Gene expression during early development and genetic variation among recently evolved sympatric Arctic charr (*Salvelinus alpinus*) morphs in Lake Þingvallavatn, Iceland

Jóhannes Guðbrandsson

Faculty of Life- and Environmental Sciences
University of Iceland
2019

Gene expression during early development and genetic variation among recently evolved sympatric Arctic charr (*Salvelinus alpinus*) morphs in Lake Þingvallavatn, Iceland

Jóhannes Guðbrandsson

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Biology

Advisor
Arnar Pálsson

PhD Committee
Zophonías Oddur Jónsson
Páll Melsted
Einar Árnason
Sigurður S. Snorrason

Opponents
Kathryn Rebecca Elmer
Jón Hallsteinn Hallsson

Faculty of Life- and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, April 2019

Gene expression during early development and genetic variation among recently evolved sympatric Arctic charr (*Salvelinus alpinus*) morphs in Lake Þingvallavatn, Iceland
Gene expression and genetic variation among Arctic charr morphs in Lake Þingvallavatn
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Biology

Faculty of Life- and Environmental Sciences
School of Engineering and Natural Sciences
University of Iceland
Askja, Sturlugötu 7
101, Reykjavik
Iceland

Telephone: 525-4000

Bibliographic information:
Jóhannes Guðbrandsson, 2019, *Gene expression during early development and genetic variation among recently evolved sympatric Arctic charr (Salvelinus alpinus) morphs in Lake Þingvallavatn, Iceland*, PhD dissertation, Faculty of Life- and Environmental Sciences, University of Iceland, 234 pp.

ISBN 978-9935-9438-5-9

Printing: Samskipti ehf.
Reykjavik, Iceland, April 2019

Abstract

The four morphs of Arctic charr (*Salvelinus alpinus*) in Lake Þingvallavatn, Iceland, which differ in many phenotypic traits related to morphology, life history, and feeding ecology, are believed to have evolved locally within the lake after the retreat of the Ice-age glacier. The principal aims of the study were to get a handle on the molecular mechanisms underlying the phenotypic differences and assess genetic differentiation among the morphs. Gene expression during early development was surveyed to look for differential expression of genes thereby identifying developmental processes affecting divergent phenotypes. Genetic variation among the morphs was studied and patterns of biological function (gene ontology) and genomic position of highly differentiated variants were examined. Both candidate gene and transcriptome-wide profiling (RNA-sequencing) approaches were used. The results reveal both differentiation in expression and genetic composition of the sympatric morphs. Gene expression differences in multiple genes and biological pathways were discovered, pointing to substantial differentiation of morphs and that many genes may have been under selection. However, RNA degradation during sample handling created technical problems that complicated data analysis. The transcriptome-wide profiling was nonetheless useful as it revealed expression differences in genes relating to extra-cellular matrix formation and skeletogenesis and highlights differentiation in more biological pathways. The data show extensive genome-wide differentiation in allele frequencies, pointing to reproductive isolation of the morphs and/or natural selection operating on multiple parts of the genome. Strong genetic differentiation among morphs was for example discovered in immunological genes using a candidate gene approach and in genes involved in collagen metabolism and environmental sensing, by gene ontology enrichment tests of differentiated variants in the transcriptome. The expression differences in embryos and the clear genetic differentiation of the morphs suggest that they have advanced along "the speciation continuum" towards becoming reproductively isolated.

Útdráttur

Fjögur afbrigði bleikju (*Salvelinus alpinus*) finnast í Þingvallavatni. Afbrigðin eru talin hafa þróast innan vatnsins frá lokum síðustu ísaldar og eru ólík hvað varðar m.a. stærð, útlit, lífsferla og fæðuöflun. Meginmarkmið verkefnisins var að auka skilning á hvaða sameinda- og þroskunarferlar liggja að baki mismunandi svipgerð afbrigðanna og meta erfðafræðilega aðgreiningu þeirra á milli. Könnuð var genatjáning í snemmþroskun í leit að mismunandi tjáðum genum sem gætu þannig bent á hvaða ferlar móta þroskun ólíkra svipgerða. Erfðabreytileiki meðal afbrigðanna var metinn og athugað hvort að erfðaset með mikinn mun í tíðni samsæta, meðal afbrigða, fyndust í genum með svipuð líffræðileg hlutverk og/eða væru bundin við ákveðna staði innan erfðamengisins. Tvenns konar aðferðarfræði var beitt, í fyrsta lagi að kanna áhugaverð gen með sértækum aðferðum og í öðru lagi að rannsaka allt umritunarmengið (RNA-raðgreiningu). Niðurstöðurnar afhjúpa ólíka genatjáningu og erfðamun milli afbrigðanna. Munur í tjáningu finnst á genum í mörgum líffræðilegum ferlum sem bendir til töluverðar aðgreiningar afbrigðanna og að mögulega hafi mörg gen verið undir náttúrulegu vali. RNA-niðurbrot í sýnum, líklega vegna tæknilegra vandkvæða í meðhöndlun þeirra, hækkaði flækjustig greininga á RNA-raðgreiningargögnum. Greining umritunarmengisins var engu að síður gagnleg þar sem tjáningarmunur fannst í genum sem tengjast myndun utanfrumugrindar og beinmyndun. Gögnin sýna mikinn mun í tíðni samsæta meðal afbrigða og var mikinn mun að finna víða í erfðamenginu. Það getur bent til æxlunarlegrar einangrunar afbrigðanna og/eða náttúrulegs vals á mörgum litningasvæðum. Mikill erfðamunur fannst til dæmis í genum tengdum ónæmiskerfinu með sértækum aðferðum og aðgreinandi erfðabreytileiki í genum tengdum kollagen-efnaskiptum og umhverfisskynjum (sjón, heyrn) var áberandi í umritunarmenginu. Genatjáningarmunur í fóstrum og skýr munur í erfðasamsetningu afbrigðanna bendir til að þau séu komin áleiðis inn á "veg tegundamyndunar".

Dedication

Tileinkað ástkonu minni Jónínu Svavarsdóttur og börnunum okkar:
Steinunni Láru, Jónasi Emil, Brynhildi Eyju og Örnólfi Kára

Table of Contents

List of Figures

List of Tables

List of Original Papers

Paper I: Ehsan Pashay Ahi, **Jóhannes Guðbrandsson**, Kalina H. Kapralova, Sigríður R. Franzdóttir, Sigurður S. Snorrason, Valerie H. Maier and Zophonías O. Jónsson. 2013. Validation of reference genes for expression studies during craniofacial development in Arctic charr. *PLOS ONE* **8(6):** e66389.

Paper II: Kalina H. Kapralova, **Jóhannes Guðbrandsson**, Sigrun Reynisdottir, Cristina B. Santos, Vanessa C. Baltanas, Valerie H. Maier, Sigurdur S. Snorrason and Arnar Palsson. 2013. Differentiation at the MHCIIα and Cath2 Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn. *PLOS ONE* **8(7):** e69402.

Paper III: **Jóhannes Gudbrandsson**, Ehsan P. Ahi, Kalina. H. Kapralova, Sigridur R. Franzdottir, Bjarni K. Kristjansson, S. Sophie Steinhaeuser, Isak M. Johannesson, Valerie H. Maier, Sigurdur S. Snorrason, Zophonias O. Jonsson and Arnar Palsson. 2016. The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations]. *F1000Research* **4:** 136.

Paper IV: **Jóhannes Guðbrandsson**, Sigríður R. Franzdóttir, Bjarni K. Kristjánsson, Ehsan P. Ahi, Valerie H. Maier, Kalina H. Kapralova, Sigurður S. Snorrason, Zophonías O. Jónsson and Arnar Pálsson. 2018. Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs. *PeerJ* **6:** e4345

Paper V: **Jóhannes Guðbrandsson**, Kalina. H. Kapralova, Sigríður R. Franzdóttir, Völundur Hafstað, Þóra M. Bergsveinsdóttir, Zophonías O. Jónsson, Sigurdur S. Snorrason and Arnar Pálsson. 2018. Extensive genetic divergence between recently evolved sympatric Arctic charr morphs. *bioRxiv preprint* doi:10.1101/489104.

Other papers (not included in the thesis)

1: Ehsan Pashay Ahi, Kalina Hristova Kapralova, Arnar Pálsson, Valerie Helene Maier, **Jóhannes Gudbrandsson**, Sigurdur S. Snorrason, Zophonías O. Jónsson and Sigrídur Rut Franzdóttir. 2014. Transcriptional dynamics of a conserved gene expression network associated with benthic-limnetic craniofacial divergence in Arctic charr. *EvoDevo* **5:** 40

2: Sigurður Guðjónsson, Sigurður Már Einarsson, Ingi Rúnar Jónsson and **Jóhannes Guðbrandsson**. 2015. Marine feeding areas and vertical movement of Atlantic salmon (*Salmo salar* L.) as inferred from recoveries of Data Storage Tags. *Canadian Journal of Fisheries and Aquatic Sciences* **72(7):** 1087–1098.

3: Leese F, Altermatt F, Bouchez A, Ekrem T, Hering D, Meissner K, Mergen P, Pawlowski J, Piggott J, Rimet F, Steinke D, Taberlet P, Weigand A, Abarenkov K, Beja P, Bervoets L, Björnsdóttir S, Boets P, Boggero A, Bones A, Borja Á, Bruce K, Bursić V, Carlsson J, Čiampor F, Čiamporová-Zatovičová Z, Coissac E, Costa F, Costache M, Creer S, Csabai Z, Deiner K, DelValls Á, Drakare S, Duarte S, Eleršek T, Fazi S, Fišer C, Flot J, Fonseca V, Fontaneto D, Grabowski M, Graf W, **Guðbrandsson J**, Hellström M, Hershkovitz Y, Hollingsworth P, Japoshvili B, Jones J, Kahlert M, Kalamujic Stroil B, Kasapidis P, Kelly M, Kelly-Quinn M, Keskin E, Kõljalg U, Ljubešić Z, Maček I, Mächler E, Mahon A, Marečková M, Mejdandzic M, Mircheva G, Montagna M, Moritz C, Mulk V, Naumoski A, Navodaru I, Padisák J, Pálsson S, Panksep K, Penev L, Petrusek A, Pfannkuchen M, Primmer C, Rinkevich B, Rotter A, Schmidt-Kloiber A, Segurado P, Speksnijder A, Stoev P, Strand M, Šulčius S, Sundberg P, Traugott M, Tsigenopoulos C, Turon X, Valentini A, van der Hoorn B, Várbíró G, Vasquez Hadjilyra M, Viguri J, Vitonytė I, Vogler A, Vrålstad T, Wägele W, Wenne R, Winding A, Woodward G, Zegura B, Zimmermann J. 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes* **2:** e11321.

4: Sigurður Már Einarsson, Sigurður Guðjónsson, Ingi Rúnar Jónsson and **Jóhannes Guðbrandsson**. 2018. Deep-diving of Atlantic salmon (*Salmo salar* L.) during their marine feeding migrations. *Environmental Biology of Fishes* **101(12):** 1707–1715.

Acknowledgments

I would like to thank my supervisor Arnar Pálsson for his patience and support through the years and getting lost in discussing exciting scientific projects and the wonders of nature. My team of supervisors Sigurður Snorrason and Zophonías Oddur Jónsson have been a great help through my studies. Thanks to Einar Árnason and Páll Melsted for serving on my PhD committee and I would also like to think Páll for collaboration on transcriptome assembly.

I am grateful for all the members of the Charr group, past and present. Thanks to Kalina Kapralova, Sigríður Rut Franzdóttir, Valerie Helene Maier and Ehsan Pashay Ahi for your collaboration on this project. Thanks to Hákon Jónsson for taking the first steps in bioinformatics along my side. The people at Hólar University and Verið have collaborated and supported this work and I am very grateful for their part. deCODE Genetics was kind enough to run our sequencing library on their machines.

Thanks to all my collaborators at the Marine and Freshwater Research Institute (former Veiðimálastofnun), especially Sigurður Guðjónsson and Sigurður Már Einarsson for giving me the freedom to find my own way in data analysis and boosting my confidence as a scientist.

Funding was provided by the University of Iceland Doctoral Fund and Teaching Assistance Fund and the Icelandic Center for Research (RANNÍS). I would like to thank the taxpayers of Iceland for financing my studies from when I started as a six-year-old until today. It would not have been possible without their support. Thanks to all the people who develop open statistical and bioinformatics tool and to the people who take time to explain the functionality of software and help other users.

Ég vil þakka fjölskyldu og vinum sem hafa stutt mig til að verða vísindamaður og betri manneskja í gegnum árin. Sameindasystur: Takk fyrir að hafa strák með í stelpuhópnum og slúðra, borða kökur, nördast, syngja og bara hafa gaman. Takk Sara og Sigga fyrir nördið, Ásdís fyrir rolluspjall og Arna sýna að maður má/á fara að vinna við það sem maður hefur ástríðu fyrir.

Takk mamma og pabbi fyrir að vera alltaf til staðar skilyrðislaust. Það er alltaf gott að koma heim og finna að maður er velkominn. Það voru forréttindi að fá að alast upp í sveit og að geta verið með í vinnu ykkar og geta leikið sér út í náttúrunni. Takk Áslaug, Úlli, Gunna, Bjargey, Gunnhildur, Nonni og Lukka fyrir að vera frábær stóru systkini. Það er ómetanlegt að vita að það sé alltaf einhver til staðar sem maður getur leitað til. Takk fyrir akademískt slúður Bjargey og takk Nonni fyrir að horfa á Attenborourgh og aðra náttúrulífsþætti og smita mig af áhuga.

Takk elsku Jónína fyrir að fara í gegnum þetta með mér. Þú, Steinunn Lára, Jónas Emil, Brynhildur Eyja og Örnólfur Kári hafið kennt mér meira um sjálfan mig og lífið en nokkur háskólagráða.

1 Introduction

> Hvaðan kom hann!? Hvert er hann að fara!? Hvað er hann!?
>
> Adolf Ingi Erlingsson - 31. Jan. 2010

The quote above, which can be translated as; "Where did he come from!? Where is he going!? What is he!?" was used by a sports presenter to describe one of the most memorable moments in Icelandic handball history. However, these questions also resonate with scientists and their urge and approach to understand natural phenomena. These questions reflect the urge to describe nature (What is he!?), the search for explanations of the origins of phenomena (Where did he come from!?) and how science aims to make predictions about the future (Where is he going!?).

The nature, origin, and delimitation of species have been topics of intense investigations in biology, and the questions above have also been asked about species. A system for naming and organizing species was developed in the 18th century (von Linné, 1758). But the foundations for the modern theory on species, their formation, properties, diversification, and extinction, was laid in the 19th century (Darwin, 1859). Questions about the nature and formation of species and their adaptations are central to the field of evolutionary biology. Evolutionary changes in populations occur because of the action and interplay of evolutionary forces, the most important being mutation, genetic drift, migration and natural selection (Hartl and Clark, 2006). Mutations generate new variation that other evolutionary forces operate on. The Random sampling of alleles from generation to generation is referred to as genetic drift, which given enough time reduces genetic variation within populations. The smaller a population the stronger this effects is. With population subdivision or structure, drift can lead to differentiation of populations, while migration between sub-populations counteracts this and promotes genetic homogenization of populations. Thus an interplay between population structure and migration can influence evolutionary trajectories. Last but not least, adaptive changes in organisms are brought about by natural selection, which alters allele frequencies over generations because of differences in fitness. All these forces operate at the same time and influence the evolution of populations and the formation of species.

Natural selection operates on the variation of phenotypes within populations. In each generation, the variation of phenotypes is influenced by how individuals develop through the influence and interactions of genetic- and environmental factors and chance, and how, at each developmental timepoint or stage, phenotypes may vary in fitness. Over generations, natural selection can only alter the variation and distribution of phenotypes if they have a heritable component (Futuyma, 2005).

Mutations causing changes in protein sequence, regulatory elements or deletions/ duplications of genes can lead to phenotypic changes. The interactions among genetic factors and the relationship between genetic factors and phenotypes are complex (Chandler et al., 2013). In the bodies (and embryos) of individual organisms, genes interact in biochemical and signaling pathways were transcription, translation, and protein degradation are initiated, terminated and tweaked to produce the right amount of the genetic products at the right time and place(s) (cells, tissues, organs). For instance, in embryos, the induction of a specific developmental event requires the activation of triggers, which directly or indirectly activate gene networks in competent cells. Development is essentially a complex symphony of such actions and interactions between molecules, tissues and cellular populations (Gilbert, 2006). Every population has segregating genetic variation, some of which in housekeeping genes and genes influencing the formation of developmental and functional phenotypes. Furthermore, the development of organisms always occurs in an environment, and variation in environmental factors (or influence of chance) can perturb development and lead to different outcomes. Therefore, it should not come as a surprise that the interactions of genes in the organism and numerous environmental factors influencing molecular pathways can have a significant effect on the amount and nature of phenotypic changes caused by characterized mutations (Chandler et al., 2013). Transcription factors, for example, influence the expression of other genes and can, therefore, control various cellular functions (Latchman, 2015). Mutations leading to deletions or additions of transcription binding sites can also have drastic effects on the timing and placement of gene expression in development (Wray, 2007). For example, the *Hox* gene family of transcription factors regulates the anterior/posterior development in animals, and in tetrapods, these genes were co-opted, as new control regions evolved that turn on *Hox* genes in the development and orientation of limbs (Tschopp and Duboule, 2011). Another example of transcriptional mechanisms influencing phenotypes is seen in how the expression of the signaling molecules *bone morphogenetic protein 4 (Bmp4)* and *calmodulin (CaM)*, in the craniofacial development of Darwin finches (*Geospiza*), influences beak depth and length respectively (Abzhanov et al., 2004, 2006). Both are important components of fitness in many finch species (Grant and Grant, 2002). Curiously, when genetic variation among finches was dissected at the genome level, no polymorphism in or close to *Bmp4* was found to correlate with beak depth, indicating that expression differences are not controlled by *cis*-regulatory elements of this gene, but by factors further upstream in the molecular pathway (Lamichhaney et al., 2015). The transcription factor *Alx*1 appeared to be a good candidate, as it showed strong differentiation among species and individuals, but the nature or existence of a relationship between these genes is not known. On the other hand, indications of genetic differentiation between finch species and forms, and positive selection could be found in the same genomic locality as *CaM* (Lamichhaney et al., 2015).

While genetic changes are clearly important for phenotypic evolution, phenotypic diversity can also emerge without an underlying genetic diversity. Phenotypic traits, that can change during individual lifetime, in response to the environment are called plastic traits. Traits can vary in plasticity, and species and even populations differ in their level of plasticity (Nordeng, 1983; Parsons et al., 2011). Theory predicts that species in unstable environments should be more plastic, and it has been suggested that

plasticity could play a crucial role in the first stages of divergence between populations (West-Eberhard, 1989; Snorrason and Skúlason, 2004; Pfennig et al., 2010). One form of plasticity is when genetic variation with no or little effect on phenotypic traits under normal conditions reveals its influence on phenotypes under extreme circumstances. This is cryptic genetic variation which maybe important for adaptation to new environments (Chandler et al., 2013; Paaby and Rockman, 2014). Plasticity can also be viewed as a trait in itself that can be under selection (Cousyn et al., 2001; de Jong, 2005; Ergon and Ergon, 2017). However, it is currently not clear how large a role plastic responses, genetic accommodation, or genetic assimilation, play in evolution (Ehrenreich and Pfennig, 2016).

It is inevitable that the divergence of populations and speciation will involve changes in developmental pathways (Wray, 2007). How genetic changes in populations modulate and lead to changes in development and its underlying genetic hierarchy is an exciting field within evolutionary biology. Recently diverged species or populations represent excellent systems to tackle questions relating to these processes.

This thesis focuses on studying and identifying the molecular background influencing the formation of four distinct morphs of Arctic charr in Lake Þingvallavatn in Iceland. A better understanding of such divergences is of paramount interest as they may reflect the early steps of speciation event(s). The introduction of this thesis will start by reviewing ecological speciation in freshwater fishes and zooms in on the highly polymorphic Arctic charr, and then gives an overview of existing knowledge about the morphs in Lake Þingvallavatn.

1.1 Ecological speciation in Freshwater fishes

Speciation has been defined in several ways and can be studied from various angles (e.g. geographic or population genetic). One view on speciation centers on how patterns of geographical distribution of populations shape the emergence of new species, such as geological barriers between diverging populations (Coyne and Orr, 2004; Futuyma, 2005). For instance, in allopatric speciation, strong geographical barriers exist and block migration between populations, which, in the long run, will eventually lead to divergence of populations into true species. On the other hand, in sympatric speciation, no geographical barriers exist to migration between populations during speciation so barriers to gene flow are of a different nature. Sympatric speciation is thought to be much rarer than allopatric speciation and is only considered possible by strong divergent selection or sexual selection (Coyne and Orr, 2004). Genetic drift can play a role in allopatric speciation and as does positive selection when isolated populations adapt to their respective environments. Furthermore, natural selection is believed to be important during secondary contact if geographical barriers break down (Coyne and Orr, 2004; Futuyma, 2005).

From another angle, speciation can be studied at the population genetic level, as the consequence of evolutionary forces, like drift, mutation and natural selection. Speciation can occur when reproductive isolation evolves as a consequence of divergent selection on traits in contrasting environments. The environment may refer to the habitat or

different resources available to the same population (see review by Schluter, 2001). This is commonly referred to as ecological speciation and happens most readily in allopatry but also possibly in sympatry. The divergence of populations into separate species does not happen spontaneously and diverging populations can be thought of as being located on different positions along a theoretical "speciation continuum" (Via, 2009; Seehausen et al., 2014; Lowry and Gould, 2016), which spans from a homogeneous single population all the way to reproductively isolated species. This "speciation continuum" is not a one-way street as populations may diverge during one time period only later to stall or become more similar, or even collapse into one intermixing population. This could, for example, be caused by environmental fluctuations (Grant and Grant, 2008), invasion of exotic species (Taylor et al., 2006; Bhat et al., 2014), or eutrophication (Seehausen, 1997; Jacobs et al., 2018a). Some data suggests that the process of speciation may be rapid and the continuum can have sudden non-linear changes or tipping points (Nosil et al., 2017). One way to study the causes, patterns, and mechanisms of speciation would be to study populations that are at different levels of divergence, either because they have been adapting to particular environments for different lengths of time, or because the selective forces act more strongly in some populations than in others (for instance in large vs. small populations) (Via, 2009; Nosil et al., 2017).

For several reasons, freshwater fish species offer excellent opportunities to study speciation. First, many freshwater systems offer discrete habitat diversity that can drive diverging adaptations. Second, many species form local (semi-)isolated populations. And third, the topography and history of watersheds are often relatively well understood making the estimation of the age of populations and lineage splits possible. Curiously, although freshwater habitats are by nature fragmented and often isolated, the diversity of freshwater fishes cannot entirely be explained by allopatric speciation. Multiple examples of sympatric morphs or species are known (Seehausen and Wagner, 2014). Cichlids have, for example, proved a good model system for studying evolutionary genetics and genomic methods have shown that much of the genetic variation can be shared among species, selection often acts on standing genetic variation, some of which can be gained by introgression or hybridization, which may be important in differentiation (Henning and Meyer, 2014). Discrete sympatric polymorphisms within species (morphs), which often correlate with resources, have been linked to "open niches" and the lack of interspecific competition within isolated lakes or other habitats (Smith and Skúlason, 1996). This seems to be the case for numerous Nordic freshwater fishes, most probably because after de-glaciation at the end of the last ice age (10,000 to 15,000 years ago), new rivers and lakes were formed and colonized by a limited number of anadromous freshwater species. With the subsequent rising of land (isostatic rebound)(Norðdahl et al., 2008) and the erosive power of rivers many populations became isolated above waterfalls and in some cases, long distances to ice age refugia have prevented colonization at lower elevations. Nordic fishes may, therefore, be ideal to study ecological speciation at both early and late stages of divergence as similar adaptations (forms) seem to have evolved in parallel on a rather well defined and short timescale (Snorrason and Skúlason, 2004) . The divergence of resident freshwater and marine or anadromous populations seems to have been common in sticklebacks and salmonids. In lakes adaptation to benthic or limnetic lifestyle is a

repeated theme (Seehausen and Wagner, 2014), for instance in lake whitefish (*Coregonus spp.*) (Bernatchez et al., 2010) and three-spined stickleback (*Gasterosteus aculeatus*) (Willacker et al., 2010). Studies on three-spined sticklebacks have also shown the importance of intrinsic factors such as chromosomal structure and recombination in divergence, as genomic areas with lower recombination appear to be more likely to harbor variants important for adaptation (Samuk et al., 2017), and McGuigan et al. (2011) revealed the role of cryptic genetic variation adaptation of marine populations to freshwater habitat.

Multiple examples exist of within-species divergence and polymorphism among salmonids. Whitefish frequently diverge into benthic and limnetic forms and morphs and the molecular mechanism behind the divergence has been extensively studied (Bernatchez et al., 2010). Beach, river and stream spawning morphs of sockeye salmon (*Oncorhynchus nerka*) show divergence in morphology and genetics (Larson et al., 2017). The genus *Salvelinus* contains very polymorphic species. Arctic charr (*Salvelinus alpinus*) has been called "the most variable vertebrate on earth" as sympatric and allopatric polymorphisms are abundant in that species (Klemetsen, 2013). Lake charr (*Salvelinus namaycush*) shows multiple examples of divergence, especially within large and deep lakes (Muir et al., 2016), and Dolly Varden charr (*Salvelinus malma*) is no exception as seven sympatric morphs can be found within the same lake (Markevich et al., 2018).

Is adaptive evolution or radiation the product of ecological opportunity or does genetic constitution also play a role? It has been hypothesized that whole genome duplications promoted the diversification of vertebrates (Holland et al., 1994; Sidow, 1996). The ancestor of salmonids underwent a whole genome duplication approximately 95 million years ago (Allendorf and Thorgaard, 1984; Macqueen and Johnston, 2014; Berthelot et al., 2014). This is in addition to three earlier whole genome duplications, two occurred in an ancestor of vertebrates, and an additional one in the lineage leading to the teleosts (Glasauer and Neuhauss, 2014). Evolution by neo-functionalization (Ohno, 1970) or sub-functionalization (Force et al., 1999) has been suggested to play a role in the retention of paralogous genes in salmonids. Large sections of the genomes of salmonids remained under tetrasomic inheritance for a significant proportion of their evolutionary history and Robertson et al. (2017) suggested that salmonid radiations co-occurred with re-diploidization of parts of the genome. Analyses of this kind (*i.e.* Robertson et al., 2017) have been enabled by advantages in DNA sequencing and related methods, that have transformed research in multiple fields in this young century (Goodwin et al., 2016). Genomics has allowed scholars to expand evolutionary and genetic research to non-model organism (da Fonseca et al., 2016), and in salmonids genomic analyses are used in studies of population structure (Bourret et al., 2013; Moore et al., 2017), adaptation (Barson et al., 2015; Kjærner-Semb et al., 2016), and estimation of evolutionary history of populations and conservation of gene expression (Jacobs et al., 2018b,a) (see Methods for more details).

1.2 The polymorphic Arctic Charr (*Salvelinus alpinus*, Linnaeus, 1758)

Arctic charr has a Holarctic distribution and is often found as the only fish species occupying rivers and lakes in the high Arctic (Reist et al., 2013, Figure 1.1). The species is highly variable in morphology and life history and in many lakes, distinct, sympatric morphs have been recorded. This extensive variation and propensity for polymorphism was problematic for taxonomists of the 20th century and was often referred to as the "Charr problem" (Reisinger, 1953; Nordeng, 1983; Klemetsen, 2010).

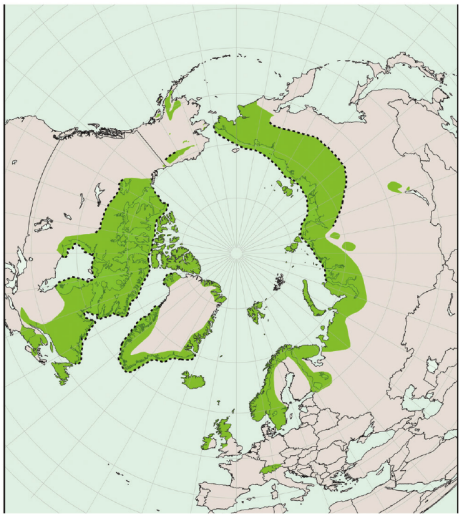


Figure 1.1. The species distribution of Arctic charr. The distribution is restricted to the Arctic and cold lakes at high altitudes in the northern hemisphere. It has a Holarctic distribution. Map from Reist et al. (2013), ©Taylor & Francis

Differences between charr populations or morphs, in morphology, migration, spawning time, feeding ecology, growth and life history characteristics have been reviewed extensively and the importance of behavior, plasticity, and genetics for morph formation have also been discussed (Skúlason and Smith, 1995; Jonsson and Jonsson, 2001; Noakes, 2008; Klemetsen, 2013). Population genetic data show that phenotypically different sympatric morphs are in most cases more closely related to each other than to morphs of similar phenotype in geographically different areas, pointing to sympatric origin of morphs (in Iceland (Gíslason et al., 1999; Kapralova et al., 2011), in Norway (Hindar et al., 1986), in Transbaikalia (Gordeeva et al., 2015), in Alaska (May-Mcnally et al., 2015) and in both Atlantic and Siberian linages (Jacobs et al., 2018b)).

There are exceptions to this pattern. In Loch Rannoch, Scotland, which harbours three different morphs of Arctic charr, genetic data indicates that the morphs are derived from at least two colonization events (Verspoor et al., 2010) and in Eastern Canada two examples suggesting an allopatric origin of morphs have been recorded (Gomez-Uchida et al., 2008; Power et al., 2009). The Canadian examples might not be surprising as mitochondrial haplotypes from three different glacial refugia meet in that geographic area (Brunner et al., 2001).

Like many other places in the northern hemisphere, Iceland was covered by ice until about 12,000 years ago (Norðdahl et al., 2008). As the ice retreated Arctic charr (presumably anadromous fish) is thought to have colonized rivers and lakes. Icelandic lakes and streams harbor multiple forms of Arctic charr (Skúlason et al., 1992; Woods et al., 2012a) and in several lakes polymorphism is observed, the occurrence of which has been linked to low temperature, absence or low density of brown trout, and the

size of the limnetic habitat (Woods et al., 2012b). Variation in morphology between monomorphic lake or spring populations has been linked to water origin, bedrock age, and fish community structure (Kristjánsson et al., 2011). A rather unique small benthic morphotype is commonly found in spring-fed streams and ponds in the neo-volcanic zone in Iceland. These fish mature early, retain parr marks and reach sexual maturity at a small size, typically at around 8 – 15cm fork length (FL). Despite these similarities in life history and morphology subtle differences in morphology have been reported related to habitat, i.e. whether they live in streams or in ponds, which suggests parallel local adaption to these contrasting habitats (Kristjánsson et al., 2012). Population genetic microsatellite data suggest that this small charr evolved independently in many locations (Kapralova et al., 2011).

In Iceland, Arctic charr has been fished for human consumption for centuries and in recent decades recreational angling has been on the rise. Aquaculture of Arctic charr was initiated by the government in Iceland 1961 (Gunnarsson and Guðbergsson, 1988), and in 1989 a breeding program was initiated in order to establish an aquaculture stock for distribution to producers. Multiple wild populations were sampled and mixed in the breeding program to produce a fast growing, late maturing fish (Gunnarsson, 2006; Svavarsson, 2007). Having gone through 6–7 generations of selection for the desired characteristics the aquaculture charr presently differs considerably from wild charr in various respects, especially in traits that relate to the physiological bases of the life-history characteristics selected for, and in morphology.

Polymorphism in Icelandic charr is at its peak in the island's biggest natural waterbody (Lake Þingvallavatn) where four sympatric morphs reside (Figure 1.2).

1.2.1 The four morphs of Lake Þingvallavatn

Lake Þingvallavatn was formed as the Icelandic ice-cap receded $\sim$10,000 years ago. After major eruptions, approximately 10,000 years ago, northeast of the lake, lava flows entered the lake shaping the northern and eastern shores and finally dammed the lake at the outflow. Carbon dating estimated the age of the lava to be $9,130 \pm 260$ years old (Kjartansson, 1964). After its formation, the lake has been influenced by tectonic movements, causing extensive subsidence and horizontal extension with extensive rift-forming in the central graben along the northeast-southwest axis of the present lake. Two smaller eruptions (at $\sim$5,700 and 2,000 years from the present) caused small lava flows into the southwest corner of the lake. During the latter episode, a sub-surface eruption formed the crater island, Sandey, and caused a considerable fall-out of volcanic ash east of the island (Saemundsson, 1992). At present, the outflowing river, River Sog, has a series of three hydroelectric dams barring any upward migration of fish to the lake and before the erection of the dams impassable waterfalls had the same effect. The actual time when the lake became closed to upwards migration of fish is not known but may have happened early (> 10,000 years ago) at the time of fast ice melt. At that time the lake is believed to have been a glacial lagoon in which case the River Sog would have been a large glacial river (Saemundsson, 1992), with a powerful waterfall-building potential.

The lake harbours three fish species, threespine stickleback (*Gasterosteus aculeatus*), brown trout (*Salmo trutta*), and Arctic charr, which, as previously mentioned, can be

categorized into four distinct morphs, differing in morphology and feeding ecology (Malmquist et al., 1985; Snorrason et al., 1989; Sandlund et al., 1992). Although the existence of Arctic charr morphs in Þingvallavatn had been known by local farmers for centuries, this knowledge first found its way into the scientific literature by the work of Bjarni Sæmundsson (1897; 1904; 1917). In his first publication mentioning Þingvallavatn Sæmundsson (1897) describes the exploitation of fish stocks in the lake and advanced the idea that one of the morphs, *murta* (which means small fish in Icelandic), was a separate species or a variety of Arctic charr. A few years later Sæmundsson (1904) described four morphs or varieties designated by the local farmers, with observations on the diet and habitats; *murta*, a small plankton feeder, *netbleikja* (or just *bleikja*), feeding on benthic invertebrates, *djúpbleikja*, a piscivorous fish which are caught in the deeper parts of the lake, and *depla* a spotted, medium-sized charr, which Sæmundsson considered to be young and immature *bleikja* or *djúpbleikja*. In addition, he described a small benthic morph *gjámurta*, found in fissures at Thingvellir on the north shore of the lake ("gjá" means fissure in Icelandic). Sæmundsson (1904) also presented results of tagging experiment of *murta* which showed that 43 out of 67 recaptures had not grown to be considered another morph even after 3 years, 19 had reached similar size as *depla* (9.5 – 12 inches), and 4 had reached the size of *bleikja*. He concluded that the *murta* catch was a mix of young charr and a special variety. In the first edition of his book on Icelandic fishes Sæmundsson (1908, in Danish) describes the *murta* and the *gjámurta* as special varieties, *S. alpinus* var. *murta* and *S. alpinus* var. *Thingvallensis*, respectively. In the updated version of the book (Sæmundsson, 1926, in Icelandic), the *murta* has lost its variety status, perhaps due to age readings showing that most of the specimens he studied were young (3–5 years), but those readings might be unreliable as scales were used and not otoliths (Sæmundsson, 1917). Later work by Friðriksson (1939) on the age of spawning *murta* (from otoliths) showed the presence of older fish. He also described a benthic morph (*svart-murta*) which in morphology was similar to *gjámurta* but larger (FL 18 – 26cm). The *murta* was according to his research also larger in size (FL 18 – 32cm) than estimates from the 1980's (Snorrason et al., 1992), which might suggest good growing condition in the lake or alternatively some sampling or methodological bias. Friðriksson (1939) concluded, from counts of vertebrae, that *murta* and *svart-murta* were new charr varieties, *S. alpinus forma murta* and *S. alpinus forma niger*, respectively.

Research on the variability of charr in the lake resumed in the 1980s and based on multi-mesh gill-net surveys covering all size ranges and different habitats, the four morphs were described and defined in more detail (Malmquist et al., 1985; Snorrason et al., 1989; Sandlund et al., 1992). From then on in the literature, and in this thesis, the morphs are named as follows. The *murta* is referred to as the planktivorous morph (PL-charr, with a mean adult fork length of 175 and 188mm for males and females respectively, and a maximum length of $\sim$260mm). It is the most abundant morph in the lake and mainly feeds on zooplankton in the limnetic zone. It has a pointed snout and a terminal mouth similar to the ancestral anadromous charr (Sandlund et al., 1992). The PL-charr matures at 4–5 years (Fork length around 15–20cm) but length at maturation and growth rate seem to depend on food availability and population size (Snorrason et al., 1992). The rarest and least studied is the Piscivorous morph (PI-charr, with a mean adult fork length of 268 and 295mm for males and females respectively, and a

maximum length > 500mm). It is similar in morphology to PL-charr but matures later, grows larger (Jonsson et al., 1988), and mainly feeds on sticklebacks in the littoral zone (Malmquist et al., 1992). Snorrason et al. (1989) proposed that the PI and PL-charr represent two alternative ontogenic trajectories of the same population because their juvenile growth is similar up to the point where some individuals switch to fish eating, thereby becoming PI-charr. The lake is also the home to two benthic morphs. The large benthic morph (LB-charr, with a mean adult fork length of 240 and 274mm for males and females respectively, and a maximum length > 500mm) matures late and continues to grow throughout its life (Sandlund et al., 1992). The small benthic morph (SB-charr, mean adult fork length of 96 and 109mm for males and females respectively, maximum length 310mm) matures early (maturing at 2–3 years, males earlier than females). It lives in between stones and in crevasses on the stony lava bottom, a habitat that dominates the littoral zone in the lake. Both benthic morphs have blunt snouts, short lower jaws, and their pectoral fins are relatively larger compared to the other morphs. Both feed exclusively on benthic invertebrates, the gastropod mollusc *Radix peregra* being the most important prey species for both morphs (Sandlund et al., 1992; Malmquist et al., 1992; Jonsson et al., 1988).

How do these descriptions of the morphs fit the descriptions of Bjarni Sæmundsson and Árni Friðriksson? According to Malmquist et al. (1985) LB-charr clearly fits the the description of (*netbleikja*). Similarly, PI-charr is the same as *djúpbleikja* described by Sæmundsson (1904) and the *murta* is the same as PL-charr. As discussed in Malmquist et al. (1985) the dwarf morph *gjá-murta* described by Sæmundsson (1904) from the fissures at Thingvellir, and the *svart-murta* described by Friðriksson (1939), is most likely the same as SB-charr.

Due to the differences in feeding ecology and habitat use the morphs also differ in their parasitic fauna (Frandsen et al., 1989, **Paper II**) and colouration, with the benthic morphs being darker (Sandlund et al., 1992). There is some separation in spawning time and location, most dramatically as the LB-charr spawns in August and the other morphs in September and October, though the peaks of spawning vary (Skúlason et al., 1989). Although there is overlap in spawning time and spawning location of the other morphs, detailed studies of spawning behaviour *in natura* have only been performed for LB-charr to date (Sigurjónsdóttir and Gunnarsson, 1989). It would be interesting to test if the morphs exhibit certain preferences for microhabitat, like types of gravel/rocks, spawning depth or even diel variation in mating behaviour. Skúlason et al. (1993, 1996) showed that some aspects of foraging behaviour, growth and maturation are heritable traits in the morphs but genetic studies have shown that the morphs are more related to each other than to other Arctic charr populations in Iceland indicating sympatric origin (Volpe and Ferguson, 1996; Gíslason, 1998). However micro-allopatric origin has also been suggested (Kapralova et al., 2011). Significant genetic differences were found between some of the morphs but the exact phylogenetic relationship between them has not been resolved (Magnusson and Ferguson, 1987; Danzmann et al., 1991; Volpe and Ferguson, 1996; Gíslason, 1998; Kapralova et al., 2011). Partly this could reflect that the morphs are environmentally induced, as suggested for PL and PI-charr (see above), or if they are genetically distinct, this could stem from limited resolution of genetic markers, a short time since separation, incomplete lineage sorting or incomplete reproductive isolation between morphs. Parsons et al. (2010) performed a common

garden experiment with offspring of PL- and LB-charr from Lake Þingvallavatn offering two different types of food, limnetic food presented at the surface and benthic food presented at the bottom. The results revealed significant developmental plasticity, where the food type influenced the shape of the head and jaws of juveniles. Notably, the level of plasticity varied between morphs, LB-charr being less plastic (Parsons et al., 2011).

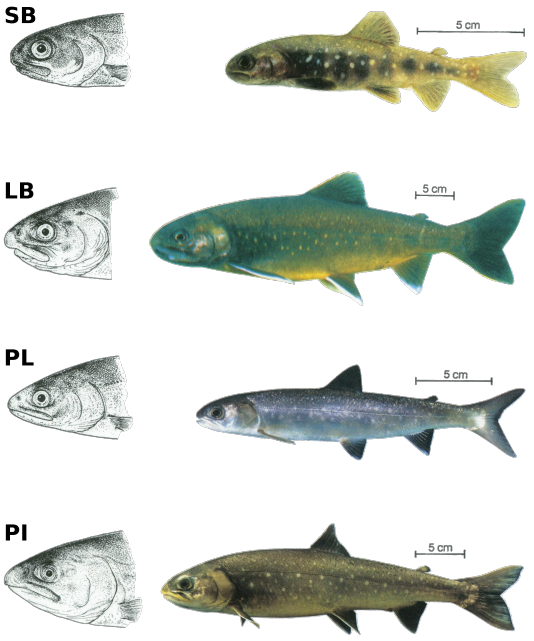


Figure 1.2. The four morphs of Arctic charr (Salvelinus alpinus) in Lake Þingvallavatn. Small benthic (SB), Large benthic (LB), Planktivorous (PL) and Piscivorous (PI). The morphs differ in body shape, craniofacial structures, coloration, and adult size. Multiple traits in their ecology and life history differ (see text). Drawings by Eggert Pétursson and photographs by Karl Gunnarsson and Arnþór Garðarson, reprinted from Sandlund et al. (1992).

Finding genetic differences that separate the morphs, and/or molecular factors that influence variation in morph development could help us understand their origin and the nature of their phenotypic differences. For instance, do the differences stem mainly from genetic factors or do environmental differences during development play a more important role? Genes involved in craniofacial development could be expected to be differentially expressed or contain genetic differentiation between morphs because the morphs differ in their trophic morphology (Figure 1.2). Also, as the immunological systems of animals are under constant selective pressures (Bernatchez and Landry, 2003), genes involved in pathogen and parasite defence may have evolved to reflect the diet (and associated parasites) and habitat differences among morphs. Thus looking at gene expression during development of important morphological features and genetic variation among the morphs can bring us closer to understanding what role molecular pathways play in morph differentiation and how and where they operate in developing individuals.

2 Aims and structure of the thesis

The overall aim of the thesis is to gain insight into the molecular and genetic basis of the phenotypic diversity and rapid evolution of Arctic charr resource morphs in Lake Þingvallavatn. The objectives were,

1. to uncover the molecular mechanisms that influence the phenotypic differences observed among the four morphs.

2. to assess to what extent the differences among the morphs are due to genetic differences rather than being environmentally induced, and furthermore

3. to look for indications of genetic differentiation in the genomes of the four morphs, that might reflect diversifying selection at specific loci and the degree of reproductive isolation between them.

To address these aims five research questions were tackled using different molecular methods. The research questions studied in this thesis were;

i) Which genes/transcripts are differentially expressed among the morphs in early development?

ii) Do the differentially expressed genes/transcripts observed share molecular functions or biological roles?

iii) Is there a genetic separation between morphs, and, if so, what is their relationship? For instance, are the benthic morphs more genetically similar to each other than to the other morphs?

iv) Which genetic variants show the strongest differentiation among morphs?

v) Do such variants show specific patterns of genetic differentiation? First, do they have a function related to phenotypes distinguishing the morphs or do multiple variants share molecular function(s)? Second, are these variants found in few distinct regions of the genome, or are they found in multiple regions?

The thesis consist of four published papers and one manuscript (referred to as **Paper V** hereafter). The papers focused on differential gene expression in development and/or genetic variation among the four morphs (Table 2.1).

By gaining better insights into the molecular mechanism of the phenotypic difference we hope to learn more about how natural selection operates on the developmental and molecular level to produce variation and divergence between populations.

Table 2.1. Overview of the research questions each paper tackles. The research questions are lined out in the text above. An X marks if a question is addressed in the corresponding paper. The questions focus on differential expression (DE) or genetic variation (GV).

	Paper				
Question	**I**	**II**	**III**	**IV**	**V**
i) Differential expression (DE)	X		X	X	
ii) Patterns in DE			X	X	
iii) Genetic distance between morphs					X
iv) Genetic variation (GV) with high differentiation		X	X		X
v) Patterns of GV with high differentiation			X		X

3 Methods

The aims and research questions were tackled with two experimental strategies (candidate genes vs. transcriptome/genome-wide profiling) and from two molecular angles (by estimating mRNA levels or identifying genetic variants). First, the candidate gene strategy was used to study specific genes or developmental processes, implicated by previous work and/or literature to be relevant for the observed phenotypic differences among the morphs (candidate gene strategy). Secondly, we analyzed the developmental transcriptomes, and thus indirectly the genomes, of charr morphs for differences among them. For both strategies, we looked at gene expression and genetic variation. The content of each paper with respect to these topics is summarized in (Table 3.2).

Table 3.2. The experimental strategies and molecular angles studied in the papers

	Candidate gene	Transcriptome
Gene expression	**Paper I, III**	**Paper III, IV**
Genetic variation	**Paper II, III**	**Paper III, V**

Because the Arctic charr of Þingvallavatn exhibit such a clear divergence in trophic morphology the members of the laboratory that I worked in have been much interested in studying differences in craniofacial development. Comparison of the early development of the morphs in common garden environment is one approach to find out when morphological differences arise and to test if the differences are more likely to be caused by genetic or environmental factors. Kapralova (2014) showed that craniofacial differences do indeed arise early and that the morphology of hybrids between two of the morphs differs from the parental morphs, some are intermediate while others show transgressive morphology. Furthermore, the hybrid embryos grew slower than progeny of pure morphs suggesting some kind of hybrid incompatibility. Studying which genes are differentially expressed at developmental stages preceding the development of craniofacial structures could reveal particular molecular pathways responsible for the observed morphological differences. Furthermore, the discovery of alleles or genes with strong differentiation between morphs could shed further light on how the morphological divergences have evolved. Genes with such variation would be of utmost interest, for instance, if their function is linked to processes that can be expected to show differentiation between morphs such as craniofacial development or pathogen resistance.

3.1 Developmental series of morphs in a common garden

One way to detect the genetic effects on phenotypic differences of morphs is to grow them under the same environmental conditions. Temperature has a large effect on the developmental rate of salmonids and Gorodilov (1996) found a strong relationship between the time it takes for a somite pair to form and temperature, which enables normalization of temperature effects on development (Gorodilov, 1996). We utilized this method to correct for temperature effects on development in **Papers III** and **IV**. However, the developmental rate can be affected by other factors and can even change during the divergence of populations (Rogers and Bernatchez, 2006). Observed changes in gene expression at developmental time points estimated from the time of conception and temperature can, therefore, be caused by changes in developmental rate (heterochrony) rather than tweaks in transcript abundance at homologous developmental events. Such changes are none the less interesting as they might be, as was previously noted, related to divergence. The head and jaw phenotype of benthic Arctic charr morphs has been classified as paedomorphic and it has been suggested that it might be the result of heterochronic changes in early development (Eiriksson et al., 1999). In a common garden set-up, differences between morphs in concentrations of a specific transcript measured at given developmental stages might, therefore, stem from at least three types of differences or changes: i) morph-specific difference in abundance of a particular transcript within the same cell type or cell population; ii) morph-specific differences in proportional sizes of specific tissues or organs; iii) changes in developmental rate between morphs. However, as noted above, this can lead to time specific changes in the expression of the transcript, *i.e.* the same expression pattern, but the timing of expression peaks, or changes, delayed or accelerated in different morphs.

3.2 The sequencing revolution and RNA-sequencing

As was stated above, short read sequencing methods developed after the year 2000 (especially Illumina sequencing) have revolutionized the fields of evolutionary ecology and developmental biology, particularly in the last decade (Goodwin et al., 2016). Instead of being limited to few model species (Watson et al., 2004) genomic studies and analyses of gene expression can now be performed in practically every species of interest (da Fonseca et al., 2016). These methods can be used to study DNA sequences and genomes of species and populations, and to study the expression of these genomes through RNA-sequencing (RNA-seq). First, whole genome sequencing has been used to do molecular population genetics on an unprecedented scale, like for instance in the model species yeast *Saccharomyces cerevisiae* and fruit fly *Drosophila melanogaster* (Liti et al., 2009; MacKay et al., 2012). But for non-model species without a reference genome, the changes have been more drastic. Here, new methods, such as RAD-sequencing and RNA-seq, have enabled affordable research on evolutionary and ecological questions by studying genomes and transcriptomes (da Fonseca et al., 2016).

For larger genomes, restriction site associated DNA (RAD) sequencing (Miller et al.,

2007) has enabled researches to study genetic diversity in multiple species (Andrews et al., 2016). For Arctic charr the method has been used for reconstructing phylogenetic relationships (Recknagel et al., 2015), generating a linkage map, delineating chromosome synteny and evolution in salmonids (Nugent et al., 2017), discovering sex-linked markers (Benestan et al., 2017) and population structure of anadromous populations in Canada (Moore et al., 2017). RNA sequencing is another powerful way to study the molecular mechanism behind phenotypic differences and adaption in non-model species (see for example (Bernatchez et al., 2010; Jacobs et al., 2018a)). Examples of RNA sequencing in Arctic charr include generation of a reference transcriptome (Magnanou et al., 2016), a study of gene expression of salinity tolerance (Norman et al., 2014), analyses of the influence of temperature on daily expression rhythms (Prokkola et al., 2018), and in conjunction with RAD-seq, estimates of parallelism in genetic and gene expression divergence (Jacobs et al., 2018b). Besides surveying variation in gene expression, RNA-seq can identify differential usage of splicing isoforms. Furthermore, genetic variation in expressed sequences can also be assessed from the same data (Cloonan et al., 2008; Morin et al., 2008). RNA-seq using mRNA, by nature focuses on the transcribed (mostly coding) parts of the genome which may make alleles influencing traits under positive selection easier to discover compared to RAD sequencing (assuming that positive selection acts primarily on coding regions and/or regulatory sequences near or embedded within transcribed genes, see Lowry et al., 2017).

RNA sequencing was the method used in **Papers III, IV** and **V**. Illumina sequencing can only amplify fragments up to 1000bp in length and sequence a maximum of 200–300bp of each end (paired-end reads). This poses certain challenges in the analysis of the output of RNA-sequencing runs. There are challenges in assembling short reads into full-length transcripts and in distributing the reads among transcripts. Also, biases caused by variation in library preparation and RNA degradation can complicate the data analysis (Sigurgeirsson et al., 2014).

For species that still lack a reference genome, the task of assembling short sequencing reads into full-length transcripts is a challenging task. Algorithms that use de Brüijn graphs have been found to be suitable for Illumina sequencing and have been popular for transcriptome assemblies (Grabherr et al., 2011; Schulz et al., 2012). Tracing the correct path for specific isoforms of the same gene in the graph can be difficult and when recent paralogs are present, distinguishing between them can be close to impossible using only short reads. For species with recent whole genome duplication such as Arctic charr and its relatives, it might be better to work on the gene family level rather than at the gene or transcript level, as we opted for in **Paper III**. The majority of RNA molecules in cells are ribosomal RNA (rRNA), but messenger RNA (mRNA) that encodes proteins are more interesting for most researchers. For this reason, it is important to increase the proportion of mRNA in the RNA pool after RNA isolation and before sequencing. This can be achieved in two ways. First, mRNA can be isolated by binding the poly-A tail of the RNA to oligo-T probes, called poly-A pulldown. Second, by hybridizing rRNA to DNA probes and either extracting it with magnetic beads (Illumina, 2016) or by digesting the double-stranded RNA/DNA hybrid (NEB, 2017), a process referred to as rRNA depletion. Following mRNA enrichment, the RNA is fragmented and reverse transcribed into cDNA, adapters are ligated and amplified by PCR before sequencing. After the sequencing, the next step is finding which transcripts/genes the reads came

from, commonly called read-mapping. Importantly, if poly-A pull-down is used, and the RNA has degraded before the sequencing, the sequencing reads will not be uniformly distributed over the length of the transcription unit. The reads will be more concentrated towards the 3′-end and can lead to reduced coverage for longer transcripts. From this follows, that if the level of degradation varies among samples this can also lead to false positives in differential expression (DE) analysis (Sigurgeirsson et al., 2014; Love et al., 2016). For well-annotated organism (with most full-length transcripts known) methods exists to account for 3′-bias. Such as modelling position bias in expression estimates (Love et al., 2016) or by only using the 3′-end of transcripts as a reference for read mapping (Sigurgeirsson et al., 2014). Wang et al. (2016) suggested calculating an integrity number for each sample and correcting expression estimates by this number. In **Paper IV** we dealt similarly with the problem of RNA degradation by estimating the severity of the 3′-bias for each sample by the proportion of reads in the 3′-half of house-keeping genes (genes with good annotation, and with stable and high expression). We subsequently took that estimate into account as a 2nd degree polynomial in the linear models used for DE analysis. This greatly reduced the number of transcripts with statistically significant morph effects, which most likely were false positives due to confounding of biological and technical variables. For species where the full-length transcripts are unknown, 3′-bias can cause serious problems. Variation in the coverage along the transcripts, due to, in addition to 3′-bias, sequence composition, shared exons among isoforms *etc.*, may lead to the breakup of one transcript into two or more contigs during assembly. Because of the 3′-bias, contigs belonging to the 3′-end of an mRNA get inflated expression estimates in degraded samples while contigs from the 5′-end are underestimated (Fig. 3.3).

When multiple paralogs are present, as in Arctic charr and other salmonid species, distributing reads between them and estimating paralog-specific expression can be solved using the expectation-maximization algorithm (EM-algorithm). Computationally the problem is similar or identical to estimating isoform specific expression and the same software packages can be used (*i.e.* Li and Dewey, 2011; Roberts and Pachter, 2012; Bray et al., 2016). The EM-algorithm iteratively estimates which reads belong to which transcript (paralog), given the expression of the transcript, and, consecutively, estimates the expression of transcripts, given which reads originate from it. The estimation continues until convergence is attained. This is convenient for distributing reads between similar sequences when the cannot be separated on sequence composition alone, as in the case of shared exons for isoforms and conserved areas for paralogs. Bray et al. (2016) utilized de Brüijn graphs for distributing reads to transcripts significantly reducing the computational time in expression estimation. The increased speed has made it possible to estimate uncertainty in expression estimated by using bootstrapping, thus hopefully, reducing false positives in DE analyses.

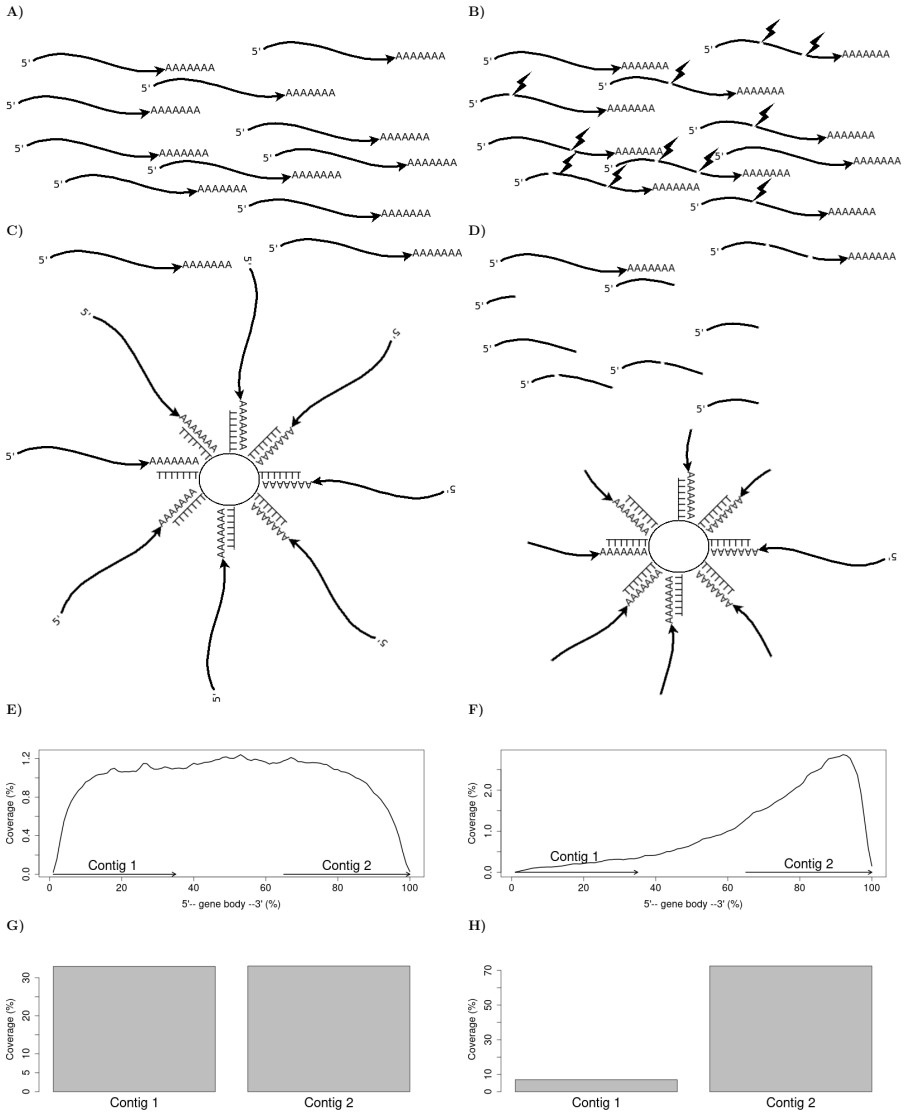


Figure 3.3. Illustration of the effect of differences in the level of RNA degradation on expression estimates in RNA-sequencing with incomplete reference transcriptome. A) Undegraded RNA molecules. B) Degraded RNA-molecules (lightning break the molecules). C) Poly-A pulldown with poly-T beads of an undegraded sample. D) Poly-A pulldown with poly-T beads of a degraded sample. E) Coverage along a transcript from an undegraded sample. F) Coverage along a transcript from a degraded sample. G) Expression estimate of two contigs from the 5′ and 3′ half the same sample from an undegraded sample and H) a degraded sample. RNA-degradation leads to underestimation of expression at the 5′-end and over-estimation at the 3′-end. If RNA-degradation varies between samples, this can lead to a contig showing differential expression that actually stems from technical variation but is not biologically real. If the full-length transcripts are not known for a species, then this problem is compounded.

Recent developments in long read sequencing can possibly solve some of these problems as full-length RNA-transcripts can be read directly (without a cDNA step) in a single pass (Garalde et al., 2018). Similarly, genome assemblies become less challenging as reads up to megabases in length could span regions with repeated sequences (Jain et al., 2018). More and more resources have become available for genomics and transcriptomics of salmonids. Notable were the reference set of expression sequence tags, SNP panels (Koop and Davidson, 2008; Sánchez et al., 2009) and fully sequenced genomes (Berthelot et al. 2014; Lien et al. 2016; Varadharajan et al. 2018. Most pertinently for this thesis is the recently published Arctic charr genome (Christensen et al., 2018). These resources, and others in development (see, Macqueen et al., 2017), will simplify further genetic work on the species and open up new scientific avenues.

3.3 Genetic diversity in expressed sequences

RNA-sequencing has the extra utility over other methods for estimating gene expression (like qPCR and microarrays), in that sequence polymorphism can also be estimated (Cloonan et al., 2008; Morin et al., 2008). Admittedly, variation in the level of expression between transcripts and treatment groups is a complicating factor which can influence the confidence in variant calling and estimates of allele frequencies between highly and lowly expressed transcripts. Sequencing errors can, for example, have higher read coverage in highly expressed genes than minor alleles in lowly expressed genes. Coverage thresholds for variant calling have, therefore, to take a local rather than a global coverage into account. Differential expression between groups (like morphs) can affect the discovery of variants, for example, fixed variants in genes that are only expressed in one morph cannot be discovered from RNA-seq data. Finally, allele-specific expression can influence allele frequency estimates in pooled sequencing (Konczal et al., 2014). Thus, cautious interpretation must be applied for differences in genes that appear to vary both in sequence variation and gene expression between morphs.

