Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs

Jóhannes Guðbrandsson[1,2], Sigríður Rut Franzdóttir[1,3], Bjarni Kristófer Kristjánsson[4], Ehsan Pashay Ahi[1,5], Valerie Helene Maier[1,3], Kalina Hristova Kapralova[1], Sigurður Sveinn Snorrason[1], Zophonías Oddur Jónsson[1,3] and Arnar Pálsson[1,3]

[1] Institute of Life and Environmental Sciences, University of Iceland, Reykjavík, Iceland
[2] Freshwater Division, Marine and Freshwater Research Institute, Reykjavík, Iceland
[3] Biomedical Center, University of Iceland, Reykjavík, Iceland
[4] Hólar University College, Sauðárkrókur, Iceland
[5] Karl-Franzens-Universität, Graz, Austria

Corresponding authors
Jóhannes Guðbrandsson, jog7@hi.is
Arnar Pálsson, apalsson@hi.is

ABSTRACT

Phenotypic differences between closely related taxa or populations can arise through genetic variation or be environmentally induced, leading to altered transcription of genes during development. Comparative developmental studies of closely related species or variable populations within species can help to elucidate the molecular mechanisms related to evolutionary divergence and speciation. Studies of Arctic charr (*Salvelinus alpinus*) and related salmonids have revealed considerable phenotypic variation among populations and in Arctic charr many cases of extensive variation within lakes (resource polymorphism) have been recorded. One example is the four Arctic charr morphs in the ∼10,000 year old Lake Thingvallavatn, which differ in numerous morphological and life history traits. We set out to investigate the molecular and developmental roots of this polymorphism by studying gene expression in embryos of three of the morphs reared in a common garden set-up. We performed RNA-sequencing, *de-novo* transcriptome assembly and compared gene expression among morphs during an important timeframe in early development, i.e., preceding the formation of key trophic structures. Expectedly, developmental time was the predominant explanatory variable. As the data were affected by some form of RNA-degradation even though all samples passed quality control testing, an estimate of 3′-bias was the second most common explanatory variable. Importantly, morph, both as an independent variable and as interaction with developmental time, affected the expression of numerous transcripts. Transcripts with morph effect, separated the three morphs at the expression level, with the two benthic morphs being more similar. However, Gene Ontology analyses did not reveal clear functional enrichment of transcripts between groups. Verification via qPCR confirmed differential expression of several genes between the morphs, including regulatory genes such as *AT-Rich Interaction Domain 4A (arid4a)* and *translin (tsn)*. The data are consistent with a scenario where genetic divergence has contributed to differential expression of multiple genes and systems during early development of these sympatric Arctic charr morphs.

INTRODUCTION

Phenotypic diversity provides the raw material for evolution and is influenced by variation in gene expression during development and the lifespan of individuals. Variation in gene expression is both influenced by genetics (*Jin et al., 2001*; *Oleksiak, Churchill & Crawford, 2002*) and environmental factors (*Giger et al., 2006*; *Danzmann et al., 2016*). Gene expression can change because of neutral evolution, as well as positive and purifying selection (*Romero, Ruvinsky & Gilad, 2012*). In the context of development the combined effects of purifying or stabilizing selection on existing traits and genetic drift, may lead to developmental system drift (*True & Haag, 2001*), that is alterations in gene expression and the functions of developmental circuits. Analyses of gene expression in developing organisms can reveal variation in the developmental circuits and the phenotypes they influence (*Garfield et al., 2013*) and alterations in the parameters of these networks (*Ludwig et al., 2005*). Evolutionary developmental biology seeks answers to questions like which developmental and cellular systems influence variation in adaptive traits and are some developmental processes, time points or tissues more prone/amenable to natural selection than others (*Kopp, Duncan & Carroll, 2000*; *Carroll, 2008*; *Stern & Orgogozo, 2008*)?

To address questions about the interplay of natural selection, developmental biology and drift in evolutionary divergence, we can study the developmental and molecular basis of natural diversity in recently diverged species or diverging populations within species. For example, studies of the Galapagos finches (*Geospiza* spp.) revealed that expression of *bone morphogenetic protein 4* and *calmodulin* during beak development has strong effects on beak depth and width (*Abzhanov et al., 2004*; *Abzhanov et al., 2006*), which are important characteristics for fitness (*Grant, 1999*; *Grant & Grant, 2008*). At the population level it was found that differential expression of the *Agouti* gene in hair follicles in deer mice (*Peromyscus* spp.) correlated with differences in coat color which varies among populations (*Linnen et al., 2009*). Here we set out to study gene expression during early development, in recently diverged populations with profound phenotypic separation, with the broad aim to understand molecular mechanisms related to phenotypic variation and adaptation.

Polymorphic and sympatric Arctic charr *Salvelinus alpinus* as a model to study evolution

After the last glaciation (∼12,000 years ago) salmonid species and threespined sticklebacks (*Gasterosteus aculeatus*) were prominent among fish species that colonized newly formed lakes and rivers of the northern hemisphere (*Wootton, 1984*; *Noakes, 2008*; *Klemetsen, 2010*).

Several fish species of northern freshwaters have diverged locally to form polymorphic systems, usually related to utilization of different resources (resource polymorphism, *Skúlason & Smith, 1995*; *Smith & Skúlason, 1996*; see additional refs. in *Snorrason & Skúlason, 2004*). This is seen in many salmonids (*Robinson & Parsons, 2002*; *Muir et al., 2016*)

and in Arctic charr many cases of phenotypically distinct sympatric morphs have been reported in post glacial lakes, for instance in Norway, Scotland and Iceland (*Telnes & Sægrov, 2004*; *Adams et al., 2007*; *Klemetsen, 2010*). In Iceland, Arctic charr is found as anadromous or non-anadromous resident populations in rivers and lakes. Many of the resident populations have become landlocked. The anadromous charr usually grow large and have pointed snouts with a terminal mouth resembling limnetic morphology. Many landlocked populations differ in feeding morphology, some feed on zooplankton or fish (limnetic morphs) while others utilize benthic prey (benthic morphs, *Skúlason et al., 1992*), as is common in northern polymorphic freshwater fish species (*Bernatchez et al., 2010*). Although somewhat variable in morphology, benthic charr are distinct from limnetic charr, with typically darker body, blunt snout and sub-terminal mouth. In Iceland they are most commonly found as dwarf morphs (adult length less than 15 cm) in isolated spring habitats in the neo-volcanic zone (*Kristjánsson et al., 2012*). Population genetics suggest that these benthic dwarfs have evolved repeatedly in groundwater springs across the island (*Kapralova et al., 2011*). Larger benthic forms do exist, with similar phenotypic characters as the dwarfs but larger adult size (*Skúlason et al., 1992*; *Kristjánsson et al., 2011*).

Sympatric Arctic charr morphs, found in several lakes, most often separate into benthic or limnetic morphotypes varying in many traits (morphology, behavior, color, life history characteristics, habitat use) (*Snorrason & Skúlason, 2004*). A well studied example of polymorphic Arctic charr are the four charr morphs of Lake Thingvallavatn (Fig. 1A). They differ distinctly in various traits, e.g., adult size, age at maturity, head and body morphology, coloration, behavior and habitat use (*Sandlund et al., 1992*). In the lake there are two limnetic morphs, the smaller planktivorous morph (PL, 15–25 cm adult length) that feeds on zooplankton, and the larger piscivorous morph (PI, 25–60 cm adult length) that mainly feeds on threespined stickleback (*Snorrason et al., 1989*; *Malmquist et al., 1992*). The lake harbors two benthic morphs, small benthic charr (SB, 12–20 cm adult length) and large benthic charr (LB, 25–60 cm adult length) both feeding on bottom-dwelling invertebrates in the lava substrate habitat along the shores (*Sandlund et al., 1992*). Rearing experiments showed that morphological and behavioral differences among the morphs arise early in development (*Skúlason et al., 1993*; *Skúlason et al., 1996*), and subsequent studies of developing embryos and juveniles showed significant differences in cartilage and bone formation (*Eiriksson, Skulason & Snorrason, 1999*; *Eiriksson, 1999*). Recently *Ahi et al. (2014)* used geometric morphometrics to capture variation in craniofacial structures among progeny of three of the morphs (PL-, LB- and SB-charr) soon after hatching (280−285 τs, see Materials and Methods for explanation of relative age measured in τs). For the ventral shape of the lower jaw and hyoid arch, distinct differences between the morphs were found at 305 τs, (*Ahi et al., 2014*). Experiments corroborate the contribution of genetic differences, but also demonstrated significant plastic potential of these morphs. The phenotypic plasticity of Arctic charr, and related salmonids is well documented (*Nordeng, 1983*; *Hindar & Jonsson, 1993*; *Skúlason, Snorrasson & Jónsson, 1999*). Studies on developing charr have revealed plastic responses to environmental factors like temperature, water velocity and food type (*Adams & Huntingford, 2004*; *Grünbaum et al., 2007*; *Jonsson & Jonsson, 2014*). Studies of limnetic and benthic charr morphs in Iceland show food type
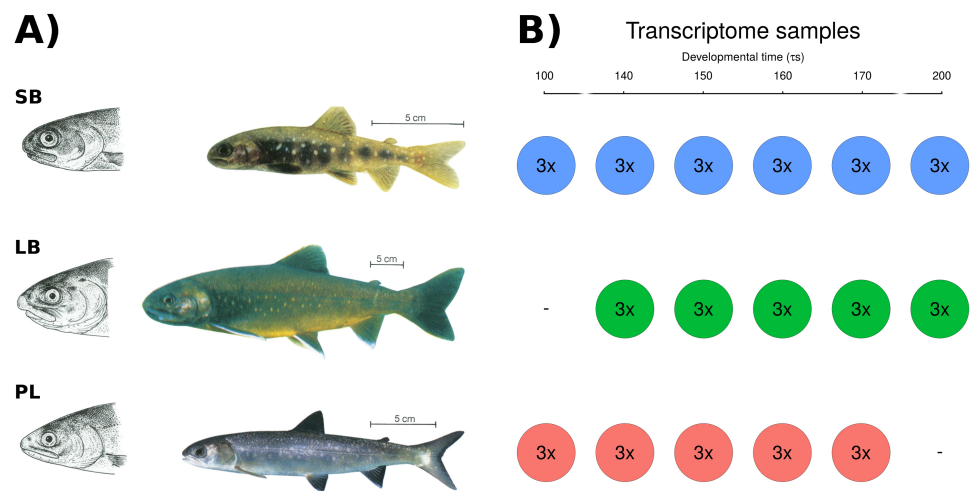
Figure 1 The phenotypically distinct sympatric Arctic charr and the experimental set-up. (A) Four sympatric morphs inhabit Lake Thingvallavatn, three of which are studied and pictured here: small benthic (SB), large benthic (LB) and planktivorous (PL) charr. They differ in size (size bars = 5 cm), the shape of the head and jaws (see drawings) and pigmentation. Adapted from *Sandlund et al. (1992)*, ©Wiley-Blackwell, drawings by Eggert Pétursson. (B) Embryos from pure crosses of the three morphs were sampled at six developmental timepoints prior to hatching, from 100 τs to 200 τs (circles) for RNA sequencing. During this period of development somatogenesis is complete and gill arches, jaws and many other structures are forming (Fig S1). Three biological replicate samples (3×) were taken for each morph and developmental timepoint, each sample being a pool of mRNA from three embryos. Six timepoints were sampled in SB-charr, and five in LB- and PL-charr. In total there were 48 samples, composed of 144 individual charr embryos. The coloring scheme indicating morphs (blue: SB, green: LB, red: PL) will be retained throughout the manuscript.

can affect growth and the shape of the feeding apparatus in early feeding juveniles (*Parsons, Skúlason & Ferguson, 2010*; *Parsons et al., 2011*; *Küttner et al., 2013*). Furthermore, egg volume, which varies considerably within and among females, is positively correlated to yolk depletion rate and fork length at hatching and at first feeding in aquaculture charr (*Leblanc, Kristjánsson & Skúlason, 2016*). Here we study gene expression during the early development of sympatric morphs, reared in a common garden that reduces the influence of environmental variations. Note however, the experimental design can not distinguish between genetic and parental effects on embryonic gene expression.

