Extensive genetic differentiation between recently evolved sympatric Arctic charr morphs

Jóhannes Guðbrandsson[1,2]*, Kalina H. Kapralova[1], Sigríður R. Franzdóttir[1,3],
Þóra Margrét Bergsveinsdóttir[1], Völundur Hafstað[1], Zophonías O. Jónsson[1,3],
Sigurður S. Snorrason[1] and Arnar Pálsson[1,3]

[1] Institute of Life and Environmental Sciences, University of Iceland, 101 Reykjavik, Iceland
[2] Marine and Freshwater Research Institute, 101 Reykjavik, Iceland
[3] Biomedical Center, University of Iceland, 101 Reykjavik, Iceland

Abstract

The availability of diverse ecological niches can promote adaptation of trophic specializations and related traits, as has been repeatedly observed in evolutionary radiations of freshwater fish. The role of genetics, environment and history in ecologically driven divergence and adaptation, can be studied on adaptive radiations or populations showing ecological polymorphism. Salmonids, especially the *Salvelinus* genus, are renowned for both phenotypic diversity and polymorphism. Arctic charr (*Salvelinus alpinus*) invaded Icelandic streams during the glacial retreat (about 10,000 years ago) and exhibits many instances of sympatric polymorphism. Particularly well studied are the four morphs in Lake Þingvallavatn in Iceland. The small benthic (SB), large benthic (LB), planktivorous (PL) and piscivorous (PI) charr differ in many regards, including size, form and life history traits. To investigate relatedness and genomic differentiation between morphs we identified variable sites from RNA-sequencing data from three of those morphs, and verified 22 variants in population samples. The data reveal genetic differences between the morphs, with the two benthic morphs being more similar and the PL-charr more genetically different. The markers with high differentiation map to all linkage groups, suggesting ancient and pervasive genetic separation of these three morphs. Furthermore, GO analyses suggest differences in collagen metabolism, odontogenesis and sensory systems between PL-charr and the benthic morphs. Genotyping in population samples from all four morphs confirms the genetic separation and indicates that the PI-charr are less genetically distinct than the other three morphs. The genetic separation of the other three morphs indicates certain degree of reproductive isolation. The extent of gene flow between the morphs and the nature of reproductive barriers between them remain to be elucidated.

1 Introduction

Organismal diversity reflects the process of evolution and highlights the importance of natural selection in building and maintaining adaptations (Darwin 1859). While purifying selection preserves adaptations and biological functions, positive selection alters phenotypic traits and frequencies of genetic variations influencing them (Via 2001, 2009). Feeding is a primary function in all animals and in nature, we see many examples of spectacular adaptive radiation where natural selection has generated a range of adaptations in specific feeding structures and foraging behaviour (Losos and Ricklefs 2009; Seehausen and Wagner 2014). Recently, studies have demonstrated how positive selection can shift phenotypic distributions on generational time scales (Grant and Grant 2002; Nosil 2012). The recurrent ecological specializations, like for feeding (Seehausen and Wagner 2014), found in many species and populations of freshwater fish, represent interesting study systems and set of functional phenotypes for analyses of evolutionary change, convergence and parallel adaptation (Muschick et al. 2012).

1.1 Genome wide divergence or islands of differentiation?

The genomics revolution has spawned powerful tools for studying relatedness of populations and species, the connections between genetic and phenotypic variation, e.g. how variations in the form of feeding structures are

---

*Corresponding author – Tel: +354-575-2615 – johannes.gudbrandsson@hafogvatn.is

caused by differential expression of key developmental genes (Abzhanov et al. 2006; Guðbrandsson et al. 2018), and to determine to what extent these differences are due to genetic differences (and how they are distributed in genomes) (Wolf and Ellegren 2016). These methods have enabled studies evaluating feedback between the organism and its environment, the role of plasticity in generating functional variation and possibly promoting adaptation (Morris and Rogers 2014; Abouheif et al. 2014). Central to this study, genomic methods can reveal differences between populations and ecotypes, and highlight genes and pathways that may be under positive selection in natural populations (Malinsky et al. 2015; Wolf and Ellegren 2016).

The differentiation and divergence between related groups (populations, sub-species or species) are influenced by genomic parameters, population history and evolutionary forces (Seehausen et al. 2014; Vijay et al. 2017). The genetic separation may be rather uniform over the entire genome or localized to "genomic islands of differentiation", depending on multiple factors, e.g. the strength of selection, the nature of the adaptive traits, genomic structure, population genetic parameters and history (Malinsky et al. 2015; Wolf and Ellegren 2016). Studies of closely related groups can detect the genetic correlates of adaptive traits (Pease et al. 2016), but this is not straightforward for a number of reasons. Selection operates on multiple genes influencing all aspects of fitness, at the same time, for instance standing variation and recent variants (Vijay et al. 2017), alleles derived from introgression, even possibly balanced ancestral polymorphism (Pardo-Diaz et al. 2012; Guerrero and Hahn 2017). Furthermore, recently separated groups adapting to different environments can be viewed as being scattered along a "speciation continuum" (Theis et al. 2014; Seehausen et al. 2014). Under this view, changes in environmental or population genetic parameters can move a population along this continuum (Grant and Grant 2008). Specifically, such changes can lead to the merging of previously distinct populations, speed up their divergence or produce reproductively isolated species (Hendry et al. 2009; Seehausen et al. 2014; Lowry and Gould 2016). We are keen to explore the genetic differentiation of recently evolved populations within a species, that coexist in the same lake, with the aim to study the genetics of ecological specializations.

1.2 Resource polymorphism along a benthic - limnetic axis

It is widely acknowledged that varying resource utilization of allopatric populations can lead to divergence of traits which, given time, can result in adaptive divergence among these populations. However, discrete variation in resource use among individuals in the same area, e.g. foraging in different habitats, can also generate divergent traits within populations (Skúlason and Smith 1995; Bernatchez et al. 2010). The occurrence of discrete phenotypes (morphs) of a species living in sympatry and diverging in traits that relate to the utilization of different resources (food, breeding grounds, etc) has been called resource polymorphism. Typically, diverging traits of resource morphs involve morphology, behaviour and/or life history characteristics (Skúlason and Smith 1995; Smith and Skúlason 1996). Resource polymorphism can arise through developmental plasticity within a homogenic population, by natural selection on genetically encoded traits or a combination of these mechanisms. For instance, the broad and narrow-headed morphs of the European eel (*Anguilla anguilla*) were found to be determined to a large extent by diet (De Meyer et al. 2016) and ecomorphs of killer whales (*Orcinus orca*) were hypothesized to have originated through plastic responses of a small founder population (Foote et al. 2016). Examples of genetically determined resource polymorphism can be found in crater lake cichlid fishes, e.g. were a single locus effects jaw and body shape (Fruciano et al. 2016). Likewise, in benthic and limnetic "species" of threespine stickleback (*Gasterosteus aculeatus*), the same loci show signs of differentiation (Jones et al. 2012).

Salmonids are renowned for their phenotypic diversity, both among and within populations, with multiple examples of resource polymorphism. The most phenotypically diverse and polymorphic species seem to be in the *Salvelinus* genus, including Arctic charr *Salvelinus alpinus* (Klemetsen 2013), Lake charr (also called Lake trout) *Salvelinus namaycush* (Muir et al. 2016) and Dolly Varden charr (*Salvelinus malma*) (also called Dolly Varden trout) with as many as seven morphs found in the same lake (Markevich et al. 2018). Arctic charr colonized lakes and rivers on the northern hemisphere after the last glaciation period (approx 9,000-12,000 years ago) (Snorrason and Skúlason 2004; Noakes 2008; Klemetsen 2010). In Iceland, multiple lakes harbour polymorphic Arctic charr (Skúlason et al. 1992; Snorrason and Skúlason 2004; Wilson et al. 2004; Woods et al. 2012b) and a unique small benthic morphotype is found in many streams and ponds across the country, especially in cold springs with a lava-rock bottom in the geologically younger parts of the island (Kapralova et al. 2011; Kristjánsson et al. 2012). One of four sympatric charr morphs found in Lake Þingvallavatn, Iceland's largest lake, is of this type. Population genetics show that populations of Arctic charr in Iceland group by geography, not morphotype (Gíslason 1998; Kapralova et al. 2011), supporting the notion that small benthic morphs and other derived forms have evolved independently in different locations. For instance, the four Lake Þingvallavatn morphs are more closely related to one another than to other charr populations (Kapralova et al. 2011). Genetic analysis also indicates closely related

2

sympatric morphs in Norway, Scotland and Transbaikalia (Hindar et al. 1986; Gordeeva et al. 2015; Jacobs et al. 2018), while in other cases sympatric charr morphs seem to have emerged by more than one invasion (Verspoor et al. 2010).

1.3   The four sympatric charr morphs in Lake Þingvallavatn

Lake Þingvallavatn formed in a rift zone as the Icelandic ice-cap receded ∼10,000 years ago and was shaped by volcanic activity and the isostatic rebound. The lake has been influenced by tectonic movements, causing extensive subsidence and horizontal extension with extensive rift-forming in the central graben (Saemundsson 1992). The lake harbours four distinct Arctic charr morphs: Small benthic (SB), Large benthic (LB), Planktivorous (PL) and Piscivorous (PI) charr, that differ along a benthic - limnetic axis and this is reflected in their form, size, habitat use, diet, life history characteristics and parasite loads (Sandlund et al. 1987; Jonsson et al. 1988; Malmquist et al. 1992; Frandsen et al. 1989; Kapralova et al. 2013) and represent genuine resource polymorphism (Skúlason and Smith 1995) (Figure 1). The morphs also differ extensively in spawning times (Skúlason, Snorrason, Noakes, Ferguson and Malmquist 1989). Common garden experiments on embryos/offspring indicated heritable differences in various traits between morphs, e.g. morphology, sexual maturation rates, foraging behaviour (Skúlason, Noakes and Snorrason 1989; Skúlason et al. 1993, 1996; Kapralova et al. 2015) but plasticity was also found to be significant (Parsons et al. 2010, 2011).
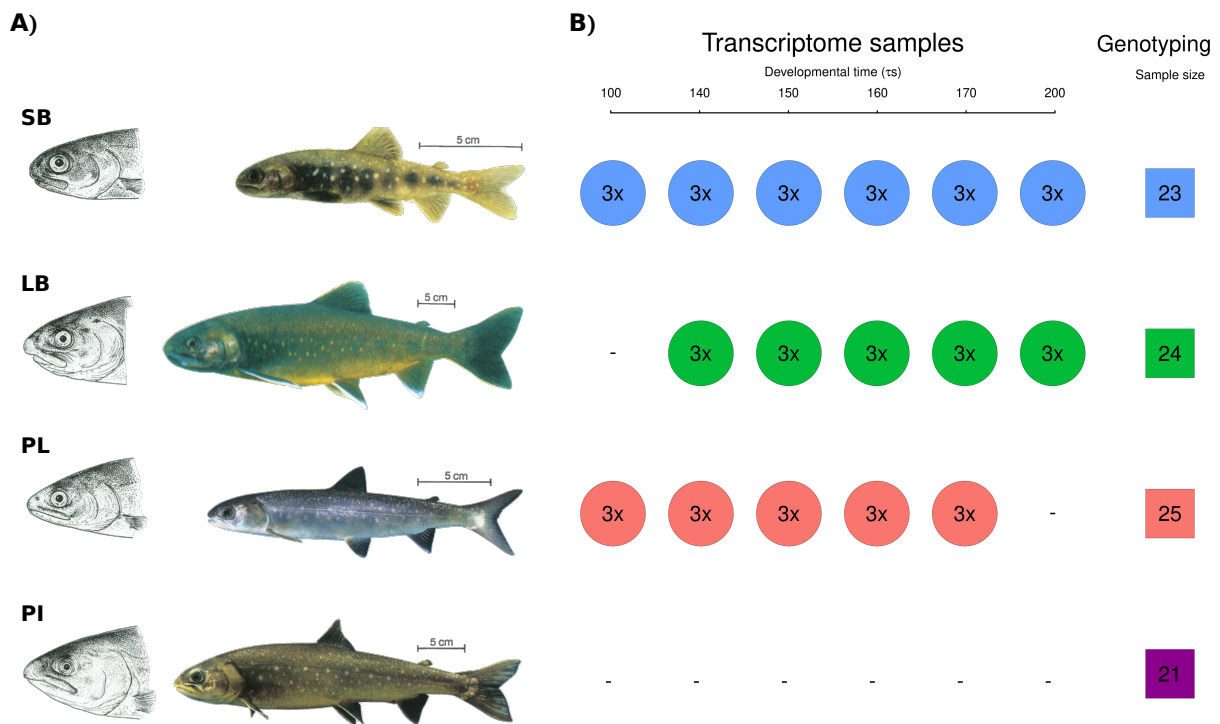

Figure 1: The phenotypically distinct sympatric Arctic charr from Lake Þingvallavatn and the sampling strategy. **A)** The four sympatric morphs are Small benthic (SB), Large benthic (LB), Planktivorous (PL) and Piscivorous (PI) charr. They differ in size (size bars = 5 cm), head and feeding morphology and pigmentation. Adapted from Sandlund et al. (1992) ©Wiley-Blackwell, drawings by Eggert Pétursson. **B)** Sampling of charr for transcriptome sequencing (circles) and genotyping (squares) of population samples. The top 3 morphs were mined for genetic variation in the transcriptome and population samples were studied from all four, to confirm genetic variants. The transcriptome samples came from embryos at 6 developmental stages prior to hatching, from $100\tau s$ to $200\tau s$, in the three morphs (circles)(Guðbrandsson et al. 2018). The sampling of each morph and developmental timepoint combination was replicated three times (biological replicates), each sample is a pool of mRNA from three embryos. Six timepoints were sampled of SB-charr, and five of LB- and PL-charr embryos. The population samples (squares) were obtained by gill netting on the spawning grounds, see methods. The morph colouring scheme (SB: blue, LB: green, PL: red and PI: purple) will be retained throughout the manuscript.

The earliest population genetic studies of the Lake Þingvallavatn charr found variation in some markers, but like in the RFLP analyses on mtDNA, generally weak if significant separation of morphs (Magnusson and Ferguson 1987; Danzmann et al. 1991; Volpe and Ferguson 1996). The estimated relationship of the morphs varied by studies (and markers), SB-charr (Magnusson and Ferguson 1987; Gíslason 1998), LB-charr (Kapralova 2008) and PL-charr (Volpe and Ferguson 1996) have all been estimated as the outgroup. PI-charr, on the other hand, tended clustered with both PL- and LB-charr. A study of 9 microsatellite markers in SB- and PL-charr from five spawning sites and LB-charr revealed significant but weak morph difference (average $F_{ST} = 0.039$) but did not resolve their relationship (Kapralova et al. 2011). Importantly, coalescence simulations on those data best supported a short initial phase of PL- and SB-charr allopatry followed by coexistence, rather than a pure sympatric origin of these morphs (Kapralova et al. 2011). More recent studies found loci with strong genetic differences ($F_{ST} = 0.13$ to 0.67) in two immunological genes (Kapralova et al. 2013) and rRNA genes within the mtDNA between morphs (Gudbrandsson et al. 2016).

Recent studies have studied the developmental and molecular correlates of morph divergence in the lake, by analyzing the morphology of embryos, miRNA and mRNA expression (Kapralova et al. 2014; Ahi et al. 2014; Kapralova et al. 2015; Gudbrandsson et al. 2016; Guðbrandsson et al. 2018). Most recently, RNA-sequencing of embryos of three of the four morphs (excluding PI-charr) reared in a common garden, revealed about 2000 transcripts with expression differences between morphs (Guðbrandsson et al. 2018). The PL-, LB- and SB-charr differed at the transcriptome level and the results suggested a closer relationship between the two benthic morphs (LB- and SB-charr), with PL-charr more distinct. Also, qPCR approaches have revealed differential expression between the morphs in the lake in genes related to extracellular matrix organization and skeletogenesis in early development (Ahi et al. 2013, 2014, 2015) and genes in the *mTOR*-pathway in muscles of SB-charr in Iceland (Macqueen et al. 2011).

Relatively few genomic resources have been available for Arctic charr, until recently when a reference genome was published (Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop 2018). Reference genomes for other salmonid species have been made available in recent years, such as for Rainbow trout (*Oncorhynchus mykiss*, Berthelot et al. 2014; Pearse et al. 2018), Atlantic salmon (*Salmo salar*, Lien et al. 2016), Grayling (*Thymallus thymallus*, Varadharajan et al. 2018; Sävilammi et al. 2019), Chinhook salmon (*Oncorhynchus tshawytscha*, Christensen, Leong, Sakhrani, Biagi, Minkley, Withler, Rondeau, Koop and Devlin 2018) and one more (Coho salmon, *Oncorhynchus kisutch*) is available on NCBI (Genebank assembly GCA_002021735.1). About 88–103 million years ago an ancestor of Salmonids underwent a whole genome duplication (Ss4R), the fourth on the vertebrate lineage (Allendorf and Thorgaard 1984; Macqueen and Johnston 2014; Berthelot et al. 2014). Comparisons of several salmonids found significant synteny in their linkage groups (Danzmann et al. 2005; Sutherland et al. 2016), despite rearrangements (Timusk et al. 2011; Nugent et al. 2017). Thus a substantial fraction of genes in salmonids are paralogs in large syntenic regions (Nugent et al. 2017; Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop 2018).

Here we mined the developmental transcriptome (Guðbrandsson et al. 2018) of three of the morphs (LB-, SB- and PL-charr) to test for genetic differences between them, elucidate their evolutionary relationship and identify loci and developmental pathways that may relate to morph differentiation. The data were used to evaluate three hypotheses about the causes of morph separation:

I Because the salmonid's homing behaviour leads offspring to spawn in the same location as their parents, heterogeneity in spawning places and micro-environments can lead to environmentally induced gene expression and phenotypic differences. In this scenario, morphs are environmentally induced, and genetic differences between them minor.

II Ecological selection on specific traits has led to genetic separation of morphs, seen as allele frequency differences at variants in key genes related to fitness traits, in the face of gene flow between morphs (Seehausen et al. 2014; Wolf and Ellegren 2016). Under this model, genomic islands of differentiation are expected, with limited genetic differences between morphs in the rest of the genome.

III Pre-zygotic barriers such as spatial and temporal separation in spawning between morphs (Skúlason, Snorrason, Noakes, Ferguson and Malmquist 1989) or behavioural/mating differences have kept the morphs reproductively isolated for some time (as has been indicated for SB- and PL-charr (Kapralova et al. 2011)). Under this scenario, modest genetic separation among the morphs is expected, on all linkage groups.

These three hypotheses are not mutually exclusive, and some may apply to specific pairs of morphs. The data enabled analyses of the genetic separation of three of the sympatric morphs, the origin of the rare piscivorous charr,

the genomic patterns of differentiation between morphs and the genes and molecular systems that may associate with their specializations.