In species with many paralogs, such as Arctic charr, extra steps are needed to separate real polymorphism from paralogous sequence divergence. Filtering variants on read depth, haplotype structure, or excess heterozygosity are applicable in single individual DNA sequencing (Willis et al., 2017). The first two of those criteria are not suitable in RNA-sequencing from pooled samples (where more than one individual is sequenced) as gene expression varies and the number of haplotypes per individual cannot be assessed in pooled sequencing. However, in such data variants with excess heterozygosity can be filtered using deviation from Hardy-Weinberg equilibrium or by using F-statistics (F_{IS} or F_{IT}) (Hohenlohe et al., 2011). In **Papers III** and **V** we isolated RNA from multiple small embryos to get the required amount of RNA for sequencing. Identifying false positive sequence variants caused by paralogs can be detected using the same principle of excess heterozygosity in pooled samples as in samples from single individuals. As Hohenlohe et al. (2011) we used an F-statistic to identify those variants which we referred to as F_{PT} in **Paper V**. This metric can be considered analogous to F_{IT}. It estimates the diversity between samples and if the diversity is low the potential variant is most likely a false positive. We performed computer simulations to confirm

this notion and to decide on a cut-off for the analyses.

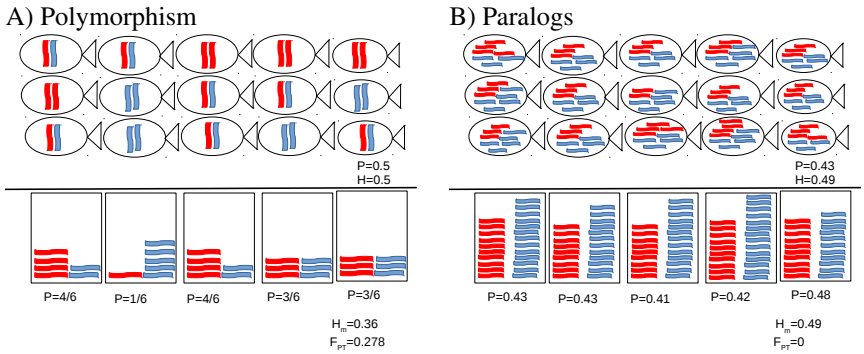


*Figure 3.4. Comparison of "allele frequencies" of candidate variants in pooled samples from two situations, A) true bi-allelic marker and B) a "candidate variant" due to sequence divergence between paralogs. A) Since each pool in the experimental setup in **Paper V** includes 3 individuals, the random sampling of true allelic polymorphism generates variation in allele frequencies between samples. B) When "candidate variation" is caused by differences between paralogs, all pooled samples should have similar observed frequencies. The exact value depends on variation in expression of the two paralogs between samples. Note, that if the expression of the paralogs is stable overall samples the variation in "allele frequencies" among samples should be low.*

As Variants with low minor allele frequency are hard to distinguish from errors in pooled RNA-seq (Konczal et al., 2014) and since we opted for using strong filters in variant calling in **Papers III** and **V**, estimating demographic histories from the site frequency spectrum was not feasible. Relative genetic distances between morphs can, on the other hand, be estimated from pooled RNA-seq data, and variants that show high divergence can be detected. Both estimates of genetic distances and highly differentiated variants can shed light on the molecular mechanism behind the morph divergence.

4 Results and Discussion

As was noted above we studied the morphs from two molecular angles: gene expression during development and genetic variation. Here, the results from the thesis will be discussed according to these perspectives, instead of by the chronological order of the papers. The next two sections start by summarizing the appropriate material of each paper followed by a synthesis of their results. First, I will discuss variation in gene expression and subsequently, I will discuss the genetic variation of the morphs.

4.1 Variation in gene expression during early charr development

Gene expression was addressed in **Papers I, III** and **IV**. The aim of **Paper I** was to identify good reference genes for qPCR-assays comparing gene expression in early development. Twelve genes were tested for stable expression both among morphs and over a developmental period. In addition to looking at their expression with qPCR, the RNA-seq data reported on in **Paper III** was used for gene expression stability analysis and comparison to the qPCR measurements in **Paper I**. All of the genes tested, except one, could be considered good reference genes. The geometric mean of two genes (*actb* and *if5a1*) were later used for gene expression analysis of Arctic charr heads in Ahi et al. (2014, 2015) and in **Paper III**. Three reference genes (*actb, UB2L3* and *ef1*α) were used in the analysis of whole embryos in **Paper III** and *actb* and *UB2L3* were used again in **Paper IV**.

The aim of **Paper III** was to create a starting point for discovering the molecular mechanisms relating to polymorphism in Icelandic Arctic charr, especially the morphs in Lake Þingvallavatn. Therefore, we decided to compare the transcriptome SB-charr from Lake Þingvallavatn to that of Aquaculture charr (AC) during development, with the purpose of looking for differentially expressed genes and genetic variation. AC-charr was selected because parent fish for making crosses were easily accessible and also because the aquaculture fish, with fast growth and limnetic craniofacial morphology, provided a dramatic contrast to SB-charr. Only eight samples were used for RNA sequencing and because the reads were short (36bp) differential expression analysis were performed on the gene family level. Interesting differential expression results were followed up on with qPCR. Expression differences in two genes with immunological function (*lysozyme II C* and *natterin-like*) were confirmed but detailed analysis of the *natterin* gene family showed that paralogs had distinct expression patterns. Differences in expression were confirmed in genes related to craniofacial development in embryonic heads. The RNA-seq data had already been used to select candidates for gene expression

analysis in embryonic heads (Ahi et al., 2014) (see further discussion below).

In **Paper IV** we surveyed gene expression in early development in three of the four morphs from Lake Þingvallavatn (PL- SB- and LB-charr). The developmental time points were chosen to precede the formation of craniofacial structures *e.g.* ossification of jaw bones. Multiple transcripts belonging to diverse biological processes were found to be differentially expressed. Gene ontology analysis of these transcripts did not indicate a significant excess of genes related to craniofacial development among the differently expressed transcripts, but RNA degradation and subsequent 3′-bias complicated the analysis and interpretation of results. Transcripts found to have time invariant expression differences between the morphs indicated more similar expression among the two benthic morphs compared to PL-charr. But the benthic morphs could also be separated based on gene expression. These data suggest genetic separation of morphs, and that the benthic morphs may be more related to each other than to PL-charr.

The results presented in **Papers III** and **IV** do not suggest that a single developmental pathway or physiological system is responsible for the phenotypic differences among the morphs. RNA degradation highly influenced transcript and gene expression analyses in Paper **IV**. For future studies, RiboZero, or similar products based on rRNA hybridization could be used to reduce the effect of degradation on DE analysis. Commercial kits have been shown to effectively remove rRNA in rainbow trout (Abernathy and Overturf, 2016). Freezing samples instead of using RNAlater is another option that has been used for Arctic charr parr with good results (Prokkola et al., 2018). It is uncertain that higher quality RNA would be isolated after freezing unhatched embryos compared to embryos stored in RNAlater, as the yolk of salmonids eggs is rich in proteins and fat (Berg et al., 2001), that could affect RNA isolation.

Analyses of gene expression of candidate genes, informed by these transcriptome data, were done by a fellow student in the lab using qPCR (Ahi et al., 2014, 2015). Differences were observed between limnetic (PL- and AC-charr) and benthic (SB- and LB-charr) forms in the expression of genes involved in extra-cellular matrix formation and skeletogenesis (Ahi et al., 2014). The genes studied by Ahi et al. (2014, 2015) were chosen based on putative function as well as differential expression in the transcriptome datasets present in **Papers III** and **IV**. Therefore, despite the biases described above, the data have been useful in our quest to discover molecular pathways that may relate to morph development and divergence.

Taken together our results have been useful, and although we have not been able to pinpoint a specific molecular/developmental pathway as the main mediator of morph divergence, our data has suggested the potential involvement of multiple candidate pathways. Having to rely on isolating RNA from whole embryos greatly restricted our studies. For example, important signals concerning differences in craniofacial development may have been swamped by expression in other body parts. Great methodological advances have now opened possibilities for a much sharper focus of enquiry as far as timing and positioning (in the embryo) of key developmental events are concerned. Thus detailed analysis, looking at specific sets of genes or repeating the transcriptome experiment focusing on specific organs, tissues or cell types will be a much more powerful approach. Recent advantages in single-cell sequencing, both in the technology and data analysis, will no doubt aid in this regard (Eberwine et al., 2014; Dal Molin and Di Camillo, 2018). Future studies on gene expression call for strict control of RNA

quality or the use of rRNA removal methods based on hybridization. This is essential to increase trust in results and enable stronger interpretation of analyses.

4.2 Genetic variation in the Lake Þingvallavatn morphs

Genetic variation in Icelandic Arctic charr, focusing mainly on three of the Lake Thingvallavatn morphs, was studied in **Papers II, III** and **V**. **Paper II** was a survey of genetic variation in candidate immunological genes (*Cath2, Leap-2a, Hamp* and *MHCII*α) and one region in the mitochondrial genome (D-loop). As immunological genes are often under strong selection pressure they are known to evolve relatively faster than the rest of the genome. Therefore immunological genes are good candidates to study if genetic divergence among the morphs exists at all. Two of the four immunological genes studied showed differentiation among the morphs (*MHCII*α and *Cath2*). *MHCII*α showed strong differentiation with PL-charr diverging from the two benthic morphs. PI-charr was not sampled in this study. Parasite infection rates and prevalence differed among the morphs, as was expected (Frandsen et al., 1989), but no association was found between parasite infections and genotype at the *MHCII*α locus. The *MHCII*α haplotype most common in PL-charr was rare in populations from other parts of the country and the variation in *Cath2* did not correlate to phenotype in other populations. This paper showed the first evidence for strong genetic differentiation between the morphs in Lake Þingvallavatn, suggesting i) that the morphs may be genetically distinct (indicating that other genomic regions might show similar patterns) and ii) that natural selection may have played a role in influencing the allele frequencies as was previously postulated (Kapralova et al., 2011).

As was noted above **Paper III** describes a scan of the transcriptome of SB-charr and Aquaculture charr (AC) during development. Genetic variants were called from the RNA-seq data and a handful of interesting variants were re-sequenced in an independent sample. Over 20.000 putative variants were discovered and over 1500 of them had considerable allele frequency differences between SB and AC-charr. Many SNPs with large frequency difference between SB and AC-charr were found in the mitochondrial genome, which could be the results of opposite selection for increased or reduced growth in AC- and SB-charr, respectively. Analysis of three mitochondrial variants in the 12s and 16s rRNA genes (positions 1829, 3211 and 3411) among SB, PL and LB-charr from Þingvallavatn indicated substantial differentiation in this chromosome, possibly reflecting differences in the function of mitochondria between the morphs. Morph specific demographic changes, such as population bottlenecks, could also increase the allele frequency of mitochondrial haplotypes if gene flow is limited between the morphs.

In **Paper V** we used the same data as in **Paper IV** to look at genetic variability among three morphs from Lake Þingvallavatn. Since the RNA-seq data was derived from pooled samples conventional approaches to variant calling and filtering were modified to analyze the data.A high level of genetic differentiation was discovered between the morphs and the samples clustered completely by morph in the first two axes from principal component analysis. Gene ontology analysis showed an excess

of variants separating PL-charr from the two benthic morphs in genes related to environmental sensing, tooth development and the extra-cellular matrix. Variants with high allele frequency differences between morphs were spread all over the genome and not restricted to a single or few genomic regions. The three mitochondrial variants observed in **Paper III** were also present in this data and two more, which differentiated the LB-charr from the other morphs, were also discovered in the mitochondrial genome. Using Kompetitive Allele Specific PCR (KASP) genotyping assays (Semagn et al., 2014) 22 variants were confirmed to be polymorphic in an independent population sample and strong differentiation between the morphs was confirmed in 17 of them. Some of these variants were located in genes that may potentially influence some of the phenotypic differences between morphs. The high F_{ST}-values in various part of the genome indicated that the morphs appear to be, at least partially, reproductively isolated in nature (Seehausen et al., 2014), although viable hybrid offspring can be generated (Kapralova, 2014).

The RNA-seq data gathered as a part of this study was not originally intended for discovering sequence variation. However, the result from **Paper V** show that RNA-seq data can be a reliable source for extracting information about genetic variants and genetic differentiation of populations or morphs. Unpublished RAD-seq data in house confirms the observed pattern of genetic distance between morph and multiple regions with high F_{ST} spread over all chromosomes (Sigurgeirsson, Xiao, Jónsson et al. unpublished data). Although more work is needed to confirm the genome-wide separation of morphs, it seems that multiple regions may have been under positive selection and that the Þingvallavatn morphs may be relatively far along the speciation continuum and might be reproductively isolated. Further investigations into the potential post- and prezygotic reproductive barriers is the logical step forward. LB-charr is known to be temporally isolated in spawning (Skúlason et al., 1989) but taking a closer look at the spawning behaviour of the other morphs might reveal pre-zygotic barriers such as spatial preferences in spawning or mate choice towards the same morph. It would be interesting to test if chemicals play a role in pre-zygotic barriers between the morphs and Sveinsson and Hara (1995) described an interesting experimental setup that could be used. Reconstructing the demographic histories of the morphs from population genetic data of the sympatric morphs, and geographically proximal populations would also help us understand how the morphs originated and evolved.

5 Conclusions and future perspectives

Let me revisit the questions that Adolf Ingi raised, cited at the beginning of the thesis, and how the work presented here has brought us closer to answering them for the evolutionary genetics of Arctic charr in Lake Þingvallavatn.

5.1 Where did they come from?

5.1.1 What is the origin of the morphs in Lake Þingvallavatn?

The focal questions of this study were, if, and then how much, have the four morphs diverged within Lake Þingvallavatn? This rests on the assumption that the sympatric morphs are of the same origin. Previous genetic results indicated that the morphs all descended from a single colonization event after the glacial retreat and formation of the lake (Gíslason, 1998; Kapralova et al., 2011). The data gathered for this thesis do not address this question directly, as the sampling was within the lake and not from other Arctic charr populations. The most prevalent haplotype, in PL-charr, at the *MHCII*α loci was uncommon in other populations outside the lake in Iceland (**Paper II**), pointing to intra-lake divergence. The data from **Paper III** show that the SB-charr and the Aquaculture charr (originated from crosses of wild charr from southern and northern Iceland (Gunnarsson, 2006)) differ genetically in many loci. But what is needed to address the origin of the morphs more fully? First, we need to understand the geology and history of the lake, and then plan studies of gene flow and genetic connectedness of Lake Þingvallavatn charr with other populations. Considering the dynamics of tectonic movements in the area, it seems likely that a part or parts of the population of colonizing charr could have been temporarily isolated from the main lake in fissures at the northern side of the lake. Fissures probably appeared soon after the lake assumed it's present characteristics ca. 10ky ago and have sequentially been connected to the lake as the graben floor subsided and the lake expanded to the north (Saemundsson, 1992). Thus a continuous series of peripheral cold spring habitats have been and are still being formed at the northern shores of the lake. As noted by Sæmundsson (1904) the fissures at Þingvellir ("the Parliament Plains") harbour small charr which, judging by his description of morphology, colouration and size at maturity, are phenotypically congruent with SB-charr in the lake proper (Malmquist et al., 1985). We since have ascertained that these SB-like charr are found in many fissures along the north shore of the lake some of which appear to be isolated from the lake. Indeed, Kapralova et al. (2011) concluded that an early allopatric epoch was the most likely scenario explaining the genetic differentiation between the SB- and PL-charr. Conceivably,

the small benthic morph in the lake could have evolved by parapatric or peripatric divergence (see Mayr, 1954, regarding peripatric speciation) and later expanded into the lake to occupy the spatial niche between stones of the shallow bottom, a niche that can only be effectively utilized by a small fish (Snorrason et al., 1989). The genetic composition of small isolated population in fissures would be dominated by genetic drift causing genetic differentiation between isolated population. If the process of isolation in fissures and re-colonization has occurred repeatedly SB-charr should be more variable than the other morphs and might contain sub-populations that migrated into the lake at different times. This is not supported by population genetic analyses of SB- and PL-charr from 5 spawning locations around the banks of the lake, which found slightly more differentiation by locations for PL-charr (Kapralova et al., 2011). Spawning site fidelity of PL-charr might, on the other hand explain the genetic structure within that morph. Looking at individual migration of different morphs within the lake, for instance by using acoustic tags (*i.e.* Dick et al., 2009), could illuminate differences in residency and homing of contrasting morphs. The other morphs (LB-, PL- and PI-charr) have most probably evolved in sympatry, perhaps after a short allopatric stage during the formation of the lake. It is important to gather extensive genetic polymorphism data from multiple individuals of each morph, to estimate demographic history of the morphs and to compare different models of morph separation and gene flow to get clearer answers about the origins of the morphs in the lake.

5.1.2 What biological processes are important for their differentiation?

The developmental pathways and molecular mechanism responsible for the phenotypic differences among the morphs are still being studied. The work presented here and by Ahi et al. (2014, 2015) found expression differences in genes related to craniofacial development and extra-cellular matrix organization, *i.e.* transcription factors *ets2* (Ahi et al., 2014) and *ahr2b* (Ahi et al., 2015). If these two genes are the causative agents driving differential expression of these processes, one might expect to observe genetic differentiation within their transcription units (except if they are influenced by distant regulatory elements). We did not discover variants with allele frequency differences close to *ets2* in **Paper V** (Figure 5.5 A). On the other hand, *ahr2b* is 440kb upstream of one marker validated to be differentiated between morphs (in the *sox21b* gene) but the variants closest to *ahr2b* show no differentiation. The expression differences among morphs, of those two genes, might, therefore, not be due to genetic variations in their regulatory regions. Results of gene ontology analysis in **Paper V** suggest an enrichment of variants in *i.e.* collagen organization and processes, which is consistent with the result of qPCR studies (Ahi et al., 2014, 2015). Further work is needed to identify the genetic causes of these expression differences and the very distinct jaw and head phenotypes of these morphs.

The enrichment of variants, with large differences in allele frequency, in genes related to environmental sensing is intriguing. This might suggests previously unobserved differences in the physiology of sight and hearing between the PL-charr and the two benthic morphs. Such differences might reflect the stark contrasts in the foraging environment and prey of the adult fish, the benthic morphs searching for and capturing benthic invertebrates in the rocky bottom littoral zone of the lake, while the PL-charr

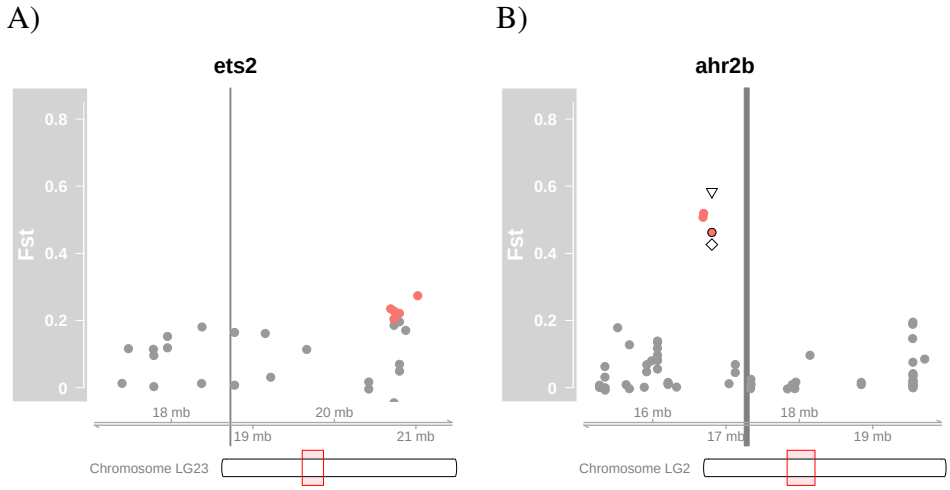


*Figure 5.5. Detailed view of F_{ST} values for variants from **Paper V** close to ets2 **A)** and ahr2b **B)**. The grey dots represent variants with F_{ST} below 0.2 and the red dots variants with F_{ST} above 0.2 were PL differs from the benthic morphs in allele frequency. The grey vertical lines represent the location of ets2 and ahr2b. A variant chosen for validation in **Paper V** (in the sox21b gene) is marked with a filled circle (●). The triangle (▽) shows the F_{ST} value from the KASP-assay for the three transcriptome morphs (PL, SB and LB) and the diamond (◇) F_{ST} for all morphs (including PI) for that marker. The red box in the figure below shows the location of the linkage group we are focusing on.*

are picking cladocerans and copepods out in open water (Malmquist et al., 1985, 1992; Skúlason et al., 1993). Feeding in open water also requires different strategies of alertness and evasive movements in response to threats of predators like the brown trout (*Salmo trutta*) or fish-eating birds, *e.g.* the great northern diver (*Gavia immer*) and the red-breasted merganser (*Mergus serrator*).

Results from the gene expression analysis in **Papers III** and **IV** indicated that there might be numerous other biological processes or physiological systems that could play a vital role in the phenotypic differences between morphs. Genes within high F_{ST}-peaks in **Paper V** need to be analyzed in more details to determine their functional roles, nature of the molecular lesions, and patterns of divergence. The publication of a Arctic charr draft genome (Christensen et al., 2018) and advantages in technologies in molecular biology will, without a doubt, stimulate research of which molecular systems show functional differences between the morphs and how they relate to phenotypic divergence.

5.2 Where are they going?

5.2.1 What does the future hold for the morphs? How will they be conserved?

Lake Þingvallavatn is not only the home of four sympatric Arctic charr morphs. Its location in a rift valley close to active volcanoes and the large influx of groundwater, has created diverse structures and a unique environment for freshwater organisms. A population of brown trout, renowned for its big size due to feasting on PL-charr, inhabits the lake (Sturlaugsson and Malmquist, 2011) and two morphs of three-spined stickleback have been described (Kristjánsson et al., 2002). Two freshwater amphipod species, endemic to Iceland, were first identified in the lake (Kristjánsson and Svavarsson, 2004; Svavarsson and Kristjánsson, 2006). The biota and geology together form a fascinating entity and its future protection status should be considered.

On a geological time scale (1000–10,000 years) the future of Þingvallavatn and its biota seems uncertain. Considering the location of the lake, in a geologically active zone, major geological events like volcanic eruptions are to expected in the near future. As stated above, the last eruption occurred approximately 2000 years ago (Saemundsson, 1992). It would be interesting to see if this eruption influenced the effective population sizes of the morphs, but predicting influences of future volcanic activity is, to say the least, challenging.

On a shorter time scale (decades) the biota of the Þingvallavatn is threatened by various human activities. In 1959 a hydropower dam was built at the lake outlet. This destroyed the immensely productive stretch of the River Efra Sog between Lake Þingvallavatn and Lake Úlfljótsvatn and is thought to have severely decreased the spawning and nursery grounds of brown trout at the outlet. This also prevented back migration of trout and charr from Lake Úlfljótsvatn. The migration of brown trout and charr between the lakes has not been measured. Similar charr morphs are found in Lake Úlfljótsvatn but they have not been studied to the same extent as the morphs in Lake Þingvallavatn. In 1990 a geothermal power plant was built at Nesjavellir on the NE side of Hengill. The plant utilizes high-temperature steam to produce hot water for district heating and household utilities in Reykjavík and the surrounding communities (capacity $\sim$300 MWt) and simultaneously generating electricity (capacity $\sim$120 MWe). At the present time, the plant discharges considerable amounts of excess heated water into subterraneous streams that drain into the lake at temperatures well above 20°C, causing local thermal stress in periods of calm weather and in periods of ice-cover during the winter (Snorrason, 2011). The effluents from the geothermal steam contain significant amounts of potentially toxic elements such as Hg, As, and Al (Wetang'ula and Snorrason, 2005). Presently this effluent is partly disposed of into drill-holes that reach beyond the layers of groundwater thus reducing the amounts of these elements reaching the lake.

The growth of phytoplankton in the lake is nitrogen-limited, making it vulnerable to the influx of nitrogen, e.g from increased use of nitrogen-containing fertilizer or organic sewage (Jónasson et al., 1992). The sharp increase in traffic due to increased tourism in the area in the last decade has increased the probability of pollution accidents in the vicinity of the lake, which could be harmful to its biota. Fortunately, the sewage from

a hotel and tourist services close to the lake is transported out of the lake catchment, but accidents during the transport of sewage still pose a pollution risk (Jónsson, 2016). Atmospheric nitrogen deposition might also be a possible pollutant source (de Wit and Lindholm, 2010). Increasing temperature due to global climate change has already affected the annual thermal cycle of the lake and further increases in global temperature can cause fundamental change the biological dynamics of the lake (Malmquist et al., 2012).

Because of its historic importance, the old parliament site and an area around it were protected by law already in 1928 (law nr. 59/1928). This law was changed in 2002 thereby including an extended area north of the lake and simultaneously, Lake Þingvallavatn and its catchment were protected by law (law nr 47/2004 and 85/2005). This law states that the biota of the lake shall be protected and refers specifically to the breeding grounds of the morphs of Arctic charr and brown trout in the lake. A project monitoring the ecological quality of the lake was initiated in 2007 (Malmquist et al., 2012) and with strong legislation, the future evolutionary trajectory of the morphs will hopefully involve minimum human impact. Environmental protection does not only aim at preserving stocks and species but also natural processes that maintain or generate diversity, such as speciation (Brodersen and Seehausen, 2014).

5.2.2 What further research is needed to characterize the differentiation of the morphs?

The work presented in this thesis has added to our knowledge on what biological processes might play a role in the phenotypic differentiation of the morphs in Lake Þingvallavatn. We have also gained information about their genetic differentiation and indications about the level of reproductive isolation between them. But how do we advance our understanding of their evolution and the molecular mechanism that shape the traits that differentiate them? It seems clear that the morphs do not constitute one panmictic population. What barriers are keeping the morphs apart and on what level are they operating (pre- or post-zygotic)? Based on present knowledge more specific questions can be asked. For example, which factors control the unusual spawning time of LB-charr, which spawn in late July-early August, which is earlier than other morphs in the lake and charr, in general, in Iceland. These questions call for studies (some already initiated) of breeding behavior and hybrid traits, and their relations to hybrid fitness.

Estimation of demographic models from genetic data have increased in popularity in recent years using the site frequency spectrum (Kamm et al., 2018) or coalescent hidden Markov models (Spence et al., 2018). How geological activity and changes in the lake topology have influenced the population sizes and perhaps gene flow between morphs are questions that could be answered with the above methods given the appropriate data. The relatedness of PI-charr to the other morphs and it's origin is worth exploring further. These questions are in part being addressed with whole genome or RAD-sequencing of populations samples in the laboratory.

Looking in detail at specific genes or molecular pathways discovered in this thesis, such as regions that show high differentiation in **Paper V** or transcripts with expression differences in **Paper IV**, are logical next steps. One would like to have detailed

information about when (in development) and where (in which tissue/organ/structural element) these candidate genes are expressed and how this differs among the morphs. Phenotypic variation is not only observed between morphs but variation in phenotypic traits within morphs that can be linked to genetic factors could help us understand how the same traits might differ between morphs and how they might be involved in the initial steps of morph evolution. Quantitative trait loci mapping is a powerful approach to determine the molecular causes of phenotypic traits (Laporte et al., 2015). However, due to long generation times (3–4 years) and relatively short grant periods (3 years) QTL analyses have not been applied to the morphs of Lake Þingvallavatn to date. Such an undertaking would without a doubt shed new light on the genetic underpinnings of the phenotypic differences among morphs and variation among individuals within morphs. Using hybrids of the morphs would be beneficial to mapping traits of interest but low survival and reduced fecundity might hinder such studies.

5.3 What are the morphs?

5.3.1 Genetically or environmentally induced?

The results in **Paper V** shows clear genetic differentiation between the morphs in Lake Þingvallavatn. It seems likely that there is already a certain degree of reproductive isolation between the three morphs studied. The genetic standing of PI-charr has not been determined but RAD-seq data in-house suggest that PI- and PL-charr may be environmentally induced forms from the same genetic population (Sigurgeirsson, Xiao, Jónsson et al. unpublished data) as was previously suggested by Snorrason et al. (1989). A similar scenario appears to be the case in landlocked Sockeye salmon (*Oncorhynchus nerka*) in Jo-Jo Lake, Alaska, where planktivorous and piscivorous ecotypes are present but no genetic difference was observed in a detailed study with over 5500 markers (Limborg et al., 2018).

Although it seems clear that differences in craniofacial development among the morphs is affected by genetic factors, we can not exclude the potential influence of maternal effects and the role of microenvironment at spawning sites (temperature, oxygen concentration, etc) in directing their early development. Presumably juvenile and adult habitats will also feature in shaping of their full phenotypes. Some traits may be more likely to be environmentally sensitive, such as adult size in PL-charr which has been hypothesized to be influenced to a large extent by the interaction of cohort size and food availability, and limits set by the efficiency of the method of feeding on small, crustacean plankton (Snorrason et al., 1992). Yet, considering the extensive differences in growth patterns and in size and age at maturity and results from an earlier common garden experiment involving all four morphs (Skúlason et al., 1996), it seems clear that these differences would also have some genetic basis.

But how do we discover the underlying genes and their effect on phenotypic traits? Curiously, in Atlantic salmon, a single variant in the transcription factor *vgll3* has a drastic effect on size at maturity (Barson et al., 2015). This gene was not expressed in the transcriptomes studied here, but fine-scale analyses of variants in nearby genes and or whole genome sequencing could address the question whether this gene is also

associated with maturity and size in charr. Larson et al. (2017) discovered genetic differentiation between spawning morphs of Sockeye salmon in Alaska in the transcription factor *tulp4* but did not relate the variation to a specific phenotype. The differentiation in *tulp4* was not confirmed in morphs from British Columbia (Veale and Russello, 2017). From the work presented in this thesis, three of the morphs show substantial genetic differentiation. It is, therefore, likely that a large fraction of the differences in phenotypes stems from genetic factors. But in order to link candidate genes discovered here to specific traits more work is needed. As noted above QTL-studies could help us find links between chromosomal regions and phenotypic traits such as size at maturity in charr.

5.3.2 Should we call them species?

Extensive research in the 1980s and 1990s (see Introduction) led Kottelat (1997) to give three of the morphs the status of separate species, the PL- and PI-charr as one species (*S. murta*) and SB-charr (*S. thingvallensis*) and the LB-charr (without a scientific name) as separate ones. Two of these new charr species, (*S. murta* and *S. thingvallensis*), are still listed on the International Union for Conservation of Nature (IUCN) European red list of freshwater fishes (Freyhof and Brooks, 2011). The proposed species status of the morphs and descriptions of multiple other *Salvelinus* "species" by (Kottelat and Freyhof, 2007) was received with a grain of salt by the Arctic charr community (Klemetsen, 2010) as most (including myself in this thesis) are more interested in understanding the evolutionary or developmental processes that create intraspecific variation in "the most variable vertebrate on earth" (Snorrason and Skúlason, 2004; Klemetsen, 2013).

Which species concept to use and how species are delimited (De Queiroz, 2007) has been a heated debate in biology (Wheeler and Meier, 2000). Researcher's choices of species concepts are often related to their field, philosophy, and goals. When one is working on evolutionary processes the exact criteria when to call populations species is perhaps not of great concern (Seehausen and Wagner, 2014; Lowry and Gould, 2016). However, when one wishes to use the number of species to estimate biodiversity and conserve variation in nature, matters might be different (Kottelat, 1997) as legislation does not always recognize intraspecific variation to be important for conservation (Coates et al., 2018). Modern research on evolution and variation has in some way changed our view on speciation, for instance by discovering the role of hybridization (see for example Henning and Meyer, 2014; Coates et al., 2018) leading to the argument that it is important to conserve intraspecific variation and monitor evolutionary processes (Brodersen and Seehausen, 2014; Coates et al., 2018).

The charr morphs in Lake Þingvallvatn fulfill several criteria used in many species concepts for delimitation, such as pronounced ecological, morphological and genetic separation and reproductive barriers (De Queiroz, 2007). The same applies to multiple Arctic charr morphs all over Europe (Kottelat and Freyhof, 2007). Multiple examples of sympatric morphs of Arctic charr are known (Klemetsen, 2013) and knowledge about the exact number or extent of polymorphism is probably limited (Woods et al., 2012a). Although complicated and speciose genera is not a valid excuse for simplifying taxonomy, the taxonomy should reflect phylogenetic relationships. Grouping multiple "species" with highly variable divergence and phylogenetic connectivity under one

genus does not seem like a good practice. Retaining the morph concept and informing legislators on intraspecific diversity and the evolutionary processes occurring in nature seems to be a more valid strategy for the future (Brodersen and Seehausen, 2014).

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.

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.

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.

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.

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.

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

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2):81–92.

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

Benestan, L., Moore, J.-S., Sutherland, B. J. G., Le Luyer, J., Maaroufi, H., Rougeux, C., Normandeau, E., Rycroft, N., Atema, J., Harris, L. N., Tallman, R. F., Greenwood, S. J., Clark, F. K., and Bernatchez, L. (2017). Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Molecular Ecology*, 26(24):6767–6783.

Berg, O. K., Hendry, A. P., Svendsen, B., Bech, C., Arnekleiv, J. V., and Lohrmann, A. (2001). Maternal provisioning of offspring and the use of those resources during ontogeny: Variation within and between Atlantic Salmon families. *Functional Ecology*, 15(1):13–23.

Bernatchez, L. and Landry, C. (2003). MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, 16(3):363–377.

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

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

Bhat, S., Amundsen, P.-A., Knudsen, R., Gjelland, K. O., Fevolden, S.-E., Bernatchez, L., and Præbel, K. (2014). Speciation Reversal in European Whitefish (Coregonus lavaretus (L.)) Caused by Competitor Invasion. *PLoS ONE*, 9(3):e91208.

Bourret, V., Kent, M. P., Primmer, C. R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P., Verspoor, E., Bernatchez, L., and Lien, S. (2013). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 22(3):532–551.

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

Brodersen, J. and Seehausen, O. (2014). Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. *Evolutionary Applications*, 7(9):968–983.

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

Chandler, C. H., Chari, S., and Dworkin, I. (2013). Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in Genetics*, 29(6):358–366.

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

Cloonan, N., Forrest, A. R. R., Kolle, G., Gardiner, B. B. A., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J., and Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, 5(7):613–619.

Coates, D. J., Byrne, M., and Moritz, C. (2018). Genetic Diversity and Conservation Units: Dealing With the Species-Population Continuum in the Age of Genomics. *Frontiers in Ecology and Evolution*, 6:e165.

Cousyn, C., De Meester, L., Colbourne, J. K., Brendonck, L., Verschuren, D., and

Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, 98(11):6256–6260.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer Associates, Sunderland, Mass.

da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., and Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, 30:3–13.

Dal Molin, A. and Di Camillo, B. (2018). How to design a single-cell RNA-sequencing experiment: Pitfalls, challenges and perspectives. *Briefings in Bioinformatics*, page bby007.

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.

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

de Jong, G. (2005). Evolution of phenotypic plasticity: Patters of plasticity and the emergence of ecotypes. *New Phytologist*, 166:101–118.

De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, 56(6):879–886.

de Wit, H. A. and Lindholm, M. (2010). Nutrient enrichment effects of atmospheric N deposition on biology in oligotrophic surface waters - a review. Technical Report ICP Waters reports 101/2010, Norwegian Institute for Water Research, Norway.

Dick, T. A., Gallagher, C. P., and Yang, A. (2009). Summer habitat use of Arctic char (*Salvelinus alpinus*) in a small Arctic lake, monitored by acoustic telemetry. *Ecology of Freshwater Fish*, 18(1):117–125.

Eberwine, J., Sul, J.-Y., Bartfai, T., and Kim, J. (2014). The promise of single-cell sequencing. *Nature Methods*, 11(1):25–27.

Ehrenreich, I. M. and Pfennig, D. W. (2016). Genetic assimilation: A review of its potential proximate causes and evolutionary consequences. *Annals of Botany*, 117(5):769–779.

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.

Ergon, T. and Ergon, R. (2017). When three traits make a line: Evolution of phenotypic plasticity and genetic assimilation through linear reaction norms in stochastic environments. *Journal of Evolutionary Biology*, 30(3):486–500.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4):1531–1545.

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

Freyhof, J. and Brooks, E. (2011). *European Red List of Freshwater Fishes*. IUCN

Red List of Threatened Species. Regional Assessments. Publications Office of the European Union ; IUCN, Luxembourg.

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

Futuyma, D. J. (2005). *Evolution*. Sinauer Associates, Sunderland, Mass.

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E. J., Jayasinghe, L., Wright, C., Blasco, J., Young, S., Brocklebank, D., Juul, S., Clarke, J., Heron, A. J., and Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, 15(3):201–206.

Gilbert, S. F. (2006). *Developmental Biology*. Sinauer Associates, Inc., Sunderland, Massachusetts, USA, 8th edition.

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

Gíslason, D., M Ferguson, M., Skúlason, S., and Snorrason, S. S. (1999). Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic char (*Salvelinus alpinus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 56(12):2229–2234.

Glasauer, S. M. K. and Neuhauss, S. C. F. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics*, 289(6):1045–1060.

Gomez-Uchida, D., Dunphy, K. P., O'Connell, M. F., and Ruzzante, D. E. (2008). Genetic divergence between sympatric Arctic charr *Salvelinus alpinus* morphs in Gander Lake, Newfoundland: Roles of migration, mutation and unequal effective population sizes. *Journal of Fish Biology*, 73(8):2040–2057.

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6):333–351.

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

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.

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.

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.

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

Gunnarsson, V. I. (2006). Staða bleikjueldis á Íslandi, samkeppnishæfni og stefnumótun rannsókna og þróunarstarfs. *Sjávarútvegurinn - Vefrit um sjávarútvegsmál*, 6(2).

Gunnarsson, V. I. and Guðbergsson, G. (1988). Bleikja - Eldisfiskur með framtíð? Technical Report VMST-R/88032, Institute of Freshwater Fisheries, Reykjavík, Iceland.

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

Henning, F. and Meyer, A. (2014). The Evolutionary Genomics of Cichlid Fishes: Explosive Speciation and Adaptation in the Postgenomic Era. *Annual Review of Genomics and Human Genetics*, 15(1):417–441.

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

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., and Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout: SNP DISCOVERY: NEXT GENERATION SEQUENCING. *Molecular Ecology Resources*, 11(Suppl. 1):117–122.

Holland, P. W. H., Garcia-Fernàndez, J., Williams, N. A., and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development. Supplement*, 1994:125–33.

Illumina (2016). Ribo-Zero rRNA Removal Kit Reference Guide. Document # 15066012 v02, Illumina, Inc., San Diego, California, USA.

Jacobs, A., Carruthers, M., Eckmann, R., Yohannes, E., Adams, C. E., Behrmann-Godel, J., and Elmer, K. R. (2018a). Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nature Ecology & Evolution*.

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

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Snutch, T. P., Tee, L., Paten, B., Phillippy, A. M., Simpson, J. T., Loman, N. J., and Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, 36(4):338–345.

Jónasson, P. M., Adalsteinsson, H., and Jónsson, G. S. (1992). Production and Nutrient Supply of Phytoplankton in Subarctic, Dimictic Thingvallavatn, Iceland. *Oikos*, 64(1/2):162–187.

Jonsson, B. and Jonsson, N. (2001). Polymorphism and speciation in Arctic charr. *Journal of Fish Biology*, 58(3):605–638.

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.

Jónsson, G. S. (2016). Þingvallavatn - ákoma og afrennsli, Skýrsla tekin saman fyrir umhverfis- og auðlindaráðneytið. Technical report, Umhverfis- og auðlindaráðneytið,

Reykjavík, Iceland.

Kamm, J. A., Terhorst, J., Durbin, R., and Song, Y. S. (2018). Efficiently inferring the demographic history of many populations with allele count data. *bioRxiv*.

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

Kapralova, K. H., Morrissey, M. B., Kristjánsson, B. K., Ólafsdóttir, G. A., 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.

Kjærner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemelä, E., Ozerov, M., Vähä, J. P., Glover, K. A., Rubin, C. J., Wargelius, A., and Edvardsen, R. B. (2016). Atlantic salmon populations reveal adaptive divergence of immune related genes - A duplicated genome under selection. *BMC Genomics*, 17(610):1–12.

Kjartansson, G. (1964). Aldur nokkurra hrauna á Suðulandi. *Náttúrufræðingurinn*, 34(3):101–126.

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

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

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

Koop, B. F. and Davidson, W. S. (2008). Genomics and the Genome Duplication in Salmonids. In Tsukamoto, K., Kawamura, T., Takeuchi, T., Beard, T. D., and Kaiser, M. J., editors, *Fisheries for Gloval Welfare and Environment - Memorial Book of the 5th World Fisheries Congress 2008*, pages 77–86, Tokyo, Japan. TERRAPUB.

Kottelat, M. (1997). European freshwater fishes - An heuristic checklist of the freshwater fishes of Europe (exclusive of former USSR), with an introduction for non-systematists and comments on nomenclature and conservation. *Biologia, Bratislava*, 52(Supplement 5):1–271.

Kottelat, M. and Freyhof, J. (2007). *Handbook of European Freshwater Fishes*. Kottelat and Freyhof, Cornol, Switzerland and Berlin, Germany.

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.

Kristjánsson, B. K., Skúlason, S., and Noakes, D. L. G. (2002). Morphological segregation of Icelandic threespine stickleback (*Gasterosteus aculeatus* L). *Biological Journal of the Linnean Society*, 76:247–257.

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

Kristjánsson, B. K. and Svavarsson, J. (2004). Crymostygidae, a new family of subterranean freshwater gammaridean amphipods (Crustacea) recorded from subarctic

Europe. *Journal of Natural History*, 38(15):1881–1894.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C. J., Wang, C., Zamani, N., Grant, B. R., Grant, P. R., Webster, M. T., and Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539):371–375.

Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-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.

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

Latchman, D. S. (2015). *Gene Control*. Garland Science, Taylor & Francis Group, New York, USA, 2nd edition.

Li, B. and Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1):323.

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.

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

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J., Van Oudenaarden, A., Barton, D. B., Bailes, E., Nguyen, A. N., Jones, M., Quail, M. A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., and Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, 458(7236):337–341.

Love, M. I., Hogenesch, J. B., and Irizarry, R. A. (2016). Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nature Biotechnology*, 34(12):1287–1291.

Lowry, D. and Gould, B. (2016). Speciation Continuum. In *Encyclopedia of Evolutionary Biology*, volume 4, pages 159–165. Elsevier.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., and Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 17(2):142–152.

MacKay, T. F., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., Cridland, J. M., Richardson, M. F., Anholt,

R. R., Barrón, M., Bess, C., Blankenburg, K. P., Carbone, M. A., Castellano, D., Chaboub, L., Duncan, L., Harris, Z., Javaid, M., Jayaseelan, J. C., Jhangiani, S. N., Jordan, K. W., Lara, F., Lawrence, F., Lee, S. L., Librado, P., Linheiro, R. S., Lyman, R. F., MacKey, A. J., Munidasa, M., Muzny, D. M., Nazareth, L., Newsham, I., Perales, L., Pu, L. L., Qu, C., Ràmia, M., Reid, J. G., Rollmann, S. M., Rozas, J., Saada, N., Turlapati, L., Worley, K. C., Wu, Y. Q., Yamamoto, A., Zhu, Y., Bergman, C. M., Thornton, K. R., Mittelman, D., and Gibbs, R. A. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, 482(7384):173–178.

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.

Macqueen, D. J., Primmer, C. R., Houston, R. D., Nowak, B. F., Bernatchez, L., Bergseth, S., Davidson, W. S., Gallardo-Escárate, C., Goldammer, T., Guiguen, Y., Iturra, P., Kijas, J. W., Koop, B. F., Lien, S., Maass, A., Martin, S. A., McGinnity, P., Montecino, M., Naish, K. A., Nichols, K. M., Ólafsson, K., Omholt, S. W., Palti, Y., Plastow, G. S., Rexroad, C. E., Rise, M. L., Ritchie, R. J., Sandve, S. R., Schulte, P. M., Tello, A., Vidal, R., Vik, J. O., Wargelius, A., and Yáñez, J. M. (2017). Functional Annotation of All Salmonid Genomes (FAASG): An international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics*, 18(484):1–9.

Magnanou, E., Noirot, C., Falcón, J., and Jørgensen, E. H. (2016). Sequencing and characterization of a multi-organ Arctic charr transcriptome: A toolbox for investigating polymorphism and seasonal life in a high Arctic fish. *Marine Genomics*, 29:45–53.

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.

Malmquist, H. J., Ingimarsson, F., Ingvason, H. R., Stefánsson, S. M., and Hrafnsdóttir, T. (2012). Vöktun á lífríki og vatnsgæðum Þingvallavatns - Yfirlit yfir fimm fyrstu vöktunarárin 2007–2011 og samanburður við eldri gögn. Technical Report Fjölrit nr. 3-2012, Náttúrufræðistofa Kópavogs, Kópavogur, Iceland.

Malmquist, H. J., Snorrason, S. S., and Skúlason, S. (1985). Bleikjan í Þingvallavatni - I. Fæðuhættir. *Náttúrufræðingurinn*, 55(4):195–217.

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

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

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

Mayr, E. (1954). Change of genetic environment and evolution. In Huxley, J., Hardy, A. C., and Ford, E. B., editors, *Evolution as a Process*, pages 157–180. Unwin Brothers, London.

McGuigan, K., Nishimura, N., Currey, M., Hurwit, D., and Cresko, W. A. (2011).

Cryptic genetic variation and body size evolution in threespine stickleback. *Evolution*, 65(4):1203–1211.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(2):240–248.

Moore, J.-S., Harris, L. N., Le Luyer, J., Sutherland, B. J. G., Rougemont, Q., Tallman, R. F., Fisk, A. T., and Bernatchez, L. (2017). Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char. *Molecular Ecology*, 26(24):6784–6800.

Morin, R. D., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T. J., McDonald, H., Varhol, R., Jones, S. J., and Marra, M. A. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *BioTechniques*, 45(1):81–94.

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.

NEB (2017). NEBNext rRNA Depletion Kit (Human/Mouse/Rat). Instruction Manual Version 4.0, New England BioLabs, Inc., Ipswich, Massachusetts, USA.

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

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.

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

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.

Nosil, P., Feder, J. L., Flaxman, S. M., and Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature Ecology & Evolution*, 1(2):0001.

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.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Paaby, A. B. and Rockman, M. V. (2014). Cryptic genetic variation: Evolution's hidden substrate. *Nature Reviews Genetics*, 15(4):247–258.

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.

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.

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., and Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology and Evolution*, 25(8):459–467.

Power, M., Power, G., Reist, J. D., and Bajno, R. (2009). Ecological and genetic differentiation among the Arctic charr of lake Aigneau, northern Québec. *Ecology of Freshwater Fish*, 18(3):445–460.

Prokkola, J. M., Nikinmaa, M., Lewis, M., Anttila, K., Kanerva, M., Ikkala, K., Seppänen, E., Kolari, I., and Leder, E. H. (2018). Cold temperature represses daily rhythms in the liver transcriptome of a stenothermal teleost under decreasing day length. *The Journal of Experimental Biology*, page jeb.170670.

Recknagel, H., Jacobs, A., Herzyk, P., and Elmer, K. R. (2015). Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Molecular Ecology Resources*, 15(6):1316–1329.

Reisinger, E. (1953). Zum Saiblingsproblem. *Carinthia II*, 143/63(2):74–102.

Reist, J. D., Power, M., and Dempson, J. B. (2013). Arctic charr (*Salvelinus alpinus*): A case study of the importance of understanding biodiversity and taxonomic issues in northern fishes. *Biodiversity*, 14(1):45–56.

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

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.

Rogers, S. M. and Bernatchez, L. (2006). The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeaformis*). *Journal of Evolutionary Biology*, 19(6):1979–1994.

Sæmundsson, B. (1897). Um fiskirannsóknir 1896. *Andvari*, 22(1):96–172.

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

Sæmundsson, B. (1908). *Oversigt over Islands fiske : med oplysninger om deres forekomst, vigtigste biologiske forhold og økonomiske betydning*. Number 5 in Skrifter / udgivne af Kommissionen for Havundersøgelser. Kommissioner for Havundersøgelser, København.

Sæmundsson, B. (1917). Fiskirannsóknir 1915 og 1916. *Andvari*, 42(1):71–129.

Sæmundsson, B. (1926). *Fiskarnir*. Number I in Íslensk Dýr. Bókaverzlun Sigfúsar Eymundssonar, Reykjavík, Iceland.

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

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., and Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, 26(17):4378–4390.

Sánchez, C. C., Smith, T. P. L., Wiedmann, R. T., Vallejo, R. L., Salem, M., Yao, J., and Rexroad, C. E. (2009). Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics*, 10:1–8.

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

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution*, 16(7):372–380.

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28(8):1086–1092.

Seehausen, O. (1997). Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. *Science*, 277(5333):1808–1811.

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

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

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*, 33(1):1–14.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. *Current Opinion in Genetics and Development*, 6(6):715–722.

Sigurgeirsson, B., Emanuelsson, O., and Lundeberg, J. (2014). Sequencing Degraded RNA Addressed by 3' Tag Counting. *PLoS ONE*, 9(3):e91851.

Sigurjónsdóttir, H. and Gunnarsson, K. (1989). Alternative mating tactics of arctic charr, *Salvelinus alpinus*, in Thingvallavatn, Iceland. *Environmental Biology of Fishes*, 26(3):159–176.

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

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

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.

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.

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

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.

Snorrason, S. (2011). Effects of geothermal effluents on macrobenthic communities in a pristine sub-arctic lake. *Inland Waters*, 1(3):146–157.

Snorrason, S. S., Jónasson, P. M., Jonsson, B., Lindem, T., Malmquist, H. J., Sandlund, O. T., and Skúlason, S. (1992). Population dynamics of the planktivorous arctic charr *Salvelinus alpinus* ("murta") in Thingvallavatn. *Oikos*, 64(1/2):352–364.

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.

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.

Spence, J. P., Steinrücken, M., Terhorst, J., and Song, Y. S. (2018). Inference of population history using coalescent HMMs: Review and outlook. *Current Opinion in Genetics & Development*, 53:70–76.

Sturlaugsson, J. and Malmquist, H. J. (2011). Brown trout in Lake Thingvallavatn. In Hersteinsson, P. and Jónasson, P. M., editors, *Thingvallavatn - A Unique World Evolving*, pages 200–206. Opna Publishing, Reykjavík, Iceland.

Svavarsson, E. (2007). árangur í kynbótum á bleikju og næstu skref. In *Fræðaþing Landbúnaðarsins 4*, pages 121–125.

Svavarsson, J. and Kristjánsson, B. K. (2006). *Crangonyx Islandicus* sp. nov., a subterranean freshwater amphipod (Crustacea, Amphipoda, Crangonyctidae) from springs in lava fields in Iceland. *Zootaxa*, 1365:1–17.

Sveinsson, T. and Hara, T. J. (1995). Mature males of Arctic charr, *Salvelinus alpinus*, release F-type prostaglandins to attract conspecific mature females and stimulate their spawning behaviour. *Environmental Biology of Fishes*, 42(3):253–266.

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., and Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2):343–355.

Tschopp, P. and Duboule, D. (2011). A Genetic Approach to the Transcriptional Regulation of *Hox* Gene Clusters. *Annual Review of Genetics*, 45(1):145–166.

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

Veale, A. J. and Russello, M. A. (2017). Genomic Changes Associated with Reproductive and Migratory Ecotypes in Sockeye Salmon (Oncorhynchus nerka). *Genome Biology and Evolution*, 9(10):2921–2939.

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

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

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.

von Linné, C. (1758). *Systema Naturæ per Regna Tria Naturæ, Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Differentiis, Synonymis, Locis*. Lau-

rentius Salvius, Holmiæ, 10th edition.

Wang, L., Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J. E., Dasari, S., Vedell, P. T., Barman, P., Wang, L., Weinshiboum, R., Jen, J., Huang, H., Kohli, M., and Kocher, J. P. A. (2016). Measure transcript integrity using RNA-seq data. *BMC Bioinformatics*, 17(1):1–16.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., and Losick, R. (2004). Chapter 21: Model Organisms. In *Molecular Biology of the Gene*, pages 681–711. Pearson / Benjamin Cummings, San Francisco, California, 5th international edition.

West-Eberhard, M. J. (1989). Phenotypic Plasticity and the Origins of Diversity. *Annual Review of Ecology and Systematics*, 20:249–278.

Wetang'ula, G. N. and Snorrason, S. (2005). Geothermal Wastewater Disposal: Chemical Stress Assessment - Lake Thingvallavatn, Iceland. In *Proceedings World Geothermal Congress 2005*, Antalya, Turkey.

Wheeler, Q. D. and Meier, R., editors (2000). *Species Concepts and Phylogenetic Theory: A Debate*. Columbia University Press, New York.

Willacker, J. J., von Hippel, F. A., Wilton, P. R., and Walton, K. M. (2010). Classification of threespine stickleback along the benthic – limnetic axis. *Biological Journal of the Linnean Society*, 101(3):595–608.

Willis, S. C., Hollenbeck, C. M., Puritz, J. B., Gold, J. R., and Portnoy, D. S. (2017). Haplotyping RAD loci: An efficient method to filter paralogs and account for physical linkage. *Molecular Ecology Resources*, 17(5):955–965.

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

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

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3):206–216.

Paper I

Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr

Ehsan Pashay Ahi, Jóhannes Guðbrandsson, Kalina H. Kapralova, Sigríður R. Franzdóttir, Sigurður S. Snorrason, Valerie H. Maier, Zophonías O. Jónsson

PLOS ONE **8(6):** e66389

JG analyzed the RNA-seq data, assisted with statistical analyses of qPCR data and contributed to the writing.

OPEN ACCESS Freely available online

PLOS ONE

Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr

Ehsan Pashay Ahi[1], Jóhannes Guðbrandsson[1], Kalina H. Kapralova[1], Sigríður R. Franzdóttir[2], Sigurður S. Snorrason[1], Valerie H. Maier[1]*, Zophonías O. Jónsson[1]

1 Institute of Life- and Environmental Sciences, University of Iceland, Reykjavik, Iceland, **2** Biomedical Centre, University of Iceland, Reykjavik, Iceland

Abstract

Arctic charr (*Salvelinus alpinus*) is a highly polymorphic species and in Lake Thingvallavatn, Iceland, four phenotypic morphs have evolved. These differences in morphology, especially in craniofacial structures are already apparent during embryonic development, indicating that genes important in the formation of the craniofacial features are expressed differentially between the morphs. In order to generate tools to examine these expression differences in Arctic charr, the aim of the present study was to identify reference genes for quantitative real-time PCR (qPCR). The specific aim was to select reference genes which are able to detect very small expression differences among different morphs. We selected twelve candidate reference genes from the literature, identified corresponding charr sequences using data derived from transcriptome sequencing (RNA-seq) and examined their expression using qPCR. Many of the candidate reference genes were found to be stably expressed, yet their quality-rank as reference genes varied considerably depending on the type of analysis used. In addition to commonly used software for reference gene validation, we used classical statistics to evaluate expression profiles avoiding a bias for reference genes with similar expression patterns (co-regulation). Based on these analyses we chose three reference genes, ACTB, UB2L3 and IF5A1 for further evaluation. Their consistency was assessed in an expression study of three known craniofacially expressed genes, sparc (or osteonectin), matrix metalloprotease 2 (mmp2) and sox9 (sex-determining region Y box 9 protein) using qPCR in embryo heads derived from four charr groups at three developmental time points. The three reference genes were found to be very suitable for studying expression differences between the morphotypes, enabling robust detection of small relative expression changes during charr development. Further, the results showed that sparc and mmp2 are differentially expressed in embryos of different Arctic charr morphotypes.

Citation: Ahi EP, Guðbrandsson J, Kapralova KH, Franzdóttir SR, Snorrason SS, et al. (2013) Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr. PLoS ONE 8(6): e66389. doi:10.1371/journal.pone.0066389

Editor: Zhang Zhang, Beijing Institute of Genomics, Chinese Academy of Sciences, China

Received February 8, 2013; **Accepted** May 5, 2013; **Published** June 13, 2013

Funding: This project was supported by The Icelandic Centre for Research (RANNIS #100204011) and the University of Iceland Research Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: valerie@hi.is

Introduction

The head of teleost fish, particularly their trophic apparatus, contains many movable elements which make it one of the most complex and integrated musculo-skeletal systems in vertebrates [1]. The development of these elements requires complicated interactions between derivatives of all three germ layers in setting up and tuning the relevant molecular pathways [2]. Fish exhibit tremendous functional diversity in their craniofacial skeleton and provide an interesting model for studying evolution of those features using developmental genetics. Phenotypic variation in the shape and formation of the trophic apparatus among related species has been most extensively studied in cichlids and zebrafish [1,3,4]. Intraspecies comparisons are also of great interest here, especially in systems where phenotypically distinct morphs have evolved [5–7]. Arctic charr (*Salvelinus alpinus*) is amongst the most thoroughly studied systems of polymorphism in fish [8]. In Lake Thingvallavatn in Iceland four residential morphs of Arctic charr are found, large benthivorous (LB), small benthivorous (SB), a pelagic planktivorous (PL) and piscivorous (PI) charr [9]. The morphs differ in diet, morphology, behaviour and life history characteristics [9–12]. The two benthivorous morphs, feeding largely on snails, have an overshot mouth (a benthic morphotype), while the pelagic morph, feeding mainly on zooplankton, and the piscivorous morph have a terminal mouth (a pelagic morphotype) [11]. The adaptive nature of morph formation among the Arctic charr of Lake Thingvallavatn has been demonstrated in a series of laboratory rearing experiments [13–15]. These studies show a strong genetic component with a significant maternal effect on the development of trophic morphology and feeding behaviour. On a population level, recent studies of 10 microsatellite markers in the two most abundant morphs have demonstrated low, but significant, genetic differentiation between these morphs, consistent with strong reproductive isolation throughout the Holocene [16].

Heterochrony is thought to be an important mechanism in the evolution of the morphs illustrated in a study where bones in the small benthivorous morph were shown to start ossifying earlier and/or faster than in the pelagic morph [15]. The low level of genetic differentiation amongst the morphs, but distinct phenotypic differences suggests a mechanism based on a few regulatory factors operating early in the development of key trophic traits. To date little is known about the expression of such regulatory elements in charr. Differences in temporal expression of the

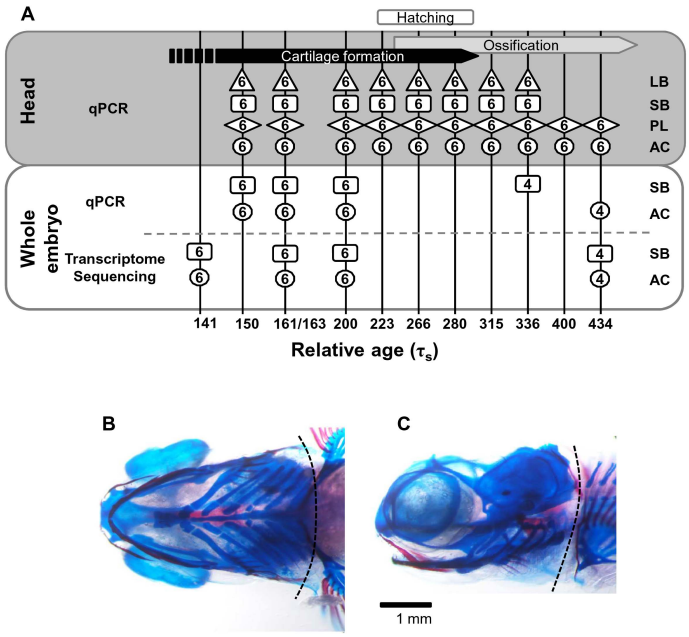


Figure 1. A scheme of sampling and analyses of Arctic charr development. (**A**) Embryos were collected at the indicated relative age (represented by vertical lines). Either whole embryos or heads of the indicated charr groups were used for RNA extraction (LB: large benthivorous charr; SB: small benthivorous charr; PL: planktivorous charr; AC: aquaculture charr). The numbers (in boxes, circles and diamonds) indicate the number of individuals pooled for each extraction. The RNA was reverse-transcribed and the cDNA used for qPCR or transcriptome sequencing as shown. Bars at top display approximate time points of cartilage formation, ossification and hatching (unpublished data). (**B–C**) Ventral and lateral views of a planktivorous head at 336 τ_s with a dashed line representing the decapitation line in front of the pectoral fin for head sample collection. Embryos were stained with alcian blue (for cartilage) and alizarin red (for bone) according to the described method [74] with some alterations.
doi:10.1371/journal.pone.0066389.g001

transcription factor Pax 7 between SB and LB morphs have been observed [17], but large scale comparisons at the expression level are lacking.