Genetic variation in polymorphic and sympatric Arctic charr

The earliest population genetic studies found little genetic separation of the sympatric morphs in Lake Thingvallavatn (*Magnusson & Ferguson, 1987*; *Danzmann et al., 1991*; *Volpe & Ferguson, 1996*). The first microsatellite study detected subtle differences (*Gíslason, 1998*) and the second study with 10 markers estimated overall F_{ST}'s = 0.039, between the LB-, SB- and PL-charr (*Kapralova et al., 2011*). More recently, we detected F_{ST}'s larger than 0.25 between morphs for variants in two immunological genes (*Kapralova et al., 2013*) and a few other loci (*Guðbrandsson et al., 2016*), suggesting substantial genetic separation at specific loci among those sympatric charr morphs. There is a need to study underlying developmental mechanisms, e.g., how differential expression or function of genes promotes
differences in charr development and phenotypes. To date, few studies have addressed these issues. The candidate gene approach illustrates how embryonic morphogenetic mechanisms may influence phenotypic diversity and speciation (*Abzhanov et al., 2004*; *Abouheif et al., 2014*). A qPCR study on muscle tissues in charr, showed that expression of three genes in the *mTOR*-pathway distinguishes five small benthic morphs from two limnetic morphs in Iceland (*Macqueen et al., 2011*). On the other hand the myogenic *paired box protein 7 (Pax7)* gene was not differentially expressed during development in Lake Thingvallavatn morphs (*Sibthorpe et al., 2006*).

Genome wide methods are the new norm, for example population genomics (*Pease et al., 2016*) or transcriptome screens (*Perry et al., 2012*) to investigate patterns of divergence or loci of adaptation. In this context it is worth stressing that salmonids, due to the fourth whole genome duplication of the linage (Ss4R) 88–103 million years ago (*Moghadam, Ferguson & Danzmann, 2011*; *Macqueen & Johnston, 2014*; *Berthelot et al., 2014*; *Lien et al., 2016*), have quite complex genomes. The extra paralogs and chromosomal changes (*Macqueen & Johnston, 2014*; *Nugent et al., 2017*) complicate genome and transcriptome assemblies and analyses (*Norman, Ferguson & Danzmann, 2014*; *Lien et al., 2016*). To date the genome of two salmonids, Atlantic salmon (*Salmo salar*, hereafter salmon) (*Lien et al., 2016*) and rainbow trout (*Oncorhynchus mykiss*) (*Berthelot et al., 2014*), have been sequenced and annotated, but comparable resources are not available for Arctic charr.

We are interested in elucidating the developmental and molecular basis of trophic diversity in Arctic charr. Previously we deployed high throughput sequencing on embryos of SB-charr from Lake Thingvallavatn and an Icelandic aquaculture-charr breeding strain, to identify expression differences in microRNA and protein coding genes (*Kapralova et al., 2014*; *Guðbrandsson et al., 2016*). The miRNA sequencing revealed differential expression in 72 microRNAs, including some related to development of the brain and sensory epithelia, skeletogenesis and myogenesis (*Kapralova et al., 2014*). Similarly, the mRNA transcriptome (*Guðbrandsson et al., 2016*) indicated differences in the function of several pathways and genes, including metabolic, structural and regulatory genes. In that study we hypothesized that the observed expression divergence in mitochondrial functions (*Guðbrandsson et al., 2016*) reflected either strong artificial selection for growth rate in aquaculture-charr or altered life history of SB-charr in Lake Thingvallavatn by selection for early maturation with the trade-off in energy allocation highly favouring the production of gonads rather than body growth (*Jonsson et al., 1988*). Based on the transcriptome data from *Guðbrandsson et al. (2016)* and known craniofacial expression in other species we chose candidate genes to analyze gene expression with qPCR in limnetic and benthic morphs. Briefly, the data showed that a number of genes with conserved co-expression, most of which are involved in extracellular matrix organization and skeletogenesis (and *ETS proto-oncogene 2, transcription factor, Ets2*), differed in expression between benthic and limnetic morphs (*Ahi et al., 2013*; *Ahi et al., 2014*). Furthermore, employing the candidate gene approach on preliminary analysis of the data presented here, linked the Aryl-hydrocarbon receptor pathway to benthic-limnetic divergence in charr (*Ahi et al., 2015*).

Here we study the early developmental transcriptome of three of the four sympatric morphs from Lake Thingvallavatn (LB-, SB- and PL-charr) with the aim of identifying

genes and molecular systems that have featured in the divergence of the Thingvallavatn morphs. The expression divergence can also shed light on the evolutionary relationship of the three morphs under study. Our previous developmental RNA-sequencing study of Arctic charr (*Guðbrandsson et al., 2016*) provided a useful start off for analysing gene expression and developmental pathways associated with the benthic vs. limnetic differences (*Ahi et al., 2014*; *Ahi et al., 2015*). The study described here differs from *Guðbrandsson et al. (2016)* in several aspects: (i) it focuses on an earlier window of development in higher temporal resolution (six time points at 100–200 τs vs four from 141–433 τs). This window of development precedes and covers the formation of key craniofacial structures, e.g., those required for feeding apparatus functions (gill arches and elements of the jaws) leading up to 200 τs when most of the viscerocranium is in place (Fig S1) (*Kapralova et al., 2015*). The developmental pathways related to these structures lay the ground well before they become visible. (ii) The present study compares expression in three Thingvallavatn morphs whereas in the previous study the comparison was between Thingvallavatn SB-charr and an aquaculture stock of mixed origin, which has a typical limnetic-like head morphology but has been subjected to strong artificial selection for growth. (iii) Because of the high coverage and length of the reads in the current study (101 bp, paired-end) we were able to perform *de-novo* transcriptome assembly, which was not possible with the short (36 bp) reads of the previous study. We were therefore able to map reads onto a charr transcriptome instead of making use of *S. salar* EST's.

Based on the documented differences in jaw morphology soon after hatching (*Ahi et al., 2014*), we anticipated substantial expression differences in systems related to growth and development of craniofacial structures. However as RNA was isolated from whole embryos, we also expect differences in genes related to physiological systems and development of other body parts. As expected, the data reveal substantial changes in gene expression during early development and importantly also morph specific expression differences in a large number of transcripts. In sum, multiple genes in many pathways were found to be differentially expressed in early development of these recently evolved sympatric charr morphs. The data set the stage for detection of genetic and environmental underpinnings of the observed phenotypic and developmental differences between the morphs.

MATERIALS AND METHODS

Sampling, rearing and developmental series

Embryos from crosses of wild caught fish were reared in a common garden environment (see below) at Hólar University College aquaculture facility in Verið (Sauðárkrókur, Iceland) as in previous studies (*Ahi et al., 2013*; *Guðbrandsson et al., 2016*). Embryos from three morphs from Lake Thingvallavatn were studied (Fig. 1).

Parents were fished in Lake Thingvallvatn with the permissions 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 55/2013).

Embryos were reared at ∼5 °C with constant water flow and in complete darkness. As the morphs spawn at different times, slight fluctuations in water temperature could

not be avoided. Water temperature was recorded twice daily and the average was used to estimate the relative age (*RA*) of the embryos using τ-somite units (τs) (*Gorodilov, 1996*). The following formula was used to calculate the relative age (*RA*) at days post fertilization (*n*) using the average daily temperature (t_i).

$$RA_n = \sum_{i=1}^{n} 1440 \cdot (1/10^{3.0984 - 0.0967 t_i + 0.00207 t_i^2})$$

Sampling of embryos for RNA extraction was performed by Holar University College aquaculture Research Station (HUC-ARC) personnel. Embryos were sampled at designated timepoints, placed in RNAlater (Ambion), stored at +4 °C overnight and frozen at −20 °C. 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.

Embryos from pure multi-parent crosses of the three morphs were sampled at six developmental timepoints prior to hatching (see below and Fig S1), from 100 τs to 200 τs for RNA sequencing (circles in Fig. 1). Three biological replicate samples (3×) were taken for each morph and developmental timepoint, each sample containing three embryos, where each embryo came from the same cross. Six timepoints were sampled in the SB-charr, and five in the LB- and PL-charr. Hence in total 48 samples were sequenced, composed of 144 individual charr embryos.

Most of the samples came from offspring of crosses created in the 2010 spawning season (SB 150–200 τs, PL 140–170 τs, LB 140–200 τs). For SB- and PL-charr, eggs from 10 females were pooled and fertilized with milt from 10 males from the same morph. For LB-charr the same setup was used except that five females and five males were used. Because of laboratory failure (samples destroyed), we had to replace three morph and timepoint combinations. For 100 τs in PL-charr, we used samples from the 2011 spawning season (generated with the identical crossing setup). Similarly, SB-charr samples from timepoints 100 and 140 τs were replaced with material from two single parent crosses generated 2011. Samples SB100A and SB100B came from the one cross but sample SB100C and all samples for timepoints 140 τs were from the second cross. The samples from 2011 did not show aberration from other samples in principal component analyses (PCA) of the expression data (Fig S2). For qPCR two timepoints (150 τs and 170 τs) were sampled for all three morphs with the same setup, all from crosses made in 2010.

Staining of embryos for developmental series

Samples of LB-charr embryos from all timepoints were fixed in 4% PFA. Samples from 140–200 τs were stained for cartilage (alcian blue) and bone (alizarin red) using a modified protocol from *Walker & Kimmel (2007)*. All samples were stained simultaneously. Stained individuals were placed in a petri dish containing 50 ml of 1% agarose gel and immobilized with insect needles to ensure the correct positioning of the embryo. The head of each individual was photographed ventrally using a Leica (MZ10) stereomicroscope. Between 140 τs and 200 τs major craniofacial elements appear as clear units of cartilage for example at 150 τs the formation in the ventral aspect of the two trabeculae, the Meckel's cartilages and palatoquatrates can be observed, shortly followed by the emergence of major elements

of the hyoid and branchial arches (160–170 τs) (Fig. S1B). The minor elements (the hypo- and basi-branchials) of these arches start to appear later (200 τs) (Fig. S1B). The ethmoid plate starts forming around 180 τs and is almost fully fused centrally at 200 τs. Rudiments of the maxillae can be seen as early as 200 τs.

RNA extraction and transcriptome sequencing

For RNA extraction embryos were dechorionated and homogenized with a disposable Pellet Pestle Cordless Motor tissue grinder (Kimble Kontes, Vineland, NJ, USA) and RNA was extracted using TRIzol reagent (Thermo Fisher, Waltham, MA, USA) according to the manufacturers instructions. RNA quantity was examined using a NanoDrop ND1000 (Labtech, East Sussex, UK) spectrophotometer. An Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) was used to assess RNA quality and samples with high RNA integrity number (RIN, an estimate of RNA quality, *Schroeder et al., 2006*) were selected. Only four samples had RIN below 9 (Table S1). Sequencing libraries were prepared using the TruSeq RNA Sample Preparation kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol (Release 15008136, November 2010). mRNA was purified on oligo-(dT) attached magnetic beads, eluted and fragmented at 94 °C for 2 min, to generate fragments of c.a. 130–290 bases. First strand cDNA synthesis was performed using random hexamer primers, followed by RNase treatment and second strand synthesis. The cDNA ends were repaired and adenylated before the ligation of indexing adapters. The libraries were PCR amplified (15 cycles). Samples were quantified with NanoDrop and quality estimated with BioAnalyzer before they were pooled and sequenced on Hiseq 2000 at deCODE genetics (Reykjavik, Iceland), yielding 101 bp paired-end reads. The raw reads were deposited into the NCBI SRA archive under BioProject identifier PRJNA391695 and with accession numbers: SRS2316381 to SRS2316428.

Assembly, abundance estimation and annotation

The sequencing reads were quality trimmed and adapters removed using Trim Galore (version 0.3.3, *Krueger, 2012*) before assembly. Bases with Phred-quality below 10 were trimmed off. Reads that were less than 20 bp after trimming were removed and the mate of the read was also removed from downstream analysis. The quality filtered reads from all samples were assembled using Trinity (version v2.1.0, *Grabherr et al., 2011*) with the default parameters, except the ''min_kmer_cov'' was set to two to reduce memory use. Preliminary analysis using salmon EST contigs (*Di Génova et al., 2011*) as reference indicated extensive RNA degradation and subsequent 3′ bias in all samples for one timepoint (160 τs) in two (LB and PL) out of the three morphs. This timepoint was thus excluded from gene expression analyses as 3′ bias can have drastic effects on expression estimations (*Sigurgeirsson, Emanuelsson & Lundeberg, 2014*). RNA degradation also affected other samples, see below. We used Kallisto (version v0.42.4, *Bray et al., 2016*) to estimate the abundance of transcripts. Kallisto was run with default parameters and 30 rounds of bootstrapping. Only transcripts with more than 200 estimated reads total in the samples, were retained for annotation and expression analysis.

The transcripts were annotated using the Trinotate pipeline (version 2.0.2, *Haas, 2015*). Trinotate runs the assembled contigs through a few programs for detecting coding

sequences, protein structures and rRNA genes as well as running `blast` on SwissProt and TrEMBL databases for ortholog detection (see http://trinotate.github.io/). Trinotate was run with the default parameters except that we set the E-value cutoff for `blast` searches to 10^{-20}. If two or more open reading frames (ORFs) were predicted for a transcript we excluded ORFs that did not `blast` to the trEMBL database. If ORFs from the same transcript overlapped we excluded the one with higher E-value.

