerham (1984) F_{ST}s between all landlocked (-L) and anadromous (-A) populations, as well as putatively sea-accessible populations from Salisbury et al. (2020) which contain small and big morphs (Ramah-R, Brooklyn-B, Esker North-E) based on N=21201 SNPs. F_{ST}s estimated between landlocked and anadromous populations are bolded. Those F_{ST}s highlighted in red correspond to the lowest F_{ST} calculated between a given landlocked population and any of the anadromous populations. F_{ST} values estimated between populations within the same drainage are within bolded black boxes.

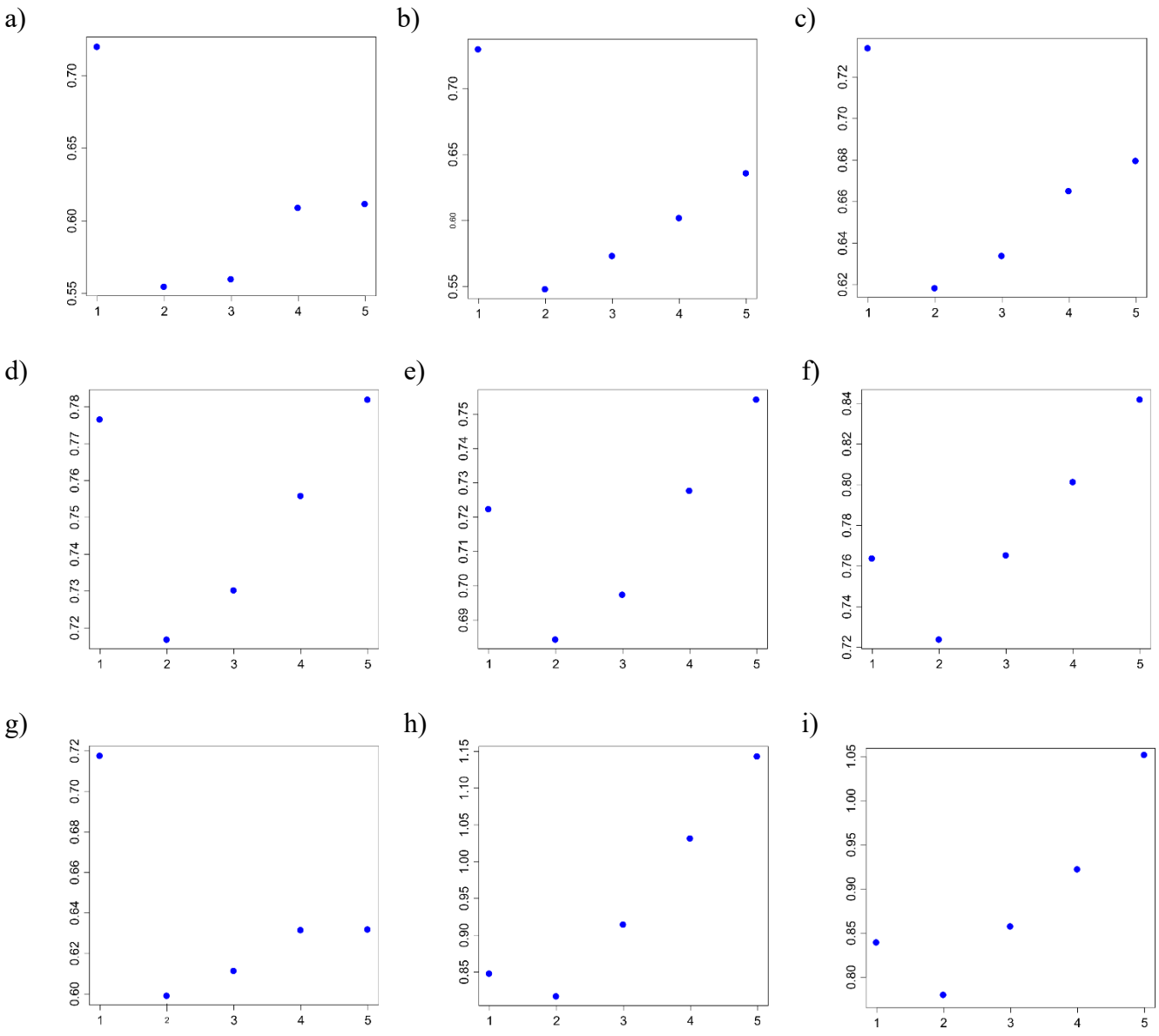


Fig.S13 SNMF cross-entropy values vs K-values for 9 landlocked vs. anadromous population comparisons: a) small morph WP-L vs. SWA-A, b) big morph WP-L vs. SWA-A, c) HEB-L vs. IKA-A, d) BS-L vs. K05-A, e) small morph LO-L vs. K05-A, f) big morph LO-L vs. K05-A, g) KNU-L vs. ANA-A, h) SLU-L vs. REI-A, i) GB-L vs. REI-A.