In recent years high-throughput transcriptome sequencing (RNA-Seq) has emerged as a way to profile gene expression [18] and can be used as a powerful tool to contrast expression for example in the pelagic and benthic morphotypes. The method has clear advantages. There is no need for probes designed from previously-known transcripts, novel transcripts can be detected in organisms without a sequenced genome, such as Arctic charr, and expression levels can be quantified [18]. Expression profiles from RNA-Seq data can be validated using PCR based approaches [19,20]. Quantitative real-time PCR (qPCR) is a widely used method to study gene expression and a cost effective way to examine expression patterns of key candidate genes, e.g. genes identified from RNA-seq developmental profiling. Measuring and comparing expression levels of genes of interest requires normalisation against the expression levels of reference genes [21,22]. Some of the classical reference genes, e.g. the genes encoding ACTB, GAPDH, EF1α and ribosomal proteins, have been examined in fish [23–29]. There is however general agreement that no perfect reference gene exists for assessing differential expression levels at various developmental stages in different tissues, body compartments and organisms [30]. Therefore the validation of reference genes under defined experimental conditions or at defined developmental stages is crucial [31,32].

The aim of this study was to identify and validate stable reference genes for quantitative real-time PCR during craniofacial development of Arctic charr embryos. To this end we selected a number of potential reference genes originating from independent pathways and based on previous studies in other fish species [23–29]. We used published sequences of these genes and transcriptome sequencing data obtained from Arctic charr embryos to design primers for quantitative real-time PCR. Candidate reference genes were analysed in a number of ways; (i) by testing for stability in expression levels among developmental time points and Arctic charr groups, (ii) by testing for consensus between expression levels derived from qPCR and RNA-seq data, and (iii) by testing consistency of results, when used to normalise expression levels of three developmental genes.

Methods

Sampling of parents and setting up of embryo groups

The embryo series come from four parental groups sampled in two years. In 2009 embryos from the Holar aquaculture stock (AC) and the small benthivorous charr (SB) from Lake Thingvallavatn were collected. In 2010 embryos from AC and SB were collected as well as the small planktivorous (PL) and the large benthivorous (LB) charr from Lake Thingvallavatn. Fishing permissions were obtained from the Thingvellir National Park Commission and the owner of the farm Mjóanes. Fish were killed by a sharp blow to the head and for each group, eggs from several females were pooled and fertilized using milt from several males from the same group. Eggs were reared at approximately 5°C in a hatching tray (EWOS, Norway) under constant water flow and in complete darkness at the Holar University College experimental facilities in Verið, Sauðárkrókur. The water temperature was recorded twice daily and the average was used to estimate the relative age of the embryos using tau-somite (τ_s) units defined as the time it takes for one somite pair to form at a given temperature [33]. Embryos were collected directly into RNA-later solution (Ambion) at the indicated relative age (Figure 1) and stored at −20°C until further use. The rearing and collection of the embryos was performed according to Icelandic regulations (licence granted to Holar University College aquaculture and experimental facilities in Verið, Sauðárkrókur).

RNA extraction and cDNA synthesis

Embryos were dechorionated under the light microscope (Leica S6E) and the yolk was discarded. Embryos sampled in 2009 were used for transcriptome sequencing. SB and AC whole embryos, at the relative age indicated in Figure 1A (lower panel, sequencing), were homogenized with a disposable Kontes Pellet Pestle Cordless Motor tissue grinder (Kimble Kontes) and RNA was extracted into two size-fractions using the mirVana kit (Ambion). The high molecular weight fraction was further used for mRNA-seq and RNA quality was analysed using an Agilent 2100 Bioanalyzer (Agilent Technologies). First and 2nd strand cDNA synthesis, fragmentation, adapter ligation and amplification were performed using the mRNA-Seq 8-Sample Prep Kit (Illumina) according to manufacturer's instructions.

The embryos reared in 2010 were used for real-time PCR analysis. For RNA extraction from heads, embryos were dechorionated and then decapitated in front of the pectoral fin (Figure 1 B–C). Whole embryos or heads were placed in TRI Reagent (Sigma) and homogenized as described above. RNA was prepared according to manufacturer's instructions and dissolved in 30 µl RNase-free water. To minimise DNA contamination, RNA was treated with DNase (New England Biolabs). Quantity and quality of the resulting RNA were assessed using a NanoDrop ND-1000 UV/Vis-Spectrophotometer (NanoDrop Technologies). The quality of the RNA from half of the samples was further evaluated by agarose gel electrophoresis or using Bioanalyzer. All samples displayed intact 28 S and 18 S rRNA without high molecular weight genomic DNA contamination. cDNA was prepared from 200 ng of RNA using the High capacity cDNA RT kit (Applied Biosystems), according to manufacturer's protocol. The absence of genomic DNA was confirmed by preparing several samples without addition of reverse transcriptase. cDNA was diluted 5 fold in water for further use in quantitative real-time PCR.

RNA sequencing and assembly of Arctic charr reference gene homologues

The whole mRNA transcriptome from the two 2009 charr groups (AC and SB- charr) at 4 developmental time points was sequenced, yielding single end reads of 36 base pairs. Sequencing was performed at DeCode genetics (Reykjavík, Iceland) using SOLEXA GAII technology (Illumina Inc.) and the sequencing depth ranged from 49 to 58 million reads with a mean depth of 55 Million reads per sample. The reads were pre-assembled into contigs using Velvet assembler [34], and further assembly steps were performed in CLC Genomics Workbench (CLC bio, Aarhus, Denmark). In order to obtain sequences for Arctic charr reference genes we selected likely candidates from related species, i.e. ESTs and FLIcs (full-length sequenced inserts from cDNAs) from Atlantic salmon (*Salmo salar*) and rainbow trout, (*Oncorhynchus mykiss*) [35–37]. The selected reference candidates were used for reference assembly of the charr homologues. The individual nucleotide mismatch score and the total mismatch score limit was set to 98% identity. All 12 consensus sequences of the charr candidate reference genes (Table 1) and the three developmental genes examined in this study have been deposited at NCBI. GenBank Accession numbers for sparc, sox9a and mmp are KC538874, JQ624876 and KC538875, respectively.

To quantify the expression levels of the candidate reference genes, reads were aligned to salmon mRNA sequences (total of 16727 sequences) from the NCBI-nucleotide database, using bowtie, version 0.12.7 [38]. The number of reads for each sequence was extracted using a python script. Subsequently a filter step was performed to exclude sequences that had less than 20 reads aligned and sequences to which only reads from post hatching samples aligned, in order to be able to estimate parameters in subsequent regression analysis. 15396 sequences passed this filtering step. Reads per million aligned per kilobase (RPMK-values) were calculated as expression measurements for the genes. Mean, standard deviation and coefficient of variation were calculated for the RPMK-values of the candidate reference genes

Primer design

The assembled Arctic charr consensus sequences were used to design primers for the candidate reference and developmental genes. We aimed to make qPCR primers overlapping exon boundaries or located in separate exons (Table S1 in File S1). As the charr genome has not been sequenced but gene exon/intron boundaries are for the most part well conserved between orthologues [39], the exon/intron borders were assumed to be similar to zebrafish. The NCBI Spidey software (www.ncbi.nlm.nih.gov/spidey) was used and the consensus sequences of Arctic charr candidate genes were aligned against zebrafish homologue genes which were retrieved from the Ensemble database (http://www.ensembl.org/Danio_rerio). Primers were designed using OligoPerfect Designer (Invitrogen) and Primer Express 3.0 software (Applied Biosystems, Foster City, CA, USA). Primers were also checked for self-annealing, hetero-dimers and hairpin structures by OligoAnalyzer 3.1 (Integrated DNA Technology).

Real-time quantitative PCR

Real-time PCR was performed in 96 well-PCR plates on an ABI 7500 real-time PCR System (Applied Biosystems) using Power SYBR green PCR Master Mix as recommended by the manufacturer (Applied Biosystems) with the exception of using 10 µl final reaction volume. Reactions were run in duplicate together with no-template control (NTC) in each run for each

Table 1. The reference genes selected in this study, their abbreviation and putative function.

Gene Name	Symbol	Function	Accession no. Arctic charr
Actin, cytoplasmic 1(beta-actin)	ACTB	Cytoskeletal structure protein	JR540730
beta-2-microglobulin	b2m/B2MG	Beta chain of major histocompatibility complex	JR540731
elongation factor 1 alpha	EF1α	Protein synthesis	JR540732
Glyceraldehyde-3-phosphate dehydrogenase	G3P/GAPDH	Glycolytic protein	JR540733
Hypoxanthine-guanine phosphoribosyltransferase	HPRT	Enzyme in purine metabolic pathway	JR540734
Eukaryotic initiation factor 5A isoform 1	IF5A1	Protein synthesis	JR540735
60S ribosomal protein L7	RL7	Member of ribosome proteins	JR540736
40S ribosomal protein S9	RS9	Member of ribosome proteins	JR540737
Ribosomal Protein S20	RS20	Member of ribosome proteins	JR540738
Tubulin alpha chain	TBA	Cytoskeletal protein	JR540739
Ubiquitin-conjugating enzyme E2 L3	UB2L3	Protein degradation	JR540740
Ubiquitin	UBIQ	Protein degradation	JR540741

doi:10.1371/journal.pone.0066389.t001

gene. Experimental set-up per run followed the preferred sample maximization method described by Hellemans et al. [40], in order to decrease run-to-run variation. The PCR was started with a 2 min hold at 50°C followed by a 10 min hot start at 95°C. Subsequently the amplification was performed with 40 cycles of 15 sec denaturation at 95°C and 1 min annealing/extension at 60°C. For each sample a dissociation step (60°C–95°C) was performed at the end of the amplification phase to identify a single, specific melting temperature for each primer set (Table S1 in File S1). Primer efficiencies (E) were calculated with 7 points of 2–4 fold serial dilutions using pooled cDNA (700 ng RNA input) from different developmental stages as well as different charr groups. The slope of the standard curve in the equation; $E\% = (10^{1/slope} - 1)\times 100$, was used for PCR efficiency calculation [41]. The range of PCR efficiencies and linear correlation coefficient (R^2) are shown in Table S1 in File S1. The background-corrected fluorescence values from the real-time PCR were imported into LinReg PCR software [42] and the individual PCR efficiency of each reaction was determined.

Data analysis

(i) Ranking candidate reference genes. First we employed three ranking analyses to detect the most stably expressed candidates. Two Excel-based programs BestKeeper [43] and NormFinder [44], as well as an improved version of the classical geNorm [45], called GeNorm^{Plus} (Biogazelle, Ghent, Netherlands). Two sources of input were used for the analysis with BestKeeper. The raw Cq values and the logarithmic N_0 values calculated by LinReg PCR [42], which takes the individual qPCR efficiency into account. The standard deviation (S.D.) based on Cq values of the developmental stages and groups was calculated by BestKeeper to determine the expression variation for each reference gene. A standard deviation higher than 1 leads to the rejection of the candidate as reference gene. In addition, BestKeeper determines the stability of reference genes based on correlation to other candidates through calculation of BestKeeper index (B.I.). In order to decrease the effect of co-regulation in BestKeeper, we used the average B.I. for all genes compared to the first four, three and two genes with lowest standard deviations. For GeNorm and NormFinder the relative expression ratios are used as input. GeNorm measures the expression stability (*M* value) which is the mean pairwise variation between each gene and other candidates and it excludes the gene with the highest *M* value (least stable) from subsequent analysis in a stepwise manner. GeNorm^{Plus} on the other hand is able to determine the best among the last two remaining genes. GeNorm assumes candidates with $M>1.5$ as unreliable and $M<0.5$ is characteristic for very stable reference genes. GeNorm^{Plus} can also determine the optimal number of reference genes to use. For this a pairwise variation coefficient Vn/n+1 between two sequential normalisation factors (NFn and NFn+1) is calculated and extra reference genes are added until the variation drops below the recommended threshold of 0.15 [45]. NormFinder ranks the most stable genes (lowest expression stability values) based on analysis of the sample subgroups (time and charr group) and estimation of inter- and intra-group variation in expression levels.

Secondly an analysis of variance (ANOVA) followed by post hoc Tukey's honest significant difference (HSD) test was implemented using relative expression ratios in R [46] (http://www.r-project.org/) with developmental time points and progeny groups as categorical variables. Relative expression ratios were calculated using primer efficiencies (E). For this the highest expressed sample point (the lowest Cq = *Min Cq*) in each primer pair was set to one and the other sampling points were calculated in relation to Min Cq, according to $E^{\Delta Cq}$, where ΔCq = Min Cq–Cq sample. The best reference genes were considered those that showed no (or very little) significant difference in relative expression among time points or embryo groups.

(ii) Comparison of qPCR and RNA-seq data. To compare the RNA-seq and qPCR expression data we tested the correlation between RPMK (transformed to a log2-scale) and quantification cycle (Cq)-values, for the 8 corresponding samples for group and relative age (Figure 1A, lower panel), using linear regression on the RPMK values.

A generalized linear model was applied to the raw RNA-seq read counts and tested for group and time effect with a likelihood ratio test using the edgeR-package in R [47]. The coefficient of variation was also calculated as a stability measure.

(iii) Consistency of normalisation. Expression levels of the three putative developmental genes, sparc, mmp2, and sox9a, in four charr groups at three developmental time points were calculated using either individual reference genes or a combination of reference genes (i.e. geometric average Cq of two and three

reference genes) for normalisation. Fold changes were calculated by comparing expression in three charr morphs from Thingvallavatn (SB, LB and PL) to expression in aquaculture charr. The consistency of normalisation with the three reference genes was examined running three full model ANOVAs (one for each developmental gene), especially examining the interaction terms for (morph)×(developmental time point)×(reference gene) and (morph)×(reference gene). Furthermore we tested for consistency in reflecting variation (as measured by coefficients of variation, CVs, for two biological replicates) of the mean expression levels for each developmental gene, morph and developmental time point. This was done by testing for positive correlations of the CVs between the three reference genes. Statistical differences between benthic (SB and LB) and pelagic (PL and AC) groups in the expression of target genes were determined using Student's t test.

Results

In order to identify key regulator elements responsible for the phenotypic differences in the charr morphotypes we sequenced the transcriptome of two charr groups at four points during development. We focused on the developmental stages covering cartilage formation and the beginning of ossification (Figure 1). Transcriptome sequencing was carried out on RNA from whole embryos. In the present study this data was used to examine the expression of selected reference genes and to compare these results to the qPCR data.

Twelve possible reference genes originating from independent pathways were selected for validation, based on previous studies in other fish species [23–29] (Table 1). The sequences of these genes from salmon or other fish species were used for reference based assembly of the corresponding charr mRNA sequences. Quantitative real-time PCR primers were designed (Table S1 in File S1) and, with the exception of the GAPDH primers, all primer efficiencies were shown to lie within the 90–110% range. Melting curve analysis revealed the absence of primer dimers and different size of amplification products for all primer pairs

Transcription profiling of candidate reference genes

Quantitative real-time PCR for the 12 reference gene candidates was performed on cDNA generated from head and whole embryo samples as described in Figure 1. Candidate reference gene expression levels during head development were profiled in the charr groups using Cq values (Figure 2). Out of the 12 genes, GAPDH showed increasing expression during development, which, combined with high primer efficiency, led us to reject this gene as a reference gene.

The remaining eleven candidates covered a broad range of expression levels, varying from ACTB, with the highest expression (lowest Cq) (Figure S1A in File S2,), to b2m with the lowest expression (highest Cq). When the raw Cq values were transformed to relative expression ratios, seven genes showed significant difference ($P<0.05$) in expression between head and whole embryo (Figure S1B in File S2), illustrating the differences in expression of genes between different body parts and the importance of validating reference genes in the tissue of interest. Interestingly some genes had lower, other higher expression in head compared to whole embryo indicating the robustness of the reference genes chosen.

Reference gene analyses and ranking

The candidate reference genes were ranked using three known algorithms (BestKeeper, GeNorm and NormFinder) and based on standard deviation (SD). For simplicity the ranking of genes using all 4 methods is only shown in Table 2, while detailed results can be found in supplementary tables (Tables S2 in File S1). In charr heads TBA was found to be the most stably expressed reference gene across all charr groups as well as within each group (Table S3 in File S1). ACTB was shown to be the second most stable gene in both analyses and UBIQ and UB2L3 were among the 4 best reference genes both in heads in general and when examining each group separately. GeNorm suggested the use of only two reference genes to be sufficient for accurate normalisation (Figure S2 in File S2). This data reflects the high stability of the candidate reference genes expression and suggests that TBA and ACTB are sufficient and suitable reference genes to quantify gene expression in Arctic charr heads.

We further performed a two-way ANOVA followed by a Tukey's test (Figure 3) to select reference genes which are not expressed at significantly different levels between group/time – the main criteria for a stable reference gene. These analyses identified six candidate reference genes that are stably expressed between groups, but of those only ACTB showed constant expression during the developmental stages examined. The post hoc Tukey's test revealed the expression pattern of the genes over the time examined. Several genes, e.g., ribosomal protein genes, EF1α and UB2L3, were found to be highly expressed at the earlier stages, while others were more highly expressed later in development e.g. b2m. Based on these results ACTB was found to be the overall most stable reference gene both over time and between the 4 different groups.

Testing consensus of transcriptome and qPCR data

RNA-seq and qPCR were used to estimate the expression levels of the candidate reference genes in whole embryos (samples used see Figure 4 insert) and the two methods were compared. As expected the expression estimates from RNA-seq data correlated significantly with the expression estimates from qPCR ($p<1^{-10}$) (Figure 4).

Candidate reference genes were ranked for expression stability in whole embryos using both the qPCR and the RNA-seq data (Table 3). Overall UB2L3 and ACTB were found to be most stable. Furthermore, UB2L3 showed no significant differences between groups or during development, as determined using a likelihood-ratio test, qualifying this gene as the best reference gene in whole embryos. Interestingly UB2L3, which was one of the four best reference genes in head samples (Table 2), is also the best reference gene for comparing head and whole embryo gene expression (Figure S1B in File S2).

Consistency of normalised expression levels of three craniofacial target genes

For a test-run of our validated qPCR reference genes, we selected three genes which have a well established craniofacial expression pattern during zebrafish development [48–50] and showed expression differences between charr groups in our transcriptome data. Arctic charr homologues of sparc, mmp2 and sox9a all showed elevated expression in SB compared to AC at 200 τ_s (unpublished data). The expression of these genes was examined in the head of all four charr groups and at three developmental time points. To normalise the qPCR data, we used ACTB, UB2L3 and IF5A1 separately, the geometric average expression of all three genes (NF = 3), or the geometric average expression of IF5A1 and ACTB alone (NF = 2) (Figure 5). The three ANOVAs of normalised expression values show that expression patterns among developmental time points and morphs are the same for all three reference genes. P values for the three-way interaction term involving the reference genes were non-

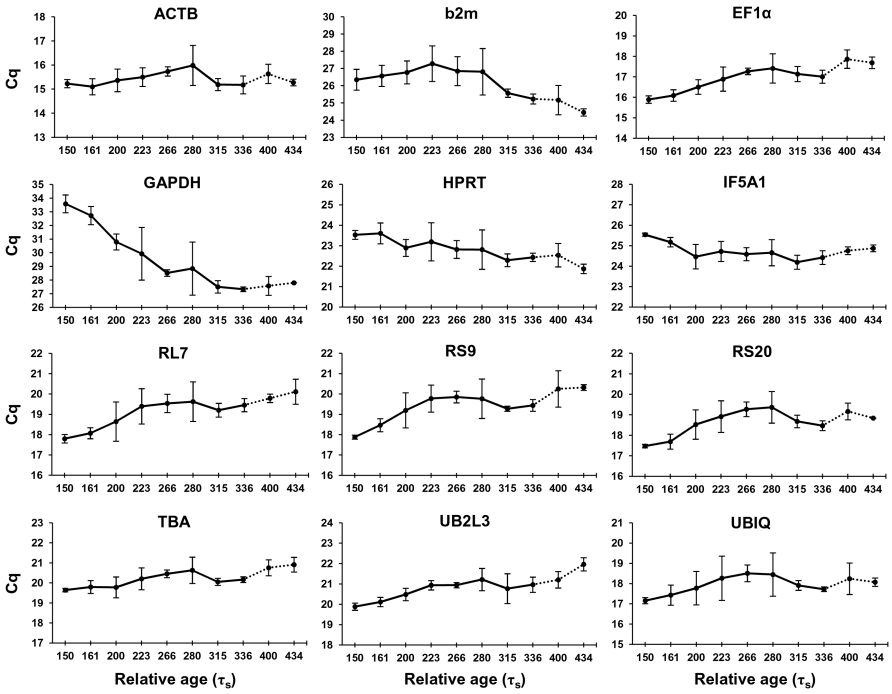


Figure 2. Expression levels of reference genes in the head of four charr groups during development. Expression profiles of 12 candidate reference genes based on quantitative real time PCR performed on embryonic heads from four charr groups at the relative ages 150 to 434 τ_s. Expression levels are shown as mean Cq (quantification cycle) values in the four charr groups at corresponding relative age, except for the two last time points (dashed line), which are based on samples of only two groups(AC and PL charr). Error bars represent standard deviation.
doi:10.1371/journal.pone.0066389.g002

Table 2. Ranking of the candidate reference genes in Arctic charr head homogenates using BestKeeper (BK), geNorm (gN), NormFinder (Nf) and standard deviation (SD).

Gene	BK	gN	Nf	SD
ACTB	3	8	1	2
b2m	9	11	11	11
EF1α	5	3	7	5
HPRT	10	10	4	6
IF5A1	11	9	5	3
RL7	7	5	10	10
RS9	4	4	9	9
RS20	6	6	8	8
TBA	1	1	2	1
UB2L3	8	2	6	4
UBIQ	2	7	3	7

doi:10.1371/journal.pone.0066389.t002

significant in all three ANOVAs ($P = 0.9961$, $P = 0.5895$, $P = 0.7715$ for sox9a, mmp2 and sparc, respectively). Furthermore, the CVs of mean expression values showed significant correlations among reference genes (Correlation coefficients and Bonferroni adjusted P values: $r = 0.692$, $P<0.001$ for IF5A1 versus UB2L3; $r = 0.556$, $P<0.008$ for IF5A1 versus ACTB; $r = 0.557$, $P<0.008$ for ACTB versus UB2L3). Interestingly, sparc and mmp2 showed significantly higher expression in the heads of the two benthic morphs (SB and LB) compared to the AC and PL groups at all three time points (Figure 5). Standard deviations of the normalised expression levels were generally low and the expression differences between morphotypes were consistent among the three reference genes (Figure 5) showing the robustness of their use for normalisation. In the case of sparc all normalisation methods detected a 1.7 fold difference in expression levels between the morphotypes. Hence, we can conclude that all three genes are suitable as reference genes for qPCR studies of Arctic charr development. Although the use of one reference gene already gave consistent results, the geometric mean of two or three reference genes further decreased variations in the relative expression.

Gene	ANOVA groups	HSD groups	ANOVA time	HSD time (τ_a)
ACTB	$P = 0.575$	*NS*	$P = 0.151$	*NS*
b2m	$P = 0.004$	AC PL SB LB	$P = 3.7*10^{-5}$	150 161 200 223 266 280 315 336
EF1α	$P = 0.340$	*NS*	$P = 3.1*10^{-6}$	150 161 200 223 266 280 315 336
HPRT	$P = 0.052$	*NS*	$P = 0.007$	150 161 200 223 266 280 315 336
IF5A1	$P = 0.219$	*NS*	$P = 0.008$	150 161 200 223 266 280 315 336
RL7	$P = 0.001$	AC PL SB LB	$P = 2.2*10^{-5}$	150 161 200 223 266 280 315 336
RS9	$P = 0.027$	AC PL SB LB	$P = 1.2*10^{-8}$	150 161 200 223 266 280 315 336
RS20	$P = 0.089$	*NS*	$P = 9.5*10^{-7}$	150 161 200 223 266 280 315 336
TBA	$P = 0.026$	AC PL SB LB	$P = 0.001$	150 161 200 223 266 280 315 336
UB2L3	$P = 0.578$	*NS*	$P = 2.7*10^{-5}$	150 161 200 223 266 280 315 336
UBIQ	$P = 0.043$	AC PL SB LB	$P = 0.008$	150 161 200 223 266 280 315 336

Figure 3. Candidate reference gene expression differences and patterns in the head of Arctic charr groups during development. Relative expression ratios, calculated from the qPCR data, were subjected to an analysis of variance (ANOVA) to test the expression differences amongst four charr groups and eight time points (numbers are relative age in τ_a). Subsequently a post hoc Tukey's honestly significant difference test (HSD) was performed to analyse the expression pattern of candidates in groups and during development. White boxes represent low expression, while black boxes represent high expression. A two or more shade difference in the boxes represents significant different expression between the samples (alpha = 0.05). *NS* = not significant.
doi:10.1371/journal.pone.0066389.g003

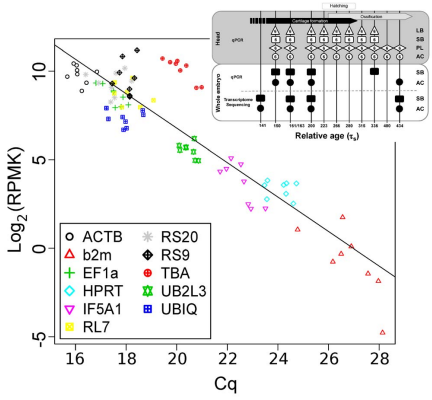


Figure 4. Comparison between expression values from RNA-seq and qPCR. Reads per million aligned per kilobase (RPMK) transformed to a log2-scale were plotted against equivalent Cq-values for eleven candidate reference genes. The compared samples were from the same groups and at the same or similar relative age (insert, black spots represent samples used for analysis). The line is a least squares linear fit to the data ($y = 26.23–0.97x$, $R^2 = 0.815$).
doi:10.1371/journal.pone.0066389.g004

Table 3. Ranking of reference gene candidates, based on stability of expression in Arctic charr whole embryos using either qPCR or RNA-seq.

	qPCR		RNA-seq		
Gene	**Nf**	**SD**	**C.V.**	**p M**	**p T**
ACTB	2	3	4	*	-
b2m	11	11	11	**	**
EF1α	3	4	6	-	-
HPRT	5	5	3	-	*
IF5A1	10	10	9	-	**
RL7	9	9	7	-	-
RS9	1	2	8	-	*
RS20	7	8	10	*	**
TBA	8	6	5	-	-
UB2L3	4	1	1	-	-
UBIQ	6	7	2	-	-

Abbreviations: Nf _ NormFinder, SD = standard deviation, C.V. = coefficient of variation (used to rank), p M = significant differences between morphological groups, p T = significant differences between developmental time points. **p = <0.01; * = p<0.05; - = no significant difference.
doi:10.1371/journal.pone.0066389.t003

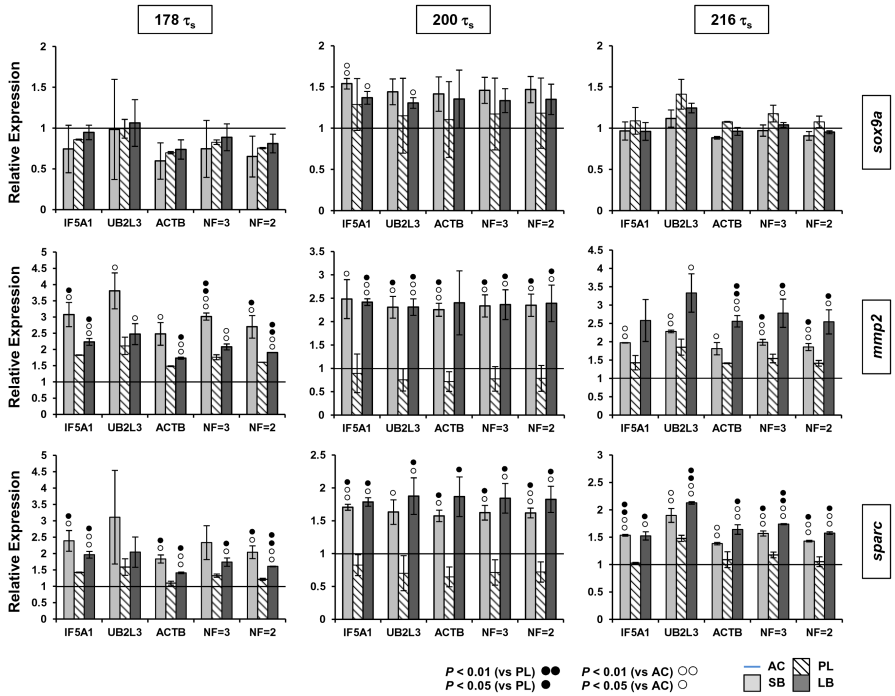


Figure 5. Comparison of different reference genes for normalising the expression of sox9a, mmp2 and sparc in charr heads at three embryonic stages. Expression of sox9a, mmp2 and sparc was examined with qPCR and normalised using either individual or two combinations of reference genes. Normalisation factors (NF) were based on geometric means of either two or three genes (NF = 2: ACTB and IF5A1; NF = 3: ACTB, IF5A1 and UB2L3). In each analysis (panel column) relative expression levels of the three genes in small benthivorous (SB), planktivorous (PL) and large benthivorous charr (LB), are compared to expression levels in aquaculture (AC) charr (horizontal line) at the same embryonic stage. Statistical differences of SB or LB gene expression versus expression in either PL (black circles) or AC (white circles) are indicated. Error bars represent standard deviation calculated from two biological replicates. Each biological replicate contains homogenate of six heads.
doi:10.1371/journal.pone.0066389.g005

Discussion

Numerous studies have discussed the importance of proving the stability of reference genes under the relevant study conditions [30,32]. Previously reference genes validated in salmon and trout have been used to study charr gene expression [51,52]. We examined several of these genes, but excluded them in further analyses due to technical problems, such as low primer efficiencies and non-specific amplification (data not shown). One of our aims in this study was to select genes from a wide variety of pathways in order to ensure a robust normalisation strategy for further gene expression analysis. This will enable us to detect small changes in the expression of developmental genes important for the formation of the craniofacial morphology of Arctic charr. Using a combination data from RNA sequencing and qPCR we have established a suite of reference genes suitable for studying gene expression in charr embryos. All of the candidate reference genes except GAPDH can be considered suitable for comparative analyses of qPCR data. Overall ACTB, TBA, UBIQ and UB2L3 were found to be the most stably expressed genes, but the ranking of the different genes varied according to body part and charr group examined, as well as by the method of analysis used. Other studies of reference genes have discussed the tendency of reference gene validation programs to rank co-regulated genes with similar expression patterns as the most stably expressed genes [23,44]. In this study we found that GeNorm and BestKeeper rankings favour genes that are co-regulated, whereas genes such as b2m, HPRT and IF5A1 are often at the bottom of the rankings. This is due to the fact that these three genes have expression patterns that are entirely different from the rest of the candidates (Figure 3) affecting the ranking with these two algorithms. Our aim was to select reference genes which are not co-regulated. EF1α and the ribosomal protein genes, for example, have a role in protein biosynthesis and show very similar expression patterns. Similarly the two genes involved in protein degradation, *i.e.* UBIQ and UB2L3, conform to the expression of ribosomal protein genes.

Genes like b2m, HRPT and IF5A1 on the other hand were found to be more highly expressed at the later points of development examined here. Therefore either HPRT or IF5A1 in combination with ACTB, TBA or UB2L3 would be good reference genes. This is confirmed in our pilot study of the expression profile of the three developmental genes. Our approach using ANOVA and post hoc HSD tests to analyse expression profiles of genes provides important advances over the commonly used programs for reference gene selection and validation. This is illustrated when considering TBA as reference gene. TBA was found to be the highest ranking gene in charr heads, but showed significantly different expression among groups and the relative ages examined (Figure 3). This suggests that although TBA is very stable within each charr group, it might not be the most suitable reference gene when comparing gene expression between charr morphs/groups. For this reason we did not include TBA in our pilot analysis of developmental genes. These findings illustrate the importance of understanding the background of the algorithms used, in order to choose reference genes and to clarify which genes are suitable for the task at hand, instead of relying on one method of reference gene selection.

When comparing the transcriptome and qPCR data we found that in general the two methods recorded similar gene expression levels. An exception to this is TBA with higher transcript levels seen in the RNA-seq data, than determined by qPCR (Figure 4). This result might be explained by the presence of gene paralogues. Salmonids, including Arctic charr, have undergone a recent genome duplication event [53] and this has led to the evolution of gene paralogues [54,55]. The TBA primers used here, only bind one of at least 3 paralogues of TBA in charr and this may have led to an underestimation of the expression of TBA using qPCR compared to the sequenced reads (Figure 4).

When examining our sequencing data in detail we found that all reference genes except RS20 and HRPT have paralogues in Arctic charr (unpublished data). In contrast to TBA these other primer pairs are thought to amplify all paralogues for the respective gene. The amplification of several paralogues with a single primer pair could explain the high expression stability observed for most genes and interestingly this did not result in a broader melting curve for the PCR products, reflecting identical lengths and GC content of the paralogues. These results underline the importance of considering the presence of paralogues when studying gene expression in salmonids, but for selection of stable qPCR references their presence may actually be an advantage.

Further evaluations of the consistency of normalisation using three of our newly validated reference genes (IF5A1, UB2L3 and ACTB) were made in a pilot study examining the expression of three developmental genes (sox9a, sparc and mmp2) at three time points in developing Arctic charr heads. The analyses showed that each of the three reference genes could be used individually with consistent results, but the use of two or three reference genes decreased the small observed variation in expression even further. Therefore, in future comparative studies of the development of divergent trophic morphologies in Arctic charr, we will use the geometric mean of ACTB and IF5A1.

The developmental results of the pilot study are of considerable interest. While sox9a expression varied significantly through time, variation among morphs was not significant. Sox9 is a member of the Sry-related HMG-box gene family and encodes a transcription factor with an important and highly conserved role in cartilage formation [56–60], Two co-orthologues of sox9 (sox9a/b) with overlapping expression pattern have been reported during craniofacial cartilage formation of teleost fish [50,61–63]. Sox9a was differentially expressed in our transcriptome analysis, but we could not detect similar differences with qPCR analysis. This might be caused by the fact that transcriptome sequencing was performed on whole embryos, whereas qPCR was focused on charr heads and sox9a might not be differentially expressed in the head.

Expression of sparc and mmp2 varied significantly both in time and among the morphs. Sparc/osteonectin is a highly conserved collagen-binding glycoprotein which plays important roles in extracellular matrix (ECM) remodelling and craniofacial morphogenesis [49,64–66]. Similarly, matrix metalloproteases, including mmp2, have important roles during craniofacial morphogenesis through precise regulation of ECM degradation [48,67]. Sparc has been suggested to act downstream of sox9 during cartilage formation of the pharyngeal arches [49]. In our data, however, sparc expression levels, which are higher in the benthic morphs, do not go hand in hand with sox9a levels (Figure 5). An association between sparc up-regulation and increased mmp2 expression and activity, has been shown in various studies [64,68–73]. In the present study both genes are consistently expressed at higher levels in the head of benthic than pelagic groups, suggesting a role of these genes in the observed differences in trophic morphology between the charr morphotypes.

In conclusion we have, using data from transcriptome sequencing and qPCR, identified several suitable reference genes for the analysis of gene expression in developing Arctic charr embryos. Furthermore, we have used these to confirm putative expression differences between the charr morphotypes in two craniofacially expressed genes. The tools generated here will be of great use in further analyses of gene expression in Arctic charr embryogenesis and will be instrumental in our search for genes that play key roles in inducing different trophic morphotypes. Finally, the use of ANOVA for reference gene selection as we have demonstrated will be useful for validation of reference genes in other species.

Supporting Information

File S1 Contains: Table S1 qPCR Primer sequences and information. Table S2 A–F Descriptive statistical analysis of the candidate reference gene expression in Arctic charr using three algorithms. Table S3 Ranking of the candidate reference genes in the heads or the different Arctic charr groups using NormFinder (Nf) and Standard deviation (SD).
(DOC)

File S2 Contains: Figure S1 Comparison of expression levels of the eleven candidate reference genes in heads and whole embryos using qPCR. The genes are ranked from left to right as most to least differentially expressed between whole embryos (whole) and head (corresponding P-values are shown below the x-axis). Insert in A displays samples (black filled spots) used in both analyses. (**A**) Boxplot shows the range of Cq values for each candidate reference gene in whole embryo (white) and head (gray) homogenates. Displayed are the median, the 25th and 75th percentiles and the minimum and maximum Cq values for each gene. (**B**) Relative quantity for each reference gene candidate in whole embryos (open circles) and head (grey diamond) homogenates. The whiskers represent ±0.95 confidence interval of the mean. **Figure S2 Optimal number of reference genes for normalisation.** Genorm^{PLUS} was used to determine the optimal number of reference genes in head and whole embryo homogenates. LB: large benthic; SB: small benthic/dwarf; PL: planktivorous; AC: aquaculture. Average pair-wise variations ($V_{n/n+1}$) were calculated using the genes ranked according to GeNorm. The

recommended cut-off value of 0.15 is shown by a dashed line and below this line the benefit of using an extra reference gene is limited.
(DOCX)

Acknowledgments

We thank Hákon Jónsson for his help with the assembly of the transcriptome data. We acknowledge Bjarni K. Kristjánsson, Einar Svavarsson and Soizic Le Deuff for assisting with the sampling of parents, generation and maintenance of and sampling from embryo groups. We thank Droplaug N. Magnúsdóttir, Guðbjörg þ. Örlygsdóttir, Steinunn Snorradóttir and Ólafur þ. Magnússon at deCODE Genetics for technical support.

Author Contributions

Conceived and designed the experiments: SSS SRF ZOJ VHM EPA. Performed the experiments: EPA KHK JG KHK VHM. Analyzed the data: EPA JG KHK VHM. Wrote the paper: EPA KHK SRF SSS ZOJ VHM.

References

1. Hulsey CD, Fraser GJ, Streelman JT (2005) Evolution and development of complex biomechanical systems: 300 million years of fish jaws. Zebrafish 2: 243–257.
2. Szabo-Rogers HL, Smithers LE, Yakob W, Liu KJ (2010) New directions in craniofacial morphogenesis. Dev Biol 341: 84–94.
3. Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. Heredity 97: 211–221.
4. Danos N, Staab KL (2010) Can mechanical forces be responsible for novel bone development and evolution in fishes? Journal of Applied Ichthyology 26: 156–161.
5. Bolnick DI, Svanback R, Fordyce JA, Yang LH, Davis JM, et al. (2003) The ecology of individuals: Incidence and implications of individual specialization. American Naturalist 161: 1–28.
6. Ward AJW, Webster MM, Hart PJB (2006) Intraspecific food competition in fishes. Fish and Fisheries 7: 231–261.
7. Whiteley AR (2007) Trophic polymorphism in a riverine fish: morphological, dietary, and genetic analysis of mountain whitefish. Biological Journal of the Linnean Society 92: 253–267.
8. Skulason S, Smith TB (1995) Resource Polymorphisms in Vertebrates. Trends in Ecology & Evolution 10: 366–370.
9. Skulason S, Snorrason SS, Noakes DLG, Ferguson MM, Malmquist HJ (1989) Segregation in Spawning and Early Life-History among Polymorphic Arctic Charr, Salvelinus-Alpinus, in Thingvallavatn, Iceland. Journal of Fish Biology 35: 225–232.
10. Malmquist HJ, Snorrason SS, Skulason S, Jonsson B, Sandlund OT, et al. (1992) Diet Differentiation in Polymorphic Arctic Charr in Thingvallavatn, Iceland. Journal of Animal Ecology 61: 21–35.
11. Sandlund OT, Gunnarsson K, Jonasson PM, Jonsson B, Lindem T, et al. (1992) The Arctic Charr Salvelinus-Alpinus in Thingvallavatn. Oikos 64: 305–351.
12. Jonsson B, Skulason S, Snorrason SS, Sandlund OT, Malmquist HJ, et al. (1988) Life-History Variation of Polymorphic Arctic Charr (Salvelinus-Alpinus) in Thingvallavatn, Iceland. Canadian Journal of Fisheries and Aquatic Sciences 45: 1537–1547.
13. Parsons KJ, Skulason S, Ferguson M (2010) Morphological variation over ontogeny and environments in resource polymorphic arctic charr (Salvelinus alpinus). Evol Dev 12: 246–257.
14. Skulason S, Noakes DLG, Snorrason SS (1989) Ontogeny of Trophic Morphology in 4 Sympatric Morphs of Arctic Charr Salvelinus-Alpinus in Thingvallavatn, Iceland. Biological Journal of the Linnean Society 38: 281–301.
15. Eiriksson GM, Skulason SS, Snorrason SS (1999) Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn, Iceland. Journal of Fish Biology 55: 175–185.
16. Kapralova KH, Morrissey MB, Kristjansson BK, Olafsdottir GA, Snorrason SS, et al. (2011) Evolution of adaptive diversity and genetic connectivity in Arctic charr (Salvelinus alpinus) in Iceland. Heredity 106: 472–487.
17. Sibthorpe D, Sturlaugsdottir R, Kristjansson BK, Thorarensen H, Skulason S, et al. (2006) Characterisation and expression of the paired box protein 7 (Pax7) gene in polymorphic Arctic charr (Salvelinus alpinus). Comp Biochem Physiol B Biochem Mol Biol 145: 371–383.
18. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57–63.
19. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18: 1509–1517.
20. Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J (2011) The zebrafish transcriptome during early development. BMC Dev Biol 11: 30.
21. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods 25: 402–408.
22. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.
23. Saele O, Nordgreen A, Hamre K, Olsvik PA (2009) Evaluation of candidate reference genes in Q-PCR studies of Atlantic cod (Gadus morhua) ontogeny, with emphasis on the gastrointestinal tract. Comp Biochem Physiol B Biochem Mol Biol 152: 94–101.
24. Jorgensen SM, Kleveland EJ, Grimholt U, Gjoen T (2006) Validation of reference genes for real-time polymerase chain reaction studies in Atlantic salmon. Mar Biotechnol (NY) 8: 398–408.
25. Tang R, Dodd A, Lai D, McNabb WC, Love DR (2007) Validation of zebrafish (Danio rerio) reference genes for quantitative real-time RT-PCR normalization. Acta Biochim Biophys Sin (Shanghai) 39: 384–390.
26. Zhong Q, Zhang Q, Wang Z, Qi J, Chen Y, et al. (2008) Expression profiling and validation of potential reference genes during Paralichthys olivaceus embryogenesis. Mar Biotechnol (NY) 10: 310–318.
27. Zheng WJ, Sun L (2010) Evaluation of housekeeping genes as references for quantitative real time RT-PCR analysis of gene expression in Japanese flounder (Paralichthys olivaceus). Fish Shellfish Immunol 30: 638–645.
28. Overgard AC, Nerland AH, Patel S (2010) Evaluation of potential reference genes for real time RT-PCR studies in Atlantic halibut (Hippoglossus Hippoglossus L.); during development, in tissues of healthy and NNV-injected fish, and in anterior kidney leucocytes. BMC Mol Biol 11: 36.
29. Ye X, Zhang L, Dong H, Tian Y, Lao H, et al. (2010) Validation of reference genes of grass carp Ctenopharyngodon idellus for the normalization of quantitative real-time PCR. Biotechnol Lett 32: 1031–1038.
30. Derveaux S, Vandesompele J, Hellemans J (2010) How to do successful gene expression analysis using real-time PCR. Methods 50: 227–230.
31. Bustin SA, Beaulieu JF, Huggett J, Jaggi R, Kibenge FS, et al. (2010) MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. BMC Mol Biol 11: 74.
32. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55: 611–622.
33. Gorodilov YN (1996) Description of the early ontogeny of the Atlantic salmon, Salmo salar, with a novel system of interval (state) identification. Environmental Biology of Fishes 47: 109–127.
34. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821–829.
35. Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, et al. (2008) A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. Bmc Genomics 9.
36. Andreassen R, Lunner S, Hoyheim B (2009) Characterization of full-length sequenced cDNA inserts (FLIcs) from Atlantic salmon (Salmo salar). Bmc Genomics 10.
37. Rexroad CE, Lee Y, Keele JW, Karamycheva S, Brown G, et al. (2003) Sequence analysis of a rainbow trout cDNA library and creation of a gene index. Cytogenetic and Genome Research 102: 347–354.
38. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10.
39. Henricson A, Forslund K, Sonnhammer EL (2010) Orthology confers intron position conservation. BMC Genomics 11: 412.
40. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8: R19.
41. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, et al. (2004) Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 313: 856–862.
42. Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339: 62–66.
43. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnol Lett 26: 509–515.
44. Andersen CL, Jensen JL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64: 5245–5250.
45. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.
46. Team RDC (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
47. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140.

48. Hillegass JM, Villano CM, Cooper KR, White LA (2008) Glucocorticoids Alter Craniofacial Development and Increase Expression and Activity of Matrix Metalloproteinases in Developing Zebrafish (Danio rerio). Toxicological Sciences 102: 413–424.
49. Rotllant J, Liu D, Yan Y-L, Postlethwait JH, Westerfield M, et al. (2008) Sparc (Osteonectin) functions in morphogenesis of the pharyngeal skeleton and inner ear. Matrix Biology 27: 561–572.
50. Chiang EFL, Pai C-I, Wyatt M, Yan Y-L, Postlethwait J, et al. (2001) Two Sox9 Genes on Duplicated Zebrafish Chromosomes: Expression of Similar Transcription Activators in Distinct Sites. Developmental Biology 231: 149–163.
51. Macqueen DJ, Kristjansson BK, Johnston IA (2010) Salmonid genomes have a remarkably expanded akirin family, coexpressed with genes from conserved pathways governing skeletal muscle growth and catabolism. Physiological Genomics 42: 134–148.
52. Macqueen DJ, Kristjansson BK, Paxton CG, Vieira VL, Johnston IA (2011) The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. Mol Ecol 20: 3167–3184.
53. Allendorf FW, Thorgaard GH (1984) Tetraploidy and the Evolution of Salmonid Fishes. Evolutionary Genetics of Fishes Edited by B.J. Turner: 1–53.
54. Macqueen DJ, Bower NI, Johnston IA (2010) Positioning the expanded akirin gene family of Atlantic salmon within the transcriptional networks of myogenesis. Biochem Biophys Res Commun 400: 599–605.
55. Macqueen DJ, Johnston IA (2006) A novel salmonid myoD gene is distinctly regulated during development and probably arose by duplication after the genome tetraploidization (vol 580, pg 4996, 2006). Febs Letters 580: 6286–6287.
56. Healy C, Uwanogho D, Sharpe PT (1999) Regulation and role of Sox9 in cartilage formation. Developmental Dynamics 215: 69–78.
57. Zhao Q, Eberspaecher H, Lefebvre V, de Crombrugghe B (1997) Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis. Developmental Dynamics 209: 377–386.
58. Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrugghe B (1999) Sox9 is required for cartilage formation. Nat Genet 22: 85–89.
59. Yan YL, Miller CT, Nissen RM, Singer A, Liu D, et al. (2002) A zebrafish sox9 gene required for cartilage morphogenesis. Development 129: 5065–5079.
60. Bell DM, Leung KK, Wheatley SC, Ng LJ, Zhou S, et al. (1997) SOX9 directly regulates the type-II collagen gene. Nat Genet 16: 174–178.
61. Klüver N, Kondo M, Herpin A, Mitani H, Schartl M (2005) Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. Development Genes and Evolution 215: 297–305.
62. Takamatsu N, Kanda H, Ito M, Yamashita A, Yamashita S, et al. (1997) Rainbow trout SOX9: cDNA cloning, gene structure and expression. Gene 202: 167–170.
63. Koopman P, Schepers G, Brenner S, Venkatesh B (2004) Origin and diversity of the SOX transcription factor gene family: genome-wide analysis in Fugu rubripes. Gene 328: 177–186.
64. Chlenski A, Cohn SL (2010) Modulation of matrix remodeling by SPARC in neoplastic progression. Seminars in Cell & Developmental Biology 21: 55–65.
65. Renn J, Schaedel M, Volff J-N, Goerlich R, Schartl M, et al. (2006) Dynamic expression of sparc precedes formation of skeletal elements in the Medaka (Oryzias latipes). Gene 372: 208–218.
66. Lane TF, Sage EH (1994) The biology of SPARC, a protein that modulates cell-matrix interactions. FASEB J 8: 163–173.
67. Mosig RA, Dowling O, DiFeo A, Ramirez MC, Parker IC, et al. (2007) Loss of MMP-2 disrupts skeletal and craniofacial development and results in decreased bone mineralization, joint erosion and defects in osteoblast and osteoclast growth. Hum Mol Genet 16: 1113–1123.
68. McClung HM, Thomas SL, Osenkowski P, Toth M, Menon P, et al. (2007) SPARC upregulates MT1-MMP expression, MMP-2 activation, and the secretion and cleavage of galectin-3 in U87MG glioma cells. Neuroscience Letters 419: 172–177.
69. Gilles C, Bassuk JA, Pulyaeva H, Sage EH, Foidart J-M, et al. (1998) SPARC/Osteonectin Induces Matrix Metalloproteinase 2 Activation in Human Breast Cancer Cell Lines. Cancer Research 58: 5529–5536.
70. Seet L-F, Su R, Toh LZ, Wong TT (2012) In vitro analyses of the anti-fibrotic effect of SPARC silencing in human Tenon's fibroblasts: comparisons with mitomycin C. Journal of Cellular and Molecular Medicine 16: 1245–1259.
71. Yamanaka M, Kanda K, Li N-C, Fukumori T, Oka N, et al. (2001) ANALYSIS OF THE GENE EXPRESSION OF SPARC AND ITS PROGNOSTIC VALUE FOR BLADDER CANCER. The Journal of Urology 166: 2495–2499.
72. Socha MJ, Manhiani M, Said N, Imig JD, Motamed K (2007) Secreted protein acidic and rich in cysteine deficiency ameliorates renal inflammation and fibrosis in angiotensin hypertension. Am J Pathol 171: 1104–1112.
73. Fujita T, Shiba H, Sakata M, Uchida Y, Nakamura S, et al. (2002) SPARC stimulates the synthesis of OPG/OCIF, MMP-2 and DNA in human periodontal ligament cells. Journal of Oral Pathology & Medicine 31: 345–352.
74. Javidan Y, Schilling TF (2004) Development of cartilage and bone. Zebrafish: 2nd Edition Cellular and Developmental Biology 76: 415–436.

File S1:

Table S1. qPCR Primer sequences and information

Gene		Primer Sequence (5'- 3')	Product Size (bp)	Efficiency (E %)	R^2*	Melting Temperature (°C) Head	Melting Temperature (°C) Whole Embryo	Location** (Exon No.)
ACTB	F	GAAGATCAAGATCATCGCCC	122	95.5	0.998	80.5 ± 0.7	80.52 ± 0.7	4:5
	R	CAGACTCGTCGTACTCCTGCT						5
b2m	F	CGAACAGGGATGGCAGTT	105	95.5	0.998	80.92 ± 0.4	81.31 ± 0.2	1
	R	TAGGTCTTCAGATTCTTCAGGTGG						1
EF1α	F	GAAGATCGGCTATAACCCTGC	111	94.6	0.998	81.36 ± 0.4	81.11 ± 0.1	3
	R	ACCTTCCATCCCTTGAACC						4
GAPDH	F	GGTCTGATGAGCACCGTTC	110	114.8	0.96	81.3 ± 0.3	81.46 ± 0.2	5:6
	R	GCAGGGATGATGTTCTGGC						6
HPRT	F	TTCTCAAACAGTACAACCCAAAAA	95	96.3	0.994	76.57 ± 0.4	76.05 ± 0.2	5
	R	TCCTATGAAGTCTGGTGTGTAGC						6:7
IF5A1	F	GGCTTCGTGGTGCTGAAG	91	91.5	0.999	80.76 ± 0.6	81.1 ± 0.1	1
	R	CCATGTGGACCTTAGCGTG						1:2
RL7	F	CATCAGGATCAGGGGTATCAA	109	91.9	0.998	81.29 ± 0.2	80.67 ± 0.5	1:2
	R	AGCCTTGTTCAGTTTGACGAA						2
RS9	F	GAGGTGTGGAGGGTGAAGTT	113	95.4	0.998	82.03 ± 0.4	82.18 ± 0.4	2
	R	CTGAGCAGGGCGTTACCTT						2:3
RS20	F	AGCCGCAACGTCAAGTCT	110	96.7	0.999	79.9 ± 0.8	80.5 ± 0.3	1
	R	CGCAGAGTCTTTGTGGGC						2:3
TBA	F	GTCACTACACCATTGGCAAAGA	104	103.5	0.994	78.9 ± 0.4	79.02 ± 0.2	2
	R	GCTGTGGAAGATGAGGAATCC						3
UB2L3	F	CGAGAAGGGACAGGTGTGTC	96	93.4	0.999	79.62 ± 0.3	79.5 ± 0.3	3
	R	ACCAACGCAATCAGGGACT						4
UBIQ	F	GACTACAACATCCAGAAAGAGTCCA	120	92.6	0.999	79.4 ± 0.3	79.36 ± 0.1	2:3
	R	GCGGCAGATCATTTTGTC						3
sox9a	F	CCTGAGTGGAGGTGGAGG	74	102.4	0.999	79.3 ± 0.3	-	N-S
	R	GCTCCGTCTTGATCTGAGTG						
sparc	F	GTTCTGGTCACCCTGTACGAG	100	91.6	0.999	79.9 ± 0.1	-	N-S
	R	GCAGTCTCTTCTCATTCTCATAGATC						
mmp2	F	ATGGATGGAGAGGCTGACATC	110	93.7	0.999	79.73 ± 0.3	-	N-S
	R	GGTCCAGGAGCAAAGGCAT						

(*)R^2, correlation coefficient of the slope of the standard curve (**) Putative exon borders based on zebrafish homologues. The numbers are referring to the exon(s) in which the primers are located, (n:n) and (N-S) indicate "exon/exon" boundaries and "non-specified", respectively.

Table S2A. Descriptive statistical analysis of the candidate reference gene expression in Arctic charr heads using three algorithms.

Algorithm		ACTB	b2m	EF1a	GAPDH	HPRT	IF5A1	RL7	RS9	RS20	TBA	UB2L3	UBIQ
BestKeeper	B.I. all (r)	0.795	0.587	0.708	-0.039	0.447	0.188	0.735	0.773	0.77	0.823	0.613	0.932
	B.I. 4 genes (r)	0.794	0.278	0.784	-0.235	0.195	0.167	0.785	0.782	0.751	0.932	0.801	0.815
	B.I. 3 genes (r)	0.838	0.441	0.646	0.017	0.42	0.383	0.618	0.614	0.644	0.822	0.557	0.778
	B.I. 2 genes (r)	0.917	0.392	0.787	-0.254	0.2	-0.01	0.747	0.763	0.76	0.885	0.716	0.869
	overall B. I.* (r)	0.836	0.425	0.731	-0.128	0.316	0.182	0.721	0.733	0.731	0.866	0.672	0.849
	S.D.	0.476	0.951	0.643	2.411	0.674	0.554	0.872	0.822	0.789	0.472	0.558	0.742
GeNorm	M	0.494	0.742	0.378	1.042	0.663	0.591	0.429	0.418	0.451	0.354	0.366	0.463
NormFinder	stability	0.13	0.584	0.398	1.753	0.359	0.374	0.516	0.476	0.403	0.2	0.39	0.236

Abbreviations: S.D. = Standard deviation, B.I. = BestKeeper Index, *r* = Pearson product-moment correlation coefficient.
*average B.I.

Table S2B. Descriptive statistical analysis of the candidate reference gene expression in aquaculture charr heads using three algorithms.

Algorithm		ACTB	b2m	EF1a	GAPDH	HPRT	IF5A1	RL7	RS9	RS20	TBA	UB2L3	UBIQ
BestKeeper	B.I. all (r)	0.748	0.738	0.62	0.166	0.238	0.201	0.231	0.356	0.63	0.608	0.601	0.762
	B.I. 4 genes (r)	0.706	0.291	0.905	-0.42	-0.375	-0.175	0.65	0.605	0.954	0.905	0.864	0.905
	B.I. 3 genes (r)	0.384	0.009	0.985	-0.638	-0.539	-0.307	0.882	0.815	0.935	0.957	0.947	0.868
	B.I. 2 genes (r)	0.393	0.002	0.957	-0.618	-0.507	-0.239	0.87	0.768	0.912	0.96	0.958	0.84
	overall B. I.* (r)	0.558	0.260	0.867	-0.378	-0.296	-0.130	0.658	0.636	0.858	0.858	0.843	0.844
	S.D.	0.463	1.076	0.401	2.258	0.665	0.549	0.588	0.599	0.695	0.349	0.352	0.465
GeNorm	M	0.397	0.68	0.16	0.957	0.584	0.497	0.276	0.31	0.34	0.168	0.173	0.214
NormFinder	stability	0.162	0.592	0.26	1.602	0.385	0.333	0.489	0.462	0.462	0.222	0.228	0.242

Abbreviations: S.D. = Standard deviation, B.I. = BestKeeper Index, *r* = Pearson product-moment correlation coefficient.
*average B.I.

Table S2C. Descriptive statistical analysis of the candidate reference gene expression in planktivorous charr heads using three algorithms.

Algorithm		ACTB	b2m	EF1a	GAPDH	HPRT	IF5A1	RL7	RS9	RS20	TBA	UB2L3	UBIQ
BestKeeper	B.I. all (r)	0.86	0.95	0.813	0.149	0.736	0.409	0.838	0.877	0.926	0.972	0.866	0.954
	B.I. 4 genes (r)	0.939	0.785	0.962	-0.23	0.465	0.071	0.919	0.903	0.958	0.968	0.958	0.926
	B.I. 3 genes (r)	0.958	0.832	0.914	-0.131	0.552	0.168	0.879	0.875	0.939	0.982	0.947	0.915
	B.I. 2 genes (r)	0.876	0.799	0.915	-0.161	0.543	0.129	0.926	0.932	0.959	0.988	0.987	0.933
	overall B. I.* (r)	0.908	0.842	0.901	-0.093	0.574	0.194	0.891	0.897	0.946	0.978	0.940	0.932
	S.D.	0.668	1.266	0.835	2.562	0.820	0.650	1.020	1.076	1.012	0.600	0.602	1.069
GeNorm	M	0.442	0.53	0.336	1.01	0.602	0.675	0.228	0.223	0.238	0.411	0.39	0.278
NormFinder	stability	0.25	0.401	0.403	1.845	0.295	0.462	0.472	0.456	0.357	0.105	0.27	0.335

Abbreviations: S.D. = Standard deviation, B.I. = BestKeeper Index, *r* = Pearson product-moment correlation coefficient.
*average B.I.

Table S2D. Descriptive statistical analysis of the candidate reference gene expression in small benthic charr heads using three algorithms.

Algorithm		ACTB	b2m	EF1a	GAPDH	HPRT	IF5A1	RL7	RS9	RS20	TBA	UB2L3	UBIQ
BestKeeper	B.I. all (r)	0.811	-0.313	0.872	-0.651	-0.583	-0.735	0.897	0.806	0.775	0.86	0.828	0.979
	B.I. 4 genes (r)	0.867	-0.003	0.743	-0.593	-0.492	-0.81	0.822	0.796	0.614	0.831	0.766	0.923
	B.I. 3 genes (r)	0.892	0.174	0.629	-0.479	-0.372	-0.797	0.704	0.718	0.449	0.742	0.635	0.809
	B.I. 2 genes (r)	0.896	0.431	0.339	-0.075	-0.03	-0.639	0.324	0.344	0.079	0.338	0.266	0.518
	overall B. I.* (r)	0.867	0.072	0.646	-0.450	-0.369	-0.745	0.687	0.666	0.479	0.693	0.624	0.807
	S.D.	0.430	0.358	0.653	2.488	0.471	0.527	0.742	0.759	0.707	0.434	0.729	0.460
GeNorm	M	0.429	0.512	0.37	1.018	0.599	0.669	0.317	0.351	0.385	0.292	0.334	0.293
NormFinder	stability	0.093	0.247	0.371	1.919	0.35	0.444	0.447	0.478	0.439	0.167	0.441	0.144

Abbreviations: S.D. = Standard deviation, B.I. = BestKeeper Index, *r* = Pearson product-moment correlation coefficient.
*average B.I.

Table S2E. Descriptive statistical analysis of the candidate reference gene expression in large benthic charr heads using three algorithms.