Orthologs of the transcripts in salmon and rainbow trout mRNA and protein sequences were found using `blastn` and `blastx` respectively. The annotations for the rainbow trout genome were obtained from *Berthelot et al. (2014)*, (http://www.genoscope.cns.fr/trout/data/, version from 2014-05-19). The annotation for the salmon genome came from two different sources; NCBI *Salmo salar* Annotation Release 100 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Salmo_salar/100/, retrieved 2015-12-17) and SalmoBase (*Samy et al., 2017*; http://salmobase.org/), version from 2015-09-18. For each reference dataset we only retained the best match for each transcript. We set the E-value cutoff for `blastn` searches to 10^{-50}, minimum percent identity to 85% and the transcript was required to cover at least 50% of the reference transcript. For `blastx` searches we set the E-value cutoff to 10^{-20}, minimum percent identity to 75% and mandated that the transcript should cover at least 20% of the reference protein. Scripts from the Trinity suite (*Grabherr et al., 2011*) were used to group discontinuous alignments and calculate the alignment coverage of reference transcripts.

Estimation of RNA degradation and 3′-bias

To estimate read coverage across the length of transcripts we supplied pseudobam files from Kallisto to eXpress (version 1.5.1; *Roberts & Pachter, 2012*). eXpress uses an expectation-maximization (EM) algorithm for read placement based on sequence composition and transcript expression. We used the default parameters except for the 'batch' option which was set to 10 to get more EM-rounds. To estimate 3′-bias we chose 381 long transcripts with high read coverage and which spanned almost full length genes. In more detail, the transcripts were chosen if at least 90% of their sequence aligned to over 90% of a salmon transcript (based on SalmoBase annotation). We restricted the analysis to transcripts between 2,000 and 6,000 bp in length, with high read coverage and little variation between samples. The coverage was estimated in 100 bins over the length of each transcript. The 3′-bias was estimated as a percentage of coverage for the 3′half of each transcript compared to the total transcript, and the average for all of 381 transcript calculated for each of the 48 samples. The calculations were performed in the R environment (*R Core Team, 2017*). This quantity will be referred to as 3′ coverage hereafter and used as an estimate of 3′-bias for each sample.

Estimating expression differences among morphs

Kallisto (*Bray et al., 2016*) was used to estimate transcripts abundance per sample. Transcripts with at least 200 mapped reads were subjected to expression analysis, using the R-package Sleuth (*Pimentel et al., 2017*) to fit linear models. The full model (FM) included morph (M) and developmental time (T) and the interaction of morph and developmental

time ($M \times T$). We also fitted three reduced models excluding different factors of the full model to test for influences of that factor. In addition we took the 3′ coverage (described above, z in formulas below) into account. We fitted the 3′ coverage as a second degree polynomial to allow the effect on expression to be non-linear while keeping the model as parsimonious as possible. We compared the full model to model R1 to test for the interaction term or morph effect within time-points. We compared R1 to R2 to test for overall morph effect and finally we compared R1 to R3 to check for influences of developmental time on gene expression. The models were compared with a likelihood ratio test to check for significance of variables.

$$y_{ijk} = M_i + T_j + (M \times T)_{ij} + \beta_1 z_k + \beta_2 z_k^2 \tag{FM}$$

$$y_{ijk} = M_i + T_j + \beta_1 z_k + \beta_2 z_k^2 \tag{R1}$$

$$y_{ijk} = T_j + \beta_1 z_k + \beta_2 z_k^2 \tag{R2}$$

$$y_{ijk} = M_i + \beta_1 z_k + \beta_2 z_k^2. \tag{R3}$$

To gauge the effect of including 3′ coverage as an explanatory variable, we also ran models excluding 3′ coverage. We tested if 3′-bias had an effect on expression (model FM vs R4). We also tested for interaction, morph and time effect without taking 3′-bias into account (R4 vs R5, R5 vs R6 and R5 vs R7).

$$y_{ijk} = M_i + T_j + (M \times T)_{ij} \tag{R4}$$

$$y_{ijk} = M_i + T_j \tag{R5}$$

$$y_{ijk} = T_j \tag{R6}$$

$$y_{ijk} = M_i. \tag{R7}$$

Sleuth uses false discovery rate (fdr) to adjust for multiple testing (*Benjamini & Hochberg, 1995*). Transcripts with significant morph/time interaction or morph effect ($fdr < 0.01$) were classified into 16 clusters using the Mfuzz-package (*Futschik, 2015*). For clustering we used log-transformed estimates of transcripts per million (tpm) normalized by 3′-bias, with the fuzzification parmeter (m) set to 1.1. To visualize the differences between morphs we performed principle component analysis (PCA) in R on the expression estimates, only for transcripts in clusters with morph effects and time-invariant expression differences between morphs (clusters A–E).

The goseq-package in R (*Young et al., 2010*) was used to test for enrichment of Gene Ontology (GO) categories of biological processes within each cluster. The annotation from SalmoBase was used and transcripts were also mapped to all the ancestors of annotated GO categories using the GO.db-package in R (version 3.2.2; *Carlson, 2015*). To get an overall signal and increase statistical power, rather than trying to get a specific signal from incompletely annotated data, we decided to focus on GO-categories at specific positions in the GO-category relationship tree. For enrichment tests we used only categories with the longest path to the root of the GO-tree at least three steps and the shortest path to root no longer than four steps. Note that different paths from a specific category to root can be of different lengths. For each cluster we ran two enrichment tests. First on the

transcript level where length bias was taken into account (*Young et al., 2010*). Second we ran enrichment test for salmon genes (based on SalmoBase annotation). A gene was considered to belong to a cluster if a transcript annotated to it belonged to the cluster. For the gene GO-enrichment tests we used a Hypergeometric test without any length correction. A GO-category was only considered significant if significance (*fdr* < 0.01) was found on both transcript and gene level. The gene level was also used to correct for genes with multiple isoforms or incomplete assemblies, which can lead to false positive categories. We clustered significant GO-categories for each cluster using semantic similarity between categories in the zebrafish genome according to the GOSemSim-package in R (*Yu et al., 2010*) as a distance measurement. The distance matrix for GO-categories was supplied to the hclust function in R and a cutoff of 0.8 was used to categories the GO-categories in to super categories.

qPCR verification of gene expression

Candidate genes for verification by qPCR were picked based on differential expression between morphs in the transcriptome and in some cases prior data on biological functions. Reference genes to study Arctic charr development have previously been identified (*Ahi et al., 2013*). Primer3 (*Untergasser et al., 2012*) was used to design primers (Table S4) and the primers were checked for self-annealing and heterodimers in line with MIQE guidelines (*Bustin et al., 2009*). Primers for genes with several paralogs were designed for regions conserved among paralogs. RNA extraction followed the same steps as for samples used in the transcriptome. cDNA synthesis followed the same steps as in *Ahi et al. (2015)*: DNA contamination was removed using DNases treatment (New England Biolabs, Ipswich, MA, USA) and cDNA was synthesized with 1 μg of RNA using the High Capacity dDNA RT kit (Applied Biosystems, Foster City, CA, USA) in 20 μl reaction volume.

Real-time PCR was performed in 96 well-PCR plates on an ABI 7500 real-time PCR System (Applied Biosystems, Foster City, CA, USA). The normalized relative expression of genes in whole embryos was estimated from the geometric mean expression of two reference genes, *β-actin (actb)* and *ubiquitin-conjugating enzyme E2 L3 (ub2l3)*. To visualize differences among morphs and time, the normalized expression was presented as relative to the expression of one of three samples in PL at 150 τs (calibration sample). Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method of *Livak & Schmittgen (2001)*. Statistical analysis was performed using the ΔC_T-values with a two-way ANOVA with GLM function in R.

$$y_{ijk} = M_i + T_j + (M \times T)_{ij} + \varepsilon_{ijk}.$$

The residuals were normally distributed for all data. Genes with significant morph effect was followed up on by performing Tukey's post-hoc test, on relative expression ratios (ΔC_Ts).

RESULTS AND DISCUSSION

Transcriptome sequencing, assembly and annotation

The number of sequenced paired-end reads varied among the 48 samples, from 4.5 to 86.9 million. No bias in read number among lanes, indexes, morphs or developmental

Table 1 Summary statistics for the transcriptome assembly, from the raw Trinity output and filtering out transcripts with less than 200 reads mapped. Lengths (in basepairs) of all transcripts and the longest transcript (isoform) for each gene are tabulated.

	Raw	Raw long iso[a]	Filtered	Filtered long iso[a]
Total Trinity 'genes'	449,681		78,667	
Total Trinity transcripts	581,474		129,388	
GC-content (%)	45.93		47.41	
N10	4,818	3,830	5,858	5,457
N20	3,527	2,417	4,598	4,132
N30	2,685	1,551	3,822	3,317
N40	2,015	1,031	3,218	2,707
N50	1,441	718	2,709	2,197
Median contig length	364	328	1,270	851
Average contig length	757.94	559.54	1,737.01	1,338.73
Total assembled bases	440,720,391	251,613,073	224,748,860	105,235,409

Notes.

[a] Longest isoform for each Trinity gene.

timepoints was detected, except that timepoint 160 τs in LB-charr had low coverage for all three replicates (Table S1). Trinity (*Grabherr et al., 2011*) *de-novo* assembly yielded 581,474 transcripts which grouped into 449,681 "genes". After filtering on coverage (minimum of 200 reads aligned) the numbers of transcripts and "genes" decreased to 129,388 and 78,667 respectively. All estimators of length increased with this filtering step, e.g., the N10–N50 statistics (Table 1).

Blastn revealed that the majority of the transcripts had homology with sequences in Atlantic salmon (72% for the NCBI database and 83% for SalmoBase) and rainbow trout (53%). Similar analyses at the protein level (blastx or blastp) found a lower proportion with homology, 43% to 55% depending on the database in the two salmonids and other organisms (Table 2). Even though Arctic charr is considered more closely related to rainbow trout than salmon (*Koop et al., 2008*; *Crête-Lafrenière, Weir & Bernatchez, 2012*; *Alexandrou et al., 2013*) a larger number of transcripts had significant blast hits to salmon. Most likely this reflects the more conservative approach used for annotation of the rainbow trout genome, e.g., requiring genes to have orthology in other vertebrates (*Berthelot et al., 2014*).

We searched reference databases with Arctic charr transcripts, using blastx and blastn, to estimate the number and length of the assembled genes and proteins (Table 3). Hits to 19,122–35,685 proteins were found (depending on database) but with more stringent filters on length (requiring more than 90% coverage) these numbers ranged from 9,367 to 18,593 (Table 3). Using BLAST to align against salmon transcripts (SalmoBase) recovered up to 48,916 hits in the databases (Table 3). Again, more transcripts show homology to salmon than to rainbow trout, which again likely reflects differences in the annotation strategies. We retrieved more hits for transcripts and proteins from the SalmoBase annotation than the Salmon NCBI annotation. The transcripts in the SalmoBase annotation are longer on average compared to the NCBI annotation, therefore our Arctic charr transcripts cover less of each SalmoBase transcripts although more hits are retrieved (Table 3). More than half

Table 2 The number and percent of Trinity transcripts and genes with significant blast hits in different databases, using different blast programs (blastn, blastx and blastp).

Database	Program	Transcripts	Genes	Transcripts (%)	Genes (%)
Ssal NCBI	blastn	93,239	49,281	72.06	62.65
Ssal SalmoBase	blastn	107,068	61,185	82.75	77.78
Omyk	blastn	68,476	33,505	52.92	42.59
Ssal NCBI	blastx	62,548	26,843	48.34	34.12
Ssal SalmoBase	blastx	63,310	27,652	48.93	35.15
Omyk	blastx	55,862	24,533	43.17	31.19
SwissProt	blastx	59,763	24,130	46.19	30.67
TrEMBL	blastx	71,156	30,927	54.99	39.31
SwissProt	blastp	57,702	22,737	44.60	28.90
TrEMBL	blastp	64,442	26,198	49.81	33.30
Total transcripts		129,388	78,667	100	100

of the genes covered 90–100% of the predicted protein length, with minimal difference depending on database, while less than half covered more than 90% of the predicted transcript length. This probably reflects the higher divergence between *S. alpinus* and its relatives in the untranslated regions of the transcripts.

