The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs

Jóhannes Gudbrandsson[†], Ehsan P. Ahi[†], Kalina. H. Kapralova[†], Sigrídur R. Franzdottir[†], Bjarni K. Kristjánsson[§], S. Sophie Steinhaeuser[†], Ísak M. Jóhannesson[†], Valerie H. Maier[†], Sigurdur S. Snorrason[†], Zophonías O. Jónsson[†], Arnar Pálsson[† *]

[†]Institute of Life and Environmental Sciences
University of Iceland
101 Reykjavik, Iceland
[§] Holar University College
551 Saudarkrokur, Iceland

Abstract

Species showing repeated evolution of similar traits can help illuminate the molecular and developmental basis of diverging traits and specific adaptations. Following the last glacial period, dwarfism and specialized bottom feeding morphology evolved rapidly in several landlocked Arctic charr (*Salvelinus alpinus*) populations in Iceland. In order to study the genetic divergence between small benthic morphs and larger morphs with limnetic morphotype, we conducted an RNA-seq transcriptome analysis of developing charr. We sequenced mRNA from whole embryos at four stages in early development of two stocks with very different morphologies, the small benthic (SB) charr from Lake Thingvallavatn and Holar aquaculture (AC) charr.

The data reveal significant differences in expression of several biological pathways during charr development. There is also a difference between SB- and AC-charr in mitochondrial genes involved in energy metabolism and blood coagulation genes. We confirmed expression difference of five genes in whole embryos with qPCR, including *lysozyme* and *natterin* which was previously identified as a fish-toxin of a lectin family that may be a putative immunopeptide. We verified differential expression of 7 genes in developing heads, and the expression associated consistently with benthic v.s. limnetic charr (studied in 4 morphs total). Comparison of Single nucleotide polymorphism (SNP) frequencies reveals extensive genetic differentiation between the SB- and AC-charr (60 fixed SNPs and around 1300 differing more than 50% in frequency). In SB-charr the high frequency derived SNPs are in genes related to translation and oxidative processes. Curiously, several derived SNPs reside in the 12s and 16s mitochondrial ribosomal RNA genes, including a base highly conserved among fishes.

*Corresponding author – Tel: +354-525-4265 – Fax: 354-525-4632 – apalsson@hi.is

The data implicate multiple genes and molecular pathways in divergence of small benthic charr and/or the response of aquaculture charr to domestication. Functional, genetic and population genetic studies on more freshwater and anadromous populations are needed to confirm the specific loci and mutations relating to specific ecological or domestication traits in Arctic charr.

Todo list

INTRODUCTION

Historical contingencies and chance shape organism during evolution (Gould 1977; Jacob 1977), but convergence in phenotype and molecular systems indicates that evolution is to some extent predictable (Stern and Orgogozo 2009). Identification of genes and nucleotide changes influencing evolved differences is not a trivial task (Rockman 2012). However by identifying genes and developmental pathways shaped by evolution, we can test models of repeatability or contingency (Stern and Orgogozo 2009). An ideal system to study the role of chance and necessity in ecological evolution would be a species with i) readily observable phenotypic variation, ii) tractable ecological settings, and most crucially iii) parallel evolution of key traits within this species or few related species. Finches of the Galapagos islands and cichlids in the African great lakes are exciting multi-species systems in this respect (Abzhanov *et al.* 2004; Abzhanov and Tabin 2004; Albertson and Kocher 2006) and threespine stickleback has emerged as model "single species" systems (Cresko *et al.* 2007). The diversity in feeding specialization of fishes provides great opportunities for studying adaptation and divergence at the developmental and genetic level.

In this respect some northern freshwater fish species are of great interest as they exhibit frequent parallelism in trophic structures and life history and in many water bodies they are found as distinct resource morphs (Skúlason and Smith 1995; Snorrason and Skúlason 2004; Cresko *et al.* 2007; Klemetsen 2010; Bernatchez *et al.* 2010; Merilä 2013). In Iceland one of these species, Arctic charr (*Salvelinus alpinus*), is particularly well suited for studying the developmental underpinnings of trophic divergence and parallel evolution. A better understanding of molecular and developmental systems in non model species helps us elucidate to what extent they can respond to directional or disruptive selection and which genes are most likely to play central roles in adaption, including trophic, divergence (Stern and Orgogozo 2008, 2009).

Transcriptomic analyses of fish biology and divergence

Multiple methods can be used to identify genes affecting adaptations (Gibson and Weir 2005; Flint and Mackay 2009; Wittkopp 2011; Rockman 2012). In the past decade, transcriptomic methods have been used on many fish species to address various evolutionary and ecological questions. For example microarrays have been used to compare gene expression in anadromous and resident populations of brown trout (*Salmo trutta*), revealing that life history was a better predictor of gene expression in the liver than relatedness (Giger *et al.* 2006). The rapid development in Next generation sequencing (NGS) technologies has spawned many suitable tools and methods for studying transcriptomes of non-model species (Vijay *et al.* 2013; Qian *et al.* 2014).

Next generation sequencing showed that the liver transcriptome in crimson spotted rainbowfish (*Melanotaenia duboulayi*) responded to higher temperature with systematic differences in expression of metabolism genes (Smith *et al.* 2013). RNA-seq methods have been used to study non-model species such as the Mexican cavefish (*Astyanax mexicanus*), cod (*Gadus morhua*) brook charr (*Salvelinus fontinalis*)and Atlantic Salmon (*Salmo salar*) (Gross *et al.* 2013; Lanes *et al.* 2013; Bougas *et al.* 2013; Micallef *et al.* 2012; Wang *et al.* 2014), addressing different questions (concerning evolution, molecular genetics, development and aquaculture) (Qian *et al.* 2014). RNA-seq has been applied to adult Canadian Arctic charr to study salinity tolerance, linking expression and quantitative trait loci (Norman *et al.* 2014). Several studies on benthic and limnetic lake whitefish (*Coregonus cupleaformis*) highlight the potential for investigating the evolution of trophic diversity with genomic techniques. Microarray studies of adult lake whitefish pointed to parallel expression differences between benthic and limnetic forms (Derome *et al.* 2006). Filteau *et al.* (2013) found that a set of coexpressed genes differentiated the two whitefish morphotypes, implicating BMP signaling in the development of ecological differences in tropic morphology. An alternative approach in identifying pathways related to morphological differences is to study gene expression as the tropic apparatus or another organ of interest is developing. Analyses of gene expression in embryos can be a powerful approach to identify systems required for specific developmental stages and processes (Domazet-Lošo and Tautz 2010; Bozinovic *et al.* 2011).

Local adaptation has been extensively studied in the salmonid family, to which Arctic charr belongs (Fraser *et al.* 2011). The family is estimated to be between 63.2 and 58.1 million year old (Allendorf and Thorgaard 1984; Crête-Lafrenière *et al.* 2012).

Assembly and annotation of paralogous genes and multigene families is a major challenge. This task is made even more challenging because a whole genome duplication occurred before the radiation of the salmonid family (Davidson *et al.* 2010; Moghadam *et al.* 2011; Macqueen and Johnston 2014; Berthelot *et al.* 2014) which has provided time for divergence of ohnologous genes (paralogous genes originated by whole genome duplication event). Furthermore, recent estimates from the rainbow trout (*Oncorhynchus mykiss*) genome suggest that ohnologous genes were lost at a rate of about 170 ohnologous genes per million years and by utilizing multiple data sources one can solve the genome assembly problem of this family (Berthelot *et al.* 2014).

The highly polymorphic Arctic charr

Arctic charr can be found as anadromous or lake/stream residents and exhibit high level of within species polymorphism (Noakes 2008; Klemetsen 2010). Following the end of the last gracial period, about 10.000 years ago, Arctic charr colonized northern freshwater systems (Noakes 2008). Resource polymorphism in charr correlates with ecological attributes (Adams *et al.* 2007; Kristjánsson *et al.* 2011; Woods *et al.* 2012). For instance small charr with benthic morphology, are found in multiple lavaspring and pond habitats in Iceland (Kristjánsson *et al.* 2012), and in a comparative study of Icelandic lakes Woods *et al.* (2012) found that lakes with greater limnetic habitat, fewer nutrients, and greater potential for zooplankton consumption appeared to promote resource polymorphism. Some of the larger lakes contain two or more distinct morphs, typically limnetic and benthic forms. Multiple lines of evidence show that these differences stem both from environmental and genetic causes (Skúlason *et al.* 1989, 1993, 1996; Parsons *et al.* 2010, 2011). The best studied and most extreme example of sym-

patric charr morphs are the four morphs in Lake Thingvallavatn (Sandlund *et al.* 1992), two of which belong to a benthic morphotype, a large benthivorous (LB-charr) and a small benthivorous (SB-charr), and two limnetic morphs, a large piscivorous morph (PI-charr) and small planktivorous morph (PL-charr) (Snorrason *et al.* 1994). PL- and PI-charr operate in open water and feed on free-swimming prey, planktonic crustaceans and small fish, respectively. The two benthic morphs mainly reside on the bottom, feeding mostly on benthic invertebrates. The SB-charr can utilize interstitial spaces and crevices in the littoral zone typically consisting of submerged lava, which due to its porous surface, offers a richer source of benthic invertebrate prey than do stones with smoother surfaces (Malmquist *et al.* 1992). Several population genetics studies, using allozymes and mtDNA revealed no differences among charr populations (Magnusson and Ferguson 1987; Danzmann *et al.* 1991; Pálsson and Árnason 1994), while studies on microsatelite markers and nuclear genes, reveled both subtle (Volpe and Ferguson 1996; Wilson *et al.* 2004; Kapralova *et al.* 2011) and strong genetic differences among morphs (Kapralova *et al.* 2013). In particular Kapralova *et al.* (2011) concluded that small benthic morphs have evolved repeatedly in Iceland and that gene flow has been reduced between the PL and SB morphs in Lake Thingvallavatn since its formation approximatly 10.000 years ago (Saemundsson 1992). These previous studies relied on few markers, but as individual genes have distinct histories (Doiron *et al.* 2002; Miller *et al.* 2012), genome wide methods are needed to identify genomic regions associated with divergence.

RNA-sequencing of charr at specific developmental stages can both identify genetic separation and reveal differential expression of genes that influence the trophic diversity in Arctic charr. Previous work showed that expression of mTOR pathway components in skeletal muscle correlates with the SB-charr form (Macqueen *et al.* 2011). In studies, relying in part on the data presented here, we found that expression of several cartilage and bone development genes is higher in the heads of developing benthic compared to limnetic charr morphs (Ahi *et al.* 2013, 2014). Also our sequencing of the miRNA transcriptome in the same charr morphs as are studied here show that the expression of several miRNAs correlate with morph differences (Kapralova *et al.* 2014a).

In this study we conduct RNA-sequencing using samples of developing offspring of two contrasting Arctic charr morphs, a small benthic charr from Lake Thingvallavatn and Icelandic aquaculture charr conforming to a limnetic morphotype (Figure 1). For each morph we sequenced RNA samples from embryos at three pre-hatching and one post-hatching stage. Firstly, in order to find genes and pathways that are involved in the development of phenotypic differences between these morphs, the RNAseq data were screened for expression differentiation between morphs (and among developmental stages). Secondly, we screened for signals of genetic differentiation between morphs that may relate to divergence and adaptations. Thirdly, using qPCR and genotyping of population samples we also set out to verify a subset of the signals.

The comparisons of transcriptomes revealed significant expression differences between morphs, both involving specific loci and particular molecular systems. This has enabled us to identify candidate developmental genes that may affect jaw and craniofacial traits which separate benthic and limnetic morphotypes in charr (see our results and related studies, Ahi *et al.* 2013, 2014). The data also reveal dynamic expression of natterin paralogs, a putative immunological gene, during development and in distinct tissues. Interestingly, the data implicate divergence of mitochondrial function among the small benthic and aquaculture charr, perhaps reflecting contrasting selection pressures, i.e. the effects of adaptation to very cold,

nutrient poor habitats (SB-charr) as opposed to strong artificial selection for fast growth and delayed maturity (aquaculture charr). These results emphasize the broad utility of comparative transcriptomics in guiding research, e.g. studies that aim to further our understanding of the genetics and developmental aspects of morphological divergence.

MATERIALS AND METHODS

Sampling, rearing and developmental series.

In order to study gene expression during charr development, we set up crosses and reared embryos in the laboratory as previously described (Ahi *et al.* 2013). Embryos from four charr morphs were studied, an aquaculture charr (hereafter called AC-charr) from the Hólar breeding program (Svavarsson 2007), and three natural morphs from Lake Thingvallvatn; small benthivorous (SB), large benthivorous (LB) and small planktivorous (PL) charr (Snorrason *et al.* 1989). The first two, AC and SB, which exhibit contrasting adult size and morphology (Figure 1), were used for the developmental transcriptome analyses, and the latter two added for qPCR and SNP studies of selected genes. Briefly, spawning SB-charr were collected in September of 2009, via gill netting in Olafsdrattur in Lake Thingvallavatn. Similarly in 2010 spawning SB-, LB- and PL-charr were collected from Lake Thingvallavatn. Fishing permissions were obtained from the Thingvellir National Park Commission and the owner of the Mjóanes farm. For each parent group eggs from several females were pooled and fertilized using milt from several males from the same group. The aquaculture charr crosses were set up identically in 2009 and 2010 (Ahi *et al.* 2013). Embryos were reared at ~ 5°C under constant water flow and in complete darkness at the Holar University College experimental facilities in Verið, Sauðárkrókur. The water temperature was recorded twice daily and the average was used to estimate the relative age of the embryos using tau-somite (τs) (Gorodilov 1996). Embryos and juveniles were sampled at designated time points, placed in RNAlater (Ambion) and frozen at -20°C. For the investigation of different tissues of adult aquaculture charr (AC) from Hólar (fish size 20-25 cm) were used. Six randomly selected individuals were killed (cutting through spinal cord) and dissected and samples were taken from the skin, heart, liver, gills, spleen, intestine and kidney of each fish. The samples were placed in RNAlater (Ambion) and stored at -20°C. DNA for population genetic analyses was from our previous study (Kapralova *et al.* 2013).