2 Materials and Methods

2.1 Sampling for transcriptome and population genetics

The fishing of parents for crosses from Lake Þingvallavatn and sampling of embryos for developmental transcriptome analysis is described in Guðbrandsson et al. (2018). Briefly, we caught SB-, LB- and PL-charr on the spawning grounds, ready to release eggs and milt, made crosses and reared embryos from each cross separately in a common garden at the Hólar Aquaculture station, Verið (Sauðárkrókur). Samples of embryos were taken at 5-6 developmental timepoints depending on morph (see Figure 1). The relative age in τs-units was defined be Gorodilov (1996) and is a non-linear function of time and temperature (see formula in Guðbrandsson et al. 2018). For each morph and developmental timepoint three biological replicates were analyzed, each of those isolated from pools of 3 embryos to a total of 48 samples (Figure 1). The crosses were made using multiple parents from the same morph. For most samples, the embryos came from crosses created in the 2010 spawning season (SB 150-200τs, PL 140-170 τs, LB 140-200τs). For SB- and PL-charr gametes from ten males and ten females were mixed for the cross but five parents of each sex for LB. Due to poor RNA quality of some samples, we had to add samples collected in 2011. For PL timepoint 100τs we used a cross from the 2011 spawning season with a pool of gametes, from 10 males and 10 females. For SB timepoints 100τs and 140τs we had to rely on two single pair crosses (one male and one female) due to difficulties of finding running SB-charr. Samples SB100A and SB100B came from one cross (referred to as SB-2011_A) and SB100C and the three samples for timepoint 140τs came from another (cross SB-2011_B).

For verification of single nucleotide polymorphism's (SNP's) and comparisons between morphs we sampled sexually mature individuals at the spawning grounds of the three morphs; 23 SB-, 24 LB- and 25 PL-charr, by gillnet fishing in Ólafsdrattur and Mjóanes in 2015. To gain insight the genetic status of the PI-charr we genotype 21 individuals which have been collect as by-catch in other surveys in the period 2010–2016. They could not be collected in a single season as PI-charr is the rarest morph and spawning grounds not as well defined as for the other morphs.

As before (Guðbrandsson et al. 2018), all fishing in Lake Þingvallavatn was with permissions obtained from the farmers in Mjóanes and from the Þingvellir National Park Commission. Ethics committee approval is not needed for regular or scientific fishing in Iceland (The Icelandic law on Animal protection, Law 15/1994, last updated with Law 55/2013).

2.2 RNA isolation and sequencing

Total mRNA was isolated from a pool of three embryos for each sample. The RNA was quality checked and used as a template for 48 cDNA libraries subjected to Illumina sequencing. The sequencing was performed on Hiseq2000 at deCODE genetics (Reykjavík, Iceland) yielding 1,188 million 101 paired-end reads. The sequencing reads from the 48 samples were deposited into the NCBI SRA archive under BioProject identifier PRJNA391695 and with accession numbers: SRS2316381 to SRS2316428. For a more detailed description about RNA isolation, quality checks and library construction see Guðbrandsson et al. (2018).

2.3 Transcriptome assembly, Abundance estimation and Annotation

The transcriptome assembly and annotation were described previously (Guðbrandsson et al. 2018), but briefly, the reads were quality trimmed and adapters removed using Trim Galore! (version 0.3.3, Krueger 2012). The filtered reads from all samples were assembled with Trinity (version v2.1.0, Grabherr et al. 2011). We used kallisto (version v0.42.4, Bray et al. 2016) to estimate transcripts abundance. To speed up the annotation process transcripts with fewer than 200 mapped reads were not retained for annotation, another reason being that lowly expressed transcripts are unlikely to supply variants with enough coverage to surpass quality filters. The transcripts were annotated with the Trinotate pipeline (version 2.0.2, Haas 2015). Orthologs of the transcripts in salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) mRNA and protein sequences were found using `blastn` and `blastx` respectively (Altschul et al. 1990). Annotation from NCBI *Salmo salar* Annotation Release 100 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Salmo_salar/

5

100/, retrieved 2015-12-17) and SalmoBase (Samy et al. 2017, http://salmobase.org/, version from 2015-09-18) were both used for salmon and annotation for rainbow trout came from Berthelot et al. (2014, http://www.genoscope.cns.fr/trout/data/, version from 2014-05-19). Only the best match was retained for each reference database. For further details about the annotation process and parameters, we refer to Guðbrandsson et al. (2018).

2.4   Variant calling and quality filtering

To identify genetic variation in the transcriptome we mapped the quality-trimmed reads to the complete Trinity assembly (annotated and non-annotated transcripts). Pseudobam files generated by kallisto were supplied to eXpress (version 1.5.1 Roberts and Pachter 2012) to get single alignment for multi-mapping reads. eXpress uses expectation–maximization (EM) algorithm to get posterior probabilities for read placements and assigns reads to transcripts by the posterior probability. We used the default eXpress parameters except we set the batch option to 10 to get more EM-rounds and better assignment of reads. Reads with more than 10 mismatches were removed using samtools (version 1.1, Li et al. 2009) and bamtools (version 2.3.0, Barnett et al. 2011).

Candidate variants (hereafter variants) were called with FreeBayes (version v1.0.1-2-g0cb2697, Garrison and Marth 2012) on all the samples simultaneously. The coverage threshold was set to 480 reads and the threshold for a variant allele to be called to 48 reads. Due to our experimental setup of pooled sampling of three individuals we added the options `--pooled-discrete` and `--ploidy 6`. As we were sequencing RNA and to reduce run time other options used for FreeBayes were; `--use-duplicate-reads` and `--use-best-n-alleles 4`. Only bi-allelic variants were processed further. Variants were filtered based on coverage and allele frequency. We only consider positions with a minimum coverage of 10 reads in at least 30 samples, thereof the minimum of eight samples in each morph. Minimum allele frequency of the alternative allele was set to 10% and the maximum to 70%.

Due to the genome duplication in the salmonid ancestor (Allendorf and Thorgaard 1984; Moghadam et al. 2011) some of the candidate variants might reflect sequence divergence of paralogous genes rather than true segregating variants. We removed such candidates by identifying variants with negligible differences in allele frequency between samples (and thus individuals). Recall, each sample represents 3 individuals, with 6 copies of each chromosome. Thus, true biallelic segregating markers only have seven possible genotype combinations (0:6, 1:5, 2:4, 3:3, etc). This should result in allele frequency differences among samples, as it is highly unlikely that the same combination of genotypes will be found in all cases. In contrast, candidate variants due to fixed differences in paralogous genes will have similar allele frequency in all samples. To estimate this we calculated hierarchical F-statistics (Hartl and Clark 2006, p. 280-283) in R (R Core Team 2015). First, we estimated the variation between samples, with a statistic termed F_{PT}; were P stands for pool (each sample is a pool of three individuals) and T the total population. We also calculated a traditional F_{ST}, to capture variation in allele frequencies between morphs (subpopulations). Note, because of the sampling of related individuals the values of F_{ST} are not comparable to other studies (see results).

The F_{PT} statistic is analogous to the classical F_{IT}, so low F_{PT} values indicate a low variation between samples just as low F_{IT} values indicate low variation among individuals (excess heterozygosity). Filtering on F_{PT} is therefore similar to removing markers that deviate from Hardy-Weinberg equilibrium when genotypes are on an individual basis. To determine a reasonable cutoff for F_{PT} we simulated 288 chromosomes with eight different alternative allele frequencies (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7). Chromosomes were distributed randomly among samples with the same experimental setup as in our study and hierarchical F-statistics were calculated. For each alternative allele frequency, 10,000 replicates were simulated. Based on this simulation (see figure S1) we chose $F_{PT} = 0.1$ as a cutoff and all variants below this value were removed. Samples with coverage of five reads or less were not included in the calculations of F-statistics.

We ran FreeBayes again on the filtered variants in order to phase fully coupled variants into longer haplotypes. We used a haplotype length window of 70 basepairs (bp, option `--haplotype-length 70`) and only asked for biallelic variants (`--use-best-n-allels 2`). Nevertheless, nine variants in the output were tri-allelic and therefore removed. We also removed variants if the new estimate of F_{PT} was the below the cutoff (0.1) in the new run. No further filtering of variants was performed.

2.5   Analysis of variant functionality and distribution among morphs

Open reading frame prediction with Transdecoder (Haas and Papanicolaou 2015) in the Trinotate pipeline (see above) was used to estimate the position (UTR, CDS) of the variants and determine if they affected the protein

sequence. Further statistical analysis was performed in the R environment (R Core Team 2015) using the VariantAnnotation package (Obenchain et al. 2014) to handle the variant data. The alternative allele frequency (based on read counts) for each sample was used for further analysis of genetic separation between groups. To study the distribution of variation among samples we did principal component analysis (PCA) with the built in prcomp-function in R. Missing values in the dataset were populated with the mean alternative allele frequency for the variant (21087 out of 1848192 values were missing or 1.14%), prior to the PCA analyses. We calculated the mean allele frequency of each variant for each morph, and also the deviation from the other two as the sum of the allele frequency difference between them. For example the deviation for PL is $d_{PL} = d(PL, SB) + d(PL, LB)$ where $d(PL, SB)$ is the difference in mean allele frequency for PL- and SB-charr. To screen for morph separation, we calculated F_{ST} comparing allele frequency by morphs (see above). Variants with F_{ST} above 0.2 were analyzed for gene ontology (GO) enrichment. The GO tests were performed with the goseq-package in R (Young et al. 2010), separately for variants with the largest deviation of mean allele frequency for each morph. The goseq-package accounts for transcript length as more variants are expected in longer transcripts. All transcripts with variants (8,961 transcripts) were used as the reference set for GO-enrichment tests. We also ran GO tests on the gene level using annotation to Salmobase. Gene length was not taken into account in that case and the reference set consisted of 7317 genes. GO categories were also mapped to their ancestor using the GO.db-package in R (Carlson 2015). We only tested biological processes, omitting categories closer than 3 steps from the root of the annotation tree. The gene ontology annotation was based on SalmoBase, see above and Guðbrandsson et al. (2018). Categories with false discovery rate (Benjamini and Hochberg 1995) below 0.01 were considered significant. We used 0.01 instead of the classic 0.05 as six tests were conducted (each morph on transcript and gene level). We cataloged variants private to one morph, the criterion being less than 1% frequency of the alternative allele in the other morphs. Morph specific lists of private alleles were analyzed for GO enrichment the same way.

2.6 Genomic distribution of candidate variants

With the Arctic charr reference genome (Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop 2018, Assembly `GCA_002910315.2 ASM291031v2`) it became possible to assign variants to linkage groups. A sequence 200 bp upstream and downstream of each variant in the transcriptome contigs were mapped to the genome with `blastn` within R (Hahsler and Nagar 2017) using: `-max_target_seqs 2 -max_hsps 12 -culling_limit 2`. In the case of more than one blast hit for each sub-sequence, hits with the highest bit score that included the variant was chosen. If no hit included the variant the hit with highest bit score was used. If more than one hit was equally likely, hits to chromosomes were priorities to hits to contigs or scaffolds. If equally likely hits mapped to the same chromosome or scaffold in the genome and were within 50kb from each other the hit with the lower position was chosen. The remaining variants were left unplaced.

2.7 Genotyping of candidate variants in a population sample

We genotyped 93 adult fish from the four morphs, PI- (21) PL- (24), LB- (24) and SB-charr (24). DNA was extracted with phenol chloroform according to standard protocols. Markers for genotyping were chosen based on high F_{ST} values, predicted biological functions of the genes they affected and for some to evaluate LD between linked markers. To design primers for genotyping we aligned regions around candidate variants, contigs from the assembly and similar regions in *O. mykiss* and *S. salar* (retrieved by `blast`). Locations where updated when the *S. alpinus* genome became available (Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop 2018). Markers were also chosen to tag independent linkage groups, but a handful was picked to survey variation in specific chromosomal regions. Sequences surrounding 23 variants (Table S5) were submitted to LGC genomics which designed KASP TM primers (He et al. 2014). The reactions were run on an Applied Biosystems 7500 Real-Time PCR machine, with the standard KASP program. All but one set of primers (*mrpl52_T76A*) passed test runs and were run on 93 samples (with 3 blank controls). Analyses of genotype and allele frequencies, correspondence of genotypes to Hardy Weinberg proportions, and the correlation of allele frequencies in the transcriptome and charr populations (PL, LB and SB-charr) were conducted with R packages (`pegas`, `adegenet` and `hierfstat`) and custom made scripts (Jombart 2008; Paradis 2010; Goudet 2005).

3 Results

3.1 Genetic variation separating three sympatric Arctic charr morphs

To estimate the relatedness of sympatric charr, to study genome wide patterns of differentiation and to look for candidate genes related to morph separation, we screened for genetic variation in developmental transcriptomes of three Þingvallavatn morphs (SB, LB and PL-charr) (Guðbrandsson et al. 2018). Because of the extra whole genome duplication in the ancestor of salmonids (Allendorf and Thorgaard 1984; Moghadam et al. 2011) and as each sample was a pool from three individuals, we developed a F-statistic based filter (F_{PT}, see methods) to remove spurious variants caused by sequence divergence of paralogous genes. Variants with similar frequency in all samples most likely reflect sequence differences between stably expressed paralogs, but true polymorphisms should differ in frequency among samples. Simulations confirmed this assumption (see Figure S1), and from them we decided on $F_{PT} = 0.1$ as a threshold. It should be noted that paralogs differing in expression levels between morphs and/or timepoints escaped this F_{PT}-filter and some variants in the dataset could be of that nature, thus a subset of variants was subject to validation (see below). After filtering, 19,252 variants remained, in 8,961 transcripts of 7,968 genes (Tables 1 and S1 and Supplementary file S2). As the data came from genetically related samples (each sample a pool of 3 embryos, from families produced by multiparent crosses), we could not apply standard population genetic analyses. Instead, we conducted principal component analyses, calculated F_{ST} between groups, tested for GO enrichment and mapped variants to linkage groups in order to characterize the patterns of genetic variation.

Table 1: The number of variants after each filtering step in the variant-calling pipeline.

	Variants
Freebayes first run	143,744
Cov and freq filters	59,292
F_{PT} filter	23,408
Freebayes second run	19,575
Final set	19,252

The three morphs separated at the genetic level. The first principal component (based on all variants) distinguished the planktivorous morph (PL-charr) from the two benthic morphs (LB- and SB-charr), which in turn separated on the second axis (Figure 2 A). The first two axes explain 19.5% and 12.7% of the variation, and other axes 4.2% at the most each (data not show). As the samples came from pooled crosses with 5-10 parents of each sex (see methods), the individuals sequenced may be more related than a random sample from the lake. But the different crosses for PL- and SB-charr cluster together in the first two PC-axis indicating that the separation of morphs is not an artifact of the experimental setup (see 2 A). To further investigate the genetic difference between morphs we studied variants with high F_{ST}-values and private alleles in each morph (note that due to the sequencing of pooled samples from sibling embryos, the F_{ST} could be inflated). Using $F_{ST} = 0.2$ as a cutoff yielded 2331 variants (Figure 2 B). Variants were categorized by the magnitude of the absolute allele frequency difference between one morph and the other two, and called LB-, SB- or PL-charr specific variants. For instance, the SB-specific variants differed most strongly in allele frequency in SB-charr versus the average allele frequency of LB- and PL-charr. A significant excess (χ^2=140.3, df = 2, p < 0.0001) of variants (1174) belonged to the PL-specific category (separating PL-charr and the two benthic morphs), while 605 and 552 variants associated with SB- and LB-charr, respectively.

A more stringent F_{ST} cut-off (0.4) exacerbated the PL-charr vs. benthic division (Figure S4). A higher fraction of variants (χ^2=56.6, df=2, p > 0.0001) separated PL-charr from the other two (202 PL-specific, 71 SB-specific, and 51 LB-specific variants). The same pattern was observed for private variants (alleles with frequency <1% were considered absent from a morph), 56, 18 and 13 were private to PL-, SB- and LB-charr, respectively (Table S3). Note, differences in allele frequencies were between groups, and not evolutionarily polarized (ancestral versus derived). Based on these data we concluded that the largest genetic separation of the sympatric charr in Lake Þingvallavatn was between PL-charr and the two benthic morphs.
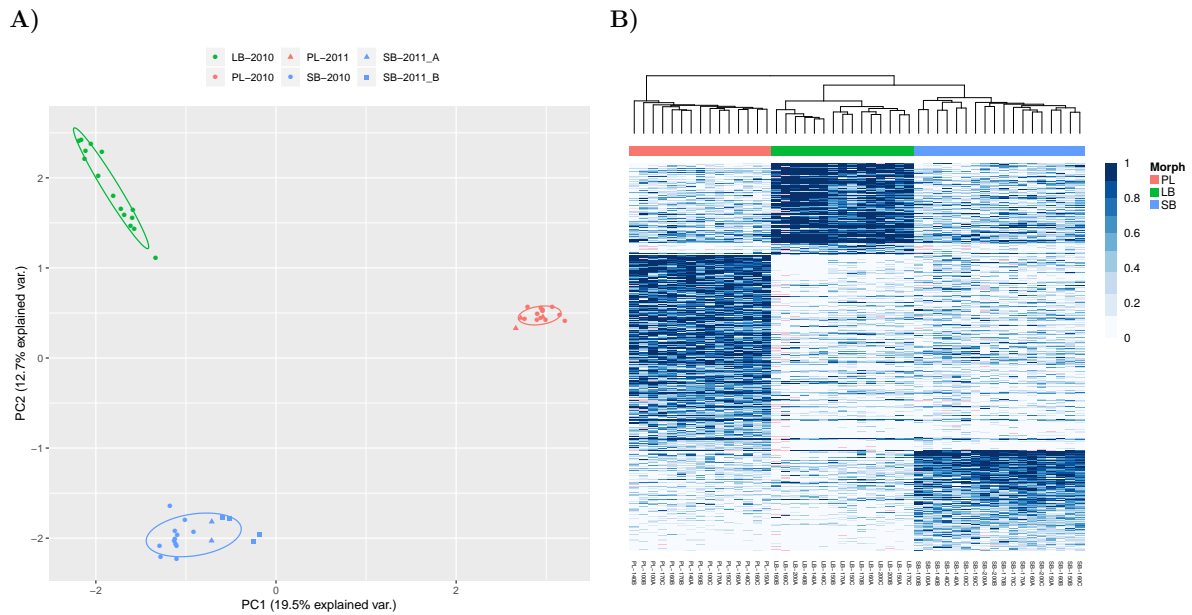
A)

B)

Figure 2: **A)** Genetic separation of samples from three sympatric Arctic charr morphs, based on principal component analysis of 19,252 transcriptome variants. The first and second principal components are depicted with the proportion of variance explained. Colours and shapes indicate which morphs (SB: blue, LB: green and PL: red) and crosses the embryos originated from (see legend). Note, while the samples from the three SB crosses do not overlap entirely, then they are very distinct from the PL and LB specimens. Overlaid are 68% normal data ellipses for each morph. **B)** Separation of morphs based on the 2331 variants with $F_{ST} > 0.2$. Each column represents a sample, name below indicates morph, developmental timepoint and biological replicate. The white-blue scale depicts allele frequency, higher frequency of the alternative allele with darker blue. Hierarchical clustering grouped samples and variants. This separated the morphs (abscissa) and similar variants by allele frequency (ordinate). A total of 1174 variants had a higher frequency in PL-charr, 552 in LB- and 605 in SB-charr. Missing values are indicated by pink. Coloring of individuals by morph, SB: blue, LB: green and PL: red.

3.2 Genetic differences between benthic and limnetic morphs in collagen metabolism and environmental sensing

In order to gauge if certain biological systems differed between the morphs, we tested for GO-enrichment in variants associating with specific morphs. As some variants were annotated to different transcript isoforms of the same gene, testing was done both on transcripts and gene level using annotation of salmon genes. No GO categories were significant for SB-specific and LB-specific variants. However, PL-specific variants were significantly enriched in 10 GO-categories (Table 2). Nine categories were found at the transcripts level and seven on the gene level. Only one of the gene level categories significant was not at the transcript level (*collagen fibril organization*) but it was close to being significant ($FDR = 0.015$). Also, variants were enriched in one other category related to the extracellular matrix *collagen catabolic process*. That may have led to two higher level categories to be significant at the transcript level but not at the gene level (*multicellular organismal catabolic process* and *macromolecule metabolic process*). Interestingly, four of the categories relate to environmental sensing and responses, *i.e.* light and sound, and showed a strong signal both on transcript and gene level, *e.g. inner ear morphogenesis* and *visual perception*. The final two categories were *tooth mineralization* and *odontogenesis*, both related to tooth development. GO-enrichment tests on private variants only revealed three GO-categories related to tRNA aminoacylation significant at the transcript level for variants private to PL-charr. The signal is most likely because three transcripts are annotated to the same gene rather than being a real biological pattern (Table S4). Together, the GO analyses of biological functions point to genetic differences in sensing, collagen metabolism and mineralization, between benthic and limnetic morphs. This could be linked to distinct differences in the main feeding habitats and principal prey species of PL- versus LB- and SB-charr and to differences in craniofacial features of these morphs.