To the best of our knowledge, only two other mRNA-sequencing studies have been conducted on Arctic charr (*Norman, Ferguson & Danzmann, 2014*; *Guðbrandsson et al., 2016*). Our previous study of SB-charr and Icelandic aquaculture charr did not involve transcriptome assembly (*Guðbrandsson et al., 2016*). However *Norman, Ferguson & Danzmann (2014)* assembled a transcriptome, in their investigation of salinity tolerance in the gills of Canadian aquaculture charr. Their assembly yielded 108,645 assembled contigs, with $N50 = 2,588$ and around 80% of contigs annotated (using both *S. salar* and *O. mykiss* databases). Our assembly yields fewer "genes" (78,667) after the quality filtering steps, but for downstream analyses we retain more than one transcript per gene. The N50 values of both datasets are similar, but *Norman, Ferguson & Danzmann (2014)* achieve slightly higher annotation percentage. Our current study provides new data on the transcriptome of Arctic charr from embryos in early development. Integration of these data with genomic sequence data, will be valuable to assemble the complete charr transcriptome and fuel studies of gene gains and losses among salmonid species and populations (*Robertson et al., 2017*).

RNA degradation and 3′-bias in the transcriptome

Preliminary expression analysis with reads mapped to salmon EST's that showed clear indication of 3′-bias at one timepoint (160 τs) led us to remove these samples from the dataset, and take a closer look at position bias. Uneven distribution of reads over transcripts can profoundly influence estimates of expression and subsequent analyses (*Wu, Wang & Zhang, 2011*). To explore and estimate this bias, we constructed an estimator of 3′ coverage bias and incorporated it into the linear models (see Materials and Methods). The 3′-bias per sample was estimated from 381 nearly full length transcripts in the 2,000–6,000 bp range that had high sequencing coverage in all samples. The patterns of read coverage over the transcripts varied greatly between samples (Fig. 2A). Many samples showed

Table 3 **Estimated number of protein coding genes in the *de-novo* assembly.** Arctic charr transcripts were compared to different protein databases (using blastx, upper table) and Salmonids mRNA databases (using blastn, lower table). The tables shows the cumulative number of proteins or transcripts covered in each database, ranked by degree of coverage.

	Proteins				
Percent covered	TrEMBL	SwissProt	Ssal NCBI	Ssal SalmoBase	Omyk
90–100	15,788	9,367	18,376	18,593	12,829
80–90	18,287	11,610	20,178	20,899	15,070
70–80	20,150	13,163	21,814	23,072	16,476
60–70	21,978	14,404	23,596	25,255	17,713
50–60	23,822	15,484	25,332	27,466	18,885
40–50	25,484	16,478	27,018	29,628	19,850
30–40	26,977	17,380	28,656	31,866	20,718
20–30	28,299	18,219	30,180	33,680	21,353
10–20	29,204	18,907	31,517	35,119	21,775
0–10	29,477	19,122	32,082	35,685	21,888
Total peptides			97,555	195,069	46,585

	mRNA Transcripts		
Percent covered	Ssal NCBI	Ssal SalmoBase	Omyk
90–100	12,418	4,189	10,335
80–90	15,671	6,805	13,644
70–80	18,812	9,516	16,182
60–70	22,051	12,820	18,461
50–60	25,748	16,854	20,656
40–50	29,387	21,594	22,113
30–40	33,044	27,691	23,426
20–30	36,822	35,080	24,621
10–20	40,202	43,626	25,671
0–10	41,284	48,916	25,915
Total transcripts	109,584	195,072	46,585

considerable 3′-bias, but more disappointingly the bias was confounded with a variable of chief interest (Morph). The 3′ coverage correlates with the RIN-values of the RNA isolates (Pearson $r = -0.83$, $p = 6.75e - 13$) but samples with higher 3′ coverage than expected are apparent (e.g., PL160B and SB200A, Fig. 2B). This clearly demonstrates the importance of maintaining high and consistent RNA quality for RNA sequencing if poly-A pulldown is used and the importance of checking for 3′ bias in RNA-seq datasets.

Analyses of differential expression (see below) revealed that the estimated 3′-bias was the second most important factor after developmental time with 32,395 significant transcripts ($\alpha = 0.01$, Table 4). Crucially, the results differed considerably if the 3′-bias term was not included; then more transcripts had significant Morph by Time interaction effect ($M \times T$) and fewer significant developmental time effect (Time) (Table 4, Fig S3C and Fig S3D). Many transcripts with significant $M \times T$ interaction effect in a model without a 3′-bias term had significant Time effect after normalizing for 3′-bias (Fig S3A). Thus we concluded that
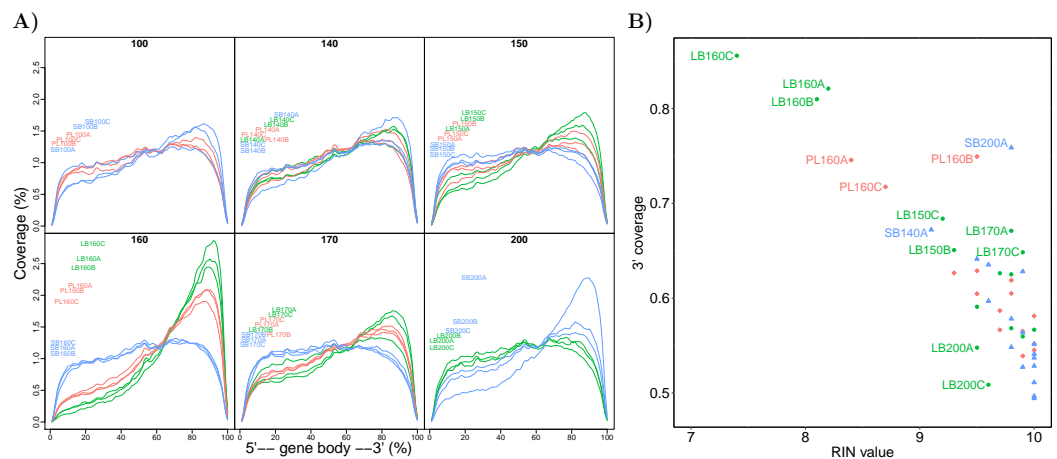
Figure 2 Confounding of 3′-bias with morphs and timepoints in the charr developmental transcriptome data. (A) Average coverage over the length of transcripts for each sample. The coverage was estimated from 381 transcripts that were highly expressed in all samples. The mean coverage for these genes was estimated in 100 windows over the gene body, from the 5′-end to the 3′-end. Samples from different timepoints are graphed separately and colored by morph: LB in green, PL in red and SB in blue. (B) Comparison of RIN-values and 3′ coverage (coverage at the 3′-half divided by total coverage) for each sample, colored by morph.

Table 4 Number of differentially expressed transcripts for each effect at different *fdr* cutoffs when taking 3′-bias into account (upper half) and when not taking 3′-bias into account (lower half).

		With 3′-bias correction		
fdr	3′-bias	$M \times T$	Morph	Time
< 0.05	46,274	14,293	3,381	60,491
< 0.01	32,395	8,407	2,002	42,879
< 0.001	20,834	3,977	1,075	28,039

	Without 3′-bias correction		
fdr	$M \times T$	Morph	Time
< 0.05	32,259	2,806	44,710
< 0.01	15,789	1,711	27,442
< 0.001	4,874	946	16,083

involving 3′-bias in the linear models decreased the number of transcripts with potentially false positive $M \times T$ interaction effect.

Degradation of RNA is an issue for RNA-sequencing. Particularly poly-A pull-down of degraded mRNA will lead to higher fraction of reads from 3′ end of transcripts (*Sigurgeirsson, Emanuelsson & Lundeberg, 2014*). Methods for estimating variation in coverage along transcripts, rely on full length sequences (*Wu, Wang & Zhang, 2011*). Correction for 3′ bias by restricting analyses to 200 bp at the 3′end of transcripts (*Sigurgeirsson, Emanuelsson & Lundeberg, 2014*) also requires full length sequences or reliable identification of 3′-ends. Neither of those methods were applicable to the current

data, as minority of transcripts are of full length, e.g., only 15,671 salmon transcripts in the NCBI database out of 41,284 with homology to Arctic charr are spanned to more than 80% by our contigs (Table 3). *Abernathy & Overturf (2016)* tested different methods for ribosomal-RNA removal on rainbow trout and concluded that Ribo-Zero (Illumina), which is based on hybridization, gave the best results and should therefore be the method of choice for future studies on incomplete transcriptomes. The use of the estimate of 3′-bias as a covariate reduced the number transcripts with, potentially false, Morph by Time effect. We do not claim that this approach accounts fully for transcript to transcript variation in 3′-bias, so we interpret the following differential expression results cautiously.

Differential transcript expression between sympatric Arctic charr morphs

While developmental time was the most commonly significant factor (42,879 transcripts, Table 4), we were most interested in expression divergence between the three charr morphs. Importantly the 3′-bias correction (above) had limited effect on the number of transcripts with significant overall Morph effect (Fig S3B). We conclude that more than one thousand genes are differentially expressed between developing embryos of the three sympatric morphs. Of the 2,002 transcripts with morph differences (at $fdr < 0.01$), 1,370 were only significant for Morph and no other terms. Further 632 had other terms also significant (some even all), but only 131 transcripts were significant for both Morph and Morph by Time ($M \times T$) interaction (Fig. 3). A considerably larger number of transcripts (8,407) had a significant $M \times T$ term, with the majority (4,684) also having significant Time and 3′-bias effects. As the 3′-coverage estimator is unlikely to control entirely for the 3′-bias, we suspect the number of transcripts with interaction of Morph by Time may be overestimated. To analyze the differences and changes in the transcripts with Morph and Morph by Time interaction we conducted clustering, yielding 16 co-expression clusters with 176 to 1,320 transcripts each (Fig. 4). Five clusters (A–E) had mostly transcripts with time-invariant Morph effects, but the remaining 11 clusters (F–P) had mainly transcripts with combinations of $M \times T$ and Time effects (Table 5). The data suggest separation between all three morphs at the expression level, for instance in cluster B (334 transcripts). Two of the five Morph effect clusters (C and D) show persistent expression difference between the two benthic (SB and LB) and PL-charr. These contain 797 transcript, while cluster A (with lower expression in SB-charr compared to the other two) has 353 transcripts and 499 were in cluster E (lower expression in LB-charr). To visualize this we performed PCA on the transcripts from these five clusters. This showed all three morphs separate at the transcriptional level (Fig. 5). Some separation of samples based on morph is expected as the genes used for the PCA were selected due to having a significant morph effect, however, importantly in this PCA all three morphs separated completely from each other. Furthermore, the PL-charr separate from the benthic morphs in PC1 (explaining 26.8% of the variance) and the two benthic morphs separate in PC2 (17.9% of the variance).

As transcriptional divergence and genetic divergence tend to be associated (*Whitehead & Crawford, 2006*), this suggests closer relation of the two benthic morphs, with PL-charr as a more distant relative, consistent with one population genetic study
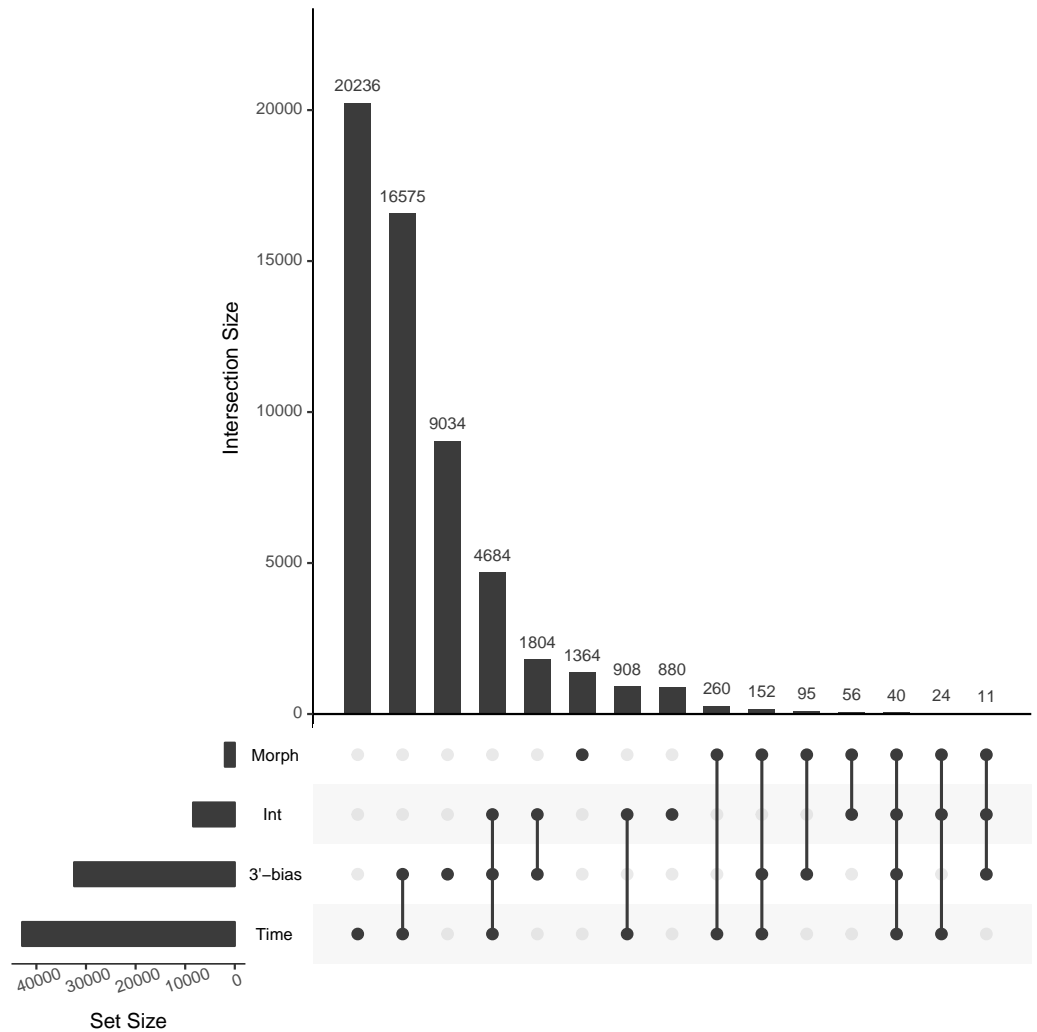
Figure 3 **The number of transcripts differentially expressed according to the linear models on developmental timepoint (Time), Morph, 3′-bias and interaction of Morph and Time (int).** The set size barplot (sideways) shows cumulated number of transcripts for each of the four main factors, while the intersection size barplot (vertical) shows the number of transcripts significant for each one or a combination of two or more factors. The dots indicate the significant factors or their combinations. For example 20,236 transcripts are only significant for Time effect but 16,575 are significant for both Time and 3′-bias effect.