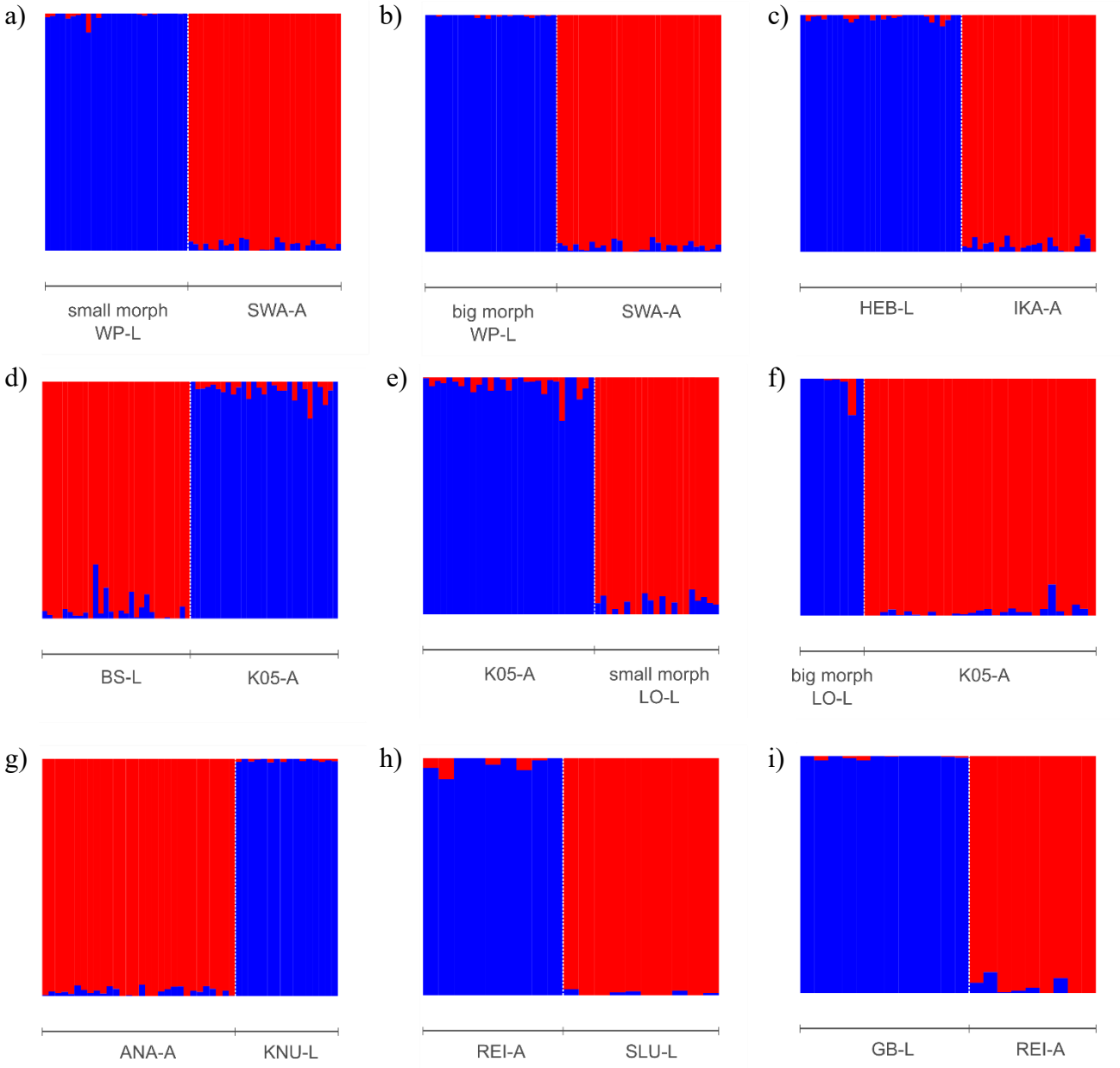


Fig.S14 SNMF plots for K = 2 for 9 landlocked vs. anadromous population comparisons: a) small morph WP-L vs. SWA-A, b) big morph WP-L vs. SWA-A, c) HEB-L vs. IKA-A, d) BS-L vs. K05-A, e) small morph LO-L vs. K05-A, f) big morph LO-L vs. K05-A, g) KNU-L vs. ANA-A, h) SLU-L vs. REI-A, i) GB-L vs. REI-A.

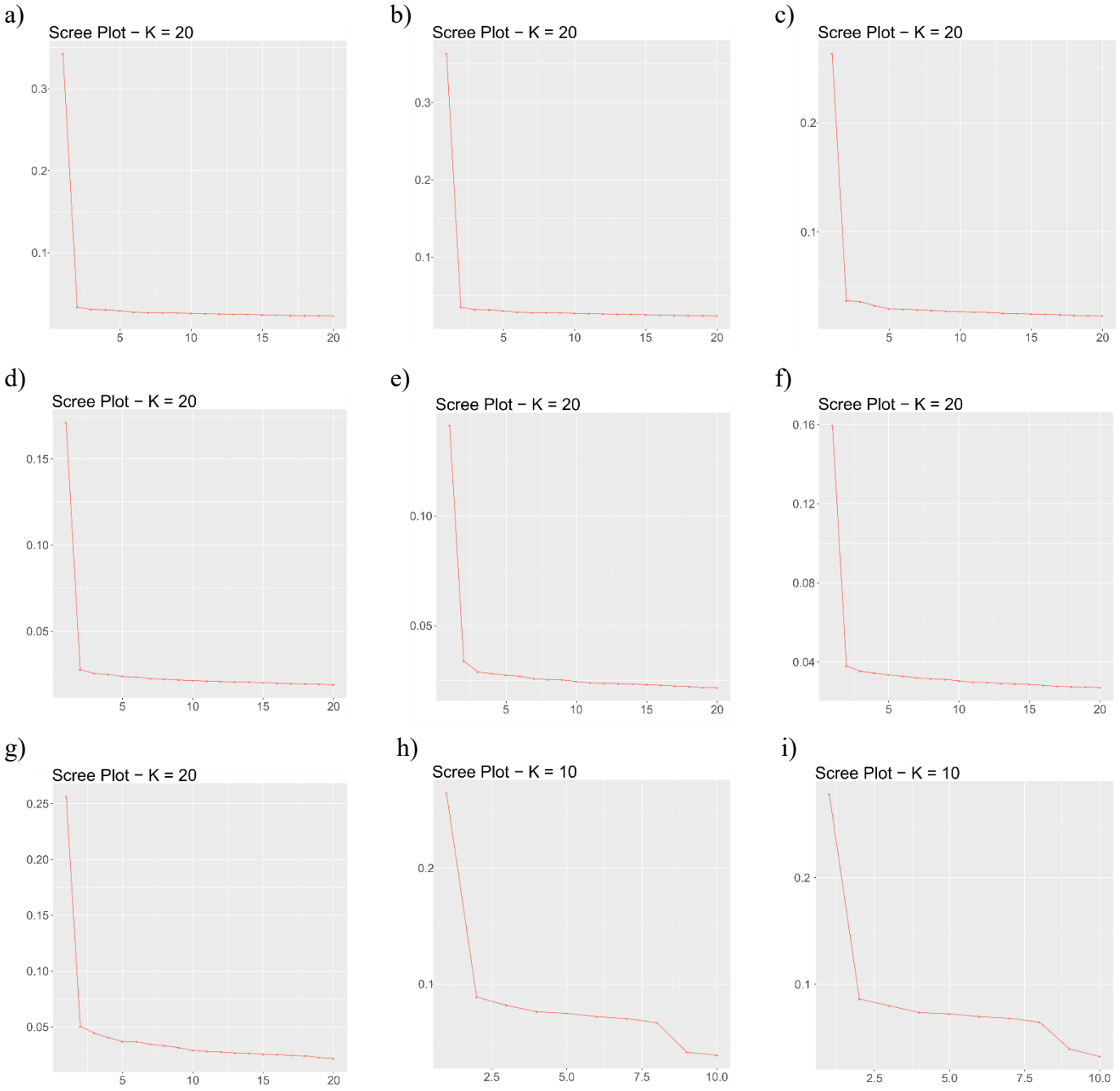


Fig.S15 Proportion of explained variance vs PC of PCAdapt population structure analysis for a) small morph WP-L vs. SWA-A, b) big morph WP-L vs. SWA-A, c) HEB-L vs. IKA-A, d) BS-L vs. K05-A, e) small morph LO-L vs. K05-A, f) big morph LO-L vs. K05-A, g) KNU-L vs. ANA-A, h) SLU-L vs. REI-A, i) GB-L vs. REI-A.

