

**A peer-reviewed version of this preprint was published in PeerJ on 7 February 2018.**

[View the peer-reviewed version](https://peerj.com/articles/4345) (peerj.com/articles/4345), which is the preferred citable publication unless you specifically need to cite this preprint.

Guðbrandsson J, Franzdóttir SR, Kristjánsson BK, Ahi EP, Maier VH, Kapralova KH, Snorrason SS, Jónsson ZO, Pálsson A. 2018. Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs. PeerJ 6:e4345  
<https://doi.org/10.7717/peerj.4345>

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

Jóhannes Gudbrandsson<sup>1,2</sup>, Sigríður R. Franzdóttir<sup>1,3</sup>, Bjarni K. Kristjánsson<sup>4</sup>, Ehsan P. Ahi<sup>1,5</sup>, Valerie H. Maier<sup>1,3</sup>, Sigurdur S. Snorrason<sup>1</sup>, Zophonías O. Jónsson<sup>1,3</sup>, and Arnar Pálsson<sup>1,3</sup>

<sup>1</sup>Institute of Life and Environmental Sciences, University of Iceland, 101 Reykjavik, Iceland

<sup>2</sup>Marine and Freshwater Research Institute, 101 Reykjavik, Iceland - current address

<sup>3</sup>Biomedical Center, University of Iceland, 101 Reykjavik, Iceland

<sup>4</sup>Holar University College, 551 Saudarkrokur, Iceland

<sup>5</sup>Karl-Franzens-Universität Graz, Austria - current address

Corresponding author:

Jóhannes Gudbrandsson

Email address: johannes.gudbrandsson@hafogvatn.is

## ABSTRACT

Phenotypic differences between closely related taxa or populations can arise through genetic variation or be environmentally induced, in both cases leading to altered transcription of genes during the structural and functional development of the body. Comparative developmental studies of closely related species or variable populations of the same species can help to elucidate the molecular mechanisms related to population 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 a timeframe in early development.

Expectedly, developmental time was the predominant explanatory variable. As the data were affected by RNA-degradation, an estimate of 3'-bias was the second most common explanatory variable. Morph, both as a independent variable and as interaction with developmental time, affected the expression of numerous transcripts. The majority of transcripts with significant morph effects separated the limnetic and the benthic morphs. 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 *Arid4a* and *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 et al., 2002) and environmental factors (Giger et al., 2006; Danzmann et al., 2016). Gene expression can change because of neutral evolution, positive and purifying selection (Romero et al., 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 and Haag, 2001), that is alterations in gene expression and the functions of developmental circuits. Analyses

of gene expression in developing organisms can reveal changes in the developmental circuits and the phenotypes they influence (Garfield et al., 2013), and also alterations in the parameters of these networks (Ludwig et al., 2005). Some open questions of evolutionary developmental biology (Kopp et al., 2000; Carroll, 2008; Stern and Orgogozo, 2008) are: i) How does natural selection operate on developmental variation to polish existing adaptations and generate new traits? ii) Which developmental and cellular systems influence variation in adaptive traits during divergence? iii) Are some developmental processes, timepoints or tissues more prone/amenable to natural selection than others?

To address those questions, and estimate the influence of positive selection, purifying selection and developmental system drift in evolutionary divergence, we can identify 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 factor 4* and *Calmodulin* during beak development has strong effects on beak depth and width (Abzhanov et al., 2004, 2006), which are important characteristics for fitness (Grant, 1999; Grant and 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 of gene expression during key events in 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 retreat of the last ice-age (~12,000 years ago) salmonid species and stickleback were prominent among fish species that colonized newly formed lakes and rivers of the northern hemisphere (Wootton, 1984; Noakes, 2008; Klemetsen, 2010). Being the most cold tolerant salmonid, anadromous Arctic charr (*Salvelinus alpinus*) are assumed to have been the earliest fish colonizing these habitats (Wilson et al., 2004; Noakes, 2008; Klemetsen, 2013). Presently this species shows extensive phenotypic variation (different forms/morphs) among locations (allopatry), which may reflect adaptation or plastic responses to diverse environment (Noakes, 2008). Several fish species of northern lakes have diverged locally to form polymorphic systems, usually related to utilization of different resources (resource polymorphism, Skúlason and Smith, 1995; Smith and Skúlason, 1996) (see additional refs. in Snorrason and Skúlason, 2004). This is seen in many salmonids (Robinson and 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 and Sægrov, 2004; Adams et al., 2007; Klemetsen, 2010). In Iceland, several lakes harbor sympatric Arctic charr morphs, some varying distinctly in a number of traits (morphology, behavior, color, life history characteristics, habitat use) while in other cases the polymorphism is more subtle (Woods et al., 2012). As with some other northern polymorphic freshwater fish species (Bernatchez et al., 2010) trait differences between Icelandic charr morphs often correlate well with major axis of ecological variation as some utilize benthic prey (benthic morphs) while others feed on zooplankton or fish (limnetic morphs). Anadromous Arctic charr are found in rivers and lakes in Iceland, that are accessible from the sea. These fish usually grow to a large size and have pointed snouts with a terminal mouth (limnetic morphology). Many landlocked populations have retained this ancestral form and are opportunist or generalist in diet, whereas other limnetic morphs are more specialized on zooplankton or other fish as prey (Skúlason et al., 1992). Although somewhat variable in morphology, benthic charr are distinct from limnetic charr, with typically dark body, blunt snout and sub-terminal mouth. 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 are also found in Icelandic lakes, with similar phenotypic characters as the dwarfs but larger adult size (Skúlason et al., 1992; Kristjánsson et al., 2011).

One of the best studied examples of polymorphic Arctic charr are the four charr morphs of Lake Thingvallavatn (Figure 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 relatively small planktivorous morph (PL, 15-25 cm adult length) that feeds on zoo-plankton, 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 benthivorous charr (LB, 25-60 cm

adult length) that both feed on bottom-dwelling invertebrates in the lava substrate habitat along the shores (Sandlund et al., 1992). Early rearing experiments showed that morphological and behavioral differences between morphs in Lake Thingvallavatn arose early (Skúlason et al., 1993, 1996), and subsequent studies of early development showed significant differences in cartilage and bone formation (Eiriksson et al., 1999; Eiriksson, 1999). More recently Kapralova et al. (2015) used geometric morphometrics to capture variation in craniofacial structures among progeny of three of the morphs (PL-, LB- and SB-charr) prior to and after hatching. Focusing on the ventral shape of the gill arches and jaws distinct differences between the morphs were found at several developmental timepoint from 200 $\tau$ s to 336 $\tau$ s, (Kapralova et al., 2015, see Materials and Methods for explanation of  $\tau$ s). Recent 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 and Jonsson, 1993; Skulason et al., 1999). Experiments on developing charr have revealed plastic responses to environmental factors like temperature and water velocity (Grünbaum et al., 2007; Jonsson and Jonsson, 2014), and several studies have shown how growth and shape of the feeding apparatus is influenced by available food types (Adams and Huntingford, 2004; Parsons et al., 2011). Studies of landlocked limnetic and benthic charr in Iceland showed that, growth and the shape of the feeding apparatus in early feeding juveniles can be significantly affected by the type of food offered (Parsons et al., 2010, 2011; Küttner et al., 2013). Furthermore, maternal effects can affect important juvenile traits, e.g. egg volume, which varies considerably within and among females, is positively correlated to yolk depletion rate and standard length at hatching and at first feeding in Aquaculture charr (Leblanc et al., 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. The experimental design can not distinguish between genetic and parental effects on embryonic gene expression, but the data cited above suggest the genetic component will outweigh the parental effects.

### Genetic variation in polymorphic and sympatric Arctic charr

The first population genetic studies found little genetic separation of the sympatric morphs in Lake Thingvallavatn (Magnusson and Ferguson, 1987; Danzmann et al., 1991; Volpe and Ferguson, 1996). Subtle difference were detected in a later study (Gíslason, 1998) and a recent microsatellite study showed significant differences between the LB-, SB- and PL-charr (Kapralova et al., 2011), with overall  $F_{ST}$ 's = 0.039. More recently, we detected  $F_{ST}$ 's larger than 0.25 between morphs in two immunological genes (Kapralova et al., 2013) and few other loci (Gudbrandsson et al., 2016), suggesting substantial genetic separation of those sympatric charr morphs. While these data suggest genetic influence on morph differentiation there is a need for studies of the underlying developmental mechanisms, e.g. how differential expression or function of multiple genes promotes these differences early in charr development. To date, few studies have addressed these issues. A qPCR study of muscle tissues 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 no difference was found in *Pax7* expression during development in Lake Thingvallavatn morphs (Sibthorpe et al., 2006).

Studies of candidate genes have been instrumental in illustrating how embryonic morphogenetic mechanisms may play a role in phenotypic diversity and speciation (Abzhanov et al., 2004; Abouheif et al., 2014) particularly when guided by QTL studies (Zimmerman et al., 2000; Shapiro et al., 2004; Palsson et al., 2005). Genome wide methods are however the new norm, either via population genome screening (Pease et al., 2016) or transcriptome analyses (Perry et al., 2012). In this context it is worth stressing that Salmonids, due to the fourth whole genome duplication of the lineage (Ss4R) 88-103 million years ago (Moghadam et al., 2011; Macqueen and Johnston, 2014; Berthelot et al., 2014; Lien et al., 2016), have quite complex genomes. The extra paralogs and chromosomal changes (Macqueen and Johnston, 2014; Nugent et al., 2017) complicate genome and transcriptome assemblies and analyses (Norman et al., 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 contrasting 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; Gudbrandsson 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 (Gudbrandsson 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 (Gudbrandsson et al., 2016) reflected the fact that the Aquaculture-charr has been affected by strong artificial selection for growth rate and delayed sexual maturation whereas the life history of SB-charr in Lake Thingvallavatn may have been shaped 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 Gudbrandsson 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, differed in expression between benthic and limnetic morphs (Ahi et al., 2013, 2014). The 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*. Furthermore, employing the candidate gene approach, preliminary analysis of the data presented in this paper linked the Aryl-hydrocarbon receptor pathway to the benthic-limnetic divergence (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 (Gudbrandsson 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, 2015). The study described here differs in several aspects: i) It focuses on an earlier window of development in higher temporal resolution (six timepoints at 100-200  $\tau$ s vs four timepoints at 141-433  $\tau$ s). These are important timepoints preceding 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  $\tau$ s when most of the viscerocranium is in place (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 head and jaw morphology between morphs prior to hatching (Kapralova et al., 2015), we anticipate 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, we also found morph specific differences in the expression of 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.