(*Volpe & Ferguson, 1996*) but incongruent with others (*Gíslason, 1998*; *Kapralova et al., 2011*). Preliminary analyses of genetic variation in this transcriptome separates the morphs, and supports closer relation of the benthic morphs (J Guðbrandsson et al., 2018, unpublished data).

We next gauged the functions of the differentially expressed transcripts by Gene Ontology (GO) enrichment analyses, run separately on the 16 co-expression clusters. Note, the GO results should be interpreted cautiously, as mere indications of functional divergence between groups. The analyses were restricted to biological processes and lower level
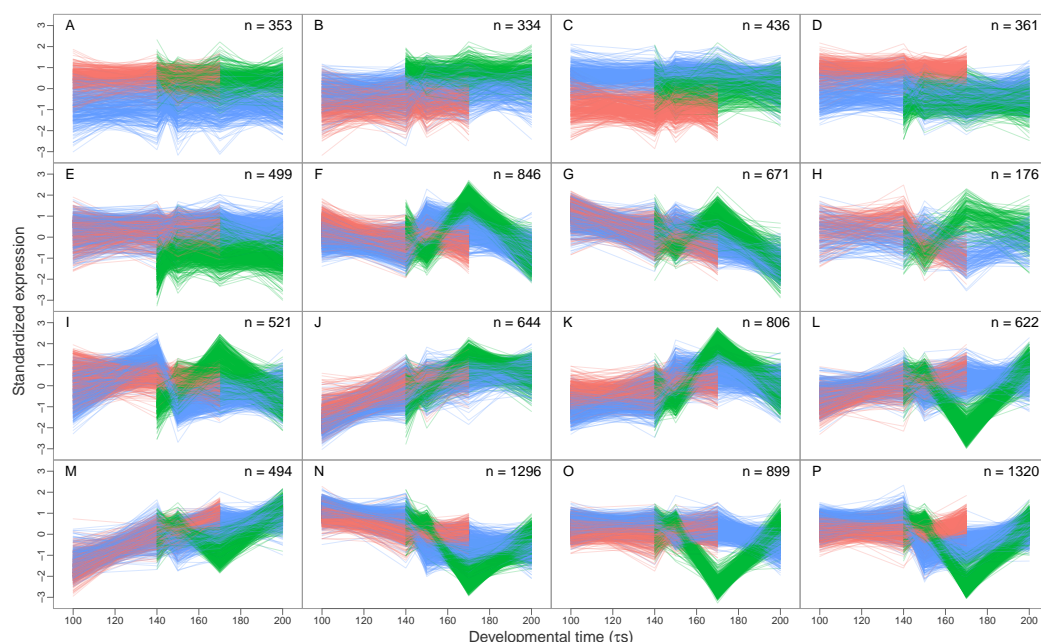
Figure 4 **Expression profiles of 16 co-expression clusters. Depicted are transcripts with significant Morph and/or Morph by Time interaction effect, during early charr development (timepoints 100 τs to 200 τs).** Standardized expression normalized by 3′-coverage is plotted against developmental time. Each line is the profile for one transcript. The first five clusters (A–E) capture mainly differences between morphs, while the remaining clusters (F–P) contain almost exclusively transcripts with *MxT* effects (Table 5). The number of transcripts in each cluster is indicated. The morphs are represented by color, SB: blue, LB: green and PL: red.

categories. The number of significantly enriched GO categories varied between clusters. Five clusters (A, B, C, E and H) did not have any significant GO enrichment (Table S3), in part reflecting low statistical power as those clusters had the fewest transcripts (176 to 499). The clusters with the largest number of significant GO categories (N, O and P) contained the largest number of transcripts. As was noted above, the five co-expression clusters of transcripts with temporally stable expression that varied between morphs (A–E) had no GO enrichment with the exception of cluster D (transcripts with higher expression in PL-charr, than either LB and SB) which had two GO categories (GO:0097360 and GO:0061450, involved in cell migration and proliferation). Combining all the transcripts in these five clusters in GO-enrichment did not yield any significant GO-categories. The same was true for GO analyses of all transcripts with only Morph effect.

Just under 700 GO categories were enriched for clusters of genes with significant Morph by Time interactions (Table S3). While dozens up to a hundred GO categories associated with each co-expression cluster, no general pattern emerged. Many different biological processes were enriched in the co-expression clusters, for instance cluster F was enriched for regulation of growth (e.g., GO:0040008) and antigen processing and presentation (e.g., GO:0048002) and cluster L cartilage condensation (e.g., GO:0001502) and limb bud formation (e.g., GO:0060174). A number of categories showed up in three or more
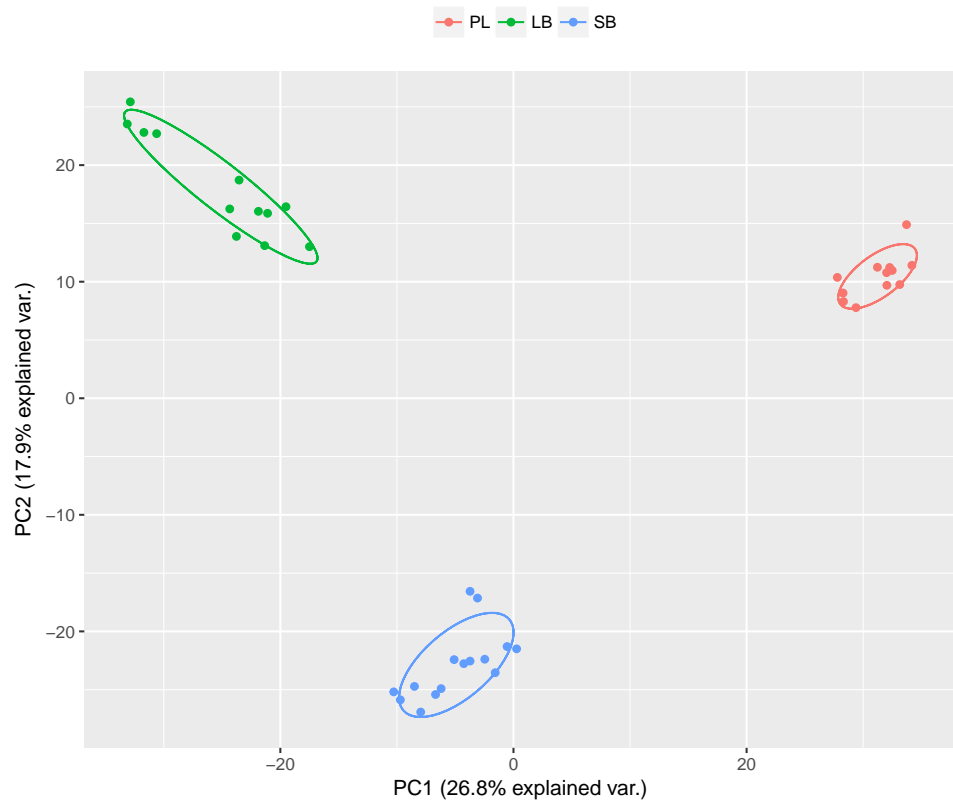
Figure 5 **Visualization of the expression differences between the three morphs, with PCA of transcripts in the first five expression clusters (A–E, 1983 transcripts).** These clusters were selected as they show an effect of morph, and therefore some separation of morphs in the PCA is expected, but importantly all three morphs are separated from each other. The position of the different morphs in the PCA is informative and indicates that the benthic morphs are more similar in expression as the first axis separates all the morphs with the benthic morphs being close to each other. The second axis separates small benthic from the other two morphs. Standardized expression normalized by 3′-coverage was used as input. Individual samples are graphed (colored by morph, SB: blue, LB: green and PL: red), and overlaid are 68% normal data ellipses for each morph. Figure prepared using the ggbiplot package in R (version 0.55; *Vu, 2011*).

Table 5 **Number of transcripts differentially expressed (*fdr* < 0.01) for Morph, Time or Morph by Time interaction ($M \times T$) for co-expression clusters A–E and F–P.**

	Clusters	
Effect	**A–E**	**F–P**
Morph	1,768	234
Time	307	5,761
$M \times T$	296	8,111
clusters, for example; GO:1903047, mitotic cell cycle process (clusters F, G, N, O and P), GO:0022613, ribonucleoprotein complex biogenesis (clusters F, G and N) and GO:0007507, heart development (clusters M, O and P). The diversity of GO categories to us suggests that multiple systems are differentially expressed during early development in these three charr morphs.

Our published data (*Guðbrandsson et al., 2016*) had revealed higher expression of genes related to mitochondrial and energy metabolism in aquaculture compared to SB-charr. We hypothesized that this might reflect higher metabolism in the aquaculture charr (due to artificial selection for increased growth) or reduced metabolism in the small benthic charr (adapting to the spring habitat). The current data support the former explanation, because only one GO category functionally related to those processes is significant in our analysis (GO:0022900, electron transport chain) in a cluster were SB does not stand out (cluster K).

In summary, the data revealed considerable expression separation of these three sympatric morphs, during early development. The expression divergence was seen in multiple genes and diverse biological systems. This suggests that the morphs differ in many aspects of development and physiology and that these differences manifest in embryos well prior (100–200 τs) to hatching (about 270–280 τs).

Verification of differential expression with qPCR

In order to verify morph specific differences in expression indicated in the data we queried a subset of genes from several of the co-expression clusters with qPCR in whole embryos. We studied the same three morphs (PL-, LB- and SB-charr) and tested seven candidate genes at two developmental timepoints (150 and 170 τs) with different expression in the benthic morphs (LB- and SB-charr) and limnetic morph (PL-charr) in the transcriptome. Note, the primers amplified mRNA of paralogous genes, which will be less sensitive if the two paralogs differ in expression (as was seen for *natterin-like genes* (*Guðbrandsson et al., 2016*)). Expression of six genes *MAM Domain Containing 2 (mamdc2)*, *delta(4)-desaturase, sphingolipid 2 (des2)*, *translin (tsn)*, *glucose 6-phosphate isomerase (gpi)*, *protein regulator of cytokinesis 1 (prc1)* and *AT-Rich Interaction Domain 4A (arid4a)*, differed significantly among morphs ($p < 0.05$). The seventh gene *eukaryotic translation initiation factor 4E binding protein 1 (eif4ebp1)* showed a suggestive limnetic and benthic separation in the qPCR (only formally significant at 170 τs) (Fig. 6). Notably, *arid4a* showed the same Morph by Time interaction in both the transcriptome and qPCR. In sum, the general agreement between the transcriptome results and the qPCR verification tests on whole embryos, suggests the majority of the ~2,000 morph effect transcripts represent true differences in expression.