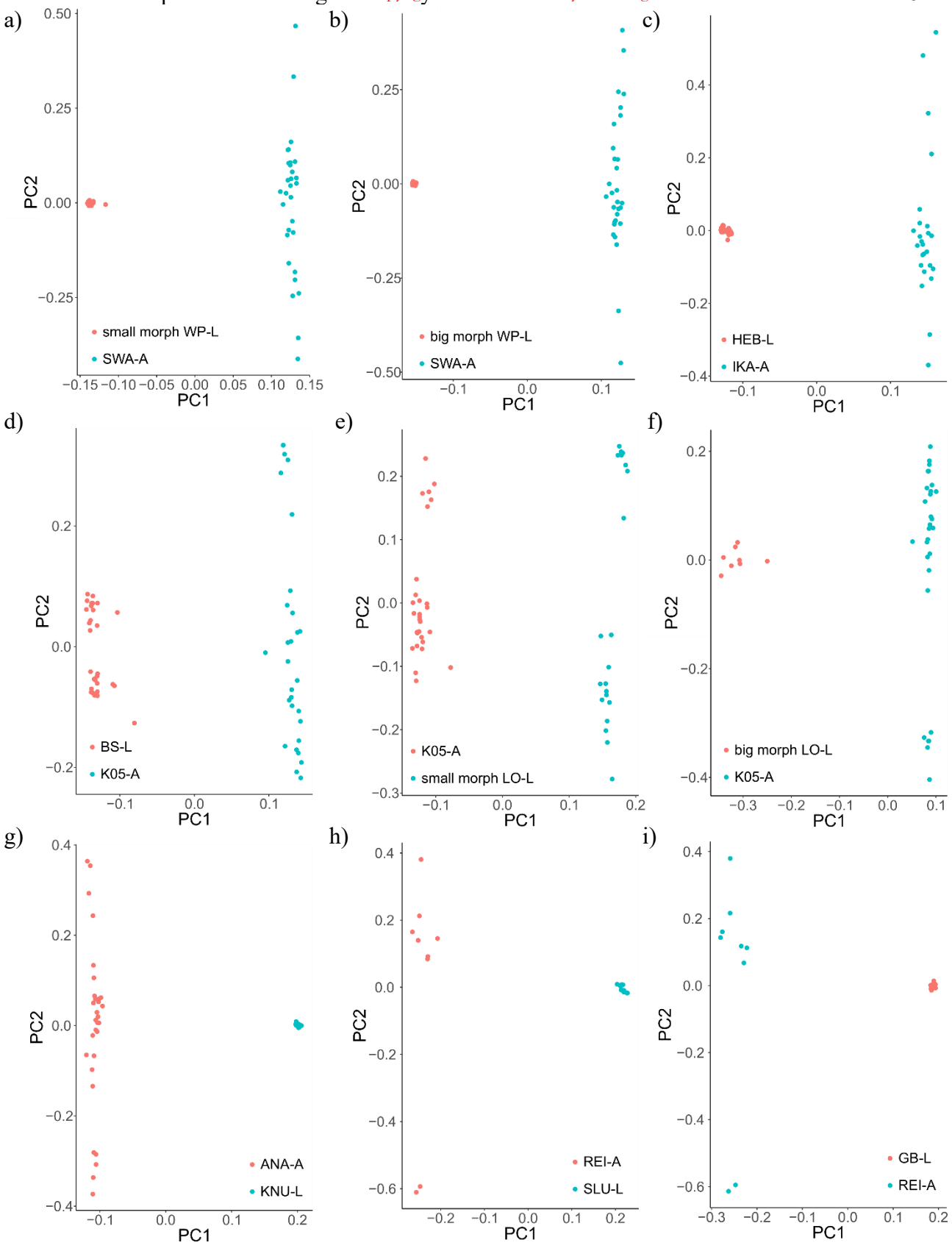


Fig.S16 PCAdapt plots for a) small morph WP-L vs. SWA-A, b) big morph WP-L vs. SWA-A, c) HEB-L vs. IKA-A, d) BS-L vs. K05-A, e) small morph LO-L vs. K05-A, f) big morph LO-L vs. K05-A, g) KNU-L vs. ANA-A, h) SLU-L vs. REI-A, i) GB-L vs. REI-A

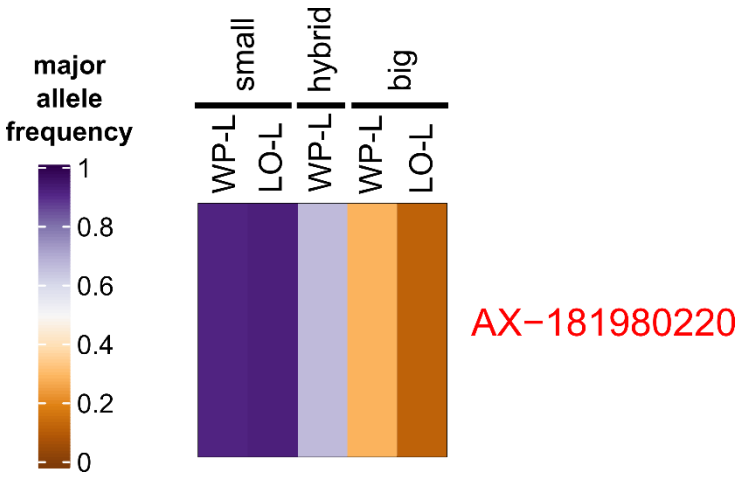


Fig.S17 Heatmap of allele frequencies of loci detected as outliers within both WP-L and LO-L. The names of SNPs which show parallel allelic trends across locations are highlighted in reds.