Fishing in Lake Thingvallavatn was with permissions obtained both from the owner of the land in Mjóanes and from the Thingvellir 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 157/2012). Sampling was performed with University College Aquaculture Research Station (HUC-ARC) personnel. HUC-ARC has an operational license according to Icelandic law on aquaculture (Law 71/2008), which includes clauses of best practices for animal care and experiments.

RNA extraction and transcriptome sequencing

Embryos of AC- and SB-charr sampled in 2009 were used for transcriptome sequencing. For this we focused on the time covering development of pharyngeal arches and morphogenesis of the head, at 141, 163, 200 and 433 τs (post fertilization). For each combination of groups and timepoints we pooled RNA from approximately 6 individuals. RNA extraction and following

steps were performed as described earlier (Ahi *et al.* 2013; Kapralova *et al.* 2014a). The embryos were dechorionated under the light microscope (Leica S6E) and the yolk was discarded. The embryos were homogenized with a disposable Pellet Pestle Cordless Motor tissue grinder (Kimble Kontes) and RNA was extracted into two size-fractions using the mirVana kit (Ambion). The high molecular weight fraction was further used for mRNA-seq and RNA quality was analysed using an Agilent 2100 Bioanalyzer (Agilent Technologies). First and 2nd strand cDNA synthesis, fragmentation, adapter ligation and amplification were performed using the mRNA-Seq 8-Sample Prep Kit (Illumina) according to manufacturer's instructions. Sequencing was performed at DeCode genetics (Reykjavík, Iceland) using SOLEXA GAII technology (Illumina Inc.). The sequencing depth ranged from 49 to 58 million reads with a mean depth of 55 Million reads per sample.

The embryos reared in 2010 were used for qPCR expression analyses. RNA was extracted from six whole embryos, in two replicates (2 X 3 fish) (AC and SB sampled at 161 and 200 τs). For the extraction of RNA from heads of AC, SB, LB and PL, 12 embryos (2 X 6) at 178, 200 and 216 τs were used. Embryos were dechorionated and decapitated in front of the pectoral fin. RNA extraction and cDNA preparation were performed as described previously (Ahi *et al.* 2013). Similarly RNA was extracted from a small piece (ca. 2 mm^2) of skin, heart, liver, gill, spleen, intestine and liver from 6 adult AC-charr.

Analyses of RNA-seq data and mapping to Salmon EST contigs

As no *S. alpinus* genome is available we assessed expression and genetic variation by mapping the 36 bp reads to 59336 *S. salar* EST contigs from the SalmonDB (Di Génova *et al.* 2011, downloaded 22. March 2012) and the Arctic charr mitochondrial genome (Doiron *et al.* 2002, NC_000861).

To estimate expression, the reads were aligned to the salmon contigs with RSEM vs. 1.1.18, which distributes reads that map to multiple locations to the most likely contig, using expectation maximization (Li and Dewey 2011). The read counts for contigs with the same annotation were pooled, because some genes were represented by more than one contig and salmonids do have several closely related paralogous genes (Moghadam *et al.* 2011; Berthelot *et al.* 2014). Thus the expression tests are done on gene or paralog group level, instead of the contig level. In the remainder of the paper, gene will have this broader meaning, some genes are represented by one contig and others by two or more (indicated in all relevant tables). This brought the total number of genes(paralogous genes) tested down to 16851. Also genes with fewer than 800 mapped reads in the entire dataset, were excluded from the analyses. 10496 genes passed this filtering step.

A generalized linear model (GLM) with morph and developmental time as explanatory variables was used to find genes with different expression levels between the two charr morphotypes (groups) using the edgeR-package in R (Robinson *et al.* 2010).

$$Y = Morph + Time + Error$$

It should be pointed out that we couldn't test for interaction, as biological replicates were not available. To obtain further insight into the expression profiles of genes with significant expression differences, we also preformed clustering analyses on log-transformed cpm-values (counts per million, cpm-function in edgeR). The values for each gene were scaled by mean and standard deviation and the euclidean distance used for the hclust-function in R with the

default settings. We used the Hypergeometric-test in goseq (Young *et al.* 2010) to test for gene ontology enrichment. Since we pooled the read-count from different contigs we could not take gene length into account in those tests as would have been optimal.

The sequencing reads were deposited into the NCBI SRA archive – with accession number ZZZZ. ENA

1: SRA - ENA ↑

Assessment of differential expression with qPCR

We have previously identified suitable reference genes to study Arctic charr development (Ahi *et al.* 2013). Here we examined the expression of several genes in whole charr embryos, embryonic heads and adult tissues. For this, primers were designed using the Primer3 tool (Untergasser *et al.* 2012) and checked for self-annealing and heterodimers according to the MIQE guidelines (Bustin *et al.* 2009) (Table S1). Primers for genes with several paralogs were designed for regions conserved among paralogs. For natterin, primers for the different paralogs were designed to match regions differing in sequence. Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). For the calculation of relative expression of genes in whole embryos, the geometric mean expression of three reference genes, β-*Actin*, *elongation factor* 1α and *Ubiquitin-conjugating enzyme E2 L3*, was used for normalization. For visual comparisons among samples the normalized expression was presented as relative to the expression in AC at 161 τs (calibration sample). For the embryonic head samples *IF5A1* and *ACTB* were used as reference genes and a biological replicate of AC at 178 (τs) as the calibrator sample. Standard errors of relative expression were calculated from the standard errors (SE) of the ΔC_T-values with the formula $2^{-(\Delta\Delta Ct+SE)}$ = minimum fold expression and $2^{-(\Delta\Delta Ct-SE)}$ = maximum fold expression. The statistical analysis was performed using the ΔC_T-values with a two-way ANOVA with GLM function in R.

$$Y = Morph + Time + MxT + Error$$

Normal distribution of residuals was confirmed for all data. For the study of expression in the embryonic head we followed a significant Morph effect in the ANOVA with Tukey's post hoc honest significant difference test, on relative expression ratios (ΔC_Ts).

Polymorphisms in the Arctic charr transcriptome

For analysis of genetic variation we mapped the reads again to the salmon contigs, but this time using BWA (Li and Durbin 2009) with a seed length of 25, allowing two mismatches. We re-mapped the reads since BWA allows short indels, but disregarding them lead to many false positive SNPs close to indels. To extract candidate polymorphic sites from the Arctic charr transcriptome we ran VarScan2 (Koboldt *et al.* 2012) with minimum coverage of 50 reads and minimum minor allele frequency 0.1 on reads mapped to each *S. salar* contig for all of the 8 timepoints and morph combinations. This was done separately for reads that mapped uniquely to one contig only (UNI) and reads that mapped to two or more contigs (REP). These SNP-candidates were further processed in the R statistical environment. SNP-candidates at 90% frequency or higher in all samples were removed, they reflect difference between Arctic charr and *S. salar* and are not the focus of this study. SNP-candidates with poor coverage in specific samples, that is coverage of 5 or fewer reads in 3 or 4 samples of each morph, were removed. As the SNP analysis was done on individual contigs, differences among paralogs appear in the

data. But, since each sample is a pool of few individuals it is very unlikely that we have the same frequency of true SNPs in all the samples. This property was used to remove variants that are most likely due to expressed paralogs. Using Fisher exact tests to evaluate differences between samples only SNPs that were significantly different between samples with p-value lower than 0.05 (no multiple testing correction) were selected for further examination. As equal cDNA input from individuals in sample cannot be assumed,due to expression differences among them and stochastic processes in sample preparation, read numbers were summed over the four samples for each morph for the comparison between the two groups. A conservative approach was taken to look for difference between morphs. We focused on SNP-candidates that showed difference in frequency between morphs without adjusting for multiple testing (Fisher exact test, $p > 5\%$). We extracted the most interesting candidates by filtering on frequency difference between the morphs (delta). SNP-candidates with the highest frequency difference (delta $> 95\%$) were manually processed and redundant candidates removed. A similar approach was used to mine for polymorphism in the Arctic charr mtDNA (NC_000861).

We wrote a python script to predict the impact of SNPs within the mRNA sequences. Polymorphisms where categorized according to their location (3'UTR, coding, 5'UTR), and those within the coding region into synonymous or non-synonymous. We also categorized amino-acid substitutions according to changes in biochemical properties.

Verification of candidate SNPs

We chose 12 candidate SNPs for verification (see below). The candidates were verified using a similar approach as in our previous work (Kapralova *et al.* 2013). First we conducted genomic comparisons of the Salmon genome, ESTs and short contigs from the preliminary assembly of the Arctic charr transcriptome. This allowed us to infer the placement of the putative polymorphism in the locus, and design paralog specific primers for PCR (less than 1 kb amplicons) for verification of the 12 candidate SNPs. Each individual was genotyped by first amplifying the region of interest using PCR, followed by ExoSAP, direct sequencing (BigDye) and finally run on an Applied Biosystems 3500xL Genetic Analyzer (Hitachi). Raw data was base-called using the Sequencing Analysis Software v5.4 with KBTMBasecaller v1.41 (Applied Biosystems). Ab1 files were run through Phred and Phrap and imported to Consed for visual editing of ambiguous bases and putative polymorphisms and trimming primer. The fasta files were aligned with ClustalW online (Larkin *et al.* 2007, http://www.ebi.ac.uk/Tools/msa/clustalw2/) and manually inspected in Genedoc (Nicholas *et al.* 1997). All sequences where deposited to Genebank as Popsets under the accession numbers KP019972-KP020026.

Comparative genomic analyses of sequence polymorphism

As few of the verified SNPs where located in functionally interesting or highly conserved genes, we studied the putative effects of the observed variants with comparative genomics. Two approaches were used, blasting salmon EST's to NCBI (May 2013) and blasting on the Human genome in UCSC genome browser (in September 2013). This yielded several hundred sequences from related fish and other vertebrates. The list was reduced to 20 sequences, aligned with ClustalW and manually adjusted within Genedoc. The species and genome versions used are; Human (*Homo sapiens*, hg19), Lamprey (*Petromyzon marinus*, petMar1), Fugu (*Takifugu rubripes*, fr2), Medaka (*Oryzias latipes*, oryLat2), Stickleback (*Gasterosteus aculea-*

tus, gasAcu1), Tetraodon (*Tetraodon nigroviridis*, tetNig2), Zebrafish (*Danio rerio*, danRer6). We also downloaded from NCBI the sequence of whole or partial mtDNA from several fish species; Brown trout (*Salmo trutta*, JQ390057 and AF148843), Broad whitefish (*Coregonus nasus*, JQ390058), Legless searsid (*Platytroctes apus*, AP004107), Pacific menhaden (*Ethmidium maculatum*, AP011602), Icefish (*Salanx ariakensis*, AP006231 and HM151535), Chain pickerel (*Esox niger*, AP013046) and Western Pacific roughy (*Hoplostethus japonicus*, AP002938). The three mitochondrial variants are (numbered by the *S. alpinus* mtDNA - NC_000861) are; m1829G>A (CCACGTTGTGAAACCAAC[G/A]TCCGAAGGTGGATTTAGCAGT), m3211T>C (CGTGCAGAAGCGGGCATAAG[T/C]ACATAAGACGAGAAGACCCT) and m3411C>T (CTCTAAGCACCAGAATTT[C/T]TGACCAAAAATGATCCGGC). The images were produced in GIMP version 2.8.6. for Windows (Kimball *et al.* 1995).

RESULTS

RNA sequencing characteristics

Our general aim was to get a handle on molecular and developmental systems that relate to the frequent ecological and phenotypic divergence seen in Arctic charr. More specifically we aimed to identify genes that relate to divergence in craniofacial features that emerge some weeks before and after hatching of embryos. To this end we analysed transcriptomes of pooled embryonic samples at four time-points (141, 163, 200 and 433 τs) derived from pure crosses of two Arctic charr groups with contrasting phenotypes; the small benthic morph from Lake Thingvallavatn (SB) and aquaculture charr (AC). From each sample we obtained good quality data with an average of 55 million reads. To quantify the expression levels, the reads were aligned to a salmon EST-assembly (Di Génova *et al.* 2011) as the Arctic charr genome has not been sequenced. Around 20% of the reads mapped uniquely to the EST data (Table S2). A further 30% mapped to two or more contigs, probably representing paralogous genes, recent duplications or repeat like elements within transcribed regions. A substantial fraction of the RNA-sequencing reads did not map to the contigs from *S. salar* (Table S2). Analyses of those reads require a full genome or assembly of the transcriptome with longer and paired end reads from *S. alpinus*.

Differential expression during Arctic charr development

The scope of the sampled developmental timepoints and morphs only allows contrasts of expression between morphs, or expression differences between time points. For the expression analysis we collapsed ESTs into 16851 genes or paralog groups (hence "gene" has here this broader meaning, see Materials and Methods). We only considered genes to which 800 or more reads mapped, or a total of 10496 genes. We tested for differential expression between time points, or morphs (SB and AC) using the edgeR-package (Robinson *et al.* 2010).

As was expected we detected considerable changes in the transcriptome during Arctic charr development (Figure 2A). The expression of 1603 and 2459 genes differed significantly between developmental timepoints at the 1% and 5% levels of false discovery rate (FDR), respectively (Supplementary file S1). The difference was most pronounced between the pre-hatching embryos (timepoints: 141, 163, 200 τs) and the post hatching embryos (timepoint 433 τs),

as more than 70% of the genes with FDR below 1% had higher expression in the post hatching embryos (Figure 2A). According to Gene Ontology analyses, six separate GO categories are significant (below 10%FDR). The most drastic changes were seen in processes related to glycolysis (GO:0006096, FDR = 0.0009), were the expression of 19 out of 25 genes changed during this developmental period. The other five classes of biological functions that were differentially expressed during Arctic charr development are, ion transport (GO:0006811, FDR = 0.027), blood coagulation (GO:0007596, FDR = 0.03), DNA repair (GO:0006281, FDR = 0.08) and two immune related categories (GO:0019882, FDR = 0.08, GO:0006955, FDR = 0.09). Those transcriptome changes probably reflect developmental changes and the difference in the environment of embryos before and after hatching.