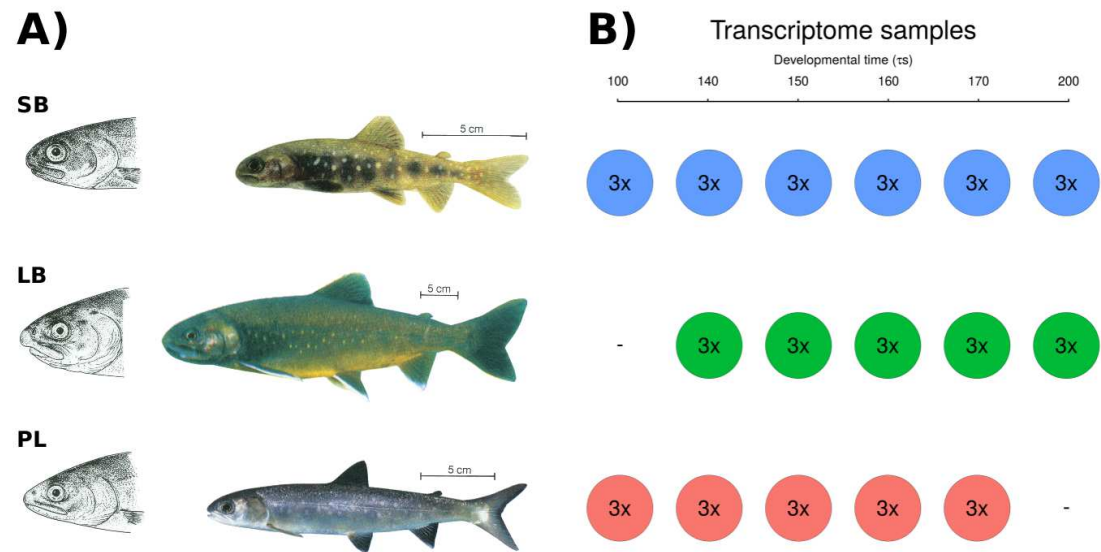
## MATERIALS AND METHODS

### 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 (Sauðárkrúkur, Iceland) as in previous studies (Ahi et al., 2014; Gudbrandsson et al., 2016). Embryos from three morphs from Lake Thingvallavatn were studied (Figure 1).

For this 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





**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), drawings by Eggert Pétursson. **B)** Embryos from pure crosses of the three morphs were sampled at six developmental timepoints prior to hatching, from  $100\tau s$  to  $200\tau s$  (circles) for RNA sequencing. Three biological replicate samples (3x) 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. Hence in total 48 samples were sequenced, composed of 144 individual charr embryos. The coloring scheme indicating morphs (blue: SB, green: LB, red: PL) will be retained throughout the manuscript.

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  $\sim 5^{\circ}\text{C}$  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  $\tau$ -somite units ( $\tau s$ ) (Gorodilov, 1996). The following formula was used to calculate the relative age (RA) at days post fertilization ( $n$ ) using the average daily temperature ( $t_i$ ).

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

Sampling of embryos for RNA extraction was performed by Holar University College Aquaculture Research Station (HUC-ARC) personnel. Embryos were sampled at designated time points, placed in RNAlater (Ambion) and frozen at  $-20^{\circ}\text{C}$ . 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, from  $100\tau s$  to  $200\tau s$  for RNA sequencing (circles in Figure 1). During this period multiple major events take place, including formation of the gill arches and jaws, the eyes migrate to the sides of the head, cartilage forms and ossification begins in certain regions of the head. Three biological replicate samples (3x) 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. For qPCR two timepoints ( $150\tau s$  and  $170\tau s$ ) were sampled for all three morphs with the same setup.

Most of the samples came from offspring of crosses created in the 2010 spawning season (SB 150-200 $\tau$ s, PL 140-170 $\tau$ s, LB 140-200 $\tau$ s). For SB- and PL-charr, eggs from 10 females were pooled and fertilized using milt from 10 males from the same morph. For LB-charr the same setup was used except that 5 females and 5 males were used. Because of technical failure with timepoint 100 $\tau$ s in PL-charr, we used samples from the 2011 spawning season (with the same setup). And due to poor quality of samples from SB-charr at timepoints 100 and 140 $\tau$ s we used material from two single parent crosses from 2011. Samples SB100A and SB100B came from the one cross but sample SB100C and all samples for timepoints 140 $\tau$ s came from the second cross. All samples used for the qPCR work were from the same set of crosses, set up in 2010.

### 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) according to the manufacturers instructions. 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) 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 minutes, 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). The quality of individual samples was assessed on BioAnalyzer before they were pooled for sequencing. The samples were sequenced on Hiseq 2000 at deCODE genetics (Reykjavik, Iceland) yielding 101 bp paired-end reads. The sequencing reads from the 48 samples were deposited into the NCBI SRA archive under BioProject identifier PRJNA391695 and with accession numbers: SRS2316381 to SRS2316428

### 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 RNA degradation and subsequent 3' bias in all samples for one timepoint (160  $\tau$ s) 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 et al., 2014](#)). RNA degradation also affected other timepoints and 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/](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 and 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 which spanned almost full length genes and had high read coverage. 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 2000 and 6000 kb 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, 2015). This quantity will be referred to as 3' coverage hereafter and used as an estimate of 3'-bias for each sample.

### Estimating expression differences between 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., 2016) to fit linear models. The full model (FM) included morph ( $M$ ) and developmental time ( $T$ ) and the interaction of morph and developmental time ( $MxT$ ). 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,  $x$  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 timepoints. 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.

$$y_{ijk} = M_i + T_j + (MxT)_{ij} + \beta_1 x_k + \beta_2 x_k^2 \quad (\text{FM})$$

$$y_{ijk} = M_i + T_j + \beta_1 x_k + \beta_2 x_k^2 \quad (\text{R1})$$

$$y_{ijk} = T_j + \beta_1 x_k + \beta_2 x_k^2 \quad (\text{R2})$$

$$y_{ijk} = M_i + \beta_1 x_k + \beta_2 x_k^2 \quad (\text{R3})$$

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).

$$y_{ijk} = M_i + T_j + (MxT)_{ij} \quad (\text{R4})$$

$$y_{ijk} = M_i + T_j \quad (\text{R5})$$

$$y_{ijk} = T_j \quad (\text{R6})$$

$$y_{ijk} = M_i \quad (\text{R7})$$

Sleuth uses false discovery rate ( $fdr$ ) to adjust for multiple testing (Benjamini and 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 parameter ( $m$ ) set to 1.1. We



performed principle component analysis (PCA) in R, with the same expression estimates, of transcripts in clusters with time-invariant expression differences between morphs (first five clusters), to get a clearer picture of morph differences.

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 length. 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. 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) and cDNA was synthesized with 1  $\mu$ g of RNA using the High Capacity dDNA RT kit (Applied Biosystems) in 20  $\mu$ l reaction volume.

Real-time PCR was performed in 96 well-PCR plates on an ABI 7500 real-time PCR System (Applied Biosystems). The normalized relative expression of genes in whole embryos was estimated from the geometric mean expression of two reference genes,  $\beta$ -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  $\tau$ s (calibration sample). Relative expression was calculated using the  $2^{-\Delta\Delta C_t}$  method of (Livak and Schmittgen, 2001). Statistical analysis was performed using the  $\Delta C_T$ -values with a two-way ANOVA with GLM function in R.

$$y_{ijk} = M_i + T_j + (MxT)_{ij} + \epsilon_{ijk}$$

The residuals were normally distributed for all data. Significant morph effect was followed up on by performing Tukey's post-hoc test, on relative expression ratios ( $\Delta C_T$ s).

## 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 timepoints was detected, except that time point 160 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 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

**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 <sup>a</sup>	Filtered	Filtered long iso <sup>a</sup>
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

<sup>a</sup> Longest isoform for each Trinity gene

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 et al., 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 (Berthelot et al., 2014).