Algorithm		ACTB	b2m	EF1a	GAPDH	HPRT	IF5A1	RL7	RS9	RS20	TBA	UB2L3	UBIQ
BestKeeper	B.I. all (r)	0.887	0.493	0.701	-0.456	0.142	-0.434	0.833	0.791	0.271	0.797	0.617	0.893
	B.I. 4 genes (r)	0.859	0.44	0.696	-0.42	0.178	-0.412	0.801	0.747	0.191	0.847	0.578	0.947
	B.I. 3 genes (r)	0.907	0.567	0.579	-0.314	0.29	-0.371	0.757	0.718	0.116	0.746	0.54	0.842
	B.I. 2 genes (r)	0.911	0.175	0.682	-0.495	0.063	-0.503	0.869	0.803	0.146	0.907	0.74	0.889
	overall B. I.* (r)	0.891	0.419	0.665	-0.421	0.168	-0.430	0.815	0.765	0.181	0.824	0.619	0.893
	S.D.	0.274	0.418	0.671	2.732	0.525	0.481	0.830	0.727	0.671	0.348	0.557	0.512
GeNorm	M	0.295	0.545	0.446	1.043	0.606	0.654	0.431	0.412	0.498	0.258	0.368	0.283
NormFinder	stability	0.093	0.105	0.435	2.061	0.25	0.343	0.535	0.461	0.512	0.054	0.354	0.211

Abbreviations: S.D. = Standard deviation, B.I. = BestKeeper Index, *r* = Pearson product-moment correlation coefficient.
*average B.I.

Table S2F. Descriptive statistical analysis of the candidate reference gene expression in whole charr embryos using three algorithms.

Algorithm		ACTB	b2m	EF1a	GAPDH	HPRT	IF5A1	RL7	RS9	RS20	TBA	UB2L3	UBIQ
BestKeeper	B.I. all (r)	0.608	0.696	0.843	-0.196	0.801	0.144	0.427	0.733	0.806	0.446	0.799	0.497
	B.I. 4 genes (r)	0.542	0.479	0.938	-0.527	0.836	-0.151	0.596	0.889	0.813	0.598	0.844	0.654
	B.I. 3 genes (r)	0.512	0.593	0.904	-0.424	0.803	-0.155	0.549	0.905	0.832	0.446	0.893	0.491
	B.I. 2 genes (r)	0.639	0.668	0.9	-0.34	0.808	-0.035	0.491	0.802	0.799	0.442	0.825	0.477
	overall B. I.* (r)	0.575	0.609	0.896	-0.372	0.812	-0.049	0.516	0.832	0.813	0.483	0.840	0.530
	S.D.	0.353	1.094	0.421	1.292	0.496	0.580	0.568	0.325	0.567	0.516	0.289	0.534
GeNorm	M	0.403	0.613	0.288	0.761	0.326	0.509	0.439	0.247	0.272	0.379	0.353	0.223
NormFinder	stability	0.172	0.664	0.174	0.993	0.224	0.406	0.399	0.148	0.276	0.36	0.204	0.254

Abbreviations: S.D. = Standard deviation, B.I. = BestKeeper Index, *r* = Pearson product-moment correlation coefficient.
*average B.I.

Table S3. Ranking of the candidate reference genes in the heads or the different Arctic charr groups using NormFinder (Nf) and Standard deviation (SD)

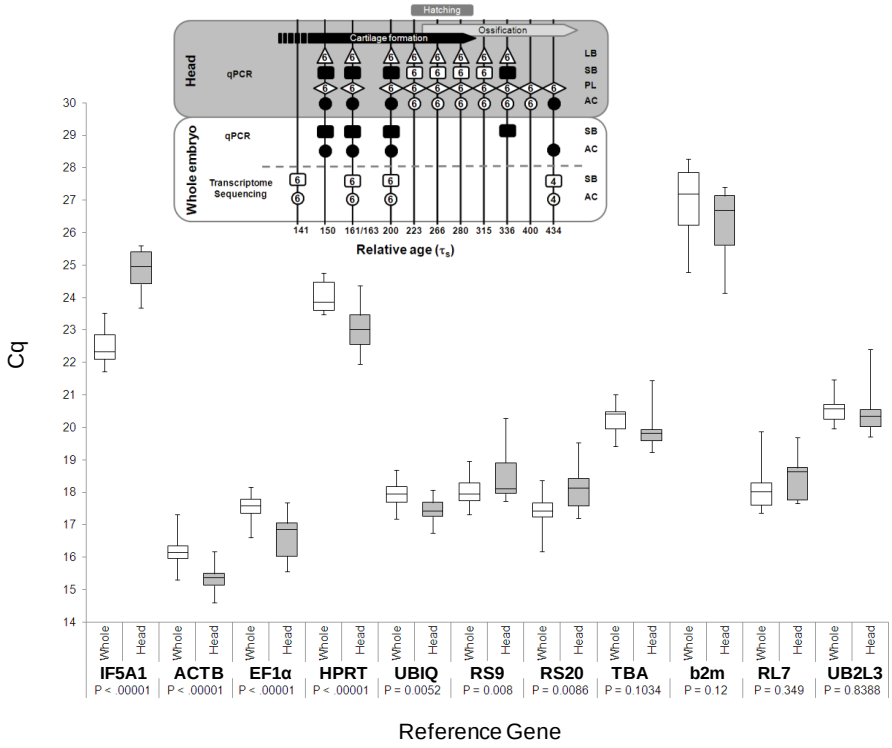
Gene	AC heads		PL heads		SB heads		LB heads	
	Nf	SD	Nf	SD	Nf	SD	Nf	SD
ACTB	1	4	2	4	1	2	2	1
b2m	11	11	7	11	4	1	3	3
EF1α	5	3	8	6	6	7	8	8
HPRT	7	9	4	5	5	5	5	6
IF5A1	6	6	10	3	9	6	6	4
RL7	10	7	11	8	10	10	11	11
RS9	9	7	9	10	11	11	9	10
RS20	8	10	6	7	7	8	10	9
TBA	2	1	1	1	3	3	1	2
UB2L3	3	2	3	2	8	9	7	7
UBIQ	4	5	5	9	2	4	4	5

Abbreviations: AC = aquaculture, PL = planktivorous, LB = large benthic, SB = small benthic/dwarf

File S2

Figure S1

A



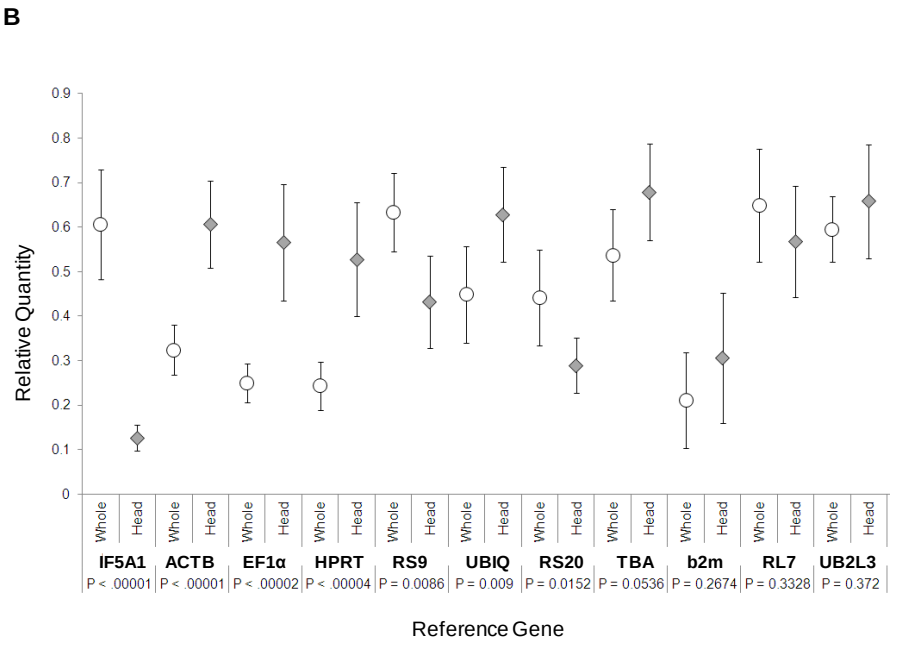


Figure S1 Comparison of expression levels of the eleven candidate reference genes in heads and whole embryos using qPCR. *The genes are ranked from left to right as most to least differentially expressed between whole embryos (whole) and head (corresponding P-values are shown below the x-axis). Insert in A displays samples (black filled spots) used in both analyses. (**A**) Boxplot shows the range of Cq values for each candidate reference gene in whole embryo (white) and head (gray) homogenates. Displayed are the median, the 25th and 75th percentiles and the minimum and maximum Cq values for each gene. (**B**) Relative quantity for each reference gene candidate in whole embryos (open circles) and head (grey diamond) homogenates. The whiskers represent ±0.95 confidence interval of the mean.*

Figure S2

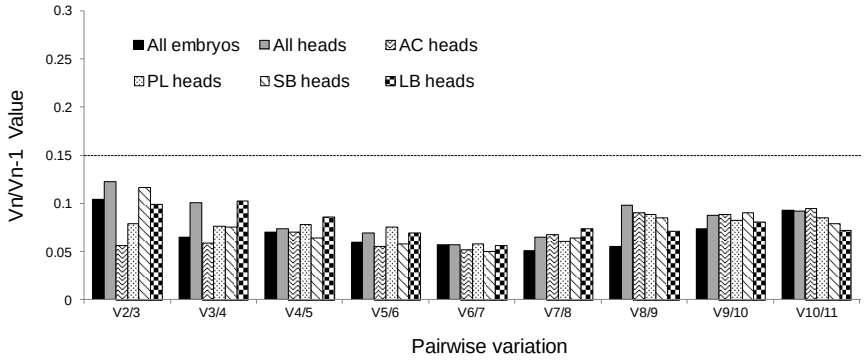


Figure S2 Optimal number of reference genes for normalisation. *Genorm^{PLUS} was used to determine the optimal number of reference genes in head and whole embryo homogenates. LB: large benthic; SB: small benthic/dwarf; PL: planktivorous; AC: aquaculture. Average pair-wise variations ($V_{n/n+1}$) were calculated using the genes ranked according to GeNorm. The recommended cut-off value of 0.15 is shown by a dashed line and below this line the benefit of using an extra reference gene is limited.*

Paper II

Differentiation at the *MHCII*α and *Cath2* Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn

Kalina H. Kapralova, Jóhannes Guðbrandsson, Sigrun Reynisdottir, Cristina B. Santos, Vanessa C. Baltanas, Valerie H. Maier, Sigurdur S. Snorrason and Arnar Palsson

PLOS ONE **8(7):** e69402

JG participated in sampling of fish, parasite analyses and writing. JG performed the statistical analysis of the parasite and phenotypic data.

OPEN ACCESS Freely available online

PLOS ONE

Differentiation at the *MHCIIα* and *Cath2* Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn

Kalina H. Kapralova, Johannes Gudbrandsson, Sigrun Reynisdottir, Cristina B. Santos, Vanessa C. Baltanás, Valerie H. Maier, Sigurdur S. Snorrason, Arnar Palsson*

Institute of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland

Abstract

Northern freshwater fish may be suitable for the genetic dissection of ecological traits because they invaded new habitats after the last ice age (~10.000 years ago). Arctic charr (*Salvelinus alpinus*) colonizing streams and lakes in Iceland gave rise to multiple populations of small benthic morphotypes, often in sympatry with a pelagic morphotype. Earlier studies have revealed significant, but subtle, genetic differentiation between the three most common morphs in Lake Thingvallavatn. We conducted a population genetic screen on four immunological candidate genes *Cathelicidin 2* (*Cath2*), *Hepcidin* (*Hamp*), *Liver expressed antimicrobial peptide 2a* (*Leap-2a*), and *Major Histocompatibility Complex IIα* (*MHCIIα*) and a mitochondrial marker (D-loop) among the three most common Lake Thingvallavatn charr morphs. Significant differences in allele frequencies were found between morphs at the *Cath2* and *MHCIIα* loci. No such signal was detected in the D-loop nor in the other two immunological genes. In *Cath2* the small benthic morph deviated from the other two (F_{ST} = 0.13), one of the substitutions detected constituting an amino acid replacement polymorphism in the antimicrobial peptide. A more striking difference was found in the *MHCIIα*. Two haplotypes were very common in the lake, and their frequency differed greatly between the morphotypes (from 22% to 93.5%, F_{ST} = 0.67). We then expanded our study by surveying the variation in *Cath2* and *MHCIIα* in 9 Arctic charr populations from around Iceland. The populations varied greatly in terms of allele frequencies at *Cath2*, but the variation did not correlate with morphotype. At the *MHCIIα* locus, the variation was nearly identical to the variation in the two benthic morphs of Lake Thingvallavatn. The results are consistent with a scenario where parts of the immune systems have diverged substantially among Arctic charr populations in Iceland, after colonizing the island ~10.000 years ago.

Citation: Kapralova KH, Gudbrandsson J, Reynisdottir S, Santos CB, Baltanás VC, et al. (2013) Differentiation at the *MHCIIα* and *Cath2* Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn. PLoS ONE 8(7): e69402. doi:10.1371/journal.pone.0069402

Editor: Gen Hua Yue, Temasek Life Sciences Laboratory, Singapore

Received January 3, 2013; **Accepted** June 9, 2013; **Published** July 24, 2013

Funding: The Palsson laboratory is supported by Icelandic Research foundation and the University of Iceland research fund. Icelandic research foundation (grant of excellence: nr 100204011) to S.S. Sigurdsson, A. Palsson, B.K. Kristjansson, Zophonias O. Jonsson and Ian A. Johnston paid for part of this work. Kalina H. Kapralova and Johannes Gudbrandsson were supported by the University of Iceland doctoral fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: apalsson@hi.is

Introduction

Processes of divergence and adaptation reflect evolutionary forces that alter the genetic make-up of populations over time [1]. While the bulk of these changes must be neutral, some are likely driven by natural selection. By identifying genes relating to adaptation we may be able to disentangle history, neutral forces and the contribution of positive and purifying selection on these evolutionary processes [2,3]. One approach to identify such loci is to dissect the molecular genetics of major adaptations in highly divergent species [4], another is to compare genetic architecture of adaptive traits between closely related species or populations [5]. One of the advantages in studying recent (or ongoing) divergence is that relatively few genetic changes differentiate populations or sibling species, compared to the vast number of changes separating major taxa. A potential downside to this approach is that, on short evolutionary time scale, divergence is mainly shaped by drift and fine tuning of preexisting adaptations. However, certain study systems have the advantage of rapid evolution, for instance when species respond to geographic catastrophes or when they colonize novel habitats [6,7].

Following the retreat of the last ice age cap (~10,000 years ago) anadromous and freshwater fishes in the northern hemisphere invaded and explored new habitats [8]. In some cases streams and lakes provided novel niches, which the colonizing populations may have adapted to. Multiple species (white fish, three spine sticklebacks, several salmonids) show signs of repeated adaptive changes in independent waterbodies [9–13], some of which have been dissected genetically [14–17].

Evolutionary Immunology of Fishes

The invasion into new habitats, changes from anadromous to "freshwater only" lifestyle, and sharing of habitat with other fishes provides novel challenges to the immune system of fishes [8]. The adaptive significance of immunological genes has been clearly illustrated. There are data supporting the role of frequency dependent selection, importance of local adaptation, the role of

generalist vs. specialist lifestyle and parasites, involvement in assortative/disassortative mating and even magic trait sympatric speciation as defined by [18], see [19] for review.

Fish possess both an adaptive and an innate immune system. The Major Histocompatibility complex (MHC) are cell surface molecules (class I on most cells and class II on specialized cells) that are involved in pathogen recognition and are central to adaptive immunity [20–22]. The MHCII is a heterodimer protein made of an α and a β chain, each with two domains (α1 and α2, β1 and β2 respectively). MHC genes have been identified in many teleost species and in general the β chain tends to be highly polymorphic [23]. The favoured explanation is that the multitude of infectious agents and environmental heterogeneity favours heterozygotes and rare alleles, which through balancing or frequency dependent selection result in high MHC diversity [19]. MHC allele diversity can be reduced in fish populations, as a consequence of local adaptation [24,25]. The distribution of *MHCIIα* alleles in Arctic charr is consistent with some degree of local adaptation [26], which will be studied further in this paper. Similarly data from brown trout (*Salmo trutta)* and Atlantic salmon (*S. salar*) how population differentiation in immunological genes, including TAP (Transporter associated with antigen processing) and interleukin-1 beta [27,28]. Curiously MHCII genes have been lost in Atlantic cod and related species [29], whereas in the Salmonidae they were duplicated along with the whole genome about 25–100 millon years ago [30]. There are two MHCII regions in Salmonids (observed in Atlantic salmon and rainbow trout (*Oncorhynchus mykiss*)), and evidence suggests at least four *MHCIIα* copies can be expressed [31].

The innate immunity system constitutes an evolutionarily old defense strategy, as the majority of gene families involved in it are present throughout the animal kingdom [32]. Innate immunity depends on a wide array of recognition, signal transduction and defence molecules, which are thought to evolve fast in response to pathogens. For instance, a comparison of 12 Drosophila species genomes revealed signs of positive selection on protein sequence and gene copy number in the sensory and effector genes of the innate immunity [33]. Innate immunity is considered to be of key importance in combating infections in fish [21,22]. Antimicrobial peptides (AMPs) play a major role in this system and in mammals these cationic peptides not only kill bacteria, but are multifunctional effectors of the innate immune system [34,35]. Many AMPs have been identified in fish including Cathelicidins (Cath), liver expressed antimicrobial peptides (LEAP) and hepcidins (HAMP) [36–40]. In salmonids two types of Cathelicidins have been identified; Cathelicidin 1 and 2 [39–41]. Cathelicidins are generally encoded by four exons with the exception of *Cathelicidin 2* (*Cath2*) in the *Salvelinus* genus, which have lost exon 3. In fish Cathelicidins expression increases due to bacterial infection and the mature antimicrobial peptide has been shown to have bactericidal activity [39,40,42–44]. Several studies have shown signs of positive selection on AMPs (reviewed by Tennessen [45]), specifically on the charged amino-acids. Population genetic studies of the AMPs and other innate immunity genes are needed to elucidate the distinct selection pressures that shape these ancient defense systems.

Arctic Charr Diversity and Resource Polymorphism

Arctic charr is a widespread circumpolar species. While it's distribution reaches south along the coastal areas of the N-Atlantic it is best described as an Arctic species and indisputably the most cold tolerant of the salmonids [46]. In the high north Arctic charr is often found in very cold waters and lakes with limited productivity and with few or no other fish species present. A body of ecological studies document high diversity among Arctic charr populations (e.g. refs. in [46–48]), and many instances of resource polymorphism within lakes (see refs. in [8,49,50]). The favored explanation is that diversity arises via ecological specialization in habitat use and diet, facilitated by relaxed inter-specific competition, leading to morphological divergence among and within lakes [8,51].

Icelandic Arctic charr descend from European charr [52] that colonized the island after the glacial retreat. Large parts of Iceland are constantly shaped by tectonic and volcanic activity which appear to have created special habitats for dwarf forms of Arctic charr that typically inhabit streams, ponds and lakes in the neo-volcanic zone that traverses Iceland from the south-west to the north-east. Kristjansson and coworkers have shown that in these habitats these small fish show similar phenotypes across locations, *e.g.* a typically benthic morphology, thus retaining a juvenile morphotype [53]. However, their evidence also shows that the morphological parallelism is incomplete [54,55]. In lakes with two or more distinct morphs they usually conform to two types in terms of morphology (i.e. morphotypes), a pelagic and a benthic type, that typically reflect their modes of habitat utilization. Multiple lines of evidence show that these differences stem both from environmental and genetic causes [56–58].

The best studied and most extreme example of sympatric charr morps are the four morphs in Lake Thingvallavatn [59]. Two large morphs are found, a large benthivorous (LB-charr) and a piscivorous morph (PI-charr), and two small forms (morphs), a small benthivorous (SB-charr) and planktivorous morph (PL-charr). PL- and PI-charr, display a pelagic morphotype and are more inclined to operate in open water and feed on free swimming prey, planktonic crustaceans and small fish, respectively. The two benthic morphs show a benthic morphotype and mainly reside on the bottom, feeding exclusively on benthic invertebrates. The very small size of the SB-charr also allows them to utilize interstitial spaces and crevices in the littoral zone typically consisting of submerged lava which offers a rich source of benthic invertebrate prey. As would be expected from the clear cut ecological diversification of the morphs their macroparasitic fauna differs distinctively [60].

Population genetic studies based on variation in mtDNA revealed a common ancestry of Arctic charr in the Nordic countries, Ireland and Iceland [52]. Within Iceland, allozyme, mtDNA and microsatellite data reveal significant genetic differences between localities and in some cases between sympatric morphs, like the four morphs in Lake Thingvallavatn [61–63]. The genetic differentiation among the Thingvallavatn morphs is rather weak however, the average F_{ST} over 10 microsatellites being 0.03, and a coalescence model suggests a scenario of early divergence with subsequent barriers to gene flow [63]. The strongest indication of genetic differentiation between sympatric charr morphs is a fixed difference in one microsatellite marker between two morphotypes in Lake Galtabol [64]. On a larger scale the available data suggest repeated evolution of dwarf forms (small fish with a benthic phenotype) in numerous Icelandic lakes and stream habitats in the neo-volcanic zone [53,63].

Molecular genetics have also been used to address the developmental basis of morphotype differences in Icelandic Arctic charr [65,66]. Macqueen and colleagues [66] conducted a study of the expression of 21 mTOR and growth regulation genes in 7 distinct Icelandic charr populations (thereof 5 with a small benthic morphotype), and revealed substantial divergence in gene expression of many pathway components. For instance mTOR is less and 4E-BP-1 more highly expressed in the populations of small benthic populations compared to other populations, a

finding consistent with the role of these genes in protein synthesis and growth regulation [55,66]. It is not clear whether those pathways are the foci of selection for changes in size and form, or realisitors of genes that promote dwarfism. Notably, considering our focus on immunological genes, the mTOR pathway is also involved in regulation of innate immunity [67,68].

We hypothesized that local differences in habitat use and diet between the morphs in Lake Thingvallavatn and among other Arctic charr populations and morphotypes in Iceland could impact variation in important immunological genes. Using samples from all major phylo-geographic groups of Arctic charr [52] Conejeros and colleagues [26] reported on rich allelic variation at the *MHCIIα* locus within and between charr populations. Their data showed considerable shared diversity within populations and across a broad geographic range, but are also consistent with differentiation among populations reflected in unique haplotypes and frequency differences. Here we present a study on a smaller geographic scale analyzing variation in *MHCIIα* and four other innate immunity genes in Icelandic Arctic charr. Our focus was on the three most common sympatric morphs from Lake Thingvallvatn and 9 populations of small benthic, anadromous and lake resident charr from the neo-volcanic zone (south, west and north) in Iceland – that we studied previously with 9 microsatellites [63]. Thus in this study we could interrogate local differences in gene frequencies and probe geographic patterns in these loci in small benthic charr in Iceland. The results indicate marked differentiation between sympatric morphotypes in Lake Thingvallavatn in two loci, *Cath2* and *MHCIIα* that we investigated further. Our findings have bearing on the understanding of those unique sympatric Arctic charr morphotypes, and immune system diversity in organisms with evolutionarily recent resource polymorphism.

Materials and Methods

Sampling

Specimens came from three collections of Arctic charr from Icelandic lakes and rivers. First, we utilized a sample of 30 large bentivorous charr (LB-charr, not sexed) caught on their spawning grounds at Olafsdrattur, and a total of 406 spawning small benthivorous charr (SB-charr, 102 females/83 males) and plantkivorous charr (PL-charr, 83 females/115 males) caught at Olafsdrattur and four other spawning locations in Lake Thingvallavatn in October 2005 (Table 1, Figure 1, inset) (for details see Kapralova *et al.* [63]). Second, we used another sample of 76 SB-charr (17 females/59 males), 102 PL-charr (51 females/males) and 17 LB-charr (1 female/16 males) collected in Olafsdrattur and Mjoanes, in September and October 2010 respectively. These two samples were pooled as our previous results [63] and the data from 2005, did not suggest genetic differentiation by location. The sampling in Lake Thingvallavatn focused on the SB and PL morphs, and the LB morph was mainly used for reference (hence the relatively lower sample size). For the 2010 sample, sex, fork length, weight, maturity and age were documented and parasite load (see below) assessed for every individual. DNA was extracted from a fin clip following a standard phenol-chloroform protocol. Third, we utilized samples from 9 populations of Arctic charr selected from a larger survey throughout Iceland collected in 2003–2006 (Table 1, Figure 1) previously described [63]. Those specimens were not sexed.

Fishing in Lake Thingvallavatn was with permissions obtained both from the owner of the land in Mjóanes and from the Thingvellir National Park commission. Ethics committee approval is not needed for regular or scientific fishing in Iceland (The Icelandic law on Animal protection, Law 15/1994, last updated with Law 157/2012). However, sampling was performed with University College Aquaculture Research Station (HUC-ARC) personnel. HUC-ARC has an operational license according to Icelandic law on aquaculture (Law 71/2008), that includes clauses of best practices for animal care and experiments.

Molecular Work and Data Processing

We screened for sequence variation in four immunological genes: *Cath2*, *Leap-2a*, *Hamp* and *MHCIIα* among the three Thingvallavatn morphs (SB-, PL- and LB-charr). Moreover we studied a 510 bp region of the D-loop (starting at base 25 in the *S. alpinus* mtDNA reference genome, accession number NC_000861.1) as a putative neutral marker or marker of maternal lineage sorting. Loci were amplified by PCR with TEQ polymerase (Prokaria-Matis). We used previously published primers for *MHCIIα* [26] and new primers for *D-loop*, *Leap-2a*, *Hamp* and *Cath2* (Table S1), designed with Primer3 (http://primer3.wi.mit.edu/[69]). The following PCR program was used for all primer pairs, except *MHCIIα*. Denaturation at 95°C for 5 min; 35 cycles of 95°C for 45 seconds; 45 seconds at a marker specific annealing temperature (Table S1); 1 min at 72°C, then a final step of 10 min at 72°C. For *MHCIIα* we used touchdown PCR, initial denaturation at 94°C for 5 min; 16 cycles of 94°C for 45 seconds, 62°C for 45 seconds (decreasing by 0.5°C every cycle), 1 min at 68°C; followed by 25 cycles of 94°C for 45 seconds, 53°C for 45 seconds, 1 min at 68°C; then a final step of 10 min at 68°C. PCR products were ExoSap purified, sequenced (BigDye) and run on an Applied Biosystems 3500xL Genetic Analyzer (Hitachi).

Raw sequencing data was base-called by Sequencing Analysis Software v5.4 with KBTMBasecaller v1.41 (Applied Biosystems), and run through Phred and Phrap [70], prior to trimming primer sequences, visual editing of ambiguous bases and putative polymorphisms in Consed [71]. Fasta files were exported and aligned with ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/, [72]) and manually inspected for alignment errors in Genedoc (www.psc.edu/biomed/genedoc) [73]. All sequences where deposited as Popsets in Genebank under the accession numbers KC590653-KC591103, KC591105-KC591218, KC591220-KC591303, KC591303-KC591626 and KC596075-KC596117.

Genotyping *MHCIIα*

Due to potential duplications or deletions of *MHC* genes and the ancestral genome duplications in salmonids [30] the presence of *MHC* paralogous genes has to be investigated in charr. Initially we used the SAALDAA primers from Conejeros *et al.* [26], (Table S1) that pick up part of exon 2 and intron 2 of *MHCIIα*, but obtained several satellite bands. To confirm the amplification of MHC, bands of various sizes (from a non-optimized PCR) were cloned into a TOPO vector (Invitrogen) and sequenced. Blastn was used to find related sequences in Genebank (NCBI – nucleotide collection – at latest in April 2013). We obtained bands from 4 size ranges. Most importantly, a ~400 bp fragment sequenced from 2 individuals (10 clones from each) yielded 3 different fragments of *MHCIIα* (Table S2). One of these fragments, represented by 5 clones from each individual, was 99% identical to Saal-DAA*0801 [26]. The other two versions, each restricted to one individual, had 98% and 99% identity to Saal-DAA*0305/0306/0307 and Saal-DAA*0305 [26], respectively (Table S2). The largest band (~720bp) was only present in ~1% of the samples and all ten clones from this band were identical to *MHCIIα* haplotype Saal-DAA*0104 (intron haplotype hap1 as defined by [26]. Two smaller fragments, ~250 bp and ~150 bp, contained mixed products of various origins unrelated to *MHCII*.

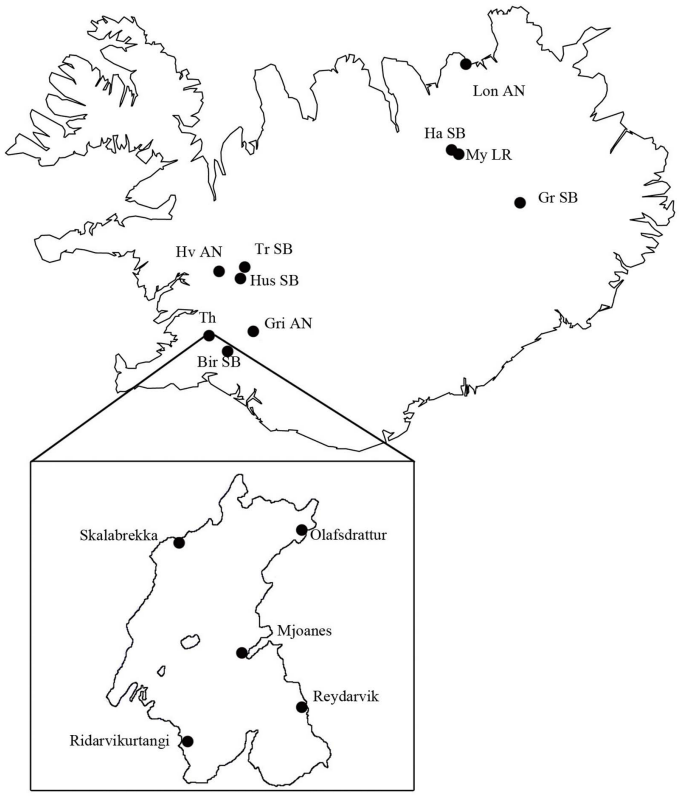


Figure 1. Sampling locations of Arctic charr in Lake Thingvallavatn and around Iceland. Fishes where collected in five locations within Lake Thingvallavatn (left), and from 9 other locations and populations around Iceland. In Lake Thingvallavatn, O: Olafsdrattur, M: Mjoanes, Re: Reydarvik, R: Ridvikurtangi and S: Skalabrekka. Around the island, either small benthic (SB) and lake resident (LR) or anadromous (AN) charr in Myvatn (My, LR), Haganes (Ha, SB), Lon (Lo, AN), Grafarlond (Gr, SB), Grimsnes (Gr, AN), Birkilundur (Bir, SB), Hvita (Hv, AN), Trussa (Tr, SB) and Husafell (Hus, SB).
doi:10.1371/journal.pone.0069402.g001

The PCR protocol was optimized to reduce unspecific small auxiliary bands (see above) and we proceeded with PCR and direct sequencing. The first 32 *MHCIIα* sequences from Lake Thingvallavatn (2005 sample) were amplified with the SAALDAA primers, and sequenced with both forward and reverse primers (error rate of Single nucleotide polymorphisms (SNP's) called was <0.1%). Subsequently only the forward primer was used to sequence the PCR products. In total PCR and direct sequencing of 413 individuals from the 2005 sample gave sequences of three major types. Those corresponded to the large fragment (intron haplotype hap1) and the two versions (similar to Saal-DAA*0303 and Saal-DAA*0305), that we denote as second intron haplotypes 14 and 15. The fragment identical to Saal-DAA*0801 was never observed. PCR and direct sequencing clearly revealed individuals heterozygotic for a single base insertion/deletion polymorphism (indel) in the intron. To us the data suggest that two *MHCIIα* paralogous genes are present in Arctic charr, with hap14, hap15 and possibly hap1 being alleles of one paralog. The optimized PCR preferentially amplifies this paralog. This is supported by two observations. First, in the direct sequencing we never observe Saal-DAA*0801 (the two suspected paralogs are easy to distinguish) and second, the indel in the second intron conforms to Hardy Weinberg Equilibrium, within each morph (see below).

Because of low DNA availability and degradation in the 2005 Icelandic lake samples, we designed new primers (Table S1.) that gave a shorter amplicon and none of the satellite bands. With

Table 1. Details on sampling locations and the number of individuals collected in 2005 and 2010.

Location	Morphotype	Code	Latitude	Longitude	2005	2010
Thingvallavatn	Large benthic	TH_LB	64°11	21°08	30	17
Thingvallavatn	Small benthic	TH_SB	64°11	21°08	185	76
Thingvallavatn	Planktivorous	TH_PL	64°11	21°08	198	102
Grimsnes	Anadromous	Gri_AN	64°00	20°53	27	
Birkilundur	Small benthic	Bir_SB	64°01	20°57	30	
Hvita	Anadromous	Hv_AN	64°42	20°59	35	
Trussa	Small benthic	Tr_SB	64°43	20°46	29	
Husafell	Small benthic	Hus_SB	64°41	20°52	31	
Lon	Anadromous	Lon_AN	66°05	16°55	27	
Grafarlond	Small benthic	Gr_SB	65°15	16°09	31	
Myvatn	Lake resident	My_LR	65°37	17°03	34	
Myvatn-Haganes	Small benthic	Hag_SB	65°37	17°03	35	

doi:10.1371/journal.pone.0069402.t001

those primers fragments of *MHCIIα* from 6 individuals were amplified, cloned and sequenced (as before). We sequenced on average 8 clones per individual and in all cases the genotyping was in perfect concordance with the genotyping from PCR and direct sequencing. The suspected paralogous copy of *MHCIIα* (similar to Saal-DAA*0801) was found in a low proportion of the clones (5/45 sequences). The 2010 sample from Lake Thingvallavatn and the 9 Iceland wide populations were amplified and sequenced with these primers. Although there is a potential for ascertainment bias, as samples from two years (2005 and 2010) were genotyped with different primers, the results do not indicate a bias; the frequency of the indel variation was not statistically different between years (tested within morphs, see details below). Finally, we also did a restriction enzyme analysis, that could distinguish hap14 and hap15 on basis of a G/A polymorphism 13 bp down stream of the indel (TGAATGAATCAATAGGATTAATGTAGTAAA(A/−)TAGTCACCTCACT(G/A)TAACCTCTCACATGTTG-TATCATCTGTGGTATGG). These two polymorphisms were fully coupled in the sequencing data. This restriction digest of 28 individuals (equal number from 2005 and 2010) was in perfect concordance with the PCR and sequencing data.

Population Genetic Analyses

Tassel version 2.0.1 (www.maizegenetics.org) [74] and DNAsp 4 (www.ub.edu/dnasp/) [75] were used to calculate and analyze population genetic statistics. Tests of Hardy Weinberg proportions, allele and genotype frequencies between morphs, locations and were implemented in R (version 2.12, R Development Core Team, 2011). Arlequin v3.5.1.2 was also used to estimate F_{ST} [76–78]. We tested determinants of genetic differentiation between morphs within Lake Thingvallavatn with analyses of molecular variance (AMOVA) using Arlequin. We analyzed variation in 3 amplicons (*D-loop*, *Cath2* and *MHCIIα*), within Lake Thingvallavatn with a two level AMOVA with morph (LB, SB, SP) as a categorical variable, split by sex or sampling location.

The genetic relationships between and within morphs were estimated with an unrooted neighbor-joining tree. The tree was constructed using Cavalli–Sforza's genetic distances obtained from nine microsatellite loci [63] with the program NEIGHBOUR available in PHYLIP3.69 [79]. Confidence intervals were estimated by 1000 bootstrap replicates.

Parasite Analyses

The 2010 samples from Lake Thingvallavatn were used to assess infection rates and loads of the eye parasite *Diplostomum sp.*, the intestine parasite *Eubothrium salvelini*, Nematodes and *Diphyllobothrium sp*. Both eyes were extracted from each individual. The contents of each eye was poured on a flat slide, covered with a slip and processed under a Leica KL200 LED microscope at 2X magnification. The slide field was divided into 45 blocks, and the average number of metacercaria of *Diplostomum sp*. was estimated. We first screened all blocks, and in case of even distribution among them, counted the metacercaria in 5 randomly selected blocks, and then calculated average infection rate. In case of non-uniform distribution or low infection we counted the parasites in all 45 blocks. We recorded both counts and used an infection scale [60] ; 0 = total absence of parasites; 1 = 1 or fewer parasites per blocks; 2 = 1 to 3 individuals per block; 3 = 4 to 10 parasites per block and 4 represented more than 10 *Diplostomum sp*. individuals per block. The estimation was done by a single observer (S. Reynisdottir) on a single eye per specimen. The correlation of infection rate between eyes was high (Pearson $r = 0.75$, $p<0.005$, for 25 pairs of eyes studied).

Infections by *Eubothrium salvelini* were assessed by carefully extracting the liver, stomach and intestine and documenting the presence or absence of the adult tapeworm. Infections of nematodes and plerocercoids of *Diphyllobothrium sp*. were estimated by counting individual nematodes and *Diphyllobothrium* cysts internal cavities and linings of flesh [60,80]. The *Diphyllobothrium sp*. infection rate was scored using the following infection scale: 0 = the total absence of parasites; 1 = 1 to 3 per individual; 2 = 4 to 7 per individual and 3 equaled more than 8 parasites per individual. For Nematodes the number per individual was recorded. All data on intestinal parasites were obtained by a single observer (C. B. Santos). Data of the 2010 and 2005 samples from Lake Thingvallavatn were deposited in the Dryad Repository: http://dx.doi.org/10.5061/dryad.81884.

Statistical Analyses of Parasite Infections

Statistical analyses were performed in R. The effects of morph, sex and weight on the load of individual parasite species was investigated with multivariate regression. Summary statistics were calculated for weight, age and parasite loads separately for each morph. Sex ratio was also calculated. For *Diphyllobothrium sp*. and

Table 2. Polymorphism in the mitochondrial D-loop and three immunological genes.

Gene/region	Morph	Size (bp)	N	S	Indel	π	θ	Haplotypes
D-loop	All	509	406	4	0	0.001	0.001	7
	PL	509	190	3	0	0.001	0.001	4
	SB	509	216	4	0	0.001	0.001	7
Hamp 5′ UTR	PL/SB*	454	12	0	0	0.000	0.000	1
Leap-2a 3′ UTR	All	559	15	4	1	0.001	0.004	3
	PL	559	8	2	0	0.001	0.003	2
	SB	559	7	3	1	0.002	0.003	3
Cath2 (intron 2)	PL/LB/SB*	219	258	0	0	0.000	0.000	1
Cath2 (peptide)	All	396	258	3	0	0.001	0.001	4
	PL	396	138	2	0	0.000	0.001	3
	LB	396	35	1	0	0.000	0.001	2
	SB	396	86	3	0	0.001	0.001	4
Cath2 (3′ UTR)	All	407	17	2	1	0.002	0.002	3
	PL	407	6	1	0	0.002	0.001	2
	SB	407	11	2	1	0.002	0.002	3

S: Segregating sites. Indel: Segregating insertion/deletion polymorphism. π: The average number of nucleotide differences per site. θ: Wattersons estimator of diversity per site. *The data from different morphs are summarized together as no differences in frequence were observed.

doi:10.1371/journal.pone.0069402.t002

Diplostomum sp. mean relative density (MRD) was calculated [60]. Statistical models for parasite load were applied to morph pairs to test for difference between the morphs. As parasite loads turned out to be different between morphs tests for other factors affecting the load were applied to the morphs separately. The models had the general structure:

$$Parasite\ load = Sex + Weight + Age + Error.$$

A term for genotype was also added to evaluate the impact of *MHCIIα* variation within morphotypes. The ANOVA function from the *car* package [81] was used to perform F-tests and log-likelihood tests. Raw counts of *Diphyllobothrium sp.* and *Diplostomum sp.* were analyzed by multivariate linear regression and variable effects tested with an F-test. The infections were also summarized with an infection scale [60] and analyzed using multinomial logit regression fitted with neural networks [82], with consistent results. Effects were tested with log-likelihood tests. Logistic regressions were applied to Nematodes and *Eubothrium salvelini* occurrence and effects were tested with log-likelihood test.

Results

Nucleotide Polymorphism in Arctic Charr Morphs in Lake Thingvallavatn

Different molecular markers have revealed significant but weak genetic differentiation among the Lake Thingvallvatn charr morphs [61–63]. Here we make use of genetic material from individuals previously typed for 9 microsatellite markers [63] to explore variation in four immunological loci, and test for indications of population differentiation.

Four segregating sites were observed in the mitochondrial *D*-loop, but nucleotide diversity was rather low (Table 2). Of the four substitutions only one (m38A>G) had significant difference in frequency between PL and SB ($\chi^2_{[1]} = 9.36$, $p = 0.002$). The $F_{ST} = 0.001$, which was lower than the F_{ST} for microsatellites between charr morphs in Lake Thingvallavatn [63]. A comparison with *S. alpinus D*-loop in genebank [52,83] shows that none of the four *D*-loop sites are restricted to Iceland. Analyzes of molecular variance (AMOVA) confirm that the observed variation in this part of the mtDNA of Lake Thingvallavatn charr is not affected by morph, sex or sampling location (Table 3).

We screened three innate immunity genes *Hamp*, *Leap-2a* and *Cath2* for nucleotide variation. The 454 bp *Hamp* amplicon, positioned in the untranslated 5′-region, proved invariant in a set of 12 specimens (6 PL- and 6 SB-charr). Four segregating sites and one insertion/deletion polymorphism (indel) were found in the 3′UTR of *Leap-2a*. These were at approximately equal frequency in SB- and PL-charr. The *Hamp* and *Leap-2a* genes where not studied further. Of the three regions surveyed in the *Cathelicidin* gene (spanning ~1 kb), only the peptide region showed frequency differences between morphs (Table 2) urging further investigation. The sequenced part of intron 2 was invariant in the sample, whereas the three mutations (one indel and two SNPs) in the 3′UTR were at about the same frequency in both morphs.

Sequencing of the antimicrobial peptide encoding region of *Cath2* in 264 individuals from Lake Thingvallavatn 2005 revealed three variant sites (including one singleton). One mutation (g558C>A) was found in intron 2. Another (g819C>A) was found in the exon encoding the mature antimicrobial peptide (in cathelicidins this region is on exon 4, but due to the lack of exon 3 in charr *Cath2* [40], it is encoded by the third exon in *S. alpinus*, Figure 2A). This mutation is predicted to lead to an amino acid replacement in the mature peptide (replacement of arginine by serine at position 115, Figure 2B). This alters the charge of the peptide, from +8 to +7.

We compared the frequency of the two mutations among morphs, sex and sampling locations in Lake Thingvallavatn. The g558C>A is largely restricted to the SB morph (11.3% frequency); it is not found in the LB-charr and only present in two of 134 PL-charr. The more common g819C>A variant shows significant frequency differences between morphs ($\chi^2_{[2]} = 43.91$, $p<0.0001$). The A allele is at 27% frequency in SB-charr, but is rarer in LB-

Table 3. Analyses of molecular variance (AMOVA) of three loci by morphotypes (PL, LB and SB collected in 2005) and either location or sex.

Gene	Terms	d.f.	Sum of squares	Variance	Variation (%)	Fixation index	p-value
D-loop*	Among morphs	1	0.3	0	0.83	FSC : −0.01	ns.
	Among locations within morphs	7	0.39	0	−1.44	FST : −0.01	ns.
	Within locations	389	58.58	0.15	100.62	FCT : 0.01	****
	Total	397	59.27	0.15			
	Among morphs	1	0.17	0	0.4	FSC : −0.01	ns.
	Among sexes within morphs	2	0.11	0	−0.66	FST : 0	ns.
	Within sexes	393	58.17	0.15	100.26	FCT : 0	****
	Total	396	58.45	0.15			
Cath2	Among morphs	2	4.56	0.02	12.64	FSC : 0.03	****
	Among locations within morphs	7	2.2	0.01	3.02	FST : 0.16	**
	Within locations	253	42.18	0.17	84.34	FCT : 0.13	*
	Total	262	48.94	0.2			
	Among morphs	2	4.56	0.03	13.48	FSC : 0.01	****
	Among sexes within morphs	2	0.45	0	0.47	FST : 0.14	ns.
	Within sexes	258	43.94	0.17	86.05	FCT : 0.13	****
	Total	262	48.94	0.2			
MHCIIα	Among morphs	2	50.76	0.22	63.2	FSC : 0.03	****
	Among locations within morphs	8	0.88	0	−0.13	FST : 0.63	ns.
	Within locations	402	50.93	0.13	36.92	FCT : 0.63	***
	Total	417	102.57	0.34			
	Among morphs	2	51.44	0.22	64.06	FSC : −0.01	****
	Among sexes within morphs	2	0.04	0	−0.33	FST : 0.64	ns.
	Within sexes	408	50.91	0.12	36.28	FCT : 0.64	****
	Total	412	102.39	0.34			

*Only PL and SB were sequenced for the D-loop. d.f.: Degrees of freedom. Significance: ns. $p>0.05$,
*$p<0.05$,
**$p<0.01$,
***$p<0.001$,
****$p<0.001$.
doi:10.1371/journal.pone.0069402.t003

(5.7%) and PL-charr (6.4%). This translates into an F_{ST} of 0.17 ($p<0.0001$) between the SB- and PL morphs, and $F_{ST}=0.13$ ($p<0.0001$) between the LB and SB samples. No differences in allele frequency where found between PL- and LB-charr, sexes or sampling locations. Analyses of Molecular Variance (AMOVA) confirmed these patterns (Table 3).

MHCIIα Variation in Lake Thingvallavatn

Due to the structural richness of MHC regions and the fact that the common ancestor of salmonids underwent a whole genome duplication, studies of MHC variation in those species are rather complicated. We tackled this by genotyping with PCR and direct sequencing, and assessed the specificity and reproducibility of this genotyping method by cloning and restriction enzyme assays.

We concentrated on the highly variable intron 2 of *MHCIIα* [26], by DNA sequencing of 413 charr (LB, SB and PL) from Lake Thingvallavatn. There was high degree of polymorphism, with many segregating mutations (10 SNPs and 2 indels in ~300 bp). Two major and two minor versions of *MHCIIα* were identified. The two major haplotypes hap 14 and hap 15 are quite distinct, being separated by 6 segregating sites and 1 indel. These polymorphism were were described by Conejeros *et al.* [26], but the haplotypes involving them are unique and probably arose by recombination. In addition two rare versions were observed, hap16 (just one site diverged from hap14) and hap1 (Saal-DAA*0104) which contains a Hpa retrotransposon [26]. The hap1 and hap16 haplotype were extremely rare in all morphs, for instance hap1 was found in four SB-charr from 3 sampling locations (1.08%) and one LB-charr (1.67%). Our analyses focused on the two dominant haplotypes, hap14 and hap15.

As described in Materials and Methods, the cloning results suggest the presence of two distinct *MHCIIα* paralogs in Arctic charr in Iceland. One of these was never observed with the PCR and direct sequencing, but only detected in the cloning (prior to PCR optimization). The hap14 and hap15 haplotypes are readily distinguishable based on several markers, such as the indel in the intron. We are quite certain that these are allelic variations (true haplotypes, not paralogous genes) because Hardy Weinberg proportions are respected for the indel polymorphism in *MHCIIα* intron in all three morphs in Lake Thingvallavatn (LB: $\chi^2_{[1]}=0$, $p=1$, SB: $\chi^2_{[2]}=1.77$, $p=0.4$, PL: $\chi^2_{[2]}=6.2$, $p=0.05$). Furthermore restriction enzyme analysis of 28 individuals was in perfect concordance with the PCR and sequencing data.

As predicted [19,27] the nucleotide diversity was higher in *MHCIIα* than in the other sequences studied; π was an order of magnitude higher than for *Cath2* and the D-loop (Table 2 and

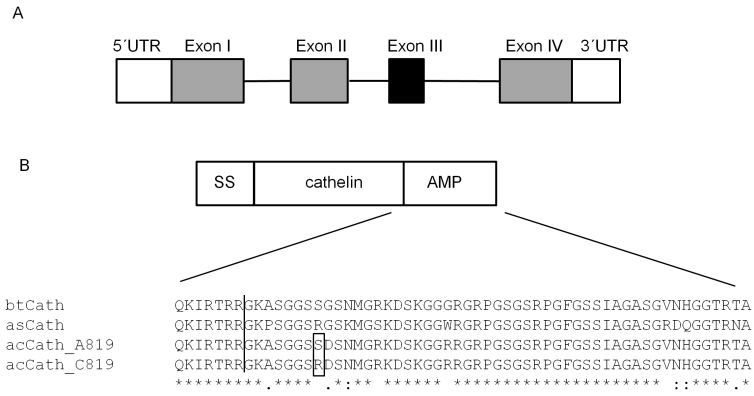


Figure 2. Polymorphism in the antimicrobial peptide Cathelicidin 2. *Cathelicidins* have a conserved 4 exon structure (A) with the exception of the *Salvelinus* Cathelicidins type 2 which have lost exon 3 (marked black). The peptides (B) are produced as pre-pro-peptides, where exon 1–3 encode the signal sequence (SS) and the conserved Cathelin region, while exon 4 encodes the processing site and the mature antimicrobial peptide (AMP). An amino acid alignment of this region for Cathelicidin 2 of Atlantic salmon (asCath), brook trout (btCath) and Arctic charr (acCath) shows the predicted processing site (vertical line) and the observed polymorphism (predicted peptide position 115) in Icelandic Arctic charr. Identical amino acids are marked with *, amino acids with a similar and somewhat similar function are marked with : and. respectively.
doi:10.1371/journal.pone.0069402.g002

below). We found large differences in *MHCIIα* frequencies among the three morphs studied from Lake Thingvallavatn (Table 4), with hap15 being dominant in both benthic morphs, 93.5% and 88.3% in SB- and LB-charr respectively. In contrast hap15 was at 22% frequency in the pelagic morph (PL). This translates into an F_{ST} of 0.56 ($p<0.0001$) between PL- and LB-charr, 0.67 between PL- and SB-charr ($p<0.0001$), and unsignificant F_{ST} between the two benthic morphs. This represents the strongest genetic differentiation reported to date between any of these three sympatric morphs. These findings were further supported by AMOVA, the effect of morphotype (benthic versus pelagic) dominating the explained variance (above 60%), while sex and sampling location did not have significant effects (Table 3).

This strong difference in *MHCIIα* frequency between morphotypes prompted several questions. Is the frequency difference consistent between years? What is the geographic distribution of variation in *MHCIIα* within Iceland? Do the haplotypes correlate with phenotypic attributes? We set out to answer these questions. Some hypotheses of MHC evolution involve temporal dynamics, for instance due to frequency dependent selection [19]. To evaluate this we used two approaches. We first compared the frequency of *MHCIIα* hap14 in the three morphs (PL-, LB- and SB-charr) in two cohorts sampled in 2005 and 2010 (Table 4). On all three morphs the haplotype frequencies were similar for the two years, $\chi^2_{[1]} = 0.0301$, $p = 0.9$, $\chi^2_{[1]} = 0.08$, $p = 0.8$, $\chi^2_{[1]} = 3.65$, $p = 0.06$, for PL-, LB- and SB-charr, respectively. The age distribution was similar in the fishes collected, for instance the average age in PL sampled in 2005 and 2010 was 6.94 and 6.96 years respectively (weighted t-test, $p = 0.96$). No significant differences in haplotype frequency between years ($\chi^2_{[2]} = 0.59$, $p = 0.74$) or age-classes were observed within morphs (Figure 3) ($\chi^2_{[12]} = 17.4$, $p = 0.13$ for 2005 and $\chi^2_{[12]} = 10.5$, $p = 0.57$ for 2010).

Cath2 and *MHCIIα* Polymorphism Across Morphotypes and Geographic Regions

As the frequency of variants both in *Cath2* and *MHCIIα* deviated significantly between morphs within Lake Thingvallavatn, we

Table 4. The frequency of the three most common *MHCIIα* haplotypes in the arctic charr morphotypes from Lake Thingvallavatn sampled in 2005 and 2010.

	LB		SB		PL	
Haplotypes	2005	2010	2005	2010	2005	2010
hap1	1 (1.7%)	0 (0.0%)	4 (1.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
hap14	6 (10.0%)	5 (14.7%)	20 (5.4%)	2 (1.3%)	309 (78.0%)	155 (76.0%)
hap15	53 (88.3%)	29 (85.3%)	346 (93.5%)	150 (98.7%)	87 (22.0%)	49 (24.0%)
N	30	17	185	76	198	102

N: the total number of fishes genotyped in each sample.
doi:10.1371/journal.pone.0069402.t004

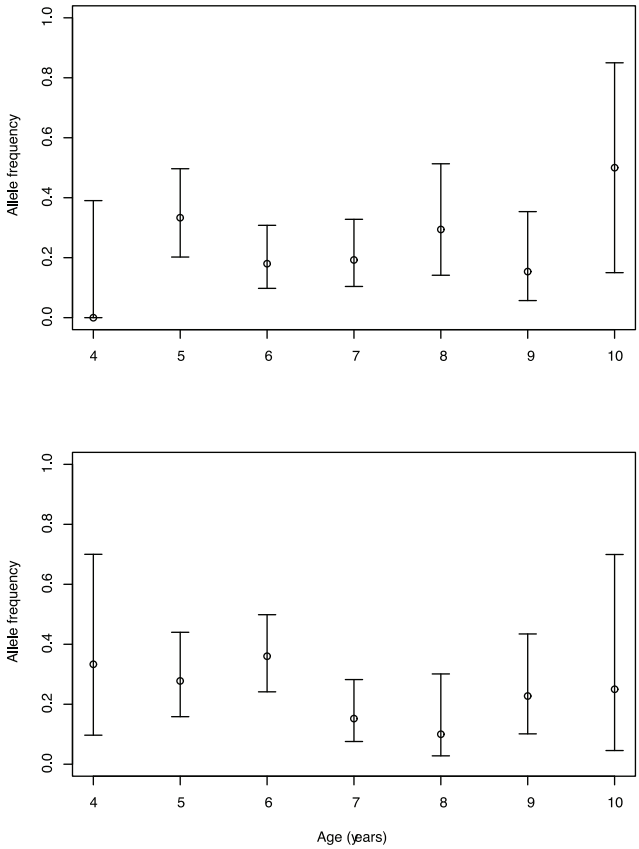


Figure 3. Frequency of *MHCIIα*variations in PL-charr from 2005 and 2010 by age classes. The frequency of *MHCIIα* hap14 (with 95% confidence intervals) by age of PL charr, collected in years 2005 (A) and 2010 (B) at the spawning grounds in Lake Thingvallavatn.
doi:10.1371/journal.pone.0069402.g003

wanted to know if the observations reflect a local or a broader geographic or ecological pattern. Our previous microsatellite study [63] enabled inference of relatedness among 9 Arctic charr populations from the north, west and south of Iceland (Figure 1 and 4A). We surveyed variations in both genes in those small benthic, anadromous and lake resident populations and superimposed on the microsatellite based tree.

There was very little polymorphism in *MHCIIα* in other populations and lakes, at maximum 3 haplotypes in each population (Table 5). The hap14 haplotype which dominated in the PL in Lake Thingvallavatn was only found in one other population (SB from Husafell), at 3% in 2 individuals (Figure 4B). The other haplotype (hap15), most common in the LB and SB morphs in Lake Thingvallavatn, dominated all other populations (average frequency 94%, lowest 81%). Several other haplotypes were observed, but all are one or few bases removed from hap15 and at very low frequency. The results show clearly reduced variation in this locus in Icelandic stocks of Arctic charr, except in the sympatric morphs in Lake Thingvallavatn. Summaries of nucleotide diversity reveal this pattern, as π (which responds to frequency and diversity of haplotypes) is larger in PL-charr from Lake Thingvallavatn than in the other charr populations surveyed (Table 5).

The *Cath2* g819C>A was genotyped in 7 populations (105 individuals total) and its frequency differed significantly between them (χ^2 [6] = 91.92, p<0.0001, Figure 4C). The g819C>A was dominant and even fixed in several small benthic charr populations (Birkilundur 100%, Haganes 86% and Grafarlond

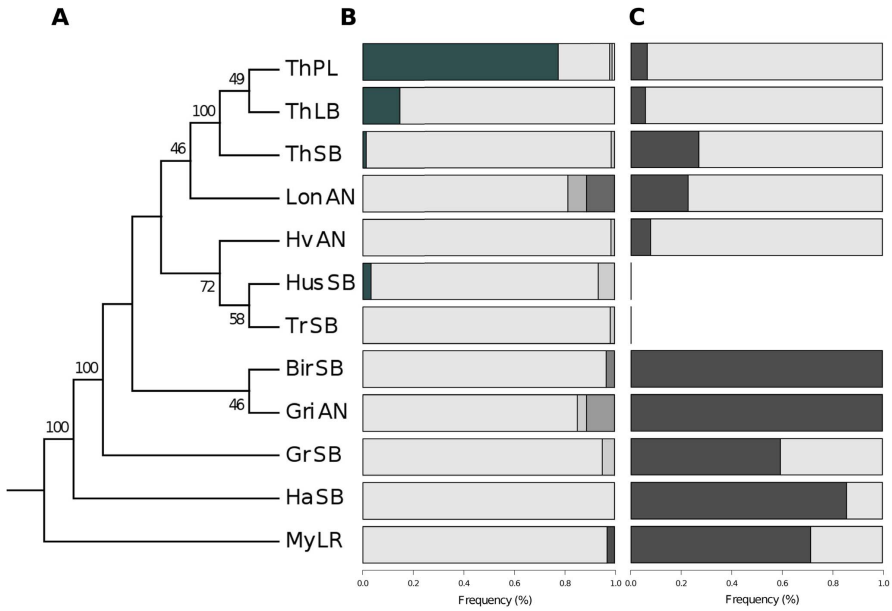


Figure 4. Arctic charr population history and variation in *Cath2* and *MHCIIα*. A) A genealogy of the sampled populations was built from 9 microsatellite markers and the confidence intervals were estimated by 1000 bootstrap replicates [63]. B) Frequencies of the *MHCIIα* intron haplotypes (hap 14 is dark, hap15 light gray, rare haplotypes are in intermediate shades of gray). C) The frequency of *Cath2* g819A (dark). Due to limited DNA available, the marker could not be typed in Husafell and Trussa. The same individuals where genotyped for all markers.
doi:10.1371/journal.pone.0069402.g004

59%). Recall, within Lake Thingvallavatn the variant was at highest frequency in the SB morph (27%), but lower in the other two. However g819C>A was also fixed in the anadromous Grimsnes population in the south of Iceland, and at high frequency in the lake resident population of large charr in Myvatn (71%) in the north. This translates into high interlocal F_{ST}, for instance 0.85 between the anadromous populations in Hvita and Grimsnes. The average F_{ST} for *Cath2* among all the populations was 0.29, while the average F_{ST} for microsatellites was 0.245 [63]. While the frequencies of the *Cath2* g819A certainly differ between the populations, the *Cath2* locus is not associated with morphotype, as for instance g819A is fixed in both anadromous and small benthic populations. However, the *Cath2* variation may correspond, to some extent, to the relatedness of populations (Figure 4). Note however that not all branches in the tree have strong bootstrap support. Finally, there is no concordance between the variation in the two loci (*MHCIIα* and *Cath2*), and no linkage disequilibrium was observed between *Cath2* and *MHCIIα* variations and the microsatellites (χ^2 [2] = 0.11, p = 0.94).

Tests of Association between *MHCIIα* Variation and Macroscopic Parasitic Infections

Frandsen and colleagues [60] reported a difference in parasite infection rate and prevalence between the four morphs in Lake Thingvallavatn. Can the differences in *MHCIIα* allele frequencies between the PL morph and the benthic morphs in Lake Thingvallvatn be driven by habitat-specific selection, caused by marked differences of infectious agents in habitat and diet? In immunity MHCII presents antigens of pathogens such as parasites [20], which may lead to evolutionary change [19]. We tested whether the *MHCIIα* variation is related to infection rate/prevelance of four classes of macroscopic parasites (*Diphyllobothrium sp.*, *Diplostomum sp.*, parasitic nematodes and *Eubothrium salvelini*), in Lake Thingvallavatn charr. We sampled PL- (102), SB- (76) and LB charr (17) in the fall of 2010, screened for parasites and ascertained *MHCIIα* haplotypes. The pattern of parasite infection rate and prevalence (Table 6) is consistent with previous reports [60], with the *Diplostomum sp.* being most common in LB- and SB-charr, but the other three parasites infecting a very high fraction of PL-charr. This was confirmed by a generalized linear models analyses (Table 7), which also revealed the effects of age (*Eubothrium salvelini* in PL charr, *Diplostomum sp.* in SB- and LB charr), weight (*Diplostomum sp.* in SB- and LB charr and *Diphyllobothrium sp.* in PL charr) and sex (only significant for Nematodes in PL charr). We added a term for the genotype, to test the effects of *MHCIIα* on each of those parasite types. This was only done for the PL morph as there was almost no segregating variation in the benthic morphs. The genotype terms were not

Table 5. Nucleotide diversity in *MHCIIα* in Lake Thingvallavatn 2010 sample and 9 other populations around Iceland.