Differential expression between Arctic charr morphs

We were especially interested in genes showing expression differences between the two Arctic charr morphs as such differences may relate to pathways and gene networks that play a causative role in phenotypic divergence. Testing for heterochrony in expression was not feasible, as only one sample per morph and developmental time was sequenced. 296 genes were differentially expressed (FDR < 5%) between the morphs (141 higher in SB and 152 higher in AC). Among genes with higher expression in SB-charr two biological GO categories were enriched: blood coagulation (GO:0007596, p = 0.001) and proteolysis (GO:0006508, p = 0.002). Notably expression of blood coagulation factors also differed between developmental stages (see above). In AC-charr, genes in three categories: respiratory electron transport chain (GO:0022904, p = 0.0006), ATP synthesis coupled electron transport (GO:0042773, p = 0.002) and neurotransmitter transport (GO:0006836, p = 0.009) have higher expression. The first two GO categories both relate to energy generation in mitochondria and may imply higher expression of genes with mitochondrial functions in AC-charr or more mitochondria per cell.

Using more stringent FDR (1%), a total of 31 genes had significantly higher expression in SB and 40 genes in AC (Figure 2B, Tables 1 and 2). The identified genes have diverse functional annotations, not elaborated on here. Few signaling pathway components were among the top entries, but curiously two Histones appear on each list, *Histone 3 embryonic* (1) is 2.7 fold higher in SB and *Histone H1* (3) is 4.5X higher in AC. The genes with higher expression in each morph were clustered into 4 groups, using hierarchical clustering in R. This did aggregate genes of similar function.

For instance SB cluster 3 has three immune related genes: *Complement factor D* (9), *H-2 class I histocompatibility antigen L-D alpha chain* (2) and *Sushi domain-containing protein 2* (4) and one gene with unknown function (Table 1). Note, however, that immune genes were not significantly enriched in the GO comparison of morphs.

The transcriptome data suggest that mitochondrial function and blood coagulation genes are differentially expressed between the morphs, but due to small sample size in the RNA-sequencing, qPCR verification was needed.

Validation of gene expression differences in whole embryos and paralog specific expression of *natterin* genes

The transcriptome study highlights genes likely to differ in expression between embryos of SB- and AC-charr. We showed previously that these RNA-seq expression estimates correlate with

expression estimated from qPCR (Ahi *et al.* 2013). Nine genes with expression differences between the morphs were selected from the transcriptome for validation by qPCR in whole embryos. Of those five genes were differentially expressed between AC and SB at 161 or 200 τs (Figure 3, Table S3). Three of these genes, *Nattl*, *Alkaline phosphatase (Alp)* and *Lysozyme (Lyz)*, had significantly higher expression in SB. The other two, *Keratin-associated protein 4-3 (Krtap4-3)* and *Poly polymerase 6 (Parp6)* had higher expression in AC embryos (Figure 3, Table S3). No significant morph and time interaction was detected.

As some of the genes are represented by different contigs or even paralogs, we set out to disentangle the expression of one paralog group in detail. The qPCR primers used above were matched conserved gene regions and thus are likely to estimate the combined expression of several paralogs. We chose to measure the expression of three different *natterin* paralogs (*nattl1, 2 and 3*), in part because this understudied gene was first characterized as a toxin produced by a tropical fish (Magalhães *et al.* 2005, 2006). We studied *nattl* expression in several developmental stages in AC-, SB- and PL-charr as well as in selected tissues of adult AC-charr. The expression level of the three paralogs differed between morphs and timepoints (Figure 4 and table S4). Overall *nattl2* had the highest expression in all morphs (data not shown). The *nattl1* had higher expression in embryos of PL-charr than in AC- and SB-charr, while *nattl2* and *nattl3* were more expressed in SB-embryos.

In adult tissues (in AC-charr) expression estimates of the entire gene family (referred to as *nattl*) was highest in the gills, followed by expression in kidney, skin and spleen. Low expression levels were detected in liver, intestine and heart (Figure S1 and table S4). The three *nattl* paralogs followed different patterns, whilst each of them showed significant expression differences among tissues. *Nattl1* was mainly expressed in spleen and kidney, while *nattl2* showed a significantly higher expression in skin, liver and in gills. Similarly, the relative expression of *nattl3* was highest in the gills and skin. This indicates that the three *nattl* paralogs are expressed in a tissue specific manner, and also differently during the development of the three charr morphs studied here.

Expression differences in the heads of benthic and limnetic charr morphs

To study the craniofacial divergence between sympatric Arctic charr morphs we used qPCR to study 8 genes with expression difference in the RNA-seq data (all higher in SB) and with craniofacial expression during zebrafish development (Sprague *et al.* 2006). We focused on charr heads at three time-points (178, 200 and 218 τs) as this period overlaps with early stages of craniofacial skeletal formation in Arctic charr (Eiriksson *et al.* 1999; Kapralova *et al.* 2014b, in review). The qPCR confirmed the higher expression of seven out of these eight genes in the head of SB-charr compared to AC-charr (Figure 5 and S2). These seven genes are *Claudin 4 (Cldn4)*, *adseverin (Scin)*, *Junction plakoglobin (Jup)*, *Lipolysis stimulated lipoprotein receptor (Lsr)*, *Major vault protein (Mvp)*, *Transforming growth factor beta receptor II (Tgfbr2)* and *Vitamin D receptor a (Vdra)*. The eight gene, *Retinoic acid receptor gamma-A (Rarg)* gave a small but significant response in the head, but the differences were reversed, i.e. the expression was higher in AC, from what was seen in the whole embryo transcriptome (higher in SB).

In order to test whether the expression differences in the head follow the line of benthic and limnetic divergence, we added two more morphs, LB- and PL-charr to the qPCR analysis. Interestingly, the same 7 genes displayed higher expression in the other benthic morph (LB), whereas expression in the limnetic PL morph was more similar to AC levels (Figure 5). The

expression difference of the seven genes was, in almost all cases, consistent over the three time-points studied (See Figure S2). These results show that differential expression of several pathways associate with the divergence in craniofacial elements in limnetic and benthic charr in Lake Thingvallavatn.

Analyses of polymorphism in RNA sequencing data

The RNA-seq data also revealed segregating variations with large frequency differences between charr morphs. To uncover candidate SNPs we mapped the reads to all of the *S. salar* EST-contigs. Filtering on coverage yielded 165,790 candidate-SNPs (Table 3); of those 66.569 came from reads that mapped uniquely (UNI) and 57.009 candidate-SNPs from reads that mapped to more than one contig (REP), with very little overlap between lists. The salmonid ancestor underwent whole genome duplication, generating long blocks of synteny and paralogous genes (Berthelot *et al.* 2014). Differences among paralogs appear as SNPs that are at similar frequency in all samples (if the expression of paralogous genes is stable). By requiring variant frequency differences ($p < 0.05$, uncorrected) between samples we reduced the list of candidates by two thirds, yielding over 20.000 candidate SNPs. Note, that as cDNA from charr families was sequenced (not a population sample), estimates of SNP frequencies are imprecise. To err on the side of caution, we only selected further SNP candidates with 50% or higher frequency difference between morphs for further studies. The candidate SNPs were also summarized by frequency of the derived allele, in reference to the *S. salar* sequence. This gave 672 and 872 SNPs at higher frequency, in AC-charr and SB-charr, respectively. The uniquely and multiply mapped reads, revealed approximately similar numbers of candidate SNPs. Gene ontology analysis showed that for derived SNPs in SB, there was an excess of variants in genes related to translation, both as a broad category and specific subgroups (Table S5). There was also enrichment of SNPs in genes related to DNA-mediated transposition, DNA integration, DNA replication and oxidation-reduction process. No GO categories were enriched for high frequency derived SNPs in AC. Furthermore, functional effects of the candidate SNPs (UTR, synonymous and non-synonymous) were predicted. The distribution among those categories did not differ between variants detected by uniquely or repeatedly mapped reads, $\chi^2_{[3]} = 2.59, p = 0.46$ (Table S6).

We next focused on candidate SNPs that are nearly fixed in this data, with frequency differences above 95% frequency. A total of 63 candidate SNPs fall into this category, after manual inspection of contigs and SNP position three candidates were removed since they represented the same SNP. This brought the total down to 60 SNP candidates, 46 SNPs from uniquely mapped reads and 14 from reads that mapped more than twice (Table 4 and 5). For the SNPs from uniquely mapped reads, 17 are fixed in AC-charr and 29 in SB-charr. The few genes with more than one polymorphic site were; *Keratin type II cytoskeletal 3 (KRT3)*, *Cysteine sulfinic acid decarboxylase (CSAD)* and *DNA-directed RNA polymerase I subunit RPA12 (RPA12)* which have 5, 5 and 2 SNP candidates respectively. *KRT3* and *CSAD* had significant differentiation in both SB and AC. In AC-charr various genes are implicated, including *Eukaryotic translation initiation factor 2-alpha kinase 1 (E2AK1)* and *Transcription cofactor HES-6 (HES6)*. Similarly, 14 SNPs with large differentiation between morphs were predicted from reads that mapped on two or more contigs (Table 5). Of these, we found two variants in the mitochondrial *60S ribosomal protein L36 (RPL36)* and variants in 4 other mitochondrial genes (*28S ribosomal protein S18a mitochondrial (MRPS18A)*, *Apoptosis-inducing factor 1 mi-*

tochondrial (AIFM1), *Isocitrate dehydrogenase [NADP] mitochondrial (acIDH1)* and *Protein S100-A1 (S100A1)*) at higher frequency in AC-charr. PCR and Sanger sequencing was used to confirm SNPs in *DNA2-like helicase (DNA2)*, a gene with nuclear and mitochondrial function, and two other genes *Uroporphyrinogen decarboxylase (UROD)*, and *Mid1-interacting protein 1-like (MID1IP1)* (Table S7). The candidate variant *Eukaryotic translation initiation factor 4 gamma 2 (EIF4G2)* was not substantiated by the PCR/sequencing. Furthermore, there was very limited overlap between the genes with SNPs that had large differences in frequency and genes with differential expression between morphs (data not shown).

In sum, the data suggest substantial genetic separation between the two charr morphs studied here, the small benthic from Lake Thingvallavatn and the aquaculture charr from Holar. While individual SNPs clearly need to be verified e.g. by Sanger sequencing or SNP assays on more individuals, the results suggests genetic differentiation between morphs in genes of various molecular systems. Considering the enrichment of differentially expressed genes related to mitochondrial energy metabolism (above), and high frequency candidate SNPs in several genes with mitochondrial function in AC-charr we decided to study the mitochondrial transcriptome further.

Polymorphism and expression of Arctic charr mtDNA

Mitochondria are key to the energy metabolism of eukaryotes. Both charr types studied here reflect metabolic extremes, the aquaculture charr has been bred for rapid growth while the small benthic morph is thought to have experienced natural selection for slow metabolism and retarded growth (Jonsson *et al.* 1988; Snorrason *et al.* 1994). Although mRNA preparation protocols were used for generating cDNA for the RNA-sequencing, a substantial number of reads came from non-polyadenylated sequences. By mapping the reads to mtDNA sequence of Arctic charr we could estimate expression of mitochondrial genes and infer polymorphism both in genes and intergenic regions. There was a clear difference in sequencing coverage, with more than twice as many reads mapped from the AC- compared to SB-charr (mean fold difference 2.27, significant by a Wilcoxon test, $p < 0.0004$). Note, as only two types of fish are compared, it is impossible to determine the polarity of expression divergence.

Using an appropriate outgroup it is possible to determine ancestral and derived states for DNA polymorphism data. The mapped RNA-reads were used to identify polymorphism and divergence in the entire mitochondrial chromosome. The mtDNA reference was sequenced in a Canadian *S. alpinus* (Doiron *et al.* 2002) , so differences between Icelandic an Canadian Arctic charr could also be identified. Bioinformatics revealed 82 candidate sites, including 35 that most likely represent divergence between Icelandic and Canadian charr. For 17 of the candidate SNPs the derived allele had higher frequency in SB and for in 21 the derived allele was higher in AC (Figure 6). Note however, the frequency distribution is most irregular as we sequenced embryos of related individuals (see Materials and Methods), not a population sample. The divergence between Iceland and Canada is particularly small in the ribosomal RNA genes, but curiously there were clear morph differences in that region. To confirm and better estimate the frequency of variants in the ribosomal genes, we PCR amplified and sequenced two ~550 bp regions in the 12s and 16s genes comparing three morphs (PL, LB and SB) from Lake Thingvallavatn (Figure 7A, C and E). The 12s polymorphism (m1829G>A) differed significantly between the morphs ($\chi^2_{[2]} = 8.6, p = 0.014$), and was at highest frequency in the SB (0% in PL, 12.5% in LB and 75% in SB). Similarly m3411C>T in the 16s was enriched in SB (62.5%) but

found at lower frequency in PL (0%) and LB (12.5%) (it deviated significantly between morphs, $\chi^2_{[2]} = 9.3333, p = 0.009$). The Sanger sequencing also revealed three other polymorphisms in the amplified region, not found in the NGS data. Among those m3211T>C in the 16s gene was at 75% frequency in LB, but not found in the other morphs ($\chi^2_{[2]} = 19.76, p < 0.0001$).