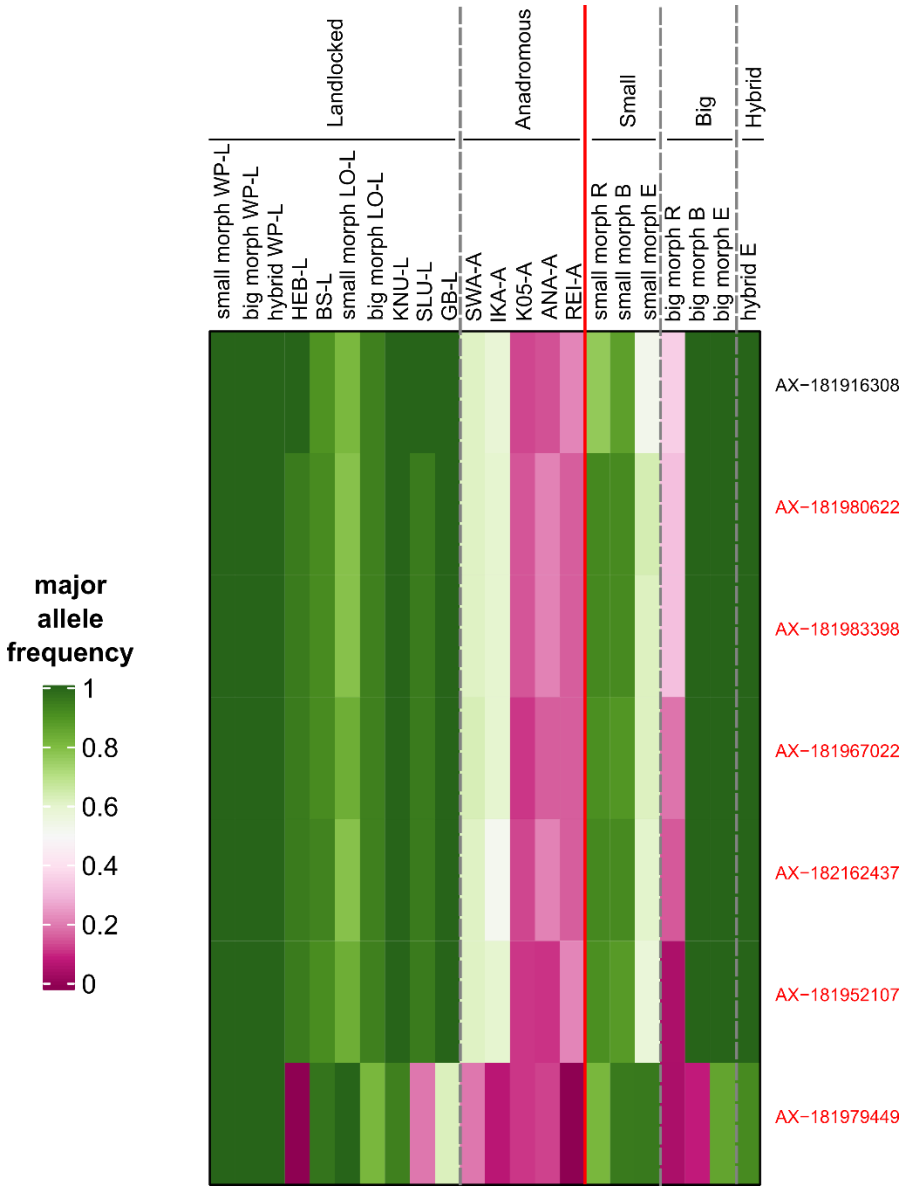


Fig.S18 Heatmap of allele frequencies for those SNPs detected as outliers for five or more of seven paired landlocked and anadromous populations (1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs. K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A) and at least one location with sympatric small (s), and big (b morphs from Salisbury et al. (2020). The names of SNPs which show parallel allelic trends across the locations in which a SNP was detected as an outlier are highlighted in red.