Of those genes, three (*arid4a*, *tsn* and *eif4ebp1*) have known regulatory functions. *Arid4a* encodes a Retinoblastoma binding protein, that has been demonstrated to repress transcription and induce growth arrest in human cell culture (*Lai et al., 1999*). *Tsn* encodes a protein which positively influences the activity of the RISC complex (*Liu et al., 2009*). *Eif4ebp1* encodes a repressor of translation initiation, and is a target of *mTOR* (*Wang et al., 2005*; *Dowling et al., 2010*). The other genes have diverse functions, *prc1* is a cell cycle related gene (*Li, Shridhar & Liu, 2003*), *gpi* a glycolytic enzyme differentially expressed in
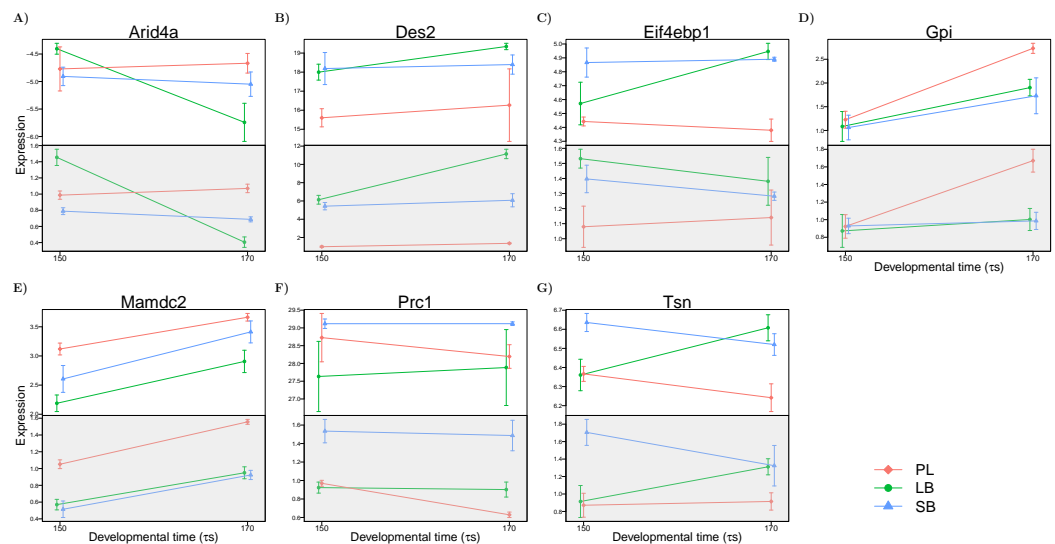
Figure 6 **Expression of seven genes that differed between benthic (SB and LB) and limnetic (PL) charr in the transcriptome.** (A–G) shows the expression of one gene at developmental timepoints 150 and 170 τs, in the transcriptome (white background) and measured with qPCR (gray background). Colors indicate morph (blue: SB, green: LB, red: PL). The upper panels show expression in transcripts per million (tpm) on log-scale, normalized by the effect of 3′ coverage in the linear model (see Methods). The qPCR expression is normalized to the geometric mean of two reference genes (*actb* and *ub2l3*) and represented relative to one replicate of the PL morph at 150 (τs). Error bars represent 2 standard errors of the mean calculated from three biological replicates each made from a homogenate of three whole embryos.

zebrafish development (*Lin et al., 2009*), *des2* is involved in sphingolypid synthesis (*Omae et al., 2004*) whereas the function of *mamdc2* is poorly characterized.

In the light of prior data we focus the discussion on the benthic-limnetic patterns of *eif4ebp1* expression. The gene had higher expression in the benthic charr (formally significant in the transcriptome but only the later timepoint with qPCR). *Macqueen et al. (2011)* found similarly higher expression of this gene and two other *mTOR* pathway related genes in muscles of five small benthic vs two limnetic morphs from south Iceland. Preliminary analyses of this transcriptome (J Gudbrandson et al., 2018, unpublished data) indicate differences in allele frequency of variants in *eif4ebp1* between SB- and PL-charr. These observations do not prove the involvement of *eif4ebp1* in morph differentiation, but call for further study of mTOR pathway genes in different Thingvallavatn morphs and benthic vs. limnetic charr. It must be emphasized that the data presented here are correlative, and do not prove causal influence of these genes on charr development or divergence.

Previously (*Ahi et al., 2014*; *Ahi et al., 2015*) we screened for candidate genes involved in craniofacial development, utilizing our published data (*Guðbrandsson et al., 2016*) and this dataset. We focused on genes with differential expression between limnetic and benthic morphs involved in bone and cartilage development or with craniofacial expression in Zebrafish, and also mined online databases for conserved patterns of co-expression among candidates (*Ahi et al., 2014*; *Ahi et al., 2015*). Several genes showed clearly overlapping

expression in perichondrial regions of the pharyngeal arches during their formation. Interestingly, binding sites for the transcription factor *ets2*, which shows the same expression pattern, are conserved upstream of the co-expressed genes in species as distantly related as *Oryzias latipes* and *Drosophila melanogaster* (*Ahi et al., 2014*). A second study revealed more genes with clear benthic-limnetic separation in expression, and pointed to transcription factors in the glucocorticoid and Aryl hydrocarbon pathways as potential modulators of benthic-limnetic diversity (*Ahi et al., 2015*).

These results and the current data suggest that multiple developmental systems have diverged in these three sympatric morphs, likely reflecting substantial genetic differentiation at multiple loci. Therefore an obvious next step is to ascertain genome-wide data on the genetic separation of the morphs, for instance by mining this transcriptome for sequence polymorphisms (already in progress, J Guðbrandsson et al., 2018, unpublished data). Alternative approaches could be whole genome scans of divergence e.g., (*Jones et al., 2012*; *Halldórsdóttir & Árnason, 2015*) or quantitative trait loci (QTL)/association studies e.g., (*Zimmerman, Palsson & Gibson, 2000*; *Palsson et al., 2005*) of specific ecological traits to identify putative causative factors and variants that differentiate these sympatric morphs. Furthermore as dwarf charr are found in multiple locations, it would be interesting to study their transcriptomes, perhaps at finer developmental resolution to test the reproducibility of developmental changes in evolution. Also, while the sympatric morphs of Lake Thingvallavatn are clearly demarcated phenotypically, subtler signs of polymorphism are found in several lakes (*Woods et al., 2012*). Molecular and genetic analyses of several morph pairs varying in degree of divergence would be exciting.

CONCLUSIONS

The differences in trophic morphology, habitat use and life history traits among the sympatric charr morphs in Lake Thingvallavatn have intrigued students of fish biology and evolution for more than a century (*Sæmundsson, 1904*; *Frioriksson, 1939*; *Snorrason et al., 1989*; *Skúlason et al., 1996*; *Ahi et al., 2015*). Genes, environment and parental effects are known to contribute to the morph differences (*Snorrason et al., 1994*; *Skúlason, Snorrasson & Jónsson, 1999*; *Leblanc, Kristjánsson & Skúlason, 2016*). The LB-, SB- and PL-charr differ significantly at the genetic level, but the estimates of relatedness and phylogenetic relationships of the three morphs vary by studies (*Volpe & Ferguson, 1996*; *Gíslason, 1998*; *Kapralova et al., 2011*). With the current experimental design parental effects can not be excluded. We still postulate that large fraction of the expression differences between morphs stem from genetic differences. The observed pattern at the expression level, that all morphs are separated and the benthic morphs are more similar (this data), suggests that it is important to follow this work with investigation of the polymorphism trends in the transcriptome (Johannes Guðbrandsson et al., in preparation). A population genomic screen may be needed to evaluate these relationships and the origin of the Lake Thingvallavatn morphs. We find that expression of multiple genes differs between the three charr morphs during early development and prior to hatching. This observation and previous studies on co-expressed genes (*Ahi et al., 2014*; *Ahi et al., 2015*) indicate that

during development, upstream regulatory mechanisms may be acting differently in these morphs. Thus differential expression of regulators such as *tsn*, *ahr2* (*Ahi et al., 2015*) or *eif4ebp1* (*Macqueen et al., 2011*), lead us to speculate that they may influence expression at multiple loci and cause differences in ecologically important traits, e.g., concerning the structure and function of the feeding apparatus and muscle growth (*Sandlund et al., 1992*; *Macqueen et al., 2011*).

Although the genes identified here and in our previous studies (*Ahi et al., 2014*; *Ahi et al., 2015*) may constitute key links in developmental cascades that through differential expression (timing and pattern) induce morph differences, the underlying genetic differences have not been identified. They may reside in the *cis*-elements of some of these genes, but more likely in up-stream members of pathways that regulate development. Identifying the causative molecular changes associated with evolutionary divergence is not straightforward (*Santure et al., 2015*), in part because of the pleiotropic nature of metabolic, homeostatic and developmental systems (*Paaby & Rockman, 2013*). One intriguing question is whether the heritable expression differences between morphs is due to variation in one gene, few genes or many QTLs? Our combined data (*Ahi et al., 2014*; *Ahi et al., 2015*; *Guðbrandsson et al., 2016*) including the present data, argues against a monogenic model, i.e., where a single gene is responsible for the observed morph differences. The data is, in our opinion, more consistent with divergence in multiple systems and thus in many genomic regions among morphs (polygenic model). To disentangle the molecular systems responsible for morph divergence the anatomical focus must be sharpened by studying gene expression in specific tissues (head or jaw) or cell types at particular developmental time-points. Another option is a genomic scan of divergence that may implicate specific loci or systems. The intersection of genes or systems that show both genetic and expression difference between morphs is naturally interesting. Although several studies have found one or few genes that contribute heavily to key traits among closely related morphs/species (*Shapiro et al., 2004*; *Johnston et al., 2013*; *Kunte et al., 2014*) in many cases divergence in numerous genes influencing multiple cellular, developmental and physiological systems is a more likely scenario (*Flint & Mackay, 2009*; *Coolon et al., 2014*; *Laporte et al., 2015*), as seems to be the case for the Arctic charr morphs in Lake Thingvallavatn.

ACKNOWLEDGEMENTS

We are very grateful to Droplaug N. Magnúsdóttir, Gudbjörg Th. Örlygsdóttir, Steinunn Snorradóttir and Ólafur Th. Magnússon at deCODE Genetics for help with the Illumina sequencing. Special thanks to Soizic le Deuff and the staff at Verið for caring for and sampling embryos. We thank Eggert Pétursson kindly for allowing use of his drawings of Arctic charr. We thank Páll Melsted for advice on the transcriptome assembly, Benjamín Sigurgeirsson for advice on 3′-bias detection, Fredrik Holm for helping with images and all members of the charr group for support, comments and suggestions. We like to thank Ben Sutherland and an anonymous reviewer for their comments and suggestions that improved the paper. We are grateful to all the fish for their parts.

Funding

The project was funded by The Icelandic Center for Research (grant number: 100204011) to Sigurður S. Snorrason, Arnar Pálsson, Zophonías O. Jónsson and Bjarni K. Kristjánsson. The University of Iceland Doctoral Fund to Jóhannes Guðbrandsson and University of Iceland research fund to Arnar Pálsson, Sigurður S. Snorrason and Zophonías O. Jónsson. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Icelandic Center for Research: 100204011.
The University of Iceland Doctoral Fund.
University of Iceland research fund.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Jóhannes Guðbrandsson conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper, conducted transcriptome assembly, annotation and analyses.
- Sigríður Rut Franzdóttir conceived and designed the experiments, performed the experiments, wrote the paper, reviewed drafts of the paper, prepared sequencing libraries.
- Bjarni Kristófer Kristjánsson performed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Ehsan Pashay Ahi performed the experiments, analyzed the data, prepared figures and/or tables, reviewed drafts of the paper, extracted RNA and did qPCR confirmation.
- Valerie Helene Maier performed the experiments, reviewed drafts of the paper, extracted RNA.
- Kalina Hristova Kapralova performed the experiments, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper, stained embryos and photographed.
- Sigurður Sveinn Snorrason conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.
- Zophonías Oddur Jónsson conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper, prepared sequencing libraries.
- Arnar Pálsson conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Ethics committee approval is not needed for regular or scientific fishing in Iceland (The Icelandic law on Animal protection, Law 15/1994, last updated with Law 55/2013). Sampling of embryos for RNA extraction was performed by Holar 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.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Fishing was performed with the permissions both from the owner of the land in Mjóanes and from the Thingvellir National Park commission.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The sequencing reads from the 48 samples were deposited into the NCBI SRA archive under BioProject identifier PRJNA391695 and with accession numbers: SRS2316381 to SRS2316428.

Data Availability

Table S5: Guðbrandsson, Jóhannes (2018): Annotation of de novo assembly from Arctic charr in Lake Thingvallavatn, Iceland. figshare. https://doi.org/10.6084/m9.figshare.5445610.v2.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4345#supplemental-information.

REFERENCES

Abernathy J, Overturf K. 2016. Comparison of Ribosomal RNA removal methods for transcriptome sequencing workflows in teleost fish. *Animal Biotechnology* **27(1)**:60–65 DOI 10.1080/10495398.2015.1086365.