In order to gauge the potential functionality of those variants we aligned the rRNA genes from nearly hundred fishes and several vertebrates. The position affected by m1829G>A and m3211T>C, in the 12s and 16s rRNAs, are not well conserved in fishes or vertebrates (Figure 7B and D). However m3411C>T, in the 16s rRNA, alters a position that is nearly invariant in 100 fish genomes (Figure 7F). The only exception is Pacific menhaden, which curiously also has T in this position. This region could not be aligned properly in other vertebrates. Thus m3411C>T alters a conserved position, but probably not very drastically as the introduced allele is tolerated in another fish species.

In summary, the results indicate substantial differentiation of mitochondrial function between the small benthic charr of Lake Thingvallavatn and Icelandic aquaculture charr. This is both seen in the apparent twofold higher expression of mtDNA in aquaculture and SNPs that differentiate the morphs.

DISCUSSION

The broad aim of this study was to get a handle on genetic and molecular systems that associate with the repeated evolution of small benthic Arctic charr in Icelandin. To this end we performed comparative transcriptome analysis contrasting the development of embryos from SB-charr and aquaculture charr. While our focus is on SB-charr and its special features and adaptations, the data can also illustrate how charr and possibly other salmonids respond to selection for growth and maturity traits during aquaculture breeding.

Developmental transcriptome of Arctic charr morphs

Here we generated an Illumina based RNA-seq transcriptomes from four timepoints during early development, in two Arctic charr morphs (SB-charr and AC-charr). As no reference genome is available for Arctic charr, we mapped reads to *S. salar* EST-contigs (Di Génova *et al.* 2011) in order to estimate expression and identify candidate genetic polymorphisms. As many of the contigs are short or with overlap in annotation, we collapsed related genes into paralogous genes when appropriate for the expression analysis. The main advantage is that this limits the number of statistical tests (and hence increases statistical power). The downside is that paralog specific expression patterns are masked, as our qPCR results of the *natterin* gene family show (Figure 3 and S1). Recent analysis from rainbow trout shows that most paralogs from the latest whole genome duplication event retrain the same expression pattern (Berthelot *et al.* 2014) indicating that this scenario is probably uncommon, but on the other hand it is of considerable interest when two paralogs show distinct expression patterns (Zou *et al.* 2009). In their analysis of Arctic charr gill transcriptome Norman *et al.* (2014) also used the Illumina sequencing technology to evaluate expression. Their read length was considerably longer than in this study and they were therefore able to assemble their data into contigs. They did however not consider the vast amount of paralogs in their approach and also merged similar contigs in their assembly. Thus the complexity of Arctic charr transcriptome still remains a mystery that advances in sequencing technology, assembly algorithms as well as genome sequencing of this species could aid in revealing.

Our data reflect differential deployment of several gene classes during Arctic charr development. Studies in salmonids and other fish have demonstrated large changes in expression during early development, including coordinated changes in many cellular and developmental systems (Domazet-Lošo and Tautz 2010; Jantzen *et al.* 2011; Drivenes *et al.* 2012; Bougas *et al.* 2013; Piasecka *et al.* 2013). Several blood coagulation factors genes showed significant changes during charr development, and were also more highly expressed in the SB-charr. This might reflect differences in the rate of development of blood composition, or tissue composition, in the two morphs. But further work is needed to confirm those patterns.

Energy metabolism tuning under domestication or adaptation in SB-charr?

By comparing Aquaculture charr with SB-charr, that represents a small benthic resource morph that has evolved repeatedly in Icelandic stream and pond habitats (Kapralova *et al.* 2011), we expected to increase chances of finding genes and pathways involved in adaptation to these special habitats. As might be expected our data highlight differences in systems related to the energy metabolism of SB- and AC-charr. First, there is 2X higher expression of respiratory electron transport chain components in AC compared to SB and 100% more mitochondrial derived reads are found in the AC-charr samples. Note though that the direction of divergence is unknown, that is whether expression was up in AC or down in SB. Second, many derived candidate-SNPs in genes related to mitochondrial function were at high frequency on the AC branch. For instance in *S100A1*, which has been implicated in mitochondrial regulation in cardiac tissue in humans (Völkers *et al.* 2010), but its expression is probably not exclusive to this tissue. Third, while the mitochondrial ribosomal genes generally evolve slowly, we do see derived variants at high frequency in the SB and large benthic charr in Lake Thingvallavatn. Specifically, m3411C>T in SB affects a position that is highly conserved among fish, and could affect function of the 16s rRNA. Considering the high frequency of mitochondrial variants on the aquaculture branch, domestication of Arctic charr may have led to selection on mitochondrial function, most probably for increased metabolic performance. It is also possible that evolution of small benthic morph in Lake Thingvallavatn has led to divergence in mitochondrial function, for instance via the m3411C>T in the 16s ribosomal gene. Earlier studies of mitochondrial markers in *S. alpinus* have not found large signals of divergence within in Iceland (Danzmann *et al.* 1991; Volpe and Ferguson 1996; Kapralova *et al.* 2013), probably because they have not sampled the same genes.

The mitochondrion is more than a powerhouse, it integrates for instance metabolism, cell cycle and apoptosis (McBride *et al.* 2006). The number of mitochondria and its functions are known to correlate with environmental attributes. For instance in Antarctic fishes under extreme cold, higher numbers of mitochondria are found in muscle and heart cells (O'Brien and Mueller 2010). Furthermore, illustrating long-term evolutionary pressure on mitochondrial function under extreme conditions, genes related to those functions are more likely to be duplicated in such species (Coppe *et al.* 2013). The data presented here suggest an expression difference between morphs that could reflect differences in total number of mitochondrion, the number of mtDNA copies per mitochondrion or cell, or difference in RNA expression from the mtDNA, possibly due to coevolution of mtDNA with diet or temperature (Ballard and Pichaud 2014). Further work is needed to map out the expression differences of mitochondrial related genes in more SB and anadromous charr morphs. The genetic signals should also be investigated by comparison of more populations and/or along ecological clines (e.g. temperature) or with

respect to life history. For instance, Teacher *et al.* (2012) found evidence of positive selection in the mitochondrially encoded *ND2*, *ND4* and *ND5* in Atlantic herring. However the correlations of genetic variation to environmental attributes (temperature, salinity, latitude) were subtle, perhaps because of the ease of migration and low population structure. In summary, the results suggest divergence in mitochondrial function, either due to the domestication of aquaculture charr or possibly reflecting adaptation of the small benthic charr to specific habitats.

Higher expression of *lyzozyme* and *natterin-like* in SB-charr

Because of the genetic separation in immunity genes (*MHCII*α and *cath2*) reported among the sympatric morphs in Lake Thingvallavatn (Kapralova *et al.* 2013) we examined and verified the expression of *Lyz* and *nattl*. Both genes are likely to play important roles in immune defense and were found to be higher expressed in SB. The substrate of lysozyme (Fleming 1922) is the bacterial cell wall peptidoglycan and it acts directly on Gram-positive bacteria (Chipman and Sharon 1969). Lysozyme also promotes the degradation of the outer membrane by complement and other enzymes exposing the peptidoglycan layer and therefore indirectly acting also on Gram-negative bacteria (Subramanian *et al.* 2007). Therefore, lysozyme plays an essential role in immune defense in most eukaryotes. Another gene that caught our attention, with significantly higher expression in SB, was *natterin-like*. Natterins were first discovered from the venom gland of the tropical toxic fish species *Thalassophryne nattereri* (Magalhães *et al.* 2005, 2006), and are found by sequence similarity in e.g. zebrafish, Atlantic salmon and here in Arctic charr. The predicted Natterin proteins contain a mannose-binding lectin-like domain (Jacalin-domain) and a pore-forming toxin-like domain and can cause edema and pain due to kiniogenase activity (Magalhães *et al.* 2005). Jacalin is known to stimulate functions of T- and B-cells and therefore can play an important role for immune system functions (Jeyaprakash *et al.* 2002). In fish, the immune defensive role of lectin-like proteins is known. Mannose-binding lectins are pathogen recognition proteins (antibodies) and therefore are important for the acute phase response of fish (Magnadóttir 2006; Magnadottir 2010). Our data suggest an immune related function of *nattl* genes in charr, as the highest expression was found in skin and kidney. In sum, the expression of those two genes in SB-charr might reflect differential preparation of those juveniles for the bottom dwelling habitats, which may be rich in bacteria or more challenging for their immune system.

In this study we collapsed contigs into gene or paralog groups for the transcriptome analyses. The disadvantage of this approach is that differential expression in one paralog, can be masked by other related genes that do not differ between groups or have contrasting expression patterns. We looked at this by studying the expression of three paralogs of the *natterin like* genes in different morphs during Arctic charr development, and among tissues of adult AC-charr. The data show that the three *nattl* genes are expressed differentially between the morphs, thus it is not divergence in the expression of one paralog that explains the general *nattl* expression disparity in the transcriptome. Certainly, other scenarios could apply to other genes in the transcriptome, and the relative contribution of different paralogs should be evaluated systematically with deeper RNA-sequencing and longer reads.

Expression divergence in craniofacial genes in benthic morphs

Given the distinct difference in feeding apparatus and jaw morphology between benthic and limnetic charr, we investigated further genes involved in these processes that were differentially expressed in the transcriptome. Study of the skulls of charr post-hatching embryos and juveniles from Lake Thingvallvatn, showed that some elements of the developing head ossified earlier in SB-charr than in PL-charr (Eiriksson 1999). Guided by our developmental transcriptome we found two extra-cellular matrix (ECM) remodeling genes, *Mmp2* and *Sparc* and a conserved co-expression module of genes with known roles in craniofacial morphogenesis, to have higher expression in developing heads of benthic Arctic charr morphs than in limnetic morphs (Ahi *et al.* 2013, 2014). These studies and the current data confirm the utility of the contrasting developmental transcriptomes for selecting relevant candidate genes with differential expression in head development, as 7 out of 8 candidates were confirmed by qPCR. These same genes had higher expression in the developing head of two benthic morphs, and lower in fish with more limnetic phenotypes.

We focused on a few targets of Tgf-β and Ahr signaling pathways because of the endogenous role of both pathways in craniofacial morphogenesis as well as their known transcriptional connection (Chai *et al.* 2003; Puga *et al.* 2005; Goodale *et al.* 2012). *Adseverin (Scin)* was one of the top differentially expressed genes (Table 1) and has roles in rearrangements of the actin cytoskeleton, chondrocyte differentiation and skeletal formation (Nurminsky *et al.* 2007; Vieira *et al.* 2013). In mice, *Scin* is demonstrated as a direct target of the Ahr pathway (Svensson *et al.* 2002) and in Arctic charr *Scin* shows higher expression in the developing head of benthic morph embryos. Also, in the transcriptome *Lsr*, *Cldn4* and *Tgfbr2* had significantly higher expression in SB-charr, and we confirm that higher expression of those genes associated with the benthic morphotype. *Lsr* (LISCH7) is a molecular component of tri-cellular tight junctions (Furuse *et al.* 2012) and has been shown to be suppressed upon Tgf-β1 stimulation (Jazag *et al.* 2005) in a human cell line. Similarly, *Cldn4*, a tight junction protein with unknown role during embryonic morphogenesis, is also a common target of the Tgf-β and Ahr signaling pathways (Planchart and Mattingly 2010; Hering *et al.* 2011). Finally, the expression of *Tgfbr2*, encoding a receptor of Tgf-β was slightly but significantly higher in the head of benthic morphs. Previous studies suggest a crucial role of *Tgfbr2* in craniofacial morphogenesis (Ito *et al.* 2003).

We also confirmed the expression of several other genes, including two genes with higher expression in SB-charr. *Mvp* is the predominant component of cytoplasmic ribonucleoprotein structures called vaults (Kedersha *et al.* 1990), which is highly conserved across eukaryotes. The vaults have been something of an enigma, but are implicated in several processes from signal transmission and immune response (Berger *et al.* 2009). The *Jup* gene also showed higher expression in SB-charr. Finally, higher expression of *Vdra*, encoding the vitamin D receptor A, was found in the heads of benthic forms. The receptor regulates mineral homeostasis, osteoblast differentiation and bone metabolism (van Driel *et al.* 2004).

We could not confirm the differential expression of *Rarg* and the expression difference was actually opposite in the head to what we saw in the transcriptome sequencing of the whole embryos. This might be caused by the expression differences in head being masked out by opposite differences in another tissue in the RNA-seq data. We decided to not look further into tissue specific expression of *Rarg* for the time being.

To summarize, the differential expression of several genes was confirmed between Icelandic benthic and limnetic charr. Some of these genes belong to signaling pathways involved in bone

mineralization and craniofacial morphogenesis (see above). It would be most interesting to see if expression of the same genes associates with SB-charr from other lakes and habitats, or even in other species with similar trophic diversity. In a parallel study we used the same transcriptome data to guide the selection of genes for detailed spatio-temporal expression analysis during the head development in Arctic charr (Ahi *et al.* 2014). Taken together, the results supports the notion that our transcriptome analyses of early development provides a reliable list of candidate genes for further study of craniofacial development and divergence in Arctic charr.

Conclusion

The results presented here set the stage for more detailed investigations of the molecular and developmental systems involved in the evolution of the suits of adaptive traits illustrated in the highly polymorphic and rapidly evolving Arctic charr. This project has focused on the unique small benthic charr, typically found in cold springs and small pond habitats in Iceland, particularly those with lava substratum (Kristjánsson *et al.* 2011; Kapralova *et al.* 2011). The availability of charr populations at different stages of divergence enables direct comparison of the role of genetic, environmental and plastic factors in the generation of diversity and divergence. Expression differences between groups may reflect genetic changes, e.g. in regulatory sequences or post transcriptional modifiers, or reverberations in homeostatic or developmental cascades (Rockman 2012). The utility of our approach obviously depends on expression levels of genes, genes with low and/or time-restricted expression are harder to detect than highly expressed genes. Although the candidate SNPs detected are restricted to transcribed regions they may reflect signal of adaptive differentiation in the relevant genes or linked regions. Our work provides a valuable glimpse into the developmental systems that construct the trophic phenotype of Arctic charr, inparticular the SB-charr. At the same time this paves the way for future studies and emphasizes the need to sample more individuals and populations, track families, correlate genetic variation and expression at specific developmental stages or in cell types.