**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

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). Blasting to salmonid transcripts recovered up to 48,916 hits (SalmoBase) in the databases (Table 3). Again, more transcripts show homology to salmon than to the 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 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 et al., 2014; Gudbrandsson et al., 2016). Our previous study of SB-charr and

**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.

<b>Proteins</b>					
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

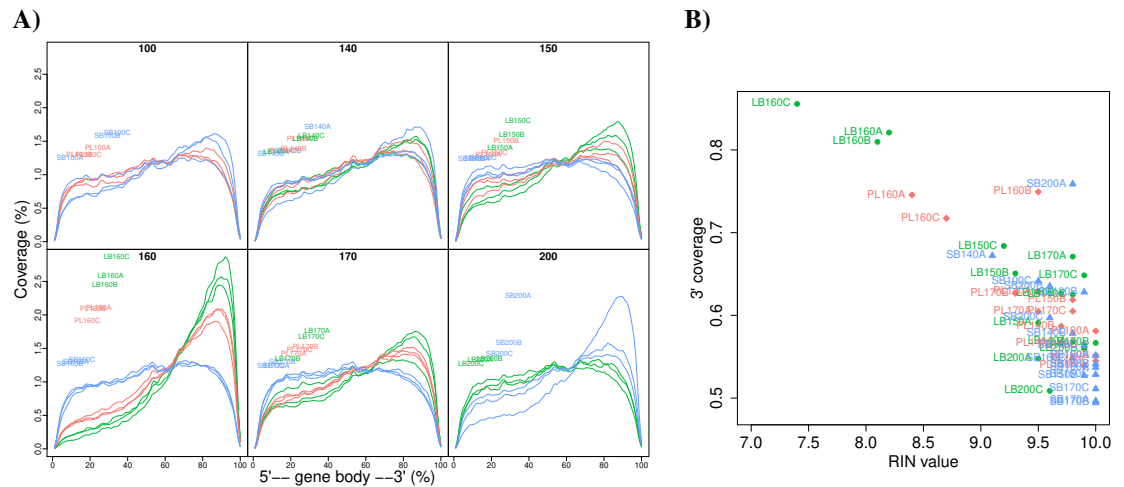
<b>mRNA Transcripts</b>				
Percent covered	Ssal NCBI	Ssal SalmoBase	Omyk	
90-100	12,418	4,189	10,335	
80-90	15,671	6,805	13,644	
70-80	18,812	9,516	16,182	
60-70	22,051	12,820	18,461	
50-60	25,748	16,854	20,656	
40-50	29,387	21,594	22,113	
30-40	33,044	27,691	23,426	
20-30	36,822	35,080	24,621	
10-20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

Icelandic Aquaculture charr did not involve transcriptome assembly (Gudbrandsson et al., 2016). However Norman et al. (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 et al. (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  $\tau$ s) led us to take a closer look at position bias. Uneven distribution of reads over transcripts can profoundly influence estimates of expression and subsequent analyses (Wu et al., 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 2000-6000 bp range that had high sequencing coverage in all samples. The patterns of read coverage over the transcripts varied greatly between samples (Figure 2A). Many samples showed 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  $r = -0.83$ ,  $p = 6.746e - 13$ ) but samples with lower 3' coverage than expected are apparent

(e.g. PL160B and SB200A, Figure 2B). This clearly demonstrates the importance of maintaining high and consistent RNA quality for RNA sequencing if poly-A pull-down is used and the importance of checking for 3' bias in RNA-seq datasets.



**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.

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 ( $MxT$ ) and fewer significant developmental time effect (Time)(Table 4, Figure S1C and S1D). Many transcripts with significant  $MxT$  interaction effect in a model without a 3'-bias term had significant Time effect after normalizing for 3'-bias (Figure S1A). Thus we concluded that involving 3'-bias in the linear models decreased the number of transcripts with potentially false positive  $MxT$  interaction effect.

Degradation of RNA is an issue for RNA-sequencing, particularly if poly-A pull-down is used (Sigurgeirsson et al., 2014). Methods for estimating variation in coverage along transcripts, rely on full length sequences (Wu et al., 2011). Correction for 3'-bias by restricting analyses to 200 bp at the 3' end of transcripts (Sigurgeirsson et al., 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 and 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 is therefore the method of choice for further studies on salmonid transcriptomes. The approach to use 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 (Figure S1B). We conclude that more than thousand genes are differentially expressed between developing

**Table 4.** Number of differently 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)).

<b>With 3'-bias correction</b>				
<i>fdr</i>	3'-bias	<i>MxT</i>	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

<b>Without 3'-bias correction</b>			
<i>fdr</i>	<i>MxT</i>	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

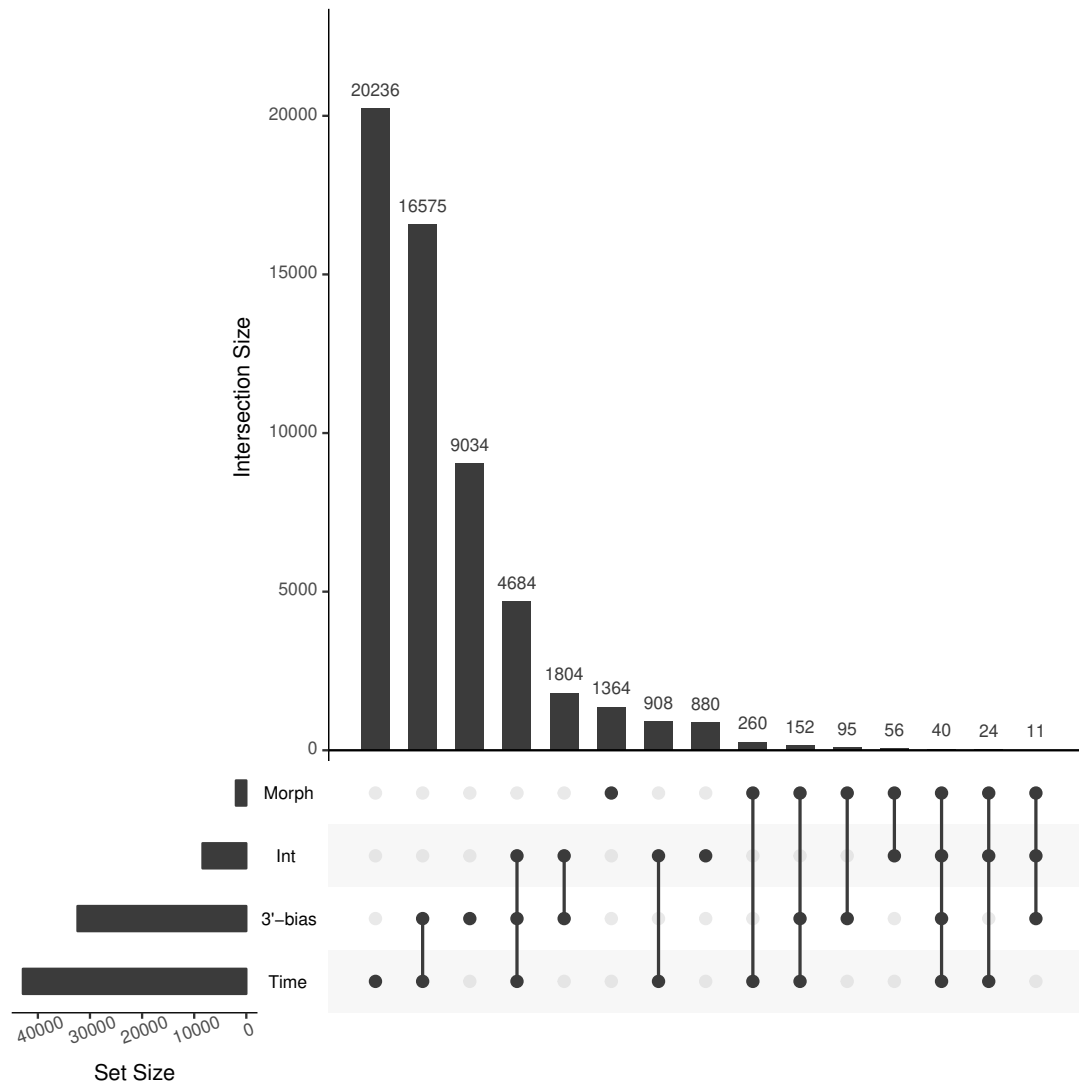
embryos of the three sympatric morphs. Of the 2,002 transcripts with morph differences (at *fdr* < 0.01), 1370 were only significant for Morph and no other terms. Further 632 had other terms also significant (some even all of them), but only 131 transcripts were significant for both Morph and Morph by Time (*MxT*) interaction (Figure 3). A considerably larger number of transcripts (8,407) had a significant *MxT* term, with the majority (4684) also having significant Time and 3'-bias effects. Because the 3'-bias estimation 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 1320 transcripts each (Figure 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 *MxT* and Time effects (Table 5). The data suggest separation between the two benthic charr and PL-charr at the gene expression level, as three of the five clusters with Morph effects show persistent expression difference between the two benthic (SB and LB) and PL-charr. These three clusters (B-D) contain 1131 transcript, while the SB specific cluster A has 353 transcripts and the LB specific cluster E 499. To analyze this further we performed PCA on the transcripts from these five clusters. This also showed closer transcriptional resemblance of the two benthic morphs (along PC1 which explained 26.8% of the variance, see Figure 5). However, the two benthic morphs are separated by PC2 (explains 17.9% of the variance), on which the LB and PL-charr seem more similar. With positive association of transcriptional divergence and genetic divergence (Whitehead and Crawford, 2006), this might suggest that the two benthic morphs are more closely related, and PL-charr more distantly, consistent with one population genetic study (Volpe and Ferguson, 1996) but incongruent with others (Gíslason, 1998; Kapralova et al., 2011). Preliminary analyses of genetic variation in this transcriptome clearly separates the morphs, and supports the close relation of the benthic morphs (Johannes Gudbrandson *et al.* unpublished data).

**Table 5.** Number of transcripts differently expressed (*fdr* < 0.01) for Morph, Time or Morph by Time interaction (*MxT*) for co-expression clusters A-E and F-P.