Abouheif E, Favé M-J, Ibarrarán-Viniegra AS, Lesoway MP, Rafiqi AM, Rajakumar R. 2014. Eco-Evo-Devo: the time has come. In: Landry CR, Aubin-Horth N, eds. *Ecological genomics—ecology and the evolution of genes and genomes.* Dordrecht: Springer Netherlands, 107–125.

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442(7102)**:563–567 DOI 10.1038/nature04843.

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. 2004. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* **305**(**5689**):1462–1465 DOI 10.1126/science.1098095.

Adams CE, Fraser D, Wilson AJ, Alexander G, Ferguson MM, Skulason S. 2007. Patterns of phenotypic and genetic variability show hidden diversity in Scottish Arctic charr. *Ecology of Freshwater Fish* **16**(**1**):78–86 DOI 10.1111/j.1600-0633.2006.00182.x.

Adams CE, Huntingford FA. 2004. Incipient speciation driven by phenotypic plasticity? Evidence from sympatric populations of Arctic charr. *Biological Journal of the Linnean Society* **81**(**4**):611–618 DOI 10.1111/j.1095-8312.2004.00314.x.

Ahi EP, Guðbrandsson J, Kapralova KH, Franzdóttir SR, Snorrason SS, Maier VH, Jónsson ZO. 2013. Validation of reference genes for expression studies during craniofacial development in arctic charr. *PLOS ONE* **8**(**6**):e66389 DOI 10.1371/journal.pone.0066389.

Ahi EP, Kapralova KH, Pálsson A, Maier VH, Guðbrandsson J, Snorrason SS, Jónsson ZO, Franzdóttir SR. 2014. Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr. *EvoDevo* **5**(**1**):40 DOI 10.1186/2041-9139-5-40.

Ahi EP, Steinhäuser SS, Pálsson A, Franzdóttir SR, Snorrason SS, Maier VH, Jónsson ZO. 2015. Differential expression of the aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr. *EvoDevo* **6**(**1**):27 DOI 10.1186/s13227-015-0022-6.

Alexandrou MA, Swartz BA, Matzke NJ, Oakley TH. 2013. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution* **69**(**3**):514–523 DOI 10.1016/j.ympev.2013.07.026.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(**1**):289–300 DOI 10.2307/2346101.

Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, Lu G, Nolte AW, Ostbye K, Rogers SM, St-Cyr J. 2010. On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **365**(**1547**):1783–1800 DOI 10.1098/rstb.2009.0274.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, Aury J-M, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff J-N, Genêt C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**:3657 DOI 10.1038/ncomms4657.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**(**5**):525–527 DOI 10.1038/nbt.3519.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines:
minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55(4)**:611–622 DOI 10.1373/clinchem.2008.112797.

Carlson M. 2015. GO.db: a set of annotation maps describing the entire Gene Ontology. R package version 3.2.2 DOI 10.18129/B9.bioc.GO.db.

Carroll SB. 2008. Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134(1)**:25–36 DOI 10.1016/j.cell.2008.06.030.

Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Research* **24(5)**:797–808 DOI 10.1101/gr.163014.113.

Crête-Lafrenière A, Weir LK, Bernatchez L. 2012. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLOS ONE* **7(10)**:e46662 DOI 10.1371/journal.pone.0046662.

Danzmann RG, Ferguson MM, Skulason S, Snorrason SS, Noakes DLG. 1991. Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn, Iceland. *Journal of Fish Biology* **39(5)**:649–659 DOI 10.1111/j.1095-8649.1991.tb04395.x.

Danzmann RG, Kocmarek AL, Norman JD, Rexroad CE, Palti Y. 2016. Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics* **17(1)**:60 DOI 10.1186/s12864-016-2363-5.

Di Génova A, Aravena A, Zapata L, González M, Maass A, Iturra P. 2011. SalmonDB: a bioinformatics resource for *Salmo salar* and *Oncorhynchus mykiss*. *Database : The Journal of Biological Databases and Curation* **2011**:bar050 DOI 10.1093/database/bar050.

Dowling RJO, Topisirovic I, Alain T, Bidinosti M, Fonseca BD, Petroulakis E, Wang X, Larsson O, Selvaraj A, Liu Y, Kozma SC, Thomas G, Sonenberg N. 2010. mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs. *Science* **328(5982)**:1172–1176 DOI 10.1126/science.1187532.

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

Eiriksson GM, Skulason S, Snorrason SS. 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 DOI 10.1111/j.1095-8649.1999.tb01054.x.

Flint J, Mackay TFC. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research* **19(5)**:723–733 DOI 10.1101/gr.086660.108.

Friðriksson Á. 1939. Um murtuna í þingvallavatni með hliðsjón af öðrum silungstegundum í vatninu. *Náttúrufræðingurinn* **9(1)**:1–15.

Futschik M. 2015. Mfuzz: soft clustering of time series gene expression data. R package version 2.30.0. *Available at http://www.sysbiolab.eu/software/R/Mfuzz/index.html*.

Garfield DA, Runcie DE, Babbitt CC, Haygood R, Nielsen WJ, Wray GA. 2013. The impact of gene expression variation on the robustness and evolvability

of a developmental gene regulatory network. *PLOS Biology* **11(10)**:e1001696 DOI 10.1371/journal.pbio.1001696.

Giger T, Excoffier L, Day PJR, Champigneulle A, Hansen MM, Powell R, Largiadèr CR. 2006. Life history shapes gene expression in salmonids. *Current Biology* **16(8)**:R281–282.

Gíslason D. 1998. Genetic and morphological variation in polymorphic arctic charr *Salvelinus alpinus*, from Icelandic lakes. Master's thesis, The University of Guelph.

Gorodilov YN. 1996. Description of the early ontogeny of the Atlantic salmon, *Salmo salar*, with a novel system of interval (state) identification. *Environmental Biology of Fishes* **47(2)**:109–127 DOI 10.1007/BF00005034.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29(7)**:644–652 DOI 10.1038/nbt.1883.

Grant B, Grant PR. 2008. Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363(1505)**:2821–2829 DOI 10.1098/rstb.2008.0051.

Grant PR. 1999. Ecology and evolution of Darwin's Finches. Second Edition. Princeton: Princeton University Press.

Grünbaum T, Cloutier R, Mabee PM, Le François NR. 2007. Early developmental plasticity and integrative responses in arctic Charr (*Salvelinus alpinus*): effects of water velocity on body size and shape. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **308B(4)**:396–408 DOI 10.1002/jez.b.21163.

Guðbrandsson J, Ahi EP, Franzdottir SR, Kapralova KH, Kristjansson BK, Steinhaeuser SS, Maier VH, Johannesson IM, Snorrason SS, Jonsson ZO, Palsson A. 2016. The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs (version 3; referees: 2 approved, 1 approved with reservations). *F1000Research* **4**:136 DOI 10.12688/f1000research.6402.3.

Haas BJ. 2015. Trinotate: transcriptome functional annotation and analysis. Version 2.0.2. *Available at http://trinotate.github.io/*.

Halldórsdóttir K, Árnason E. 2015. Whole-genome sequencing uncovers cryptic and hybrid species among Atlantic and Pacific cod-fish. *BioRxiv* DOI 10.1101/034926.

Hindar K, Jonsson B. 1993. Ecological polymorphism in Arctic charr. *Biological Journal of the Linnean Society* **48(1)**:63–74 DOI 10.1111/j.1095-8312.1993.tb00877.x.

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. 2001. The contributions of sex, genotype and age to transcriptonal variance in *Drosophila melanogaster*. *Nature Genetics* **29(4)**:389–395 DOI 10.1038/ng766.

Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM, Slate J. 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502(7469)**:93–95 DOI 10.1038/nature12489.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E, Searle S, Schmutz J, Grimwood J, Dickson MC,

Myers RM, Miller CT, Summers BR, Knecht AK, Brady SD, Zhang H, Pollen AA, Howes T, Amemiya C, Baldwin J, Bloom T, Jaffe DB, Nicol R, Wilkinson J, Lander ES, Di Palma F, Lindblad-Toh K, Kingsley DM. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484(7392)**:55–61 DOI 10.1038/nature10944.

Jonsson B, Jonsson N. 2014. Early environment influences later performance in fishes. *Journal of Fish Biology* **85(2)**:151–188 DOI 10.1111/jfb.12432.

Jonsson B, Skúlason S, Snorrason SS, Sandlund OT, Malmquist HJ, Jónasson PM, Cydemo R, Lindem T. 1988. Life history variation of polymorphic arctic charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland. *Canadian Journal of Fisheries and Aquatic Sciences* **45(9)**:1537–1547 DOI 10.1139/f88-182.

Kapralova KH, Franzdóttir SR, Jónsson H, Snorrason SS, Jónsson ZO. 2014. Patterns of MiRNA expression in arctic charr development. *PLOS ONE* **9(8)**:e106084 DOI 10.1371/journal.pone.0106084.

Kapralova KH, Guðbrandsson J, Reynisdottir S, Santos CB, Baltanás VC, Maier VH, Snorrason SS, Palsson A. 2013. Differentiation at the *MHCIIα* and *Cath2* Loci in sympatric *salvelinus alpinus* resource morphs in lake Thingvallavatn. *PLOS ONE* **8(7)**:e69402 DOI 10.1371/journal.pone.0069402.

Kapralova KH, Jónsson ZO, Pálsson A, Franzdóttir SR, Le Deuf S, Kristjánsson BK, Snorrason SS. 2015. Bones in motion: ontogeny of craniofacial development in sympatric Arctic charr morphs. *Developmental Dynamics* **244(9)**:1168–1178 DOI 10.1002/dvdy.24302.

Kapralova KH, Morrissey MB, Kristjánsson BK, Olafsdóttir GÁ, Snorrason SS, Ferguson MM. 2011. Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity* **106(3)**:472–487 DOI 10.1038/hdy.2010.161.

Klemetsen A. 2010. The charr problem revisited: exceptional phenotypic plasticity promotes ecological speciation in postglacial lakes. *Freshwater Reviews* **3(1)**:49–74 DOI 10.1608/FRJ-3.1.3.

Koop BF, Von Schalburg KR, Leong J, Walker N, Lieph R, Cooper GA, Robb A, Beetz-Sargent M, Holt RA, Moore R, Brahmbhatt S, Rosner J, Rexroad CE, McGowan CR, Davidson WS. 2008. A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics* **9(1)**:545 DOI 10.1186/1471-2164-9-545.

Kopp A, Duncan I, Carroll SB. 2000. Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* **408(6812)**:553–559 DOI 10.1038/35046017.

Kristjánsson BK, Malmquist HJ, Ingimarsson F, Antonsson T, Snorrason SS, Skúlason S. 2011. Relationships between lake ecology and morphological characters in Icelandic Arctic charr, *Salvelinus alpinus*. *Biological Journal of the Linnean Society* **103(4)**:761–771 DOI 10.1111/j.1095-8312.2011.01670.x.

Kristjánsson BK, Skúlason S, Snorrason SS, Noakes DL. 2012. Fine-scale parallel patterns in diversity of small benthic Arctic charr (*Salvelinus alpinus*) in relation to the ecology of lava/groundwater habitats. *Ecology and Evolution* **2(6)**:1099–1112 DOI 10.1002/ece3.235.

Krueger F. 2012. Trim Galore! Version 0.3.3. Babraham Bioinformatics. *Available at* http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, Mullen SP, Kronforst MR. 2014. *doublesex* is a mimicry supergene. *Nature* **507(7491)**:229–232 DOI 10.1038/nature13112.

Küttner E, Parsons KJ, Robinson BW, Skúlason S, Danzmann RG, Ferguson MM. 2013. Effects of population, family, and diet on craniofacial morphology of Icelandic Arctic charr (*Salvelinus alpinus*). *Biological Journal of the Linnean Society* **108(3)**:702–714 DOI 10.1111/j.1095-8312.2012.02038.x.

Lai A, Marcellus RC, Corbeil HB, Branton PE. 1999. RBP1 induces growth arrest by repression of E2F-dependent transcription. *Oncogene* **18(12)**:2091–2100 DOI 10.1038/sj.onc.1202520.

Laporte M, Rogers SM, Dion-Côté A-M, Normandeau E, Gagnaire P-A, Dalziel AC, Chebib J, Bernatchez L. 2015. RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish (*Coregonus clupeaformis*) species pairs. *G3: Genes, Genomes, Genetics* **5(7)**:1481–1491 DOI 10.1534/g3.115.019067.

Leblanc CA-L, Kristjánsson BK, Skúlason S. 2016. The importance of egg size and egg energy density for early size patterns and performance of Arctic charr *Salvelinus alpinus*. *Aquaculture Research* **47(4)**:1100–1111 DOI 10.1111/are.12566.

Li C, Shridhar K, Liu J. 2003. Molecular characterization of oncostatin M-induced growth arrest of MCF-7 cells expressing a temperature-sensitive mutant of p53. *Breast Cancer Research and Treatment* **80(1)**:23–37 DOI 10.1023/A:1024483017549.