We are confident that further work on charr and other fresh-water fishes will yield results that constitute significant steps towards a synthesis and synergy of the principles of evolutionary and developmental genetics of ecologically important traits and resource polymorphism.

ACKNOWLEDGMENTS

This project was supported by The Icelandic Center for Research (RANNIS #100204011) to SSS, AP, ZOJ and BKK and The University of Iceland Research/Doctoral Fund to SSS, AP, ZOJ, JG and KHK. We would like to thank Baldur Kristjánsson for help with genomic region and alignments, and especially Droplaug N. Magnúsdóttir, Guðbjörg Þ. Örlygsdóttir, Steinunn Snorradóttir and Ólafur Þ. Magnússon at deCODE Genetics for help with the Illumina sequencing.

AUTHOR CONTRIBUTIONS

Conceived and designed the study: JG, AP, ZOJ, SSS, SRF, VHM, EPA
Sampling, crosses and rearing was done by SSS, BKK, ZOJ, KHK, AP

RNA extraction and RNA sequencing, by SRF
Analyses of RNA sequencing data, JG, AP
qPCR work, EPA, SSS2, VHM
SNP analyses JG, AP
SNP confirmation IMJ, KHK, AP
Comparative genomic analysis AP
Writing: AP, JG, EPA, VHM, SSS
Analyses: JG, AP, EPA, SSS2
Gathered the data: ZOJ, SRF, EPA, IAJ, KHK, SSS2

FIGURES

Figure 1: The two contrasting morphs used in this study. Adult individuals of the two morphs; the Holar aquaculture charr above and the small benthic charr from Lake Thingvallavatn below. Differences in size, coloration and head morphology are apparent.

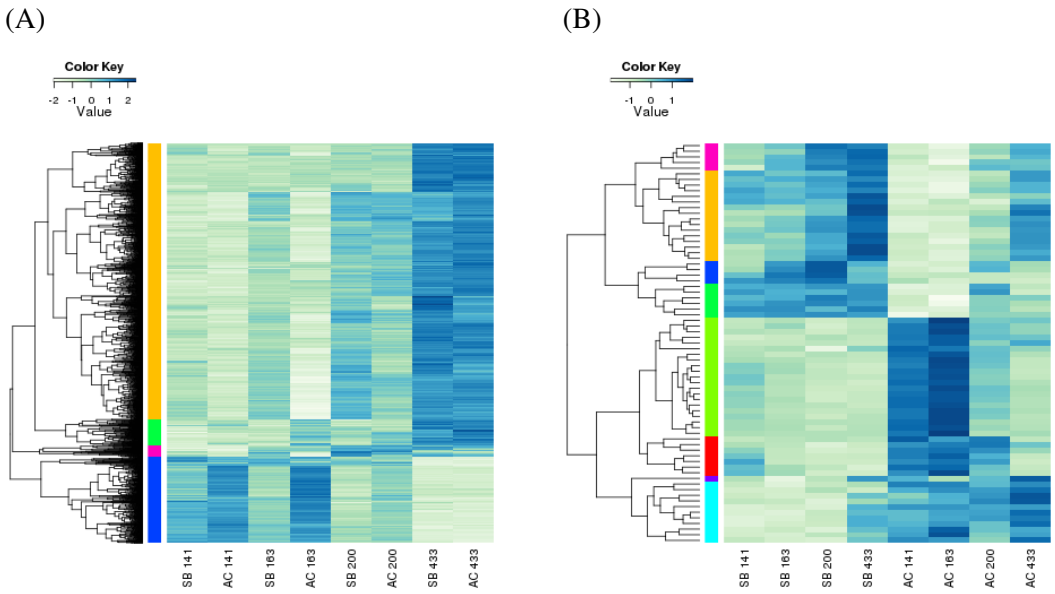

Figure 2: Heatmap of differentially expressed genes in the Arctic charr developmental transcriptome. Two morphs (SB and AC) are represented, at four timepoints. (A) The 1603 genes with expression difference among time points, here clustered into four groups. (B) The 71 genes differentially expressed between morphs, are clustered into 4 groups for each of the two morphs. High expression is indicated by blue and low expression by beige. Pictures of animals

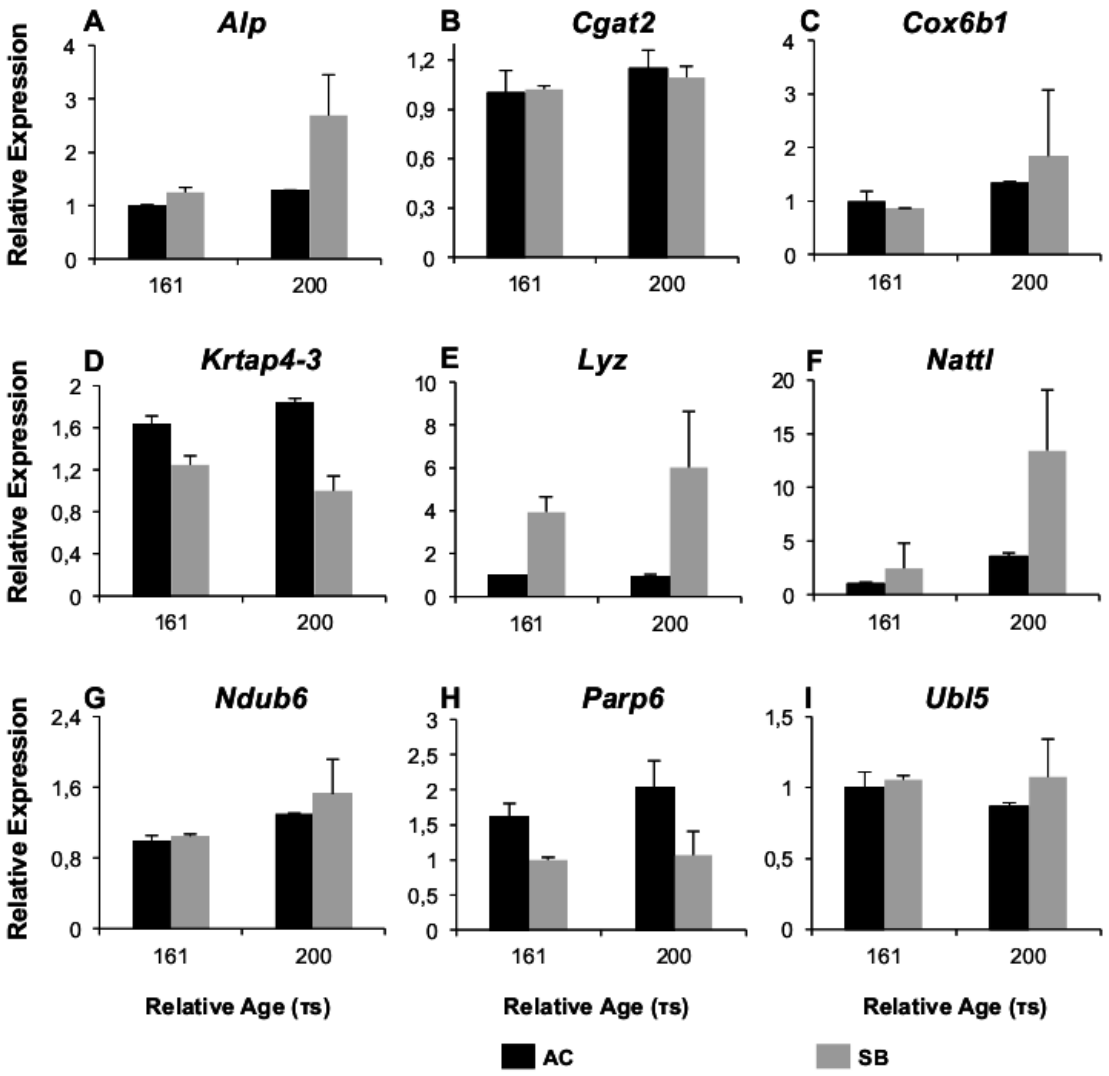

Figure 3: qPCR validation of candidates from transcriptome in whole embryos of Arctic charr. Relative expression of 9 genes (A-I) analysed by qPCR in the small benthic (SB) charr from Lake Thingvallavatn and aquaculture (AC) charr at two different developmental timepoints (161 and 200 τs). 5 genes were differentially expressed between the two morphs *(Alp, Krtap4-3, Lyz, Nattl, Parp6)*, while 4 further genes did not show significant expression differences between morphs *(Cgat2, Cox6B1, Ndub6, Ubl5)*, see Table S3. Error bars represent standard deviation calculated from two biological replicates.

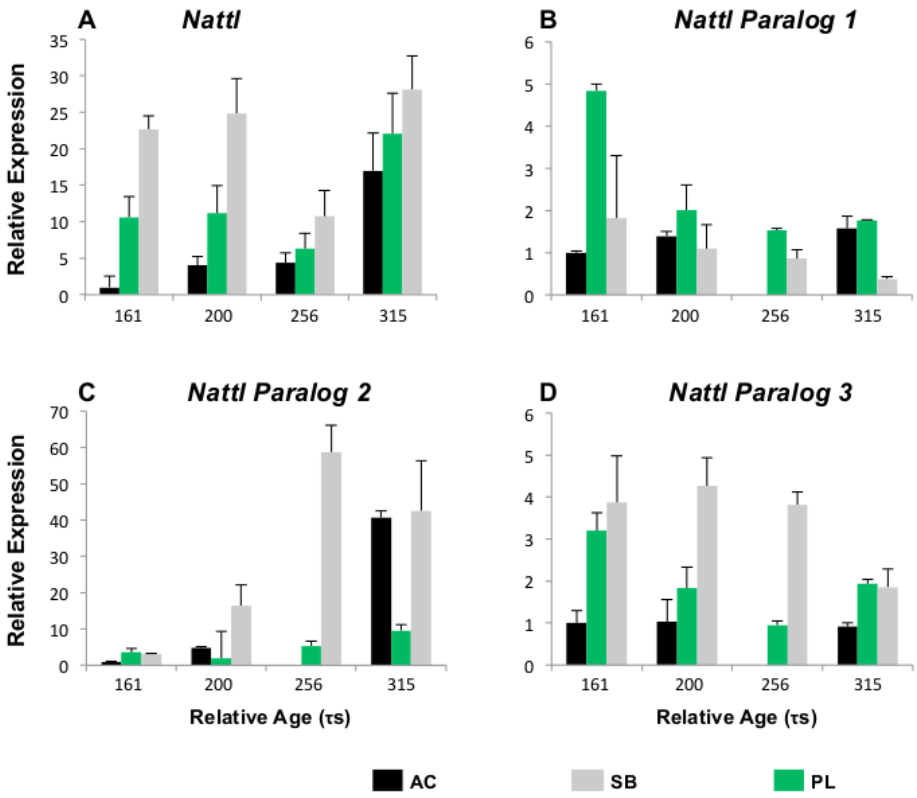

Figure 4: Relative expression of Nattl and three of its paralogs during charr development in different morphs. The expression is graphed for different morphs (SB, AC and PL) at four developmental timepoints (161, 200, 256 & 315 τs, relative to AC-charr at timepoint 161. A) General *Nattl* expression along charr development. B-D) Expression of *Nattl paralogs* 1-3. ANOVA showing the variation among morphs is summarized in Table S4.

Gene	Morph (p-value)	HSD morph				Time (p-value)	HSD time (Ts)			M x T (p-value)
Cldn4	5.30e-07	AC	PL	SB	LB	4.77e-05	178	200	216	2.64e-03
Jup	7.18e-05	AC	PL	SB	LB	8.18e-04	178	200	216	0.754
Lsr	1.6e-06	AC	PL	SB	LB	2.10e-04	178	200	216	0.462
Mvp	1.25e-05	AC	PL	SB	LB	1.91e-03	178	200	216	0.94
Rarg	0.031	AC	PL	SB	LB	1.37e-03	178	200	216	0.321
Scin	2.43e-07	AC	PL	SB	LB	0.26	NS			3.53e-04
Tgfbr2	1.85e-05	AC	PL	SB	LB	8.48e-07	178	200	216	0.021
Vdra	2.20e-04	AC	PL	SB	LB	3.68e-06	178	200	216	0.285

Figure 5: Expression differences of craniofacial candidate genes in developing head of Arctic charr morphs. Relative expression ratios, calculated from the qPCR data, were subjected to an ANOVA to test the expression differences amongst four charr groups and three close time points (τs). The underlined gene names reflect significant difference between the Small Benthic and aquaculture Limnetic charr. A post hoc Tukey's test (HSD) was performed to determine the effects of morphs, time and morph-time interaction (M X T). White boxes represent low expression, while black boxes represent high expression. The shading represents significant different expression between the samples (α = 0.05, NS = not significant).

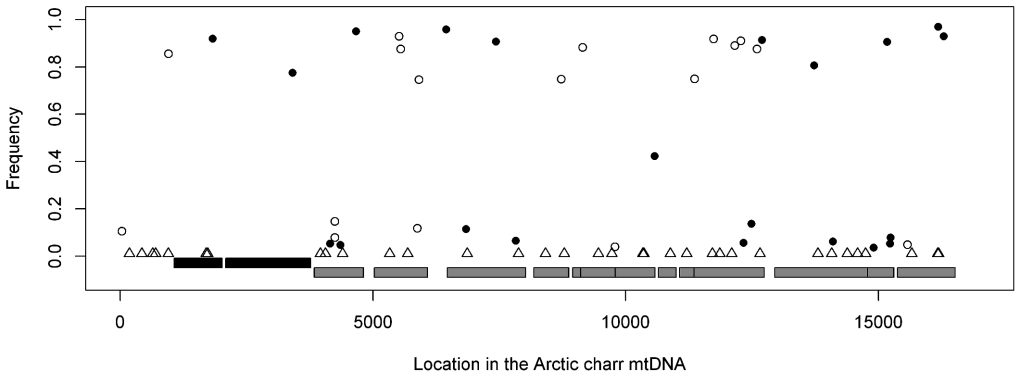

Figure 6: Genetic divergence in the mtDNA between SB- and AC-charr. The frequency differences between morphs of candidate SNPs estimated from the RNA-sequencing, graphed along the mtDNA chromosome. The SNPs indicate whether the derived allele is higher in SB (black dots) or AC (open circles). Sites of divergence between the Icelandic stocks and the Canadian reference sequence are indicated by triangles. The two black boxes represent the 12s (left) and 16s (right) rRNA genes, and gray boxes the 14 coding regions.