Effect	Clusters	
	A-E	F-P
Morph	1768	234
Time	307	5761
<i>MxT</i>	296	8111

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 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

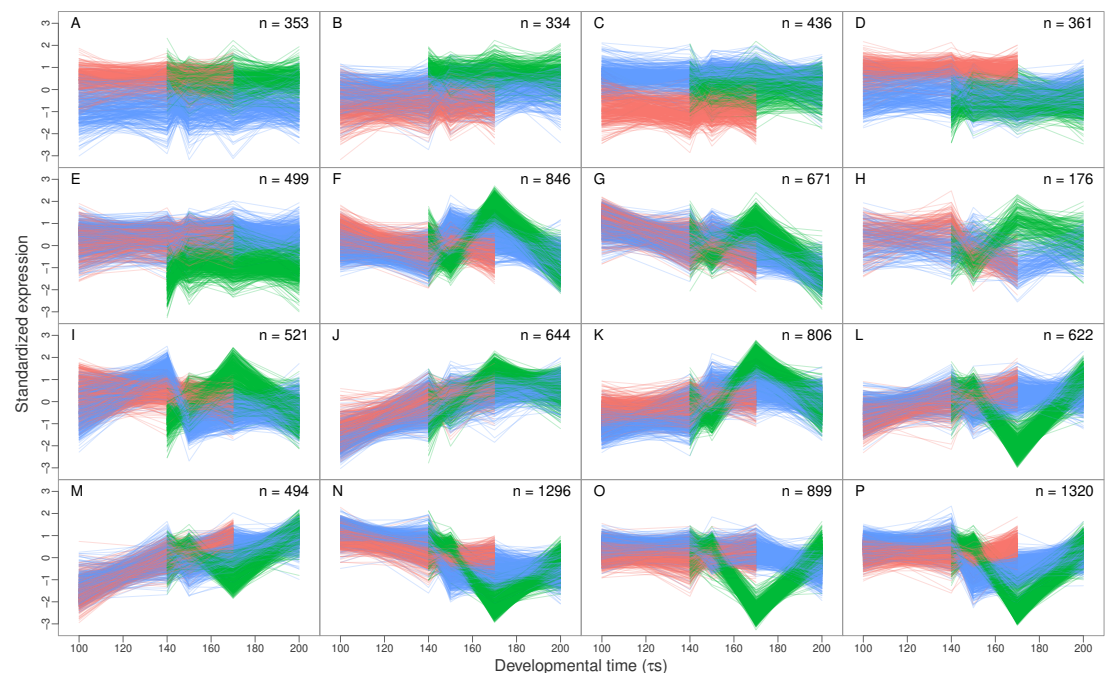




**Figure 3.** The number of transcripts differently expressed according to the linear models on developmental timepoint (Time), Morph, 3'-bias and interaction of Morph and Time (int). The set size barplot (sideways, lower left) shows cumulated number of transcripts for each of the four main factors, while the intersection size barplot (upper) 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.

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 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



**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  $100\tau s$  to  $200\tau s$ ). 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 contain almost exclusively transcripts with  $M \times T$  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.

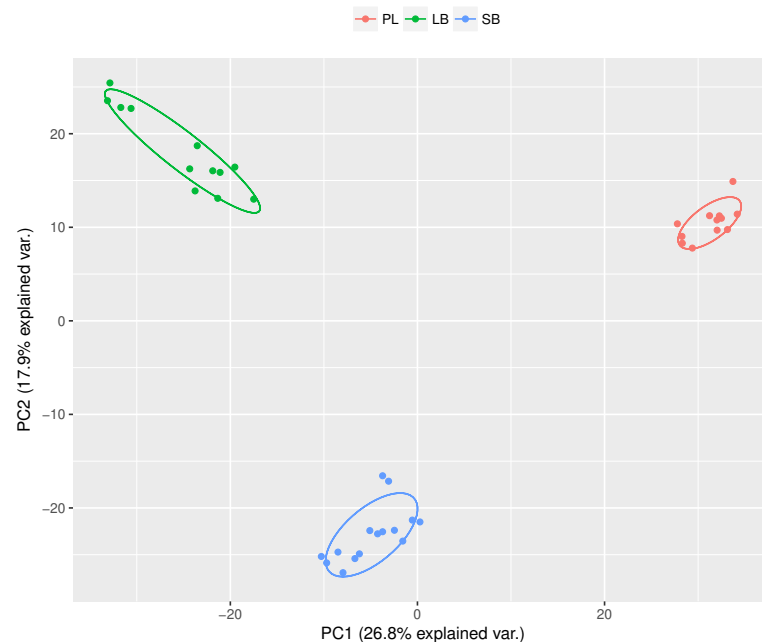
three or more clusters, for example; GO:1903047, mitotic cell cycle process (clusters F, G, N, O and P), GO:0022613, ribonucleoprotein complex biogenesis (clusters F, G and N) and GO:0007507, heart development (clusters M, O and P). The diversity of GO categories to us suggests that multiple systems are differentially expressed during early development in these three charr morphs.

Our published data (Gudbrandsson 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) and in a cluster were SB does not especially 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 the embryos well prior to hatching.

### 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 7 candidate genes at two developmental timepoints (150 and 170  $\tau s$ ) with different expression in the benthic morphs (LB- and SB-charr) and limnetic morph (PL-charr) in the transcriptome. 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*



**Figure 5.** Principle component analysis (PCA) of transcripts in the first five expression clusters (A-E, 1983 transcripts). Standardized expression normalized by 3'-coverage was used as input. 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. 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](#)).

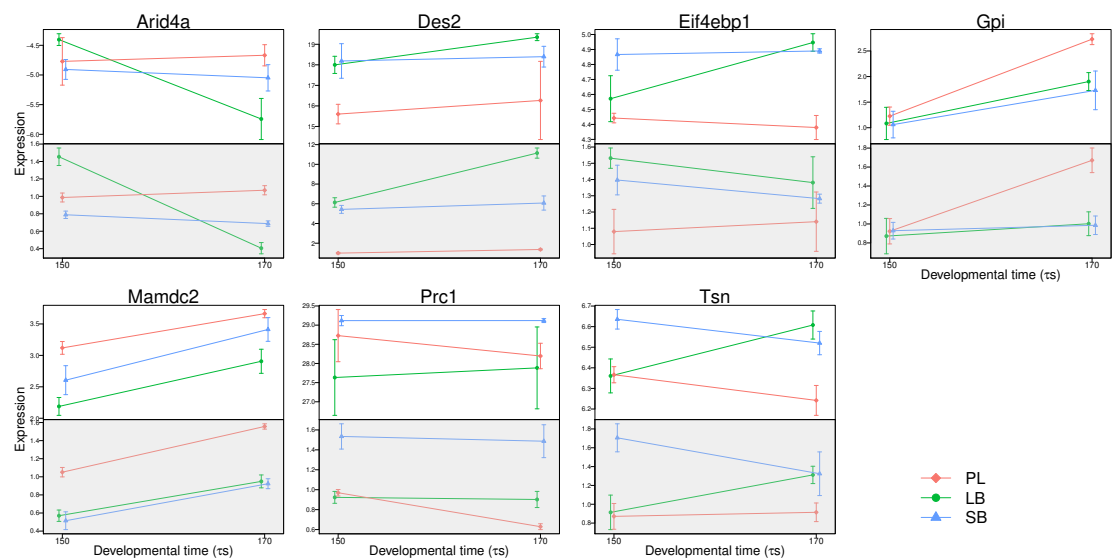
(*Eif4ebp1*) showed a suggestive limnetic and benthic separation in the qPCR (only formally significant at 170  $\tau$ s) (Figure 6). Notably, *Arid4a* showed the same Morph by Time interaction in both the transcriptome and qPCR. In sum, there was general consistence between the transcriptome results and the qPCR verification tests on whole embryos.

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 et al., 2003](#)), *Gpi* a glycolytic enzyme differentially expressed in zebrafish development ([Lin et al., 2009](#)), *Des2* is involved in sphingolipid synthesis ([Omae et al., 2004](#)) whereas the function of *Mamdc2* is poorly characterized.

In the light of prior data we will 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 ([Johannes Gudbrandson et al.](#) unpublished data) indicate differences in allele frequency of variants in *Eif4ebp1* between SB- and PL-charr. These facts do not prove the involvement of *Eif4ebp1* in morph differentiation, but calls 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 we screened for candidate genes involved in craniofacial development, utilizing our

published data (Gudbrandsson et al., 2016) and this dataset. There 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, 2015). This revealed multiple genes with clear benthic - limnetic separation in expression, and pointed to *Ets2* and transcription factors in the glucocorticoid and Aryl hydrocarbon pathways as potential modulators of benthic-limnetic diversity. These results and the current data suggest that multiple developmental systems are diverging between these three sympatric morphs, most 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, Johannes Gudbrandsson *et al.*, unpublished). Alternative approaches could be whole genome scans of divergence (Jones et al., 2012; Halldórsdóttir and Árnason, 2015), QTL or association studies (Zimmerman et al., 2000; Palsson et al., 2005) of specific ecological traits to identify putative causative factors and variants that differentiate these sympatric morphs.



**Figure 6.** Expression of 7 genes that differed between benthic (SB and LB) and limnetic (PL) charr in the transcriptome. Each panel shows the expression of one gene at developmental timepoints 150 and 170  $\tau$ s, in the transcriptome (above) and measured with qPCR (below). Colors indicate morph (blue: SB, green: LB, red: PL). The upper panels show expression is 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 ( $\tau$ s). Error bars represent 2 standard errors of the mean calculated from three biological replicates each made from a homogenate of three whole embryos.

## 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; 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; Skulason et al., 1999; Leblanc et al., 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 and Ferguson, 1996; Gíslason, 1998; Kapralova et al., 2011). Though parental effects can not be excluded, we postulate that the majority of the expression differences between morphs stem from genetic differences. The observed pattern at the expression level, that all morphs are separate and that the benthic morphs are more similar (this data) is in concordance with genetic differences estimated from this

transcriptome (Johannes Gudbrandsson *et al.* unpublished data). 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, 2015) indicate that during development, upstream regulatory mechanisms are acting differently in these morphs. Thus differential expression of regulators such as *Tsn*, *Ahr2* (Ahi *et al.*, 2015) or *Eif4ebp1* (Macqueen *et al.*, 2011), 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.