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, Von Schalburg K, Rondeau EB, Di Genova A, Samy JKA, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Inge Våge D, De Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJM, Jonassen I, Maass A, Omholt SW, Davidson WS. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature* **533(7602)**:200–205 DOI 10.1038/nature17164.

Lin W-W, Chen L-H, Chen M-C, Kao H-W. 2009. Differential expression of zebrafish *gpia* and *gpib* during development. *Gene Expression Patterns* **9(4)**:238–245 DOI 10.1016/j.gep.2008.12.006.

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009. On the origin and spread of an adaptive allele in deer mice. *Science* **325(5944)**:1095–1098 DOI 10.1126/science.1175826.

Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q. 2009. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* **325(5941)**:750–753 DOI 10.1126/science.1176325.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. *Methods* **25(4)**:402–408 DOI 10.1006/meth.2001.1262.

Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, Kreitman M. 2005. Functional evolution of a *cis*-regulatory module. *PLOS Biology* **3(4)**:e93 DOI 10.1371/journal.pbio.0030093.

Macqueen DJ, Johnston IA. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences* **281(1778)**:20132881 DOI 10.1098/rspb.2013.2881.

Macqueen DJ, Kristjánsson BK, Paxton CGM, Vieira VLA, Johnston IA. 2011. The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. *Molecular Ecology* **20(15)**:3167–3184 DOI 10.1111/j.1365-294X.2011.05172.x.

Magnusson KP, Ferguson MM. 1987. Genetic analysis of four sympatric morphs of Arctic charr, *Salvelinus alpinus*, from Thingvallavatn, Iceland. *Environmental Biology of Fishes* **20(1)**:67–73 DOI 10.1007/BF00002026.

Malmquist HJ, Snorrason SS, Skulason S, Jonsson B, Sandlund OT, Jonasson PM. 1992. Diet differentiation in polymorphic arctic charr in thingvallavatn, Iceland. *The Journal of Animal Ecology* **61(1)**:21–35 DOI 10.2307/5505.

Moghadam HK, Ferguson MM, Danzmann RG. 2011. Whole genome duplication: challenges and considerations associated with sequence orthology assignment in Salmoninae. *Journal of Fish Biology* **79(3)**:561–574 DOI 10.1111/j.1095-8649.2011.03030.x.

Muir AM, Hansen MJ, Bronte CR, Krueger CC. 2016. If Arctic charr *Salvelinus alpinus* is the most diverse vertebrate', what is the lake charr *Salvelinus namaycush*? *Fish and Fisheries* **17(4)**:1194–1207 DOI 10.1111/faf.12114.

Noakes DLG. 2008. Charr truth: sympatric differentiation in *Salvelinus* species. *Environmental Biology of Fishes* **83(1)**:7–15 DOI 10.1007/s10641-008-9379-x.

Nordeng H. 1983. Solution to the ''Char Problem'' based on Arctic Char (*Salvelinus alpinus*) in Norway. *Canadian Journal of Fisheries and Aquatic Sciences* **40(9)**:1372–1387 DOI 10.1139/f83-159.

Norman JD, Ferguson MM, Danzmann RG. 2014. Transcriptomics of salinity tolerance capacity in Arctic charr (*Salvelinus alpinus*): a comparison of gene expression profiles between divergent QTL genotypes. *Physiological Genomics* **46(4)**:123–137 DOI 10.1152/physiolgenomics.00105.2013.

Nugent CM, Easton AA, Norman JD, Ferguson MM, Danzmann RG. 2017. A SNP based linkage map of the arctic charr (*Salvelinus alpinus*) genome provides insights into the diploidization process after whole genome duplication. *G3: Genes, Genomes, Genetics* **7(2)**:543–556 DOI 10.1534/g3.116.038026.

Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. *Nature Genetics* **32(2)**:261–266 DOI 10.1038/ng983.

Omae F, Miyazaki M, Enomoto A, Suzuki A. 2004. Identification of an essential sequence for dihydroceramide C-4 hydroxylase activity of mouse DES2. *FEBS Letters* **576(1–2)**:63–67 DOI 10.1016/j.febslet.2004.08.060.

Paaby AB, Rockman MV. 2013. The many faces of pleiotropy. *Trends in Genetics* **29(2)**:66–73 DOI 10.1016/j.tig.2012.10.010.

Palsson A, Dodgson J, Dworkin I, Gibson G. 2005. Tests for the replication of an association between *Egfr* and natural variation in *Drosophila melanogaster* wing morphology. *BMC Genetics* **6(1)**:44 DOI 10.1186/1471-2156-6-44.

Parsons KJ, Sheets HD, Skúlason S, Ferguson MM. 2011. Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (*Salvelinus alpinus*). *Journal of Evolutionary Biology* **24(8)**:1640–1652 DOI 10.1111/j.1420-9101.2011.02301.x.

Parsons KJ, Skúlason S, Ferguson M. 2010. Morphological variation over ontogeny and environments in resource polymorphic arctic charr (*Salvelinus alpinus*). *Evolution and Development* **12(3)**:246–257 DOI 10.1111/j.1525-142X.2010.00410.x.

Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology* **14(2)**:e1002379 DOI 10.1371/journal.pbio.1002379.

Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, Pritchard JK, Gilad Y. 2012. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research* **22(4)**:602–610 DOI 10.1101/gr.130468.111.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* **14(7)**:687–690 DOI 10.1038/nmeth.4324.

R Core Team. 2017. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Roberts A, Pachter L. 2012. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* **10(1)**:71–73 DOI 10.1038/nmeth.2251.

Robertson FM, Gundappa MK, Grammes F, Hvidsten TR, Redmond AK, Lien S, Martin SAM, Holland PWH, Sandve SR, Macqueen DJ. 2017. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology* **18(1)**:111 DOI 10.1186/s13059-017-1241-z.

Robinson BW, Parsons KJ. 2002. Changing times, spaces, and faces: tests and implications of adaptive morphological plasticity in the fishes of northern postglacial lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **59(11)**:1819–1833 DOI 10.1139/f02-144.

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics* **13(7)**:505–516 DOI 10.1038/nrg3229.

Sæmundsson B. 1904. Fiskirannsóknir 1902. *Andvari* **29(1)**:79–119.

Samy JKA, Mulugeta TD, Nome T, Sandve SR, Grammes F, Kent MP, Lien S, Våge DI. 2017. SalmoBase: an integrated molecular data resource for Salmonid species. *BMC Genomics* **18(1)**:482 DOI 10.1186/s12864-017-3877-1.

Sandlund OT, Gunnarsson K, Jónasson PM, Jonsson B, Lindem T, Magnússon KP, Malmquist HJ, Sigurjónsdóttir H, Skúlason S, Snorrason SS, Jonasson PM,

Magnusson KP, Sigurjonsdottir H, Skulason S. 1992. The arctic charr *Salvelinus alpinus* in Thingvallavatn. *Oikos* **64(1/2)**:305–351 DOI 10.2307/3545056.

Santure AW, Poissant J, De Cauwer I, Van Oers K, Robinson MR, Quinn JL, Groenen MAM, Visser ME, Sheldon BC, Slate J. 2015. Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology* **24(24)**:6148–6162 DOI 10.1111/mec.13452.

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**:3 DOI 10.1186/1471-2199-7-3.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428(6984)**:717–723 DOI 10.1038/nature02415.

Sibthorpe D, Sturlaugsdóttir R, Kristjansson BK, Thorarensen H, Skúlason S, Johnston IA. 2006. Characterisation and expression of the paired box protein 7 (*Pax7*) gene in polymorphic Arctic charr (*Salvelinus alpinus*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **145(3–4)**:371–383 DOI 10.1016/j.cbpb.2006.08.013.

Sigurgeirsson B, Emanuelsson O, Lundeberg J. 2014. Sequencing degraded RNA addressed by 3′ tag counting. *PLOS ONE* **9(3)**:e91851 DOI 10.1371/journal.pone.0091851.

Skúlason S, Antonsson T, Gudbergsson G, Malmquist HJ, Snorrason SS. 1992. Varibility in Icelandic Arctic charr. *Icelandic Agricultural Sciences* **6**:143–153.

Skúlason S, Smith TB. 1995. Resource polymorphisms in vertebrates. *Trends in Ecology & Evolution* **10(9)**:366–370 DOI 10.1016/S0169-5347(00)89135-1.

Skúlason S, Snorrason S, Noakes D, Ferguson M. 1996. Genetic basis of life history variations among sympatric morphs of Arctic char *Salvelinus alpinus*. *Canadian Journal of Fisheries and Aquatic Sciences* **53(8)**:1807–1813 DOI 10.1139/cjfas-53-8-1807.

Skúlason S, Snorrason S, Ota D, Noakes D. 1993. Genetically based differences in foraging behaviour among sympatric morphs of arctic charr (Pisces: Salmonidae). *Animal Behaviour* **45(6)**:1179–1192 DOI 10.1006/anbe.1993.1140.

Skúlason S, Snorrasson SS, Jónsson B. 1999. Sympatric morphs, populations and speciation in freshwater fish with emphasis on Arctic charr. In: Magurran AE, May RM, eds. *Evolution of biological diverstiy*. Oxford: Oxford University Press, 70–92.

Smith TB, Skúlason S. 1996. Evolutionary significance of resource polymorphisms in fishes, amphibians, and birds. *Annual Review of Ecology and Systematics* **27**:111–133 DOI 10.1146/annurev.ecolsys.27.1.111.

Snorrason SS, Skúlason S. 2004. Adaptive speciation in northern freshwater fishes. In: Dieckmann U, Doebeli M, Metz JAJ, Tautz D, eds. *Adaptive speciation*. Cambridge: Cambridge University Press, 210–228.

Snorrason SS, Skúlason S, Jonsson B, Malmquist HJ, Jónasson PM, Sandlund OT, Lindem T. 1994. Trophic specialization in Arctic charr *Salvelinus alpinus* (Pisces;

Salmonidae): morphological divergence and ontogenetic niche shifts. *Biological Journal of the Linnean Society* **52(1)**:1–18 DOI 10.1111/j.1095-8312.1994.tb00975.x.

Snorrason SS, Skúlason S, Sandlund OT, Malmquist HJ, Jonsson B, Jonasson PM. 1989. Shape polymorphism in arctic charr, *Salvelinus alpinus*. *Physiology and Ecology Japan* **1**:393–404.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* **62(9)**:2155–2177 DOI 10.1111/j.1558-5646.2008.00450.x.

Telnes T, Sægrov H. 2004. Reproductive strategies in two sympatric morphotypes of Arctic charr in Kalandsvatnet, west Norway. *Journal of Fish Biology* **65(2)**:574–579 DOI 10.1111/j.0022-1112.2004.00432.x.

True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evolution & Development* **3(2)**:109–119 DOI 10.1046/j.1525-142x.2001.003002109.x.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research* **40(15)** DOI 10.1093/nar/gks596.

Volpe JP, Ferguson MM. 1996. Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. *Molecular Ecology* **5(6)**:763–772 DOI 10.1111/j.1365-294X.1996.tb00372.x.

Vu VQ. 2011. ggbiplot: a ggplot2 based biplot. R package version 0.55. *Available at http://github.com/vqv/ggbiplot*.

Walker MB, Kimmel CB. 2007. A two-color acid-free cartilage and bone stain for zebrafish larvae. *Biotechnic & Histochemistry* **82(1)**:23–28 DOI 10.1080/10520290701333558.

Wang X, Beugnet A, Murakami M, Yamanaka S, Proud CG. 2005. Distinct signaling events downstream of mTOR cooperate to mediate the effects of amino acids and insulin on initiation factor 4e-binding proteins. *Molecular and Cellular Biology* **25(7)**:2558–2572 DOI 10.1128/MCB.25.7.2558-2572.2005.

Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology* **15(5)**:1197–1211 DOI 10.1111/j.1365-294X.2006.02868.x.

Woods PJ, Skúlason S, Snorrason SS, Kristjánsson BK, Malmquist HJ, Quinn TP. 2012. Intraspecific diversity in Arctic charr, *Salvelinus alpinus*, in Iceland: II. Which environmental factors influence resource polymorphism in lakes? *Evolutionary Ecology Research* **14(8)**:993–1013.

Wootton RJ. 1984. A functional biology of sticklebacks. Boston: Springer US, 266.

Wu Z, Wang X, Zhang X. 2011. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics* **27(4)**:502–508 DOI 10.1093/bioinformatics/btq696.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11(2)**:R14 DOI 10.1186/gb-2010-11-2-r14.

Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**(7):976–978 DOI 10.1093/bioinformatics/btq064.

Zimmerman E, Palsson A, Gibson G. 2000. Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics* **155**(2):671–683.