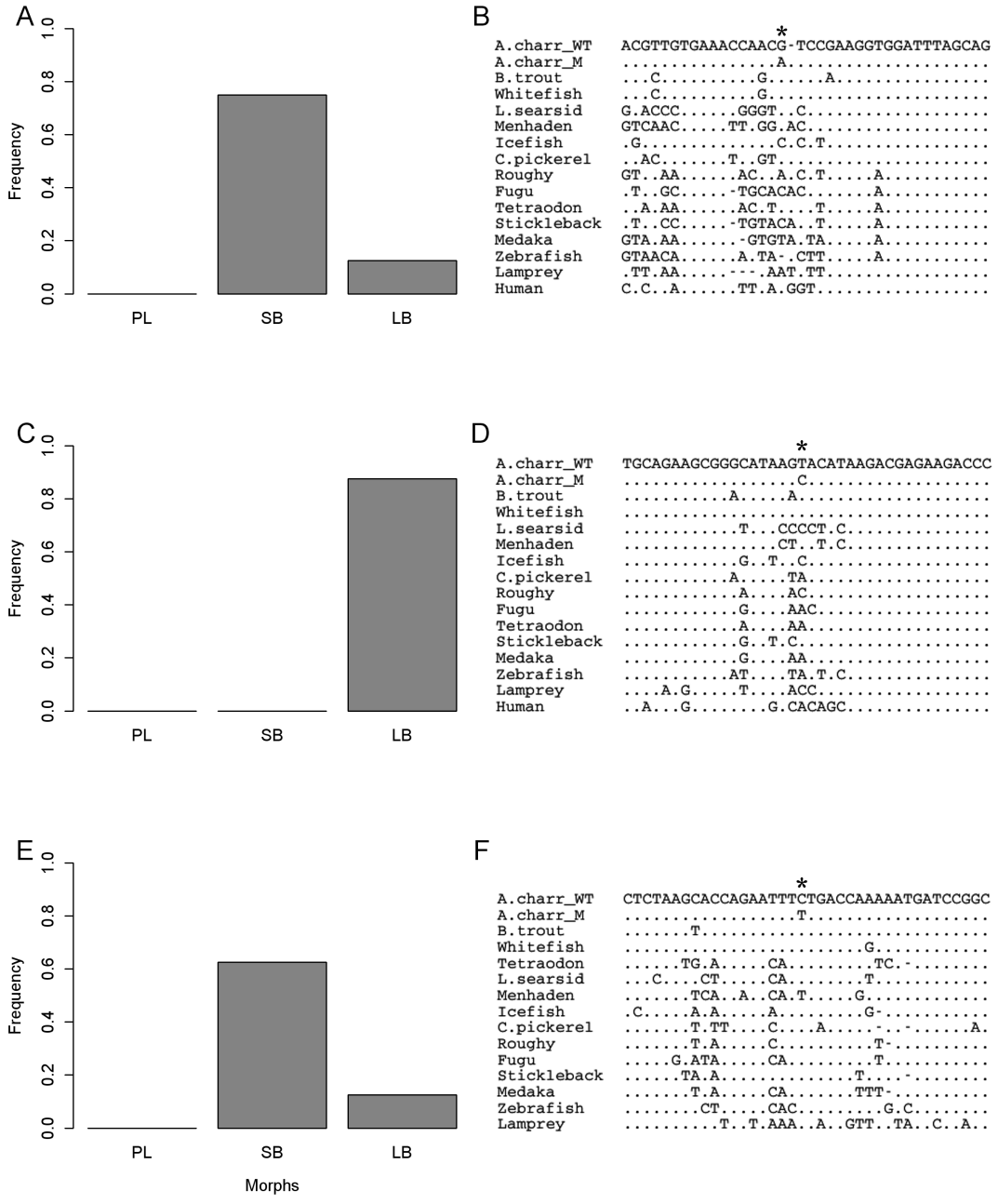

Figure 7: Comparative genomics and population genetic differentiation in Arctic charr at 3 mtDNA locations. Aligned are several fish genomes, with Lamprey or humans as outgroups, reflecting a 38 bp window around each of the 3 positions (asterix). A, C, E) Frequency of each of those variants in three Arctic charr populations from Lake Thingvallavatn (PL, LB and SB). A total of 8 individuals were genotyped from each morph, see methods. B) Alignment of variant m1829G>A in the 12s rRNA gene in fishes, using humans as an outgroup. D) Similar alignment of a 16s variant, m3211T>C and F) alignment of variant m3411C>T in the 16s rRNA gene.

TABLES

Table 1.
Differentially expressed genes, with higher expression in the SB morph from Lake Thingvallavatn.

NR	Unigene.Description	NR.contigs	logFC	logCPM	FDR	Cluster
3766	Histone H3 embryonic	1	8.71	2.74	7.80E-035	S-1
5103	Natterin-like protein	6	2.75	7.12	7.76E-007	S-2
356	A7J6M9 Putative uncharacterized protein n175R	1	2.33	4.66	3.30E-006	S-1
6697	Q1KY05 Main olfactory receptor-like protein	5	3.12	6.92	9.96E-005	S-1
8151	Sushi domain-containing protein 2	4	2.20	5.55	0.0001	S-3
1682	Carcinoembryonic antigen-related cell adhesion molecule 1	3	2.55	3.83	0.0002	S-1
6228	Protein FAM98A	2	1.96	4.76	0.0003	S-1
7531	STAM-binding protein-like	2	2.07	2.62	0.0005	S-1
6712	Q1M160 Myc-regulated DEAD box protein	1	1.67	3.23	0.0009	S-1
2300	Cytosolic sulfotransferase 3	3	1.73	2.13	0.0009	S-1
2063	Complement factor D	7	1.79	6.42	0.0016	S-3
3326	Galectin-3-binding protein A	4	1.79	3.85	0.0017	S-4
3169	Flocculation protein FLO11	2	1.86	4.05	0.0017	S-1
1203	B5XDY0 H-2 class I histocompatibility antigen L-D alpha chain	2	1.70	2.12	0.0028	S-3
9183	UPI000065D844 UPI000065D844 related cluster	2	1.97	5.55	0.0028	S-1
2909	Epidermis-type lipoxygenase 3	4	1.68	4.84	0.0029	S-1
4884	Myeloperoxidase	4	2.20	6.78	0.0029	S-1
10003	Uridine phosphorylase 1	4	1.51	3.00	0.0047	S-1
2513	Desmoglein-1-alpha	1	1.59	2.80	0.0054	S-2
377	A7SJA8 Predicted protein (Fragment)	1	1.73	2.50	0.0055	S-3
9204	UPI00006A2900 UPI00006A2900 related cluster	2	6.38	3.26	0.0064	S-1
9642	UPI00017B1B0F UPI00017B1B0F related cluster	1	2.00	1.92	0.0064	S-2
1965	Coiled-coil domain-containing protein 136	2	2.15	2.32	0.0064	S-2
9260	UPI0000F1D4BA PREDICTED: hypothetical protein	1	1.80	2.41	0.0065	S-2
738	Adseverin	8	1.58	5.51	0.0073	S-1
9678	UPI00017B4479 UPI00017B4479 related cluster	1	2.18	1.97	0.0074	S-4
8339	Testin	4	1.50	4.93	0.0080	S-2
6840	Q4SNH3 Chromosome 8 SCAF14543 whole genome shotgun sequence	1	1.42	4.00	0.0080	S-1
1668	Carbohydrate sulfotransferase 6	1	2.09	2.08	0.0090	S-4
8341	Testisin	2	2.01	2.76	0.0090	S-4
6373	Protein asteroid homolog 1	6	1.29	4.24	0.0090	S-4

logFC – log Fold Change
logCPM – log Counts Per Million
FDR – False Discovery Rate
The number of the cluster corresponds to Figure 1.

Table 2.
Differentially expressed genes, with higher expression in the AC morph.

NR	Unigene.Description	NR.contigs	logFC	logCPM	FDR	Cluster
3465	Glutathione S-transferase P 1	1	-8.35	2.45	1.12E-019	A-2
2475	Dehydrogenase/reductase SDR family member 7	2	-4.88	2.15	9.67E-014	A-3
6945	Q6NWE8 Sb:cb283 protein (Fragment)	3	-6.08	3.02	2.15E-013	A-2
399	A8DW32 Predicted protein	1	-5.32	6.38	4.27E-010	A-1
9682	UPI00017B4B48 UPI00017B4B48 related cluster	2	-3.70	2.81	2.61E-008	A-2
9817	Uncharacterized protein ART2	5	-12.63	6.89	8.23E-008	A-2
6724	Q2L0Z2 Putative ATP-dependent RNA helicase	1	-3.41	1.89	1.88E-007	A-2
1197	B5XD10 Vacuolar proton pump subunit G 1	1	-4.30	2.10	1.84E-006	A-2
5325	Nucleoside diphosphate kinase B	1	-9.85	7.63	2.51E-006	A-1
9205	UPI0000D5B923 PREDICTED: myelin basic protein isoform 1	3	-2.49	3.45	9.18E-006	A-3
6377	Protein broad-minded	1	-2.11	2.74	4.75E-005	A-1
5711	Pistil-specific extensin-like protein	1	-2.16	2.60	0.0002	A-3
3203	Formin-like protein 20	7	-1.98	1.95	0.0002	A-3
9315	UPI0000F2EC69 PREDICTED: hypothetical protein	2	-5.60	4.57	0.0005	A-2
363	A7RFV0 Predicted protein (Fragment)	2	-1.74	4.96	0.0010	A-3
6937	Q6AZT1 MGC81677 protein	3	-2.06	3.81	0.0014	A-2
3756	Histone H1	3	-2.26	4.54	0.0017	A-2
1133	B5DGN9 Creatine kinase-1	7	-4.72	5.50	0.0017	A-3
309	A1IMH7 CD80-like protein	12	-1.94	4.29	0.0017	A-2
7651	Serine protease ami	2	-1.54	5.90	0.0017	A-3
9935	Uncharacterized protein C7orf63 homolog	1	-1.87	1.91	0.0025	A-2
5219	Nostrin	2	-2.55	3.38	0.0029	A-2
1855	Chondroitin sulfate N-acetylgalactosaminyltransferase 2	5	-2.56	6.14	0.0034	A-1
10203	Xylose isomerase	6	-1.55	2.43	0.0035	A-3
2249	Cytochrome c oxidase subunit 3	11	-1.78	11.15	0.0035	A-3
180	40S ribosomal protein S3-B	2	-5.31	8.67	0.0050	A-1
1227	B6NBL3 Putative uncharacterized protein	3	-1.59	2.95	0.0050	A-2
5055	NADH-ubiquinone oxidoreductase chain 6	2	-1.48	2.65	0.0061	A-3
4634	Metallothionein A	1	-3.33	5.44	0.0064	A-2
342	A5C0J4 Putative uncharacterized protein	2	-2.58	2.47	0.0064	A-2
9698	UPI00019258B4 PREDICTED: similar to epithelial cell transforming sequence 2 oncogene protein partial	1	-2.06	2.94	0.0064	A-2
5878	Pro-opiomelanocortin B	1	-2.04	5.60	0.0065	A-2
2248	Cytochrome c oxidase subunit 2	9	-2.21	9.83	0.0074	A-2
1246	B8JI87 Novel protein similar to vertebrate collagen type VI alpha 3 (COL6A3) (Fragment)	1	-1.69	3.16	0.0080	A-3
7994	Sperm-associated antigen 5	1	-2.07	3.71	0.0080	A-2
9515	UPI000175F90F PREDICTED: similar to pleckstrin homology domain containing family A member 7	1	-2.00	1.87	0.0090	A-2
1124	B5DDZ4 Acta1 protein	1	-1.52	2.62	0.0090	A-2
1127	B5DG94 2-peptidylprolyl isomerase A	2	-2.56	5.67	0.0090	A-1
9175	UPI000054A3C0 PREDICTED: apolipoprotein B	3	-1.32	4.09	0.0090	A-4
9671	UPI00017B3C62 UPI00017B3C62 related cluster	1	-1.51	1.92	0.0096	A-1

The number of the cluster corresponds to Figure 1.

Table 3. Candidate SNPs in the Arctic charr transcriptome, filtered by coverage, difference between sample and morphs and frequency difference between morphs. For Delta > 0.95 we show the number of SNP-candidates before the redundant ones were removed.

SNP-candidates	Morph	Uni	Rep	**Total**
Total		96,231	74,341	**165,790**
Filter coverage		66,569	57,009	**113,776**
Diff. Bwn. samples		21,417	22,252	**42,869**
Diff. Bwn morphs		11,385	12,953	**23,974**
Delta > 0.5	AC	396	285	**672**
Delta > 0.5	SB	526	353	**872**
Delta > 0.75	AC	95	68	**159**
Delta > 0.75	SB	155	95	**248**
Delta > 0.95[a]	AC	17	13	**30**
Delta > 0.95[a]	SB	29	4	**33**

SNP-candidates are found by mapping to *S. salar* ESTs.
From UNIquely or REPeadely mapped RNA-reads.
Delta: Differences in allele frequency between morphs, categorized by which morph had the higher derived allele frequency.

Table 4A.
SNP candidates with higher frequency in AC morph from uniquely mapped reads.