Although the genes identified here and in our previous studies (Ahi *et al.*, 2014, 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 driving ecological divergence is not straightforward (Santure *et al.*, 2015), in part because of the pleiotropic nature of metabolic, homeostatic and developmental systems (Paaby and Rockman, 2013). One intriguing question is whether the heritable expression differences between morphs is due to one gene, few genes or many QTLs? Our combined data (Ahi *et al.*, 2014, 2015; Gudbrandsson *et al.*, 2016, present study) argues against a major gene 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. 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 to perform 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 and 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. Magnusdottir, Gudbjorg Th. Orlygsdottir, Steinunn Snorraddottir and Olafur Th. Magnusson 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 Kalina H. Kapralova for help with RNA extractions, quality controls and discussions, Pall Melsted for advice on the transcriptome assembly, Benjamin Sigurgeirsson for advice on 3'-bias detection and all members of the charr group at the University for support, comments and suggestions. We are grateful to all the fish for their parts. The project was funded by The Icelandic Center for Research (grant number: 100204011) to SSS, AP, ZOJ and BKK, The University of Iceland Doctoral Fund to JG and University of Iceland research fund to AP, SSS and ZOJ.

## AUTHOR CONTRIBUTIONS

Conceived and designed the study: JG, AP, ZOJ, SSS, SRF  
Sampling: BKK, SSS, SRF, ZOJ, JG  
RNA extraction, library preparation and sequencing: VHM, SRF, ZOJ, EPA  
Transcriptome assembly, annotation and analyses: JG  
qPCR confirmation EPA  
Writing: JG, AP, SSS, SRF, ZOJ  
Analyses: JG, AP  
Gathered the data: ZOJ, SRF, EPA



## REFERENCES

- Abernathy, J. and Overturf, K. (2016). Comparison of Ribosomal RNA Removal Methods for Transcriptome Sequencing Workflows in Teleost Fish. *Animal Biotechnology*, 27(1):60–65. **11**
- Abouheif, E., Favé, M.-J., Ibarrarán-Viniegra, A. S., Lesoway, M. P., Rafiqi, A. M., and Rajakumar, R. (2014). Eco-Evo-Devo: The Time Has Come. In Landry, C. R. and Aubin-Horth, N., editors, *Ecological Genomics - Ecology and the Evolution of Genes and Genomes*, chapter 6, pages 107–125. Springer Netherlands, Dordrecht. **3**
- Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R., and Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, 442(7102):563–567. **2**
- Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., and Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, 305(5689):1462–1465. **2, 3**
- Adams, C. E., Fraser, D., Wilson, A. J., Alexander, G., Ferguson, M. M., and Skulason, S. (2007). Patterns of phenotypic and genetic variability show hidden diversity in Scottish Arctic charr. *Ecology of Freshwater Fish*, 16(1):78–86. **2**
- Adams, C. E. and Huntingford, F. A. (2004). Indipient speciation driven by phenotypic plasticity? Evidence from sympatric populations of Arctic charr. *Biological Journal of the Linnean Society*, 81(4):611–618. **3**
- Ahi, E. P., Gudbrandsson, J., Kapralova, K. H., Franzdóttir, S. R., Snorrason, S. S., Maier, V. H., and Jónsson, Z. O. (2013). Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr. *PloS ONE*, 8(6):e66389. **4, 8**
- Ahi, E. P., Kapralova, K. H., Pálsson, A., Maier, V. H., Gudbrandsson, J., Snorrason, S. S., Jónsson, Z. O., and Franzdóttir, S. R. (2014). Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr. *EvoDevo*, 5(1):40. **4, 16, 17**
- Ahi, E. P., Steinhäuser, S. S., Pálsson, A., Franzdóttir, S. R., Snorrason, S. S., Maier, V. H., and Jónsson, Z. O. (2015). Differential expression of the aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr. *EvoDevo*, 6(1):27. **4, 8, 16, 17**
- Alexandrou, M. A., Swartz, B. A., Matzke, N. J., and Oakley, T. H. (2013). Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution*, 69(3):514–523. **9**
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300. **7**
- Bernatchez, L., Renaut, S., Whiteley, A. R., Derome, N., Jeukens, J., Landry, L., Lu, G., Nolte, A. W., Ostbye, K., Rogers, S. M., and St-Cyr, J. (2010). On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1547):1783–800. **2**
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G. H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Roest Crolius, H., and Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications*, 5:3657. **3, 6, 9**
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527. **6, 7**
- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., and Wittwer, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, 55(4):611–22. **8**
- Carlson, M. (2015). GO.db: A set of annotation maps describing the entire Gene Ontology. **8**
- Carroll, S. B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell*, 134(1):25–36. **2**
- Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., and Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome research*, 24(5):797–808. **17**
- Crête-Lafrenière, A., Weir, L. K., and Bernatchez, L. (2012). Framing the Salmonidae Family Phylogenetic

- Portrait: A More Complete Picture from Increased Taxon Sampling. *PLoS ONE*, 7(10):e46662. 9
- Danzmann, R. G., Ferguson, M. M., Skulason, S., Snorrason, S. S., and Noakes, D. L. G. (1991). Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn, Iceland. *Journal of Fish Biology*, 39(5):649–659. 3
- Danzmann, R. G., Kocmarek, A. L., Norman, J. D., Rexroad, C. E., and Palti, Y. (2016). Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics*, 17(1):60. 1
- Di Génova, A., Aravena, A., Zapata, L., González, M., Maass, A., and Iturra, P. (2011). SalmonDB: a bioinformatics resource for *Salmo salar* and *Oncorhynchus mykiss*. *Database : the journal of biological databases and curation*, 2011:bar050. 6
- Dowling, R. J. O., Topisirovic, I., Alain, T., Bidinosti, M., Fonseca, B. D., Petroulakis, E., Wang, X., Larsson, O., Selvaraj, A., Liu, Y., Kozma, S. C., Thomas, G., and Sonenberg, N. (2010). mTORC1-Mediated Cell Proliferation, But Not Cell Growth, Controlled by the 4E-BPs. *Science*, 328(5982):1172–1176. 15
- Eiriksson, G. M. (1999). *Heterochrony in bone development and growth in two morphs of Arctic charr (Salvelinus alpinus) from Thingvallavatn, Iceland*. Master's thesis, University of Iceland. 3
- Eiriksson, G. M., Skulason, S., and Snorrason, S. S. (1999). Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn, Iceland. *Journal of Fish Biology*, 55(sa):175–185. 3
- Flint, J. and Mackay, T. F. C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome research*, 19(5):723–33. 17
- Futschik, M. (2015). Mfuzz: Soft clustering of time series gene expression data. 7
- Garfield, D. A., Runcie, D. E., Babbitt, C. C., Haygood, R., Nielsen, W. J., and Wray, G. A. (2013). The Impact of Gene Expression Variation on the Robustness and Evolvability of a Developmental Gene Regulatory Network. *PLoS Biology*, 11(10):e1001696. 2
- Giger, T., Excoffier, L., Day, P. J. R., Champigneulle, A., Hansen, M. M., Powell, R., and Largiadèr, C. R. (2006). Life history shapes gene expression in salmonids. *Current Biology*, 16(8). 1
- Gíslason, D. (1998). *Genetic and morphological variation in polymorphic arctic charr Salvelinus alpinus, from Icelandic lakes*. Master's thesis, The University of Guelph. 3, 12, 16
- Gorodilov, Y. N. (1996). Description of the early ontogeny of the Atlantic salmon, *Salmo salar*, with a novel system of interval (state) identification. *Environmental Biology of Fishes*, 47(2):109–127. 5
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7):644–52. 6, 7, 8
- Grant, B. and Grant, P. R. (2008). Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1505):2821–2829. 2
- Grant, P. R. (1999). *Ecology and Evolution of Darwin's Finches*. Princeton University Press, New Jersey, second edition. 2
- Grünbaum, T., Cloutier, R., Mabee, P. M., and Le François, N. R. (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. 3
- Gudbrandsson, J., Ahi, E. P., Franzdóttir, S. R., Kapralova, K. H., Kristjánsson, B. K., Steinháuser, S. S., Maier, V. H., Johannesson, I. M., Snorrason, S. S., Jonsson, Z. O., and Palsson, A. (2016). The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations]. *F1000Research*, 4(136). 3, 4, 9, 10, 14, 16, 17
- Haas, B. J. (2015). Trinotate: Transcriptome Functional Annotation and Analysis. 6
- Halldórsdóttir, K. and Árnason, E. (2015). Whole-genome sequencing uncovers cryptic and hybrid species among Atlantic and Pacific cod-fis. *bioRxiv*. 16
- Hindar, K. and Jonsson, B. (1993). Ecological polymorphism in Arctic charr. *Biological Journal of the Linnean Society*, 48(1):63–74. 3
- Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature*