9

Table 2: Gene ontology categories enriched in variants with F_{ST} above 0.2 and highest mean allele frequency deviation for PL-charr on the transcripts (tr) and gene (ge) level. The number of transcripts and genes with high divergence is shown (PL_{tr} and PL_{ge}) and the total number of transcripts and genes tested in the category as well (Tot_{tr} and Tot_{ge}). The multiple testing corrected p-value or false discovery rate (FDR) is also shown for both levels.

Category	Term	PL_{tr}	Tot_{tr}	FDR_{tr}	PL_{ge}	Tot_{ge}	FDR_{ge}
GO:0007601	visual perception	30	122	0.0020	32	114	0.0014
GO:0050953	sensory perception of light stimulus	30	122	0.0020	32	114	0.0014
GO:0007605	sensory perception of sound	34	156	0.0025	36	137	0.0014
GO:0034505	tooth mineralization	10	18	0.0025	9	13	0.0031
GO:0030574	collagen catabolic process	23	89	0.0069	20	60	0.0071
GO:0042472	inner ear morphogenesis	30	137	0.0083	32	128	0.0087
GO:0030199	collagen fibril organization	19	73	0.0150	18	51	0.0073
GO:0044243	multicellular organismal catabolic process	23	93	0.0087	20	65	0.0140
GO:0042476	odontogenesis	26	109	0.0083	24	92	0.0359
GO:0043170	macromolecule metabolic process	499	4,790	0.0083	490	4,155	0.3650

3.3 Genome-wide divergence of sympatric charr morphs

To explore if the genetic differences between morphs were restricted to specific chromosomal regions or more widely distributed, we mapped the variants to the Arctic charr genome using `blastn` (see methods). A large fraction (93%, 17,933 of 19,252) mapped to the genome. Of those variants 10,956 or 61% mapped to chromosomes but the rest to unplaced scaffolds and contigs. The number of variants mapped on each linkage group was, as expected, related to the linkage group size (Figure S2). Markers with $F_{ST} > 0.2$ were found on all linkage groups, and peaks of more pronounced differentiation were found on many chromosomes. Similarly, the three types of high F_{ST} markers (PL-, SB- and LB-charr specific) were found on almost all linkage groups (Figure 3 B).

There were indications of enrichment of high F_{ST} variants of a particular type in particular linkage groups, most dramatically on the mitochondrial chromosome. Previously we found genetic differentiation between SB-charr and Aquaculture charr in the mtDNA (Gudbrandsson et al. 2016), using transcriptome data and genotyping of population samples. There we observed a high frequency of derived alleles in LB- and SB-charr. In the current dataset, 6 variants were observed in the mtDNA (Table S1) and all but one had high F_{ST} (Figure 3). Consistent with previous results, these variants were LB- or SB-specific and three of those (m1829G>A, m3211T>C and m3411C>T) were within the 12s and 16s rRNA genes (Gudbrandsson et al. 2016). A lower fraction of markers had high F_{ST}'s on other linkage groups. Distinct chromosomal regions harboured variants with high differentiation associating with particular morphs. For instance, a high peak of PL-specific variants was observed on LG18, and a peak of SB-specific markers on LG10 (Figure 3) and on LG1 high F_{ST} variants of all categories (PL, SB, LB-specific) were found. In sum, the genetic separation between morphs was found on all linkage groups, including strong differences in the mitochondrial DNA.

3.4 Verification of variants in population sample confirms morph separation and regions of differentiation

As the candidate variants were derived from sequenced pools of embryos from multi-parent families, we wanted to verify them in population samples from the wild to address three questions: First, do estimates of allele frequencies in the transcriptome and in wild populations correspond? Second, do the three morphs differ genetically in population samples? Third, do some markers associate fully with morph status? Fourth, how broad are the regions of differentiation? Fifth, does the understudied Piscivorous (PI) charr differ genetically from the other three morphs?

Candidate variants (23 in total, Table S5) were chosen based on high F_{ST} values in the transcriptome sequencing, chromosomal location, biological functions and/or differential expression (Guðbrandsson et al. 2018). Sexually mature individuals of all four morphs in Lake Þingvallavatn were genotyped. All markers but one amplified successfully and behaved as true single nucleotide polymorphisms (SNPs, Table S5). The allele frequencies and estimated F_{ST} values in the transcriptome and population sample were highly correlated (Kendall's $\tau = 0.76$,
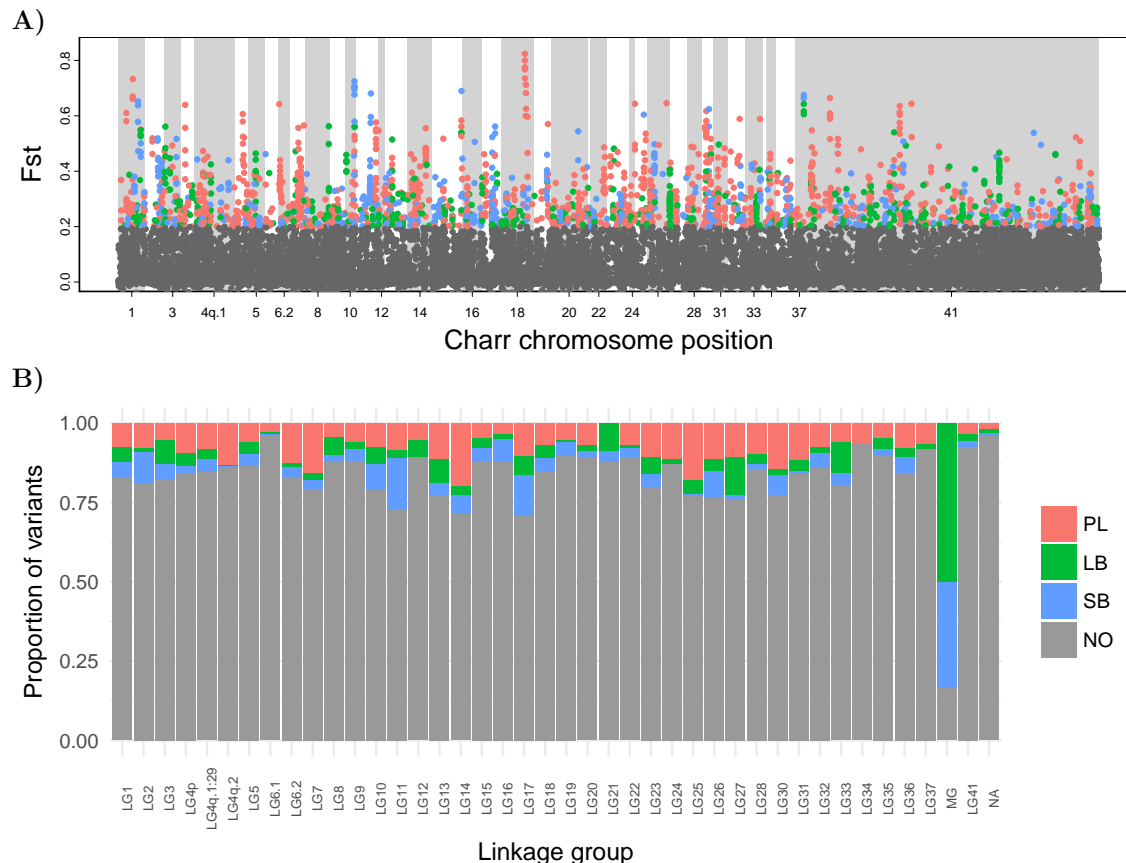
10

Figure 3: **A)** F_{ST} values plotted by position of variants on the Arctic charr genome from (Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop 2018). The colors indicate which morph differs most strongly in allele frequency from the other two for variants with F_{ST} above 0.2 (SB: blue, LB: green and PL: red) and gray represents variants with F_{ST} below 0.2. **B)** Proportion of each variant group on each linkage group (MG: mitochondrial chromosome). Unplaced scaffolds and contigs are represented by "linkage" group 41 and in **B)** NA refers to unmapped markers.

$p < 0.0001$, and $\tau = 0.58$, $p < 0.001$ respectively) (Figure S5). Allele frequencies seemed to be underestimated in the transcriptome, particularly lower frequencies (< 0.3) (Figure S5 A). The overall F_{ST}'s for individual markers ranged from 0.01 (*gnl3l_G795T*) to 0.66 (*gas1l_A3641C* and *wee1_T305A*) (Table S7). The correspondence of allele frequencies in the transcriptome and the population sample, suggests the large scale patterns in the former are real. Majority of markers (18/22) deviated significantly from Hardy-Weinberg equilibrium when tested on the 93 samples (all four morphs), but almost all were in HWE within each morph (Table S7). Consistently, the majority of the variance in genotypes associated with morph (Figure 4 A). PC1 (45.2% of variance explained) separated benthic-limnetic morphs while PC2 (14.4%) distinguished three morphs (SB, LB and PL-charr). Note, the higher proportion of variance explained by the first two PC's for the genotyped SNPs (compared to transcriptome) likely reflects the non-random selection of variants with large frequency differences for genotyping.

Eleven markers that separated benthic-limnetic morphs were genotyped (Figure 4 B), but despite strong associations, no marker coupled fully to a morph or morphotype. The 22 genotyped variants mapped to 12 separate linkage groups and 3 unplaced scaffolds (Supplemental table S5). Note that the resolution of the data is limited. For instance, the region around *calm1* on LG4.q2 had numerous variants with high differentiation between PL and the benthic morphs in the transcriptome (Figure 5 A), but it is unclear whether that reflects a single or many differentiated regions.

To ask whether there were multiple independent regions of differentiation within the same chromosome we chose several supposedly linked variants (from four specific linkage groups) from the transcriptome for genotyping in the population samples and for linkage disequilibrium (LD) within each morph. Note, the LD-estimates must
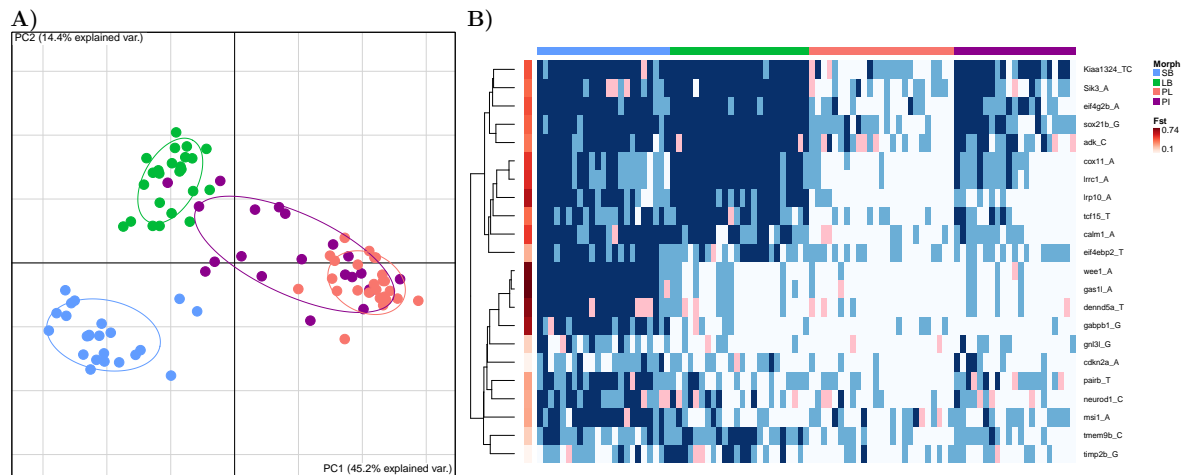
11

Figure 4: **A)** Genetic separation of the four morphs sympatric charr morphs depicted with principal component analyses of 22 variants genotyped in population samples. Individuals are graphed according to scores of the first two PC's, along with 68% normal data ellipses for each morph (SB: blue, LB: green, PL: red and PI: purple). **B)** Heatmap of genetic variation in population samples of four morphs, depicting association of variants with morphology and linkage of markers. Genotypes of the 22 variants were clustered by genes. F_{ST} values (shades of red) for morphs are graphed for each marker. The morphs are color coded as in **A)**, and genotypes, homozygous reference allele (white), heterozygous (light-blue) and homozygous alternate allele (blue), with pink indicating missing data.

be interpreted cautiously as they are imprecise because only 21 or 24 fish were sampled from each morph (Yan et al. 2009).

We surveyed four sets of putatively linked markers. Two pairs of markers were separated by more than 10 Mb, and showed very low or no association. These were PL-specific markers (*Kiaa1324_TC393AA* and *eif4g2b_G652A*) on LG26 (Figure 5 B) and two markers in (*msi1* and *tcf15*) on LG11 (Figure 5 C, Table 3).

More tightly spaced were three markers mapping to LG1 that separated PL-charr from the benthic morphs (Figure 5 D). While *lrrc1* and *cox11* are 2.8 Mb apart the markers showed tight LD ($r^2 > 0.72$) (Table 3) but the third marker (*adk_C1075T*), located 3.4 Mb downstream of *cox11*, had weak or no LD with the other two (Table 3). This suggests there are two differentiated regions within this ∼6 Mb region of the genome.

Table 3: Estimates of LD (r^2), within each morph, and genomic distance for pairs of genotyped variants that mapped to the same linkage group or showed a strong association.

Var1	Var2	Pattern of divergence[a]	Linkage group	Genomic distance (bp)	SB	LB	PL	PI
lrrc1	*cox11*	PL	LG18	2,762,681	0.90	1.00	0.73	0.90
cox11	*adk*	PL	LG18	3,370,151	0.04	NA	0.06	0.23
lrrc1	*adk*	PL	LG18	6,132,832	0.03	NA	0.00	0.32
eif4g2b	*Kiaa1324*	PL	LG26	16,793,974	0.08	0.00	0.00	0.06
msi1	*tcf15*	SB, PL	LG11	11,914,739	0.02	0.06	0.01	0.00
dennd5a	*wee1*	SB	LG10	118,070	0.82	0.92	0.52	1.00
wee1	*gas1l*	SB	LG10[b]	[b]210,891	0.51	0.85	0.51	0.95
dennd5a	*gas1l*	SB	LG10[b]	[b]328,961	0.42	0.79	0.47	0.95
gas1l	*tmem9b*	SB, PL	LG15	400,866	0.02	0.00	0.00	0.26

[a] In which morph did the variant differ most in frequency, first description for Var1 and second for Var2.
[b] Second best blast hit for *gas1l* was on LG10, the best was on LG15.
NA: Not available, because variant *adk* was not polymorphic in LB.

Three SB-specific variants (*gas1l_A3641C*, *wee1_T305A* and *dennd5a_A2555T*) on LG10 were in strong LD ($r^2 > 0.45$) in all morphs. While the variants in *dennd5a* and *wee1* mapped 118 kb apart, then the region

containing the *gas1l* variant aligned to LG15 (The second best hit was on LG10 close to the other two, Figure 5 E). LG10 and LG15 are ohnologous and have over 90% sequence similarity (Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop 2018). Notably, the *gas1l* variant was in linkage equilibrium with another variant on LG15 (in *tmem9b*, Figure 5 F, Table 3). These discrepancies may be due to assembly errors in the Arctic charr draft genome by Christensen, Rondeau, Minkley, Leong, Nugent, Danzmann, Ferguson, Stadnik, Devlin, Muzzerall, Edwards, Davidson and Koop (2018) as the regions with the three SB-specific variants all mapped to the same chromosome in *S. salar* (ssa26) and *O. mykiss* (omy06) genomes. Otherwise, there was low LD between variants (Table S8).

Finally, the phyletic relationship of the rare PI-charr to the other morphs is unknown (Gíslason 1998; Volpe and Ferguson 1996). It was hypothesized that PI-charr could be an ontogenetic morph of PL-charr that learn to utilize threespined sticklebacks as prey (Snorrason et al. 1989). No transcriptome data were available for PI-charr, but we genotyped 21 individuals. Contrary to the other three morphs, PI individuals did not group in the PC plot (Figure 4 A). This may depend on some technical shortcomings, i) too few markers, ii) ascertainment bias or iii) incorrect classification of PI-individuals. If technical biases do not dominate, then the data may reflect biological patterns. At face value, the data suggest the PI-charr are more heterogeneous genetically than the other morphs. And the data do not support the ontogenetic shift hypothesis (which predicted the PI should cluster with the PL-charr). Curiously, half of the PI-charr were in the PL-cluster but the rest (except one) fell in between the PL- and LB-clusters. Thus while the other three morphs separated clearly genetically, determination of the genetic status and nature of the piscivorous charr requires more data.

In summary, while no single variant associated fully with specific morph many distinct regions within chromosomes harbour variants separating the morphs.

4   Discussion

The four sympatric morphs of Arctic charr in Lake Þingvallavatn differ in a host of phenotypic attributes, e.g. in adult morphology, habitat choice, feeding preferences, growth pattern, size and age at maturity (see references in Sandlund et al. 1992). Common garden rearing experiments indicate these differences are based on genotype but also influenced by the environment, and emerge during embryonic and juvenile development (Skúlason, Noakes and Snorrason 1989; Parsons et al. 2010; Kapralova et al. 2015). While population genetics indicated a genetic separation of PL-, LB- and SB-charr (Magnusson and Ferguson 1987; Volpe and Ferguson 1996), the degrees of relatedness and phyletic relationships of these morphs have been unresolved. Here we utilized SNPs from RNAseq data derived from embryos of pure crosses of three morphs (Guðbrandsson et al. 2018) to assess their genetic separation. For each variant, we asked if one of the morphs was more diverged, and if so, how morph-specific variants were distributed in the genome and whether they related to particular functional systems. Although the experimental design, pools of individuals from bulk crosses of each morph, was not optimal for these purposes the data yielded many informative variants. Consistently, 22 of the 23 selected variants genotyped in independent population samples were real variants. Estimates of F_{ST} in the transcriptome may have been inflated relative to the population samples (15 out of 22 markers had lower F_{ST} in the latter, Figure S5). Notably, the frequencies of rare alleles were underestimated in the transcriptome as expected (Konczal et al. 2014) but might be exaggerated due to the sequencing of related individuals, pooling of three embryos in individual samples or Beavis effects (Beavis 1994) because high F_{ST} SNPs were chosen for verification. These facts limit the interpretation of this dataset, precluding for example estimation of the site frequency spectrum. But they do not negate the observed widespread genetic differentiation between morphs, as the correspondence of allele frequencies estimated with the two methods used (RNAseq and KASP-assay) and the genetic differentiation of morphs is stronger than the variation between crosses within the same morph, that suggest the patterns are biological and not a side effect of the experimental design.