Location	Size (bp)	S	π	θ	*Haplotypes*
TH_LB	293	8	0.013	0.014	3
TH_SB	293	8	0.003	0.010	3
TH_PL	293	10	0.018	0.014	4
All Lake Thingvallavatn	293	10	0.024	0.012	4
Gri_AN	293	7	0.008	0.011	3
Bir_SB	293	2	0.001	0.003	2
Hv_AN	293	3	0.002	0.004	2
Tr_SB	293	3	0.002	0.005	2
Hus_SB	293	8	0.009	0.013	3
Lon_AN	293	3	0.003	0.005	3
Gr_SB	293	3	0.005	0.005	2
My_LR	293	2	0.001	0.003	2
Hag_SB	293	0	0.000	0.000	1
All Iceland w/o Lake Thingvallavatn	293	12	0.003	0.011	8

S: the total number of segregating sites. π: The average number of nucleotide differences per site. θ: Wattersons estimator of diversity per site. See Table 1 for population identification code.

doi:10.1371/journal.pone.0069402.t005

significant, neither as a class or quantitative variable (Table 7). The models were evaluated both on a parasite-scoring-scale and raw counts, with consistent results (Table 7). For exploration we also tested interaction of genotype with other terms, which yield borderline significance for Genotype by Sex interaction with nematodes ($p = 0.07$). Considering the number of tests preformed and the poor replicability of genetic interaction terms [84] this is almost certainly a spurious association. In summary, the data do not suggest that infection rate (or infection intensity) of those four parasite classes is affected by the frequency of *MHCIIα* alleles in Lake Thingvallavatn charr.

Table 6. Parasite infection rate in Lake Thingvallavatn Arctic charr in 2010.

		Morph		
Parasite	**Measure**	**SB**	**LB**	**PL**
Diphyllobothrium sp.	MRD	0.07	0.02	1.25
	Prevalence	15/113	5/19	125/131
	Count	0.22	0.4	10.13
	Score	0.15	0.37	2.32
Diplostomum sp.	MRD	46.84	8.69	9.93
	Prevalence	109/113	19/19	131/131
	Count	192.1	178.3	70.3
	Score	2.10	2.26	1.53
Nematodes	Prevalence	1/82	1/15	55/105
Eubothrium salvelini	Prevalence	6/82	2/15	68/105

MRD: mean relative density. Both count and score are summarized by arithmetic means.

doi:10.1371/journal.pone.0069402.t006

Discussion

The sharp distinction in form, size and ecology between the four sympatric Arctic charr morphs in Lake Thingvallavatn [8,59] calls for explanation. Earlier studies found evidence of subtle but significant genetic differentiation among the morphs within the lake [50,61–63,85]. Here we report substantial genetic differentiation among the morphs within the lake, in two of the four immunological genes investigated (*Cath2* and *MHCIIα*). The pattern of divergence is not the same for both loci. In *Cath2* the strongest differentiation is between SB charr and the other two morphs studied (LB- and PL charr). Whereas in the case of *MHCIIα* the PL charr deviates markedly from the two benthic morphs within the lake, which have very similar haplotype frequencies. No differentiation was detected in two other innate immunity genes (*Hamp* and *Leap-2a*) nor the D-loop. The lack of association between mtDNA haplotypes and morphotypes, is consistent with results on variation in Arctic charr (dwarf and large forms) in 56 Siberian lakes [83]. Allele frequency differences can be caused by neutral and selective forces, but several studies have documented the impact of selection on immunological genes, with most focus on MHC loci [19,33,45].

Which Evolutionary Forces Shaped the *MHCIIα* and *Cath2* Variation in Iceland?

We observe large frequency differences of the *MHCIIα* haplotypes in the three sympatric morphs in Lake Thingvallavatn. The highest F_{ST} was 0.67 between PL- and SB charr, while the F_{ST} was 0.03 on average for 10 microsatellites between these morphs [63]. This is in contrast to very little difference in *MHCIIα* variation among 9 Arctic charr populations from around Iceland (Figure 4). It is quite surprising to discover large differences at the *MHCIIα* among morphs within one lake, while the populations around Iceland were very similar. The pattern for *Cath2* was different. A modest F_{ST} of 0.23 among morphs in Lake Thingvallavatn is notably (~8X) higher than the F_{ST} for microsatellites [63]. On a larger geographic scale, we observe very large F_{ST} 's at *Cath2* among populations (highest 0.85). However there is no association of *Cath2* polymorphism with morphotype, while there may be a connection between relatedness and *Cath2* variation. The extent of differentiation in this locus is however stronger than seen in any individual microsatellite marker. In the absence of population genetic data spanning the relevant genomic regions, we cannot test for positive selection on those (or neighboring) genes.

Coalescence simulations [63] based on microsatellites (on the same fish studied here) support a model of very limited gene flow among the PL- and SB morphs in Lake Thingvallavatn, for the last 10.000 years. Also, the observed variation in microsatellites among arctic charr populations in Iceland and Lake Thingvallavatn, suggests substantial standing genetic variation in the anadromous stock(s) that colonized Icelandic waters. The reduced gene flow, due to isolation of populations or morphs, and local selective pressures could thus lead to differentiation in loci with fitness consequences. Thus the observed patterns in *MHCIIα* and *Cath2* within Lake Thingvallavatn and between Icelandic populations may reflect chance, history, and/or interplay of isolation and selection.

Table 7. Generalized linear model analyses of the contribution of morph, sex, weight, age and *MHCIIα* genotype on parasite infections in Lake Thingvallavatn charr in 2010.

Parasite	N	Morph	Weight	Age	Sex	*MHCIIα*
Diphyllobothrium sp.	263	PL vs. SB***; LB vs. PL***	PL**	ns.	ns.	ns.
Diplostomum sp.	263	PL vs. SB***; LB vs. SB***	SB***; LB***	SB*; LB***	ns.	ns.
Nematodes	202	PL vs. SB***; LB vs. PL**	ns.	ns.	PL*	ns.
Eubothrium salvelini	202	PL vs. SB***; LB. vs. PL***	ns.	PL*	ns.	ns.

Significance: ns. $p>0.05$,
*$p<0.05$,
**$p<0.01$,
***$p<0.001$.
doi:10.1371/journal.pone.0069402.t007

Reduced Variation in the *MHCIIα* in Iceland?

One feature in the data demands special attention. MHC loci often exhibit extreme polymorphism and signs of balancing selection in fish systems [19]. In Iceland *MHCIIα* variation is very much reduced in all populations, except for the PL morph in Lake Thingvallavatn (which has two common haplotypes). Conejeros and colleagues [26] studied *MHCIIα* variation in 6 populations of Arctic charr across Europe, Asia and North America, and found much higher diversity (7 or more haplotypes in 5 populations; at most 14 individuals sequenced in each). Only the population from Trinité (2 haplotypes at 50% frequency in 9 individuals) had comparable level of variation to that observed in Lake Thingvallavatn PL charr. Part of the explanation may be that, we are studying a slightly shorter fragment of the *MHCIIα* locus than Conejeros and associates [26]. Many studies have documented excessive variation in MHC genes within and between fish populations, but there are also examples of local differences, in part attributable to natural selection [19].

The low diversity in *MHCIIα* among Icelandic Arctic charr populations may reflect history, for instance low diversity within the colonizing stock or a bottleneck in recent history. Alternatively strong selection for certain *MHCIIα* alleles in specific populations may also have played a role. A putative case in point is is the observation that the PL-charr is clearly distinct from the two benthic morphs in Lake Thingvallavatn. MHC driven mate choice has been extensively studied, with documented examples of both assortative and disassortative mating [86–88]. Eizaguirre and Lenz [19] conclude that under parasite mediated selection, MHC mediated assortative mate-choice could promote local adaptation and divergence. Our data cannot be used to evaluate such scenarios, but it would be interesting to test whether *MHCII* variation correlates with mating preferences of Arctic charr.

F_{ST} mapping and Putative Functional Alleles

F_{ST} mapping can reveal both loci under positive selection and genes with relaxed purifying selection in certain populations, that stand out of the distribution of neutral variation. In this study a small fraction of the genome was interrogated and candidates were selected based on prior data and focus on particular pathways. This approach, although unlikely to find genes with the strongest signal of differentiation between groups, provided curious patterns for the sequenced candidates. In future genome wide single base polymorphism [89], microsatellite [90,91], Rad-tag screens [92,93] or even next generation sequencing of transcriptomes [94,95] from distinct populations/species are interesting strategies to study this system in more detail.

The *MHCII* genomic regions have been cloned and sequenced in *S. salar* [96], but not in *S. alpinus*. In light of the results, it would be most interesting to clone and sequence the MHCII regions from Arctic charr, possibly from distinct morphs, populations or continents. Also, in salmon the regions contain several immunological genes, so differentiation at *MHCIIα* could be caused by linked variants in other genes [31]. As we studied only a part of intron 2 in *MHCIIα* it is rather unlikely that functional polymorphism(s) were surveyed in the data. The situation is different with *Cath2* where the strongest signal was a segregating polymorphism that leads to an amino acid replacement, serine to an arginine (S115R), in the predicted antimicrobial peptide region. Cathelicidins are like most AMPs cationic and target specifically the negative charged bacterial membrane, which ultimately leads to the killing of the bacteria [34]. It has been suggested that Cod cathelicidins (codCath) kill bacteria through lysis [44], but so far little is known about the functional mechanisms of other fish cathelicidins, which are less charged than codCath. Therefore it is difficult to speculate on the effect an amino acid change in the mature cathelicidin antimicrobial peptide in Arctic charr. Phylogenetic comparisons show that positive selection operates on charged amino acids in AMPs [45]. Thus it is tempting to speculate that the Cath2 S115R replacement is functional. One way to test whether *Cath2* is under positive selection is to assess F_{ST} s along the locus and neighboring regions, to identify the marker with strongest signal of genetic differentiation between morphs and test formally for positive selection [97].

Tests of Association of Genes and Ecological Attributes

Several studies in *S. salar* and related species reveal strong differentiation in immunological genes among populations or morphotypes [27,98,99], which may be in part due to differences in parasite diversity in distinct habitats. Eizaguirre et al. [100] demonstrated with an experimental set up that parasitic nematode infections change *MHCIIβ* allele frequencies in a single generation. Here we tested for association of four classes of large and prevalent parasites (*Diplostomum sp.*, *Diphyllobothrium sp.*, *Eubothrium salvelini* and Nematodes) and the *MHCIIα* haplotypes, but found no significant associations. This does not formally exclude the possibility that those parasites were not involved in shaping *MHCIIα* diversity, for methodological and other reasons. On the methodology side, the sample size is relatively small, compared to association tests in human genetics [101,102] and the phenotypes

are not measured in controlled environment as in quantitative genetics [103,104]. Also, we only tested for association in a sample of 4–10 year old fish from 2010, but an association may have been between the genotype and parasites in the past (over many generations or during episodes of high infection) or only in juveniles. Reverse quantitative genetics can identify ecological variables of importance and shed light on the interplay of history, population genetic and ecological factors. However, failure of such phenotype hunts do not devalue the genetic signatures of differentiation among groups. QTL mapping within Arctic charr populations have identified chromosome regions that relate to ecologically important traits, *e.g.* spawning time and development [58,105,106].By combining population genetic and QTL mapping techniques, loci related to adaptation can be identified [107].

Freshwater Fishes to Study Adaptation

Following the last glaciation Nordic freshwater fishes expanded into new territories. Several features, like novel habitats, geographic isolation of stocks, in some cases small population sizes or bottlenecks, reduced gene flow and the relatively simpler ecosystem of arctic areas, could lead to rapid evolution via both drift and selection. Some Arctic charr populations show dedicated resource morphotypes while others retain ancestral phenotypes [8,108]. Similar to the stickleback and Mexican cavefish [9,109] the dozens of morphologically and ecologically distinct Arctic charr populations are *de facto* natural experiments in parallel evolution [53,63]. Genome-wide markers make it possible to elucidate the history of the distinct and even sympatric populations [93,107,110] and identify genes relating to adaptation [14,15,95,111]. Northern species like Arctic charr, which have invaded similar habitats multiple times and adapted to them in relatively short evolutionary time, provide an interesting system to dissect the genetics and ecology of parallel evolution, however complicated and challenging.

Supporting Information

Table S1 Specifics of primers and annealing temperatures.
(DOC)

Table S2 MHCIIα genotyping and polymorphism.
(DOC)

Acknowledgments

We thank Moira Ferguson and Xia Yue for sending charr DNA samples, Bjarni K. Kristjansson for help with collecting the 2010 sample and members of the Einar Arnason laboratory for help with the ABI sequencer. We thank Marta Gomez Munoz for preliminary work on *Cath2*, *Hamp* and *Leap-2a*, Jetty Ramadevi for DNA isolation and members of the University of Iceland Arctic charr group for discussion and advice.

Author Contributions

Conceived and designed the experiments: AP KHK VHM SSS. Performed the experiments: KHK JG SR CBS VCB VHM SSS AP. Analyzed the data: KHK JG VHM AP. Contributed reagents/materials/analysis tools: KHK SR VHM AP. Wrote the paper: KHK JG VHM SSS AP. Designed experiments: AP KHK VHM SSS. Molecular work: KHK SR. Parasite analyses: JG CBS KHK SR. Aged the specimens: VCB SSS.

References

1. Lewontin RC (1974) The Genetic Basis of Evolutionary Change. Columbia University Press. Available: http://www.abebooks.com/Genetic-Basis-Evolutionary-Change-LEWONTIN-Columbia/657524818/bd.
2. Rockman M V (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution; international journal of organic evolution 66: 1–17. Available: http://www.ncbi.nlm.nih.gov/pubmed/22220860. Accessed 25 October 2012.
3. Phillips PC (2005) Testing hypotheses regarding the genetics of adaptation. Genetica 123: 15–24. Available: http://www.ncbi.nlm.nih.gov/pubmed/15881677.
4. Raff R (1996) The Shape of Life: Genes, Development, and the Evolution of Animal Form. University Of Chicago Press. Available: http://www.citeulike.org/group/4585/article/2931409. Accessed 14 December 2012.
5. Stern DL (2000) Evolutionary biology. The problem of variation. Nature 408: 529, 531. Available: http://dx.doi.org/10.1038/35046183. Accessed 14 December 2012.
6. Lawton-Rauh A, Robichaux RH, Purugganan MD (2003) Patterns of nucleotide variation in homoeologous regulatory genes in the allotetraploid Hawaiian silversword alliance (Asteraceae). Molecular Ecology 12: 1301–1313. Available: http://doi.wiley.com/10.1046/j.1365-294X.2003.01814.x. Accessed 14 December 2012.
7. Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends in ecology & evolution 26: 298–306. Available: http://www.ncbi.nlm.nih.gov/pubmed/21459472. Accessed 4 November 2012.
8. Snorrason SS, Skúlason S (2004) Adaptive speciation in northern fresh water fishes – patterns and processes. In: Dieckmann U, Metz H, Doebeli M, Tautz D, editors. Adaptive speciation. Cambridge University Press, Cambridge, 210–228.
9. Rundle HD (2000) Natural Selection and Parallel Speciation in Sympatric Sticklebacks. Science 287: 306–308. Available: http://www.sciencemag.org/cgi/doi/10.1126/science.287.5451.306. Accessed 29 November 2012.
10. Reimchen TE, Nosil P (2006) Replicated ecological landscapes and the evolution of morphological diversity among Gasterosteus populations from an archipelago on the west coast of Canada. Canadian Journal of Zoology 84: 643–654. Available: http://www.nrcresearchpress.com/doi/abs/10.1139/z06-036#.UMp1HIqu_lc. Accessed 14 December 2012.
11. Landry L, Vincent WF, Bernatchez L (2007) Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. Journal of evolutionary biology 20: 971–984. Available: http://www.ncbi.nlm.nih.gov/pubmed/17465908. Accessed 14 December 2012.
12. Noakes DLG (2008) Charr truth?: sympatric differentiation in Salvelinus species. Environmental Biology of Fishes: 7–15. doi:10.1007/s10641-008-9379-x.
13. Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity 106: 404–420. Available: http://dx.doi.org/10.1038/hdy.2010.167. Accessed 1 November 2012.
14. Peichel CL, Nereng KS, Ohgi K a, Cole BL, Colosimo PF, et al. (2001) The genetic architecture of divergence between threespine stickleback species. Nature 414: 901–905. Available: http://www.ncbi.nlm.nih.gov/pubmed/11780061.
15. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, et al. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428: 717–723. Available: http://www.ncbi.nlm.nih.gov/pubmed/15085123.
16. Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, et al. (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Current biology?: CB 22: 83–90. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3319444&tool=pmcentrez&rendertype=abstract. Accessed 4 November 2012.
17. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55–61. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3322419&tool=pmcentrez&rendertype=abstract. Accessed 25 October 2012.
18. Gavrilets S (2004) Fitness Landscapes and the Origin of Species. Princeton University Press. Available: http://press.princeton.edu/titles/7799.html.
19. Eizaguirre C, Lenz TL (2010) Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. Journal of fish biology 77: 2023–2047. Available: http://www.ncbi.nlm.nih.gov/pubmed/21133915. Accessed 13 August 2011.
20. Murphy K, Travers P, Walport M (2007) Janeway's Immunobiology. 7th ed. Garland Science.
21. Magnadóttir B (2006) Innate immunity of fish (overview). Fish & shellfish immunology 20: 137–151. Available: http://www.ncbi.nlm.nih.gov/pubmed/15950491. Accessed 29 October 2012.
22. Magnadottir B (2010) Immunological control of fish diseases. Marine biotechnology (New York, NY) 12: 361–379. Available: http://www.ncbi.nlm.nih.gov/pubmed/20352271. Accessed 29 October 2012.
23. Cuesta A, Angeles Esteban M, Meseguer J (2006) Cloning, distribution and up-regulation of the teleost fish MHC class II alpha suggests a role for granulocytes

as antigen-presenting cells. Molecular immunology 43: 1275–1285. Available: http://www.ncbi.nlm.nih.gov/pubmed/16168483. Accessed 29 November 2012.

24. Blais J, Rico C, Van Oosterhout C, Cable J, Turner GF, et al. (2007) MHC adaptive divergence between closely related and sympatric African cichlids. PloS one 2: e734. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1939875&tool=pmcentrez&rendertype=abstract. Accessed 17 December 2012.
25. Matthews B, Harmon LJ, M'Gonigle L, Marchinko KB, Schaschl H (2010) Sympatric and allopatric divergence of MHC genes in threespine stickleback. PloS one 5: e10948. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2886830&tool=pmcentrez&rendertype=abstract. Accessed 16 November 2012.
26. Conejeros P, Phan A, Power M, Alekseyev S, O'Connell M, et al. (2008) MH class IIalpha polymorphism in local and global adaptation of Arctic charr (Salvelinus alpinus L.). Immunogenetics 60: 325–337. Available: http://www.ncbi.nlm.nih.gov/pubmed/18488215. Accessed 8 December 2010.
27. Tonteri A, Vasemägi A, Lumme J, Primmer CR (2010) Beyond MHC: signals of elevated selection pressure on Atlantic salmon (Salmo salar) immune-relevant loci. Molecular ecology 19: 1273–1282. Available: http://www.ncbi.nlm.nih.gov/pubmed/20196809. Accessed 2 November 2012.
28. Keller I, Taverna a, Seehausen O (2011) Evidence of neutral and adaptive genetic divergence between European trout populations sampled along altitudinal gradients. Molecular ecology 20: 1888–1904. Available: http://www.ncbi.nlm.nih.gov/pubmed/21418113. Accessed 9 November 2012.
29. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, et al. (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477: 207–210. Available: http://www.ncbi.nlm.nih.gov/pubmed/21832995. Accessed 1 November 2012.
30. Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: Turner BJ, editor. Evolutionary genetics of fishes. 1–53.
31. Harstad H, Lukacs MF, Bakke HG, Grimholt U (2008) Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (Salmo salar). BMC genomics 9: 193. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2386828&tool=pmcentrez&rendertype=abstract. Accessed 2 November 2012.
32. Flajnik MF, Du Pasquier L (2004) Evolution of innate and adaptive immunity: can we draw a line? Trends in immunology 25: 640–644. Available: http://www.ncbi.nlm.nih.gov/pubmed/15530832. Accessed 6 November 2012.
33. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, et al. (2007) Dynamic evolution of the innate immune system in Drosophila. Nature genetics 39: 1461–1468. Available: http://www.ncbi.nlm.nih.gov/pubmed/17987029. Accessed 1 November 2012.
34. Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415: 389–395. Available: http://www.ncbi.nlm.nih.gov/pubmed/11807545. Accessed 26 November 2012.
35. Lai Y, Gallo RL (2009) AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. Trends in immunology 30: 131–141. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2765035&tool=pmcentrez&rendertype=abstract. Accessed 31 October 2012.
36. Shike H, Lauth X, Westerman ME, Ostland VE, Carlberg JM, et al. (2002) Bass hepcidin is a novel antimicrobial peptide induced by bacterial challenge. European Journal of Biochemistry 269: 2232–2237. Available: http://doi.wiley.com/10.1046/j.1432-1033.2002.02881.x. Accessed 12 November 2012.
37. Zhang Y-A, Zou J, Chang C-I, Secombes CJ (2004) Discovery and characterization of two types of liver-expressed antimicrobial peptide 2 (LEAP-2) genes in rainbow trout. Veterinary immunology and immunopathology 101: 259–269. Available: http://www.ncbi.nlm.nih.gov/pubmed/15350756. Accessed 29 November 2012.
38. Chang C-I, Pleguezuelos O, Zhang Y-A, Zou J, Secombes CJ (2005) Identification of a novel cathelicidin gene in the rainbow trout, Oncorhynchus mykiss. Infection and immunity 73: 5053–5064. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1201231&tool=pmcentrez&rendertype=abstract. Accessed 19 December 2012.
39. Chang C-I, Zhang Y-A, Zou J, Nie P, Secombes CJ (2006) Two cathelicidin genes are present in both rainbow trout (Oncorhynchus mykiss) and atlantic salmon (Salmo salar). Antimicrobial agents and chemotherapy 50: 185–195. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1346769&tool=pmcentrez&rendertype=abstract. Accessed 19 December 2012.
40. Maier VH, Dorn K V, Gudmundsdottir BK, Gudmundsson GH (2008) Characterisation of cathelicidin gene family members in divergent fish species. Molecular immunology 45: 3723–3730. Available: http://www.ncbi.nlm.nih.gov/pubmed/18614236. Accessed 24 October 2010.
41. Scocchi M, Pallavicini A, Salgaro R, Bociek K, Gennaro R (2009) The salmonid cathelicidins: a gene family with highly varied C-terminal antimicrobial domains. Comparative biochemistry and physiology Part B, Biochemistry & molecular biology 152: 376–381. Available: http://www.ncbi.nlm.nih.gov/pubmed/19168146. Accessed 24 October 2010.
42. Feng CY, Johnson SC, Hori TS, Rise M, Hall JR, et al. (2009) Identification and analysis of differentially expressed genes in immune tissues of Atlantic cod stimulated with formalin-killed, atypical Aeromonas salmonicida. Physiological genomics 37: 149–163. Available: http://www.ncbi.nlm.nih.gov/pubmed/19240301. Accessed 29 October 2012.
43. Broekman DC, Zenz A, Gudmundsdottir BK, Lohner K, Maier VH, et al. (2011) Functional characterization of codCath, the mature cathelicidin antimicrobial peptide from Atlantic cod (Gadus morhua). Peptides 32: 2044–2051. Available: http://www.ncbi.nlm.nih.gov/pubmed/21945422. Accessed 15 November 2012.
44. Broekman DC, Frei DM, Gylfason G a, Steinarsson A, Jörnvall H, et al. (2011) Cod cathelicidin: isolation of the mature peptide, cleavage site characterisation and developmental expression. Developmental and comparative immunology 35: 296–303. Available: http://www.ncbi.nlm.nih.gov/pubmed/20950641. Accessed 29 November 2012.
45. Tennessen J a (2005) Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. Journal of evolutionary biology 18: 1387–1394. Available: http://www.ncbi.nlm.nih.gov/pubmed/16313451. Accessed 6 November 2012.
46. Johnson L (1980) The Arctic charr, Salvelinus alpinus. In: Balon EK, editor. Charrs, Salmonid Fishes of the Genus Salvelinus. The Hague: Junk. 15–98. Available: http://www.cabdirect.org/abstracts/19801402600.html;jsessionid=0C6C3ABB8538DE9A918A3009DE2B62EE?gitCommit=4.13.8-6-g6e31ff9. Accessed 17 December 2012.
47. Behnke RJ (1972) Systematics of salmonid fishes of recently glaciated lakes | Aquaculture Association of Canada. Journal of the Fisheries Research Board of Canada 29: 639–671. Available: http://www.aquacultureassociation.ca/salmon/253.
48. Behnke RJ (1980) A systematic review of the genus Salvelinus. In: Balon E, editor. Charrs: Salmonid Fishes of the Genus Salvelinus. The Hague: Junk. 441–481. Available: http://www.powells.com/biblio?isbn=9789061937012.
49. Skúlason S, Snorrason SS, Jonsson B (1999) Sympatric morphs, populations and speciation. In: Magurran A, May R, editors. Evolution of Biological Diversity. New York: Oxford University Press. 70–92.
50. Wilson AJ, Gíslason D, Skúlason S, Snorrason SS, Adams CE, et al. (2004) Population genetic structure of Arctic charr, Salvelinus alpinus from northwest Europe on large and small spatial scales. Molecular ecology 13: 1129–1142. Available: http://www.ncbi.nlm.nih.gov/pubmed/15078451.
51. Skulason S, Smith TB (1995) Resource polymorphisms in vertebrates. Trends in ecology & evolution 10: 366–370. Available: http://www.ncbi.nlm.nih.gov/pubmed/21237070. Accessed 16 December 2012.
52. Brunner PC, Douglas MR, Osinov a, Wilson CC, Bernatchez L (2001) Holarctic phylogeography of Arctic charr (Salvelinus alpinus L.) inferred from mitochondrial DNA sequences. Evolution; international journal of organic evolution 55: 573–586. Available: http://www.ncbi.nlm.nih.gov/pubmed/11327164. Accessed 26 November 2012.
53. Kristjansson BK, MALMQUIST HJ, INGIMARSSON F, ANTONSSON T, SNORRASON SS, et al. (2011) Relationships between lake ecology and morphological characters in Icelandic Arctic charr, Salvelinus alpinus. Biological Journal of the Linnean Society 103: 761–771. Available: http://doi.wiley.com/10.1111/j.1095-8312.2011.01670.x. Accessed 14 December 2012.
54. Sigursteinsdóttir RJ, Kristjánsson BK (2005) Parallel Evolution, not Always so Parallel: Comparison of Small Benthic Charr, Salvelinus alpinus, from Grimsnes and Thingvallavatn, Iceland. Environmental Biology of Fishes 74: 239–244. Available: http://www.springerlink.com/index/10.1007/s10641-005-0499-2. Accessed 16 December 2012.
55. Macqueen DJ, Kristjánsson BK, Johnston I a (2010) Salmonid genomes have a remarkably expanded akirin family, coexpressed with genes from conserved pathways governing skeletal muscle growth and catabolism. Physiological genomics 42: 134–148. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2888561&tool=pmcentrez&rendertype=abstract. Accessed 29 November 2012.
56. Skúlason S, Noakes DLG, Snorrason SS (1989) Ontogeny of trophic morphology in four sympatric morphs of arctic charr Salvelinus alpinus in Thingvallavatn, Iceland*. Biological Journal of the Linnean Society 38: 281–301. Available: http://doi.wiley.com/10.1111/j.1095-8312.1989.tb01579.x. Accessed 16 December 2012.
57. Eiriksson GM, Skulason S, Snorrason SS (1999) Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn, Iceland. Journal of Fish Biology 55: 175–185. Available: http://doi.wiley.com/10.1111/j.1095-8649.1999.tb01054.x. Accessed 16 December 2012.
58. Parsons KJ, Skúlason S, Ferguson M (n.d.) Morphological variation over ontogeny and environments in resource polymorphic arctic charr (Salvelinus alpinus). Evolution & development 12: 246–257. Available: http://www.ncbi.nlm.nih.gov/pubmed/20565535. Accessed 27 May 2013.
59. Snorrason SS, Skúlason S, Sandlund O, Malmquist H, Jonsson B, et al. (1989) Shape polymorphism in sympatric Arctic charr, Salvelinus alpinus, in Thingvallavatn, Iceland. In: Kawanabe H, Yamazaki F, Noakes D, editors. Biology of Charrs and Masu Salmon. Physiology and Ecology of Japan, Special Vol. 1. Kyoto: Kyoto University Press. 393–404.
60. Frandsen F, Malmquist HJ, Snorrason SS (1989) Ecological parasitology of polymorphic Arctic charr, Salvelinus alpinus (L.), in Thingvallavatn, Iceland. Journal of Fish Biology 34: 281–297. Available: http://doi.wiley.com/10.1111/j.1095-8649.1989.tb03309.x. Accessed 16 December 2012.
61. Magnusson KP, Ferguson MM (1987) Genetic analysis of four sympatric morphs of Arctic charr, Salvelinus alpinus, from Thingvallavatn, Iceland.

Environmental Biology of Fishes 20: 67–73. Available: http://www.springerlink.com/index/10.1007/BF00002026. Accessed 16 December 2012.
62. Volpe JP, Ferguson MM (1996) Molecular genetic examination of the polymorphic Arctic charr Salvelinus alpinus of Thingvallavatn, Iceland. Molecular ecology 5: 763–772. Available: http://www.ncbi.nlm.nih.gov/pubmed/8981767.
63. Kapralova KH, Morrissey MB, Kristjánsson BK, Olafsdóttir GÁ, Snorrason SS, et al. (2011) Evolution of adaptive diversity and genetic connectivity in Arctic charr (Salvelinus alpinus) in Iceland. Heredity 106: 472–487. Available: http://www.ncbi.nlm.nih.gov/pubmed/21224880.
64. Gislason D, Ferguson MM, Skúlason S, Snorrason SS (1999) RAPID COMMUNICATIONS/COMMUNICATIONS RAPIDES Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic char (Salvelinus alpinus). Arctic 2234: 2229–2234.
65. Sibthorpe D, Sturlaugsdóttir R, Kristjansson BK, Thorarensen H, Skúlason S, et al. (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 & molecular biology 145: 371–383. Available: http://www.ncbi.nlm.nih.gov/pubmed/17049897.
66. Macqueen DJ, Kristjánsson BK, Paxton CGM, Vieira VLA, Johnston IA (2011) The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. Molecular ecology 20: 3167–3184. Available: http://www.ncbi.nlm.nih.gov/pubmed/21714822. Accessed 27 May 2013.
67. Weichhart T, Costantino G, Poglitsch M, Rosner M, Zeyda M, et al. (2008) The TSC-mTOR signaling pathway regulates the innate inflammatory response. Immunity 29: 565–577. Available: http://www.ncbi.nlm.nih.gov/pubmed/18848473. Accessed 16 November 2012.
68. Smith TJ (2010) Insulin-like growth factor-I regulation of immune function: a potential therapeutic target in autoimmune diseases? Pharmacological reviews 62: 199–236. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2879913&tool=pmcentrez&rendertype=abstract. Accessed 12 November 2012.
69. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Totowa, NJ: Humana Press. 365–386.
70. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome research 8: 175–185. Available: http://www.ncbi.nlm.nih.gov/pubmed/9521921. Accessed 2 November 2012.
71. Gordon D (2003) Viewing and editing assembled sequences using Consed. Current protocols in bioinformatics/editoral board, Andreas D Baxevanis. [et al] Chapter 11: Unit11.2. Available: http://www.ncbi.nlm.nih.gov/pubmed/18428695. Accessed 26 November 2012.
72. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) 23: 2947–2948. Available: http://www.ncbi.nlm.nih.gov/pubmed/17846036. Accessed 1 November 2012.
73. Nicholas KB (Bank of A, Nicholas Jr. HB (Pittsburg. SC, Deerfield II. DW (Pittsburgh SC (1997) GeneDoc: Analysis and Visualization of Genetic Variation. Available: http://www.nrbsc.org/gfx/genedoc/ebinet.htm.
74. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics (Oxford, England) 23: 2633–2635. Available: http://www.ncbi.nlm.nih.gov/pubmed/17586829. Accessed 5 November 2012.
75. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497. Available: http://bioinformatics.oxfordjournals.org/content/19/18/2496.short. Accessed 30 October 2012.
76. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105: 767–779. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1202185&tool=pmcentrez&rendertype=abstract. Accessed 26 November 2012.
77. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457–462. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1206343&tool=pmcentrez&rendertype=abstract. Accessed 26 November 2012.
78. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary bioinformatics online 1: 47–50. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2658868&tool=pmcentrez&rendertype=abstract. Accessed 6 December 2012.
79. Felsenstein J (1989) PHYLIP–phylogeny inference package (version 3.2). Cladistics 5: 164–166.
80. Kristmundsson Á, Richter S (2009) Parasites of resident arctic charr, Salvelinus alpinus, and brown trout, Salmo trutta, in two lakes in Iceland. ICEL AGRIC SCI 22: 5–18.
81. Fox J, Weisberg S (2011) An R Companion to Applied Regression. Second Edi. Available: tinyurl.com/carbook.
82. Venables WN, Ripley BD (2002) Modern Applied Statistics with S, 4th ed. Springer. Available: http://www.stats.ox.ac.uk/pub/MASS4/.
83. Alekseyev SS, Bajno R, Gordeeva N V, Reist JD, Power M, et al. (2009) Phylogeography and sympatric differentiation of the Arctic charr Salvelinus alpinus (L.) complex in Siberia as revealed by mtDNA sequence analysis. Journal of fish biology 75: 368–392. Available: http://www.ncbi.nlm.nih.gov/pubmed/20738544. Accessed 17 October 2010.
84. Palsson A, Dodgson J, Dworkin I, Gibson G (2005) Tests for the replication of an association between Egfr and natural variation in Drosophila melanogaster wing morphology. BMC genetics 6: 44. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1208880&tool=pmcentrez&rendertype=abstract. Accessed 26 November 2012.
85. Gislason D, Ferguson MM, Skulason S, Snorrason SS (1999) Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic char … Library.
86. Thünken T, Meuthen D, Bakker TCM, Baldauf S a (2012) A sex-specific trade-off between mating preferences for genetic compatibility and body size in a cichlid fish with mutual mate choice. Proceedings Biological sciences/The Royal Society 279: 2959–2964. Available: http://www.ncbi.nlm.nih.gov/pubmed/22513859. Accessed 2 November 2012.
87. Lenz TL, Eizaguirre C, Scharsack JP, Kalbe M, Milinski M (2009) Disentangling the role of MHC-dependent "good genes" and "compatible genes" in mate-choice decisions of three-spined sticklebacks Gasterosteus aculeatus under semi-natural conditions. Journal of fish biology 75: 2122–2142. Available: http://www.ncbi.nlm.nih.gov/pubmed/20738677. Accessed 2 November 2012.
88. Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. Molecular ecology 18: 3316–3329. Available: http://www.ncbi.nlm.nih.gov/pubmed/19523111. Accessed 2 November 2012.
89. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome research 12: 1805–1814. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=187574&tool=pmcentrez&rendertype=abstract. Accessed 12 November 2012.
90. Lehmann T, Licht M, Elissa N, Maega BTA, Chimumbwa JM, et al. (n.d.) Population Structure of Anopheles gambiae in Africa. The Journal of heredity 94: 133–147. Available: http://www.ncbi.nlm.nih.gov/pubmed/12721225. Accessed 26 November 2012.
91. Storz JF, Payseur BA, Nachman MW (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Molecular biology and evolution 21: 1800–1811. Available: http://www.ncbi.nlm.nih.gov/pubmed/15201398. Accessed 12 November 2012.
92. Hohenlohe P a, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular ecology resources 11 Suppl 1: 117–122. Available: http://www.ncbi.nlm.nih.gov/pubmed/21429168. Accessed 29 October 2012.
93. Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, et al. (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America 107: 16196–16200. Available: http://www.pnas.org/content/107/37/16196.long. Accessed 2 November 2012.
94. Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. Molecular ecology resources 8: 3–17. Available: http://www.ncbi.nlm.nih.gov/pubmed/21585713. Accessed 31 October 2012.
95. Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, et al. (2010) Adaptation genomics: the next generation. Trends in ecology & evolution 25: 705–712. Available: http://www.ncbi.nlm.nih.gov/pubmed/20952088. Accessed 26 October 2012.
96. Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, et al. (2003) MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210–219. Available: http://www.ncbi.nlm.nih.gov/pubmed/12811427. Accessed 18 November 2012.
97. Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ (1994) Evidence for positive selection in the superoxide dismutase (Sod) region of Drosophila melanogaster. Genetics 136: 1329–1340. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1205914&tool=pmcentrez&rendertype=abstract. Accessed 21 May 2013.
98. Evans ML, Neff BD, Heath DD (2010) MHC genetic structure and divergence across populations of Chinook salmon (Oncorhynchus tshawytscha). Heredity 104: 449–459. Available: http://www.ncbi.nlm.nih.gov/pubmed/19773808. Accessed 30 June 2010.
99. Consuegra S, De Eyto E, McGinnity P, Stet RJM, Jordan WC (2011) Contrasting responses to selection in class I and class IIα major histocompatibility-linked markers in salmon. Heredity 107: 143–154. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3178404&tool=pmcentrez&rendertype=abstract. Accessed 10 November 2012.
100. Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nature communications 3: 621. Available: http://www.nature.com/ncomms/journal/v3/n1/full/ncomms1632.html#/f1. Accessed 28 October 2012.
101. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, et al. (2006) A common variant associated with prostate cancer in European and

African populations. Nature genetics 38: 652–658. Available: http://www.ncbi.nlm.nih.gov/pubmed/16682969. Accessed 4 November 2012.
102. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, et al. (2007) Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 448: 353–357. Available: http://www.ncbi.nlm.nih.gov/pubmed/17603472. Accessed 2 November 2012.
103. Palsson A, Gibson G (2004) Association between nucleotide variation in Egfr and wing shape in Drosophila melanogaster. Genetics 167: 1187–1198. Available: http://www.genetics.org/content/167/3/1187.short. Accessed 6 November 2012.
104. Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, et al. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant journal?: for cell and molecular biology 44: 1054–1064. Available: http://www.ncbi.nlm.nih.gov/pubmed/16359397. Accessed 29 October 2012.
105. Leder EH, Danzmann RG, Ferguson MM (2006) The candidate gene, Clock, localizes to a strong spawning time quantitative trait locus region in rainbow trout. The Journal of heredity 97: 74–80. Available: http://jhered.oxfordjournals.org/content/97/1/74.short. Accessed 7 November 2012.
106. Parsons KJ, Sheets HD, Skúlason S, Ferguson MM (2011) Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (Salvelinus alpinus). Journal of evolutionary biology 24: 1640–1652. Available: http://www.ncbi.nlm.nih.gov/pubmed/21599773. Accessed 27 May 2013.
107. Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, et al. (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. Philosophical transactions of the Royal Society of London Series B, Biological sciences 367: 354–363. Available: http://www.ncbi.nlm.nih.gov/pubmed/22201165. Accessed 4 November 2012.
108. Noakes DLG (2008) Charr truth: sympatric differentiation in Salvelinus species. Environmental Biology of Fishes 83: 7–15. Available: http://apps.webofknowledge.com/full_record.do?product = WOS&search_mode = GeneralSearch&qid = 1&SID = P144M4CgPi63FGeh7lj&page = 1&doc = 1. Accessed 21 November 2012.
109. Jeffery WR (2001) Cavefish as a model system in evolutionary developmental biology. Developmental biology 231: 1–12. Available: http://www.ncbi.nlm.nih.gov/pubmed/11180948. Accessed 12 November 2012.
110. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genetics 6: e1000862. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2829049&tool = pmcentrez&rendertype = abstract. Accessed 1 November 2012.
111. Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive allele in deer mice. Science (New York, NY) 325: 1095–1098. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2736094&tool = pmcentrez&rendertype = abstract. Accessed 3 November 2012.

Supplemental table S1 for **Differentiation at the *MHCIIα* and *Cath2* loci in sympatric *Salvelinus alpinus* resource morphs in Lake Thingvallavatn** Kalina H. Kapralova, Johannes Gudbrandsson, Sigrun Reynisdottir, Cristina B. Santos, Vanessa C. Baltanás, Valerie H. Maier, Sigurdur S. Snorrason and Arnar Palsson.

Supplemental Table S1. Specifics of primers and annealing temperatures.

Locus	Primer name	Primer sequence 5'-3'	Ta (°C)
MHCIIα	*SAALDAAF**	CTGGATGCAGTGATTCAGATG	53
	*SAALDAAR**	GACGTGGCAGATGAGAGTG	
	SAMHC2a_f8	CAAGAACCCACCAGAGACAA	53
	SAMHC2a_r5	TGGGAACACATTTAGCATCA	
D-loop	SADloop_F	CCACCCTTAACTCCCAAAGC	57
	SADloop_R	GGCTTGGTGGGTAACGAAC	
Hepcidin	SAHep_F	TACGCTGGCCCTTTTCTACA	53
	SAHep_R	CTTTCTCCCTGGGTGCATTA	
Leap-2a	SALeap2a_F	GATATTGAATGCTAGCTTTTGGAC	53
	SALeap2a_R	AAAGGCCATTGCAAAGACAG	
Cath2 (3' UTR)	SACat_F5	AGCAAGGCCAACCATGTC	57
	SACat_R5	TGCAGTAAACATGAACTGGAAA	
Cath2 (peptide)	SACat_f9	GGAGACGCTCTGCAGTAAGG	57
	SACat_r9	GGTTTACTCCGCTAGCTCCA	
Cath2 (intron 2)	SACat_f7	AAATCAGCTGCTTCCTGTGG	57
	SACat_r8	GAGGACATGGTTGGCCTTG	

* From Conejeros et al 2008.

Supplemental table S2 for **Differentiation at the *MHCIIα* and *Cath2* loci in sympatric *Salvelinus alpinus* resource morphs in Lake Thingvallavatn** Kalina H. Kapralova, Johannes Gudbrandsson, Sigrun Reynisdottir, Cristina B. Santos, Vanessa C. Baltanás, Valerie H. Maier, Sigurdur S. Snorrason and Arnar Palsson.

Supplemental Table S2. *MHCIIα* genotyping and polymorphism.

Genotyping	Reference	Haplotype	S186	S236	S269	S271	I280	S294	S346	S349	S362	S375	S458	S472
Cloned	Saal-DAA*0801 (99%)		A	T	C	A	1	G	A	A	G	A	A	C
Cloned and PCR	Saal-DAA*0305/0306/0307 (98%)	hap14	C	G	A	T	1	G	A	T	A	A	A	T
Cloned and PCR	Saal-DAA*0305 (99%)	hap15	C	G	A	T	0	A	G	T	G	G	T	T

Positions in reference to EF450451.1 (Saal-DAA*0305)
Reference indicates the best hit in Genebank (% identity from Blastn)
Cloned= Cloned and sequenced
PCR = direct PCR and sequencing

Paper III

The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations]

Jóhannes Gudbrandsson, Ehsan P. Ahi, Kalina. H. Kapralova, Sigridur R. Franzdottir, Bjarni K. Kristjansson, S. Sophie Steinhaeuser, Isak M. Johannesson, Valerie H. Maier, Sigurdur S. Snorrason, Zophonias O. Jonsson and Arnar Palsson

F1000Research **4:** 136

JG performed the and designed the RNA-seq data analysis. Both regarding gene expression and genetic variation. He also participated in analysis of other data. JG participated in the writing of the paper.

F1000Research F1000Research 2016, 4:136 Last updated: 25 DEC 2016

Check for updates

RESEARCH ARTICLE

REVISED The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations]

Johannes Gudbrandsson[1], Ehsan P. Ahi[1], Sigridur R. Franzdottir[1], Kalina H. Kapralova[1], Bjarni K. Kristjansson[2], S. Sophie Steinhaeuser[1], Valerie H. Maier[1], Isak M. Johannesson[1], Sigurdur S. Snorrason[1], Zophonias O. Jonsson[1], Arnar Palsson[1]

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

v3 **First published:** 01 Jun 2015, **4**:136 (doi: 10.12688/f1000research.6402.1)
Second version: 25 Apr 2016, **4**:136 (doi: 10.12688/f1000research.6402.2)
Latest published: 02 Dec 2016, **4**:136 (doi: 10.12688/f1000research.6402.3)

Abstract

Species and populations with parallel evolution of specific traits can help illuminate how predictable adaptations and divergence are at the molecular and developmental level. Following the last glacial period, dwarfism and specialized bottom feeding morphology evolved rapidly in several landlocked Arctic charr *Salvelinus alpinus* populations in Iceland.

To study the genetic divergence between small benthic morphs and limnetic morphs, we conducted RNA-sequencing charr embryos at four stages in early development. We studied two stocks with contrasting morphologies: the small benthic (SB) charr from Lake Thingvallavatn and Holar aquaculture (AC) charr. The data reveal significant differences in expression of several biological pathways during charr development. There was also an expression difference between SB- and AC-charr in genes involved in energy metabolism and blood coagulation genes. We confirmed differing expression of five genes in whole embryos with qPCR, including *lysozyme* and *natterin-like* which was previously identified as a fish-toxin of a lectin family that may be a putative immunopeptide. We also verified differential expression of 7 genes in the developing head that associated consistently with benthic v.s.limnetic morphology (studied in 4 morphs). Comparison of single nucleotide polymorphism (SNP) frequencies reveals extensive genetic differentiation between the SB and AC-charr (~1300 with more than 50% frequency difference). Curiously, three derived alleles in the otherwise conserved 12s and 16s mitochondrial ribosomal RNA genes are found in benthic charr.

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

Open Peer Review

Referee Status: ✔ ? ✔

	Invited Referees 1	2	3
REVISED version 3 published 02 Dec 2016			✔ report
REVISED version 2 published 25 Apr 2016	✔ report		? report
version 1 published 01 Jun 2015	? report	? report	

1 **Daniel Macqueen**, University of Aberdeen UK

2 **Anne Dalziel**, Laval University Canada

3 **Örjan Östman**, Swedish University of Agricultural Sciences Sweden

Discuss this article

Comments (0)

Corresponding authors: Johannes Gudbrandsson (jog7@hi.is), Arnar Palsson (apalsson@hi.is)

How to cite this article: Gudbrandsson J, Ahi EP, Franzdottir SR *et al.* **The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2016, **4**:136 (doi: 10.12688/f1000research.6402.3)

Grant information: This project was supported by The Icelandic Center for Research (grant number: 100204011) to SSS, AP, ZOJ and BKK, The University of Iceland Research/Doctoral Fund to JG and KHK and University of Iceland research fund to AP, SSS and ZOJ.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 01 Jun 2015, **4**:136 (doi: 10.12688/f1000research.6402.1)

F1000Research 2016, 4:136 Last updated: 25 DEC 2016

REVISED **Amendments from Version 2**

The changes to the manuscript are rewriting of the introduction, results and discussion to present more clearly the biology of the Aquaculture charr used here as a reference strain, and the results from the contrast of the SB and AC transcriptomes and their implication both for evolutionary questions and also biology of the AC-charr. We also set out to reduce the emphasis on the ecological divergence in the description and interpretation of the results. We also rewrote part of the introduction, to provide better flow from general to specific background, and shortened the summary of molecular genetics of the craniofacial genes in the discussion. We clarified several issues, like the potential transgenerational effects in common garden experiment, the description of the Salmonid ancestor genome duplication and the age of the clade, and the qPCR results on both the Nattl genes and the summary of validated genes. We amended Figure 1, Figure 2 and Figure 4, and cleaned or adjusted the language/spelling in accordance with the suggestions of the reviewers. We thank them dearly for their thoughtful comments and suggestions, which have clearly improved the manuscript.

See referee reports

Introduction

Historical contingencies and chance shape organisms during evolution[1,2], but convergence in phenotype and molecular systems indicates that evolution is to some extent predictable[3,4]. Identification of genes and variants that influence evolved differences is not a trivial task[5]. Ideal systems to study the role of chance and necessity in ecological evolution would be related species or populations with readily observable phenotypic variation, living in a tractable ecological setting, and showing parallel evolution of specific traits within/among species/populations. Examples of such species complexes are provided by finches of the Galapagos islands[6], while cichlids of the African great lakes also provide an exciting multi-species system in the same respect[7]. The threespine stickleback has also emerged as a model "single species" system[8]. The amount of diversity in the feeding specializations of fish provide great opportunities for studying adaptation and divergence at the developmental and genetic level.

Some northern freshwater fish species exhibit frequent parallelism in trophic structures and life history and in several cases are found as distinct resource morphs[8–13]. Local adaptation has been extensively studied in the salmonid family, to which our study species Arctic charr (*Salvelinus alpinus*) belongs[14]. This species is well suited for studying the developmental underpinnings of trophic divergence and parallel evolution. The common ancestor to salmonids experienced a whole genome duplication 88–103 million years ago, the fourth vertebrate whole-*genome* duplication (Ss4R)[15–18]. This has provided time for divergence of ohnologous genes (paralogous genes originated by whole genome duplication event) in salmonid lineages. Estimates from the rainbow trout (*Oncorhynchus mykiss*) genome suggest that ohnologous genes are lost at a rate of about 170 genes per million years, and that around half of the original Ss4R ohnologue pairs are still functionally retained in rainbow trout[16].

One approach to identify pathways related to function or morphological differences between species, populations or ecomorphs is to study gene expression during development[19,20]. For example a microarray study of liver samples from anadromous and resident populations of brown trout (*Salmo trutta*), revealed that gene expression in juveniles was more influenced by life history than relatedness[21]. Furthermore, Filteau *et al.* (2013)[22] found a set of coexpressed genes differentiating two whitefish morphotypes, implicating Bone morphogenesis protein (BMP) signalling in the development of ecological differences in trophic morphology. Thus we were quite keen to apply RNA-sequencing to analyse ecomorphs in Arctic charr. *De novo* assembly of genomes and transcriptomes is complicated if many paralogs are present, which is the case in Arctic charr – see 23, 24. In this study we opted for mapping the reads (36 bp) to a related reference genome/transcriptome[25], instead of *de novo* assembly. Two previous studies have used RNA-seq to study salinity tolerance in adult Arctic charr, and found links between gene expression and quantitative trait loci[23,24].

Molecular studies of the highly polymorphic Arctic charr

Following the end of the last glacial period, about 10.000 years ago, Arctic charr colonized northern freshwater systems[26]. It is found as anadromous or lake/stream residents and exhibits high level of within species polymorphism[11,26]. Charr is also known to harbour substantial phenotypic plasticity, which may promote or reduce divergence[9,27]. Resource polymorphism in charr correlates with ecological attributes[28–30]. For instance small charr with benthic morphology, are found in multiple lavaspring and stream habitats in Iceland[31], and a comparative study of Icelandic lakes[30] found that lakes with greater limnetic habitat, lower nutrients levels, and greater potential for zooplankton consumption appeared to promote resource polymorphism. Some of the larger lakes contain two or more distinct morphs, typically limnetic and benthic forms. Multiple lines of evidence show that these differences stem both from environmental and genetic causes[32–36]. The best studied example of sympatric charr are the four morphs in Lake Thingvallavatn[37]; two have a benthic morphotype, a large benthivorous (LB-charr) and a small benthivorous (SB-charr), and two morphs are limnetic, a large piscivorous morph (PI-charr) and small planktivorous morph (PL-charr)[38]. Both PL and PI-charr operate in open water and feed on free-swimming prey, PL on planktonic crustaceans and PI on small fish. The PL, LB and SB-charr are presented in Figure 1.

Several population genetics studies, using allozymes or mtDNA revealed no differences among charr morphs in Lake Thingvallavatn[39–41] while other studies using microsatellite markers and nuclear genes, found significant[42–44] genetic differences among morphs in the lake[45]. Importantly Kapralova *et al.* (2011)[44] concluded that small benthic morphs have evolved repeatedly in Iceland and that gene flow has been reduced between the PL and SB morphs in Lake Thingvallavatn since its formation approximately 10,000 years ago[46]. We also discovered genetic separation in immunological genes (*MHCIIα* and *cath2*) between morphs in Iceland and within the lake[45], consistent with ecologically driven evolution

Figure 1. The Arctic charr morphs used in this study. Adult individuals of the four morphs studied here, **A**) the Holar aquaculture charr, **B**) the small benthic charr, **C**) the planktivorous charr **D**) and the large benthic charr. The latter three all come from Lake Thingvallavatn and were sexually ripe. The morphs differ in size at maturation, body and head shape - mainly lower jaw and length of maxilla and colour pattern in the wild.

of immune functions. Recently qPCR analyses showed that expression of mTOR pathway components in skeletal muscle correlates with the SB-charr form in Iceland[47], but it is unknown whether there is genetic differentiation in those genes or upstream regulators. Because individual genes have distinct histories[48,49], genome wide methods are needed to identify genes and mutation that associate with divergence.

AC charr as a reference for sympatric Arctic charr

The ideal reference populations for developmental and molecular studies of landlocked and sympatric Arctic charr in Iceland would be local anadromous charr. However, capturing running charr from the wild is not trivial. For this study we chose to use the Icelandic aquaculture charr (AC) as a reference. The AC-charr was founded with fish from the north of Iceland, and has been bred at Holar University College since 1990[50]. Body weight and age at sexual maturity have significant heritability in the Holar AC-charr, and the stock responded well to artificial selection for growth and performance characteristics. It is now the dominant charr breed in aquaculture in Iceland. While clearly a derived breed, it seems to have retained general limnetic craniofacial morphotype (Figure 1). The rationale for comparing SB-charr from Lake Thingvallavatn and AC-charr was threefold: i) SB charr represents an extensively studied and derived form of charr, that has been separated from anadromous fish for approx. 10,000 years, ii) AC charr was readily available for sampling (but wild anadromous charr was not), iii) we wanted an extreme contrast, because of budget reasons we could only sequence 8 samples at the time. Note, the transcriptome itself can only point out differences between the two morphs, but not highlight whether specific genes associate with SB or AC biology and breeding. But by focusing the verification on sympatric benthic and limnetic morphs of Lake Thingvallavatn, we could test and verify a subset of the signals found here. The contrast of SB and AC was justified as the data and studies[51–53] building on this data illustrate (see discussion).

Our long term research objectives are to investigate the genetics and developmental underpinnings of charr divergence and benthic parallelism. As a step towards this we compared the developmental transcriptome of SB charr and AC charr, reared in common lab environment to minimize the effects of environmentally induced phenotypic plasticity. The aims of this study are threefold. First, to find genes and pathways related to the development of phenotypic differences between small benthic charr from Lake Thingvallavatn and Icelandic aquaculture charr conforming to a limnetic morphotype. Second, to screen for signals of genetic differentiation between these two charr types. Third, we set out to verify a subset of the expression and genetic signals in the high-throughput sequencing data and also studying two more morphs (LB and PL) from Lake Thingvallavatn. The data reveal differential expression of genes that may affect the development of craniofacial and other phenotypic traits in charr. Genetic differences in nuclear and mitochondrial genes are also observed and provide a starting point for studying evolution of wild populations and genetics of domestication in Icelandic Arctic charr.

Methods

Sampling, rearing and developmental series

Overview of the experimental design, RNA sequencing, analyses and follow work is outlined in Figure 2. We set up crosses and reared embryos in the laboratory as described in 51. Embryos from four charr morphs were studied: an aquaculture charr (AC-charr) from the Holar breeding program[50] and three natural morphs from Lake Thingvallavatn; SB, LB and PL-charr[54]. Samples of the first two, AC and SB-charr, with contrasting adult size and morphology (Figure 1), were collected in 2009 and material sent for RNA sequencing. The latter two were sampled in 2010 and were used for qPCR and SNP studies of selected genes. Briefly, in September 2009 we got material from spawning AC-charr from the Holar breeding program[50], from single parent crosses and spawning SB-charr collected via gill netting in Olafsdrattur in Lake Thingvallavatn. Similarly, in the 2010 spawning season SB-, LB- and PL-charr were collected from Lake Thingvallavatn. For each parent group, eggs from several females (3–10) were pooled and fertilized using milt from several males (3–5) from the same group. Embryos were reared at ~ 5°C under constant water flow and in complete darkness at the Holar University College experimental facilities in Verid, Saudárkrókur. The water temperature was recorded twice daily and the average was used to estimate the relative age of the embryos using tausomite units (ts)[55]. Embryos and juveniles were sampled at designated time points, placed in RNAlater (Ambion) and frozen at −20°C. Post hatching juveniles were reared at the same temperature on standard Aquaculture food. For the investigation of different tissues of adult aquaculture charr (AC) from Hólar (fish size 20–25 cm) were used. Six randomly selected individuals were killed (by cutting through

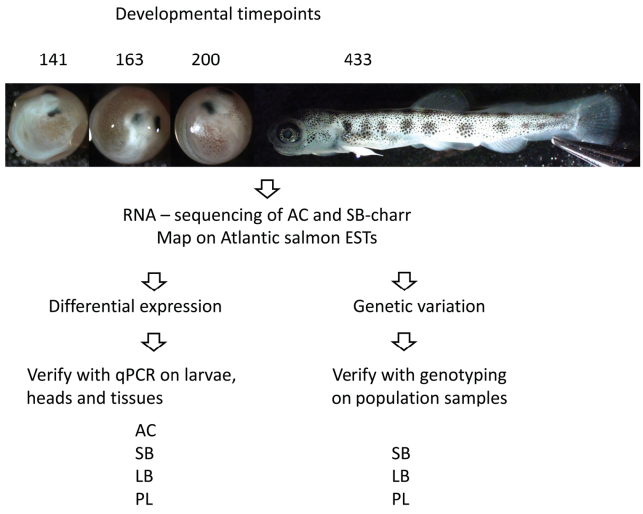


Figure 2. Schematic of RNA sequencing and follow up qPCR and population genetic work. RNA from embryos of the AC and SB charr at four stages (AC embryos pictured at top) were sequenced with Illumina technology. Reads were mapped to Atlantic salmon expressed sequence tags (ESTs). To verify differentially expressed genes we used RNA from embryos and heads of these four morphs, and tissues from adult AC charr. To verify SNPs we genotyped population samples from three Lake Thingvallavatn morphs (PL, LB and SB).

spinal cord) and dissected, and samples were taken from the skin, heart, liver, gills, spleen, intestine and kidney of each fish. The samples were placed in RNAlater (Ambion) and stored at −20°C. We used DNA for population genetic analyses from our previous study[45], eight individuals from each of the three types, PL, LB and SB-charr.

Fishing in Lake Thingvallavatn was done with permissions obtained both from the owner of the land in Mjóanes and from the Thingvellir National Park commission. Ethics committee approval is not needed for regular or scientific fishing in Iceland (The Icelandic law on Animal protection, Law 15/1994, last updated with Law 157/2012). Sampling was performed by Holar University College Aquaculture Research Station (HUC-ARC) personnel. HUC-ARC has an operational license according to Icelandic law on aquaculture (Law 71/2008), which includes clauses of best practices for animal care and experiments.

RNA extraction and transcriptome sequencing

Embryos of AC- and SB-charr sampled in 2009 were used for transcriptome sequencing. For this we focused on the time covering development of pharyngeal arches and morphogenesis of the head: at 141, 163, 200 and 433 *ts* (post fertilization). For each combination of morphs and timepoints we pooled RNA from approximately six individuals. RNA extraction and following steps were performed as described earlier[51,56]. Briefly, the embryos were dechorionated and homogenized with a disposable Pellet Pestle Cordless Motor tissue grinder (Kimble Kontes, Vineland, NJ, USA) and RNA was extracted into two size-fractions using the Ambion mirVana kit (Life Technologies, Carlsbad, CA, USA). The high molecular weight fraction was further used for mRNA-seq and RNA quality was analysed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA from samples was pooled - equal contribution of each sample - and first and second strand cDNA synthesis, fragmentation, adapter ligation and amplification were performed using the mRNA-Seq 8-Sample Prep Kit (Illumina, San Diego, CA, USA) according to manufacturer's instructions. Sequencing was performed at DeCode genetics (Reykjavík, Iceland) using SOLEXA GAII technology (Illumina, San Diego, CA, USA).

The sequencing reads were deposited into the NCBI SRA archive under BioProject identifier PRJNA239766 and with accession numbers: SRX761559, SRX761571, SRX761575, SRX761577, SRX761451, SRX761461, SRX761490 and SRX761501.