- Genetics*, 29(4):389–395. **1**
- Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., and Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502(7469):93–95. **17**
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., Brady, S. D., Zhang, H., Pollen, A. A., Howes, T., Amemiya, C., Baldwin, J., Bloom, T., Jaffe, D. B., Nicol, R., Wilkinson, J., Lander, E. S., Di Palma, F., Lindblad-Toh, K., and Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392):55–61. **16**
- Jonsson, B. and Jonsson, N. (2014). Early environment influences later performance in fishes. *Journal of Fish Biology*, 85(2):151–188. **3**
- Jonsson, B., Skúlason, S., Snorrason, S. S., Sandlund, O. T., Malmquist, H. J., Jónasson, P. M., Cydemo, R., and Lindem, T. (1988). Life History Variation of Polymorphic Arctic Charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland. *Canadian Journal of Fisheries and Aquatic Sciences*, 45(9):1537–1547. **4**
- Kapralova, K. H., Franzdóttir, S. R., Jónsson, H., Snorrason, S. S., and Jónsson, Z. O. (2014). Patterns of MiRNA Expression in Arctic Charr Development. *PLoS ONE*, 9(8):e106084. **3, 4**
- Kapralova, K. H., Gudbrandsson, J., Reynisdóttir, S., Santos, C. B., Baltanás, V. C., Maier, V. H., Snorrason, S. S., and Pálsson, A. (2013). Differentiation at the MHCII $\alpha$  and Cath2 Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn. *PLoS ONE*, 8(7):e69402. **3**
- Kapralova, K. H., Jónsson, Z. O., Pálsson, A., Franzdóttir, S. R., Le Deuf, S., Kristjánsson, B. K., and Snorrason, S. S. (2015). Bones in motion: Ontogeny of craniofacial development in sympatric Arctic charr morphs. *Developmental Dynamics, under revision*. **3, 4**
- Kapralova, K. H., Morrissey, M. B., Kristjánsson, B. K., Olafsdóttir, G. Á., Snorrason, S. S., and Ferguson, M. M. (2011). Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity*, 106(3):472–87. **2, 3, 12, 16**
- Klemetsen, A. (2010). The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes. *Freshwater Reviews*, 3(1):49–74. **2**
- Klemetsen, A. (2013). The Most Variable Vertebrate on Earth. *Journal of Ichthyology*, 53(10):781–791. **2**
- Koop, B. F., Von Schalburg, K. R., Leong, J., Walker, N., Lieph, R., Cooper, G. A., Robb, A., Beetz-Sargent, M., Holt, R. A., Moore, R., Brahmabhatt, S., Rosner, J., Rexroad, C. E., McGowan, C. R., and Davidson, W. S. (2008). A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics*, 9(1):545. **9**
- Kopp, A., Duncan, I., and Carroll, S. B. (2000). Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature*, 408(6812):553–559. **2**
- Kristjánsson, B. K., Malmquist, H. J., Ingimarsson, F., Antonsson, T., Snorrason, S. S., and 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. **2**
- Kristjánsson, B. K., Skúlason, S., Snorrason, S. S., and Noakes, D. L. (2012). Fine-scale parallel patterns in diversity of small benthic Arctic charr (*Salvelinus alpinus*) in relation to the ecology of lava/groundwater habitats. *Ecology and evolution*, 2(6):1099–112. **2**
- Krueger, F. (2012). Trim Galore! **6**
- Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P., and Kronforst, M. R. (2014). Doublesex Is a Mimicry Supergene. *Nature*, 507(7491):229–232. **17**
- Küttner, E., Parsons, K. J., Robinson, B. W., Skúlason, S., Danzmann, R. G., and Ferguson, M. M. (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. **3**
- Lai, A., Marcellus, R. C., Corbeil, H. B., and Branton, P. E. (1999). RBP1 induces growth arrest by repression of E2F-dependent transcription. *Oncogene*, 18(12):2091–2100. **15**
- Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A. A., Dalziel, A. C., Chebib, J., Bernatchez, L., Dion-Cote, A. M., Normandeau, E., Gagnaire, P.-A. A., Dalziel, A. C., Chebib, J., Bernatchez, L., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A. A., Dalziel, A. C., Chebib, J., and 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. **17**

- Leblanc, C. A.-L., Kristjánsson, B. K., and 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. **3, 16**
- Li, C., Shridhar, K., and 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. **15**
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K. A., Olav Vik, J., Vigeland, M. D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D. R., Yorke, J. A., Nederbragt, A. J., Tooming-Klunderud, A., Jakobsen, K. S., Jiang, X., Fan, D., Hu, Y., Liberles, D. A., Vidal, R., Iturra, P., Jones, S. J. M., Jonassen, I., Maass, A., Omholt, S. W., and Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602):200–205. **3**
- Lin, W.-W., Chen, L.-H., Chen, M.-C., and Kao, H.-W. (2009). Differential expression of zebrafish *gpiA* and *gpiB* during development. *Gene expression patterns : GEP*, 9(4):238–45. **15**
- Linnen, C. R., Kingsley, E. P., Jensen, J. D., and Hoekstra, H. E. (2009). On the Origin and Spread of an Adaptive Allele in Deer Mice. *Science*, 325(5944):1095–1098. **2**
- Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L. N., Grishin, N. V., and Liu, Q. (2009). C3PO, an Endoribonuclease That Promotes RNAi by Facilitating RISC Activation. *Science*, 325(5941):750–753. **15**
- Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(-Delta Delta C(T)) Method. *Methods*, 25(4):402–8. **8**
- Ludwig, M. Z., Palsson, A., Alekseeva, E., Bergman, C. M., Nathan, J., and Kreitman, M. (2005). Functional Evolution of a cis-Regulatory Module. *PLoS Biology*, 3(4):e93. **2**
- Macqueen, D. J. and Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778):20132881. **3**
- Macqueen, D. J., Kristjánsson, B. K., Paxton, C. G. M., Vieira, V. L. A., and Johnston, I. A. (2011). The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. *Molecular ecology*, 20(15):3167–84. **3, 15, 17**
- Magnusson, K. P. and Ferguson, M. M. (1987). Genetic analysis of four sympatric morphs of Arctic charr, *Salvelinus alpinus*, from Thingvallavatn, Iceland. *Environmental Biology of Fishes*, 20(1):67–73. **3**
- Malmquist, H. J., Snorrason, S. S., Skúlason, S., Jonsson, B., Sandlund, O. T., and Jonsson, P. M. (1992). Diet Differentiation in Polymorphic Arctic Charr in Thingvallavatn, Iceland. *The Journal of Animal Ecology*, 61(1):21. **2**
- Moghadam, H. K., Ferguson, M. M., and Danzmann, R. G. (2011). Whole genome duplication: Challenges and considerations associated with sequence orthology assignment in Salmoninae. *Journal of Fish Biology*, 79(3):561–574. **3**
- Muir, A. M., Hansen, M. J., Bronte, C. R., and Krueger, C. C. (2016). If Arctic charr *Salvelinus alpinus* is ‘the most diverse vertebrate’, what is the lake charr *Salvelinus namaycush*? *Fish and Fisheries*, 17(4):1194–1207. **2**
- Noakes, D. L. G. (2008). Charr truth: sympatric differentiation in *Salvelinus* species. *Environmental Biology of Fishes*, 83(1):7–15. **2**
- 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. **3**
- Norman, J. D., Ferguson, M. M., and Danzmann, R. G. (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):123–37. **3, 9, 10**
- Nugent, C. M., Easton, A. A., Norman, J. D., Ferguson, M. M., and Danzmann, R. G. (2017). A SNP Based Linkage Map of the Arctic Charr (*Salvelinus alpinus*) Genome Provides Insights into the Diploidization Process After Whole Genome Duplication. *G3: Genes|Genomes|Genetics*, 7(2):543–556. **3**
- Oleksiak, M. F., Churchill, G. A., and Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nature Genetics*, 32(2):261–266. **1**



- Omae, F., Miyazaki, M., Enomoto, A., and Suzuki, A. (2004). Identification of an essential sequence for dihydroceramide C-4 hydroxylase activity of mouse DES2. *FEBS letters*, 576(1-2):63–7. 15
- Paaby, A. B. and Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in Genetics*, 29(2):66–73. 17
- Palsson, A., Dodgson, J., Dworkin, I., and Gibson, G. (2005). No Title. *BMC Genetics*, 6(1):44. 3, 16
- Parsons, K. J., Sheets, H. D., Skúlason, S., and Ferguson, M. M. (2011). Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (*Salvelinus alpinus*). *Journal of Evolutionary Biology*, 24(8):1640–1652. 3
- Parsons, K. J., Skúlason, S., and Ferguson, M. (2010). Morphological variation over ontogeny and environments in resource polymorphic arctic charr (*Salvelinus alpinus*). *Evolution and Development*, 12(3):246–257. 3
- Pease, J. B., Haak, D. C., Hahn, M. W., and Moyle, L. C. (2016). Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLOS Biology*, 14(2):e1002379. 3
- Perry, G. H., Melsted, P., Marioni, J. C., Wang, Y., Bainer, R., Pickrell, J. K., Michelini, K., Zehr, S., Yoder, A. D., Stephens, M., Pritchard, J. K., and Gilad, Y. (2012). Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research*, 22(4):602–610. 3
- Pimentel, H. J., Bray, N., Puente, S., Melsted, P., and Pachter, L. (2016). Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRxiv*. 7
- R Core Team (2015). R: A Language and Environment for Statistical Computing. 7
- Roberts, A. and Pachter, L. (2012). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods*, 10(1):71–73. 7
- Robertson, F. M., Gundappa, M. K., Grammes, F., Hvidsten, T. R., Redmond, A. K., Lien, S., Martin, S. A. M., Holland, P. W. H., Sandve, S. R., and Macqueen, D. J. (2017). Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology*, 18(1):111. 10
- Robinson, B. W. and Parsons, K. J. (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. 2
- Romero, I. G., Ruvinsky, I., and Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13(7):505–516. 1
- Sæmundsson, B. (1904). Fiskirannsóknir 1902. *Andvari*, 29(1):79–119. 16
- Samy, J. K. A., Mulugeta, T. D., Nome, T., Sandve, S. R., Grammes, F., Kent, M. P., Lien, S., and Våge, D. I. (2017). SalmoBase: an integrated molecular data resource for Salmonid species. *BMC Genomics*, 18(1):482. 6
- Sandlund, O. T., Gunnarsson, K., Jónasson, P. M., Jonsson, B., Lindem, T., Magnússon, K. P., Malmquist, H. J., Sigurjónsdóttir, H., Skúlason, S., Snorrason, S. S., Jonasson, P. M., Magnusson, K. P., Sigurjónsdóttir, H., and Skulason, S. (1992). The arctic charr *Salvelinus alpinus* in Thingvallavatn. *Oikos*, 64(1/2):305–351. 2, 3, 5
- Santure, A. W., Poissant, J., De Cauwer, I., van Oers, K., Robinson, M. R., Quinn, J. L., Groenen, M. A. M., Visser, M. E., Sheldon, B. C., and Slate, J. (2015). Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology*, 24(24):6148–6162. 17
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology*, 7(3). 6
- Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D., and Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, 428(6984):717–723. 3, 17
- Sibthorpe, D., Sturlaugsdóttir, R., Kristjánsson, B. K., Thorarensen, H., Skúlason, S., and Johnston, I. A. (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. 3
- Sigurgeirsson, B., Emanuelsson, O., Lundeberg, J., Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Imbeaud, S., Graudens, E., Boulanger, V., Barlet, X., Zaborski, P., Opitz, L., Salinas-Riester, G., Grade, M., Jung, K., Jo, P., Mortazavi, A., Williams, B., McCue, K., Schaeffer, L., Wold, B., Wu, Z., Wang, X., Zhang, X., Li, W., Jiang, T., Roberts, A., Trapnell, C., Donaghey, J., Rinn, J., Pachter, L.,