Previously, we (Gudbrandsson et al. 2016) extracted variants from RNAseq data from SB-charr of Þingvallavatn and an aquaculture breeding stock derived from several populations (mainly from the north of Iceland (Svavarsson 2007) - an outgroup for Þingvallavatn charr). While that study also yielded ∼20,000 candidate variants, the proportion of variants with high F_{ST} was larger, probably reflecting more divergence between the aquaculture charr and the SB-charr than among the three Þingvallavatn morphs studied here. It is unknown how many of the variants observed in these RNAseq data are shared with other populations in Iceland and across the species range (Brunner et al. 2001).
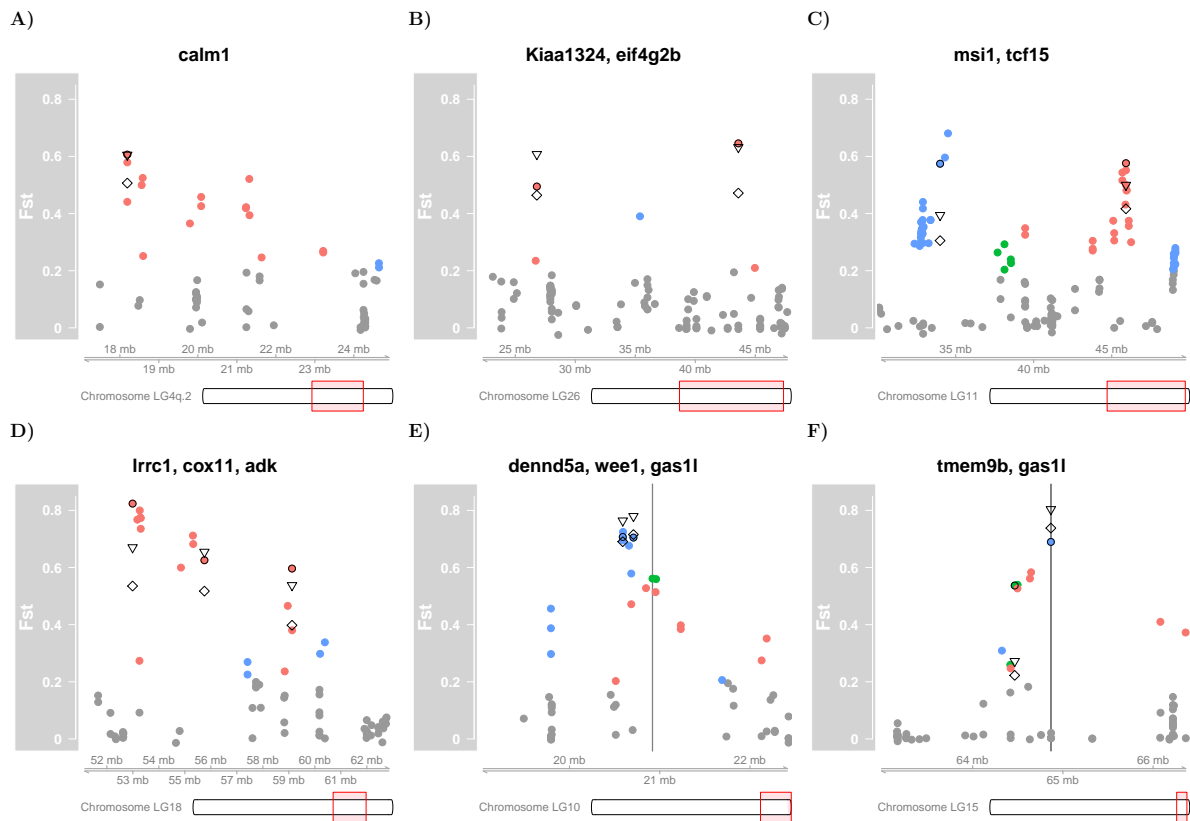
Figure 5: Detailed view of chromosomal regions and variants showing high divergence in the KASP-assay. **A)** Region on LG4q.2 that includes *calm1* and many other variants with high F_{ST} values (not validated). **B)** A large region on LG26 that includes *Kiaa1324* and *eif4g2b*. **C)** A region on LG11 that includes *msi1* and *tcf15*. **D)** A region on LG18 that includes *lrcc1*, *cox11* and *adk*. **E)** A region on LG10 that includes *dennd5a*, *wee1*. The location of the second best blast hit for the variant in *gas1* is shown with a vertical line. **F)** A region on LG15 that includes *tmem9b* and *gas1l*. The location of the best blast hit for the variant in *gas1* is shown with a vertical line. The regions on LG10 and LG15 shown in **E)** and **F)** are ohnologous and some of the markers might belong in the opposite linkage group as we suspect is the case for *gas1l*. Colored dots are values from the transcriptome as in Fig. 3 and validated variants are marked with a black circle (○). The triangles (▽) show the F_{ST} value from the KASP-assay for three morphs (PL, SB and LB) and the diamonds (◇) F_{ST} for all morphs (including PI).

4.1 Genetic separation of recently evolved sympatric morphs

Geological forces, volcanic activity, isostatic rebound as well as underground rivers and springs in the rift-zone generate specific niches for the Icelandic biota (Kornobis et al. 2010; Kristjánsson et al. 2012; Marteinsson et al. 2013; Gudmundsdóttir et al. 2017), but also perturb them. Lake Þingvallavatn formed ∼10,000 years ago but resides in a geologically active rift zone, with the most recent eruption ∼2000 years ago (Saemundsson 1992). Furthermore, the tempo of the Ice-age glacial retreat was uneven, interleaved with periods of glacial advance (Saemundsson 1992; Norðdahl et al. 2008). Thus, the history of colonization and adaptations by Arctic charr and other organisms in this geographic region may be quite complex.

Studying the relatedness of the morphs can shed some light on the history of colonization and divergence in Lake Þingvallavatn. The morphs (PL-, LB- and SB-charr) are more closely related to one another than to other Icelandic populations (Kapralova et al. 2011). The data presented here firmly reject hypothesis I, that the PL-, LB- and SB-charr morphs are formed by environmental influence on individuals of a shared gene pool (the PI-charr may be an exception, see below). The data also suggest that the benthic morphs (LB- and SB-charr) are more closely related. This is congruent with patterns of differential expression in this transcriptome (Guðbrandsson et al. 2018) and one previous study (Volpe and Ferguson 1996) but not others (Magnusson and Ferguson 1987; Danzmann et al. 1991; Gíslason 1998). This incongruence and the observed distribution of the high F_{ST} variants

14

into PL-, LB- and SB-specific categories may reflect limited resolution due to few markers in earlier studies of Lake Þingvallavatn charr, incomplete sorting of alleles or differential effects of selection on distinct traits and variants. Coalescence simulations based on microsatellite data were more consistent with a brief phase of allopatric separation of the ancestors of PL- and SB-charr Lake Þingvallavatn, rather than sympatric evolution (Kapralova et al. 2011). This study considered two demographic scenarios and two parameters (migration and population size). Analyses of other scenarios, *i.e.* with changing migration rates or introgression would be most interesting (see for example Jacobs et al. 2018). Considering the episodic nature of ice retreat and the high geological activity in the Þingvallavatn area during the formation of the lake (Saemundsson 1992; Norðdahl et al. 2008) charr may have invaded the waters more than once, or experienced intromittent isolation of populations in rifts. The current study lacks a natural outgroup, and future population genomic surveys of these morphs should include populations from the neighbouring geographic area, including the Þingvallavatn outflow river Sog and its downstream Lake Úlfljótsvatn. Lake Úlfljótsvatn is particularly interesting as it also hosts landlocked morphs of charr, including forms similar to the PL- and SB-charr (Jóhannsson et al. 1994; Woods et al. 2012a).

The *Salvelinus* genus is particularly interesting for sympatric polymorphism (Noakes 2008; Klemetsen 2010; Muir et al. 2016; Markevich et al. 2018) that seems to run counter to the predominant mode of allopatric speciation. Smith and Skúlason (1996) argued resource polymorphism and phenotypic plasticity could potentiate ecological speciation, even within a geographic area. The question remains if charr morphs in sympatry (e.g. in the Transbaikalian lakes (Gordeeva et al. 2015), Lake Galtarból in northern Iceland (Wilson et al. 2004), and Lower Tazimina Lake in Alaska (May-Mcnally et al. 2015)) differentiated in true sympatry, or originate by repeated invasions of the water bodies. The morphs in Loch Stack (Adams et al. 2008) and Loch Tay (Garduño-Paz et al. 2012) are clearly of allopatric origin. A recent genome-wide study of sympatric charr in Scotland and Siberia is consistent with sympatric origins of charr morphs (Jacobs et al. 2018). Ecological specialization can lead to differential growth, behavior and maturity that may influence both choice of spawning sites and timing of spawning. Within Lake Þingvallavatn, the timing of spawning differs between the three best-studied morphs. The LB-charr spawns in August, PL-charr in October, but SB-charr over a broader period from September to November (PI-charr spawn in October) (Skúlason, Snorrason, Noakes, Ferguson and Malmquist 1989; Sandlund et al. 1992). Although all morphs appear to spawn in the same habitat, i.e. among loose stones in the littoral zone, there may be micro-habitat preferences for spawning sites. Variation in mating behavior can in principle lead parents to choose specific nesting sites, e.g. close to inflow of cold water. Such differences in environmental conditions could, in turn, influence development of ecological functional traits and adult form of the morphs.

The ambiguous genetic status of the large, rare and piscivorous morph is quite intriguing. PI-charr was only studied by genotyping 22 markers, but curiously PI individuals did not cluster as tightly as the other morphs. PI individuals grouped with LB, PL-charr and a few in between. Previous studies found PI individuals affiliated with either or both morphs (Magnusson and Ferguson 1987; Volpe and Ferguson 1996). Several scenarios may explain the data and the origin of PI-charr. First, classification of the morph based on phenotype may not be stringent enough and thus some LB-charr might have been misidentified as PI-charr. As the LB- and PI-charr utilize different prey and differ in spawning periods (the former in August and later in October (Skúlason, Snorrason, Noakes, Ferguson and Malmquist 1989)) this misidentification is unlikely. However, it is possible that some LB-charr males might be erroneously classified as PI because of an extended jaw-hook, as some are still running in October. Secondly, it is possible that PI-charr is genetically distinct, but that the markers chosen for genotyping may not suffice to distinguish them. Thirdly, Snorrason et al. (1989) postulated that PI-charr emerges as certain PL-charr learn to eat fish. PI-charr would then be environmentally induced, genetically identical to PL-charr. The fact that PI-charr does not overlap exclusively with PL-charr argues against this theory. Fourthly, PI-charr may actually be more heterogeneous at the genetic level than the other three morphs. This might reflect recent origin or perhaps a combination of ontogenetic shift of some individuals from PL stock with repeated hybridizations with other morphs (most likely LB-charr because of the size). Currently, we can not distinguish between these possibilities.

We conclude that three of the Lake Þingvallavatn morphs differ genetically, and considering the differences in trophic traits and spawning, may even be reproductively isolated (hypothesis III). It remains to be determined if the PI-charr are genetically separated from the other morphs or if they are an environmentally induced form as may be the case of eco-morphs in Sockeye salmon *Oncorhynchus nerka* (Limborg et al. 2018).

4.2 Genome-wide separation of sympatric morphs

How extensive is the genetic separation of morphs and how is the genetic differentiation distributed in the genome? For instance, are differentiating variants localized to a few islands (Nadeau et al. 2012; Andrew and Rieseberg

2013; Malinsky et al. 2015) as stated in hypothesis II or are there multiple signals on many chromosomes (Hohenlohe et al. 2010; Jones et al. 2012) according to hypothesis III?

Variants with relatively high F_{ST} (>0.2) mapped to all 40 *S. alpinus* linkage groups (including the mitochondrial chromosome), which supports hypothesis III. The clustering of variants, some with F_{ST} >0.5, suggests specific genes/regions associated with the specializations of particular morphs. Analyses of LD between genotyped variants implied that peaks on the same linkage group were independent. The distribution of variants on linkage groups was rather even, except the mtDNA had a high fraction of LB- and SB-specific variants (consistent with earlier findings (Gudbrandsson et al. 2016)). This might reflect history or adaptive evolution of mitochondrial functions in the benthic morphs. The extensive genetic separation of these three morphs suggests many differentiating genes on many (or all) linkage groups. While the data seem to refute hypothesis II, that few genomic islands differentiate the morphs, it is plausible that ecological specialization has driven differentiation in certain genomic regions.

Previous studies of salmonids have found genome-wide differences between populations and ecological morphs. Significant genetic differences were found between migratory and non-migratory Rainbow trout (Hale et al. 2013), the lake, river and stream ecotypes of sockeye salmon (Larson et al. 2017) and ecologically different subpopulations of salmon (Vincent et al. 2013; Cauwelier et al. 2017). Sympatric ecotypes are found in several salmonid species, e.g. whitefish (Gagnaire et al. 2013), but are most pronounced in Arctic charr and Lake charr *Salvelinus namaycush*. Genomic studies of the latter revealed varying degrees of genetic separation of sympatric morphs in large lakes in North America (Harris et al. 2015; Perreault-Payette et al. 2017). A study on five pairs of benthic-limnetic whitefish morphs (Gagnaire et al. 2013) found a correlation between genetic and phenotypic differentiation and considerable overlap of genomic regions that differentiated morph pairs in each lake. Notably, the significant degree of genetic parallelism broke up at the finer level, with different haplotypes associating with morphotype, suggesting genetic and/or allelic heterogeneity of the causative loci (Gagnaire et al. 2013). Recently, genome wide analyses of Scottish and Siberian Arctic charr suggested limited genetic parallelism in benthic-limnetic specializations (Jacobs et al. 2018). Fine scale analyses of peaks of differentiation could not be conducted with the current data, but the genome wide distribution of differentiated variants suggests restricted gene flow between at least three of the morphs (PL-, LB- and SB-charr) and that they may be reproductively isolated.

Genomic differentiation reflects the history and evolution of groups, degree and age of separation of groups, extent of gene flow and hybridization events (Seehausen et al. 2014; Shapiro et al. 2016). But intrinsic factors, such as genes and their biological actions, chromosomal inversions and polymorphisms, also shape the rates of differentiation and divergence, and genomic features, such as centromeres, gene density, recombination and GC content influence estimates of sequence divergence and allele frequency differences (Seehausen et al. 2014; Burri et al. 2015; Sutherland et al. 2016; Vijay et al. 2017). Population genomic data and analyses are needed to disentangle the role of positive selection and intrinsic factors on patterns of differentiation in this system. More broadly, the confirmed synteny of large genomic regions (Nugent et al. 2017), the range of species, subspecies and ecologically distinct populations (Klemetsen 2013; Jacobs et al. 2018) sets the stage for future studies of the genomic and ecological correlates of divergence and polymorphism in salmonids.

4.3 Potential mechanisms of phenotypic and developmental differences between sympatric morphs

Which genes influence the many phenotypic differences between the morphs? Genetic variation in transcribed sequences tags large fraction of the potentially functional regions of the genome. The genome-wide differentiation between morphs implies polygenic basis of their differences, but major genes influencing specific traits may be segregating. The genomic resolution of the current data is low, and while we can describe frequency differences in variants in specific genes, it is more likely that other linked polymorphisms/loci are actually contributing to phenotypic differences.

The current data can inform future studies by bringing attention to particular chromosomal regions, genes and systems that may influence differences between charr morphs. The GO results suggest the limnetic and benthic morphs differ genetically in three systems (collagen metabolism, tooth mineralization and sensory functions). The enrichment of variants in collagen organization/catabolism and extracellular matrix categories is consistent with observed differential expression of ECM related and cartilage remodeling genes in benthic and limnetic morphs (Ahi et al. 2014, 2015). Strong differentiation was found for a variant in *calm1*. Calmodulins are broadly expressed in vertebrate tissues, including bone and articular cartilage in humans (Mototani et al. 2005), and notably higher expression of *Calm1* was found in beak primordia of finches with longer beaks (Abzhanov et al. 2006). Secondly,

we found frequency difference of a variant in *eif4ebp1* between morphs (F_{ST} = 0.28). This gene had higher expression in SB compared to PL-charr in developing embryos (this transcriptome (Guðbrandsson et al. 2018)). Curiously, the same was seen in adult muscles of SB-charr from other locations in Iceland (Macqueen et al. 2011). *eif4ebp1* is involved in the insulin receptor signaling pathway and mTOR signalling (Bidinosti et al. 2010; Gkogkas et al. 2012; Banko et al. 2007). However, the *eif4ebp1* variant associated with benthic morphotype not small benthic charr, which is at odds with the proposition that reduced growth of the SB populations is mediated through changes in the mTOR activity. Finally, three SB-specific variants on linkage group 10 (or 15) had high F_{ST}'s. This might represent long haplotypes with strong differentiation between morphs, that associate with the small benthic phenotype. Curiously one gene in the region (*Wee1*) is a key regulator of the timing of mitosis and cell size, first identified in fission yeast (Nurse and Thuriaux 1980) and another (*gas1*, *Growth arrest-specific protein 1*) associates with embryonic cranial skeleton morphogenesis and palate development (Seppala et al. 2007).

Few traits have been mapped to a gene in salmonids. The most celebrated exception being *vgll3* which associates with the age of maturity in wild populations, with very curious sex-dependent dominance of alleles (Barson et al. 2015; Ayllon et al. 2015). Unfortunately the *vgll3* gene was not transcribed in our data, but SB-specific variants with F_{ST} around 0.4 are found ∼1 Mb from the location of *vgll3* on LG2 (Figure S6 K). Further work is needed to check if variants in *vgll3*, *tulp4* (see Larson et al. 2017), *Greb1L* (see Prince et al. 2017) or other genes contribute to phenotypic differences in charr. QTL and fine-scale association mapping (Zimmerman et al. 2000; Palsson and Gibson 2004; Dworkin et al. 2005) are needed to find the causative variants. In sum, the data presented here can lead to the identification of genes influencing specific phenotypes separating the sympatric morphs, but more crucially suggest a polygenic basis for the morph differences.

4.4 Conclusion and future perspective

We mined genetic variation from three of the four morphs in Lake Þingvallavatn to address questions about their genetic separation, genome-wide differentiation of morphs and potential functionality of loci separating them. We started with three hypotheses about the putative causes of morph differences: Hypothesis I stated that the morphs are environmentally induced. According to hypothesis II, ecological specialization led to few genomic islands of strong genetic differentiation between morphs, with low background separation across the genome. Hypotheses III postulated substantial differentiation across the genome due to reduced gene flow, with the bulk of the genome showing separation between morphs. Estimates of genetic differentiation between the Þingvallavatn morphs have yield different results from low (Volpe and Ferguson 1996; Kapralova et al. 2011) to extremely high (Kapralova et al. 2013) depending on the marker used. Based on the genetic differentiation observed in this study, we conclude that three of the morphs (SB, LB and PL-charr) are distinct populations and find it highly unlikely that these morphs are environmentally induced (rejecting hypothesis I). Furthermore, we interpret the data as suggesting genome-wide differentiation over weak differentiation with a few islands of differentiation (hypothesis III over II). This is supported by the facts that the background F_{ST} in the transcriptome was rather high and high F_{ST} peaks were found on all linkage groups (and even multiple peaks on some). The fact that spawning times of certain morphs do not overlap (LB-charr spawn in August, but PL- and SB-charr in September - October) also argues that gene flow has been reduced between these morphs. Note however that hybrids of specific morphs, SB-, PL- and PI-charr can be generated in the laboratory (Kapralova 2014, Kalina H. Kapralova, Sigurdur S. Snorrason, Zophonías O. Jónsson, Arnar Pálsson *et al.* unpublished data) but their fitness and how common they are in nature is unknown. Thus we conclude that the three studied morphs in Lake Þingvallavatn are not one panmictic population and that gene flow between them has been limited. The nature of the PI-charr is still in doubt, some PI-charr may arise every generation by ontogenetic shift as suggested by Snorrason et al. (1989) while other PI may propagate, and even hybridize with the LB-charr. Population ddRAD-seq data of sexually mature/spawning charr of all morphs confirm the clear demarcation of three of the morphs and offer insights into the nature of the PI-charr (Han Xiao, Benjamín Sigurgeirsson, et al. unpublished data). Future studies of the fertilization success of hybrids and pure morphs, analyses of the development and fitness of hybrids and pure morph, and behavioural studies of spawning behaviour, reed locations and properties, and mate choice have can cast light on potential pre- and postzygotic barriers to gene flow between morphs. In the future, the phenotypic and genetic differences between sympatric and locally adapted Arctic charr populations can aid studies of ecological adaptation and the synthesis of evolutionary, ecological and developmental biology.