The embryos sampled in 2010 were used for qPCR analyses. RNA was extracted from six whole embryos, in two replicates (two repetitions *X* three fish) (AC and SB sampled at 161 and 200 *ts*). For the extraction of RNA from heads of AC, SB, LB and PL, 12 embryos (two repetitions *X* six fish) at 178, 200 and 216 *ts* were used. Embryos were dechorionated and decapitated in front of the pectoral fin. RNA extraction and cDNA preparation were performed as described previously in 51. Similarly, RNA was extracted from a small piece (approximately 2 mm^2) of skin, heart, liver, gill, spleen, intestine and liver from six adult AC-charr.

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

As no *S. alpinus* genome is available and *de novo* assembly of the 36 bp reads yielded an excessive number of short contigs we chose to assess expression and genetic variation by mapping the reads to 59336 *S. salar* expressed sequence tag (EST) contigs from the SalmonDB [57, downloaded 22. March 2012] and the Arctic charr mitochondrial genome [48, NC_000861].

To estimate expression, reads were aligned with RSEM version 1.1.18 with default parameters. RSEM distributes reads that map to multiple locations to the most likely contig, using expectation maximization[58]. The read counts for contigs with the same annotation were pooled because some genes were represented by more than one contig, and due to whole genome duplication almost the half of salmonid genes exist as ohnologs[16,18]. Thus the expression tests are done on gene or paralog group level, instead of the contig level. We acknowledge that paralogous genes are not always expressed similarly, but feel its necessary to do this pooling because of the nature of the data. In the remainder of the paper, we will refer to gene or paralog group (the number of underlying contigs is indicated in relevant tables). This brought the number of genes considered down to 16851. Lastly, paralog groups with fewer than 800 mapped reads in the entire dataset were excluded from the analyses, yielding a total of 10496.

A generalized linear model (GLM) with morph and developmental time as explanatory variables was used to find genes with different expression levels between the two charr morphotypes (groups) using the edgeR-package in R[59].

Y=Morph+Time+Error

To obtain further insight into the expression profiles of differently expressed genes, we performed clustering analyses on log-transformed cpm-values (counts per million; cpm-function in edgeR). The values for each gene were scaled by mean and standard deviation, and the euclidean distance used for the hclust-function in R[60] with the default settings. We used the hypergeometric-test in goseq[61] to test for gene ontology enrichment. Since we pooled the read-count from different contigs we could unfortunately not take gene length into account in those tests.

Tests of differential expression with qPCR

We previously identified suitable reference genes to study Arctic charr development[51]. Here we examined the expression of several genes in whole charr embryos, embryonic heads and adult tissues. Primers were designed using the Primer3 tool[62] and checked for self-annealing and heterodimers according to the MIQE guidelines[63] (S1 Table). Primers for genes with several paralogs were designed for regions conserved among paralogs, except for *natterin-like,* where primers were designed to match regions differing in sequence between paralogs. Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method[64]. For the calculation of relative expression of genes in whole embryos, the geometric mean expression of three reference genes, *β- Actin (Actb), elongation*

factor 1 α and *Ubiquitin-conjugating enzyme E2 L3*, was used for normalization. For visual comparisons among samples, the normalized expression was presented as relative to the expression in AC at 161 *ts* (calibration sample). For the embryonic head samples *Eukaryotic Translation Initiation Factor 5A (If5a1)* and *Actb* were used as reference genes and a biological replicate of AC at 178 (*ts*) as the calibrator sample, see 51, 52. Standard errors of relative expression were calculated from the standard errors (SE) of the ΔC_T-values with the formula $2^{-(\Delta\Delta Ct+SE)}$ = minimum fold expression and $2^{-(\Delta\Delta Ct-SE)}$ = maximum fold expression. The statistical analysis was performed using the ΔC_T-values with a two-way ANOVA with GLM function in R.

Y=Morph+Time+MorphxTime+Error

Normal distribution of residuals was confirmed for all data. For the study of expression in the embryonic head we followed a significant morph effect in the ANOVA with Tukey's post-hoc honest significant difference test, on relative expression ratios (ΔC_Ts). Three genes had lower efficiency (as low as 1.72). We acknowledge that the data on those genes may be weak.

Polymorphisms in the Arctic charr transcriptome

For analysis of genetic variation we mapped the reads to the salmon contigs, this time using the Burrow-Wheeler Aligner (BWA)[65] with a seed length of 25, allowing two mismatches. We re-mapped the reads, since BWA allows short indels (RSEM does not) but disregarding them leads to many false SNPs close to indels. To extract candidate polymorphic sites from the Arctic charr transcriptome we ran VarScan2[66] with minimum coverage of 50 reads and minimum minor allele frequency of 0.1 on reads mapped to each *S. salar* contig for all of the 8 timepoints and morph combinations. This was done separately for reads that mapped uniquely to one contig only (UNI) and reads that mapped to two or more contigs (REP). These SNP-candidates were further processed in R[60], following established principles for variant calling[67]. SNP-candidates at 90% frequency or higher in all samples were disregarded, as they reflect differences between Arctic charr and *S. salar* and are not the focus of this study. SNP-candidates with poor coverage in specific samples - i.e. coverage of five or fewer reads in three or four samples of each morph - were removed. As the SNP analysis was done on individual contigs, differences among paralogs appear in the data. To address this we use the fact that each sample is a pool of few individuals, thus true SNPs are unlikely to have the same frequency in all samples. However, variants reflecting differences between paralogs will have similar frequency all samples (assuming steady difference in their expression in all samples). We evaluated differences between samples with Fisher exact tests, and only SNPs significantly different between samples with a $p < 0.05$ (with no multiple testing correction) were retained. To compare morphs, read numbers were summed over the four samples from each morph. A conservative approach was taken by focusing on SNP-candidates that showed the largest differences in frequency between morphs (delta), without adjusting for multiple testing (Fisher exact test, $p > 5\%$). SNP-candidates with the highest frequency difference (delta > 95%) were manually processed and redundant candidates removed. A similar approach was used to mine for polymorphisms in Arctic charr mtDNA (NC_000861), using *S. salar* mtDNA as the outgroup (NC_001960.1).

We wrote a python script to predict the impact of SNPs within the mRNA sequences. Polymorphisms were categorized according to their location (3'UTR, coding, 5'UTR), and those within the coding region into synonymous or non-synonymous.

Verification of candidate SNPs

We chose 12 candidate SNPs for verification (see below). As the AC-charr is not a random breeding population, and because our interest is on differences between wild morphs, we took random samples of spawning SB, LB and PL-charr from Lake Thingvallavatn (8 per morph) from our earlier study[45]. Using the same PCR and DNA sequencing approach we genotyped 12 candidate SNPs (S2 Table). Briefly, we first compared the Salmon genome and ESTs [57, downloaded 22. March 2012] and short contigs from our preliminary assembly of the Arctic charr transcriptome. This allowed us to infer the placement of the putative polymorphism in the locus, and design paralog specific primers for PCR (less than 1 kb amplicons). MJ tetrad machine was used for PCR and the program was 5 min. at 95°C, followed by 35 cycles of 30 sec. at 52°C, 1 min. at 72°C, 30 sec. at 95°C, ending with 12°C while waiting on the human. Each individual was genotyped by first amplifying the region of interest using PCR, followed by ExoSAP (Affymetrix), direct sequencing (BigDye) and finally run on an Applied Biosystems 3500xL Genetic Analyzer (Hitachi). Raw data was base-called using the Sequencing Analysis Software v5.4 with KBTMBasecaller v1.41 (Applied Biosystems). Ab1 files were run through Phred and Phrap and imported to Consed for visual editing of ambiguous bases and putative polymorphisms, and for trimming primers. The FASTA files were aligned with ClustalW online [68, http://www.ebi.ac.uk/Tools/msa/clustalw2/] and manually inspected in Genedoc[69]. All sequences where deposited to Genbank as popsets under the accession numbers KP019972-KP020026.

Comparative genomic analyses of sequence polymorphisms

Two approaches were used for genomic comparisons of verified SNPs in the mitochondrial genome. Using the charr mtDNA sequence we performed both a BLAST search on salmon ESTs (May 2013) and retrieved multiZ alignments of vertebrates from the UCSC genome browser (in September 2013). This yielded several hundred sequences from related fish and other vertebrates. The list was reduced to 20 sequences for visualization, by keeping members of the major taxa but removing more closely related sequences, aligned with ClustalW and manually adjusted in Genedoc. The species and genome versions used are; Human (*Homo sapiens*, hg19), Lamprey (*Petromyzon marinus*, petMar1), Fugu (*Takifugu rubripes*, fr2), Medaka (*Oryzias latipes*, oryLat2), Stickleback (*Gasterosteus aculeatus*, gasAcu1), Tetraodon (*Tetraodon nigroviridis*, tetNig2), Zebrafish (*Danio rerio*, danRer6). We also downloaded from NCBI the sequence of whole or partial mtDNA from several fish species; Brown trout (*Salmo trutta*, JQ390057 and AF148843), Broad whitefish (*Coregonus nasus*, JQ390058), Legless searsid (*Platytroctes apus*, AP004107), Pacific menhaden (*Ethmidium maculatum*, AP011602), Icefish

F1000Research 2016, 4:136 Last updated: 25 DEC 2016

(*Salanx ariakensis*, AP006231 and HM151535), Chain pickerel (*Esox niger*, AP013046) and Western Pacific roughy (*Hoplostethus japonicus*, AP002938). The three mitochondrial variants (numbered by the *S. alpinus* mtDNA - NC_000861) are; m1829G>A (CCACGTTGTGAAACCAAC[G/A]TCCGAAGGTGGATTTAGCAGT), m3211T>C (CGTGCAGAAGCGGGCATAAG[T/C]ACATAAGACGAGAAGACCCT) and m3411C>T (CTCTAAGCACCAGAATTT[C/T]TGACCAAAAATGATCCGGC).

Results

RNA sequencing characteristics

Each sample yielded good quality data, with sequencing depth from 49 to 58 million (average: 55 million) reads. To quantify the expression levels, the reads were aligned to a salmon EST-assembly[57]. Around 20% of the reads mapped uniquely to the EST data (S3 Table). A further 30% mapped to two or more contigs, probably representing paralogous genes, recent duplications or repeat-like elements within transcribed regions. A substantial fraction of the RNA-sequencing reads did not map to the contigs from *S. salar*. Analyses of those reads require an Arctic charr genome sequence or transcriptome assembly from longer and paired end reads, currently underway in our laboratory.

Differential expression during Arctic charr development

We detected considerable changes in the transcriptome during Arctic charr development (Figure 3a). The expression of 1603 and 2459 paralog groups differed significantly between developmental timepoints at the 1% and 5% levels of false discovery rate (FDR), respectively (Dataset 1). The difference was most pronounced between prehatching (timepoints: 141, 163, 200 ᴛs) and post hatching embryos (timepoint 433 ᴛs), as more than 70% of the paralog groups with FDR below 1% had higher expression in the latter (Figure 3a). Gene Ontology analyses reveal six enriched GO categories (below 10%FDR). The most drastic changes were seen in processes related to glycolysis (GO:0006096, FDR = 0.0009), where the expression of 19 out of 25 paralog groups changed during this developmental period. The other five classes that were differentially expressed during charr development are: ion transport (GO:0006811, FDR = 0.027), blood coagulation (GO:0007596, FDR = 0.03), DNA repair (GO:0006281, FDR = 0.08) and two immune related categories (GO:0019882, FDR = 0.08, GO:0006955, FDR = 0.09). Those results probably reflect developmental changes and/or differences in the environment of embryos before and after hatching.

Differential expression between Arctic charr morphs

The embryos were reared in a common garden setting, which minimizes the impact of environmental factors, as we are interested in genes showing expression differences between the two morphs. In the data 296 paralog groups were differentially expressed (FDR < 5%) between the morphs (141 higher in SB and 152 higher in AC-charr, Dataset 1). Among genes with higher expression in SB-charr two biological GO categories were enriched: blood coagulation (GO:0007596, p = 0.001) and proteolysis (GO:0006508,

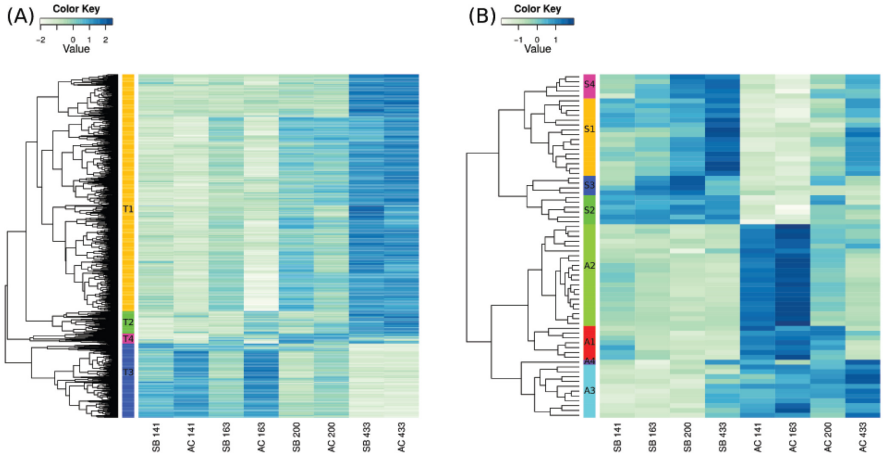


Figure 3. Heatmap of differentially expressed genes in the Arctic charr developmental transcriptome. Two morphs (SB and AC) are represented, at four timepoints. (**A**) The 1603 genes with expression difference among time points, here clustered into four groups. (**B**) The 71 genes differentially expressed between morphs are clustered into 4 groups for each morph. High expression is indicated by blue and low expression by beige.

p = 0.002). Recall, expression of blood coagulation factors also differed between developmental stages (see above). In AC-charr, genes in three categories: respiratory electron transport chain (GO:0022904, p = 0.0006), ATP synthesis coupled electron transport (GO:0042773, p = 0.002) and neurotransmitter transport (GO:0006836, p = 0.009) have higher expression. The first two GO categories both relate to energy generation in mitochondria and could reflect higher expression of genes with mitochondrial functions in AC-charr. At more stringent FDR (1%), 31 paralog groups, with diverse functional annotations, were higher expressed in SB and 40 genes higher in AC-charr (Figure 3b, Table 1 and Table 2). The higher expressed ESTs were clustered into 4 groups

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

NR	Name	Abbr	Cont	logFC	logCPM	FDR	Cluster
3766	Histone H3 embryonic		1	8.71	2.74	7.80E-035	S-1
5103	Natterin-like	*Nattl*	6	2.75	7.12	7.76E-007	S-2
356	A7J6M9 Putative uncharacterized protein n175R		1	2.33	4.66	3.30E-006	S-1
6697	Q1KY05 Main olfactory receptor-like	*Sorf*	5	3.12	6.92	9.96E-005	S-1
8151	Sushi domain-containing protein 2	*Susd2*	4	2.20	5.55	0.0001	S-3
1682	Carcinoembryonic antigen-related cell adhesion molecule 1	*Ceacam1*	3	2.55	3.83	0.0002	S-1
6228	Protein FAM98A		2	1.96	4.76	0.0003	S-1
7531	STAM-binding protein-like	*Stambpl1*	2	2.07	2.62	0.0005	S-1
6712	Q1M160 Myc-regulated DEAD box protein		1	1.67	3.23	0.0009	S-1
2300	Cytosolic sulfotransferase 3	*Sult3st1*	3	1.73	2.13	0.0009	S-1
2063	Complement factor D	*Cfd*	7	1.79	6.42	0.0016	S-3
3326	Galectin-3-binding protein A		4	1.79	3.85	0.0017	S-4
3169	Flocculation protein 11	*Flo11*	2	1.86	4.05	0.0017	S-1
1203	B5XDY0 H-2 class I histocompatibility antigen L-D alpha chain		2	1.70	2.12	0.0028	S-3
9183	UPI000065D844 related cluster		2	1.97	5.55	0.0028	S-1
2909	Epidermis-type lipoxygenase 3	*Loxe3*	4	1.68	4.84	0.0029	S-1
4884	Myeloperoxidase	*Mpo*	4	2.20	6.78	0.0029	S-1
10003	Uridine phosphorylase 1	*Upp1*	4	1.51	3.00	0.0047	S-1
2513	Desmoglein-1-alpha	*Dsg1*	1	1.59	2.80	0.0054	S-2
377	A7SJA8 Predicted protein (Fragment)		1	1.73	2.50	0.0055	S-3
9204	UPI00006A2900 related cluster		2	6.38	3.26	0.0064	S-1
9642	UPI00017B1B0F related cluster		1	2.00	1.92	0.0064	S-2
1965	Coiled-coil domain-containing protein 136	*Ccdc136*	2	2.15	2.32	0.0064	S-2
9260	UPI0000F1D4BA PREDICTED		1	1.80	2.41	0.0065	S-2
738	Adseverin	*Scin*	8	1.58	5.51	0.0073	S-1
9678	UPI00017B4479 related cluster		1	2.18	1.97	0.0074	S-4
8339	Testin	*Tes*	4	1.50	4.93	0.0080	S-2
6840	Q4SNH3 Chromosome 8 SCAF14543		1	1.42	4.00	0.0080	S-1
1668	Carbohydrate sulfotransferase 6	*Chst7*	1	2.09	2.08	0.0090	S-4
8341	Testisin	*Prss21*	2	2.01	2.76	0.0090	S-4
6373	Protein asteroid homolog 1	*Aste1*	6	1.29	4.24	0.0090	S-4

Name: name of unigene or paralog group
Abbr: Abbreviated paralog group or gene name
Cont: Number of contigs
logFC: log Fold Change
logCPM: log Counts Per Million
FDR: False Discovery Rate
The cluster numbering corresponds to Figure 3.

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

NR	Name	Abbr	Cont	logFC	logCPM	FDR	Cluster
3465	Glutathione S-transferase P 1	*Gstp1*	1	-8.35	2.45	1.12E-019	A-2
2475	Dehydrogenase/reductase SDR family member 7	*Dhrs7*	2	-4.88	2.15	9.67E-014	A-3
6945	Q6NWE8 Sb:cb283 protein		3	-6.08	3.02	2.15E-013	A-2
399	A8DW32 Predicted protein		1	-5.32	6.38	4.27E-010	A-1
9682	UPI00017B4B48 related cluster		2	-3.70	2.81	2.61E-008	A-2
9817	Uncharacterized protein ART2		5	-12.63	6.89	8.23E-008	A-2
6724	Q2L0Z2 Putative ATP-dependent RNA helicase		1	-3.41	1.89	1.88E-007	A-2
1197	B5XD10 Vacuolar proton pump subunit G 1	*Atpv1g1*	1	-4.30	2.10	1.84E-006	A-2
5325	Nucleoside diphosphate kinase B	*Nme2*	1	-9.85	7.63	2.51E-006	A-1
9205	UPI0000D5B923: myelin basic protein isoform 1	*Mbpa*	3	-2.49	3.45	9.18E-006	A-3
6377	Protein broad-minded	*Tbc1d32*	1	-2.11	2.74	4.75E-005	A-1
5711	Pistil-specific extensin-like protein		1	-2.16	2.60	0.0002	A-3
3203	Formin-like protein 20	*Fmnl2b*	7	-1.98	1.95	0.0002	A-3
9315	UPI0000F2EC69: hypothetical protein		2	-5.60	4.57	0.0005	A-2
363	A7RFV0 Predicted protein (Fragment)		2	-1.74	4.96	0.0010	A-3
6937	Q6AZT1 MGC81677 protein		3	-2.06	3.81	0.0014	A-2
3756	Histone H1	*Histh1*	3	-2.26	4.54	0.0017	A-2
1133	B5DGN9 Creatine kinase-1	*Ckm1*	7	-4.72	5.50	0.0017	A-3
309	A1IMH7 CD80-like protein	*Cd80*	12	-1.94	4.29	0.0017	A-2
7651	Serine protease ami		2	-1.54	5.90	0.0017	A-3
9935	Uncharacterized protein C7orf63 homolog		1	-1.87	1.91	0.0025	A-2
5219	Nostrin	*Nostrin*	2	-2.55	3.38	0.0029	A-2
1855	Chondroitin sulfate N-acetylgalactosaminyl-transferase 2	*Csgalnact2*	5	-2.56	6.14	0.0034	A-1
10203	Xylose isomerase		6	-1.55	2.43	0.0035	A-3
2249	Cytochrome c oxidase subunit 3	*Cox3*	11	-1.78	11.15	0.0035	A-3
180	40S ribosomal protein S3-B	*Rps3b*	2	-5.31	8.67	0.0050	A-1
1227	B6NBL3 Putative uncharacterized protein		3	-1.59	2.95	0.0050	A-2
5055	NADH-ubiquinone oxidoreductase chain 6	*Nd6*	2	-1.48	2.65	0.0061	A-3
4634	Metallothionein A	*Mta*	1	-3.33	5.44	0.0064	A-2
342	A5C0J4 Putative uncharacterized protein		2	-2.58	2.47	0.0064	A-2
9698	UPI00019258B4: similar to epithelial cell transforming sequence 2 oncogene protein partial		1	-2.06	2.94	0.0064	A-2
5878	Pro-opiomelanocortin B	*Pomcb*	1	-2.04	5.60	0.0065	A-2
2248	Cytochrome c oxidase subunit 2	*Cox2*	9	-2.21	9.83	0.0074	A-2
1246	B8JI87 Novel protein similar to vertebrate collagen type VI alpha 3 (COL6A3) (Fragment)		1	-1.69	3.16	0.0080	A-3
7994	Sperm-associated antigen 5	*Spag5*	1	-2.07	3.71	0.0080	A-2
9515	UPI000175F90F: similar to pleckstrin homology domain containing family A member 7		1	-2.00	1.87	0.0090	A-2
1124	B5DDZ4 Acta1 protein	*Actc1b*	1	-1.52	2.62	0.0090	A-2
1127	B5DG94 2-peptidylprolyl isomerase A	*Ppia1*	2	-2.56	5.67	0.0090	A-1
9175	UPI000054A3C0 PREDICTED: apolipoprotein B		3	-1.32	4.09	0.0090	A-4
9671	UPI00017B3C62 related cluster		1	-1.51	1.92	0.0096	A-1

For column header explanation, see footer of Table 1.

F1000Research 2016, 4:136 Last updated: 25 DEC 2016

for each morph, reflecting in some cases functional similarity. For instance SB cluster 3 has three immune related paralog groups: *Complement factor D* (9), *H-2 class I histocompatibility antigen L-D alpha chain* (2) and *Sushi domain-containing protein 2* (4) (Table 1). Note, however, that immune genes were not significantly enriched in the GO comparison of morphs. The results suggest genes with mitochondrial function, blood coagulation and other functions are differentially expressed between the morphs. Note, because only two morphs are compared, then those genes implicate pathways involved in either ecological divergence in SB charr or adaptation of the AC charr during breeding[50]. But as few samples were sequenced, qPCR verification was needed.

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

For validation we opted for qPCR analyses of 9 genes/paralog groups in whole embryos and 8 in embryonic heads (see next section), which showed differential expression between AC and SB-charr, with statistical support ranging from <1% to about 10% FDR. We studied paralog groups with less FDR support, in part to be able to cast a wider net (see below). Of the nine paralog groups studied in whole embryos, five were confirmed to be differentially expressed between AC and SB-charr at 161 or 200 τs (Figure 4, S4 Table and Dataset 2). Part of the reason may be that the transcriptome covered four developmental time points, but the

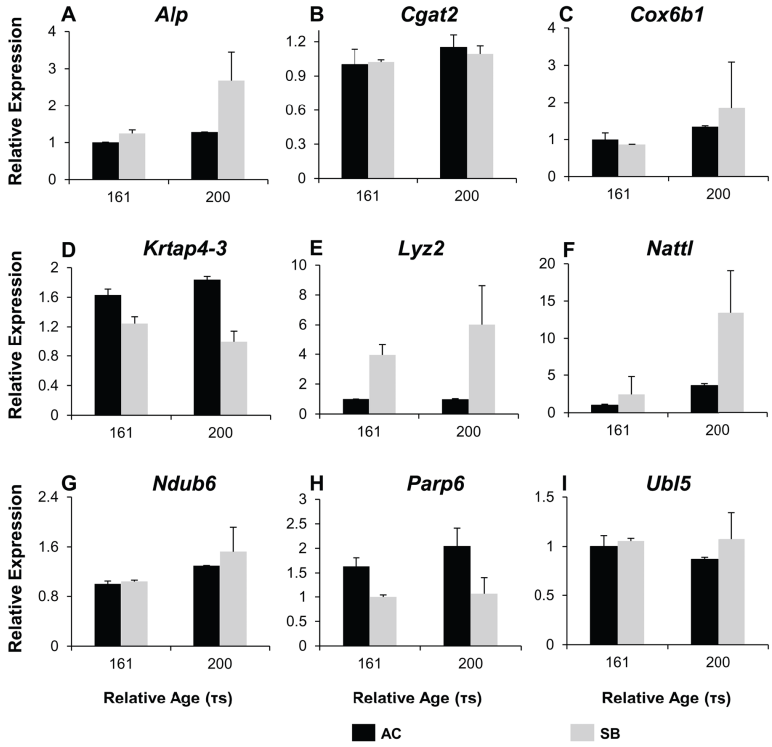


Figure 4. qPCR validation of candidates from transcriptome in whole embryos of Arctic charr. Relative expression of 9 genes (**A**– **I**) analysed by qPCR in the small benthic (SB) charr from Lake Thingvallavatn and aquaculture (AC) charr at two different developmental timepoints (161 and 200 τs). 5 genes were differentially expressed between the two morphs (*Alp, Krtap4-3, Lyz, Nattl, Parp6*), while 4 further genes did not show significant expression differences between morphs (*Cgat2, Cox6B1, Ndub6, Ubl5*), see S4 Table. Error bars represent standard deviation calculated from two biological replicates.

validation only two. Three genes, *Nattl, Alkaline phosphatase (Alp)* and *Lysozyme C II (Lyz2),* had significantly higher expression in SB. The other two, *Keratin-associated protein 4-3 (Krtap4-3)* and *Poly polymerase 6 (Parp6)* had higher expression in AC embryos (Figure 4, S4 Table). No morph and time interaction was detected for any of the genes.

As some genes are represented by different contigs or even paralogs, we set out to disentangle the expression of one paralog group *Nattern-like* (*Nattl*) in detail. We measured the expression of three *natterin* paralogs (*nattl1, nattl2 and nattl3*), by designing qPCR primers that matched divergent regions. These genes caught our interest because the only prior work implicated Natterin as a toxin produced by a tropical fish[70,71]. We studied *nattl* expression in several developmental stages in AC-, SB- and PL-charr as well as in selected tissues of adult AC-charr. The expression level of the three paralogs differed between morphs and timepoints (Figure 5 and S5 Table). Overall *nattl2* had the highest expression in all morphs. The *nattl1* had higher expression in embryos of PL-charr than in AC- and SB-charr, while *nattl2* and *nattl3* were more expressed in SB-embryos. Note however, the efficiency of the primers for the *nattl* genes ranged from 1.72 to 1.77, which suggests this data should be interpreted with caution.

In order to evaluate the hypothesis that *nattl* genes have immune-related functions we studied expression in adult tissues (in AC-charr). The *nattl* expression was highest in the gills, followed by expression in kidney, skin and spleen. Low expression levels were detected in liver, intestine and heart (S1 Figure and S5 Table). The three *nattl* paralogs followed different patterns, whilst each of them showed significant expression differences among tissues. *Nattl1* was mainly expressed in spleen and kidney, while *nattl2* showed a significantly higher expression in skin, liver and in gills. Similarly, the relative expression of *nattl3* was highest in the gills and skin. This indicates that the three *nattl* paralogs are expressed in a tissue specific manner, and also differently during the development of the three charr morphs studied here.

Expression differences in the developing heads of benthic and limnetic charr morphs

The transcriptome only compared two morphs, but we want to find genes with relationship with benthic form or ecology. Thus we next compared two benthic (SB, LB) and two limnetic charr (AC, PL). To get a handle on the craniofacial divergence between sympatric Arctic charr morphs we used qPCR to study 8 paralog groups with expression difference in the RNA-seq data (all higher in SB). We focused on those with known craniofacial expression

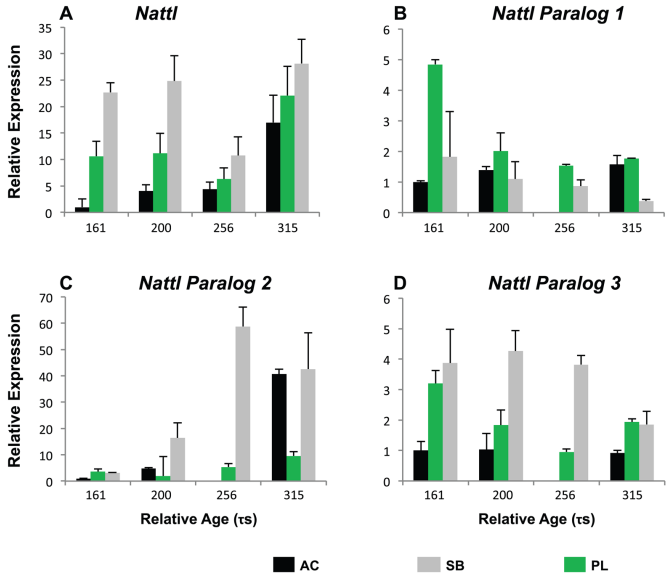


Figure 5. Relative expression of *Natterin-like* and its three paralogs during charr development in different morphs. The expression is graphed for different morphs (SB, AC and PL) at four developmental timepoints (161, 200, 256 & 315 *τs*, relative to AC-charr at timepoint 161. **A**) General *nattl* expression along charr development. **B**–**D**) Expression of *nattl* paralogs 1–3. ANOVA showing the variation among morphs is summarized in S5 Table.

in zebrafish development[72]. We analyzed heads at three time-points (178, 200 and 218 *ts*) as this period overlaps with early stages of craniofacial skeletal formation in Arctic charr[73,74]. The qPCR confirmed the higher expression of seven out of these eight genes in the head of benthic charr compared to limnetic charr (Figure 6, S2 Figure and Dataset 3). These seven genes are *Claudin 4 (Cldn4)*, *adseverin (Scin)*, *Junction plakoglobin (Jup)*, *Lipolysis stimulated lipoprotein receptor (Lsr)*, *Major vault protein (Mvp)*, *Transforming growth factor beta receptor II (Tgfbr2)* and *Vitamin D receptor a (Vdra)*. The eighth gene, *Retinoic acid receptor gamma-A (Rarg)* gave a small but significant response in the head, but the effects were reversed, i.e. the expression was higher in AC. The expression difference of the seven genes was, in almost all cases, consistent over the three timepoints studied (See S2 Figure). In summary the qPCR confirmed the differential expression of 12 of the 17 paralog groups studied (Table 3), some which had 5–10% FDR support. The data reveal notable expression differences between these two charr morphs, and can lead to hypotheses about morph specific variation in particular structures, like the developing head. However this transcriptome should not be taken at face value, because a substantial fraction of signals were false positives.

Analyses of polymorphism in Arctic charr transcriptome

The RNA-seq data also revealed segregating variations with large frequency differences between charr morphs. To uncover candidate SNPs we mapped the reads to all of the *S. salar* EST-contigs. Filtering on coverage yielded 165,790 candidate SNPs (Table 4); of those 66.569 came from reads that mapped uniquely and 57.009 candidate SNPs from reads that mapped to more than one contig; with limited overlap between lists. Assuming that the expression of paralogous genes is stable, then differences among paralogs appear as SNPs at similar frequency in all samples. By requiring variant frequency differences ($p < 0.05$, uncorrected) between samples we reduced the list of candidates by two thirds, yielding over 20.000 candidate SNPs. Note, as cDNA from charr families was sequenced (not a population sample), estimates of SNP frequencies are imprecise. To err on the side of caution, we chose SNP candidates with 50% or higher frequency difference between morphs for further study. The candidate SNPs were also summarized by frequency of the derived allele, in reference to the *S. salar* sequence. This gave 672 and 872 SNPs at higher frequency, in AC-charr and SB-charr, respectively. The uniquely and multiply mapped reads, revealed approximately similar numbers of candidate SNPs. Gene ontology analysis showed that for derived SNPs in SB, there was an excess of variants in genes related to translation, both as a broad category and specific subgroups (S6 Table). There was also enrichment of SNPs in genes related to DNA-mediated transposition, DNA integration, DNA replication and oxidation-reduction process. No GO categories were enriched for high frequency derived SNPs in AC. Furthermore, functional effects of the candidate SNPs (UTR, synonymous and non-synonymous) were predicted. The distribution among those categories did not differ between variants detected by uniquely or repeatedly mapped reads, $\chi^2_{[3]}=2.59$, $p = 0.46$ (S7 Table).

A total of 60 candidate SNPs are nearly fixed in one morph, with frequency difference between morphs above 95% (after manual inspection of contigs and SNP position three candidates were removed since they represented the same SNP). Of these "fixed" SNPs 46

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

Figure 6. Expression differences of craniofacial candidate genes in developing head of Arctic charr morphs. Relative expression ratios, calculated from the qPCR data, were subjected to an ANOVA to test the expression differences amongst four charr groups and three time points (*ts*). The underlined gene names reflect significant difference between SB and AC-charr. A post hoc Tukey's test (HSD) was performed to determine the effects of morphs, time and morph by time interaction (M X T). White boxes represent low expression, while black boxes represent high expression. The shading represents significant different expression between the samples ($\alpha = 0.05$, NS = not significant). The genes studied were, *Claudin 4 (Cldn4), adseverin (Scin), Junction plakoglobin (Jup), Lipolysis stimulated lipoprotein receptor (Lsr), Major vault protein (Mvp), Transforming growth factor beta receptor II (Tgfbr2) Vitamin D receptor a (Vdra) and Retinoic acid receptor gamma-A (Rarg).*

Table 3. Correspondence of transcriptome and qPCR verification on Arctic charr embryos.

Tissue	Name	Abbr	FDRm	FRDt	Effect	qPCR	Morph
Embryo	Alkaline phosphatase	*Alp*	0.070	0.001	0.986	*	SB
Embryo	Chondroitin sulfate N-acetylgalactosaminyltransferase 2	*Cgat*	0.004	0.331	-2.556		
Embryo	Cytochrome c oxidase subunit 6B1	*Cox6b1*	0.058	0.632	-1.208		
Embryo	B5X596 Keratin-associated protein 4-3	*Krtap4-3*	0.012	0.278	-1.986	*	AC
Embryo	Lysozyme C II	*Lyz2*	0.041	0.001	1.138	*	SB
Embryo	Natterin-like protein	*Nattl*	0.000	0.000	2.755	*	SB
Embryo	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 6	*Ndub6*	0.098	0.670	-1.175		
Embryo	Poly [ADP-ribose] polymerase 6	*Parp6*	0.108	0.379	-0.986	*	AC
Embryo	Ubiquitin-like protein 5	*Ubl5*	0.059	0.003	-1.234		
Head	Claudin-4	*Cldn4*	0.068	0.000	1.343	*	SB/LB
Head	Major vault protein	*Mvp*	0.065	0.528	0.958	*	SB/LB
Head	Junction plakoglobin	*Jup*	0.051	0.006	1.147	*	SB/LB
Head	Lipolysis-stimulated lipoprotein receptor	*Lsr*	0.013	0.043	1.369	*	SB/LB
Head	TGF-beta receptor type-2	*Tgfbr2*	0.065	0.013	1.728	*	SB/LB
Head	Vitamin D3 receptor A	*Vdra*	0.053	0.052	1.312	*	SB/LB
Head	Retinoic acid receptor gamma-A	*Rarg*	0.012	0.001	1.403		
Head	Adseverin	*Scin*	0.007	0.000	1.578	*	SB/LB

Tissue: which tissue was studied

Abbr: abbreviated paralog group or gene name

FRDm: FDR for comparison of SB and AC-charr in transcriptome

FDRt: FDR for comparison among developmental timepoints in transcriptome

Effect: logarithm of fold change between morphs, positive is higher in SB and negative higher in AC-charr in transcriptome (logFC.morph in supplemental dataset 1)

qPCR: results consistent with transcriptome (*), a blank cell reflects lack of correspondence

Morph: which morph(s) had higher expression in qPCR verification

Table 4. Candidate SNPs in the Arctic charr transcriptome, filtered by coverage, difference between sample and morphs and frequency difference between morphs.

SNP-candidates	Morph	Uni	Rep	Total
Total		96231	74341	165790
Filter coverage		66569	57009	113776
Diff. Bwn. samples		21417	22252	42869
Diff. Bwn. morphs		11385	12953	23974
Delta > 0.5	AC	396	285	672
Delta > 0.5	SB	526	353	872
Delta > 0.75	AC	95	68	159
Delta > 0.75	SB	155	95	248
Delta > 0.95[a]	AC	17	13	30
Delta > 0.95[a]	SB	29	4	33

SNP-candidates: found by mapping to *S. salar* ESTs

Uni/REP: from UNIquely or REPeatedly mapped RNA-reads

Delta: differences in allele frequency between morphs, categorized by which morph had the higher derived allele frequency

[a]The number of SNP-candidates before the redundant ones were removed

came from uniquely mapped reads and 14 from reads that mapped more than twice (Table 5 and Table 6). For the SNPs from uniquely mapped reads, 17 are fixed in AC-charr and 29 in SB-charr. The few genes with two or more polymorphic sites were; *Keratin type II cytoskeletal 3 (Krt3)*, *Cysteine sulfinic acid decarboxylase (Csad)* and *DNA-directed RNA polymerase I subunit RPA12 (Rpa12)* with 5, 5 and 2 SNPs respectively. *Krt3* and *Csad* had significant differentiation in both SB and AC. Similarly, 14 SNPs with large

Table 5. SNP candidates from uniquely mapped reads.

(a) Higher frequency in AC morph							
Contig	**Annotation**	**Pos**	**Ref**	**Var**	**Freq-SB**	**Freq-AC**	**Effect**
SS2U026955	Keratin type II cytoskeletal 3	300	A	T	0.000	0.984	synonymous
SS2U026955	Keratin type II cytoskeletal 3	309	G	A	0.000	0.996	synonymous
SS2U033960	Cysteine sulfinic acid decarboxylase	192	C	G	0.000	1.000	5prime
SS2U033960	Cysteine sulfinic acid decarboxylase	416	G	T	0.000	0.961	G to V
SS2U033960	Cysteine sulfinic acid decarboxylase	945	C	A	0.004	0.956	synonymous
SS2U043396	Eukaryotic translation initiation factor 2-alpha kinase 1	134	A	G	0.000	1.000	5prime
SS2U043886	Transcription cofactor HES-6	1308	T	C	0.000	1.000	5prime
SS2U044339	Intraflagellar transport protein 52 homolog	479	T	C	0.021	1.000	D to G
SS2U045168	Putative Peptide prediction	1275	G	A	0.000	1.000	3prime
SS2U045328	E3 ubiquitin-protein ligase DTX3L	388	G	A	0.000	0.977	synonymous
SS2U045990	Low-density lipoprotein receptor-related protein 1	135	T	C	0.000	0.969	synonymous
SS2U048125[a]	Transmembrane protein 131-like	480	G	A	0.000	1.000	synonymous
SS2U052747	Uridine 5'-monophosphate synthase	914	G	A	0.000	0.951	synonymous
SS2U054542	Mediator of RNA polymerase II transcription subunit 20	474	C	T	0.027	0.995	synonymous
SS2U056193	SUMO-conjugating enzyme UBC9	96	A	T	0.000	1.000	3prime
SS2U057101	ETS domain-containing protein Elk-3	440	C	G	0.000	1.000	3prime
SS2U058860	Voltage-dependent anion-selective channel protein 2	681	G	T	0.000	1.000	3prime
(b) Higher frequency in SB morph							
Contig	**Annotation**	**Pos**	**Ref**	**Var**	**Freq-SB**	**Freq-AC**	**Effect**
SS2U000399	Insulin-like growth factor-binding protein 7	598	C	A	1.000	0.000	3prime
SS2U004484	Titin	387	G	A	0.990	0.010	synonymous
SS2U026826	L-asparaginase	363	C	T	1.000	0.000	H to Y
SS2U026955	Keratin type II cytoskeletal 3	116	C	A	0.996	0.031	T to N
SS2U026955	Keratin type II cytoskeletal 3	264	C	T	0.970	0.008	synonymous
SS2U026955	Keratin type II cytoskeletal 3	317	C	T	1.000	0.002	T to M
SS2U033960	Cysteine sulfinic acid decarboxylase	363	C	T	1.000	0.025	5prime
SS2U033960	Cysteine sulfinic acid decarboxylase	387	C	T	1.000	0.030	synonymous
SS2U033960	Cysteine sulfinic acid decarboxylase	657	T	C	0.990	0.031	synonymous
SS2U034322	Cyclin-C	1094	A	G	1.000	0.000	3prime
SS2U034431	Dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit 2	436	G	A	0.992	0.000	G to S
SS2U036025	Nuclear receptor coactivator 4	36	G	A	1.000	0.043	5prime
SS2U040590	Glutamyl-tRNA(Gln) amidotransferase subunit A homolog	478	G	A	0.972	0.000	synonymous
SS2U045606	Superkiller viralicidic activity 2-like 2	500	C	T	1.000	0.000	synonymous
SS2U047816	Squalene synthase	1139	G	A	1.000	0.029	synonymous

(b) Higher frequency in SB morph							
Contig	**Annotation**	**Pos**	**Ref**	**Var**	**Freq-SB**	**Freq-AC**	**Effect**
SS2U048063	Lysine-specific demethylase NO66	669	C	T	1.000	0.000	synonymous
SS2U050394	UPF0542 protein C5orf43 homolog	596	G	A	1.000	0.000	synonymous
SS2U050880[a]	Transmembrane protein 131-like	901	C	T	1.000	0.000	A to V
SS2U052076	Eukaryotic translation initiation factor 3 subunit A	824	C	T	1.000	0.031	synonymous
SS2U053417	RNA polymerase-associated protein LEO1	454	G	A	1.000	0.049	synonymous
SS2U054333	Scaffold attachment factor B2	382	G	A	0.999	0.000	V to M
SS2U054705	Cell division protein kinase 4	122	A	G	0.971	0.000	3prime
SS2U054965	DNA-directed RNA polymerase I subunit RPA12	106	G	A	1.000	0.000	5prime
SS2U054965	DNA-directed RNA polymerase I subunit RPA12	411	T	G	1.000	0.000	synonymous
SS2U055120	Chromatin modification-related protein MEAF6	350	A	C	1.000	0.000	H to P
SS2U055153	Complexin-1	1191	C	A	1.000	0.031	3prime
SS2U057635	Mitogen-activated protein kinase 14B	1370	A	T	1.000	0.026	3prime
SS2U058169	Transmembrane protein 50A	1214	C	G	0.973	0.000	3prime
SS2U058802	Signal recognition particle 54 kDa protein	607	T	A	0.969	0.000	C to S

[a]Those genes are distinct paralogs

Table 6. SNP candidates with significant difference frequency between AC and SB morphs, from reads that mapped to two or more contigs.

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

[a]Those genes are distinct paralogs

differentiation between morphs were predicted from reads that mapped on two or more contigs (Table 6). Of these, we found two variants in the mitochondrial *60S ribosomal protein L36 (RpL36)* and variants in 4 other mitochondrial genes *(28S ribosomal protein S18a mitochondrial (MRPS18A)*, *Apoptosis-inducing factor 1 mitochondrial (AIFM1)*, *Isocitrate dehydrogenase [NADP] mitochondrial (acIDH1)* and *Protein S100-A1 (S100a1))*, all at higher frequency in AC-charr. PCR and Sanger sequencing of population samples confirmed SNPs in *DNA2-like helicase (Dna2)*, a gene with nuclear and mitochondrial function, and two other genes *Uroporphyrinogen decarboxylase (Urod)*, and *Mid1-interacting protein 1-like (Mid1ip1)* (S2 Table). The candidate variant *Eukaryotic translation initiation factor 4 gamma 2 (Eif4g2)* was not substantiated by the PCR/sequencing.

Polymorphism and expression of Arctic charr mtDNA

Considering the enrichment of differentially expressed genes related to mitochondrial energy metabolism (above), and high frequency candidate SNPs in several genes with mitochondrial function in AC-charr we decided to study the mitochondrial transcriptome further. The charr studied here reflect metabolic extremes, the aquaculture charr was bred for growth while the small benthic morph is thought to have experienced natural selection for slow metabolism and retarded growth[38,75]. Although mRNA preparation protocols were used for generating cDNA for the RNA-sequencing, a substantial number of reads came from non-polyadenylated sequences. By mapping the reads to mtDNA sequence of Arctic charr we could estimate expression and infer polymorphism both in genes and intergenic regions. There was a clear difference in sequencing coverage, with more than twice as many reads mapped from the AC- compared to SB-charr (mean fold difference 2.27, Wilcoxon test, $p < 0.0004$). Note, as only two types of fish are compared, the polarity of expression divergence is unknown.

The mapped RNA-reads were used to identify polymorphism and divergence in the entire mitochondrial chromosome. The polymorphisms were found by mapping to mtDNA from a Canadian *S. alpinus*[48], but ancestral vs. derived status inferred by comparison to *S. salar* mtDNA. This revealed 82 candidate sites, including 35 that represent divergence between Icelandic and Canadian charr. A total of 20 candidate SNPs had high (more than 50%) frequency difference between SB- and AC-charr (Figure 7). There was no bias in the distribution of derived SNPs, 11 on the AC branch and 9 in SB. The divergence between Iceland and Canada is particularly little in the 12s and 16s ribosomal RNA genes. Curiously two SNPs in those genes differed strongly in frequency between morphs (Figure 7). To confirm and better estimate the frequency of variants in the ribosomal genes, we PCR amplified and sequenced two ~550 bp regions in the rRNA genes. Because of our interest in the evolutionary genetics of sympatric charr, we three morphs (PL, LB and SB) from Lake Thingvallavatn (Figure 8A, C & E, S2 Table). The 12s polymorphism (m1829G>A) differed significantly between the morphs ($\chi^2_{[2]} = 8.6$, $p = 0.014$), and was at highest frequency in the SB (0% in PL, 12.5% in LB and 75% in SB). Similarly m3411C>T in the 16s was enriched in SB (62.5%) but found at lower frequency in PL (0%) and LB (12.5%) (it differed significantly between morphs, $\chi^2_{[2]} = 9.3333$, $p = 0.009$). The Sanger sequencing also revealed three other polymorphisms in the amplified region, not seen in the transcriptome. Among those m3211T>C in the 16s gene was at 75% frequency in LB, but not found in the other morphs ($\chi^2_{[2]} = 19.76$, $p < 0.0001$).

In order to gauge the potential functionality of those variants we aligned the rRNA genes from nearly hundred fishes and several vertebrates. The position affected by m1829G>A and m3211T>C, in the 12s and 16s rRNAs, are not well conserved in fishes or vertebrates (Figure 8B & D). However m3411C>T, in the 16s

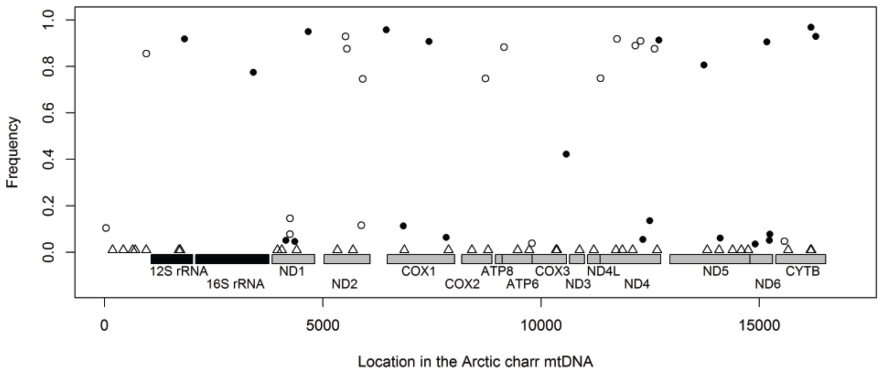


Figure 7. Genetic divergence in the mtDNA between SB- and AC-charr. The frequency differences between morphs of candidate SNPs, estimated from the RNA-sequencing, graphed along the mtDNA chromosome. The SNPs indicate whether the derived allele is of higher frequency in SB (black dots) or AC (open circles). Sites of divergence between the Icelandic stocks and the Canadian reference sequence are indicated by triangles. The two black boxes represent the rRNA genes and gray boxes the 14 coding sequences (abbreviated names underneath each gene).

F1000Research 2016, 4:136 Last updated: 25 DEC 2016

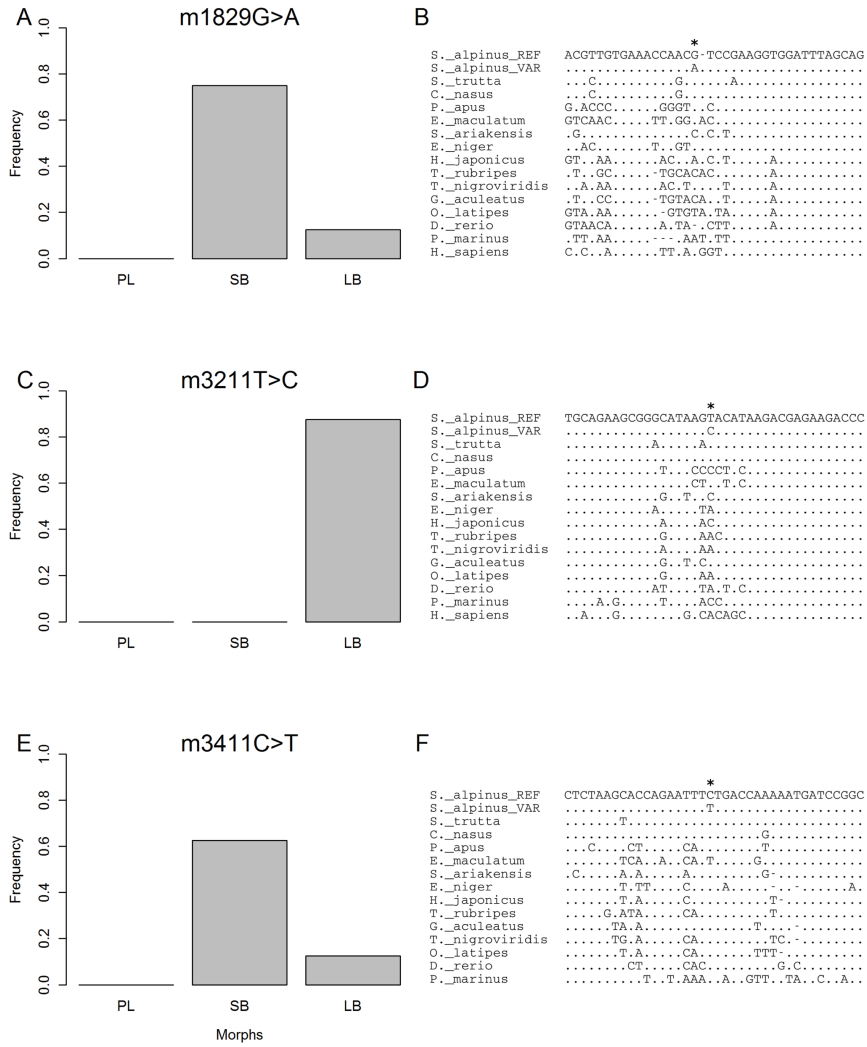


Figure 8. Comparative genomics and population genetic differentiation in Arctic charr at 3 mtDNA locations. Three variants in the 12s and 16s RNA genes are segregating in charr morphs in Lake Thingvallavatn. **A**, **C**, **E**) Frequency of each of those variants in three morphs from Lake Thingvallavatn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

rRNA, alters a position that is nearly invariant in 100 fish genomes (Figure 8F). The only exception is Pacific menhaden, which curiously also has T in this position. This region could not be aligned properly in other vertebrates. Thus m3411C>T alters a conserved position, but probably not very drastically as the introduced allele is tolerated in another fish species.

Dataset 1. Parameters and multiple testing corrected p-values for expression analysis

10.5256/f1000research.6402.d48005

The file is tab-delimited and the columns are; "Unigene.Description": the annotation for that gene/paralog group. "NR.contigs": number of contigs with this annotation. "logCPM": count per million, log-scale. "logFC.morph": Mean fold change between the morphs, log-scale. "logFC.T163", "logFC.T200", "logFC.T433": Mean fold change for each timepoints compared to timepoint 141, log-scale. "FDR.morph": P-value for morph difference, multiple testing corrected. "FDR.time": P-value for time differences, multiple testing corrected. "Contigs": SalmonDB id for the contigs with the specific annotation[109].

Dataset 2. qPCR data for tests of expression in charr developing embryos and adult tissues

10.5256/f1000research.6402.d48006

"Gene Type": Designates the reference and candidate genes. "Gene": Name of the gene. "Morph": Which charr type the sample came from. "Relative age": Developmental timepoint, and also indicates the samples from adult fish. "Biological replicate": The two or more biological replicates used. "cDNA No": Marks the cDNA isolation used. "Ct value": Estimate of gene expression. "Sample": Indicates the material used, whole embryos or distinct tissues. "Batch": Demarcates distinct collections of cDNA, applies only to nattl[110].

Dataset 3. qPCR data for tests of expression in charr developing embryo heads

10.5256/f1000research.6402.d48007

"Gene Type": Designates the reference and candidate genes. "Gene": Name of the gene. "Morph": Which charr type the sample came from. "Relative age": Developmental timepoint. "Biological replicate": The two or more biological replicates used. "cDNA No": Marks the cDNA isolation used. "Ct value": Estimate of gene expression. "Tissue": Indicates the material used[111].

Discussion

We are interested in the predictability of evolution at the molecular level, especially whether there exist principles that influence the rewiring of developmental and regulatory systems[4,76]. One way to study this is to identify genetic and developmental effects affecting key traits in species or populations which exhibit parallel evolution. The aim of this study was to find expression and genetic differences separating the small benthic morph in Lake Thingvallavatn and aquaculture charr, with the long term objective being to reveal the genetic and molecular systems that associate with benthic morphology in charr. The transcriptome reflects the biology of these two morphs, their different histories and ecology. AC-charr will also be shaped by domestication, which may explain for instance the higher expression of metabolic genes in AC-charr.

Developmental transcriptome of Arctic charr morphs

As no reference genome is available for Arctic charr, we mapped reads to *S. salar* EST-contigs[57] in order to estimate expression and identify candidate genetic polymorphisms. As many of the contigs are short or have overlapping annotations, we collapsed genes into paralogous genes when appropriate for the expression analysis. The main advantage was the reduced number of statistical tests (and hence an increase in statistical power). The downside is that paralog-specific expression patterns are masked, as the qPCR results of the *natterin like* gene family show (Figure 5 and S1 Figure). Recent rainbow trout data shows about 1/4 of paralogs from the latest whole genome duplication event retain the very similar expression patterns[16] indicating that distinct expression patterns of paralogs is quite common[77]. In their analysis of the Arctic charr gill transcriptome, Norman *et al.* (2014)[23,24] also used Illumina sequencing technology to evaluate expression. Their reads were longer (2x100 bp) than in this study (36 bp) enabling them to assemble contigs. They did not consider distinct paralogs in their approach and merged contigs based on sequence identity. Thus the complexity of Arctic charr transcriptome still remains unsolved. The data reflected differential deployment of several gene classes during Arctic charr development, which is most probably genetic in origin. We raised the embryos in a common garden, but their parents were wild so parental environments and transgenerational plasticity may also have contributed. Studies in salmonids and other fish have demonstrated large changes in expression during early development, including coordinated changes in many cellular and developmental systems[19,78–81]. Several blood coagulation factors genes showed significant changes during charr development, and were also more highly expressed in the SB-charr. This might reflect differences in the rate of development of blood composition, or tissue composition, in the two morphs. While our main interest is on the derived and repeatedly evolved small benthic charr, the data can also reflect differences due to breeding. As was reasoned in the introduction we chose to compare SB to AC-charr. This proved useful, as the data revealed differential expression of several developmental genes and regulators with differential expression between benthic and limnetic charr[51,52]. Previously we found tight correlation of RNA-seq expression and qPCR estimates - using data from this very transcriptome[51]. Furthermore, we actually used the same morphs (AC and SB) and samples in a comparison of the developmental miRNA transcriptome – which reveal that expression of several miRNAs correlates with morph differences[56].

Higher expression of *lysozyme II C* and *natterin-like* in SB-charr

Natural selection can shape variation in immunological genes. We decided to study further *Lyz2* and the putative immunological *nattl* genes that had higher expression in SB. Note, because only two charr transcriptomes studied, it was impossible to polarize the changes. It was not possible to say that these genes are upregulated in SB or downregulated in AC charr. The substrate of lysozyme[82] is the bacterial cell wall peptidoglycan and it acts directly on Gram-positive bacteria[83]. Lysozyme also promotes the degradation of the outer membrane and therefore indirectly acts also on Gram-negative bacteria[84]. Another gene that caught our attention was *natterin-like*. Natterins were first discovered from the venom

gland of the tropical toxic fish species *Thalassophryne nattereri*[70,71], and are found by sequence similarity in e.g. zebrafish, Atlantic salmon and here in Arctic charr. The Natterin proteins contain a mannose-binding lectin-like domain (Jacalin-domain). Mannose-binding lectins are pathogen recognition proteins (antibodies) and therefore are important for the acute phase response of fish[85,86], thus we hypothesized that *nattl* genes in charr may have immune related functions. The data are consistent with this as the highest expression was found in skin and kidney. This putative immune functions needs to be verified. One can speculate that higher expression of *Lyz2* and some *Nattl* paralogs in SB-charr reflect preparation of juveniles for bottom dwelling habitats, which may be rich in bacteria and challenging for immune systems. It would be interesting to study further the expression of these and other immunological genes implicated in this transcriptome, in natural charr populations and juveniles challenged with pathogens or in families of Aquaculture charr breed for pathogen resistance. An evolutionary question is, whether immunological genes are expected to show similar or less parallelism than others genes shaped by natural selection? The current data does not reflect on this question, but our population genetic work shows genetic variation in immunological genes (*MHCIIα* and *cath2*) does not correlate with the SB-charr ecotype in Iceland[45].

In this study we collapsed contigs into paralog groups for the transcriptome analyses. The disadvantage of this approach is that differential expression of a paralog, can be masked by related genes that do not differ between groups. We looked at this by studying the expression of three paralogs of the *natterin like* genes in different morphs during Arctic charr development, and among tissues of adult AC-charr. The data suggest that the three *nattl* genes are expressed differentially between the morphs, thus it is not divergence in the expression of one paralog that explains the general *nattl* expression disparity in the transcriptome. Certainly, other scenarios could apply to other genes in the transcriptome.

Expression divergence in craniofacial genes in benthic morphs

A study of the skulls of post-hatching embryos and juveniles from Lake Thingvallavatn, showed that some elements of the developing head ossified earlier in SB than in PL-charr[87]. Morphometric analyses of developing heads (same stages as studied here) demonstrate differences in craniofacial elements between AC and SB-charr, along a limnetic vs. benthic axis[74]. Based on those developmental phenotypes we investigated further genes with roles in craniofacial development that were differentially expressed in the transcriptome. Our published data[51,52] and the current data (7 out of 8 craniofacial candidates were confirmed by qPCR) demonstrate the utility of the SB and AC-charr developmental transcriptomes for identifying candidate genes with differential expression, even within specific structures like the head. All seven of the verified genes had consistently higher expression in the developing head of two benthic morphs (SB and LB), and lower in more limnetic fish (AC and PL). We must highlight the fact that three of these morphs (SB, LB and PL) are closely related and live in sympatry in Lake Thingvallavatn[44].

We focused on a several craniofacial candidate genes including a few within, or transcriptionally connected to, the Tgf-β and Ahr signaling pathways[88–90]. These are the *Lsr*, *Cldn4, Jup, Scin, Vdra, Myp* and *Tgfbr2*, here described briefly. *Adseverin (Scin)* has roles in rearrangements of the actin cytoskeleton, chondrocyte differentiation and skeletal formation[91,92]. *Lsr* encodes a component of tri-cellular tight junctions[93] and has been shown to be suppressed upon Tgf-β1 stimulation[94] in a human cell line. Similarly, *Cldn4*, a tight junction protein with unknown role during embryonic morphogenesis, is a target of the Tgf-β and Ahr signaling pathways[95,96]. The *Tgfbr2*, encoding a receptor of Tgf-β, i involved in craniofacial morphogenesis[97]. *Myp* is the predominant component of cytoplasmic ribonucleoprotein structures called vaults[98], which is highly conserved across eukaryotes and are implicated in several processes from signal transmission and immune response[99]. Finally, higher expression of *Vdra*, encoding the vitamin D receptor A, was found in the heads of benthic charr. The receptor regulates mineral homeostasis, osteoblast differentiation and bone metabolism[100]. A related study from our group, building on this transcriptome, described in more detail the differential expression of these and other coexpressed genes in limnetic and benthic charr[53].