- Wan, L., Yan, X., Chen, T., Sun, F., Brisco, M., Morley, A., Lundin, S., Stranneheim, H., Pettersson, E., Klevebring, D., Lundeberg, J., Martin, M., Trapnell, C., Pachter, L., Salzberg, S., Wang, L., Wang, S., Li, W., Bourgon, R., Gentleman, R., Huber, W., Anders, S., Huber, W., Trapnell, C., Williams, B., Perlea, G., Mortazavi, A., Kwan, G., Solnestam, B., Stranneheim, H., Hallman, J., Kaller, M., Lundberg, E., Yang, L., Duff, M., Graveley, B., Carmichael, G., Chen, L., Djebali, S., Davis, C., Merkel, A., Dobin, A., Lassmann, T., Morlan, J., Qu, K., Sinicropi, D., Shepard, P., Choi, E., Lu, J., Flanagan, L., Hertel, K., Beck, A., Weng, Z., Witten, D., Zhu, S., Foley, J., Jan, C., Friedman, R., Ruby, J., Bartel, D., Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., and Fu, X. (2014). Sequencing Degraded RNA Addressed by 3' Tag Counting. *PLoS ONE*, 9(3):e91851. **6, 11**
- Skúlason, S., Antonsson, T., Gudbergsson, G., Malmquist, H. J., and Snorrason, S. S. (1992). Variability in Icelandic Arctic charr. **2**
- Skúlason, S. and Smith, T. B. (1995). Resource polymorphisms in vertebrates. *Trends in Ecology & Evolution*, 10(9):366–370. **2**
- Skúlason, S., Snorrason, S., Noakes, D., and Ferguson, M. (1996). Genetic basis of life history variations among sympatric morphs of Arctic char *Salvelinus alpinus*. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(8):1807–1813. **3, 16**
- Skúlason, S., Snorrason, S., Ota, D., and Noakes, D. (1993). Genetically based differences in foraging behaviour among sympatric morphs of arctic charr (Pisces: Salmonidae). *Animal Behaviour*, 45(6):1179–1192. **3**
- Skulason, S., Snorrason, S., Jonsson, B., and Snorrason, S. S. (1999). Sympatric morphs, populations and speciation in freshwater fish with emphasis on Arctic charr. In Magurran, A. E. and May, R. M., editors, *Evolution of Biological Diversity*, chapter 4, pages 70–92. Oxford University Press. **3, 16**
- Smith, T. B. and Skúlason, S. (1996). Evolutionary Significance of Resource Polymorphisms in Fishes, Amphibians, and Birds. *Annual Review of Ecology and Systematics*, 27:111–133. **2**
- Snorrason, S. S. and Skúlason, S. (2004). Adaptive Speciation in Northern Freshwater Fishes. In Dieckmann, U., Doebeli, M., Metz, J. A. J., and Tautz, D., editors, *Adaptive speciation*, chapter 10, pages 210–228. Cambridge University Press, Cambridge. **2**
- Snorrason, S. S., Skúlason, S., Jonsson, B., Malmquist, H. J., Jónasson, P. M., Sandlund, O. T., and 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. **16**
- Snorrason, S. S., Skúlason, S., Sandlund, O. T., Malmquist, H. J., Jonsson, B., and Jonasson, P. M. (1989). Shape polymorphism in arctic charr, *Salvelinus alpinus*. *Physiology and Ecology Japan*, 1:393–404. **2, 16**
- Stern, D. L. and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution*, 62(9):2155–77. **2**
- Telnes, T. and Sægrov, H. (2004). Reproductive strategies in two sympatric morphotypes of Arctic charr in Kalandvatnet, west Norway. *Journal of Fish Biology*, 65(2):574–579. **2**
- True, J. R. and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution & development*, 3(2):109–19. **1**
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40(15). **8**
- Volpe, J. P. and Ferguson, M. M. (1996). Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. *Molecular ecology*, 5(6):763–72. **3, 12, 16**
- Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. **15**
- Wang, X., Beugnet, A., Murakami, M., Yamanaka, S., and Proud, C. G. (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. **15**
- Whitehead, A. and Crawford, D. L. (2006). Variation within and among species in gene expression: raw material for evolution. *Molecular ecology*, 15(5):1197–211. **12**
- Wilson, A. J., Gíslason, D., Skúlason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G., and Ferguson, M. M. (2004). Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular ecology*, 13(5):1129–42. **2**
- Woods, P. J., Skúlason, S., Snorrason, S. S., Kristjánsson, B. K., Malmquist, H. J., and Quinn, T. P. (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. [2](#)
- Wootton, R. J. (1984). *A Functional Biology of Sticklebacks*. Springer US, Boston, MA. [2](#)
- Wu, Z., Wang, X., and Zhang, X. (2011). Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics*, 27(4):502–508. [10](#), [11](#)
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*, 11(2):R14. [8](#)
- Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 26(7):976–978. [8](#)
- Zimmerman, E., Palsson, A., and Gibson, G. (2000). Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics*, 155(2):671–83. [3](#), [16](#)

## SUPPLEMENT

**Table S1.** Sequencing effort, quality trimming, mapping and estimated insert size for each sample.

Morph	Relative age	Rep	Index	Lane	Nr of reads	After trim <sup>a</sup>	KaAP <sup>b</sup>	KaFL <sup>c</sup>	RIN
LB	140	A	CCGTCC	L004	27,701,961	0.997	0.777	205.77	9.8
LB	140	B	AGTTCC	L003	14,573,957	0.996	0.795	186.08	9.7
LB	140	C	ATGTCA	L002	23,771,531	0.996	0.798	190.16	9.8
LB	150	A	GTGAAA	L008	31,473,035	0.996	0.786	215.33	9.5
LB	150	B	AGTTCC	L006	20,016,988	0.998	0.808	178.49	9.3
LB	150	C	ATGTCA	L005	30,253,324	0.996	0.814	183.35	9.2
LB	160	A	GTGAAA	L005	14,981,564	0.996	0.811	195.50	8.2
LB	160	B	AGTCAA	L001	6,613,627	0.995	0.801	183.29	8.1
LB	160	C	GTCCGC	L003	4,498,403	0.997	0.817	188.57	7.4
LB	170	A	AGTCAA	L007	28,634,115	0.996	0.808	188.17	9.8
LB	170	B	GTCCGC	L006	54,668,356	0.998	0.809	182.10	10.0
LB	170	C	AGTCAA	L004	22,356,753	0.996	0.810	188.93	9.9
LB	200	A	ATGTCA	L008	16,701,328	0.997	0.791	173.07	9.5
LB	200	B	GTGAAA	L002	8,978,602	0.996	0.777	205.14	9.9
LB	200	C	CCGTCC	L001	20,497,625	0.998	0.763	208.80	9.6
PL	100	A	AGTCAA	L008	16,674,885	0.996	0.798	195.50	10.0
PL	100	B	GTCCGC	L007	9,288,679	0.997	0.798	190.52	10.0
PL	100	C	CCGTCC	L002	40,970,653	0.997	0.787	213.78	10.0
PL	140	A	GTCCGC	L004	37,166,151	0.996	0.799	189.61	9.5
PL	140	B	ATGTCA	L003	6,519,881	0.996	0.802	180.05	9.7
PL	140	C	AGTCAA	L002	26,836,537	0.997	0.786	202.68	9.7
PL	150	A	CCGTCC	L008	27,515,495	0.998	0.779	219.85	9.9
PL	150	B	ATGTCA	L006	21,572,729	0.998	0.807	180.36	9.8
PL	150	C	AGTCAA	L005	31,276,989	0.996	0.809	191.31	9.9
PL	160	A	CCGTCC	L005	19,719,655	0.997	0.789	207.33	8.4
PL	160	B	AGTTCC	L001	46,868,956	0.998	0.795	190.61	9.5
PL	160	C	GTGAAA	L003	19,583,357	0.995	0.788	197.48	8.7
PL	170	A	AGTTCC	L007	43,829,383	0.996	0.800	183.36	9.5
PL	170	B	GTGAAA	L006	30,612,275	0.997	0.797	189.97	9.3
PL	170	C	AGTTCC	L004	13,537,568	0.996	0.797	195.55	9.8
SB	100	A	AGTTCC	L008	20,853,072	0.996	0.805	189.29	10.0
SB	100	B	GTGAAA	L007	11,073,164	0.996	0.804	188.17	9.9
SB	100	C	GTCCGC	L001	19,435,986	0.998	0.806	192.59	9.5
SB	140	A	GTGAAA	L004	11,034,246	0.995	0.802	199.26	9.1
SB	140	B	AGTCAA	L003	35,722,829	0.996	0.802	189.73	9.8
SB	140	C	AGTTCC	L002	21,053,359	0.997	0.801	196.91	9.9
SB	150	A	GTCCGC	L008	19,505,065	0.997	0.794	209.40	-
SB	150	B	AGTCAA	L006	44,039,656	0.998	0.814	169.16	9.9
SB	150	C	AGTTCC	L005	17,412,112	0.996	0.789	195.52	10.0
SB	160	A	GTCCGC	L005	30,431,301	0.997	0.805	198.43	10.0
SB	160	B	ATGTCA	L001	28,770,693	0.997	0.796	186.67	10.0
SB	160	C	CCGTCC	L003	28,085,159	0.997	0.789	199.52	9.8
SB	170	A	ATGTCA	L007	31,585,623	0.996	0.797	199.35	10.0
SB	170	B	CCGTCC	L006	17,485,747	0.999	0.793	191.56	10.0
SB	170	C	ATGTCA	L004	20,881,241	0.996	0.810	184.47	10.0
SB	200	A	CCGTCC	L007	14,769,479	0.998	0.805	194.85	9.8
SB	200	B	GTCCGC	L002	11,357,151	0.997	0.793	202.11	9.6
SB	200	C	GTGAAA	L001	86,894,303	0.997	0.769	196.70	9.6
Average					24,751,761	0.997	0.797	193.42	9.58
Min					4,498,403	0.995	0.763	169.16	7.4
Max					86,894,303	0.999	0.817	219.85	10.0