17

5 Acknowledgments

6 Author contributions

Conceived and designed the study: JG, AP, ZOJ, SSS, SRF, KHK
Sampling: SSS, KHK, SRF, ZOJ, AP, JG
Gathered the data: ZOJ, SRF, KHK, VH, ÞMB
Retrieval and analyses of variants data: JG
SNP confirmation: KHK, AP, VH, ÞMB, JG
Analyses: JG, AP, KHK
Writing: AP, JG, SSS, KHK, SRF, ZOJ

7 Data access

The sequencing reads from the 48 RNA-seq samples were deposited into the NCBI SRA archive under BioProject identifier PRJNA391695 and with accession numbers: SRS2316381 to SRS2316428. Supplementary tables and files are available on figshare https://doi.org/10.6084/m9.figshare.c.4565735.v1 (specific hyperlinks in Supplementary material).

References

Abouheif, E., Favé, M.-J., Ibarrarán-Viniegra, A. S., Lesoway, M. P., Rafiqi, A. M. and Rajakumar, R. (2014). Eco-Evo-Devo: The Time Has Come, *in* C. R. Landry and N. Aubin-Horth (eds), *Ecological Genomics - Ecology and the Evolution of Genes and Genomes*, Springer Netherlands, Dordrecht, chapter 6, pp. 107–125. 2

Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R. and Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches, *Nature* **442**(7102): 563–567. 2, 16

Adams, C. E., Wilson, A. J. and Ferguson, M. M. (2008). Parallel divergence of sympatric genetic and body size forms of Arctic charr, *Salvelinus alpinus*, from two Scottish lakes, *Biological Journal of the Linnean Society* **95**(4): 748–757. 15

Ahi, E. P., Gudbrandsson, J., Kapralova, K. H., Franzdóttir, S. R., Snorrason, S. S., Maier, V. H. and Jónsson, Z. O. (2013). Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr., *PloS ONE* **8**(6): e66389. 4

Ahi, E. P., Kapralova, K. H., Pálsson, A., Maier, V. H., Gudbrandsson, J., Snorrason, S. S., Jónsson, Z. O. and Franzdóttir, S. R. (2014). Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr, *EvoDevo* **5**(1): 40. 4, 16

Ahi, E. P., Steinhäuser, S. S., Pálsson, A., Franzdóttir, S. R., Snorrason, S. S., Maier, V. H. and Jónsson, Z. O. (2015). Differential expression of the aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr, *EvoDevo* **6**(1): 27. 4, 16

Allendorf, F. W. and Thorgaard, G. H. (1984). Tetraploidy and the Evolution of Salmonid Fishes, *in* B. J. Turner (ed.), *Evolutionary Genetics of Fishes*, Springer US, pp. 1–53. 4, 6, 8

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool., *Journal of molecular biology* **215**(3): 403–10. 5

Andrew, R. L. and Rieseberg, L. H. (2013). Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes, *Evolution* **67**(9): 2468–2482. 15

Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M. F., Dahle, G., Taranger, G. L., Glover, K. A., Almén, M. S., Rubin, C. J., Edvardsen, R. B. and Wargelius, A. (2015). The *vgll3* Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (*Salmo salar* L.) Males, *PLOS Genetics* **11**(11): e1005628. 17

Banko, J. L., Merhav, M., Stern, E., Sonenberg, N., Rosenblum, K. and Klann, E. (2007). Behavioral alterations in mice lacking the translation repressor 4E-BP2, *Neurobiology of Learning and Memory* **87**(2): 248–256. 17

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stromberg, M. P. and Marth, G. T. (2011). BamTools: A C++ API and toolkit for analyzing and managing BAM files, *Bioinformatics* **27**(12): 1691–1692. 6

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., Erkinaro, J., Lien, S. and Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon, *Nature* **528**(7582): 405–408. 17

Beavis, W. D. (1994). The power and deceit of QTL experiments: Lessons from comparitive QTL studies, *in* D. B. Wilkinson (ed.), *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference*, American Seed Trade Association, Washington DC, pp. 250–266. 13

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(1): 289–300. 7

Bernatchez, L., Renaut, S., Whiteley, A. R., Derome, N., Jeukens, J., Landry, L., Lu, G., Nolte, A. W., Østbye, K., Rogers, S. M. and St-Cyr, J. (2010). On the origin of species: Insights from the ecological genomics of lake whitefish., *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**(1547): 1783–800. 2

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G. H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Roest Crollius, H. and Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates., *Nature communications* **5**: 3657. 4, 6

Bidinosti, M., Ran, I., Sanchez-Carbente, M. R., Martineau, Y., Gingras, A.-C., Gkogkas, C., Raught, B., Bramham, C. R., Sossin, W. S., Costa-Mattioli, M., DesGroseillers, L., Lacaille, J.-C. and Sonenberg, N. (2010). Postnatal Deamidation of 4E-BP2 in Brain Enhances Its Association with Raptor and Alters Kinetics of Excitatory Synaptic Transmission, *Molecular Cell* **37**(6): 797–808. 17

Bray, N. L., Pimentel, H., Melsted, P. and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification, *Nature Biotechnology* **34**(5): 525–527. 5

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C. and Bernatchez, L. (2001). Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences, *Evolution* **55**(3): 573–586. 13

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Růžić, M., Sæther, S. A., Sætre, G. P., Török, J. and Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers, *Genome Research* **25**(11): 1656–1665. 16

Carlson, M. (2015). GO.db: A set of annotation maps describing the entire Gene Ontology. 7

Cauwelier, E., Gilbey, J., Sampayo, J., Stradmeyer, L. and Middlemas, S. J. (2017). Identification of a single genomic region associated with seasonal river return timing in adult Scottish Atlantic salmon (*Salmo salar* L.) identified using a genome-wide association study., *Canadian Journal of Fisheries and Aquatic Sciences* pp. cjfas–2017–0293. 16

19

Christensen, K. A., Leong, J. S., Sakhrani, D., Biagi, C. A., Minkley, D. R., Withler, R. E., Rondeau, E. B., Koop, B. F. and Devlin, R. H. (2018). Chinook salmon (*Oncorhynchus tshawytscha*) genome and transcriptome, *PLOS ONE* **13**(4): e0195461. 4

Christensen, K. A., Rondeau, E. B., Minkley, D. R., Leong, J. S., Nugent, C. M., Danzmann, R. G., Ferguson, M. M., Stadnik, A., Devlin, R. H., Muzzerall, R., Edwards, M., Davidson, W. S. and Koop, B. F. (2018). The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly, *PLOS ONE* **13**(9): e0204076. 4, 7, 11, 13

Danzmann, R. G., Cairney, M., Davidson, W. S., Ferguson, M. M., Gharbi, K., Guyomard, R., Holm, L.-E., Leder, E., Okamoto, N., Ozaki, A., Rexroad, C. E., Sakamoto, T., Taggart, J. B. and Woram, R. A. (2005). A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae)., *Genome* **48**(6): 1037–51. 4

Danzmann, R. G., Ferguson, M. M., Skulason, S., Snorrason, S. S. and Noakes, D. L. G. (1991). Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn, Iceland, *Journal of Fish Biology* **39**(5): 649–659. 4, 14

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.*, Vol. 5, John Murray, London. 1

De Meyer, J., Christiaens, J. and Adriaens, D. (2016). Diet-induced phenotypic plasticity in European eel (*Anguilla anguilla*), *Journal of Experimental Biology* **219**(3): 354–363. 2

Dworkin, I., Palsson, A. and Gibson, G. (2005). Replication of an *Egfr*-wing shape association in a wild-caught cohort of *Drosophila melanogaster*, *Genetics* **169**(4): 2115–2125. 17

Foote, A. D., Vijay, N., Ávila-Arcos, M. C., Baird, R. W., Durban, J. W., Fumagalli, M., Gibbs, R. A., Hanson, M. B., Korneliussen, T. S., Martin, M. D., Robertson, K. M., Sousa, V. C., Vieira, F. G., Vinař, T., Wade, P., Worley, K. C., Excoffier, L., Morin, P. A., Gilbert, M. T. P. and Wolf, J. B. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes, *Nature Communications* **7**(May): 11693. 2

Frandsen, F., Malmquist, H. J. and Snorrason, S. S. (1989). Ecological parasitology of polymorphic Arctic charr, *Salvelinus alpinus* (L.), in Thingvallavatn, Iceland, *Journal of Fish Biology* **34**(2): 281–297. 3

Fruciano, C., Franchini, P., Kovacova, V., Elmer, K. R., Henning, F., Meyer, A., Via, S., Coyne, J. A., Bird, C. E., Fernandez-Silva, I., Skillings, D. J., Toonen, R. J., Michel, A. P., Savolainen, V., Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., Meyer, A., Elmer, K. R., Bolnick, D. I., Fitzpatrick, B. M., Feder, J. L., Nosil, P., Via, S., Via, S., Conte, G., Mason-Foley, C., Mills, K., Elmer, K. R., Kusche, H., Lehtonen, T. K., Meyer, A., Franchini, P., Franchini, P., Fruciano, C., Frickey, T., Jones, J. C., Meyer, A., Schluter, D., Svanback, R., Eklov, P., Rouleau, S., Glemet, H., Magnan, P., Meyer, A., Gavrilets, S., Vose, A., Barluenga, M., Salzburger, W., Meyer, A., Flaxman, S. M., Wacholder, A. C., Feder, J. L., Nosil, P., Gel, B., Gunter, H. M., Rebsamen, M., Seehausen, O., Barluenga, M., Meyer, A., Roesti, M., Hendry, A. P., Salzburger, W., Berner, D., Fruciano, C., Franchini, P., Raffini, F., Fan, S., Meyer, A., Slate, J., Boulding, E. G., Boulton, K., Bookstein, F. L., Valentin, A. E., Penin, X., Chanut, J. P., Sévigny, J. M., Rohlf, F. J., Fruciano, C., Tigano, C., Ferrito, V., Fruciano, C., Fruciano, C., Tigano, C., Ferrito, V., Escoufier, Y., Rohlf, F. J., Corti, M., Klingenberg, C. P., Franchini, P., Xiong, P., Fruciano, C., Meyer, A., Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., Hoekstra, H. E., Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., Haas, B. J., Henning, F., Lee, H. J., Franchini, P., Meyer, A., Haley, C. S., Knott, S. A., Hu, Z., Xu, S., Bolger, A. M., Lohse, M., Usadel, B., Pertea, M., Trapnell, C., Pachter, L., Salzberg, S. L., Langmead, B., Salzberg, S. L., Li, H., Trapnell, C., Grabherr, M. G., Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., Quinlan, A. R., Hall, I. M., Larkin, M. A., Kimura, M., Needleman, S. B., Wunsch, C. D., Nei, M., Gojobori, T., Yang, Z., Nielsen, R., Yang, Z., Goldman, N., Yang, Z., Conesa, A., Kofler, R., Pandey, R. V., Schlotterer, C., Kofler, R., Castro-Conde, I. and de Una-Alvarez, J. (2016). Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish, *Nature Communications* **7**: 12736. 2

Gagnaire, P. A., Pavey, S. A., Normandeau, E. and Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing, *Evolution* **67**(9): 2483–2497. 16

Garduño-Paz, M. V., Adams, C. E., Verspoor, E., Knox, D. and Harrod, C. (2012). Convergent evolutionary processes driven by foraging opportunity in two sympatric morph pairs of Arctic charr with contrasting post-glacial origins, *Biological Journal of the Linnean Society* **106**(4): 794–806. 15

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing, *arXiv preprint* **7**(arXiv:1207.3907v2 [q-bio.GN]). 6

Gíslason, D. (1998). *Genetic and Morphological Variation in Polymorphic Arctic Charr* Salvelinus Alpinus*, from Icelandic Lakes*, Master's thesis, The University of Guelph. 2, 4, 13, 14

Gkogkas, C. G., Khoutorsky, A., Ran, I., Rampakakis, E., Nevarko, T., Weatherill, D. B., Vasuta, C., Yee, S., Truitt, M., Dallaire, P., Major, F., Lasko, P., Ruggero, D., Nader, K., Lacaille, J.-C. and Sonenberg, N. (2012). Autism-related deficits via dysregulated eIF4E-dependent translational control, *Nature* **493**: 371–377. 17

Gordeeva, N. V., Alekseyev, S. S., Matveev, A. N., Samusenok, V. P. and Taylor, E. (2015). Parallel evolutionary divergence in Arctic char *Salvelinus alpinus* complex from Transbaikalia: Variation in differentiation degree and segregation of genetic diversity among sympatric forms, *Canadian Journal of Fisheries and Aquatic Sciences* **72**(1): 96–115. 3, 15

Gorodilov, Y. N. (1996). Description of the early ontogeny of the Atlantic salmon, *Salmo salar*, with a novel system of interval (state) identification, *Environmental Biology of Fishes* **47**(2): 109–127. 5

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics, *Molecular Ecology Notes* **5**(1): 184–186. 7

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome., *Nature biotechnology* **29**(7): 644–52. 5

Grant, B. and Grant, P. R. (2008). Fission and fusion of Darwin's finches populations, *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1505): 2821–2829. 2

Grant, P. R. and Grant, B. R. (2002). Unpredictable Evolution in a 30-Year Study of Darwin's Finches, *Science* **296**(5568): 707–711. 1

Guðbrandsson, J., Franzdóttir, S. R., Kristjánsson, B. K., Ahi, E. P., Maier, V. H., Kapralova, K. H., Snorrason, S. S., Jónsson, Z. O. and Pálsson, A. (2018). Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs, *PeerJ* **6**: e4345. 2, 3, 4, 5, 6, 7, 8, 10, 13, 14, 17

Gudbrandsson, J., Ahi, E. P., Franzdottir, S. R., Kapralova, K. H., Kristjansson, B. K., Steinhaeuser, S. S., Maier, V. H., Johannesson, I. M., Snorrason, S. S., Jonsson, Z. O. and Palsson, A. (2016). The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations], *F1000Research* **4**(136). 4, 10, 13, 16

Gudmundsdóttir, R., Kornobis, E., Kristjánsson, B. K. and Pálsson, S. (2017). Genetic analysis of ciliates living on the groundwater amphipod *Crangonyx islandicus* (Amphipoda: Crangonyctidae), *Acta Zoologica* **99**(2): 188–198. 14

Guerrero, R. F. and Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism, *Molecular Ecology* **26**(20): 5362–5368. 2

Haas, B. J. (2015). Trinotate: Transcriptome Functional Annotation and Analysis. 5

Haas, B. J. and Papanicolaou, A. (2015). TransDecoder (Find Coding Regions Within Transcripts). 6

Hahsler, M. and Nagar, A. (2017). rBLAST: Interface to the Basic Local Alignment Search Tool (BLAST). 7

Hale, M. C., Thrower, F. P., Berntson, E. A., Miller, M. R. and Nichols, K. M. (2013). Evaluating Adaptive Divergence Between Migratory and Nonmigratory Ecotypes of a Salmonid Fish, *Oncorhynchus mykiss*, *G3: Genes|Genomes|Genetics* **3**(8): 1273–1285. 16

Harris, L. N., Chavarie, L., Bajno, R., Howland, K. L., Wiley, S. H., Tonn, W. M. and Taylor, E. B. (2015). Evolution and origin of sympatric shallow-water morphotypes of Lake Trout, *Salvelinus namaycush*, in Canada's Great Bear Lake, *Heredity* **114**(1): 94–106. 16

Hartl, D. L. and Clark, A. G. (2006). *Principles of Population Genetics, Fourth Edition*, Sinauer Associates, Inc., Sunderland, Massachusetts, USA. 6

He, C., Holme, J. and Anthony, J. (2014). SNP genotyping: The KASP assay., *Methods in molecular biology* **1145**: 75–86. 7

Hendry, A. P., Bolnick, D. I., Berner, D. and Peichel, C. L. (2009). Along the speciation continuum in sticklebacks, *Journal of Fish Biology* **75**(8): 2000–2036. 2

Hindar, K., Ryman, N. and Ståhl, G. (1986). Genetic differentiation among local populations and morphotypes of Arctic charr, *Salvelinus alpinus*, *Biological Journal of the Linnean Society* **27**(3): 269–285. 3

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. and Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags, *PLoS Genetics* **6**(2). 16

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N. V., Alekseyev, S., Hooker, O., Leong, J., Rondeau, E., Koop, B., Adams, C. and Elmer, K. R. (2018). Convergence in form and function overcomes non-parallel evolutionary histories in Arctic charr, *bioRxiv* p. 265272. 3, 15, 16

Jóhannsson, M., Guðbergsson, G., Antonsson, T. and Guðjónsson, S. (1994). Fiskrannsóknir á Úlfljótsvatni 1993 - VMST-S/94001X, *Technical Report VMST-S/94001X*, Veiðimálastofnun, Selfoss. 15

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers, *Bioinformatics* **24**(11): 1403–1405. 7

Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., Absher, D. M., Myers, R. M., Reimchen, T. E., Deagle, B. E., Schluter, D. and Kingsley, D. M. (2012). A Genome-wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-Pair Divergence in Sticklebacks, *Current Biology* **22**(1): 83–90. 2, 16

Jonsson, B., Skúlason, S., Snorrason, S. S., Sandlund, O. T., Malmquist, H. J., Jónasson, P. M., Cydemo, R. and Lindem, T. (1988). Life History Variation of Polymorphic Arctic Charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland, *Canadian Journal of Fisheries and Aquatic Sciences* **45**(9): 1537–1547. 3

Kapralova, K. H. (2008). *Evolutionary Origins of Widespread Small Benthic Charr (*Salvelinus Alpinus*) over Different Geographic Scales in Iceland*, MSc., University of Iceland, Reykjavík, Iceland. 4

Kapralova, K. H. (2014). *Study of Morphogenesis and miRNA Expression Associated with Craniofacial Diversity in Arctic Charr (*Salvelinus Alpinus*) Morphs*, PhD thesis, University of Iceland, Reykjavík, Iceland. 17

Kapralova, K. H., Franzdóttir, S. R., Jónsson, H., Snorrason, S. S. and Jónsson, Z. O. (2014). Patterns of MiRNA Expression in Arctic Charr Development, *PLoS ONE* **9**(8): e106084. 4

Kapralova, K. H., Gudbrandsson, J., Reynisdottir, S., Santos, C. B., Baltanás, V. C., Maier, V. H., Snorrason, S. S. and Palsson, A. (2013). Differentiation at the *MHCIIα* and *Cath2* Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn, *PLoS ONE* **8**(7): e69402. 3, 4, 17

Kapralova, K. H., Jónsson, Z. O., Palsson, A., Franzdóttir, S. R., le Deuff, S., Kristjánsson, B. K. and Snorrason, S. S. (2015). Bones in motion: Ontogeny of craniofacial development in sympatric arctic charr morphs, *Developmental Dynamics* **244**(9): 1168–1178. 3, 4, 13