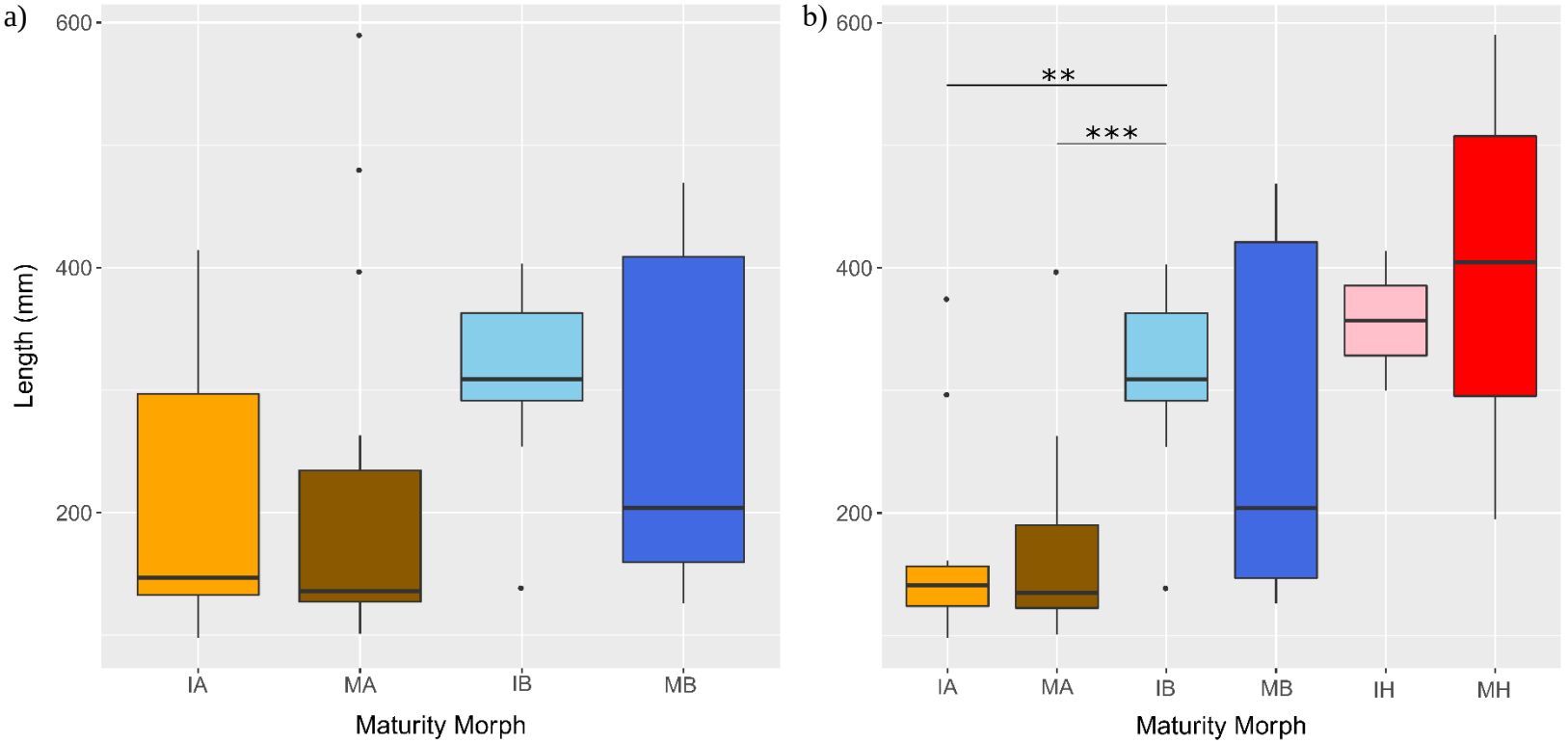


Fig.S19 Boxplots demonstrating length of fish by maturity (immature (I), mature (M)) and assigned genetic group (A or B, as assigned based on 11 microsatellites in Salisbury et al. 2018) in WP-L a) before and b) after separating individuals assigned as putative hybrids (H) detected using 6404 SNPs in this study. Lengths of the WP individuals used in this study were compared according to maturity status (N=55, as three samples had unknown maturity status), and their genetic group as assigned using microsatellites by Salisbury et al. (2018). No difference in the variance of fork-lengths was found between maturity-morph type groups before separating individuals assigned as putative hybrids (Levene test, $F_{(3,51)} = 2.77$, p-value > 0.05). A one-way ANOVA testing the interaction of microsatellite-assigned genetic group and maturity found no significant differences in length between genetic groups ($F_{3,51} = 2.51$, p-value > 0.05) (Fig. S12a). However, after separating the six individuals identified as hybrids based on our SNP analysis, more significant differences in length were observed among the microsatellite-assigned genetic groups (Levene test, $F_{(5,49)} = 4.45$, p-value < 0.01; one-way Welch's ANOVA testing for differences in lengths among six groups varying in maturity, SNP-assigned hybrid status, and microsatellite-assigned genetic group, $F_{5,7.92} = 5.51$, $p < 0.05$) (Fig. S12b). Asterisks indicate significant posthoc Games-Howell tests among groups in Fig. S12b (* = $p < 0.05$, ** = $p < 0.01$, *** $p < 0.001$). Therefore, the lack of size differences observed between genetically distinguishable morphs in Salisbury et al. (2018) was potentially due to the failure to remove putative hybrid individuals. Further discrepancies in morph assignment between our SNP data and the microsatellite data may be due to the anticipated greater assignment accuracy associated with 6404 SNPs in comparison to 11 microsatellites.

Table S1 Details of sampling locations for landlocked (L) and anadromous (A) populations.

Code	Location	Drainage	Population Type (L/A)	Latitude	Longitude	Sampling Year(s)	N	N Passing QC
SWA-A	Southwest Arm	Saglek	A	58.46825	-63.64623	2017	30	30
WP-L	WP133-L	Saglek	L	58.27167	-64.03136	2014	28	28
	WP132-L	Saglek	L	58.28016	-63.9693	2014	30	30
IKA-A	Ikarut River	Hebron	A	58.16057	-63.16141	2017	30	25
HEB-L	Hebron Lake	Hebron	L	58.14611	-63.59133	2015	30	30
K05-A	North River	Okak	A	57.50159	-62.74318	2015	30	29
BS-L	Beachy Strip Lake	Okak	L	57.66161	-62.95445	2015	30	29
LO-L	Lonely Lake	Okak	L	57.63915	-63.23292	2015	30	29
ANA-A	Anaktalik River	Anaktalik	A	56.49753	-62.93309	2017	30	30
KNU-L	Knumandi Lake	Anaktalik	L	56.58141	-63.32335	2011	20	16
REI-A	Reid Brook	Voisey	A	56.30319	-62.08522	2017	30	9
SLU-L	Slushy Lake	Voisey	L	56.41561	-64.10225	2010, 2011, 2012, 2013	11	10
GB-L	Genetics B Lake	Voisey	L	56.11067	-63.38858	2010, 2011	13	12
						Totals	**342**	**307**

Table S2 Details of sampling locations for putatively sea-accessible populations with sympatric small and big morphs as detected in Salisbury et al. (2020).

Code	Location	Drainage	Latitude	Longitude	Sampling Year(s)	Morph	N
R	Ramah	Ramah	58.8413796	-63.477406	2014	Small	32
R	Ramah	Ramah	58.8413796	-63.477406	2014	Big	28
B	Brooklyn	Okak	57.7264811	-62.473362	2015	Small	42
B	Brooklyn	Okak	57.7264811	-62.473362	2015	Big	16
E	Esker North	Tikkoatokak	57.1488411	-62.878159	2015	Small	33
E	Esker North	Tikkoatokak	57.1488411	-62.878159	2015	Big	21
E	Esker North	Tikkoatokak	57.1488411	-62.878159	2015	Hybrid	6
						Totals	**178**

Table S3 Number of SNPs which passed filtering within each landlocked location.

Location	Sample Batch	"PolyHighResolution", "NoMinorHom", "MonoHighResolution" SNPs as assigned from AxiomAnalysisSuite *after removing 321 SNPs inconsistently genotyped across technical replicates	SNPs Passing MAF = 0.01
WP-L	1	62812	6404
HEB-L	1	62812	10980
BS-L	1	62812	16069
LO-L	1	62812	16702
KNU-L	1	62812	7378
Voisey Bay Landlocked (GB-L and SLU-L)	1	62812	13886

Table S4 Average cross-validation error estimates for ADMIXTURE results for each lake and K-value. Lowest values for each lake are shaded.

K	WP-L (N = 6404 SNPs)	HEB-L (N = 10980 SNPs)	BS-L (N = 16069 SNPs)	LO-L (N = 16702 SNPs)	KNU-L (N = 7378 SNP)	Voisey Bay Landlocked (SLU-L and GB-L) (N = 13886 SNPs)
1	0.58803	0.59508	0.59270	0.63386	0.65604	0.67043
2	0.52533	0.62951	0.64830	0.61114	0.65368	0.61211
3	0.5439	0.71219	0.78784	0.66757	0.84033	0.81138
4	0.55642	0.80263	1.08615	0.80034	1.18553	0.88846
5	0.56519	0.97839	1.13428	0.95716	1.65730	1.15537

Table S5 Maturity, sex and glacial lineage information for samples from landlocked lakes with no evidence based on ADMIXTURE of genetic sub-structuring. Lineage assignments are based on mtDNA D-loop haplotypes collected in Salisbury et al. (2019).

Location	N	Immature Males	Immature Females	Mature Males	Mature Females	Atlantic Lineage	Arctic Lineage	Acadian Lineage
HEB-L	30	5	11	11	3	24	0	0
BS-L	29	7	7	8	7	0	20	0
KNU-L	16					0	0	16
SLU-L	10					5	5	0
GB-L	12					12	0	0

Table S6 Number of SNPs passing filtering for each landlocked vs. anadromous comparison. Note when two different batches were compared, only those SNPs passing AxiomAnalysisSuite QC in both batches were retained.

	small morph WP-L vs. SWA-A	big morph WP-L vs. SWA-A	HEB-L vs. IKA-A	BS-L vs. K05-A	small morph LO-L vs. K05-A	big morph LO-L vs. K05-A	KNU-L vs. ANA-A	SLU-L vs. REI-A	GB-L vs. REI-A
Batches Compared	1 vs. 2	1 vs. 2	1 vs. 2	1 vs. 1	1 vs. 1	1 vs. 1	1 vs. 2	1 vs. 2	1 vs. 2
"PolyHighResolution", "NoMinorHom", "MonoHighResolution" SNPs as assigned from AxiomAnalysisSuite *after removing 321 SNPs inconsistently genotyped across technical replicates	58073	58073	58073	62812	62812	62812	58073	58073	58073
After removing SNPs genotyped inconsistently across batches	58068	58068	58068	62691	62691	62691	58068	58068	58068
After applying MAF of 0.01	20393	20361	19613	20334	22540	22185	19994	17385	17321

Table S7 The number of outlier SNPs which were detected by a each outlier detection method within each of nine comparisons of landlocked and anadromous populations.