To summarize, the results show that RNA-sequencing of Aquaculture charr with limnetic craniofacial morphology and small benthic charr can implicate candidate genes for qPCR analyses. Those studies[52,53] have revealed genes that associate with limnetic and benthic divergence in craniofacial elements in sympatric charr morphs. It would be interesting if expression of these genes associates with benthic morphology in independently evolved charr populations, as was seen for certain mTOR-pathway genes in muscle of adult SB-charr[47], or even in other species with similar trophic diversity.

Genetics differences between the AC and SB-morphs - possibly in mtDNA function

Previous studies on microsatellite markers documented the history of charr populations in Iceland and in particular the parallel evolution of SB-charr[44]. The data confirm genetic differences between SB and AC-charr. By comparing AC and SB-charr, that represents a small benthic resource morph that has evolved repeatedly in Icelandic stream and pond habitats[44], we hoped to implicate genes and pathways involved in adaptation to these special habitats. The allele frequency differences and expression divergence observed in the transcriptome reflect neutral population genetic processes and/or selection during AC charr domestication or adaptation of SB-charr. Changes in the AC-charr are interesting as domestication over several decades led to rapid growth and increased size[50]. Morphometrics have not been used to compare the body or craniofacial shape of AC to other charr morphs, but domestication of *O. mykiss* has affected body shape and fin structure in particular[101]. By studying expression and allele frequencies in limnetic and benthic morphs from more locations, it may be possible to disentangle the role of drift and selection. We attempted to verify several SNPs, and focused mostly on variants in mtDNA because to us the data suggest interesting divergence between

AC and SB charr in systems related to energy metabolism. First, there is 2X higher expression of respiratory electron transport chain components in AC compared to SB-charr and 100% more mitochondrial derived reads are found in the AC-charr samples. Note that the direction of divergence is unknown, i.e. whether expression was up in AC or down in SB. Second, many derived candidate-SNPs in genes related to mitochondrial function were at high frequency on the AC branch. For instance in *S100A1*, which has been implicated in mitochondrial regulation in cardiac tissue in humans[102], but its expression is probably not exclusive to this tissue. Third, while the mitochondrial ribosomal genes generally evolve slowly, we do see derived variants at high frequency in the SB and large benthic charr in Lake Thingvallavatn. Specifically, m3411C>T in SB affects a position that is highly conserved among fish, and could affect function of the 16s rRNA. Earlier studies of mitochondrial markers in *S. alpinus* did not find large signals of divergence within Iceland[40,42,45], probably because they studied other genes.

The mitochondrion is more than a powerhouse, it integrates metabolism, cell cycle and apoptosis[103]. The number of mitochondria and its functions are known to correlate with environmental attributes. For instance in Antarctic fishes under extreme cold, higher numbers of mitochondria are found in muscle and heart cells[104]. Our data suggest an expression difference between morphs that could reflect differences in total number of mitochondrion, the number of mtDNA copies per mitochondrion or cell, or difference in RNA expression from the mtDNA, possibly due to evolution of mtDNA related to diet and/or temperature[105]. The results suggest divergence (adaptive or neutral) in mitochondrial function due to the domestication of aquaculture charr and/or adaptation of the small benthic charr to its habitat. Increase in mitochondrial function in AC charr embryos could reflect higher basal metabolic rate in this aquaculture stock. Alternatively, lower metabolic rate in the SB charr would also be curious in the context of their ecology. Clearly further work is needed to map out the functional differences of mitochondrial related genes in AC charr, more SB populations and hopefully anadromous charr morphs (representing the ancestral state). The mtDNA signals could also be investigated in populations along ecological clines (e.g. temperature) or with respect to life history[106].

Conclusions

The charr developmental transcriptome provides a starting point to investigate the molecular systems that associate with artificial selection during aquaculture breeding of charr or divergence among the highly polymorphic and rapidly evolving Arctic charr in Iceland. The data reveal differential expression of two immunological genes between morphs and of several craniofacial developmental genes, that may help sculpture benthic vs. limnetic heads. The genetic data suggest among other things differentiation in the charr mtDNA between the SB and AC-charr morphs. It must be acknowledged that it is not trivial to identify genes affecting variation in ecologically important phenotypes, like shape[107,108]. Our broad interest is in how natural selection tweaks genetic regulatory systems, for instance via genetic changes in regulatory sequences or post transcriptional modifiers relating to adaptations. Genetic changes affecting gene expression can be raw material for adaptation, but could also rise in frequency due to reverberations in regulatory cascades[76]. We plan to study the degree of developmental and population genetics parallelism of the small benthic charr, typically found in cold springs and small pond habitats in Iceland with lava substratum[29,44]. The availability of charr populations at different stages of divergence sets the stage for future genomic studies of the roles of genes, environment and plasticity for shaping this polymorphic species.

Data availability

The sequencing reads were deposited into the NCBI SRA archive under BioProject identifier PRJNA239766 and with accession numbers: SRX761559, SRX761571, SRX761575, SRX761577, SRX761451, SRX761461, SRX761490 and SRX761501.

All DNA sequences where deposited to Genbank as popsets under the accession numbers KP019972-KP020026.

F1000Research: Dataset 1. Parameters and multiple testing corrected p-values for expression analysis, 10.5256/f1000research.6402.d48005[109]

F1000Research: Dataset 2. qPCR data for tests of expression in charr developing embryos and adult tissues., 10.5256/f1000research.6402.d48006[110]

F1000Research: Dataset 3. qPCR data for tests of expression in charr developing embryo heads., 10.5256/f1000research.6402.d48007[111]

Author contributions

- Conceived and designed the study: JG, AP, ZOJ, SSS, SRF, VHM, EPA.
- Sampling, crosses and rearing: SSS, BKK, ZOJ, KHK, VHM, AP.
- RNA extraction and RNA sequencing: SRF.
- Analyses of RNA sequencing data: JG, AP.
- qPCR work: EPA, SSS2, VHM.
- SNP analyses: JG, AP.
- SNP confirmation: IMJ, KHK, AP.
- Comparative genomic analysis: AP.
- Writing: AP, JG, EPA, VHM, SSS.
- Analyses: JG, AP, EPA, SSS2.
- Gathered the data: ZOJ, SRF, EPA, IAJ, KHK, SSS2.

Competing interests

No competing interests were disclosed.

Grant information

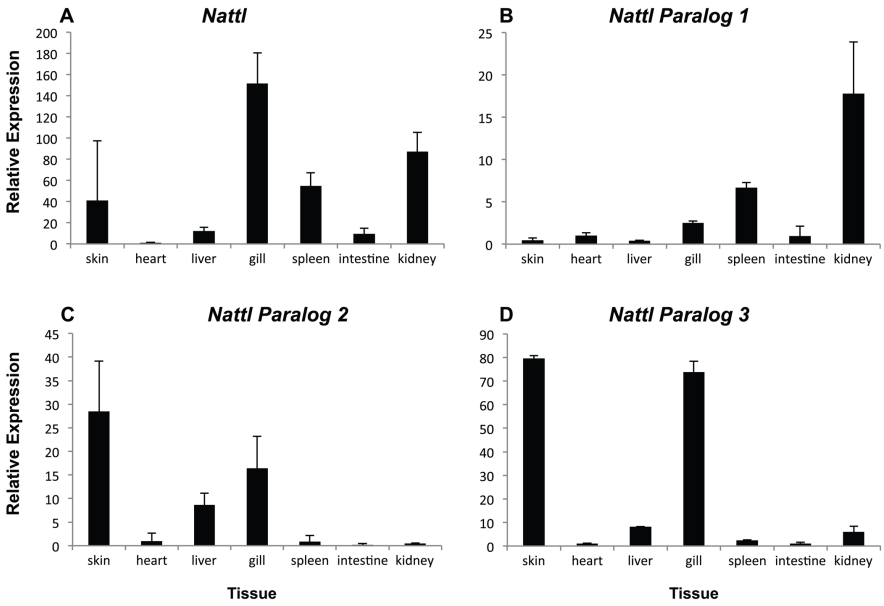
This project was supported by The Icelandic Center for Research (grant number: 100204011) to SSS, AP, ZOJ and BKK, The University of Iceland Research/Doctoral Fund to JG and KHK and University of Iceland research fund to AP, SSS and ZOJ.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

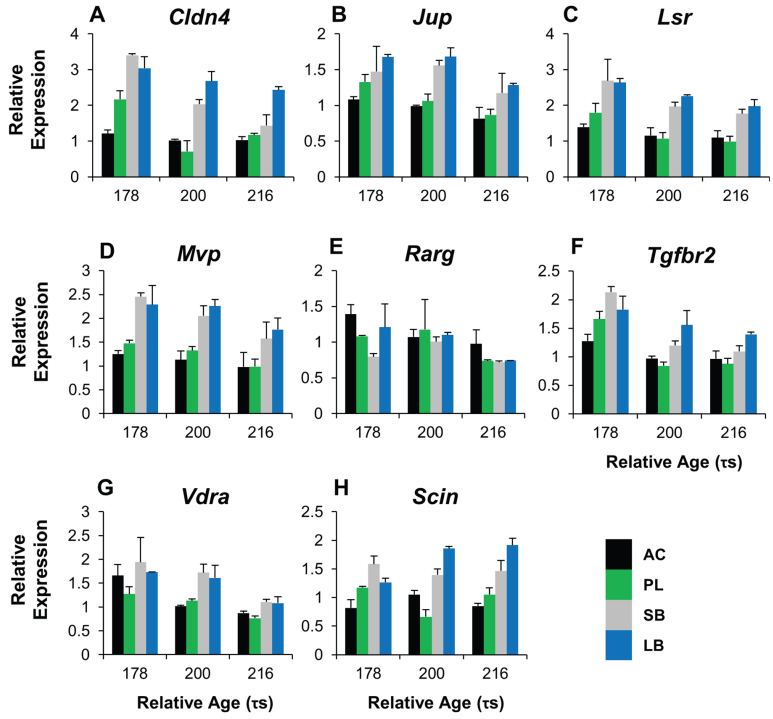
We would like to thank Baldur Kristjansson for help with genomic alignments. We are very grateful to Droplaug N. Magnusdottir, Gudbjorg Th. Orlygsdottir, Steinunn Snorradottir and Olafur Th. Magnusson at deCODE Genetics for help with the Illumina sequencing. Special thanks to Fredrik Holm for assembling Figure 1 and the three reviewers for their thoughtful comments and corrections.

Supporting Information



S1 Figure. Relative expression of *nattl* and *nattl 1–3* in tissues of adult AC-charr. Relative expression of *Natterin* (**A**) & *Natterin paralogs 1–3* (**B**–**D**) within different tissues (skin, heart, liver, gill, spleen, intestine & kidney) of adult aquaculture charr (RT-qPCR); expression plotted for different tissues, relative to heart tissue (lowest expression levels).

F1000Research 2016, 4:136 Last updated: 25 DEC 2016



S2 Figure. Relative expression of selected craniofacial candidate genes. Relative expression of 12 candidate genes with characterized craniofacial expression during zebrafish development (ZFIN website) in the head of SB, LB, PL and AC at three time points in development. In the transcriptome data all of the genes had shown higher expression in SB at 200 τs. The expression is normalized to the geometric means of two craniofacial reference genes (*ACTB* and *IF5A1*). Expression is relative to a replicate of AC morph at 200 (τs), set to one. Error bars represent standard deviation calculated from two biological replicates and each biological replicate contains homogenate of six heads.

F1000Research 2016, 4:136 Last updated: 25 DEC 2016

Supplemental Table S1 A. **qPCR primers used in this study.**

Gene	Description	Primer Sequence (5'- 3')	Product Size (bp)	PCR Efficiency	Melting Temperature (°C)	Exon Boundary
Actb	Beta Cytoskeletal Actin	F-GAAGATCAAGATCATCGCCC R-CAGACTCGTCGTACTCCTGCT	122	1.95	80.5 ± 0.7	Yes
Alp	Alkaline phosphatase	F-ACAGCATACCTCTGTGGGG R-GGTGGCATGGTTCACACG	177	1.90	85.12 ± 0.5	Yes
Cldn4	Claudin-4	F-GTGCTGTGC CATCCCAAG R-CACCACACAGGTCATCCACA	100	1.98	80.4 ± 0.6	Yes
Cgat2	Chondroitin beta-1,4-N-acetylgalactosaminyltransferase 2	F-GAGAGCCACTTTACTGAGGGG R-GAATGGACGGAAAAGAGTAACG	120	1.98	81.86 ± 0.3	Yes
Cox6b1	Cytochrome c oxidase subunit VIb isoform 1	F-GAGGGTCTACAAATCACTGTGC R-CCTGGAGTCCTACTCATACAAACAT	147	1.93	82.22 ± 0.7	Yes
Ef1α	Eukaryotic Translation Elongation Factor 1 Alpha	F-GAAGATCGGCTATAACCCTGC R-ACCTTCCATCCCTTGAACC	111	1.94	81.36 ± 0.4	Yes
If5a1	Eukaryotic Translation Initiation Factor 5A	F-GGCTTCGTGGTGCTGAAG R-CCATGTGGACCTTAGCGTG	91	1.91	80.76 ± 0.6	Yes
Jup	Junction plakoglobin	F-CACAGCAGACATACCAGGATG G R-CTGGCGATCTCTCCCCTGTT	109	1.97	81.0 ± 0.3	Yes
Krtap4–3	Keratin-associated protein 4–3	F-GCGGGACATCTACACTGCTTA R-AGAAGGCTAAAGTCTTAGTGACTATC	151	1.89	81.88 ± 0.6	Yes
Lsr	Lipolysis-stimulated lipoprotein receptor	F-TGCTGTCACTCTGGGCGA R-CCGTCTGGGCAAGGTTCA G	80	1.91	80.77 ± 0.5	Yes
Lyz	Lysozyme	F-TTCCAGATCAACAGCCGCTA R-GATCGCCACTGTGATGTCAT	111	1.94	81.87 ± 0.7	Yes
Mvp	Major vault protein	F-ACCAACTCCCAGGAGGCT R-CCTCTCCAGACGACCACG	75	1.97	78.93 ± 0.3	Yes
Nattl	Natterin-like protein	F-GTGAAAGTCACCTGCATGAATG R-CATCTCTCCTTTGTGGATACCC	104	1.98	78.81 ± 0.8	No
Nattl-1	Natterin-like protein paralog-1	F-AATCCGTGTCCTACCACAATGA R-GGTGTGTCGGTCAAAGCA	135	1.77	78.03 ± 0.1	No
Nattl-2	Natterin-like protein paralog-1	F-TGAAATVTVTGTCTCATCACAAC R-GGATCTGGTCGAGGTGGC	163	1.72	80.50 ± 0.2	No
Nattl-3	Natterin-like protein paralog-1	F-GTGACATCCGTTTCTCACCAG R-GATGTGTCGGTCAAAGCG	138	1.77	79.12 ± 0.2	No
Ndub6	NADH dehydrogenase 1 beta subcomplex subunit 6	F-TGGTGGAGTGTTCGCCTT R-CTCTCTGGGAGGTCTGGAA	171	1.89	82.40 ± 0.3	Yes
Parp6	Poly (ADP-Ribose) Polymerase Family, Member 6	F-CCGTATGAATACCGTTCCACAGG R-CACCCAGATGTTGCCGTGCTT	147	1.93	81.87 ± 0.7	Yes

Supplemental Table S1 B.

Gene	Description	Primer Sequence (5'- 3')	Product Size (bp)	PCR Efficiency	Melting Temperature (°C)	Exon Boundary
Rarg	Retinoic acid receptor gamma-A	F-AAGGCGAGCCCCTTCTTC R-TGCTCTGGGTCTCCACCG	82	1.92	78.62 ± 0.3	Yes
Scin	Scinderin/Adseverin	F-CACCTGATCCCAGACATCCAA R-CCTCACTCAACAACCTCGC	136	1.90	83.24 ± 0.7	No
Tgfbr2	TGF-beta receptor type-2	F-CTGCTCCGAGGACGAGTG R-ACCGACACCACCTGGGAG	72	1.93	79.02 ± 0.5	Yes
Ubl5	Ubiquitin-like protein 5	F-AATAAGGATGATTGAGGTGGTTTG R-ATGAGCTTCTTCAGGTCTCC	99	1.95	78.44 ± 0.3	Yes
Ub2l3	Ubiquitin-Conjugating Enzyme E2L 3	F-CGAGAAGGGACAGGTGTGTC R-ACCAACGCAATCAGGGACT	96	1.93	79.62 ± 0.3	Yes
Vdra	Vitamin D3 receptor A	F-CGTCACCAAGGCGGGTCA R-TGGAGCTTG AGTTTCTTCAGGC	81	1.93	78.12 ± 0.3	Yes

Supplemental Table S2 A. Verification of candidate polymorphisms. Primer sequences, melting temperatures and primary data.

Sequence	Position	Forward primer	Reverse primer	Tm forward	Tm reverse	Paralogs
NC_000861.1	1829	GTGCCTCAGACCCACCTAGA	TCTGTCGCCCGTACTAAGGT	60.26	59.76	No
NC_000861.1	3119	GGCCAGAGTAAACACCGAGA	CCTGGATTACTCCGGTCTGA	60.25	60.07	No
NC_000861.1	3411	GGCCAGAGTAAACACCGAGA	CCTGGATTACTCCGGTCTGA	60.25	60.07	No
NC_000861.1	8876	GACGTCCTTCACTCCTGAGC	GGGCTCATAAACTGGTCGAA	59.99	60.07	No
NC_000861.1	15240	ACCCTAAAACCGAACGATCC	TGGCTAGGAAGAGTCCGGTA	60.19	59.83	No
SS2U034121	233	CTCAACGTGCTTGACCAGTG	CCCTTACCCTCCAGGATCTC	60.5	59.89	Yes
SS2U054644	1037	AAGGACGGCCACTATGGTCT	GGGGCATAGAGTGCACAGG	60.9	61.65	Yes
SS2U054644	1188	TCAGAGATAGTGAAGAAGATGCTG	CGTACTTGATAAGACCTGTCGGTA	57.92	59.62	No
SS2U054644	1283	TCAGAGATAGTGAAGAAGATGCTG	CGTACTTGATAAGACCTGTCGGTA	57.92	59.62	No
SS2U055283	1822	TGTGTGAGGTGGTTGAGGAG	GGGTCATTGCTCCCTACAGA	59.7	60.07	No
SS2U055923	615	GTGGACCCAGAGGATGAGAA	AGAACCTGCTCCCAGTTTGA	60.05	59.84	No
SS2U058906	350	GCCAAAACCTCCACAATGAT	AACTGGCCTTCCAGATCAGA	59.8	59.8	Yes/No

Paralogs: indicates whether the PCR and sequencing yielded mixed products, indicative of paralogous genes.

Supplemental Table S2 B.

Sequence	Genome contig	Gene name	Position	Ref	Var	Freq_AC	Freq_SB	FreqP_PL	FreqP_SB	FreqP_LB
NC_000861.1	n.a.	12S ribosomal RNA	1829	G	A	0 / 53	77 / 81	0 / 6	3 / 4	1 / 8
NC_000861.1	n.a.	16S ribosomal RNA	3119	A	T	46 / 87	18 / 28	0 / 8	0 / 8	0 / 8
NC_000861.1	n.a.	16S ribosomal RNA	3411	C	T	0 / 119	26 / 33	0 / 8	5 / 8	1 / 8
NC_000861.1	n.a.	tRNA-Lys	8876	C	A	73 / 779	74 / 352	0 / 4	0 / 4	n.a.
NC_000861.1	n.a.	NADH dehydrogenase 6	15240	G	A	2 / 3608	137 / 2702	2 / 4	0 / 4	n.a.
SS2U034121	AGKD01052493.1	Eukaryotic translation initiation factor 4 gamma 2	233	C	T	0 / 95	22 / 40	2 / 4	2 / 2	n.a.
SS2U054644	AGKD01031893.1	Uroporphyrinogen decarboxylase	1037	G	A	28 / 33	0 / 56	n.a.	n.a.	n.a.
SS2U054644	AGKD01031893.1	Uroporphyrinogen decarboxylase	1188	C	T	0 / 53	19 / 25	4 / 4	4 / 4	n.a.
SS2U054644	AGKD01031893.1	Uroporphyrinogen decarboxylase	1283	G	A	4 / 60	12 / 15	4 / 4	4 / 4	n.a.
SS2U055283	AGKD01013777.1	DNA2-like helicase	1822	G	A	1 / 65	25 / 50	3 / 4	n.a.	n.a.
SS2U055923	AGKD01022586.1	Bystin	615	G	A	106 / 109	7 / 190	0 / 4	n.a.	n.a.
SS2U058906	AGKD01005918.1	Mid1-interacting protein 1-like	350	G	T	0 / 49	67 / 68	4 / 4	4 / 4	n.a.

Sequence: name of the genebank sequence or EST-contig used as reference for mapped reads.
Genome contig: name of salmon genome (ICSASG_v1) contig with best sequence match to the respective EST-contig.
Ref: Reference variant.
Var: The derived variant.
Freq_AC and Freq_SB: Frequency of variant reads as fraction of total numbers of reads mapped in Aquaculture (AC) or Small benthic (SB).
FreqP: The frequency of variant in genotyping by PCR and direct sequencing, as a fraction of total number of chromosomes sequenced.

F1000Research 2016, 4:136 Last updated: 25 DEC 2016

Supplemental Table S3. Mapping of Illumina reads to *S. salar* EST data. Numbers of reads aligning to salmon reference for each sample.

Alignment per read	SB 141	SB 163	SB 200	SB 433	AC 141	AC 163	AC 200	AC 433
0	33088778	30492314	27175901	25569628	32159386	30051365	31267710	28563169
1	6979368	11791558	11449549	11058555	11599602	11320997	11027195	10650748
2	2742358	4021683	3814418	3734404	4328402	4523686	3959198	3655786
3	2099068	2964994	2748108	2651522	3111277	3332577	2878729	2515303
4	1228292	1777846	1720902	1968251	1977738	2182392	1929818	1980420
5	914704	1317556	1284262	1434314	1471739	1679277	1447604	1426744
6	645264	946579	938290	1087959	1001350	1083025	1045157	1081063
7	425856	595785	578175	726290	657220	750523	690286	735351
8	293065	428003	424426	590100	530040	591332	527821	579860
9	206205	319401	334861	455838	296169	334264	387901	485653
10+	749074	1419362	1761275	3041930	1092980	1189781	1967857	3294222
Total reads	49372032	56075081	52230167	52318791	58225903	57039219	57129276	54968319

Supplemental Table S4. ANOVAs on qPCR data. Expression of nine genes was analyzed in whole SB- and AC-charr embryos, at two developmental timepoints (161 and 200 τs).

Gene	Term	Df	F value	p value	Significance	FDR RNA-seq
Alp	Morph	1	13.4797	0.0214	*	0.0697
	Time	1	14.9526	0.0180	*	0.0012
	M x T	1	3.9519	0.1177	.	
Cgat2	Morph	1	0.0257	0.8804	.	0.0035
	Time	1	1.5141	0.2859	.	0.3312
	M x T	1	0.1866	0.6880	.	
Cox6B1	Morph	1	0.0898	0.7793	.	0.0580
	Time	1	3.8312	0.1219	.	0.6320
	M x T	1	0.7359	0.4393	.	
Krtap4–3	Morph	1	30.0255	0.0054	**	0.0121
	Time	1	0.3902	0.5661	.	0.2784
	M x T	1	4.5225	0.1006	.	
Lyz	Morph	1	64.1566	0.0013	**	0.0406
	Time	1	1.0390	0.3657	.	0.0005
	M x T	1	1.2026	0.3344	.	
Nattl	Morph	1	8.1148	0.0465	*	7.718e-07
	Time	1	14.6659	0.0186	*	6.714e-14
	M x T	1	0.2958	0.6154	.	
Ndub6	Morph	1	0.7447	0.4368	.	0.0982
	Time	1	7.3316	0.0537	.	0.6698
	M x T	1	0.2269	0.6587	.	
Parp6	Morph	1	11.2682	0.0284	*	0.1076
	Time	1	0.7393	0.4384	.	0.3789
	M x T	1	0.2343	0.6537	.	
Ubl5	Morph	1	1.1420	0.3454	.	0.0587
	Time	1	0.2434	0.6476	.	0.0025
	M x T	1	0.3974	0.5627	.	

Significance: p > 0.05; * p < 0.05; ** p < 0.01.
FDR RNA-seq: indicates significance of Morph and Time effects in the transcriptome data.

Supplemental Table S5. ANOVAs on Natterin-like qPCR on adults. Studied were levels of *Natterin-like* and *Natterin-like Paralogs 1–3* in Arctic charr whole embryos (among SB, AC and PL morphs) and tissues from adult AC-charr.

Gene	Term	Df	F value	p value	Significance
Nattl	Morph	2	11.5515	0.0002	***
	Time	5	8.3202	3.99e-05	***
	M x T	9	4.4758	0.0007	***
Nattl1	Morph	2	19.4070	0.0001	***
	Time	3	5.9346	0.0089	**
	M x T	5	4.5761	0.0126	*
Nattl2	Morph	2	14.2921	0.0005	***
	Time	3	15.0463	0.0001	***
	M x T	5	3.2462	0.0404	*
Nattl3	Morph	2	34.4888	6.33e-06	***
	Time	3	4.4204	0.0238	*
	M x T	5	4.1843	0.0174	*
Nattl	Tissue	6	15.468	1.42e-08	***
Nattl1	Tissue	6	12.022	0.0002	***
Nattl2	Tissue	6	7.6811	0.0011	**
Nattl3	Tissue	6	46.182	8.89e-06	***

Significance: p > 0.05; * p < 0.05; ** p < 0.01.

Supplemental Table S6. Gene Ontology analyses of derived SNPs in SB-charr.

Category	Observed	In category	TERM	FDR adjusted p-value
GO:0006412	24	189	translation	4.34E-006
GO:0006396	8	32	RNA processing	0.0016
GO:0006414	6	19	translational elongation	0.0038
GO:0006313	5	20	transposition, DNA-mediated	0.0498
GO:0015074	5	21	DNA integration	0.0510
GO:0006260	6	35	DNA replication	0.0679
GO:0055114	20	285	oxidation-reduction process	0.0679

Supplemental Table S7. Predicted effect of SNP-candidates differing in frequency between charr morphs.

Effect on transcribed region	Uni_SB	Uni_AC	Rep_SB	Rep_AC
5´prime	32	19	35	24
Synonymous	232	179	176	113
Non-synonymous	112	72	81	72
3´prime	147	123	59	74

From RNA-reads that mapped to one (Uni) or more (Rep) *S. salar* ESTs.
The candidate SNPs frequencies differ more than 50% between SB and AC-charr, summarized by which morph with higher frequency of the derived allele.

References

1. Gould SJ: **Ontogeny and Phylogeny.** Harvard University Press, 1977.
Reference Source
2. Jacob F: **Evolution and tinkering.** *Science.* 1977; **196**(4295): 1161–1166.
PubMed Abstract | Publisher Full Text
3. Stern DL, Orgogozo V: **The loci of evolution: how predictable is genetic evolution?** *Evolution.* 2008; **62**(9): 2155–77.
PubMed Abstract | Publisher Full Text | Free Full Text
4. Stern DL, Orgogozo V: **Is genetic evolution predictable?** *Science.* 2009; **323**(5915): 746–751.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Rockman MV: **The QTN program and the alleles that matter for evolution: all that's gold does not glitter.** *Evolution.* 2012; **66**(1): 1–17.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Abzhanov A, Protas M, Grant BR, *et al.*: ***Bmp4* and morphological variation of beaks in Darwin's finches.** *Science.* 2004; **305**(5689): 1462–1465.
PubMed Abstract | Publisher Full Text
7. Albertson RC, Kocher TD: **Genetic and developmental basis of cichlid trophic diversity.** *Heredity (Edinb).* 2006; **97**(3): 211–221.
PubMed Abstract | Publisher Full Text
8. Cresko WA, McGuigan KL, Phillips PC, *et al.*: **Studies of threespine stickleback developmental evolution: progress and promise.** *Genetica.* 2007; **129**(1): 105–126.
PubMed Abstract | Publisher Full Text
9. Skúlason S, Smith TB: **Resource polymorphisms in vertebrates.** *Trends Ecol Evol.* 1995; **10**(9): 366–370.
PubMed Abstract | Publisher Full Text
10. Snorrason SS, Skúlason S: **Adaptive Speciation in Northern Freshwater Fishes**. In Ulf Dieckmann, Michael Doebeli, Johan A. J. Metz, and Diethard Tautz, editors, *Adaptive speciation*. Cambridge University Press, Cambridge, chapter 10, 2004; 210–228.
Publisher Full Text
11. Klemetsen A: **The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes.** *Freshwater Rev.* 2010; **3**(1): 49–74.
Publisher Full Text
12. Bernatchez L, Renaut S, Whiteley AR, *et al.*: **On the origin of species: insights from the ecological genomics of lake whitefish.** *Philos Trans R Soc Lond B Biol Sci.* 2010; **365**(1547): 1783–800.
PubMed Abstract | Publisher Full Text | Free Full Text
13. Merilä J: **Nine-spined stickleback (*Pungitius pungitius*): an emerging model for evolutionary biology research.** *Ann N Y Acad Sci.* 2013; **1289**: 18–35.
PubMed Abstract | Publisher Full Text
14. Fraser DJ, Weir LK, Bernatchez L, *et al.*: **Extent and scale of local adaptation in salmonid fishes: review and meta-analysis.** *Heredity (Edinb).* 2011; **106**(3): 404–20.
PubMed Abstract | Publisher Full Text | Free Full Text
15. Macqueen DJ, Johnston IA: **A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification.** *Proc Biol Sci.* 2014; **281**(1778): 20132881.
PubMed Abstract | Publisher Full Text | Free Full Text
16. Berthelot C, Brunet F, Chalopin D, *et al.*: **The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates.** *Nat Commun.* 2014; **5**: 3657.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Davidson WS, Koop BF, Jones SJ, *et al.*: **Sequencing the genome of the Atlantic salmon (*Salmo salar*).** *Genome Biol.* 2010; **11**(9): 403.
PubMed Abstract | Free Full Text
18. Moghadam HK, Ferguson MM, Danzmann RG: **Whole genome duplication: challenges and considerations associated with sequence orthology assignment in Salmoninae.** *J Fish Biol.* 2011; **79**(3): 561–574.
PubMed Abstract | Publisher Full Text
19. Domazet-Lošo T, Tautz D: **A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns.** *Nature.* 2010; **468**(7325): 815–8.
PubMed Abstract | Publisher Full Text
20. Bozinovic G, Sit TL, Hinton DE, *et al.*: **Gene expression throughout a vertebrate's embryogenesis.** *BMC Genomics.* 2011; **12**: 132.
PubMed Abstract | Publisher Full Text | Free Full Text
21. Giger T, Excoffier L, Day PJ, *et al.*: **Life history shapes gene expression in salmonids.** *Curr Biol.* 2006; **16**(8): R281–2.
PubMed Abstract | Publisher Full Text
22. Filteau M, Pavey SA, St-Cyr J, *et al.*: **Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish.** *Mol Biol Evol.* 2013; **30**(6): 1384–96.
PubMed Abstract | Publisher Full Text
23. Norman JD, Ferguson MM, Danzmann RG: **Transcriptomics of salinity tolerance capacity in Arctic charr (*Salvelinus alpinus*): a comparison of gene expression profiles between divergent QTL genotypes.** *Physiol Genomics.* 2014; **46**(4): 123–37.
PubMed Abstract | Publisher Full Text | Free Full Text
24. Norman JD, Ferguson MM, Danzmann RG: **An integrated transcriptomic and comparative genomic analysis of differential gene expression in Arctic charr (*Salvelinus alpinus*) following seawater exposure.** *J Exp Biol.* 2014; **217**(Pt 22): 4029–4042.
PubMed Abstract | Publisher Full Text
25. Vijay N, Poelstra JW, Künstner A, *et al.*: **Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments.** *Mol Ecol.* 2013; **22**(3): 620–34.
PubMed Abstract | Publisher Full Text
26. Noakes DLG: **Charr truth: Sympatric differentiation in *Salvelinus* species.** *Environ Biol Fishes.* 2008; **83**(1): 7–15.
Publisher Full Text
27. Ghalambor CK, Hoke KL, Ruell EW, *et al.*: **Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature.** *Nature.* 2015; **525**(7569): 372–5.
PubMed Abstract | Publisher Full Text
28. Adams CE, Fraser D, Wilson AJ, *et al.*: **Patterns of phenotypic and genetic variability show hidden diversity in Scottish Arctic charr.** *Ecol Freshwater Fish.* 2007; **16**(1): 78–86.
Publisher Full Text
29. Kristjánsson BK, Malmquist HJ, Ingimarsson F, *et al.*: **Relationships between lake ecology and morphological characters in Icelandic Arctic charr, *Salvelinus alpinus*.** *Biol J Linn Soc.* 2011; **103**(4): 761–771.
Publisher Full Text
30. Woods PJ, Skúlason S, Snorrason SS, *et al.*: **Intraspecific diversity in Arctic charr, *Salvelinus alpinus*, in Iceland: II. Which environmental factors influence resource polymorphism in lakes?** *Evolutionary Ecol Res.* 2012; **14**(8): 993–1013.
Reference Source
31. Kristjánsson BK, Skúlason S, Snorrason SS, *et al.*: **Fine-scale parallel patterns in diversity of small benthic Arctic charr (*Salvelinus alpinus*) in relation to the ecology of lava/groundwater habitats.** *Ecol Evol.* 2012; **2**(6): 1099–112.
PubMed Abstract | Publisher Full Text | Free Full Text
32. Skúlason S, Noakes DL, Snorrason SS: **Ontogeny of trophic morphology in four sympatric morphs of Arctic charr *Salvelinus alpinus* in Thingvallavatn, Iceland.** *Biol J Linn Soc.* 1989; **38**(3): 281–301.
Publisher Full Text
33. Skúlason S, Snorrason SS, Ota D, *et al.*: **Genetically based differences in foraging behaviour among sympatric morphs of Arctic charr (Pisces: Salmonidae).** *Anim Behav.* 1993; **45**(6): 1179–1192.
Publisher Full Text
34. Skúlason S, Snorrason SS, Noakes DLG, *et al.*: **Genetic basis of life history variations among sympatric morphs of Arctic char *Salvelinus alpinus*.** *Can J Fish Aquat Sci.* 1996; **53**(8): 1807–1813.
Publisher Full Text
35. Parsons KJ, Skúlason S, Ferguson M: **Morphological variation over ontogeny and environments in resource polymorphic Arctic charr (*Salvelinus alpinus*).** *Evol Dev.* 2010; **12**(3): 246–257.
PubMed Abstract | Publisher Full Text
36. Parsons KJ, Sheets HD, Skúlason S, *et al.*: **Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (*Salvelinus alpinus*).** *J Evol Biol.* 2011; **24**(8): 1640–1652.
PubMed Abstract | Publisher Full Text
37. Sandlund TO, Gunnarsson K, Jónasson PM, *et al.*: **The Arctic charr *Salvelinus alpinus* in Thingvallavatn.** *Oikos.* 1992; **64**(1/2): 305–351.
Publisher Full Text
38. Snorrason SS, Skúlason S, Jonsson B, *et al.*: **Trophic specialization in Arctic charr *Salvelinus alpinus* (Pisces; Salmonidae): morphological divergence and ontogenetic niche shifts.** *Biol J Linn Soc.* 1994; **52**(1): 1–18.
Publisher Full Text
39. Magnusson KP, Ferguson MM: **Genetic analysis of four sympatric morphs of Arctic charr, *Salvelinus alpinus*, from Thingvallavatn, Iceland.** *Environ Biol Fishes.* 1987; **20**(1): 67–73.
Publisher Full Text
40. Danzmann RG, Ferguson MM, Skulason S, *et al.*: **Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn, Iceland.** *J Fish Biol.* 1991; **39**(5): 649–659.
Publisher Full Text
41. Pálsson S, Árnason E: **Sequence variation for cytochrome b genes of three salmonid species from Iceland.** *Aquaculture.* 1994; **128**(1–2): 29–39.
Publisher Full Text
42. Volpe JP, Ferguson MM: **Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland.** *Mol Ecol.* 1996; **5**(6): 763–72.
PubMed Abstract | Publisher Full Text
43. Wilson AJ, Gíslason D, Skúlason S, *et al.*: **Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales.** *Mol Ecol.* 2004; **13**(5): 1129–42.
PubMed Abstract | Publisher Full Text
44. Kapralova KH, Morrissey MB, Kristjánsson BK, *et al.*: **Evolution of adaptive**

diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity (Edinb).* 2011; **106**(3): 472–487.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Kapralova KH, Gudbrandsson J, Reynisdottir S, *et al.*: **Differentiation at the *MHCIIα* and *Cath2* loci in sympatric *Salvelinus alpinus* resource morphs in Lake Thingvallavatn.** *PLoS One.* 2013; **8**(7): e69402.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Saemundsson K: **Geology of the Thingvallavatn Area.** *Oikos.* 1992; **64**(1/2): 40–68.
Publisher Full Text

47. Macqueen DJ, Kristjánsson BK, Paxton CG, *et al.*: **The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression.** *Mol Ecol.* 2011; **20**(15): 3167–84.
PubMed Abstract | Publisher Full Text

48. Doiron S, Bernatchez L, Blier PU: **A comparative mitogenomic analysis of the potential adaptive value of Arctic charr mtDNA introgression in brook charr populations (*Salvelinus fontinalis* Mitchill).** *Mol Biol Evol.* 2002; **19**(11): 1902–9.
PubMed Abstract | Publisher Full Text

49. Miller W, Schuster SC, Welch AJ, *et al.*: **Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change.** *Proc Natl Acad Sci U S A.* 2012; **109**(36): E2382–90.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Svavarsson E: **Árangur í kynbótum á bleikju og næstu skref [reference in Icelandic].** In *Fræðaþing landbúnaðarsins 4.* 2007; 121–125.
Reference Source

51. Ahi EP, Guðbrandsson J, Kapralova KH, *et al.*: **Validation of reference genes for expression studies during craniofacial development in arctic charr.** *PLoS One.* 2013; **8**(6): e66389.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Ahi EP, Kapralova KH, Pálsson A, *et al.*: **Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr.** *Evodevo.* 2014; **5**(1): 40.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Ahi EP, Steinhäuser SS, Pálsson A, *et al.*: **Differential expression of the aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr.** *Evodevo.* 2015; **6**(1): 27.
PubMed Abstract | Publisher Full Text | Free Full Text

54. Snorrason SS, Skúlason S, Sandlund OT, *et al.*: **Shape polymorphism in Arctic charr, *Salvelinus alpinus*.** *Physiol Ecol Japan.* 1989; **1**: 393–404.
Reference Source

55. Gorodilov YN: **Description of the early ontogeny of the Atlantic salmon, *Salmo salar*, with a novel system of interval (state) identification.** *Environ Biol Fishes.* 1996; **47**(2): 109–127.
Publisher Full Text

56. Kapralova KH, Franzdóttir SR, Jónsson H, *et al.*: **Patterns of miRNA expression in Arctic charr development.** *PLoS One.* 2014; **9**(8): e106084.
PubMed Abstract | Publisher Full Text | Free Full Text

57. Di Génova A, Aravena A, Zapata L, *et al.*: **SalmonDB: a bioinformatics resource for *Salmo salar* and *Oncorhynchus mykiss*.** *Database (Oxford).* 2011; **2011**: bar050.
PubMed Abstract | Publisher Full Text | Free Full Text

58. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics.* 2011; **12**(1): 323.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics.* 2010; **26**(1): 139–40.
PubMed Abstract | Publisher Full Text | Free Full Text

60. R Core Team: **R: A Language and Environment for Statistical Computing.** 2014.

61. Young MD, Wakefield MJ, Smyth GK, *et al.*: **Gene ontology analysis for RNA-seq: accounting for selection bias.** *Genome Biol.* 2010; **11**(2): R14.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Untergasser A, Cutcutache I, Koressaar T, *et al.*: **Primer3--new capabilities and interfaces.** *Nucleic Acids Res.* 2012; **40**(15): e115.
PubMed Abstract | Publisher Full Text | Free Full Text

63. Bustin SA, Benes V, Garson JA, *et al.*: **The MIQE guidelines: *M*inimum *I*nformation for publication of *Q*uantitative real-time PCR *E*xperiments.** *Clin Chem.* 2009; **55**(4): 611–22.
PubMed Abstract | Publisher Full Text

64. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method.** *Methods.* 2001; **25**(4): 402–8.
PubMed Abstract | Publisher Full Text

65. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics.* 2009; **25**(14): 1754–60.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Koboldt DC, Zhang Q, Larson DE, *et al.*: **VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.** *Genome Res.* 2012; **22**(3): 568–76.
PubMed Abstract | Publisher Full Text | Free Full Text

67. Piskol R, Ramaswami G, Li JB: **Reliable identification of genomic variants from RNA-seq data.** *Am J Hum Genet.* 2013; **93**(4): 641–651.
PubMed Abstract | Publisher Full Text | Free Full Text

68. Larkin MA, Blackshields G, Brown NP, *et al.*: **Clustal W and Clustal X version 2.0.** *Bioinformatics.* 2007; **23**(21): 2947–2948.
PubMed Abstract | Publisher Full Text

69. Nicholas KB, Nicholas HB, Deerfield DW: **GeneDoc: analysis and visualization of genetic variation.** *EMBNEW.NEWS.* 1997; **4**(14).
Reference Source

70. Magalhães GS, Lopes-Ferreira M, Junqueira-de Azevedo IL, *et al.*: **Natterins, a new class of proteins with kininogenase activity characterized from *Thalassophryne nattereri* fish venom.** *Biochimie.* 2005; **87**(8): 687–99.
PubMed Abstract | Publisher Full Text

71. Magalhães GS, Junqueira-de-Azevedo IL, Lopes-Ferreira M, *et al.*: **Transcriptome analysis of expressed sequence tags from the venom glands of the fish *Thalassophryne nattereri*.** *Biochimie.* 2006; **88**(6): 693–9.
PubMed Abstract | Publisher Full Text

72. Sprague J, Bayraktaroglu L, Clements D, *et al.*: **The Zebrafish Information Network: the zebrafish model organism database.** *Nucleic Acids Res.* 2006; **34**(Database issue): D581–D585.
PubMed Abstract | Publisher Full Text | Free Full Text

73. Eiriksson GM, Skúlason S, Snorrason SS: **Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn, Iceland.** *J Fish Biol.* 1999; **55**(sa): 175–185.
Publisher Full Text

74. Kapralova KH, Jónsson ZO, Pálsson A, *et al.*: **Bones in motion: Ontogeny of craniofacial development in sympatric arctic charr morphs.** *Dev Dyn.* 2015; **244**(9): 1168–1178.
PubMed Abstract | Publisher Full Text

75. Jonsson B, Skúlason S, Snorrason SS, *et al.*: **Life History Variation of Polymorphic Arctic Charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland.** *Can J Fish Aquat Sci.* 1988; **45**(9): 1537–1547.
Publisher Full Text

76. Palsson A, Wesolowska N, Reynisdóttir S, *et al.*: **Naturally occurring deletions of hunchback binding sites in the *even-skipped* stripe 3+7 enhancer.** *PLoS One.* 2014; **9**(5): e91924.
PubMed Abstract | Publisher Full Text | Free Full Text

77. Zou C, Lehti-Shiu MD, Thomashow M, *et al.*: **Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*.** *PLoS Genet.* 2009; **5**(7): e1000581.
PubMed Abstract | Publisher Full Text | Free Full Text

78. Jantzen SG, Sanderson DS, von Schalburg KR, *et al.*: **A 44K microarray dataset of the changing transcriptome in developing Atlantic salmon (*Salmo salar* L.).** *BMC Res Notes.* 2011; **4**: 88.
PubMed Abstract | Publisher Full Text | Free Full Text

79. Drivenes Ø, Taranger GL, Edvardsen RB: **Gene expression profiling of Atlantic cod (*Gadus morhua*) embryogenesis using microarray.** *Mar Biotechnol (NY).* 2012; **14**(2): 167–76.
PubMed Abstract | Publisher Full Text

80. Bougas B, Audet C, Bernatchez L: **The influence of parental effects on transcriptomic landscape during early development in brook charr (*Salvelinus fontinalis*, Mitchill).** *Heredity (Edinb).* 2013; **110**(5): 484–491.
PubMed Abstract | Publisher Full Text | Free Full Text

81. Piasecka B, Lichocki P, Moretti S, *et al.*: **The hourglass and the early conservation models--co-existing patterns of developmental constraints in vertebrates.** *PLoS Genet.* 2013; **9**(4): e1003476.
PubMed Abstract | Publisher Full Text | Free Full Text

82. Fleming A: **On a Remarkable Bacteriolytic Element Found in Tissues and Secretions.** *Proc R Soc Lond B.* 1922; **93**(653): 306–317.
Publisher Full Text

83. Chipman DM, Sharon N: **Mechanism of lysozyme action.** *Science.* 1969; **165**(3892): 454–465.
PubMed Abstract | Publisher Full Text

84. Subramanian S, MacKinnon SL, Ross NW: **A comparative study on innate immune parameters in the epidermal mucus of various fish species.** *Comp Biochem Physiol B Biochem Mol Biol.* 2007; **148**(3): 256–63.
PubMed Abstract | Publisher Full Text

85. Magnadóttir B: **Innate immunity of fish (overview).** *Fish Shellfish Immunol.* 2006; **20**(2): 137–51.
PubMed Abstract | Publisher Full Text

86. Magnadottir B: **Immunological control of fish diseases.** *Mar Biotechnol (NY).* 2010; **12**(4): 361–79.
PubMed Abstract | Publisher Full Text

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

88. Chai Y, Ito Y, Han J: **TGF-beta signaling and its functional significance in regulating the fate of cranial neural crest cells.** *Crit Rev Oral Biol Med.* 2003; **14**(2): 78–88.
PubMed Abstract | Publisher Full Text

89. Puga A, Tomlinson CR, Xia Y: **Ah receptor signals cross-talk with multiple developmental pathways.** *Biochem Pharmacol.* 2005; **69**(2): 199–207.
PubMed Abstract | Publisher Full Text

90. Goodale BC, La Du JK, Bisson WH, *et al.*: **AHR2 mutant reveals functional diversity of aryl hydrocarbon receptors in zebrafish.** *PLoS One.* 2012; **7**(1):

e29346.
PubMed Abstract | Publisher Full Text | Free Full Text

91. Nurminsky D, Magee C, Faverman L, *et al.*: **Regulation of chondrocyte differentiation by actin-severing protein adseverin.** *Dev Biol.* 2007; **302**(2): 427–37.
PubMed Abstract | Publisher Full Text | Free Full Text

92. Vieira FA, Thorne MA, Stueber K, *et al.*: **Comparative analysis of a teleost skeleton transcriptome provides insight into its regulation.** *Gen Comp Endocrinol.* 2013; **191**: 45–58.
PubMed Abstract | Publisher Full Text

93. Furuse M, Oda Y, Higashi T, *et al.*: **Lipolysis-stimulated lipoprotein receptor: a novel membrane protein of tricellular tight junctions.** *Ann N Y Acad Sci.* 2012; **1257**: 54–8.
PubMed Abstract | Publisher Full Text

94. Jazag A, Ijichi H, Kanai F, *et al.*: **Smad4 silencing in pancreatic cancer cell lines using stable RNA interference and gene expression profiles induced by transforming growth factor-beta.** *Oncogene.* 2005; **24**(4): 662–71.
PubMed Abstract | Publisher Full Text

95. Planchart A, Mattingly CJ: **2,3,7,8-Tetrachlorodibenzo-*p*-dioxin upregulates *FoxQ1b* in zebrafish jaw primordium.** *Chem Res Toxicol.* 2010; **23**(3): 480–7.
PubMed Abstract | Publisher Full Text | Free Full Text

96. Hering NA, Andres S, Fromm A, *et al.*: **Transforming growth factor-*β*, a whey protein component, strengthens the intestinal barrier by upregulating claudin-4 in HT-29/B6 cells.** *J Nutr.* 2011; **141**(5): 783–9.
PubMed Abstract | Publisher Full Text

97. Ito Y, Yeo JY, Chytil A, *et al.*: **Conditional inactivation of *Tgfbr2* in cranial neural crest causes cleft palate and calvaria defects.** *Development.* 2003; **130**(21): 5269–80.
PubMed Abstract | Publisher Full Text

98. Kedersha NL, Miquel MC, Bittner D, *et al.*: **Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes.** *J Cell Biol.* 1990; **110**(4): 895–901.
PubMed Abstract | Publisher Full Text | Free Full Text

99. Berger W, Steiner E, Grusch M, *et al.*: **Vaults and the major vault protein: novel roles in signal pathway regulation and immunity.** *Cell Mol Life Sci.* 2009; **66**(1): 43–61.
PubMed Abstract | Publisher Full Text

100. van Driel M, Pols HA, van Leeuwen JP: **Osteoblast differentiation and control by vitamin D and vitamin D metabolites.** *Curr Pharm Des.* 2004; **10**(21): 2535–55.
PubMed Abstract | Publisher Full Text

101. Pulcini D, Wheeler PA, Cataudella S, *et al.*: **Domestication shapes morphology in rainbow trout *Oncorhynchus mykiss*.** *J Fish Biol.* 2013; **82**(2): 390–407.
PubMed Abstract | Publisher Full Text

102. Völkers M, Rohde D, Goodman C, *et al.*: **S100A1: a regulator of striated muscle sarcoplasmic reticulum Ca^{2+} handling, sarcomeric, and mitochondrial function.** *J Biomed Biotechnol.* 2010; **2010**: 178614.
PubMed Abstract | Publisher Full Text | Free Full Text

103. McBride HM, Neuspiel M, Wasiak S: **Mitochondria: more than just a powerhouse.** *Curr Biol.* 2006; **16**(14): R551–60.
PubMed Abstract | Publisher Full Text

104. O'Brien KM, Mueller IA: **The unique mitochondrial form and function of Antarctic channichthyid icefishes.** *Integr Comp Biol.* 2010; **50**(6): 993–1008.
PubMed Abstract | Publisher Full Text

105. William J, Ballard O, Pichaud N: **Mitochondrial DNA: more than an evolutionary bystander.** *Functional Ecol.* 2014; **28**(1): 218–231.
Publisher Full Text

106. Teacher AG, André C, Merilä J, *et al.*: **Whole mitochondrial genome scan for population structure and selection in the Atlantic herring.** *BMC Evol Biol.* 2012; **12**: 248.
PubMed Abstract | Publisher Full Text | Free Full Text

107. Palsson A, Gibson G: **Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*.** *Genetics.* 2004; **167**(3): 1187–1198.
PubMed Abstract | Publisher Full Text | Free Full Text

108. Palsson A, Dodgson J, Dworkin I, *et al.*: **Tests for the replication of an association between *Egfr* and natural variation in *Drosophila melanogaster* wing morphology.** *BMC Genet.* 2005; **6**: 44.
PubMed Abstract | Publisher Full Text | Free Full Text

109. Gudbrandsson J, Ahi E, Franzdottir S, *et al.*: **Dataset 1 in: The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs.** *F1000Research.* 2015.
Data Source

110. Gudbrandsson J, Ahi E, Franzdottir S, *et al.*: **Dataset 2 in: The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs.** *F1000Research.* 2015.
Data Source

111. Gudbrandsson J, Ahi E, Franzdottir S, *et al.*: **Dataset 3 in: The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs.** *F1000Research.* 2015.
Data Source

F1000Research F1000Research 2016, 4:136 Last updated: 25 DEC 2016

Open Peer Review

Current Referee Status: ✔ ? ✔

Version 3

Referee Report 20 December 2016

doi:10.5256/f1000research.10982.r18642

✔ **Örjan Östman**
Department of Aquatic Resources, Swedish University of Agricultural Sciences, Uppsala, Sweden

The authors have addressed all my comments and incorporated them into the manuscript. So my opinion of the study is it will be an important contribution for further work, and thus, I approve this version of the manuscript.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 2

Referee Report 30 August 2016

doi:10.5256/f1000research.9044.r15587

? **Örjan Östman**
Department of Aquatic Resources, Swedish University of Agricultural Sciences, Uppsala, Sweden

The study by Gudbrandsson et al. reports a thorough analysis of differences in the transcriptome between different 'morphs' or 'populations' of artic charr. More specifically they have studied the transcriptome of eggs and larvae from a natural population of small benthic charr (SB) and Icelandic aquaculture charr (AC), which is fast growing and have a 'limnic-like' morphology. They find a list of potential candidate genes involved in the ecological differentiation of artic charr (and during the embryonic development). In addition they studied the transcriptome from different tissues of adult AC-charr. From the transcriptome of these populations two populations they developed 12 SNP-markers applied to other sympathric (Lake Thingvallavatn) wild morphs to study if these genes differed between other morphs. Finally they also study mtDNA expression between morphs to find they mainly differ between the benthic morphs and a limnic.

The search for genes involved in the ecological divergence of species is an important topic that has exploded the last decade with the new generation of sequencing. I find this study to be an important contribution because of the study system with artic charr is an example of relative recent and rapid

divergence into many different morphs/ecological, and the extensive and thorough investigation of the differences in the transcriptome between morphs.

However, I think the authors try to stretch their conclusions a bit too far. The study is great as a base for further research in the topic, which I guess is in the pipeline. The use of cultivated charr make sense for comparing the most extreme morphs. But to me it does not make sense for making conclusions about genes involved in the ecological niche differentiation in natural populations, which is the motivation of the study in the introduction and brought up in the discussion). The cultivated population has been selected for fast body growth and they are not from Lake Thingvallavatn and little can therefore be said about the genetics of the ecological differentiation of sympatric species. Not very surprising genes related to metabolism seemed upregulated in AC and immunogens upregulated in SB. What does that actually tells us about the genetics of ecological differentiation of natural populations??

Although the aims on p. 4 feels valid, they are not contingent with the previous text in the introduction. Thus, I suggest that the much of the earlier part of the introduction is rewritten to actually address the differences in gene expression between a cultivated morph and its extreme opposite small benthic artic charr.

As far as I understand it is only egg that are kept in the same environment, but the parents have been raised in different environments and transgenerational plasticity cannot be ruled out. This is not a major criticism (the ideal case would be to have had them in lines in a common environment of course) but needs to be addressed in the text.

The 'Nattl' paralogs provide an interesting case where the expression of different paralogs has been studied. But again, are the result difficult to interpret from an ecological niche differentiation perspective. Often is the natural small limnic morph (PL) in between SB and AC (Fig. 5A), but for Nattl1 and Nattl2 AC and SB seem most similar? The connection to the original question is weak and the authors do not conclude more than "…it is not divergence in the expression of one paralog that explains the general nattl expression disparity in the transcriptome." Fair enough, but that is more about the genetic architecture than the genetics of ecological divergence.

In the validation of the transcriptome differences with qPCR of 9 genes/paralogs only 5 was still significant. What conclusions should one make out of that, that around half of the 296 paralogs differing between SB and AC are false detection (despite FDR < 5%). I support the use of qPCR but please comment on the implication of this.

I think Fig. 6 should be converted into a bar-graph plot instead (this feels more like a table).

To conclude, I think this study is a great contribution for suggesting putative differences in the transcriptome of a fish species. However, the importance for understanding ecological differentiation of sympatric species it is, however, so far limited as it that would require more using natural morphs, replicated populations, back-crosses, investigation of plasticity and how reproductive isolation is maintained etc., which is likely to come. But until that, I suggest the this text should mainly focus on the differences between SB and AC arctic charr, and not try to squeeze in everything in one paper.

Minor comments:
In the equation on p. 6 I guess M & T is 'Morph' and 'Time'? If so spell out or use M & T consistently.

Note that on p. 8 the qPCR of 8 paralogs in embryonic heads is metioned but the results do not come until

"Expression differences in the developing heads of benthic and limnetic charr morphs".

Use "." instead of "," as decimal sign in Fig. 4.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 11 Nov 2016

Arnar Palsson, University of Iceland, Iceland

The study by Gudbrandsson et al. reports a thorough analysis of differences in the transcriptome between different 'morphs' or 'populations' of artic charr. More specifically they have studied the transcriptome of eggs and larvae from a natural population of small benthic charr (SB) and Icelandic aquaculture charr (AC), which is fast growing and have a 'limnic-like' morphology. They find a list of potential candidate genes involved in the ecological differentiation of artic charr (and during the embryonic development). In addition they studied the transcriptome from different tissues of adult AC-charr. From the transcriptome of these populations two populations they developed 12 SNP-markers applied to other sympathric (Lake Thingvallavatn) wild morphs to study if these genes differed between other morphs. Finally they also study mtDNA expression between morphs to find they mainly differ between the benthic morphs and a limnic.

The search for genes involved in the ecological divergence of species is an important topic that has exploded the last decade with the new generation of sequencing. I find this study to be an important contribution because of the study system with artic charr is an example of relative recent and rapid divergence into many different morphs/ecological, and the extensive and thorough investigation of the differences in the transcriptome between morphs.

However, I think the authors try to stretch their conclusions a bit too far. The study is great as a base for further research in the topic, which I guess is in the pipeline. The use of cultivated charr make sense for comparing the most extreme morphs. But to me it does not make sense for making conclusions about genes involved in the ecological niche differentiation in natural populations, which is the motivation of the study in the introduction and brought up in the discussion). The cultivated population has been selected for fast body growth and they are not from Lake Thingvallavatn and little can therefore be said about the genetics of the ecological differentiation of sympatric species. Not very surprising genes related to metabolism seemed upregulated in AC and immunogens upregulated in SB. What does that actually tells us about the genetics of ecological differentiation of natural populations??

Although the aims on p. 4 feels valid, they are not contingent with the previous text in the introduction. Thus, I suggest that the much of the earlier part of the introduction is rewritten to actually address the differences in gene expression between a cultivated morph and its extreme opposite small benthic arctic charr.

Reply: We thank the reviewer for his comments, we have now carefully reviewed the introduction,

results, discussion and conclusions to address his concerns. We provide below an excerpt of most of the changes made.

"However, I think the authors try to stretch their conclusions a bit too far"

Reply: We acknowledge that the discussion in particular, worded the conclusions about ecological effects too strongly and have toned those down, for example:

"The charr developmental transcriptome provides a starting point to investigate the molecular systems that associate with divergence among the highly polymorphic and rapidly evolving Arctic charr in Iceland."

"The embryos were reared in a common garden setting, which minimizes the impact of environmental factors, as we are interested in genes showing expression differences between the two morphs. Those genes might implicate pathways involved in the ecological divergence among charr populations and of course adaptation of the AC charr during breeding [50] "

The use of cultivated charr make sense for comparing the most extreme morphs. But to me it does not make sense for making conclusions about genes involved in the ecological niche differentiation in natural populations, which is the motivation of the study in the introduction and brought up in the discussion) …

Reply: We agree with the reviewer's remarks, the flow of the introduction and partly the discussion was not optimal, with the interpretations overreaching in some places. We have now restructured the introduction, added a separate section on Aquaculture charr, and improved the description of the results. For each result section we tried to make clear where the data support conclusions about the difference between SB and AC only or more general about benthic - limnetic differences (like where the follow up qPCR or SNP validation involved also samples of Lake Thingvallavatn morphs). Also, throughout the manuscript we also brought the contrast of AC and SB charr into sharper focus, and the fact that many patterns can reflect the AC charr domestication, for example:

"The aim of this study was to find expression and genetic differences separating the small benthic morph in Lake Thingvallavatn and aquaculture charr, with the long term objective being to reveal the genetic and molecular systems that associate with benthic morphology in charr. The transcriptome reflects the biology of these two morphs, their different histories and ecology. AC-charr will also be shaped by domestication, which may explain for instance the higher expression of metabolic genes in AC-charr."

"It would be interesting to study further the expression of these and other immunological genes implicated in this transcriptome, in natural charr populations and juveniles challenged with pathogens or in families of Aquaculture charr breed for pathogen resistance."

"The results suggest divergence (adaptive or neutral) in mitochondrial function due to the domestication of aquaculture charr and/or adaptation of the small benthic charr to its habitat. Increase in mitochondrial function in AC charr embryos could reflect higher basal metabolic rate in this aquaculture stock. Alternatively, lower metabolic rate in the SB charr would also be curious in the context of their ecology. Clearly further work is needed to map out the functional differences of

mitochondrial related genes in AC charr, more SB populations and hopefully anadromous charr morphs (representing the ancestral state)."