Kapralova, K. H., Morrissey, M. B., Kristjánsson, B. K., Ólafsdóttir, G. Á., Snorrason, S. S. and Ferguson, M. M. (2011). Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland., *Heredity* **106**(3): 472–87. 2, 4, 14, 15, 17

Klemetsen, A. (2010). The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes, *Freshwater Reviews* **3**(1): 49–74. 2, 15

bioRxiv preprint doi: https://doi.org/10.1101/489104; this version posted July 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Klemetsen, A. (2013). The Most Variable Vertebrate on Earth, *Journal of Ichthyology* **53**(10): 781–791. 2, 16

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J. and Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq, *Molecular Ecology Resources* **14**(2): 381–392. 13

Kornobis, E., Pálsson, S., Kristjánsson, B. K. and Svavarsson, J. (2010). Molecular evidence of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland., *Molecular ecology* **19**(12): 2516–30. 14

Kristjánsson, B. K., Skúlason, S., Snorrason, S. S. and Noakes, D. L. G. (2012). Fine-scale parallel patterns in diversity of small benthic Arctic charr (*Salvelinus alpinus*) in relation to the ecology of lava/groundwater habitats, *Ecology and evolution* **2**(6): 1099–112. 2, 14

Krueger, F. (2012). Trim Galore! 5

Larson, W. A., Limborg, M. T., McKinney, G. J., Schindler, D. E., Seeb, J. E. and Seeb, L. W. (2017). Genomic islands of divergence linked to ecotypic variation in sockeye salmon, *Molecular Ecology* **26**(2): 554–570. 16, 17

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools, *Bioinformatics* **25**(16): 2078–2079. 6

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K. A., Olav Vik, J., Vigeland, M. D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D. R., Yorke, J. A., Nederbragt, A. J., Tooming-Klunderud, A., Jakobsen, K. S., Jiang, X., Fan, D., Hu, Y., Liberles, D. A., Vidal, R., Iturra, P., Jones, S. J. M., Jonassen, I., Maass, A., Omholt, S. W. and Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization, *Nature* **533**(7602): 200–205. 4

Limborg, M. T., Larson, W. A., Shedd, K., Seeb, L. W. and Seeb, J. E. (2018). Novel RAD sequence data reveal a lack of genomic divergence between dietary ecotypes in a landlocked salmonid population, *Conservation Genetics Resources* **10**(2): 169–171. 15

Losos, J. B. and Ricklefs, R. E. (2009). Adaptation and diversification on islands, *Nature* **457**: 830–836. 1

Lowry, D. and Gould, B. (2016). Speciation Continuum, *Encyclopedia of Evolutionary Biology*, Vol. 4, Elsevier, pp. 159–165. 2

Macqueen, D. J. and Johnston, I. A. (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: Biological Sciences* **281**(1778): 20132881. 4

Macqueen, D. J., Kristjánsson, B. K., Paxton, C. G. M., Vieira, V. L. A. and Johnston, I. A. (2011). The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression., *Molecular ecology* **20**(15): 3167–84. 4, 17

Magnusson, K. P. and Ferguson, M. M. (1987). Genetic analysis of four sympatric morphs of Arctic charr, *Salvelinus alpinus*, from Thingvallavatn, Iceland, *Environmental Biology of Fishes* **20**(1): 67–73. 4, 13, 14, 15

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J. and Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake, *Science* **350**(6267): 1493–1498. 2, 16

Malmquist, H. J., Snorrason, S. S., Skulason, S., Jonsson, B., Sandlund, O. T. and Jonasson, P. M. (1992). Diet Differentiation in Polymorphic Arctic Charr in Thingvallavatn, Iceland, *The Journal of Animal Ecology* **61**(1): 21–35. 3

Markevich, G., Esin, E. and Anisimova, L. (2018). Basic description and some notes on the evolution of seven sympatric morphs of Dolly Varden *Salvelinus malma* from the Lake Kronotskoe Basin, *Ecology and Evolution* **8**(5): 2554–2567. 2, 15

Marteinsson, V. T., Rúnarsson, Á., Stefánsson, A., Thorsteinsson, T., Jóhannesson, T., Magnússon, S. H., Reynisson, E., Einarsson, B., Wade, N., Morrison, H. G. and Gaidos, E. (2013). Microbial communities in the subglacial waters of the Vatnajökull ice cap, Iceland, *ISME Journal* **7**: 427–437. 14

May-Mcnally, S. L., Quinn, T. P., Woods, P. J. and Taylor, E. B. (2015). Evidence for genetic distinction among sympatric ecotypes of Arctic char (*Salvelinus alpinus*) in south-western Alaskan lakes, *Ecology of Freshwater Fish* **24**(4): 562–574. 15

Moghadam, H. K., Ferguson, M. M. and Danzmann, R. G. (2011). Whole genome duplication: Challenges and considerations associated with sequence orthology assignment in Salmoninae, *Journal of Fish Biology* **79**(3): 561–574. 6, 8

Morris, M. and Rogers, S. M. (2014). Integrating Phenotypic Plasticity Within an Ecological Genomics Framework: Recent Insights from the Genomics, Evolution, Ecology, and Fitness of Plasticity, *in* C. R. Landry and N. Aubin-Horth (eds), *Ecological Genomics - Ecology and the Evolution of Genes and Genomes*, Springer Netherlands, Dordrecht, chapter 5, pp. 73–105. 2

Mototani, H., Mabuchi, A., Saito, S., Fujioka, M., Iida, A., Takatori, Y., Kotani, A., Kubo, T., Nakamura, K., Sekine, A., Murakami, Y., Tsunoda, T., Notoya, K., Nakamura, Y. and Ikegawa, S. (2005). A functional single nucleotide polymorphism in the core promoter region of *CALM1* is associated with hip osteoarthritis in Japanese., *Human molecular genetics* **14**(8): 1009–17. 16

Muir, A. M., Hansen, M. J., Bronte, C. R. and Krueger, C. C. (2016). If Arctic charr *Salvelinus alpinus* is 'the most diverse vertebrate', what is the lake charr *Salvelinus namaycush* ?, *Fish and Fisheries* **17**(4): 1194–1207. 2, 15

Muschick, M., Indermaur, A. and Salzburger, W. (2012). Convergent Evolution within an Adaptive Radiation of Cichlid Fishes, *Current Biology* **22**(24): 2362–2368. 1

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., Ffrench-Constant, R. H., Blaxter, M. L., Mallet, J. and Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing, *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1587): 343–353. 15

Noakes, D. L. G. (2008). Charr truth: Sympatric differentiation in *Salvelinus* species, *Environmental Biology of Fishes* **83**(1): 7–15. 2, 15

Norðdahl, H., Ingólfsson, Ó., Pétursson, H. G. and Hallsdóttir, M. (2008). Late Weichselian and Holocene environmental history of Iceland, *Jökull* **58**: 343–364. 14, 15

Nosil, P. (2012). *Ecological Speciation*, Oxford Series in Ecology and Evolution, Oxford University Press, Oxford ; New York. 1

Nugent, C. M., Easton, A. A., Norman, J. D., Ferguson, M. M. and Danzmann, R. G. (2017). A SNP Based Linkage Map of the Arctic Charr (*Salvelinus alpinus*) Genome Provides Insights into the Diploidization Process After Whole Genome Duplication, *G3: Genes|Genomes|Genetics* **7**(2): 543–556. 4, 16

Nurse, P. and Thuriaux, P. (1980). Regulatory genes controlling mitosis in the fission yeast *Schizosaccharomyces pombe*, *Genetics* **96**(3): 627–637. 17

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P. and Morgan, M. (2014). VariantAnnotation: A Bioconductor package for exploration and annotation of genetic variants, *Bioinformatics* **30**(14): 2076–2078. 7

Palsson, A. and Gibson, G. (2004). Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*, *Genetics* **167**(3): 1187–1198. 17

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach, *Bioinformatics* **26**(3): 419–420. 7

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O. and Jiggins, C. D. (2012). Adaptive Introgression across Species Boundaries in *Heliconius* Butterflies, *PLoS Genetics* **8**(6): e1002752. 2

Parsons, K. J., Sheets, H. D., Skúlason, S. and Ferguson, M. M. (2011). Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (*Salvelinus alpinus*), *Journal of Evolutionary Biology* **24**(8): 1640–1652. 3

Parsons, K. J., Skúlason, S. and Ferguson, M. (2010). Morphological variation over ontogeny and environments in resource polymorphic arctic charr (*Salvelinus alpinus*), *Evolution and Development* **12**(3): 246–257. 3, 13

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., Anderson, E. C., Rundio, D. E., Williams, T. H., Naish, K. A., Moen, T., Liu, S., Kent, M., Minkley, D. R., Rondeau, E. B., Brieuc, M. S. O., Sandve, S. R., Miller, M. R., Cedillo, L., Baruch, K., Hernandez, A. G., Ben-Zvi, G., Shem-Tov, D., Barad, O., Kuzishchin, K., Garza, J. C., Lindley, S. T., Koop, B. F., Thorgaard, G. H., Palti, Y. and Lien, S. (2018). Sex-dependent dominance maintains migration supergene in rainbow trout, *bioRxiv* . 4

Pease, J. B., Haak, D. C., Hahn, M. W. and Moyle, L. C. (2016). Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation, *PLOS Biology* **14**(2): e1002379. 2

Perreault-Payette, A., Muir, A. M., Goetz, F., Perrier, C., Normandeau, E., Sirois, P. and Bernatchez, L. (2017). Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior, *Molecular Ecology* **26**(6): 1477–1497. 16

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., Hotaling, T. J., Spidle, A. P. and Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation, *Science Advances* **3**(8): e1603198. 17

R Core Team (2015). R: A Language and Environment for Statistical Computing. 6, 7

Roberts, A. and Pachter, L. (2012). Streaming fragment assignment for real-time analysis of sequencing experiments, *Nature Methods* **10**(1): 71–73. 6

Saemundsson, K. (1992). Geology of the Thingvallavatn Area, *Oikos* **64**(1/2): 40–68. 3, 14, 15

Samy, J. K. A., Mulugeta, T. D., Nome, T., Sandve, S. R., Grammes, F., Kent, M. P., Lien, S. and Våge, D. I. (2017). SalmoBase: An integrated molecular data resource for Salmonid species, *BMC Genomics* **18**(1): 482. 6

Sandlund, O. T., Gunnarsson, K., Jónasson, P. M., Jonsson, B., Lindem, T., Magnússon, K. P., Malmquist, H. J., Sigurjónsdóttir, H., Skúlason, S. and Snorrason, S. S. (1992). The arctic charr *Salvelinus alpinus* in Thingvallavatn, *Oikos* **64**(1/2): 305–351. 3, 13, 15

Sandlund, O. T., Jonsson, B., Malmquist, H. J., Gydemos, R., Lindem, T., Skúlason, S., Snorrason, S. S. and Jónasson, P. M. (1987). Habitat use of arctic charr *Salvelinus alpinus* in Thingvallavatn, Iceland, *Environmental Biology of Fishes* **20**(4): 263–274. 3

Sävilammi, T., Primmer, C. R., Varadharajan, S., Guyomard, R., Guiguen, Y., Sandve, S. R., Vøllestad, L. A., Papakostas, S. and Lien, S. (2019). The Chromosome-Level Genome Assembly of European Grayling Reveals Aspects of a Unique Genome Evolution Process Within Salmonids, *G3: Genes|Genomes|Genetics* **9**(5): 1283–1294. 4

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., Lindholm, A. K., Lucek, K., Maan, M. E., Marques, D. A., Martin, S. H., Matthews, B., Meier, J. I., Möst, M., Nachman, M. W., Nonaka, E., Rennison, D. J., Schwarzer, J., Watson, E. T., Westram, A. M. and Widmer, A. (2014). Genomics and the origin of species, *Nature Reviews Genetics* **15**(3): 176–192. 2, 4, 16

Seehausen, O. and Wagner, C. E. (2014). Speciation in Freshwater Fishes, *Annual Review of Ecology, Evolution, and Systematics* **45**(1): 621–651. 1

Seppala, M., Depew, M. J., Martinelli, D. C., Fan, C. M., Sharpe, P. T. and Cobourne, M. T. (2007). *Gas1* is a modifier for holoprosencephaly and genetically interacts with sonic hedgehog, *Journal of Clinical Investigation* **117**: 1575–1584. 17

Shapiro, B. J., Leducq, J.-B. and Mallet, J. (2016). What Is Speciation?, *PLOS Genetics* **12**(3): e1005860. 16

Skúlason, S., Antonsson, T., Gudbergsson, G., Malmquist, H. J. and Snorrason, S. S. (1992). Varibility in Icelandic Arctic charr, *Icelandic Agricultural Sciences* **6**: 143–153. 2

Skúlason, S., Noakes, D. L. and Snorrason, S. S. (1989). Ontogeny of trophic morphology in four sympatric morphs of arctic charr *Salvelinus alpinus* in Thingvallavatn, Iceland, *Biological Journal of the Linnean Society* **38**(3): 281–301. 3, 13

Skúlason, S. and Smith, T. B. (1995). Resource polymorphisms in vertebrates, *Trends in Ecology & Evolution* **10**(9): 366–370. 2, 3

Skúlason, S., Snorrason, S., Noakes, D. and Ferguson, M. (1996). Genetic basis of life history variations among sympatric morphs of Arctic char *Salvelinus alpinus*, *Canadian Journal of Fisheries and Aquatic Sciences* **53**(8): 1807–1813. 3

Skúlason, S., Snorrason, S., Ota, D. and Noakes, D. (1993). Genetically based differences in foraging behaviour among sympatric morphs of arctic charr (Pisces: Salmonidae), *Animal Behaviour* **45**(6): 1179–1192. 3

Skúlason, S., Snorrason, S. S., Noakes, D. L. G., Ferguson, M. M. and Malmquist, H. J. (1989). Segregation in spawning and early life history among polymorphic Arctic charr, *Salvelinus alpinus*, in Thingvallavatn, Iceland, *Journal of Fish Biology* **35**(Supplement A): 225–232. 3, 4, 15

Smith, T. B. and Skúlason, S. (1996). Evolutionary Significance of Resource Polymorphisms in Fishes, Amphibians, and Birds, *Annual Review of Ecology and Systematics* **27**: 111–133. 2, 15

Snorrason, S. S. and Skúlason, S. (2004). Adaptive Speciation in Northern Freshwater Fishes, *in* U. Dieckmann, M. Doebeli, J. A. J. Metz and D. Tautz (eds), *Adaptive Speciation*, Cambridge University Press, Cambridge, chapter 10, pp. 210–228. 2

Snorrason, S. S., Skúlason, S., Sandlund, O. T., Malmquist, H. J., Jonsson, B. and Jonasson, P. M. (1989). Shape polymorphism in Arctic charr, *Salvelinus alpinus*, *Physiology and Ecology Japan* **1**: 393–404. 13, 15, 17

Sutherland, B. J. G., Gosselin, T., Normandeau, E., Lamothe, M., Isabel, N., Audet, C. and Bernatchez, L. (2016). Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps, *Genome Biology and Evolution* **8**(12): evw262. 4, 16

Svavarsson, E. (2007). Árangur í kynbótum á bleikju og næstu skref, *Fræðaþing Landbúnaðarsins 4*, pp. 121–125. 13

Theis, A., Ronco, F., Indermaur, A., Salzburger, W. and Egger, B. (2014). Adaptive divergence between lake and stream populations of an East African cichlid fish, *Molecular Ecology* **23**(21): 5304–5322. 2

Timusk, E. R., Ferguson, M. M., Moghadam, H. K., Norman, J. D., Wilson, C. C. and Danzmann, R. G. (2011). Genome evolution in the fish family salmonidae: Generation of a brook charr genetic map and comparisons among charrs (Arctic charr and brook charr) with rainbow trout., *BMC genetics* **12**: 68. 4

Varadharajan, S., Sandve, S. R., Gillard, G., Tørresen, O. K., Mulugeta, T., Hvidsten, T. R., Lien, S., Vollestad, L. A., Jentoft, S., Nederbragt, A. J. and Jakobsen, K. S. (2018). The grayling genome reveals selection on gene expression regulation after whole genome duplication, *Genome Biology and Evolution* **10**(10): 2785–2800. 4

Verspoor, E., Knox, D., Greer, R. and Hammar, J. (2010). Mitochondrial DNA variation in Arctic charr (*Salvelinus alpinus* (L.)) morphs from Loch Rannoch, Scotland: Evidence for allopatric and peripatric divergence, *Hydrobiologia* **650**(1): 117–131. 3

Via, S. (2001). Sympatric speciation in animals: The ugly duckling grows up, *Trends in Ecology & Evolution* **16**(7): 381–390. 1

Via, S. (2009). Natural selection in action during speciation, *Proceedings of the National Academy of Sciences* **106**(Supplement_1): 9939–9946. 1

Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H. and Wolf, J. B. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa, *Molecular Ecology* **26**(16): 4284–4295. 2, 16

Vincent, B., Dionne, M., Kent, M. P., Lien, S. and Bernatchez, L. (2013). Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation, *Evolution* **67**(12): 3469–3487. 16

Volpe, J. P. and Ferguson, M. M. (1996). Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland., *Molecular ecology* **5**(6): 763–72. 4, 13, 14, 15, 17

Wilson, A. J., Gíslason, D., Skúlason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G. and Ferguson, M. M. (2004). Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales., *Molecular ecology* **13**(5): 1129–42. 2, 15

Wolf, J. B. W. and Ellegren, H. (2016). Making sense of genomic islands of differentiation in light of speciation, *Nature Reviews Genetics* **18**(2): 87–100. 2, 4

Woods, P. J., Skúlason, S., Snorrason, S. S., Kristjánsson, B. K., Malmquist, H. J. and Quinn, T. P. (2012a). Intraspecific diversity in Arctic charr, *Salvelinus alpinus*, in Iceland: I. Detection using mixture models, *Evolutionary Ecology Research* **14**: 973–992. 15

Woods, P. J., Skúlason, S., Snorrason, S. S., Kristjánsson, B. K., Malmquist, H. J. and Quinn, T. P. (2012b). Intraspecific diversity in Arctic charr, *Salvelinus alpinus*, in Iceland: II. Which environmental factors influence resource polymorphism in lakes?, *Evolutionary Ecology Research* **14**(8): 993–1013. 2

Yan, J., Shah, T., Warburton, M. L., Buckler, E. S., McMullen, M. D. and Crouch, J. (2009). Genetic Characterization and Linkage Disequilibrium Estimation of a Global Maize Collection Using SNP Markers, *PLoS ONE* **4**(12): e8451. 12

Young, M. D., Wakefield, M. J., Smyth, G. K. and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias., *Genome biology* **11**(2): R14. 7

Zimmerman, E., Palsson, A. and Gibson, G. (2000). Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*., *Genetics* **155**(2): 671–83. 17

Supplementary Material

Supplementary Tables and Files

Table S1: Tab-delimited text file with the position, alleles, alternative frequencies within morphs, F-statistics between morphs and samples and predicted effect of the variant on protein composition for all the variants after the final filtering step.

On figshare: doi:10.6084/m9.figshare.8705888

id: Identifier for each variant from 1 to 19,252

transcript_id: Name of the transcript from the Trinity assembly.

start: Start position of polymorphism within contig.

end: End position of polymorphism within contig.

ref: Base(s) of the reference allele.

alt: Base(s) of the alternative allele.

var_pos: Position of polymorphism within contig.

NCBI_id: Sequence identifier for chromosome or scaffold in genome.

Chr: Name of chromosome.

Chr_pos: Position mapped to in chromosome.

Freq_LB: Frequency of alternative allele within the LB morph.

Freq_PL: Frequency of alternative allele within the PL morph.

Freq_SB: Frequency of alternative allele within the SB morph.