Outlier Detection Method	small morph WP-L vs. SWA-A	big morph WP-L vs. SWA-A	HEB-L vs. IKA-A	BS-L vs. K05-A	small morph LO-L vs. K05-A	big morph LO-L vs. K05-A	KNU-L vs. ANA-A	SLU-L vs. REI-A	GB-L vs. REI-A
Both PCAdapt and FST	74	67	201	152	270	393	213	277	168
PCAdapt Only	1542	1299	345	14	13	1857	2083	263	300
FST Only	0	0	2	204	182	0	0	6	11
Total	1616	1366	548	370	465	2250	2296	546	479

Table S8 The number of outlier and non-outlier SNPs detected between landlocked and anadromous populations that were found to be polymorphic outliers, polymorphic non-outliers or non-polymorphic in each of the landlocked vs. anadromous population comparisons.

	small morph WP-L vs. SWA-A (N=20393)		big morph WP-L vs. SWA-A (N=20361)		HEB-L vs. IKA-A (N=19613)		BS-L vs. K05-A (N=20334)		small morph LO-L vs. K05-A (N=22540)		big morph LO-L vs. K05-A (N=22185)		KNU-L vs. ANA-A (N=19994)		SLU-L vs. REI-A (N=17385)		GB-L vs. REI-A (N=17321)	
	Outlier (N=1616)	Non-Outlier (N=18777)	Outlier (N=1366)	Non-Outlier (N=18995)	Outlier (N=548)	Non-Outlier (N=19065)	Outlier (N=370)	Non-Outlier (N=19964)	Outlier (N=465)	Non-Outlier (N=22075)	Outlier (N=2250)	Non-Outlier (N=19935)	Outlier (N=2296)	Non-Outlier (N=17698)	Outlier (N=546)	Non-Outlier (N=16839)	Outlier (N=479)	Non-Outlier (N=16842)
small morph WP-L vs. SWA-A Polymorphic Outlier	1616	0	988	622	76	1455	58	1515	103	1485	339	1234	384	1192	84	1463	67	1430
small morph WP-L vs. SWA-A Polymorphic Non-Outlier	0	18777	376	18359	436	16483	296	16148	329	16932	1617	15580	1859	14627	448	14375	400	14365
small morph WP-L vs. SWA-A Non-polymorphic	0	0	2	14	36	1127	16	2301	33	3658	294	3121	53	1879	14	1001	12	1047
big morph WP-L vs. SWA-A Polymorphic Outlier	988	376	1366	0	70	1214	53	1289	89	1265	327	1016	361	976	76	1203	49	1214
big morph WP-L vs. SWA-A Polymorphic Non-Outlier	622	18359	0	18995	438	16707	301	16348	343	17123	1623	15777	1880	14820	454	14615	417	14567
big morph WP-L vs. SWA-A Non-polymorphic	6	42	0	0	40	1144	16	2327	33	3687	300	3142	55	1902	16	1021	13	1061
HEB-L vs. IKA-A Polymorphic Outlier	76	436	70	438	548	0	33	475	27	498	89	433	106	410	27	446	21	443
HEB-L vs. IKA-A Polymorphic Non-Outlier	1455	16483	1214	16707	0	19065	314	16899	398	17565	1815	16097	2106	15037	494	15132	434	15155
HEB-L vs. IKA-A Non-polymorphic	85	1858	82	1850	0	0	23	2590	40	4012	346	3405	84	2251	25	1261	24	1244
BS-L vs. K05-A Polymorphic Outlier	58	296	53	301	33	314	370	0	118	252	188	182	104	247	39	299	34	302
BS-L vs. K05-A Polymorphic Non-Outlier	1515	16148	1289	16348	475	16899	0	19964	344	19492	1861	17920	2144	15547	503	15584	436	15594
BS-L vs. K05-A Non-polymorphic	43	2333	24	2346	40	1852	0	0	3	2331	201	1833	48	1904	4	956	9	946
small morph LO-L vs. K05-A Polymorphic Outlier	103	329	89	343	27	398	118	344	465	0	261	202	145	286	38	380	36	372
small morph LO-L vs. K05-A Polymorphic Non-Outlier	1485	16932	1265	17123	498	17565	252	19492	0	22075	1953	19617	2119	16122	502	15839	438	15857
small morph LO-L vs. K05-A Non-polymorphic	28	1516	12	1529	23	1102	0	128	0	0	36	116	32	1290	6	620	5	613
big morph LO-L vs. K05-A Polymorphic Outlier	339	1617	327	1623	89	1815	188	1861	261	1953	2250	0	405	1556	103	1734	81	1722
big morph LO-L vs. K05-A Polymorphic Non-Outlier	1234	15580	1016	15777	433	16097	182	17920	202	19617	0	19935	1853	14789	434	14431	393	14456
big morph LO-L vs. K05-A Non-polymorphic	43	1580	23	1595	26	1153	0	183	2	505	0	0	38	1353	9	674	5	664
KNU-L vs. ANA-A Polymorphic Outlier	384	1859	361	1880	106	2106	104	2144	145	2119	405	1853	2296	0	166	2042	135	2049
KNU-L vs. ANA-A Polymorphic Non-Outlier	1192	14627	976	14820	410	15037	247	15547	286	16122	1556	14789	0	17698	370	13986	335	13996
KNU-L vs. ANA-A Non-polymorphic	40	2291	29	2295	32	1922	19	2273	34	3834	289	3293	0	0	10	811	9	797
SLU-L vs. REI-A Polymorphic Outlier	84	448	76	454	27	494	39	503	38	502	103	434	166	370	546	0	159	374
SLU-L vs. REI-A Polymorphic Non-Outlier	1463	14375	1203	14615	446	15132	299	15584	380	15839	1734	14431	2042	13986	0	16839	311	15938
SLU-L vs. REI-A Non-polymorphic	69	3954	87	3926	75	3439	32	3877	47	5734	413	5070	88	3342	0	0	9	530
GB-L vs. REI-A Polymorphic Outlier	67	400	49	417	21	434	34	436	36	438	81	393	135	335	159	311	479	0
GB-L vs. REI-A Polymorphic Non-Outlier	1430	14365	1214	14567	443	15155	302	15594	372	15857	1722	14456	2049	13996	374	15938	0	16842
GB-L vs. REI-A Non-polymorphic	119	4012	103	4011	84	3476	34	3934	57	5780	447	5086	112	3367	13	590	0	0

Table S9 Top Biological Processes GO terms with an unadjusted p-value < 0.01 for outlier loci within at least five of seven paired landlocked and anadromous populations (1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs. K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A). Gene universe was generated using only those SNPs which passed filtering in both sample batches (N = 58068 SNPs) and were within 5000 bp of a gene's coding sequence resulting in a total of N = 19303 genes in the gene universe.