Contig	Annotation	Position	Reference	Variant	Freq-SB	Freq-AC	Effect
SS2U026955	Keratin type II cytoskeletal 3	300	A	T	0.000	0.984	synonymous
SS2U026955	Keratin type II cytoskeletal 3	309	G	A	0.000	0.996	synonymous
SS2U033960	Cysteine sulfinic acid decarboxylase	192	C	G	0.000	1.000	5prime
SS2U033960	Cysteine sulfinic acid decarboxylase	416	G	T	0.000	0.961	G to V / Special to Hydrophobic
SS2U033960	Cysteine sulfinic acid decarboxylase	945	C	A	0.004	0.956	synonymous
SS2U043396	Eukaryotic translation initiation factor 2-alpha kinase 1	134	A	G	0.000	1.000	5prime
SS2U043886	Transcription cofactor HES-6	1308	T	C	0.000	1.000	5prime
SS2U044339	Intraflagellar transport protein 52 homolog	479	T	C	0.021	1.000	D to G / Negative to Special
SS2U045168	Putative Peptide prediction	1275	G	A	0.000	1.000	3prime
SS2U045328	E3 ubiquitin-protein ligase DTX3L	388	G	A	0.000	0.977	synonymous
SS2U045990	Low-density lipoprotein receptor-related protein 1	135	T	C	0.000	0.969	synonymous
SS2U048125[a]	Transmembrane protein 131-like	480	G	A	0.000	1.000	synonymous
SS2U052747	Uridine 5'-monophosphate synthase	914	G	A	0.000	0.951	synonymous
SS2U054542	Mediator of RNA polymerase II transcription subunit 20	474	C	T	0.027	0.995	synonymous
SS2U056193	SUMO-conjugating enzyme UBC9	96	A	T	0.000	1.000	3prime
SS2U057101	ETS domain-containing protein Elk-3	440	C	G	0.000	1.000	3prime
SS2U058860	Voltage-dependent anion-selective channel protein 2	681	G	T	0.000	1.000	3prime

Table 4B.
SNP candidates with higher frequency in SB morph, from uniquely mapped reads.

SS2U000399	Insulin-like growth factor-binding protein 7	598	C	A	1.000	0.000	3prime
SS2U004484	Titin	387	G	A	0.990	0.010	synonymous
SS2U026826	L-asparaginase	363	C	T	1.000	0.000	H to Y / Positive to Hydrophobic
SS2U026955	Keratin type II cytoskeletal 3	116	C	A	0.996	0.031	T to N / Polar to Polar
SS2U026955	Keratin type II cytoskeletal 3	264	C	T	0.970	0.008	synonymous
SS2U026955	Keratin type II cytoskeletal 3	317	C	T	1.000	0.002	T to M / Polar to Hydrophobic
SS2U033960	Cysteine sulfinic acid decarboxylase	363	C	T	1.000	0.025	5prime
SS2U033960	Cysteine sulfinic acid decarboxylase	387	C	T	1.000	0.030	synonymous
SS2U033960	Cysteine sulfinic acid decarboxylase	657	T	C	0.990	0.031	synonymous
SS2U034322	Cyclin-C	1094	A	G	1.000	0.000	3prime
SS2U034431	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	436	G	A	0.992	0.000	G to S / Special to Polar
SS2U036025	Nuclear receptor coactivator 4	36	G	A	1.000	0.043	5prime
SS2U040590	Glutamyl-tRNA(Gln) amidotransferase subunit A homolog	478	G	A	0.972	0.000	synonymous
SS2U045606	Superkiller viralicidic activity 2-like 2	500	C	T	1.000	0.000	synonymous
SS2U047816	Squalene synthase	1139	G	A	1.000	0.029	synonymous
SS2U048063	Lysine-specific demethylase NO66	669	C	T	1.000	0.000	synonymous
SS2U050394	UPF0542 protein C5orf43 homolog	596	G	A	1.000	0.000	synonymous
SS2U050880[a]	Transmembrane protein 131-like	901	C	T	1.000	0.000	A to V / Hydrophobic to Hydrophobic
SS2U052076	Eukaryotic translation initiation factor 3 subunit A	824	C	T	1.000	0.031	synonymous
SS2U053417	RNA polymerase-associated protein LEO1	454	G	A	1.000	0.049	synonymous
SS2U054333	Scaffold attachment factor B2	382	G	A	0.999	0.000	V to M / Hydrophobic to Hydrophobic
SS2U054705	Cell division protein kinase 4	122	A	G	0.971	0.000	3prime
SS2U054965	DNA-directed RNA polymerase I subunit RPA12	106	G	A	1.000	0.000	5prime
SS2U054965	DNA-directed RNA polymerase I subunit RPA12	411	T	G	1.000	0.000	synonymous
SS2U055120	Chromatin modification-related protein MEAF6	350	A	C	1.000	0.000	H to P / Positive to Special
SS2U055153	Complexin-1	1191	C	A	1.000	0.031	3prime

SS2U057635	Mitogen-activated protein kinase 14B	1370	A	T	1.000	0.026	3prime
SS2U058169	Transmembrane protein 50A	1214	C	G	0.973	0.000	3prime
SS2U058802	Signal recognition particle 54 kDa protein	607	T	A	0.969	0.000	C to S / Special to Polar

[a]Those genes are distinct paralogs.

Table 5.
SNP candidates with significant difference frequency between AC and SB morphs, from reads that mapped to two or more contigs.

Contig	Annotation	Position	Reference	Variant	Freq-SB	Freq-AC	Effect
SS2U004839	Actin alpha sarcomeric/cardiac	550	A	C	0.015	0.999	3prime
SS2U021298	28S ribosomal protein S18a mitochondrial	462	A	C	0.000	1.000	synonymous
SS2U041264	Apoptosis-inducing factor 1 mitochondrial	341	C	T	0.000	0.952	synonymous
SS2U054211[a]	Cytoplasmic dynein 1 intermediate chain 2	136	T	C	0.018	0.974	synonymous
SS2U054362[a]	Q08CA8 Dynein cytoplasmic 1 intermediate chain 2	945	A	G	0.000	1.000	synonymous
SS2U055923	Bystin	1623	A	C	0.000	0.983	3prime
SS2U058758	Protein S100-A1	253	C	T	0.000	0.984	synonymous
SS2U059000	Isocitrate dehydrogenase [NADP] mitochondrial	1654	T	C	0.000	0.975	3prime
SS2U059146	60S ribosomal protein L36	263	T	G	0.009	1.000	synonymous
SS2U059146	60S ribosomal protein L36	470	A	C	0.009	1.000	synonymous
SS2U036667	Heterogeneous nuclear ribonucleoprotein K	813	C	T	1.000	0.022	5prime
SS2U042873	RNA polymerase-associated protein LEO1	460	G	A	1.000	0.000	synonymous
SS2U058455	Adenylosuccinate lyase	1616	C	T	1.000	0.000	3prime
SS2U058906	Mid1-interacting protein 1-like	350	G	T	0.985	0.000	E to D / Negative to Negative

[a]Those genes are distinct paralogs

LITERATURE CITED

Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant, and C. J. Tabin, 2004 Bmp4 and morphological variation of beaks in Darwin's finches. Science **305**: 1462–1465. 2

Abzhanov, A., and C. J. Tabin, 2004 Shh and Fgf8 act synergistically to drive cartilage outgrowth during cranial development. Developmental Biology **273**: 134–148. 2

Adams, C. E., D. Fraser, A. J. Wilson, G. Alexander, M. M. Ferguson, *et al.*, 2007 Patterns of phenotypic and genetic variability show hidden diversity in Scottish Arctic charr. Ecology of Freshwater Fish **16**: 78–86. 3

Ahi, E. P., J. Guðbrandsson, K. H. Kapralova, S. R. Franzdóttir, S. S. Snorrason, *et al.*, 2013 Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr. PloS ONE **8**: e66389. 4, 5, 6, 7, 11, 17

Ahi, E. P., K. H. Kapralova, A. Pálsson, V. H. Maier, J. Gudbrandsson, *et al.*, 2014 Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr. EvoDevo **5**: 40. 4, 17, 18

Albertson, R. C., and T. D. Kocher, 2006 Genetic and developmental basis of cichlid trophic diversity. Heredity **97**: 211–221. 2

Allendorf, F. W., and G. H. Thorgaard, 1984 Tetraploidy and the Evolution of Salmonid Fishes. In B. J. Turner, editor, *Evolutionary Genetics of Fishes*. Springer US, 1–53. 3

Ballard, J. W. O., and N. Pichaud, 2014 Mitochondrial DNA: more than an evolutionary bystander. Functional Ecology **28**: 218–231. 15

Berger, W., E. Steiner, M. Grusch, L. Elbling, and M. Micksche, 2009 Vaults and the major vault protein: novel roles in signal pathway regulation and immunity. Cellular and molecular life sciences : CMLS **66**: 43–61. 17

Bernatchez, L., S. Renaut, A. R. Whiteley, N. Derome, J. Jeukens, *et al.*, 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**: 1783–800. 2

Berthelot, C., F. Brunet, D. Chalopin, A. Juanchich, M. Bernard, *et al.*, 2014 The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature communications **5**: 3657. 3, 6, 12, 14

Bougas, B., C. Audet, and L. Bernatchez, 2013 The influence of parental effects on transcriptomic landscape during early development in brook charr (Salvelinus fontinalis, Mitchill). Heredity **110**: 484–491. 3, 15

Bozinovic, G., T. L. Sit, D. E. Hinton, and M. F. Oleksiak, 2011 Gene expression throughout a vertebrate's embryogenesis. BMC genomics **12**: 132. 3

Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, *et al.*, 2009 The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry **55**: 611–22. 7

Chai, Y., Y. Ito, and J. Han, 2003 TGF-beta signaling and its functional significance in regulating the fate of cranial neural crest cells. Critical reviews in oral biology and medicine : an official publication of the American Association of Oral Biologists **14**: 78–88. 17

Chipman, D. M., and N. Sharon, 1969 Mechanisms of Lysozyme Action. Science **165**: 454–465. 16

Coppe, A., C. Agostini, I. A. M. Marino, L. Zane, L. Bargelloni, *et al.*, 2013 Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. Genome biology and evolution **5**: 45–60. 15

Cresko, W. A., K. L. McGuigan, P. C. Phillips, and J. H. Postlethwait, 2007 Studies of threespine stickleback developmental evolution: Progress and promise. Genetica **129**: 105–126. 2

Crête-Lafrenière, A., L. K. Weir, and L. Bernatchez, 2012 Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. PloS ONE **7**: e46662. 3

Danzmann, R. G., M. M. Ferguson, S. Skulason, S. S. Snorrason, and D. L. G. Noakes, 1991 Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, Salvelinus alpinus L., from Thingvallavatn, Iceland. Journal of Fish Biology **39**: 649–659. 4, 15

Davidson, W. S., B. F. Koop, S. J. M. Jones, P. Iturra, R. Vidal, *et al.*, 2010 Sequencing the genome of the Atlantic salmon (Salmo salar). Genome biology **11**: 403. 3

Derome, N., P. Duchesne, and L. Bernatchez, 2006 Parallelism in gene transcription among sympatric lake whitefish (Coregonus clupeaformis Mitchill) ecotypes. Molecular ecology **15**: 1239–49. 3

Di Génova, A., A. Aravena, L. Zapata, M. González, A. Maass, *et al.*, 2011 SalmonDB: a bioinformatics resource for Salmo salar and Oncorhynchus mykiss. Database : the journal of biological databases and curation **2011**: bar050. 6, 9, 14

Doiron, S., L. Bernatchez, and P. U. Blier, 2002 A comparative mitogenomic analysis of the potential adaptive value of Arctic charr mtDNA introgression in brook charr populations (Salvelinus fontinalis Mitchill). Molecular biology and evolution **19**: 1902–9. 4, 6, 13

Domazet-Lošo, T., and D. Tautz, 2010 A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature **468**: 815–8. 3, 15

Drivenes, O. y., G. L. Taranger, and R. B. Edvardsen, 2012 Gene expression profiling of Atlantic cod (Gadus morhua) embryogenesis using microarray. Marine biotechnology **14**: 167–76. 15

Eiriksson, G. M., 1999 *Heterochrony in bone development and growth in two morphs of Arctic charr (Salvelinus alpinus) from Thingvallavatn, Iceland*. Master's thesis, University of Iceland. 17

Eiriksson, G. M., S. Skulason, and S. S. Snorrason, 1999 Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn, Iceland. Journal of Fish Biology **55**: 175–185. 11

Filteau, M., S. A. Pavey, J. St-Cyr, and L. Bernatchez, 2013 Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. Molecular biology and evolution **30**: 1384–96. 3

Fleming, A., 1922 On a Remarkable Bacteriolytic Element Found in Tissues and Secretions. Proceedings of the Royal Society of London. Series B **93**: 306–317. 16

Flint, J., and T. F. C. Mackay, 2009 Genetic architecture of quantitative traits in mice, flies, and humans. Genome research **19**: 723–33. 2

Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor, 2011 Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity **106**: 404–20. 3

Furuse, M., Y. Oda, T. Higashi, N. Iwamoto, and S. Masuda, 2012 Lipolysis-stimulated lipoprotein receptor: a novel membrane protein of tricellular tight junctions. Annals of the New York Academy of Sciences **1257**: 54–8. 17

Gibson, G., and B. Weir, 2005 The quantitative genetics of transcription. Trends in genetics : TIG **21**: 616–23. 2

Giger, T., L. Excoffier, P. J. R. Day, A. Champigneulle, M. M. Hansen, *et al.*, 2006 Life history shapes gene expression in salmonids. Current Biology **16**. 2

Goodale, B. C., J. K. La Du, W. H. Bisson, D. B. Janszen, K. M. Waters, *et al.*, 2012 AHR2 mutant reveals functional diversity of aryl hydrocarbon receptors in zebrafish. PloS ONE **7**: e29346. 17