Fst: Estimate of F_{ST} between morphs.

Fpt: Estimate of F_{PT} or variation among samples (pools). See methods for further explanation.

effect: Which effect does the mutation have on the gene, 3'- and 5' UTR indicate mutations in those regions, synonymous do not change the reading frame but mutations that do so are coded as transitions from one amino acid to another (single letter a.a. code)

gene_name: Gene name based on SalmonBase annotation.

Table S2: VCF-file with the final set of variants after all filtering steps.

On figshare: doi:10.6084/m9.figshare.8718695

28

Table S3: Variants only found within one morph. Position within contig and genome, alleles, frequency within the morph, F_{ST}-values and predicted effect are shown.

Transcript id	Var pos	Ref	Alt	Chr	Chr pos	Freq	F_{ST}	F_{PT}	Effect	Gene name
Private to LB										
T143531c2g1i1	545	G	A	LG2	1,668,110	0.33	0.19	0.32		CIGSSA_083607.t3
T147645c2g1i1	571	T	A	LG18	49,590,368	0.34	0.24	0.39		AN1-type zinc finger protein 4-like, isoform X1
T149060c0g1i1	83	T	A	LG4q.1:29	13,785,703	0.46	0.16	0.41	UTR	Prenylcysteine oxidase
T149651c0g1i1	420	C	G	LG18	47,847,696	0.34	0.21	0.43	UTR	40S ribosomal protein S24, isoform X3
T161481c0g1i2	203	C	T	LG13	50,967,669	0.45	0.30	0.68	UTR	Ubiquitin-conjugating enzyme E2 D2|predicted protein
T161641c2g2i1	1119	G	A	LG23	40,674,682	0.37	0.24	0.34	UTR	Low-density lipoprotein receptor-related protein 6 |hypothetical weakly similar
T170659c2g1i1	434	T	A	LG20	55,172,526	0.34	0.19	0.40	UTR	
T171446c2g1i3	159	C	G			0.48	0.15	0.55		RNA-binding protein Nova-1-like, isoform X1
T174484c7g1i1	742	C	T	LG23	12,588,360	0.32	0.23	0.40	Syn	TRPM8 channel-associated factor homolog, isoform X2
T174834c7g3i2	629	G	C	LG4p	9,417,520	0.38	0.25	0.60	T to R	ADP-ribosylation factor 1 like
T175088c0g3i1	698	T	C	LG18	9,241,944	0.33	0.19	0.46	UTR	protein FAM46A-like
T175804c1g1i1	766	C	G	LG31	15,851,598	0.32	0.17	0.31	R to G	Cytochrome c-type heme lyase
T193474c0g1i1	1331	G	A	LG14	38,002,543	0.33	0.19	0.46	Syn	
Private to PL										
T44119c0g1i1	152	A	C	LG17	8,115,846	0.47	0.31	0.48	UTR	Glutamine–tRNA ligase
T70707c0g1i1	1085	T	C	LG14	14,039,993	0.47	0.33	0.48	UTR	Costars family protein ABRACL
T70707c0g1i1	1506	T	C	LG14	14,040,414	0.43	0.31	0.45	UTR	Costars family protein ABRACL
T83913c0g1i1	419	G	A	LG18	59,124,877	0.55	0.38	0.67	D to N	Adenosine kinase
T83913c0g1i1*	1075	C	T	LG18	59,122,320	0.66	0.60	0.73		Adenosine kinase
T106859c0g1i1	366	C	T	LG14	14,052,297	0.46	0.25	0.47		beta-taxilin, isoform X2
T140173c0g1i2	148	C	T	Scf2193	34,736	0.41	0.35	0.38	UTR	N-myc proto-oncogene protein-like, isoform X2
T148137c3g1i2	1158	G	C	Scf16416	190,147	0.45	0.39	0.56	UTR	dolichyldiphosphatase 1-like
T148137c3g1i2	1432;1440	G(N7)T	T(N7)A	Scf16416	189,864	0.44	0.30	0.50	UTR	dolichyldiphosphatase 1-like
T154282c0g1i1	142	C	G	LG17	8,115,335	0.44	0.32	0.48		Glutamine–tRNA ligase
T154282c0g1i1	252;292;304	G(N39)A-(N11)A	A(N39)T-(N11)C	LG17	8,115,195	0.52	0.37	0.58		Glutamine–tRNA ligase
T154282c0g1i1	575	G	T	LG17	8,114,902	0.41	0.28	0.43		Glutamine–tRNA ligase
T156040c0g1i1	747	C	A	LG4q.1:29	23,198,437	0.37	0.22	0.46	UTR	Vacuolar protein sorting-associated protein 29
T159843c0g1i1	1331	G	A	LG31	24,512,078	0.40	0.26	0.39	UTR	chromobox protein homolog 3-like, isoform X1
T159843c0g1i1	1796	G	T	LG31	24,512,546	0.58	0.44	0.55	UTR	chromobox protein homolog 3-like, isoform X1
T161458c2g1i1	77	A	G	LG6.2	13,677,880	0.34	0.22	0.54	UTR	cyclin A2, isoform X2
T161984c0g1i1	618	C	T	LG31	22,464,046	0.33	0.20	0.33	Syn	Trafficking protein particle complex subunit 3, isoform X1
T162197c14g1i1	1353	G	T	LG7	16,955,663	0.53	0.39	0.54	Syn	Protein LZIC
T163256c7g7i3	3892	T	A	Scf1703	56,973	0.54	0.40	0.51	UTR	CIGSSA_123112.t11
T163256c7g7i3	4065	A	C	Scf1703	56,800	0.47	0.34	0.56	UTR	CIGSSA_123112.t11
T163989c5g1i1	1293	T	C	LG7	27,364,962	0.36	0.11	0.52	UTR	fibrous sheath CABYR-binding protein-like, isoform X1
T165162c16g1i1	820	G	A	LG1	17,578,711	0.34	0.15	0.26	UTR	TOX high mobility group box family member 2-like, isoform X4
T165162c16g1i1	974	C	T	LG1	17,578,865	0.32	0.14	0.31	UTR	TOX high mobility group box family member 2-like, isoform X4

Continued on Next Page…

29

Table S3: Variants only found within one morph. Position within contig and genome, alleles, frequency within the morph, F_{ST}-values and predicted effect are shown.

Transcript id	Var pos	Ref	Alt	Chr	Chr pos	Freq	F_{ST}	F_{PT}	Effect	Gene name
T165162c16g1i1	975;984	A(N8)A	T(N8)C	LG1	17,578,870	0.33	0.13	0.38	UTR	TOX high mobility group box family member 2-like, isoform X4
T168092c2g1i5	2724;2726	TTA	ATT	LG7	26,670,845	0.35	0.23	0.51	UTR	45 kDa calcium-binding protein
T168398c0g1i4	1759	G	T	LG32	20,522,655	0.33	0.28	0.35	P to H	Kinesin-like protein
T168651c0g1i1	863	A	G	LG1	17,539,258	0.36	0.30	0.40		alpha-1-syntrophin-like, isoform X1
T168962c1g1i2	994	G	C	LG14	7,560,825	0.35	0.24	0.38	S to T	translocating chain-associated membrane protein 2-like
T169544c12g4i1	442	A	C	LG41	294,246	0.48	0.34	0.53	UTR	CIGSSA_056585.t2
T70480c1g1i1	2202	A	T	LG4q.2	9,848,462	0.34	0.16	0.49		CIGSSA_044851.t3
T170529c17g1i1	176	A	T	LG36	28,810,075	0.40	0.31	0.49		integral membrane protein 2B-like, isoform X1
T71302c3g1i1	521	C	T	LG17	8,116,418	0.42	0.24	0.39		Glutamine–tRNA ligase
T71326c3g1i2	1206	A	T	LG11	21,367,985	0.33	0.15	0.48	UTR	Peptidylarginine deiminase hypothetical
T71686c7g2i1	193	G	C	LG17	8,312,053	0.49	0.39	0.50		Zgc:110709
T71686c7g2i1	665	G	C	LG17	8,311,589	0.53	0.42	0.58		Zgc:110709
T71686c7g2i1	982	T	A	LG17	8,311,272	0.50	0.43	0.50		Zgc:110709
T71686c7g2i1	1424	T	G	LG17	8,310,831	0.42	0.33	0.46		Zgc:110709
T71162c0g1i1	425	T	A	LG7	16,736,407	0.47	0.35	0.54		TAR DNA binding protein
T72217c4g4i1	1238	T	A	LG32	6,053,989	0.53	0.36	0.57	UTR	WW domain-containing adapter protein with coiled-coil
T72217c4g4i1	2113	A	G	LG32	6,054,862	0.58	0.47	0.58	UTR	WW domain-containing adapter protein with coiled-coil
T72357c4g1i1	2197	A	C			0.42	0.41	0.38	Syn,UTR	coronin-7-like
T72967c3g1i1	199	G	C	LG35	8,330,659	0.51	0.35	0.51		RAB11a member RAS oncogene family
T72992c11g4i1	448	G	C	LG14	14,078,902	0.57	0.44	0.60	Syn	Beta-taxilin|hypothetical
T72933c13g1i1	2089	C	T	LG7	16,959,001	0.48	0.33	0.47		calsyntenin 1, isoform X4
T73769c10g1i2	742;743;748	TATTTTT	ATTTTTG	Scf1321	226,642	0.34	0.18	0.49		Inosine-5-monophosphate dehydrogenase 1
T174499c6g3i3	1150	A	G	LG7	16,917,732	0.50	0.38	0.53	S to G	ubiquitination factor E4B, UFD2 homolog (S. cerevisiae), isoform X1
T174814c3g1i1	411	T	A	LG4q.1:29	37,790,310	0.33	0.24	0.47		CMP-N-acetylneuraminate-poly-alpha-28-sialyltransferase
T174821c6g1i1	967	C	G	LG36	31,081,180	0.41	0.35	0.40		unconventional myosin-Ib-like, isoform X1
T174821c6g1i1	1199	G	A	LG36	31,080,948	0.37	0.23	0.36		unconventional myosin-Ib-like, isoform X1
T175009c8g1i2	2905	C	T	LG30	8,090,556	0.47	0.34	0.45	UTR	ATP-binding cassette sub-family F member 1-like
T175236c1g3i4	5881	T	G	LG31	24,323,360	0.61	0.45	0.56	UTR	collagen alpha-1(XVI) chain-like, isoform X1
T175348c1g10i1	258	A	G	LG20	21,264,718	0.32	0.18	0.40		ubiquitin domain containing 2
T175403c90g7i1	1245	A	C	Scf396	147,836	0.73	0.55	0.83	UTR	peptidyl-prolyl cis-trans isomerase FKBP3-like
T175479c1g2i5	407	A	C	LG35	8,345,621	0.50	0.38	0.50	V to G	Ubiquilin-4
T176050c1g1i1	5019	C	A	LG35	8,316,952	0.52	0.44	0.51	UTR	RNA-binding protein MEX3A
T210338c0g1i1	1772	G	A	LG17	8,125,541	0.50	0.36	0.48	Syn	glutaminyl-tRNA synthetase
Private to SB										
T145511c0g1i1	2002	T	C	Scf4114	11,937	0.26	0.26	0.25	Syn	Myelin expression factor 2|hypothetical
T160725c0g1i2	919	A	G	LG17	22,513,650	0.40	0.41	0.37	Syn	transcription factor 15-like
T162598c0g1i1	729;742	G(N12)C	A(N12)T	Scf1046	1,457	0.44	0.23	0.44		chromobox homolog 6
T165516c2g3i4	382	C	A	Scf4168	5,044	0.27	0.19	0.46	UTR	
T165516c2g3i4	873;876;912;924;927	CCAC(N35)A-(N10)AAAT	TCAT(N35)T-(N10)GAAC	LG18	24,320,993	0.43	0.38	0.65	Syn	
T68835c3g2i2	1142	T	A	LG4q.1:29	21,965,763	0.33	0.33	0.37	UTR	Phosphoglycerate mutase 2
T169028c43g1i1	3476	A	C	LG32	2,199,330	0.31	0.29	0.40	UTR	Muscle-related coiled-coil protein

Table S3: Variants only found within one morph. Position within contig and genome, alleles, frequency within the morph, F_{ST}-values and predicted effect are shown.

Transcript id	Var pos	Ref	Alt	Chr	Chr pos	Freq	F_{ST}	F_{PT}	Effect	Gene name
T169667c0g1i4	2617	A	G	LGI8	12,429,904	0.28	0.30	0.32	F to S	DNA topoisomerase 2-alpha
T169667c0g1i4	2835	G	A	LGI8	12,430,772	0.28	0.31	0.32	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3054	G	T	LGI8	12,431,454	0.28	0.30	0.30	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3198	A	G	LGI8	12,431,726	0.26	0.30	0.30	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3375	A	G	LGI8	12,432,251	0.29	0.31	0.31	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3447	G	A	LGI8	12,432,602	0.29	0.31	0.36	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3672	T	C	LGI8	12,433,037	0.28	0.32	0.32	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	4308	C	A	LGI8	12,435,579	0.27	0.29	0.32	Syn	DNA topoisomerase 2-alpha
T170294c11g5i4	1744	T	C	LGI7	22,247,612	0.33	0.31	0.32	Syn	ras-related protein Rap-1A-like, isoform X1
T173826c6g1i1	342	A	G	LG28	18,758,190	0.36	0.32	0.43	Syn	Si:dkey-77p13.2
T210331c21g1i1	4709	G	A	MG	3,411	0.74	0.68	0.79	UTR	

*Validated in KASP-assay

Table S4: Gene ontology categories enriched in variants private to PL-charr ($p < 0.01$ in the other morphs) on the transcripts (tr) and gene (ge) level. The number of transcripts and genes with observed private variants (PL_{tr} and PL_{ge}) and the total number of transcripts and genes tested in the category (Tot_{tr} and Tot_{ge}) are shown. The multiple testing corrected p-value or false discovery rate (FDR) is also shown for both levels.

Category	Term	PL_{tr}	Tot_{tr}	FDR_{tr}	PL_{ge}	Tot_{ge}	FDR_{ge}
GO:0006425	glutaminyl-tRNA aminoacylation	4	4	2.47e-06	1	1	1.0000
GO:0006433	prolyl-tRNA aminoacylation	4	4	2.47e-06	1	1	1.0000
GO:0006424	glutamyl-tRNA aminoacylation	4	6	2.44e-05	1	3	1.0000

Table S5: Genetic polymorphisms studied in the population sample of Arctic charr.

On figshare: doi:10.6084/m9.figshare.8719784

Transcript_ID: Name of the transcript from the Trinity assembly
Variant: Name of variant.
Gene: Short name of gene.
Gene_name: Long name of gene.
Start: Position of polymorphism within contig.
Ref: Base(s) of the reference allele.
Alt: Base(s) of the alternative allele.
Outg: Base(s) in the salmonid outgroup.
Chr_NCBI_id: Sequence identifier for chromosome or scaffold in genome.
Chr: Name of chromosome.
Chr_position: Position mapped to in chromosome.
Effect: Which effect does the mutation have on the gene, 3'- and 5' UTR indicate mutations in those regions, synonymous do not change the reading frame but mutations that do so are coded as transitions from one amino acid to another (single letter a.a. code)
Marker_ID: Abbreviated ID of each marker.
Sequence: The sequence used to design the KASP assay, the polymorphism are marked by e.g. [A/T].

Table S6: Datafile of the genotypes for the 22 markers scored in the population samples from the four sympatric charr morphs (coded by bases, missing data indicated by "NA").

On figshare: doi:10.6084/m9.figshare.8721317

Table S7: Estimates of F-statistics for the entire KASP data and tests of Hardy-Weinberg proportions for the entire KASP dataset (Total) and individual morphs.

Variant	Total (P)	SB (P)	LB (P)	PL (P)	PI (P)	F_{ST}	F_{IS}	F_{IT}
eif4g2b_G652A	0.0	1.0	1.0	0.546	0.400	0.47	0.15	0.55
timp2b_G355A	0.076	1.0	0.039	0.488	1.0	0.10	0.14	0.23
tcf15_C1080T	0.0	1.0	0.515	1.0	0.021	0.42	0.14	0.50
lrrc1_G2814A	0.0	1.0	1.0	1.0	1.0	0.54	0.00	0.54
sox21b_T317G	0.001	1.0	1.0	1.0	0.378	0.43	0.05	0.46
calm1_C1027A	0.0	1.0	0.369	1.0	0.638	0.51	0.02	0.52
tmem9b_C649T	1.0	0.436	0.343	1.0	0.619	0.22	-0.20	0.07
pairb_C2016T	0.122	0.546	1.0	1.0	1.0	0.31	-0.09	0.25
Kiaa1324_TC393AA	0.0	1.0	1.0	1.0	0.026	0.46	0.21	0.58
gas1l_A3641C	0.0	1.0	1.0	1.0	1.0	0.74	-0.12	0.71
gabpb1_G238A	0.0	1.0	1.0	1.0	1.0	0.66	-0.10	0.62
msi1_A1594T	0.002	0.328	0.586	0.373	0.597	0.31	0.10	0.37
cox11_C1202A	0.0	1.0	1.0	1.0	1.0	0.52	-0.01	0.51
gnl3l_G795T	0.013	0.399	1.0	1.0	0.444	0.21	0.14	0.32
eif4ebp2_T1083A	0.410	1.0	1.0	0.291	0.108	0.28	-0.18	0.15
lrp10_A775C	0.0	0.560	1.0	1.0	1.0	0.63	0.03	0.64
neurod1_T1187C	0.001	0.553	0.077	0.426	0.502	0.30	0.15	0.40
dennd5a_A2555T	0.0	1.0	1.0	1.0	1.0	0.69	-0.09	0.66
Sik3_A5236G	0.0	1.0	0.013	1.0	0.091	0.42	0.19	0.53
cdkn2a_A257G	0.149	1.0	1.0	1.0	0.022	0.11	0.09	0.19
wee1_T305A	0.0	1.0	1.0	1.0	1.0	0.72	-0.10	0.69
adk_C1075T	0.0	0.336	1.0	1.0	0.347	0.40	0.18	0.50

Variant: Name of variant

Total (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests, on the entire dataset