GO.ID	Term	Annotated	Significant	Expected	weight01	weight01padj
GO:1903259	exon-exon junction complex disassembly	1	1	0	0.0013	1
GO:0010390	histone monoubiquitination	51	2	0.07	0.0019	1
GO:0034402	recruitment of 3'-end processing factors...	2	1	0	0.0026	1
GO:1990091	sodium-dependent self proteolysis	2	1	0	0.0026	1
GO:0014718	positive regulation of satellite cell ac...	4	1	0.01	0.0051	1
GO:2001168	positive regulation of histone H2B ubiqu...	4	1	0.01	0.0051	1
GO:0031399	regulation of protein modification proce...	3390	5	4.33	0.0053	1
GO:0018022	peptidyl-lysine methylation	202	2	0.26	0.0074	1

Table S10 Genes for which different paralogous outlier loci were detected in different landlocked vs. anadromous population comparisons for five or more of seven comparisons (1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs. K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A). The method by which each SNP was identified as an outlier is denoted for each landlocked vs. anadromous population comparison (P – PCAdapt, F- F_{ST}). For a given gene, starred linkage groups are homeologous.

							Method by which outlier was detected								
General Protein Name	Linkage Group	Protein Code	Specific Protein Name	SNP Code	Position (Mbp)	Position Relative to CDS (kbp)	small morph WP-L vs. SWA-A	big morph WP-L vs. SWA-A	HEB-L vs. IKA-A	BS-L vs. K05-A	small morph LO-L vs. K05-A	big morph LO-L vs. K05-A	KNU-L vs. ANA-A	SLU-L vs. REI-A	GB-L vs. REI-A
adhesion G-protein coupled receptor G2	AC06.2	XP_023845604.1	adhesion G-protein coupled receptor G2-like	AX-181975480	1.9	0						P			
	AC06.2	XP_023845604.1	adhesion G-protein coupled receptor G2-like	AX-181929714	1.9	0				F					
	AC14	XP_023856439.1	adhesion G-protein coupled receptor G2 isoform X2	AX-181935811	3.9	0			P					P	
	AC14	XP_023856439.1	adhesion G-protein coupled receptor G2 isoform X2	AX-181964222	3.9	0						P			
	AC23	XP_023824679.1	adhesion G-protein coupled receptor G2	AX-181963321	39.5	1.3	P	P							
chloride channel protein 2	AC04p	XP_023838719.1	LOW QUALITY PROTEIN: chloride channel protein 2-like	AX-181945304	25.0	0	P							P	
	AC32	XP_023833454.1	chloride channel protein 2-like	AX-181978855	14.4	0	P	P	P			P	P		
E3 ubiquitin-protein ligase BRE1B	AC17	XP_023860736.1	E3 ubiquitin-protein ligase BRE1B	AX-181928576	37.4	0		P							
	AC18	XP_023862569.1	E3 ubiquitin-protein ligase BRE1B isoform X2	AX-181935230	13.0	0			P,F	P,F	P,F		P		P
	AC18	XP_023862569.1	E3 ubiquitin-protein ligase BRE1B isoform X2	AX-181935231	13.0	0			P,F	P,F	P,F		P		
ephrin type-A receptor 3	AC02	XP_023863833.1	ephrin type-A receptor 3	AX-181971250	32.1	0						P			
	AC02	XP_023863833.1	ephrin type-A receptor 3	AX-181971249	32.2	0						P			
	AC02	XP_023863833.1	ephrin type-A receptor 3	AX-181971248	32.2	0		P				P			
	AC02	XP_023863833.1	ephrin type-A receptor 3	AX-181944140	32.5	0							P		
	NW_019942998.1	XP_023994522.1	LOW QUALITY PROTEIN: ephrin type-A receptor 3-like	AX-181991583	0.2	0								P,F	P,F
extended synaptotagmin-1	AC01*	XP_023846266.1	extended synaptotagmin-1 isoform X1	AX-181925736	33.7	0	P	P					P,F		
	AC01*	XP_023846266.1	extended synaptotagmin-1 isoform X1	AX-181930078	33.7	0	P	P					P		
	AC11*	XP_023852472.2	extended synaptotagmin-1	AX-181939957	30.7	0			P	P,F	P,F	P,F	P	P,F	
gastrula zinc finger protein XlCGF57.1	AC06.1	XP_024002901.1	gastrula zinc finger protein XlCGF57.1 isoform X4	AX-181942043	1.5	0.4		P		P,F	P,F	P,F			P
	AC14	XP_023856467.1	gastrula zinc finger protein XlCGF57.1-like	AX-181937152	3.7	0						P	P		
	AC14	XP_023856467.1	gastrula zinc finger protein XlCGF57.1-like	AX-181937150	3.7	0							P		
homeobox protein MSX-2	AC23	XP_023824809.1	homeobox protein MSX-2	AX-181967650	3.2	0	P	P			P		P		
	NW_019943350.1	XP_023995838.1	homeobox protein MSX-2-like isoform X1	AX-181920182	0.1	-0.5			P,F	F					