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**: 109–127. 5

Gould, S. J., 1977 *Ontogeny and Phylogeny*. Harvard University Press. 2

Gross, J. B., A. Furterer, B. M. Carlson, and B. A. Stahl, 2013 An Integrated Transcriptome-Wide Analysis of Cave and Surface Dwelling Astyanax mexicanus. PLoS ONE **8**. 3

Hering, N. A., S. Andres, A. Fromm, E. A. van Tol, M. Amasheh, *et al.*, 2011 Transforming growth factor-β, a whey protein component, strengthens the intestinal barrier by upregulating claudin-4 in HT-29/B6 cells. The Journal of nutrition **141**: 783–9. 17

Ito, Y., J. Y. Yeo, A. Chytil, J. Han, P. Bringas, *et al.*, 2003 Conditional inactivation of Tgfbr2 in cranial neural crest causes cleft palate and calvaria defects. Development **130**: 5269–80. 17

Jacob, F., 1977 Evolution and tinkering. Science **196**: 1161–1166. 2

Jantzen, S. G., D. S. Sanderson, K. R. von Schalburg, M. Yasuike, F. Marass, *et al.*, 2011 A 44K microarray dataset of the changing transcriptome in developing Atlantic salmon (Salmo salar L.). BMC research notes **4**: 88. 15

Jazag, A., H. Ijichi, F. Kanai, T. Imamura, B. Guleng, *et al.*, 2005 Smad4 silencing in pancreatic cancer cell lines using stable RNA interference and gene expression profiles induced by transforming growth factor-beta. Oncogene **24**: 662–71. 17

Jeyaprakash, A. A., P. Geetha Rani, G. Banuprakash Reddy, S. Banumathi, C. Betzel, *et al.*, 2002 Crystal structure of the jacalin-T-antigen complex and a comparative study of lectin-T-antigen complexes. Journal of molecular biology **321**: 637–45. 16

Jonsson, B., S. Skúlason, S. S. Snorrason, O. T. Sandlund, H. J. Malmquist, *et al.*, 1988 Life History Variation of Polymorphic Arctic Charr (Salvelinus alpinus) in Thingvallavatn, Iceland. Canadian Journal of Fisheries and Aquatic Sciences **45**: 1537–1547. 13

Kapralova, K. H., S. R. Franzdóttir, H. Jónsson, S. S. Snorrason, and Z. O. Jónsson, 2014a Patterns of MiRNA Expression in Arctic Charr Development. PLoS ONE **9**: e106084. 4, 6

Kapralova, K. H., J. Gudbrandsson, S. Reynisdottir, C. B. Santos, V. C. Baltanás, *et al.*, 2013 Differentiation at the MHCIIα and Cath2 loci in sympatric Salvelinus alpinus resource morphs in Lake Thingvallavatn. PloS ONE **8**: e69402. 4, 5, 8, 15, 16

Kapralova, K. H., Z. O. Jónsson, A. Pálsson, S. R. Franzdóttir, S. Le Deuf, *et al.*, 2014b Bones in motion: Ontogeny of craniofacial development in sympatric Arctic charr morphs. submitted to Developmental Dynamics . 11

Kapralova, K. H., M. B. Morrissey, B. K. Kristjánsson, G. A. Olafsdóttir, S. S. Snorrason, *et al.*, 2011 Evolution of adaptive diversity and genetic connectivity in Arctic charr (Salvelinus alpinus) in Iceland. Heredity **106**: 472–487. 4, 15, 18

Kedersha, N. L., M. C. Miquel, D. Bittner, and L. H. Rome, 1990 Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes. The Journal of cell biology **110**: 895–901. 17

Kimball, S., P. Mattis, and GIMP Development Team, 1995 GIMP: GNU Image Manipulation Program. 9

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

Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, *et al.*, 2012 VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research **22**: 568–76. 7

Kristjánsson, B. K., H. J. Malmquist, F. Ingimarsson, T. Antonsson, S. S. Snorrason, *et al.*, 2011 Relationships between lake ecology and morphological characters in Icelandic Arctic charr, Salvelinus alpinus. Biological Journal of the Linnean Society **103**: 761–771. 3, 18

Kristjánsson, B. K., S. Skúlason, S. S. Snorrason, and D. L. Noakes, 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**: 1099–112. 3

Lanes, C. F. C., T. T. Bizuayehu, J. M. de Oliveira Fernandes, V. Kiron, and I. Babiak, 2013 Transcriptome of Atlantic Cod (Gadus morhua L.) Early Embryos from Farmed and Wild Broodstocks. Marine Biotechnology **15**: 677–694. 3

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. Mcgettigan, *et al.*, 2007 Clustal W and Clustal X version 2.0. Bioinformatics **23**: 2947–2948. 8

Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**: 323. 6

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754–60. 7

Livak, K. J., and T. D. Schmittgen, 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25**: 402–8. 7

Macqueen, D. J., and I. A. Johnston, 2014 A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proceedings of the Royal Society B: Biological Sciences **281**: 20132881. 3

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

Magalhães, G. S., I. L. M. Junqueira-de Azevedo, M. Lopes-Ferreira, D. M. Lorenzini, P. L. Ho, *et al.*, 2006 Transcriptome analysis of expressed sequence tags from the venom glands of the fish Thalassophryne nattereri. Biochimie **88**: 693–9. 11, 16

Magalhães, G. S., M. Lopes-Ferreira, I. L. M. Junqueira-de Azevedo, P. J. Spencer, M. S. Araújo, *et al.*, 2005 Natterins, a new class of proteins with kininogenase activity characterized from Thalassophryne nattereri fish venom. Biochimie **87**: 687–99. 11, 16

Magnadóttir, B., 2006 Innate immunity of fish (overview). Fish & shellfish immunology **20**: 137–51. 16

Magnadottir, B., 2010 Immunological control of fish diseases. Marine biotechnology **12**: 361–79. 16

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

Malmquist, H. J., S. S. Snorrason, S. Skulason, B. Jonsson, O. T. Sandlund, *et al.*, 1992 Diet differentiation in polymorphic Arctic charr in Thingvallavatn , Iceland. Journal of Animal Ecology **61**: 21–35. 4

McBride, H. M., M. Neuspiel, and S. Wasiak, 2006 Mitochondria: more than just a powerhouse. Current biology : CB **16**: R551–60. 15

Merilä, J., 2013 Nine-spined stickleback (Pungitius pungitius): an emerging model for evolutionary biology research. Annals of the New York Academy of Sciences **1289**: 18–35. 2

Micallef, G., R. Bickerdike, C. Reiff, J. M. O. Fernandes, A. S. Bowman, *et al.*, 2012 Exploring the Transcriptome of Atlantic Salmon (Salmo salar) Skin, a Major Defense Organ. Marine Biotechnology **14**: 559–569. 3

Miller, W., S. C. Schuster, A. J. Welch, A. Ratan, O. C. Bedoya-Reina, *et al.*, 2012 Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proceedings of the National Academy of Sciences of the United States of America **109**: E2382–90. 4

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

Nicholas, K. B., H. B. Nicholas, and D. W. Deerfield, 1997 GeneDoc: Analysis and Visualization of Genetic Variation. EMBNEW.NEWS **4**. 8

Noakes, D. L. G., 2008 Charr truth: Sympatric differentiation in Salvelinus species. Environmental Biology of Fishes **83**: 7–15. 3

Norman, J. D., M. M. Ferguson, and R. G. Danzmann, 2014 Transcriptomics of salinity tolerance capacity in Arctic charr (Salvelinus alpinus): a comparison of gene expression profiles between divergent QTL genotypes. Physiological genomics **46**: 123–37. 3, 14

Nurminsky, D., C. Magee, L. Faverman, and M. Nurminskaya, 2007 Regulation of chondrocyte differentiation by actin-severing protein adseverin. Developmental biology **302**: 427–37. 17

O'Brien, K. M., and I. A. Mueller, 2010 The unique mitochondrial form and function of Antarctic channichthyid icefishes. Integrative and comparative biology **50**: 993–1008. 15

Pálsson, S. b., and E. Árnason, 1994 Sequence variation for cytochrome b genes of three salmonid species from Iceland. Aquaculture **128**: 29–39. 4

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

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

Piasecka, B., P. Lichocki, S. Moretti, S. Bergmann, and M. Robinson-Rechavi, 2013 The hourglass and the early conservation models–co-existing patterns of developmental constraints in vertebrates. PLoS genetics **9**: e1003476. 15

Planchart, A., and C. J. Mattingly, 2010 2,3,7,8-Tetrachlorodibenzo-p-dioxin upregulates FoxQ1b in zebrafish jaw primordium. Chemical research in toxicology **23**: 480–7. 17

Puga, A., C. R. Tomlinson, and Y. Xia, 2005 Ah receptor signals cross-talk with multiple developmental pathways. Biochemical pharmacology **69**: 199–207. 17

Qian, X., Y. Ba, Q. Zhuang, and G. Zhong, 2014 RNA-Seq technology and its application in fish transcriptomics. Omics : a journal of integrative biology **18**: 98–110. 2, 3

Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–40. 6, 9

Rockman, M. V., 2012 The QTN program and the alleles that matter for evolution: All that's gold does not glitter. Evolution **66**: 1–17. 2, 18

Saemundsson, K., 1992 Geology of the Thingvallavatn Area. Oikos **64**: 40. 4

Sandlund, O. T., K. Gunnarsson, P. M. Jónasson, B. Jonsson, T. Lindem, *et al.*, 1992 The arctic charr Salvelinus alpinus in Thingvallavatn. Oikos **64**: 305–351. 4

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

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

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

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

Smith, S., L. Bernatchez, and L. Beheregaray, 2013 RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. BMC genomics **14**: 375. 3

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

Snorrason, S. S., S. Skúlason, B. Jonsson, H. J. Malmquist, P. M. Jónasson, *et al.*, 1994 Trophic specialization in Arctic charr Salvelinus alpinus (Pisces; Salmonidae): morphological divergence and ontogenetic niche shifts. Biological Journal of the Linnean Society **52**: 1–18. 4, 13

Snorrason, S. S., S. Skúlason, O. T. Sandlund, H. J. Malmquist, B. Jonsson, *et al.*, 1989 Shape polymorphism in arctic charr, Salvelinus alpinus. Physiology and Ecology Japan **1**: 393–404. 5

Sprague, J., L. Bayraktaroglu, D. Clements, T. Conlin, D. Fashena, *et al.*, 2006 The Zebrafish Information Network: the zebrafish model organism database. Nucleic acids research **34**: D581–D585. 11

Stern, D. L., and V. Orgogozo, 2008 The loci of evolution: how predictable is genetic evolution? Evolution **62**: 2155–77. 2

Stern, D. L., and V. Orgogozo, 2009 Is genetic evolution predictable? Science **323**: 746–751. 2

Subramanian, S., S. L. MacKinnon, and N. W. Ross, 2007 A comparative study on innate immune parameters in the epidermal mucus of various fish species. Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology **148**: 256–63. 16

Svavarsson, E., 2007 Árangur í kynbótum á bleikju og næstu skref [reference in Icelandic]. In *Fræðaþing landbúnaðarsins 4*. 121–125. 5

Svensson, C., A. E. Silverstone, Z.-W. Lai, and K. Lundberg, 2002 Dioxin-induced adseverin expression in the mouse thymus is strictly regulated and dependent on the aryl hydrocarbon receptor. Biochemical and biophysical research communications **291**: 1194–200. 17

Teacher, A. G., C. André, J. Merilä, and C. W. Wheat, 2012 Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. BMC evolutionary biology **12**: 248. 16

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, *et al.*, 2012 Primer3-new capabilities and interfaces. Nucleic Acids Research **40**. 7

van Driel, M., H. A. P. Pols, and J. P. T. M. van Leeuwen, 2004 Osteoblast differentiation and control by vitamin D and vitamin D metabolites. Current pharmaceutical design **10**: 2535–55. 17

Vieira, F. A., M. A. S. Thorne, K. Stueber, M. Darias, R. Reinhardt, *et al.*, 2013 Comparative analysis of a teleost skeleton transcriptome provides insight into its regulation. General and comparative endocrinology **191**: 45–58. 17

Vijay, N., J. W. Poelstra, A. Künstner, and J. B. W. Wolf, 2013 Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Molecular ecology **22**: 620–34. 2

Völkers, M., D. Rohde, C. Goodman, and P. Most, 2010 S100A1: a regulator of striated muscle sarcoplasmic reticulum Ca2+ handling, sarcomeric, and mitochondrial function. Journal of biomedicine & biotechnology **2010**: 178614. 15

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

Wang, S., T. Furmanek, H. Kryvi, C. Krossø y, G. K. Totland, *et al.*, 2014 Transcriptome sequencing of Atlantic salmon (Salmo salar L.) notochord prior to development of the vertebrae provides clues to regulation of positional fate, chordoblast lineage and mineralisation. BMC genomics **15**: 141. 3

Wilson, A. J., D. Gíslason, S. Skúlason, S. S. Snorrason, C. E. Adams, *et al.*, 2004 Population genetic structure of Arctic charr, Salvelinus alpinus from northwest Europe on large and small spatial scales. Molecular ecology **13**: 1129–42. 4

Wittkopp, P. J., 2011 Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of cis- and trans-regulatory differences. Methods in molecular biology (Clifton, N.J.) **772**: 297–317. 2

Woods, P. J., S. Skúlason, S. S. Snorrason, B. K. Kristjánsson, H. J. Malmquist, *et al.*, 2012 Intraspecific diversity in Arctic charr, Salvelinus alpinus, in Iceland: II. Which environmental factors influence resource polymorphism in lakes? Evolutionary Ecology Research **14**: 993–1013. 3

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

Zou, C., M. D. Lehti-Shiu, M. Thomashow, and S.-H. Shiu, 2009 Evolution of stress-regulated gene expression in duplicate genes of Arabidopsis thaliana. PLoS genetics **5**: e1000581. 14