<sup>a</sup> The proportion of reads retained as pairs after trimming<sup>b</sup> The proportion of raw reads aligned with kallisto<sup>c</sup> Estimated fragment length by kallisto

**Table S2.** Tab-delimited text file with detailed annotation of all the filtered transcripts in the Trinity assembly.

On figshare: [Annotation\\_combined.tsv](#)

**gene\_id:** Trinity gene identification code  
**transcript\_id:** Trinity transcript identification code  
**sprot\_Top\_BLASTX\_hit:** Top blastx hit from the SwissProt database  
**TrEMBL\_Top\_BLASTX\_hit:** Top blastx hit from the TrEMBL database  
**RNAMMER:** ribosomal RNA prediction  
**prot\_id:** Id for protein prediction  
**prot\_coords:** Coordinates for predicted proteins  
**sprot\_Top\_BLASTP\_hit:** Top blastp hit for predicted proteins from the SwissProt database  
**TrEMBL\_Top\_BLASTP\_hit:** Top blastp hit for predicted proteins from the TrEMBL database  
**Pfam:** Protein domain prediction from Pfam  
**SignalP:** Prediction of signal peptides  
**TmHMM:** Prediction for transmembrane domains  
**eggnoG:** Annotation to the eggNOG database  
**gene\_ontology\_blast:** Gene ontology categories based on blast results from SwissProt and TrEMBL  
**gene\_ontology\_pfam:** Gene ontology based on Pfam protein domains  
**SSncbi\_Top\_BLASTN\_trans:** Id for top blastn transcripts in NCBI *Salmo salar* Annotation  
**SSncbi\_Top\_BLASTX\_trans:** Id for top blastx transcripts in NCBI *Salmo salar* Annotation  
**SSncbi\_Top\_BLASTN\_gene\_name:** Gene name based on top blastn hit in the NCBI *Salmo salar* Annotation  
**SSncbi\_Top\_BLASTX\_gene\_name:** Gene name based on top blastx hit in the NCBI *Salmo salar* Annotation  
**OM\_Top\_BLASTN\_trans:** Id for top blastn transcripts in *Oncorhynchus mykiss* genome annotation  
**OM\_Top\_BLASTX\_trans:** Id for top blastx transcripts in *Oncorhynchus mykiss* genome annotation  
**SalmoBase\_Top\_BLASTN\_trans:** Id for top blastn transcripts in SalmoBase *Salmo salar* Annotation  
**SalmoBase\_Top\_BLASTX\_trans:** Id for top blastx transcripts in SalmoBase *Salmo salar* Annotation  
**SalmoBase\_Top\_BLASTN\_gene\_name:** Gene name based on top blastn hit in the SalmoBase *Salmo salar* Annotation  
**SalmoBase\_Top\_BLASTX\_gene\_name:** Gene name based on top blastx hit in the SalmoBase *Salmo salar* Annotation

**Table S3.** The results of gene ontology analyses of the transcripts with significant expression difference between morphs (or morph by time interaction) in the Arctic charr developmental transcriptome. The enrichment was tested for transcripts and genes (SalmoBase) within each expression cluster.

In Dropbox: [GO\\_Clusters\\_BP.csv](#)

**GO.ID:** Identification number for Gene Ontology categories  
**Term:** The Gene Ontology term or description of the category  
**numDE.t:** Number of transcripts within expression cluster in each GO-category  
**numIn.t:** Total number of transcripts in each GO-category  
**fdr.t:** Multiple testing corrected P-value (FDR) for enrichment based on transcripts  
**p.t:** Uncorrected P-value for enrichment based on transcripts  
**numDE.g:** Number of genes (SalmoBase) within expression cluster in each GO-category  
**numIn.g:** Total number of genes (SalmoBase) in each GO-category  
**fdr.g:** Multiple testing corrected P-value (FDR) for enrichment based on genes  
**p.g:** Uncorrected P-value for enrichment based on genes  
**Cluster:** Expression cluster  
**GOclust:** Super-GO-categories based on categories semantic similarity (see Methods)

**Table S4.** Information about genes used in qPCR. Detailed gene names, primer sequence, amplicon size and transcripts in the assembly used for comparison.

In Dropbox: [qPCR\\_primers.csv](#)

**Gene Symbol:** The symbol or short gene name used in figures and text

**Description:** Full name of each gene

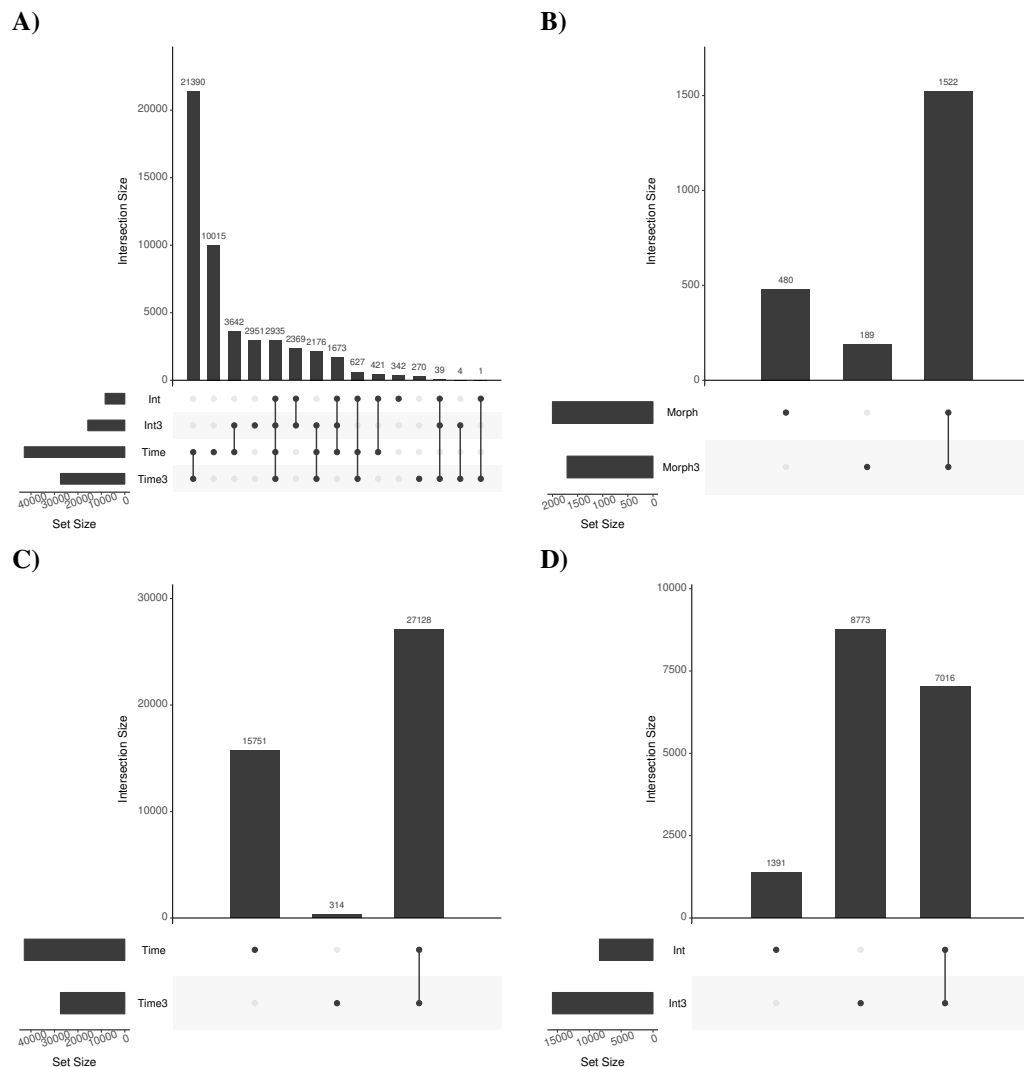
**Forward primer:** Sequence for the forward qPCR primer in 5'-3' orientation

**Reverse primer:** Sequence for the reverse qPCR primer in 5'-3' orientation

**Amplicon size:** Size of the sequence amplified in the qPCR reaction

**Transcripts:** The id of assembled transcripts in the transcriptome that the primers bind to and were used for comparison of expression. If there are more than one transcripts for each gene the ids are separated by a semicolon.





**Figure S1.** Effect of 3'-bias correction on the number of transcripts differently expressed by developmental timepoint (time), morph and interaction of morph and time (int). Each figure shows the intersection size (upper barplot) - the number of transcripts significant for each one or a combination of two or more factors (indicated by dots), while the set size barplot (lower) shows cumulated number for each factor. Indicated are the number transcripts differently expressed (DE) with ("effect") and without ("effect3") taking 3'-bias into account. For example, the Morph category in figure **B** represents the number of DE-transcripts when 3'-bias is taken into account, but Morph3 category denotes transcripts that are DE when 3'-bias estimator was dropped from the model. The different panels represent the impact of 3'-bias on **A)** the Time and Morph by Time interaction (int) terms, **B)** the Morph term, **C)** the Time term in isolation and **D)** only the  $M \times T$  interaction (int) term. The dots indicate the significant factors or their combinations.