neurexin-3a	AC04q.2	XP_023842075.1	neurexin-3a-like	AX-181947935	28.8	0								P,F		
	AC04q.2	XP_023842075.1	neurexin-3a-like	AX-181915470	28.9	0	P		P		F	P	P			
	AC04q.2	XP_023842075.1	neurexin-3a-like	AX-181937420	29.2	0	P	P								
	NW_019942794.1	XP_023993341.1	neurexin-3a-like	AX-181941664	0.1	-2.6							P			
neuronal PAS domain-containing protein 3	AC04q.2	XP_023841991.1	LOW QUALITY PROTEIN: neuronal PAS domain-containing protein 3-like	AX-181915280	20.9	0								P,F	P	
	AC05	XP_023844009.1	neuronal PAS domain-containing protein 3	AX-181952675	32.7	-0.2	P	P				P,F	P			
PAN2-PAN3 deadenylation complex catalytic subunit PAN2	AC07*	XP_023847330.1	PAN2-PAN3 deadenylation complex catalytic subunit PAN2	AX-181977426	25.0	0						P				
	AC07*	XP_023847330.1	PAN2-PAN3 deadenylation complex catalytic subunit PAN2	AX-181935260	25.0	0						P				
	AC17*	XP_023861482.1	PAN2-PAN3 deadenylation complex catalytic subunit PAN2 isoform X1	AX-181952107	22.7	-2.6				P,F	P,F	P,F	P,F	P	P,F	
parafibromin	AC19*	XP_023865178.1	parafibromin	AX-181969955	34.1	0	P	P	P,F	P	F		P			
	AC32*	XP_023833739.1	parafibromin	AX-181942618	7.9	0			P							
	AC32*	XP_023833739.1	parafibromin	AX-181934703	7.9	0			P,F			P				
piezo-type mechanosensitive ion channel component 2	AC27	XP_023829138.1	piezo-type mechanosensitive ion channel component 2-like	AX-181930959	6.9	0							P			
	NW_019944187.1	XP_023997787.1	LOW QUALITY PROTEIN: piezo-type mechanosensitive ion channel component 2-like	AX-181973456	0.1	0			P	P,F		P		P		
pro-neuregulin-3, membrane-bound	AC17	XP_023860894.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181941701	38.6	0						P				
	AC17	XP_023860894.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181922212	38.7	0						P				
	AC17	XP_023860894.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181942722	38.7	0			P							
	AC17	XP_023860894.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181913589	38.7	0						P				
	AC17	XP_023860894.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181974022	38.8	0						P				
	AC17	XP_023860894.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181914524	39.0	0	P									
	AC18*	XP_023864068.1	pro-neuregulin-3, membrane-bound isoform	AX-181914283	21.2	0						P	P			
	AC18*	XP_023864068.1	pro-neuregulin-3, membrane-bound isoform	AX-181942151	21.5	0		P				P	P			
	AC18*	XP_023864284.1	pro-neuregulin-3, membrane-bound isoform-like	AX-181943495	44.0	0	P									
	AC25*	XP_023825564.1	pro-neuregulin-3, membrane-bound isoform isoform X2	AX-181942348	20.7	0								P		
protocadherin-11 X-linked	AC08	XP_023847824.1	LOW QUALITY PROTEIN: protocadherin-11 X-linked-like	AX-181937960	0.8	0	P	P				P	P	P	P,F	
	NW_019942572.1	XP_023991446.1	protocadherin-11 X-linked-like	AX-181957245	0.8	-4.4	P	P	P,F			P				
sialic acid-binding Ig	AC06.1	XP_023844980.1	sialic acid-binding Ig-like lectin 5	AX-181937410	15.0	0	P	P					P,F	P,F	P,F	
	AC35	XP_023836226.1	sialic acid-binding Ig-like lectin 5	AX-181935420	11.7	0						P				
type I inositol 1,4,5-trisphosphate 5-phosphatase	AC18*	XP_023863329.1	type I inositol 1,4,5-trisphosphate 5-phosphatase isoform X2	AX-181926142	29.5	0	P	P			P,F	P				
	AC18*	XP_023863329.1	type I inositol 1,4,5-trisphosphate 5-phosphatase isoform X2	AX-181948291	29.5	0							P		P,F	
	AC25*	XP_023825797.1	type I inositol 1,4,5-trisphosphate 5-phosphatase-like isoform X2	AX-181939239	7.7	0						P		P		

Table S11 The number of outlier loci which were detected by a each outlier detection method between small and big morphs within each of WP-L and LO-L.

Outlier Detection Method	small morph WP-L vs. big morph WP-L	small morph LO-L vs. big morph LO-L
Both PCAdapt and FST	57	269
PCAdapt Only	4	87
FST Only	47	44
Total	108	400

Table S12 The number of outlier and non-outlier SNPs detected between morphs (small and big) within a given lake that were found to be polymorphic outliers, polymorphic non-outliers or non-polymorphic in each of WP-L and LO-L.

	WP-L (N = 6404)		LO-L (N = 16702)	
	Outlier (N = 108)	Non-outlier (N = 6296)	Outlier (N = 400)	Non-outlier (N = 16302)
WP-L Polymorphic Outlier	108	0	1	95
WP-L Polymorphic Non-outlier	0	6296*	135	5952
WP-L Non-polymorphic	0	0	264	10255
LO-L Polymorphic Outlier	1	135	400	0
LO-L Polymorphic Non-outlier	95	5952	0	16302
LO-L Non-polymorphic	12	209	0	0

*Note that 24 loci within WP-L were polymorphic when including all samples but were monomorphic after removing the six putative hybrid individuals before outlier analyses. These monomorphic loci were still retained when performing outlier analyses between small and big morphs within WP-L, but were (unsurprisingly) not identified as outliers.

Table S13 Genes containing outlier loci differentiating sympatric small and big morphs within both WP-L and LO-L. The method by which each SNP was identified as an outlier is denoted for each comparison between morphs within each lake (P – PCAdapt, F- F_{ST}).

Protein Name	Linkage Group	Protein Code	SNP Code	Position (Mbp)	Position Relative to CDS (kbp)	Method by which outlier was detected: small morph WP-L vs. big morph WP-L	Method by which outlier was detected: small morph LO-L vs. big morph LO-L
VPS10 domain-containing receptor SorCS2	AC37	XP_023837960.1	AX-181980220	14.9	0	P,F	F
	AC37	XP_023837960.1	AX-181940385	14.9	0	F	
	AC37	XP_023837960.1	AX-181940384	14.9	0	F	

Table S14 SNPs detected as outliers for five or more of seven paired landlocked and anadromous populations (1) WP-L vs. SWA-A (either small morph WP-L vs. SWA-A or big morph WP-L vs. SWA-A), 2) HEB-L vs. IKA-A, 3) BS-L vs. K05-A, 4) LO-L vs. K05-A (either small morph LO-L vs. K05-A or big morph LO-L vs.K05-A), 5) KNU-L vs. ANA-A, 6) SLU-L vs. REI-A, 7) GB-L vs. REI-A) and at least one location with sympatric small and big morphs from Salisbury et al. (2020).

						Landlocked vs. Anadromous									Small vs. Big		
General Protein Name	Linkage Group	Protein Code	SNP Code	Position (Mbp)	Position Relative to CDS (kbp)	small morph WP-L vs. SWA-A	big morph WP-L vs. SWA-A	HEB-L vs. IKA-A	BS-L vs. K05-A	small morph LO-L vs. K05-A	big morph LO-L vs. K05-A	KNU-L vs. ANA-A	SLU-L vs. REI-A	GB-L vs. REI-A	small morph R vs. big morph R	small morph B vs. big morph B	small morph E vs. big morph E
NA	AC13	NA	AX-181979449	27.2	NA	*	*		*	*	*	*		*	*	*	
partner of Y14 and mago B	AC17	XP_023861958.1	AX-181916308	22.6	0.1				*	*	*	*	*	*			*
PAN2-PAN3 deadenylation complex catalytic subunit PAN2 isoform X1	AC17	XP_023861482.1	AX-181952107	22.7	-2.6				*	*	*	*	*	*	*		
nuclear envelope integral membrane protein 1-like isoform X1	AC17	XP_023862374.1	AX-181967022	22.7	0.0				*	*	*	*	*	*	*		
LOW QUALITY PROTEIN: serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit C-like	AC17	XP_023862177.1	AX-181980622	22.9	2.7				*	*	*	*	*	*	*		
			AX-181983398	22.9	2.7				*	*	*	*	*	*	*		
inactive dipeptidyl peptidase 10	AC17	XP_023860785.1	AX-182162437	22.9	0.0				*	*	*	*	*	*	*		