SB (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for SB

LB (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for LB

PL (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for PL

PI (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for PI

F_{ST} : F-statistics for variation between morphs

F_{IS} : F-statistics for variation within population

F_{IT} : F-statistics for variation between individuals

Table S8: Datafile of the LD (r^2) for all pairs of markers, by morph

On figshare: doi:10.6084/m9.figshare.8721455
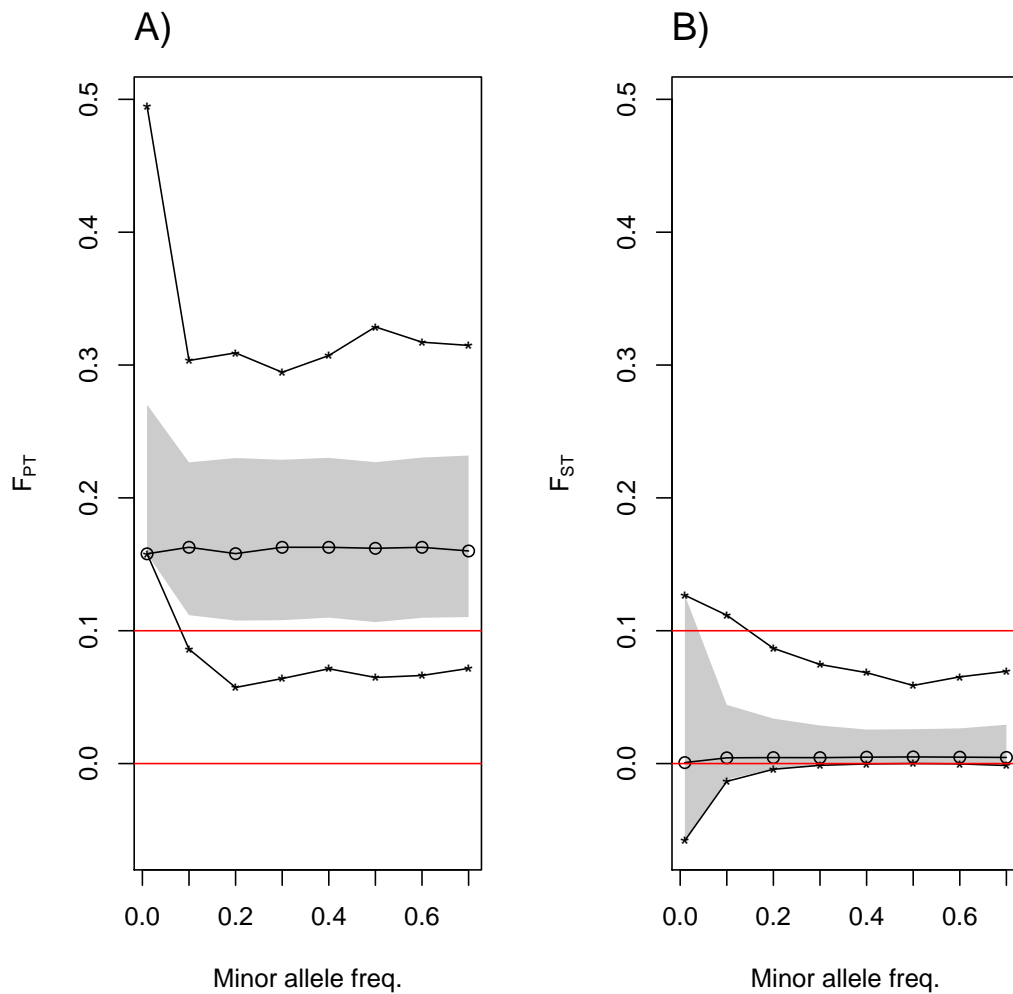
Supplementary Figures

Figure S1: Results from simulations on F-statistics for a biallelic variant with the same allele frequency among morphs. The gray area indicate the 95% confidence area. The median, minimum and maximum values are shown by dots and lines. A) Shows F_{PT} values and B) F_{ST} values. As expected F_{PT} values are high and we chose 0.1 (red horizontal line) as cutoff as it is outside the 95% confidence area in the simulations.
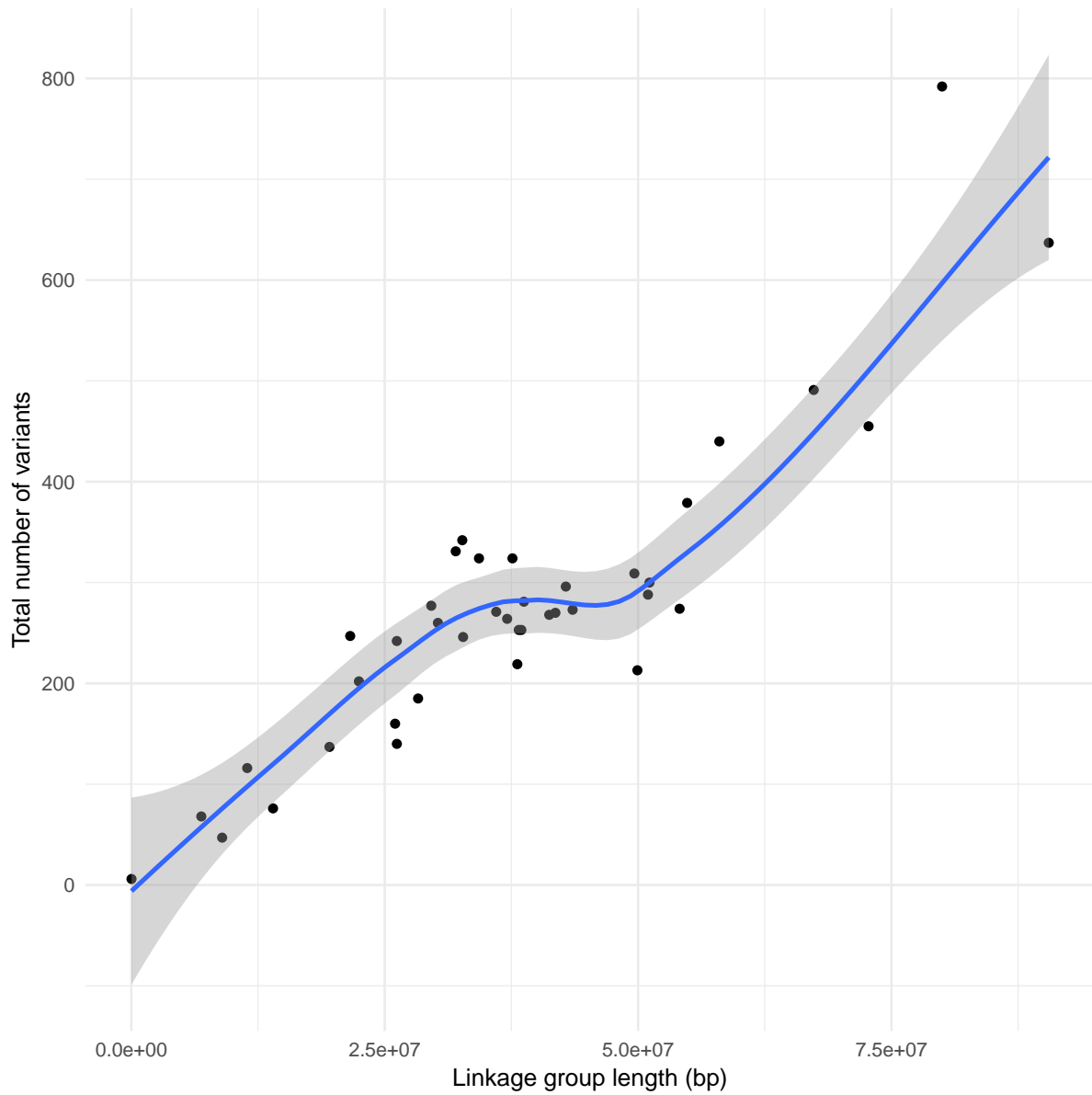
34

Figure S2: Total number of transcriptome variants by linkage group length. A loess smooth curve (blue) with 95% confident interval (gray) is also graphed.

A)

B)
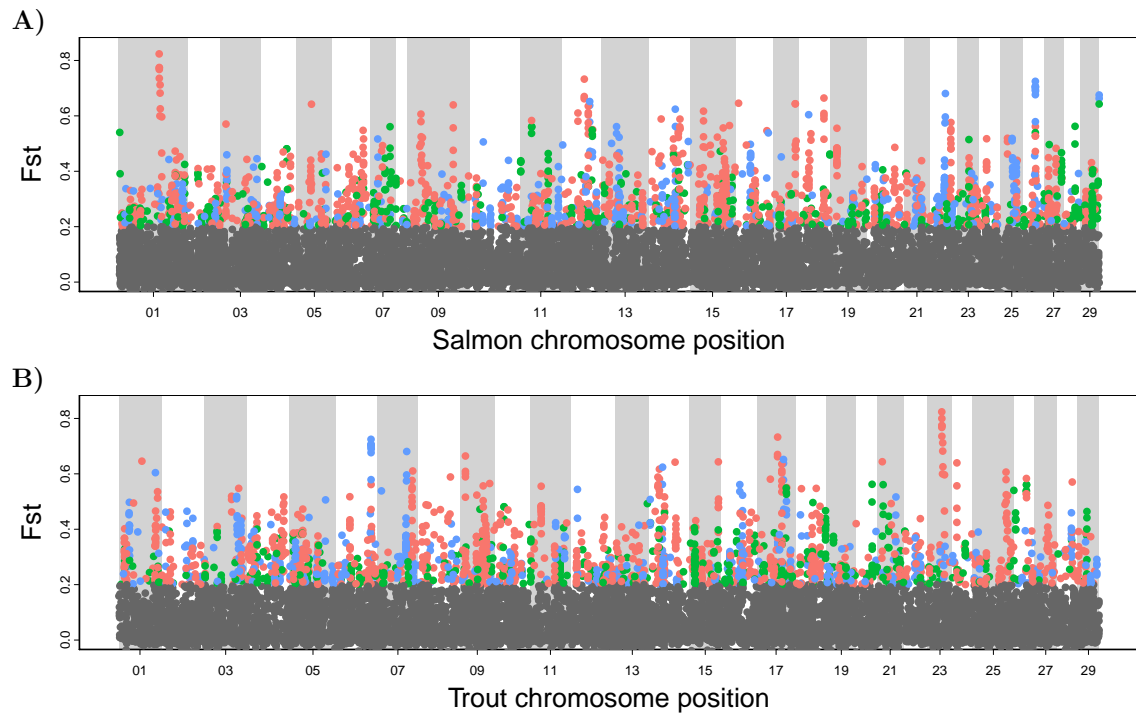
Figure S3: F_{ST} values plotted by position of variants on the salmon **A)** and rainbow trout **B)** genome. The colors indicate which morphs differs most strongly in allele frequency from the other two for variants with $F_{ST} > 0.2$. Red PL, blue SB and green LB
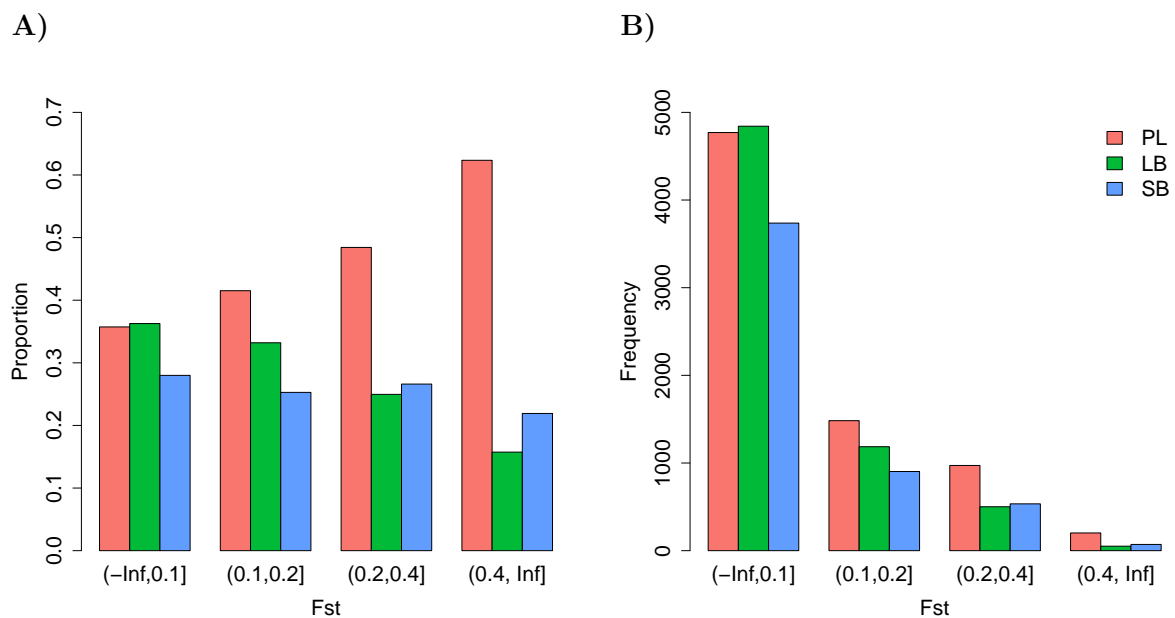
A)

B)

Figure S4: **A)** Proportion of variants (within F_{ST} groups) for different F_{ST}-values grouped by the morph with highest deviation in allele frequency from the other two. **B)** The number of variants in each F_{ST} category grouped by the morph with highest deviation in allele frequency from the other two. The legend in **B)** also applies to **A)**.
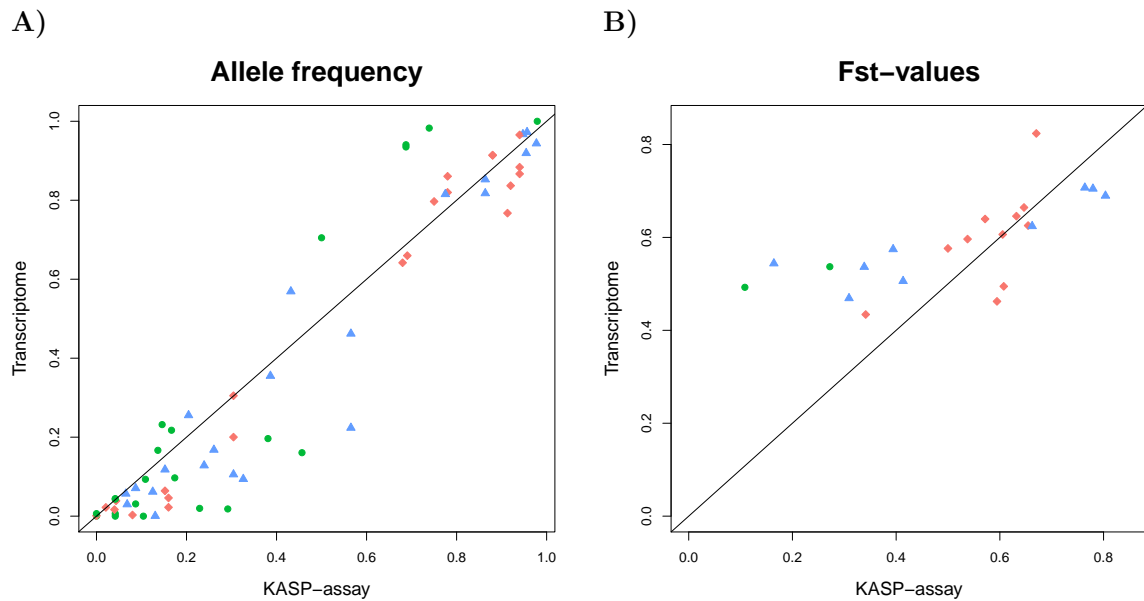
36

Figure S5: Tight relationship was found between the allele frequencies and F_{ST} values estimated from the transcriptome and the population genetic sample (Kasp assay). **A)** The allele frequencies for the 22 markers were estimated for each morph (color coded) and the estimates from the two methods show high positive correlation (Kendall's $\tau = 0.76$, Pearson's $r = 0.96$, $p < 0.0001$). The diagonal line represents the 1:1 relationship. **B)** F_{ST}-values calculated for the transcriptome and the population sample had comparatively weaker association (Kendall's $\tau = 0.58$, Pearsons's $r = 0.70$, $p < 0.001$). Notably, two markers deviated substantially between the two methods (in *timp2b* and *cdkn2a*), with higher F_{ST} in the transcriptome compared to the population sample (due to underestimation of some rarer allele frequencies in the transcriptome). The colors indicate which morphs shows the highest deviation from the other two in mean allele frequency for each marker in the transcriptome (SB: blue, LB: green and PL: red).
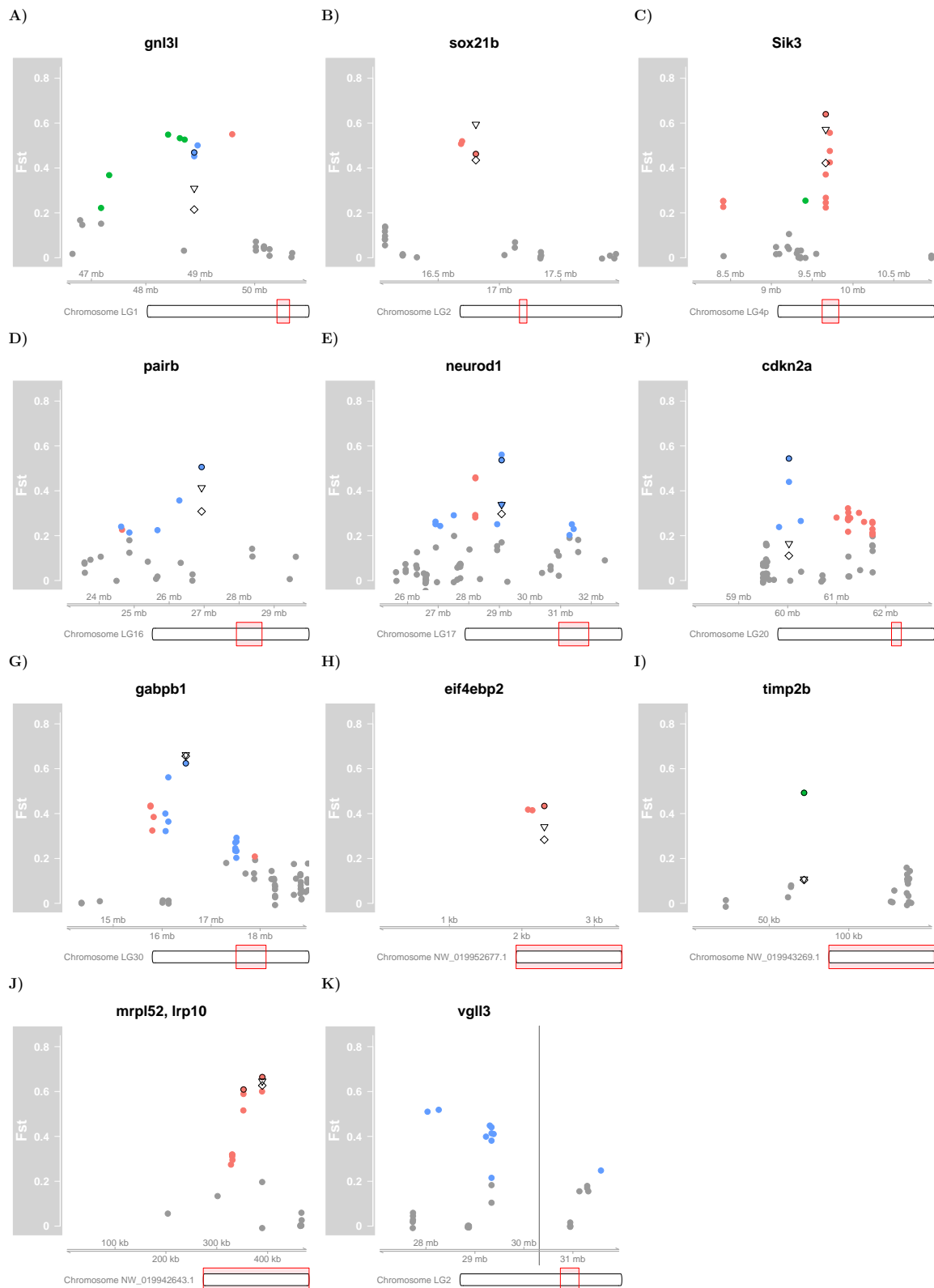
Figure S6: **A-J)** Detailed view of chromosomal regions of variants in KASP-assay and nearby transcriptome variants (not-validated). The colored dots mark F_{ST} values from the transcriptome as in Fig. 3 and the variants taken for validation are marked with a filled black circl ($\circ$). The triangles ($\triangledown$) show the F_{ST} value from the KASP-assay for the three transcriptome morphs (PL, SB and LB) and the diamonds ($\diamond$) F_{ST} for all morphs (including PI). **K)** Detailed view of variants nearby the *vgll3* locus. The vertical line represents the location of the *vgll3* gene, not transcribed and thus no variants detected, using the same color code (SB: blue, LB: green and PL: red) for variants.