As far as I understand it is only egg that are kept in the same environment, but the parents have been raised in different environments and transgenerational plasticity cannot be ruled out. This is not a major criticism (the ideal case would be to have had them in lines in a common environment of course) but needs to be addressed in the text.

Reply: We add a sentence about transgenerational plasticity in the discussion. "We raised the embryos in a common garden, but their parents were wild so parental environments and transgenerational plasticity may also have contributed."

The 'Nattl' paralogs provide an interesting case where the expression of different paralogs has been studied. But again, are the result difficult to interpret from an ecological niche differentiation perspective. Often is the natural small limnic morph (PL) in between SB and AC (Fig. 5A), but for Nattl1 and Nattl2 AC and SB seem most similar? The connection to the original question is weak and the authors do not conclude more than "…it is not divergence in the expression of one paralog that explains the general nattl expression disparity in the transcriptome." Fair enough, but that is more about the genetic architecture than the genetics of ecological divergence.

Reply: This issue is of interest to us. We wanted to work more on the Nattl genes and confess that the paragraph did not summarize the data properly. We paraphrased it, toning down the interpretation and added a more forward looking statement – about how to build on this dataset "It would be interesting to study further the expression of these and other immunological genes implicated in this transcriptome, in natural charr populations and juveniles challenged with pathogens."

In the validation of the transcriptome differences with qPCR of 9 genes/paralogs only 5 was still significant. What conclusions should one make out of that, that around half of the 296 paralogs differing between SB and AC are false detection (despite FDR < 5%). I support the use of qPCR but please comment on the implication of this.

Reply: The transcriptome was done on 4 timepoints, but only 2 of those were used for the qPCR verification. Thus we expect incomplete correlation, because of the contribution of the earliest or latest (in particular) timepoints. We explain this clarification to the results on qPCR verification (Table 3) and added a caveat, "Thus this transcriptome should not be taken at face value, because substantial fraction of signals were false positives. "

I think Fig. 6 should be converted into a bar-graph plot instead (this feels more like a table).

Reply: We opted for this heatmap-table representation, as we feel it emphasizes the benthic – limnetic separation most clearly. The other option we explored was indeed a bar-graph (Supplemental figure 2), with extra lines and stars indicating the significance of the post-hoc tests, but felt the heatmap-table captured best the pattern.

To conclude, I think this study is a great contribution for suggesting putative differences in the transcriptome of a fish species. However, the importance for understanding ecological differentiation of sympatric species it is, however, so far limited as it that would require more using natural morphs, replicated populations, back-crosses, investigation of plasticity and how reproductive isolation is maintained etc., which is likely to come. But until that, I suggest the this text should mainly focus on the differences between SB and AC arctic charr, and not try to squeeze in everything in one paper.

Reply: We acknowledge that with respect to our long term research goals, this can be viewed as a pilot study as the contrast is between the SB and AC charr. In this version we tried to focus more on describing the differences between SB and AC charr, and highlight also results that may reflect the AC-charr biology (see some sentences listed above, and more in the manuscript). By validating differential gene expression and some of the SNPs also on samples from more wild populations, the study also revealed interesting candidates for follow up studies addressing the long term objectives of the group.

Minor comments:
In the equation on p. 6 I guess M & T is 'Morph' and 'Time'? If so spell out or use M & T consistently.

Reply: We opted for spelling out Morph and Time – its more transparent.

Note that on p. 8 the qPCR of 8 paralogs in embryonic heads is mentioned but the results do not come until "Expression differences in the developing heads of benthic and limnetic charr morphs".

Reply: Good point, now the next section is referenced "… and 8 in embryonic heads (see next section)..."

Use "." instead of "," as decimal sign in Fig. 4.

Reply: Fixed.

Competing Interests: No competing interests were disclosed.

Referee Report 25 May 2016

doi:10.5256/f1000research.9044.r13702

Daniel Macqueen
Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, UK

Second review of Gudbrandsson *et al*. "*The developmental transcriptome of contrasting Arctic charr (Salvelinus alpinus) morphs*".

Overview:

The authors have addressed the comments made by myself and Anne Dalziel. They have incorporated a range of associated changes into version 2 of their paper. Readers will find these changes, along with several clarifications provided in the published response to reviewers section, to facilitate transparent interpretation of this large and diverse study, including its strengths and caveats. My overall opinion of the study remains unchanged – it is interesting and reports findings of merit that will be followed up on in future work. **I am thus happy to approve version 2 of the paper**.

I did spot a few typos or grammatical issues that the authors might address and had some final comments that might be addressed – all of a minor nature and easy to address.

- Abstract – "*energy metabolism and blood coagulation genes*" remove "*genes*"
- Abstract - "*Comparison of single nucleotide polymorphism (SNP) frequencies reveals*" change "*reveals*" to "*revealed*" (for accurate use of tense)
- Introduction – "*Examples of such species complexes are provided finches of the Galapagos island*" should be "*Examples of such species complexes are provided by finches of the Galapagos island*"
- Introduction: "*Thus we were quite keen to apply RNA-sequencing to analyze ecomorphs in our study system, Arctic charr*". The authors should add the Latin name for charr here, rather than in the next paragraph.
- Introduction: "*The family is estimated to be between 88–103 million years old[21,22]. A whole genome duplication event occurred before the radiation of the salmonid family[21–24] which has provided time for divergence of ohnologous genes (paralogous genes originated by whole genome duplication event*"

It would be simpler to just state that the common ancestor to salmonids experienced a whole genome duplication 88–103 million years ago. The actual age of the salmonid family depends on whether one considers the (extinct) direct ancestors to salmonids that didn't experience genome duplication to be salmonids.

- Introduction: "*Furthermore, recent estimates from the rainbow trout (Oncorhynchus mykiss) genome suggest that ohnologous genes are lost at a rate of about 170 genes per million years, and that on the order of 4500 were retained in rainbow trout[22]*"

This information is inaccurate. Firstly, based on the paper cited (Berthalot *et al.* 2014), this information should state that around 4,500 *pairs* of ohnologous genes were retained from Ss4R (i.e. around 9,000 separate genes). More importantly, without going into detail, the stated data represents a non-comprehensive fraction of the genome. I suggest the authors update this part of the text with accurate estimates, since the number of retained Ss4R ohnologue pairs in much larger than what is stated. The authors might also draw in more comprehensive data from the recent publication of the Atlantic salmon genome (Lien et al. Nature, 533, 200–205)[1]. The simplest way to present the information is to state that around half of the original Ss4R ohnologue pairs are still functionally retained (both stated papers are in agreement about that).

- Figure 1: It would be easier for the reader to link the text and images if the authors updated with 'a', 'b', 'c' and 'd' panels for each of the different charr morphs.
- Introduction: "*In this study, we compare SB-charr from*" should be "*In this study, we compared SB-charr from*" (again, it is correct here to use past tense – the authors should check the rest of the manuscript for similar tense issues).
- Figure 2: Minor comments – the text "*Map on salmon genes*" is vague and open to several interpretations. Better: "*Map on Atlantic salmon expressed sequence tags*"?

References

1. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, Di Genova A, Samy JK, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Inge Våge D, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles DA, Vidal R, Iturra P, Jones SJ, Jonassen I, Maass A, Omholt SW, Davidson WS: The Atlantic salmon genome provides insights into rediploidization.*Nature*. 2016; **533** (7602): 200-5 PubMed Abstract | Publisher Full Text

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 11 Nov 2016

Arnar Palsson, University of Iceland, Iceland

Overview:
The authors have addressed the comments made by myself and Anne Dalziel. They have incorporated a range of associated changes into version 2 of their paper. Readers will find these changes, along with several clarifications provided in the published response to reviewers section, to facilitate transparent interpretation of this large and diverse study, including its strengths and caveats. My overall opinion of the study remains unchanged – it is interesting and reports findings of merit that will be followed up on in future work. **I am thus happy to approve version 2 of the paper**.

I did spot a few typos or grammatical issues that the authors might address and had some final comments that might be addressed – all of a minor nature and easy to address.

- Abstract – "*energy metabolism and blood coagulation genes*" remove "*genes*"
- Abstract - "*Comparison of single nucleotide polymorphism (SNP) frequencies reveals*" change "*reveals*" to "*revealed*" (for accurate use of tense)
- Introduction – "*Examples of such species complexes are provided finches of the Galapagos island*" should be "*Examples of such species complexes are provided by finches of the Galapagos island*"
- Introduction: "*Thus we were quite keen to apply RNA-sequencing to analyze ecomorphs in our study system, Arctic charr*". The authors should add the Latin name for charr here, rather than in the next paragraph.

Reply: They have all been fixed.

- Introduction: "*The family is estimated to be between 88–103 million years old*[21,22]*. A whole genome duplication event occurred before the radiation of the salmonid family*[21–24] *which has provided time for divergence of ohnologous genes (paralogous genes originated by whole genome duplication event*"

It would be simpler to just state that the common ancestor to salmonids experienced a whole genome duplication 88–103 million years ago. The actual age of the salmonid family depends on whether one considers the (extinct) direct ancestors to salmonids that didn't experience genome duplication to be salmonids.

Reply: Good suggestion, we now use this wording.

- Introduction: "*Furthermore, recent estimates from the rainbow trout (Oncorhynchus mykiss) genome suggest that ohnologous genes are lost at a rate of about 170 genes per million years, and that on the order of 4500 were retained in rainbow trout*[22]"

This information is inaccurate. Firstly, based on the paper cited (Berthalot *et al*. 2014), this information should state that around 4,500 *pairs* of ohnologous genes were retained from Ss4R (i.e. around 9,000 separate genes). More importantly, without going into detail, the stated data represents a non-comprehensive fraction of the genome. I suggest the authors update this part of the text with accurate estimates, since the number of retained Ss4R ohnologue pairs in much larger than what is stated. The authors might also draw in more comprehensive data from the recent publication of the Atlantic salmon genome (Lien et al. Nature, 533, 200–205)[1]. The simplest way to present the information is to state that around half of the original Ss4R ohnologue pairs are still functionally retained (both stated papers are in agreement about that).

Reply: We thank the reviewer for a good point and clarification. We adopt the wording and rewrote part of this paragraph, it now reads: "The common ancestor to salmonids experienced a whole genome duplication 88–103 million years ago, the fourth vertebrate whole-*genome* duplication (Ss4R) [21–24] . This has provided time for divergence of ohnologous genes (paralogous genes originated by whole genome duplication event) in salmonid lineages. Estimates from the rainbow trout (*Oncorhynchus mykiss*) genome suggest that ohnologous genes are lost at a rate of about 170 genes per million years, and that around half of the original Ss4R ohnologue pairs are still functionally retained in rainbow trout [22] ."

- Figure 1: It would be easier for the reader to link the text and images if the authors updated with 'a', 'b', 'c' and 'd' panels for each of the different charr morphs.

Reply: This has been fixed.

- Introduction: "*In this study, we compare SB-charr from*" should be "*In this study, we compared SB-charr from*" (again, it is correct here to use past tense – the authors should check the rest of the manuscript for similar tense issues).

Reply: Fixed, we went through the manuscript and corrected a few more errors of this type.

- Figure 2: Minor comments – the text "*Map on salmon genes*" is vague and open to several interpretations. Better: "*Map on Atlantic salmon expressed sequence tags*"?

Reply: Fixed, put "*Map on Atlantic salmon ESTs" in the figure and "ESTs = expressed sequence tags*" into legend.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 09 July 2015

doi:10.5256/f1000research.6869.r9419

? **Anne Dalziel**
Institute for Systems and Integrative Biology (IBIS), Department of Biology, Laval University, Quebec City, QC, Canada

In this paper "The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs" Gudbrandsson *et al*. have tested for differential gene expression at multiple developmental time-points among a number of Artic charr morpho-types from Lake Thingvallavatn (3 wild morphs, 1 studied with RNA-seq and qPCR, the others with qPCR only) and Holar aquaculture (1 domesticated morph, RNA-seq and qPCR). They have also studied multiple tissues/body regions for a subset of the differentially expressed genes found with RNA-seq. The goal of the paper was to find candidate genes that may underlie variation in morphology, with a focus on craniofacial morphology related to benthic vs. limnetic feeding. In general, I think this goal was met and this paper contributes to our understanding of the mechanisms contributing to morphological evolution in a non-genetic model organism. The authors provide an extensive, multi-time point comparison of two morphologically divergent groups of charr reared in a common environment (reducing the influence of phenotypic plasticity) and have collected a tremendous amount of data. This information will help them to hone in on the genetic loci contributing to phenotypic evolution in this very interesting system, and on the effects of domestication. However, there are a number of major issues that do need to be more clearly addressed in the manuscript prior to final publication. I have outlined these comments below.

Major Comments

1. Introduction:

   Requires some reorganization, clarification of what phenotypes have evolved in parallel among morphs, and how the authors separate the effects of domestication (SB vs. AC) from benthic/limnetic evolution (SB/LB vs. PL/AC).

   a) At present, the introduction focuses upon the utility of instances of parallel evolution to help us determine how repeatable evolutionary change may be. This is definitely true, and the repeated evolution of the dwarf, benthic morph (SB; the focus of the introduction/abstract/discussion) in many lakes strongly argues that this phenotype has evolved via natural selection. However, it is not clear to me if true 'parallelism' seen among the SB (small benthic) and LB (large benthivorous) vs. AC (Holar aquaculture) and PL (small planktivorous) morphs because not enough information is provided for me to assess this. To support the argument for parallelism the specific traits that have

evolved in parallel among morphs must be displayed and the evolutionary history of these morphs should be clarified (e.g. in paragraph 6 and Figure 1). As well, any related non-parallelism in traits should also be discussed (i.e. how are the domesticated AC and wild PL different?). At present Figure 1 only shows the AC and SB morphs, and does not point out the specific traits they are interested in. This is critical background information for readers who are not familiar with this system.

b) The comparison of AC (domestic, limnetic-like head) vs. LB (wild, benthic like head) looks at two confounded variables: domestication and the benthic/limnetic morphology. This should be clearly stated in the introduction, and the use of the additional morphs (PL, LB) in detangling domestication vs. benthic/limnetic evolution should be noted.

c) The use of the AC morph is still a bit unclear to me. The argument for point 'ii) of the availability of abundant AC material' could be expanded by providing more information on the 'limnetic' like features of this morph and why it is an appropriate comparison to a benthic morph, the genetic divergence from the lake Thingvallavatn fish, and also the selection regime it has experienced (selection for limnetic features? What other traits vary with domestication?).

d) Paragraph 2 – Much of this paragraph, including discussing the ability to measure gene expression and relate to phenotype in fishes, is unnecessary as fish are no different from other vertebrates in this respect. Instead, the final sentence "One approach to identify pathways related to function or morphological differences is to study gene expression during development" should become the 'topic sentence' and expanded upon to explain why gene expression studies are especially relevant ways to link genotype to phenotype in evo-devo studies.

e) Better highlight the strengths – The authors have done a wonderful job of assessing multiple developmental time points and rearing fish in a common garden environment. However, they do not highlight these strengths. Some small notes on the importance of controlling for phenotypic plasticity in these traits (which are known to be quite plastic) to better study genetic differentiation would be a nice addition.

2. Methods:

a) Page 4 paragraph 1 - Clarify the number of fish used to make the crosses (this will help us determine the likelihood of selecting a full or half-sib for sequencing/qPCR).

b) I should note that I am not an expert in the analysis of RNA-seq data, but luckily the first reviewer has done an excellent job of commenting upon these aspects of the project. I fully agree with their comments and suggestions. I would also like to see more information on the methods used to pool samples and how RNA-seq data was normalized among samples, developmental times and morphs. I will also note that the authors often use *S.salar* for comparisions, not *O.mykiss*, which is a closer relative to *S.alpinus*. The reasons for this approach should be discussed.

c) I am also not trained as a population geneticist. However, from my experience studying paralogous genes in salmonids, and with respect to the author's own findings for the Nattl paralogs (Fig 4), I do not think it is prudent to "assume that the expression of paralogous genes is stable… " in the methods (page 12). In fact, Berthelot *et al*. (2014) find the opposite (see my comments for the discussion).

d) The authors should use their genetic information to test if the fish chosen are siblings with each other (full or half-sibs). This may have important implications for the population genetic analyses.

e) Page 5 - It is not appropriate to change the meaning of the word 'gene'. I think it is much clearer to use the term 'paralog group' or 'gene family' when referring to the fact that the authors do not study single genes, but instead groups of paralogs.

f) Selection of genes for qPCR – the methods by which genes for the qPCR studies (Fig 3) were selected should be clearly noted. From my reading, it seems that most of these genes do not significantly vary among SB and AC at the 1% FDR level (Tables 1 and 2; only Natterin?). Thus, I am assuming these genes are only significant at the 5% FDR level (S1 file) – why focus upon these and not those significant at 1%? As well, it would be good to include information on why different genes were selected for Figure 3 (qPCR validation of whole fish) and Figure 4 (candidate genes-qPCR validation in just the head). Finally, the abbreviations used for qPCR validation should also be listed in Table 1 for easy comparisons among figures/tables.

3. Results & Figures:

a) Include an experimental design figure - At present, it is difficult to keep track of all of the morphotypes, tissues, and developmental time points used without referring to the methods. Thus, an experimental design figure summarizing the samples used (morphotype, population, sample size, developmental time point), how they were pooled and which techniques were used to measure gene expression on each sample (RNA-seq and/or qPCR) is needed.

b) Include the LB and PL morphs in Figure 1 and clarify traits of interest – The legend states that "differences in size, coloration and head morphology are apparent", but it would be better to specifically point out the differences they are referring to. F1000 is for a general audience, and this would help non-ichthylogists better understand what ecologically-important traits the authors are interested in (e.g. those related to benthic/limnetic feeding). In addition, the two other morphs used in the qPCR studies should also be displayed (large benthivorous and small planktivorous) to facilitate phenotypic comparisons and assess parallelism in benthic/limnetic feeding and/or the effects of domestication on AC.

c) Figure 5- this is actually a table not a figure (?) and is a bit confusing. I think it is much easier to interpret Figure S2 (displaying the data as in Fig 3 and 4), and that Fig 5 and S2 should be switched. It would be great to show significant differences in mRNA content in this, and all other figures, by including symbols. Also, full gene names should be listed in all figure legends.

4. Discussion:

The discussion focuses on the SB morph (page 17 – "The objective of this study were to get a handle on genetic and molecular systems that associate with benthic morphology in charr by mainly focusing on the small benthic morph in Lake Thingvallavatn, Iceland"), while the introduction discusses parallel evolution (indicating that the comparisons should be among many morphs). These are two different topics i) mRNA content differences among benthic vs. limnetic morphs changing in parallel or ii) linking mRNA content to phenotype in SB (benthic, wild) vs. AC (limnetic head, domesticated) morphs. In particular, the role of domestication vs. wild fish divergence needs to be addressed. At present these two topics/questions are mixed in the introduction/discussion and should be addressed separately.

a) Paragraph on Immune Defences - Is immunity also expected to evolve in parallel in all benthic morphs? Is this predicted to be unique to SB vs. AC? Whatever the case, the parallelism (or not) in these genes should also be discussed, and whether this relates more to domestication in AC or differences between limnetic vs. benthic fish. Much of the functional discussion can also be cut.

b) Page 18 – The information about genes found to be differentially expressed among morphs in your prior work should also be in the introduction, as it is background work that explains why you took this transcriptomic approach. This can also be used to explain why you focused in on particular qPCR genes.

c) A discussion of domestication related differences vs. benthic/limnetic differences should be included. I think the data from head gene expression is very interesting (Figs 5, S2) and really speaks to this question.

d) In general, the role of stochastic evolutionary processes, and not just selection (artificial and natural) should be noted. For example, if the AC charr were simply taken from a stock with a different mtDNA haplotype then these differences in the mtDNA genome might not be adaptive, just random. If the AC fish has much higher mtDNA expression might this be simply a domestication issue and not indicative of selection in SB as stated? Finally, you find that not all mitochondrial transcripts (which are transcribed as a polycistronic transcript) are found at similar levels (Table 1) – what does this tell you about differential degradation/post-transcriptional processes?

e) There is no discussion about the "Analyses of polymorphism in Arctic charr transcriptome" (Table 3, 4, 5), except for the mtDNA.

Minor Comments

Introduction:

a) Paragraph 3 – "Furthermore, recent estimates from the rainbow trout….by utilizing multiple data sources the genome assembly problem of this family can be solved". I am not sure how this statement is relevant to this particular study. This and the following statement seem more appropriate for the methods/discussion to me.

b) The morphs being discussed should be clarified throughout the paper. For example, the authors often state "among morphs/among charr populations" but it is not clear which of the many morphs they are referring to (e.g. Paragraph 5, first sentence on allozymes and mtDNA and later sentence on MCHIIa – do you mean all 4 morphs of specific 2-way comparisons? Are some morphs more differentiated than others?)

Methods:

a) The authors should note why they did not use the PI (large piscivorous) morph in any qPCR studies (in the methods or discussion) as this would be a nice morph to use in their tests for parallelism.

b) Page 5 (last paragraph) – the methods used to remove particular variants needs to be clarified. In particular, why the assumptions used to remove variants are valid by referencing past studies.

Figures & Results:

a) Figure 2. The key for Figure 2 should include a specific heading for morph and time-point with the abbreviations restated [e.g. Timepoint: 141 dpf, Morph: Small Benthic (SB)].

b) Figure 6 – would be helpful to label the protein coding genes in this figure as well as the 12s and 16s RNAs.

c) Figure 7 – It is not clear to me which variant is present in which morph. Adding the nucleotide to the x-axis (i.e. frequency of m1829G for B) would make this figure easier to quickly interpret. The "A.charr_WT" and "A.charr_M" should also be defined in the legend and it would be more appropriate to use scientific names for all species.

Discussion:

a) Discussion of reference 32 – The discussion of reference 32 is not put into the proper context. Figure 6 of this paper (Berthelot *et al*. 2014) shows that there are many genes that have no correlation among expression patterns and/or differences in expression levels (1573, 1248, and 1895=4716 paralog pairs), and that together these represent more than the 1,407 correlated/similar expression level paralogs. This section of the discussion needs to be modified.

b) The Norman *et al*. (2014) paper should be mentioned earlier – if this is available why was it not used for their analyses? As well, the last sentence in this paragraph can be cut as it is evident.

c) Page 18 – "Our new data also demonstrate differences in craniofacial elements between AC- and SB-charr, along a limnetic vs. benthic axis[79]". Are you referring to ref 79 or data from this study? If you are referring to 79, clarify and note what you found. This occurs a few times in the discussion

General grammatical errors

There are a number of grammatical errors throughout this paper (e.g. "31 genes were higher expressed in SB and 40 genes higher in AC-charr"; "that may help sculpture benthic vs. limnetic heads" pg 19).

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 04 Apr 2016

Arnar Palsson, University of Iceland, Iceland

Major Comments

Introduction:

Requires some reorganization, clarification of what phenotypes have evolved in parallel among morphs, and how the authors separate the effects of domestication (SB vs. AC) from

benthic/limnetic evolution (SB/LB vs. PL/AC).

a) At present, the introduction focuses upon the utility of instances of parallel evolution to help us determine how repeatable evolutionary change may be. This is definitely true, and the repeated evolution of the dwarf, benthic morph (SB; the focus of the introduction/abstract/discussion) in many lakes strongly argues that this phenotype has evolved via natural selection. However, it is not clear to me if true 'parallelism' seen among the SB (small benthic) and LB (large benthivorous) vs. AC (Holar aquaculture) and PL (small planktivorous) morphs because not enough information is provided for me to assess this. To support the argument for parallelism the specific traits that have evolved in parallel among morphs must be displayed and the evolutionary history of these morphs should be clarified (e.g. in paragraph 6 and Figure 1). As well, any related non-parallelism in traits should also be discussed (i.e. how are the domesticated AC and wild PL different?). At present Figure 1 only shows the AC and SB morphs, and does not point out the specific traits they are interested in. This is critical background information for readers who are not familiar with this system.

Reply: These are excellent suggestions. At the end of the intro we stress the difference between the aims of our research program (study the genetics of parallel evolution) and the aims of this study (get a handle on differences between sympatric morphs, with the AC as possible outgroup). The morphs studied here do not represent parallel evolution of benthic phenotypes (SB and LB are both from the same lake and appear to be closely related - Kapralova et al 2011). Analyses of that question requires further studies. This data can implicate genes that separate PL/AC and SB/LB and may be studied in such follow up analyses of more populations. We have updated figure 1 as advised - including the 4 morphs studied, expanded on the legend and also provide an overview of research approach (part B).

b) The comparison of AC (domestic, limnetic-like head) vs. LB (wild, benthic like head) looks at two confounded variables: domestication and the benthic/limnetic morphology. This should be clearly stated in the introduction, and the use of the additional morphs (PL, LB) in detangling domestication vs. benthic/limnetic evolution should be noted.

c) The use of the AC morph is still a bit unclear to me. The argument for point 'ii) of the availability of abundant AC material' could be expanded by providing more information on the 'limnetic' like features of this morph and why it is an appropriate comparison to a benthic morph, the genetic divergence from the lake Thingvallavatn fish, and also the selection regime it has experienced (selection for limnetic features? What other traits vary with domestication?).

Reply (b and c): The reviewer is correct, AC and SB are separated by multiple traits, and the data probably reveal signals associating with most of them. Unfortunately the AC charr is not well characterized phenotypically, thus we can not address the question of other traits. We focus mainly on the head and jaw morphology, as these attributes distinguish benthic and limnetic morphs. The revised intro elaborates on the choice of AC, and how the follow up work on the morphs from Lake Thingvallavatn can help us sort this out. This point is also picked up in the discussion.

d) Paragraph 2 – Much of this paragraph, including discussing the ability to measure gene expression and relate to phenotype in fishes, is unnecessary as fish are no different from other vertebrates in this respect. Instead, the final sentence "One approach to identify pathways related to function or morphological differences is to study gene expression during development" should become the 'topic sentence' and expanded upon to explain why gene expression studies are

especially relevant ways to link genotype to phenotype in evo-devo studies.

Reply: We restructured and shortened this paragraph around this topic sentence - and gave more room for the previous RNAseq study on Arctic charr.

e) Better highlight the strengths – The authors have done a wonderful job of assessing multiple developmental time points and rearing fish in a common garden environment. However, they do not highlight these strengths. Some small notes on the importance of controlling for phenotypic plasticity in these traits (which are known to be quite plastic) to better study genetic differentiation would be a nice addition.

Reply: Great advice, we tried to integrate this into the last paragraph of the intro.

Methods:

a) Page 4 paragraph 1 - Clarify the number of fish used to make the crosses (this will help us determine the likelihood of selecting a full or half-sib for sequencing/qPCR).

Reply: We did bulk crosses, joining eggs from 5-10 females in a can and sperm from 3-5 males (SB, PL, LB) and single parent cross for AC. Each sample included RNA pooled from 3 embryos, so there is a chance that full sibs were sequenced, but unlikely. The embryos/samples for qPCR are from similar pools. Now described better in methods.

b) I should note that I am not an expert in the analysis of RNA-seq data, but luckily the first reviewer has done an excellent job of commenting upon these aspects of the project. I fully agree with their comments and suggestions. I would also like to see more information on the methods used to pool samples and how RNA-seq data was normalized among samples, developmental times and morphs. I will also note that the authors often use S.salar for comparisions, not O.mykiss, which is a closer relative to S.alpinus. The reasons for this approach should be discussed.

Reply: The RNA was isolated from individual embryos, quantified and then united (in equal concentrations) prior to cDNA synthesis. The read counts per gene are normalized per million reads in sample. Not normalized with other variables.

c) I am also not trained as a population geneticist. However, from my experience studying paralogous genes in salmonids, and with respect to the author's own findings for the Nattl paralogs (Fig 4), I do not think it is prudent to "assume that the expression of paralogous genes is stable… " in the methods (page 12). In fact, Berthelot et al. (2014) find the opposite (see my comments for the discussion).

Reply: Excellent suggestion. We corrected our misunderstanding, added this fact into the intro and discussion, and reinterpreted our data in this light.

d) The authors should use their genetic information to test if the fish chosen are siblings with each other (full or half-sibs). This may have important implications for the population genetic analyses.

Reply:The fish chosen for pop-gen work are random sample from spawning grounds - assumed to be not sibling groups. Our earlier study (Kapralova 2011) showed no family structure in charr collected this way from the lake.

e) Page 5 - It is not appropriate to change the meaning of the word 'gene'. I think it is much clearer to use the term 'paralog group' or 'gene family' when referring to the fact that the authors do not study single genes, but instead groups of paralogs.

Reply: Excellent suggestion. We amended this., and use paralog group throughout.

f) Selection of genes for qPCR – the methods by which genes for the qPCR studies (Fig 3) were selected should be clearly noted. From my reading, it seems that most of these genes do not significantly vary among SB and AC at the 1% FDR level (Tables 1 and 2; only Natterin?). Thus, I am assuming these genes are only significant at the 5% FDR level (S1 file) – why focus upon these and not those significant at 1%? As well, it would be good to include information on why different genes were selected for Figure 3 (qPCR validation of whole fish) and Figure 4 (candidate genes-qPCR validation in just the head). Finally, the abbreviations used for qPCR validation should also be listed in Table 1 for easy comparisons among figures/tables.

Reply: Very important point. We deliberately studied some genes with less statistical support (FDR between 5% and 10%), to gauge the differences in the genes with less support and in particular to have a bigger pool of candidates that may relate to the specific developmental process (like head and jaw formation). Of course we can not assert that all the genes with strongest DE signal in the transcriptome are true positives, but the data can be used for hypothesis generation. We also amended table 1 and the figure legends accordingly.

Results & Figures:

a) Include an experimental design figure - At present, it is difficult to keep track of all of the morphotypes, tissues, and developmental time points used without referring to the methods. Thus, an experimental design figure summarizing the samples used (morphotype, population, sample size, developmental time point), how they were pooled and which techniques were used to measure gene expression on each sample (RNA-seq and/or qPCR) is needed.

b) Include the LB and PL morphs in Figure 1 and clarify traits of interest – The legend states that "differences in size, coloration and head morphology are apparent", but it would be better to specifically point out the differences they are referring to. F1000 is for a general audience, and this would help non-ichthylogists better understand what ecologically-important traits the authors are interested in (e.g. those related to benthic/limnetic feeding). In addition, the two other morphs used in the qPCR studies should also be displayed (large benthivorous and small planktivorous) to facilitate phenotypic comparisons and assess parallelism in benthic/limnetic feeding and/or the effects of domestication on AC.

Reply: (a and b) Excellent suggestions. Now picture 1 has all 4 morphs, and a schematic describing the work flow and samples.

c) Figure 5- this is actually a table not a figure (?) and is a bit confusing. I think it is much easier to interpret Figure S2 (displaying the data as in Fig 3 and 4), and that Fig 5 and S2 should be switched. It would be great to show significant differences in mRNA content in this, and all other figures, by including symbols. Also, full gene names should be listed in all figure legends.

Reply: We acknowledge that this graph is not the simplest, but would like to keep it over Figure S2.

Our reasoning is that this graph illustrates the sharp differences between the limnetic (AC-PL) and benthic (SB-LB), which are the main result in this section. But we will of course switch them, or possibly join both in a single figure ?? if the reviewer insists or the editors recommend it.

Discussion:

The discussion focuses on the SB morph (page 17 – "The objective of this study were to get a handle on genetic and molecular systems that associate with benthic morphology in charr by mainly focusing on the small benthic morph in Lake Thingvallavatn, Iceland"), while the introduction discusses parallel evolution (indicating that the comparisons should be among many morphs). These are two different topics i) mRNA content differences among benthic vs. limnetic morphs changing in parallel or ii) linking mRNA content to phenotype in SB (benthic, wild) vs. AC (limnetic head, domesticated) morphs. In particular, the role of domestication vs. wild fish divergence needs to be addressed. At present these two topics/questions are mixed in the introduction/discussion and should be addressed separately.

Reply: We tried to separate these two aims more clearly in the revised discussion. The strategy was to use the AC vs SB contrast for hypothesis generation, as the first aim is central to our program. We have now added sentences on the domestication in two parts of the discussion.

a) Paragraph on Immune Defenses - Is immunity also expected to evolve in parallel in all benthic morphs? Is this predicted to be unique to SB vs. AC? Whatever the case, the parallelism (or not) in these genes should also be discussed, and whether this relates more to domestication in AC or differences between limnetic vs. benthic fish. Much of the functional discussion can also be cut.

Reply: Good question, we assume it to be so, but that may be wrong. We moved the discussion towards this question and away from functional description.

b) Page 18 – The information about genes found to be differentially expressed among morphs in your prior work should also be in the introduction, as it is background work that explains why you took this transcriptomic approach. This can also be used to explain why you focused in on particular qPCR genes.

Reply: We added a sentence in the intro about the published papers, that this transcriptome made available. In those papers we focused on genes with putative craniofacial effects, though the focus in this study was broader.

c) A discussion of domestication related differences vs. benthic/limnetic differences should be included. I think the data from head gene expression is very interesting (Figs 5, S2) and really speaks to this question.

d) In general, the role of stochastic evolutionary processes, and not just selection (artificial and natural) should be noted. For example, if the AC charr were simply taken from a stock with a different mtDNA haplotype then these differences in the mtDNA genome might not be adaptive, just random. If the AC fish has much higher mtDNA expression might this be simply a domestication issue and not indicative of selection in SB as stated? Finally, you find that not all mitochondrial transcripts (which are transcribed as a polycistronic transcript) are found at similar levels (Table 1) – what does this tell you about differential degradation/post-transcriptional processes?

e) There is no discussion about the "Analyses of polymorphism in Arctic charr transcriptome" (Table 3, 4, 5), except for the mtDNA.

Reply: (c,d,e) Excellent suggestions. We added in the final discussion section few sentences on domesticated charr vs Benthic/limnetic. Unfortunately we do not have quantitative data on the phenotypes (head shape, and jaw) of the AC charr and acknowledge that we categorize it as limnetic based on general features.

We gladly added a sentence citing neutral forces, and are acutely aware that much of the divergence is likely due to history, drift etc. The domestication can certainly be the driver for the higher expression in AC - but we need transcriptomes from more populations/morphs to address that point. And yes, the variance in RNA levels from different parts of the mtDNA do indeed suggest differential half life of the various RNA species. Some are certainly degraded and others most probably actively utilized / protected. We decided not to follow that thought further though, as the MS already consists of quite a few threads already.

We also added sentences on the genetic polymorphism, before focusing more on the mtDNA. The main reason we dont want to elaborate to much on the SNPs is that we feel these data are mainly for generating hypotheses, and that more work is needed to substantiate SNPs and study their distribution in other populations.

Minor Comments

Introduction:

a) Paragraph 3 – "Furthermore, recent estimates from the rainbow trout….by utilizing multiple data sources the genome assembly problem of this family can be solved". I am not sure how this statement is relevant to this particular study. This and the following statement seem more appropriate for the methods/discussion to me.

Reply: We deleted this sentence and simplified the paragraph.

b) The morphs being discussed should be clarified throughout the paper. For example, the authors often state "among morphs/among charr populations" but it is not clear which of the many morphs they are referring to (e.g. Paragraph 5, first sentence on allozymes and mtDNA and later sentence on MCHIIa – do you mean all 4 morphs of specific 2-way comparisons? Are some morphs more differentiated than others?)

Reply: We tried to clarify this in various places in the manuscript, but in some cases we refer to morphs in general. Genetic separation can be estimated with Fst values either between pairs or over a larger set of groups (populations, morphs). In the intro we cite the work done to date in Iceland, which highlights the need for more pop. genetic analyses.

Methods:

a) The authors should note why they did not use the PI (large piscivorous) morph in any qPCR studies (in the methods or discussion) as this would be a nice morph to use in their tests for parallelism.

Reply: The PI charr is very rare in the lake and hard to catch. We later captured few sexually mature individuals, and generated couple of families, that were used for one study (Ahi et al Evodevo 2015).

b) Page 5 (last paragraph) – the methods used to remove particular variants needs to be clarified. In particular, why the assumptions used to remove variants are valid by referencing past studies.

Reply: Many of the principles are common to most pipelines for removing spurious variants. In addition we applied filters necessitated by the properties of our dataset (pool of individuals), the mapping to an outgroup and paralogs due to salmonid genome complexity.

Figures & Results:

a) Figure 2. The key for Figure 2 should include a specific heading for morph and time-point with the abbreviations restated [e.g. Timepoint: 141 dpf, Morph: Small Benthic (SB)].
Reply: Now fixed.

b) Figure 6 – would be helpful to label the protein coding genes in this figure as well as the 12s and 16s RNAs.
Reply: Now fixed.

c) Figure 7 – It is not clear to me which variant is present in which morph. Adding the nucleotide to the x-axis (i.e. frequency of m1829G for B) would make this figure easier to quickly interpret. The "A.charr_WT" and "A.charr_M" should also be defined in the legend and it would be more appropriate to use scientific names for all species.
Reply: Now fixed

Discussion:

a) Discussion of reference 32 – The discussion of reference 32 is not put into the proper context. Figure 6 of this paper (Berthelot et al. 2014) shows that there are many genes that have no correlation among expression patterns and/or differences in expression levels (1573, 1248, and 1895=4716 paralog pairs), and that together these represent more than the 1,407 correlated/similar expression level paralogs. This section of the discussion needs to be modified.

Reply: Really valuable point, that we are especially grateful for. That we have added this fact to the intro and altered our interpretations in the discussion.

b) The Norman et al. (2014) paper should be mentioned earlier – if this is available why was it not used for their analyses? As well, the last sentence in this paragraph can be cut as it is evident.

Reply: The Norman papers are now presented more clearly in the intro. There are historical reasons for not including their data in our analyses, we had completed the analyses for this manuscript when they became available and have since then focused our data analyses efforts on another transcriptome generated in the lab (with longer reads).

c) Page 18 – "Our new data also demonstrate differences in craniofacial elements between AC- and SB-charr, along a limnetic vs. benthic axis79". Are you referring to ref 79 or data from this

study? If you are referring to 79, clarify and note what you found. This occurs a few times in the discussion

Reply: Ref 79 is a related study that built in part on the data presented here. We have now rephrased this in the manuscript, hopefully to the better.

Competing Interests: No competing interests were disclosed.No competing interests were disclosed.

Referee Report 07 July 2015

doi:10.5256/f1000research.6869.r8970

Daniel Macqueen
Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, UK

Review of Gudbrandsson *et al*. "The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs".

The work is founded on the solid premise that rapidly evolving phenotypes in nature can be underpinned by changes at the transcriptome level. The model system here is Arctic charr populations that have evolved (since the last ice age) major differences in phenotypes along the 'benthic' - 'limnetic' axis, with strong differences in head morphology linked to feeding specializations. The work provides an extensive analysis of transcriptome and genetic differences between different morphs and populations. It is interesting, generally well-written and has merit on many levels. It is also rather hard going, since so much ground is covered on diverse areas. The study also comes with a large number of caveats, of which the authors are undoubtedly aware. Overall though, I am supportive of this work, as it represents one of the most detailed analyses of molecular mechanisms linked to rapid phenotypic evolution in Arctic charr. I see it as a great start point for future work and a source of several new findings and hypotheses. I suggest that the paper be indexed in F1000 Research as long as its caveats are transparent and the authors address my comments.

I list below a number of suggestions that may help the authors improve the work, or that at least highlight study limitations for the benefit of interested readers. I also provide a number of minor comments and suggestions, which should help improve the manuscript more incrementally.

Main comments & caveats

1. **RNAseq study design.** I sympathize with the fact that the authors are trying to publish Illumina data that was generated in 2009, since (obviously) the technology has moved on greatly in the last 6 years, while its costs have been reduced dramatically. Adding to this is the fact that the authors are using a particularly complex transcriptome in terms of high content of similar paralogues (and expressed transposable elements), without a reference sequence for mapping in their species. I accept the author's argument that it is more sensible to map against a closely related species with the sequence data rather than to try and create a *de novo* assembly from 36bp reads. I also believe it is sensible to pool read counts for putative paralogous contigs in this study, since the short read length ablates any ability to separate paralogous differences in expression (yet does not preclude the generation of useful hypotheses about putative gene expression differences among morphs).

However, I do question whether the use of Atlantic salmon EST contigs is the best approach here. Firstly, reference assemblies for both Atlantic salmon and rainbow trout are now available, which distinguish paralogous variation. More importantly, using these reference genome data would provide certainty that reads are being mapped to exons from single genes, whereas many of the ESTs will provide a fragmented representation of exon sequences, presumably relying on annotation to piece them back into 'genes' *post hoc* . In addition, paired 100bp Ilumina reads are available at high coverage for Arctic charr (e.g. Norman *et al*. 2014), which could also be used to generate a specific reference transcriptome to map against in this study, although this might be underrepresented in terms of developmental genes as it is a gill study. Overall, I do wonder how much more information might have been gleaned from this dataset with a different mapping strategy?

With all the above said, I understand that the authors have built up a large study based around the original mapping to the salmon ESTs and that it would not be routine for them to repeat the study using better reference data. Furthermore, the approach used has definitely led to the generation of several valid hypotheses concerning the nature of gene expression and genetic differences among charr morphs, which have been followed up using independent approaches.

2. **Methods "Biological Replication in RNAseq" –** a general comment: obviously the design of the study is not optimal because biological variation within developmental stages is not considered in the statistics. Thus, the approach lacks power to detect differences when morph variation is restricted to different developmental stages. I wanted to explain my opinion (for the record) that the study design is nonetheless useful for identifying constitutive differences between morphs. This is especially true because gene expression variability is likely to be relatively low in embryonic stages (compared to a similar study design in adults at least). Further, the pooling of individuals will have helped to at least recapture some biological variation at different stages. Thus, as mentioned above, I see the author's use of RNAseq as a hypothesis-generating approach, which has been quite fruitful in identifying putative differences between different morphs.

3. **Methods "QPCR study design".** The authors adhere to the MIQE guidelines, but do not always follow the best approaches. Most pertinently, the authors use the $2^{-\Delta\Delta Ct}$ method (assuming PCR efficiency of 2.0) despite having gone to the effort of gaining and reporting efficiencies for each assay, which can be as low as 1.72 for some genes. The effect of failing to incorporate differences in efficiency are highly established and this is likely to have affected the author's results. The authors should consider incorporating the effect of differences in efficiency into their analyses. This is likely to have some impact on the study conclusions in my opinion.

4. **Methods "Polymorphisms in charr transcriptome".** While this is not exactly my area of expertise, I struggled to understand the methods behind filtering paralogous variants from SNPs in the data. The authors state "*As the SNP analysis was done on individual contigs, differences among paralogs appear in the data. However, since each sample is a pool of few individuals, it is very unlikely that we have the same frequency of true SNPs in the samples. This property was used to remove variants that are most likely due to expressed paralogs*". Can the authors please try to re-explain this in even simpler terms to help me get it? I don't see how this description leads to a robust identification of paralogous variation. Is there an underlying assumption of equal expression among paralogues? If so, this is likely to be routinely invalidated.

5. **Methods** "**Verification of candidate SNPs".** While it is good that the authors have attempted to verify SNPs identified from their RNAseq data, I don't believe the data is particularly well incorporated in the results section. It needs to be stated up front the extent to which the SNPs predicted from the RNAseq were independently verified. Also, the methods for this section can be improved, especially "*we conducted genomic comparisons of the Salmon genome, ESTs and short contigs from the preliminary assembly of the Arctic charr transcriptome*". None of this information is elaborated on – what is the preliminary assembly of the Arctic charr transcriptome? Which version of the salmon genome was used and how? Moreover, it would be useful to actually explain in the methods that the genotyping was done on a small number of SB, PL and PI morphs, rather than relying on the reader to extract all the required information from Table S2. I guess overall, the way this section is incorporated into the manuscript needs some thought in terms of improving the reader's experience. I struggled after reading it several times and am still not sure I have all the information I need.

6. **Results**. "*Analyses of those reads require an Arctic charr genome sequence or transcriptome assembly from longer and paired end reads*." As mentioned already, the latter is available to generate an Arctic charr transcriptome assembly to map against.

7. **Results**; Figure 3 and 4. The authors found that around half the genes studied were not differentially expressed among morphs by qPCR. Obviously this is quite a large number, but on closer inspection, I noticed that *Ndub6*, *Ubl5* and *parp6* were not even differentially expressed according to RNAseq. Thus, I am confused at the selection of genes from the RNAseq analysis for verification by qPCR. The authors should explain this selection more transparently and provide clearer indices of the correlation between RNAseq and qPCR results and associated discussion.

Minor comments, typos and suggested changes

1. Abstract: "*Species and populations with parallel evolution of specific traits can help illuminate how predictable adaptations and divergence are at the molecular and developmental level*. Grammatically – his reads better: "*….. can help illuminate the predictability of adaptations and divergence at the molecular and developmental level"*

2. Introduction: "*Examples of such a species complex are the finches of the Galapagos islands, cichlids in the African great lakes are exciting multi-species systems in this respect"*. Grammatically – reads better: "*Examples of such species complexes are provided by finches of the Galapagos islands, while cichlids of the African great lakes also provide an exciting multi-species system in the same respect"*

3. Introduction: "*Some northern freshwater fish species exhibit frequent parallelism in trophic structures and life history and in several cases are they found as distinct resource morphs*" change to "*…. are found as distinct resource morphs*"

4. Introduction: "*in the development of ecological differences in tropic morphology*" change to "*… trophic morphology*".

5. Introduction: "*The family is estimated to be between 63.2 and 58.1 million years old*". This information is not correct – it is correct to state that the age of the salmonid crown (based on the cited paper; different estimates exist in the literature, e.g. Macqueen and Johnston, 2014; Campbell *et al*. 2013) is estimated at 63.2 and 58.1 million years old, but the family dates back much further – to the origin of the WGD event in fact, which occurred more like 88-103 Ma

(Macqueen and Johnston, 2014; Berthelot *et al*. 2014). Thus, the last common ancestor to extant salmonid species is what the authors are actually referring to in this sentence.

6. Introduction: "*Furthermore, for data with short reads, mapping to a related reference genome/transcriptome is recommended over de novo assembly*". While this sentence is technically correct in the context of the work cited, I feel it is being used slightly out of context. For a start, what comprises a 'short read' is undefined. 36bp is short, but it is possible to get a sold reference transcriptome using 2*100bp, assuming the appropriate diversity of transcripts is represented and suitable depth is attained.

7. Introduction: "*nuclear genes, reveled both subtle*" change to "*nuclear genes, revealed both subtle*"

8. Minor comment – AC, PL, LB and SB were already defined in introduction.

9. Methods: "*Fishing in Lake Thingvallavatn was with permissions*" changed to "*Fishing in Lake Thingvallavatn was done with permissions*".

10. Methods: "*of differently expressed genes, we preformed clustering analyses*" change to "*…we performed clustering analyses*"

11. Results: "*The most drastic changes were seen in processes related to glycolysis (GO:0006096, FDR = 0.0009), were the expression of 19 out of 25 genes*" change to "*…. where the expression*".

12. Figure 7. What does the charr_WT vs. charr_M signify in the alignment data?

13. Discussion "*We are interested in how predictable evolution is a the molecular level and if there certain principles influence the rewiring of developmental and regulatory systems during evolution*" consider changing to "*We are interested in the predictability of evolution at the molecular level, especially whether there exist principles that influence the rewiring of developmental and regulatory systems"*.

14. Discussion. "*Recent rainbow trout data shows most paralogs from the latest whole genome duplication event retain the same expression pattern*[32] *indicating that this scenario is probably uncommon; hence it is of considerable interest when two paralogs show distinct expression patterns".* I do not agree that it is of considerable interest when two paralogs show distinct expression patterns – I could list tens of examples for salmonids.

15. Conclusions "*The results suggest genetic and expression changes in multiple systems relate to divergence among populations*." Change to "*… associated with divergence among populations*."

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 04 Apr 2016

Arnar Palsson, University of Iceland, Iceland

Main comments & caveats

RNAseq study design. I sympathize with the fact that the authors are trying to publish Illumina data that was generated in 2009, since (obviously) the technology has moved on greatly in the last 6 years, while its costs have been reduced dramatically. Adding to this is the fact that the authors are using a particularly complex transcriptome in terms of high content of similar paralogues (and expressed transposable elements), without a reference sequence for mapping in their species. I accept the author's argument that it is more sensible to map against a closely related species with the sequence data rather than to try and create a de novo assembly from 36bp reads. I also believe it is sensible to pool read counts for putative paralogous contigs in this study, since the short read length ablates any ability to separate paralogous differences in expression (yet does not preclude the generation of useful hypotheses about putative gene expression differences among morphs).

However, I do question whether the use of Atlantic salmon EST contigs is the best approach here. Firstly, reference assemblies for both Atlantic salmon and rainbow trout are now available, which distinguish paralogous variation. More importantly, using these reference genome data would provide certainty that reads are being mapped to exons from single genes, whereas many of the ESTs will provide a fragmented representation of exon sequences, presumably relying on annotation to piece them back into 'genes' post hoc . In addition, paired 100bp Ilumina reads are available at high coverage for Arctic charr (e.g. Norman et al. 2014), which could also be used to generate a specific reference transcriptome to map against in this study, although this might be underrepresented in terms of developmental genes as it is a gill study. Overall, I do wonder how much more information might have been gleaned from this dataset with a different mapping strategy?

With all the above said, I understand that the authors have built up a large study based around the original mapping to the salmon ESTs and that it would not be routine for them to repeat the study using better reference data. Furthermore, the approach used has definitely led to the generation of several valid hypotheses concerning the nature of gene expression and genetic differences among charr morphs, which have been followed up using independent approaches.

Reply: We thank the reviewer for excellent diagnosis and suggestions. The paper describes the (in our humble opinion) most sensible summary of the data, as the writing of the paper started 2 years ago. We did map on the O.mykiss cDNA collection also, got similar results, but opted for reporting on the salmon data to avoid further extending an already long manuscript. We are currently analyzing DE and SNPs on a new assembly (100 bp PE reads - 48 samples - 3 morphs - development), and may include a remapping of this dataset in that.

Methods "Biological Replication in RNAseq" – a general comment: obviously the design of the study is not optimal because biological variation within developmental stages is not considered in the statistics. Thus, the approach lacks power to detect differences when morph variation is restricted to different developmental stages. I wanted to explain my opinion (for the record) that the study design is nonetheless useful for identifying constitutive differences between morphs. This is especially true because gene expression variability is likely to be relatively low in embryonic stages (compared to a similar study design in adults at least). Further, the pooling of individuals will have helped to at least recapture some biological variation at different stages. Thus, as mentioned above, I see the author's use of RNAseq as a hypothesis-generating approach, which has been quite fruitful in identifying putative differences between different morphs.

Reply: We appreciate the reviewers careful analyses of the study and approach. We tried to emphasize the "hypothesis-generation" aspect during the rewrite.

Methods "QPCR study design". The authors adhere to the MIQE guidelines, but do not always follow the best approaches. Most pertinently, the authors use the 2−ΔΔCt method (assuming PCR efficiency of 2.0) despite having gone to the effort of gaining and reporting efficiencies for each assay, which can be as low as 1.72 for some genes. The effect of failing to incorporate differences in efficiency are highly established and this is likely to have affected the author's results. The authors should consider incorporating the effect of differences in efficiency into their analyses. This is likely to have some impact on the study conclusions in my opinion.

Reply: Great point. The qPCR primer efficiencies more than 1.90 can be easily assumed as 2 because of the negligible effects. Since we used LinReg software for efficiencies not the traditional method, it takes into account the efficiencies for each test for a given primer pair and discard those have different and lower efficiencies. However, the Natterin-like paralogues were below the cut-off. The statistical analyses were done on deltaCt values, prior to transformation based on efficiencies used for visualization. We now report the graphs of their expression adjusting for the lower efficiency, and state in the results "Note however, the efficiency of the primers for the nattl genes ranged from 1.72 to 1.77, which suggests this data should be interpreted with caution."

Methods "Polymorphisms in charr transcriptome". While this is not exactly my area of expertise, I struggled to understand the methods behind filtering paralogous variants from SNPs in the data. The authors state "As the SNP analysis was done on individual contigs, differences among paralogs appear in the data. However, since each sample is a pool of few individuals, it is very unlikely that we have the same frequency of true SNPs in the samples. This property was used to remove variants that are most likely due to expressed paralogs". Can the authors please try to re-explain this in even simpler terms to help me get it? I don't see how this description leads to a robust identification of paralogous variation. Is there an underlying assumption of equal expression among paralogues? If so, this is likely to be routinely invalidated.

Reply: We acknowledge this part is a hard read. We rewrote this part of the methods. Here is another summary. Reads from regions that are very similar in paralogous genes can map to both of them. Because we consider also reads that map to many contigs, some of the candidate variants will reflect sequence differences between paralogs, not polymorphism in either paralog. Next we deploy the population genetic argument, since we are sequencing RNA from 6 chromosomes in each sample, then it is very unlikely that a TRUE SNP will be at the same frequency in all of the 8 samples. But variants - that are due to differences bwn paralogs - are likely to be similar in frequency because they are unaffected by the population sampling. This filter is designed to toss those out.

To emphasize the objective is not to find differences between paralogs, but rather to enrich for true SNPs. This method will toss out many sites separating paralogous genes (but not all because some paralogous genes are differentially expressed between morphs or time points).

Methods "Verification of candidate SNPs". While it is good that the authors have attempted to verify SNPs identified from their RNAseq data, I don't believe the data is particularly well incorporated in the results section. It needs to be stated up front the extent to which the SNPs predicted from the RNAseq were independently verified. Also, the methods for this section can be improved, especially "we conducted genomic comparisons of the Salmon genome, ESTs and short

contigs from the preliminary assembly of the Arctic charr transcriptome". None of this information is elaborated on – what is the preliminary assembly of the Arctic charr transcriptome? Which version of the salmon genome was used and how? Moreover, it would be useful to actually explain in the methods that the genotyping was done on a small number of SB, PL and PI morphs, rather than relying on the reader to extract all the required information from Table S2. I guess overall, the way this section is incorporated into the manuscript needs some thought in terms of improving the reader's experience. I struggled after reading it several times and am still not sure I have all the information I need.

Reply: We fixed the methods section to accommodate both reviewers which brought up similar points. We highlight the sampling (8 individuals of 3 morphs), and extend the description of the genomic comparisons. We also extend the discussion of those results.

Results. "Analyses of those reads require an Arctic charr genome sequence or transcriptome assembly from longer and paired end reads." As mentioned already, the latter is available to generate an Arctic charr transcriptome assembly to map against.

Reply: Unfortunately the great Norman *et al.* 2014 data (http://www.ncbi.nlm.nih.gov/pubmed/24368751) came to our attention after we had done these analyses, and started working on our new data (see above). Thus we opted for not redoing the whole analyses for this manuscript, but focus on the verification - and of course working on a new assembly using longer reads.

Results; Figure 3 and 4. The authors found that around half the genes studied were not differentially expressed among morphs by qPCR. Obviously this is quite a large number, but on closer inspection, I noticed that Ndub6, Ubl5 and parp6 were not even differentially expressed according to RNAseq. Thus, I am confused at the selection of genes from the RNAseq analysis for verification by qPCR. The authors should explain this selection more transparently and provide clearer indices of the correlation between RNAseq and qPCR results and associated discussion.

Reply: This reflects the history of the project, and the difference between the preliminary and final analyses. We decided to report on all the data - but explain better in the manuscript the classification of genes tested with qPCR, at 1%, 5% and 10% FDR. In summary, some of the genes tested were above 5% and one even just above 10% FDR. Some of those were not corroborated by qPCR. The number of genes is insufficient to do a statistical comparison of the verification rate at the different FDR levels. A table (new Table 3) - supported with few sentences in the results, hopefully clarifies this.

Minor comments, typos and suggested changes

Abstract: "Species and populations with parallel evolution of specific traits can help illuminate how predictable adaptations and divergence are at the molecular and developmental level. Grammatically – his reads better: "….. can help illuminate the predictability of adaptations and divergence at the molecular and developmental level"
Reply: Thanks - fixed.

Introduction: "Examples of such a species complex are the finches of the Galapagos islands, cichlids in the African great lakes are exciting multi-species systems in this respect". Grammatically – reads better: "Examples of such species complexes are provided by finches of the Galapagos

islands, while cichlids of the African great lakes also provide an exciting multi-species system in the same respect"
Reply: Thanks - fixed.

Introduction: "Some northern freshwater fish species exhibit frequent parallelism in trophic structures and life history and in several cases are they found as distinct resource morphs" change to "…. are found as distinct resource morphs"
Reply: Thanks - fixed.

Introduction: "in the development of ecological differences in tropic morphology" change to "… trophic morphology".
Reply: Thanks - fixed.

Introduction: "The family is estimated to be between 63.2 and 58.1 million years old". This information is not correct – it is correct to state that the age of the salmonid crown (based on the cited paper; different estimates exist in the literature, e.g. Macqueen and Johnston, 2014; Campbell et al. 2013) is estimated at 63.2 and 58.1 million years old, but the family dates back much further – to the origin of the WGD event in fact, which occurred more like 88-103 Ma (Macqueen and Johnston, 2014; Berthelot et al. 2014). Thus, the last common ancestor to extant salmonid species is what the authors are actually referring to in this sentence.

Reply: Thanks so for pointing this out. We changed the text to "local adaptation has been extensively studied in the salmonid family, to which Arctic charr belongs {Fraser2011}. The family is estimated to be between 88-103 million years old {Macqueen2014,Berthelot2014c}. A whole genome duplication event occurred before the radiation of the salmonid family {Davidson2010,Moghadam2011,Macqueen2014,Berthelot2014c} which has provided time for divergence of ohnologous genes (paralogous genes originated by whole genome duplication event). "

*Introduction: "Furthermore, for data with short reads, mapping to a related reference genome/transcriptome is recommended over de novo assembly". While this sentence is technically correct in the context of the work cited, I feel it is being used slightly out of context. For a start, what comprises a 'short read' is undefined. 36bp is short, but it is possible to get a sold reference transcriptome using 2*100bp, assuming the appropriate diversity of transcripts is represented and suitable depth is attained.*

Reply: Great point, we opted for keeping the point (at this place in the ms) but changing the wording to: In this study we opted to map the reads (36 bp) to a related reference genome/transcriptome {Vijay2013a}, instead of conducting de novo assembly.

Introduction: "nuclear genes, reveled both subtle" change to "nuclear genes, revealed both subtle"
Reply: Thanks fixed.

Minor comment – AC, PL, LB and SB were already defined in introduction.
Reply: Thanks, removed this.

Methods: "Fishing in Lake Thingvallavatn was with permissions" changed to "Fishing in Lake Thingvallavatn was done with permissions".

Reply: Ammended.

Methods: "of differently expressed genes, we preformed clustering analyses" change to "…we performed clustering analyses"
Reply: Thanks, fixed.

Results: "The most drastic changes were seen in processes related to glycolysis (GO:0006096, FDR = 0.0009), were the expression of 19 out of 25 genes" change to "…. where the expression".
Reply: Thanks, fixed.

Figure 7. What does the charr_WT vs. charr_M signify in the alignment data?
Reply: Designates the two alleles, the legend now makes this explicit.

Discussion "We are interested in how predictable evolution is a the molecular level and if there certain principles influence the rewiring of developmental and regulatory systems during evolution" consider changing to "We are interested in the predictability of evolution at the molecular level, especially whether there exist principles that influence the rewiring of developmental and regulatory systems".
Reply: Thanks, excellent suggestion, included

Discussion. "Recent rainbow trout data shows most paralogs from the latest whole genome duplication event retain the same expression pattern32 indicating that this scenario is probably uncommon; hence it is of considerable interest when two paralogs show distinct expression patterns". I do not agree that it is of considerable interest when two paralogs show distinct expression patterns – I could list tens of examples for salmonids.
Reply: Good point, we have revisited this interpretation (see also point by rev. 1).

Conclusions "The results suggest genetic and expression changes in multiple systems relate to divergence among populations." Change to "… associated with divergence among populations."
Reply: Thanks, fixed.

Competing Interests: No competing interests were disclosed.No competing interests were disclosed.

Paper IV

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

Jóhannes Guðbrandsson, Sigríður R. Franzdóttir, Bjarni K. Kristjánsson, Ehsan P. Ahi, Valerie H. Maier, Kalina H. Kapralova, Sigurður S. Snorrason, Zophonías O. Jónsson and Arnar Pálsson

PeerJ **6:** e4345

JG participated in sampling and designing the study. He performed the transcriptome assembly, annotation and gene expression analyses. He participated in the writing of the paper.

PeerJ

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

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

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

ABSTRACT

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

Submitted 2 October 2017
Accepted 19 January 2018
Published 7 February 2018

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

Academic editor
Marc Robinson-Rechavi

Additional Information and Declarations can be found on page 24

DOI **10.7717/peerj.4345**

How to cite this article Guðbrandsson et al. (2018), Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs. **PeerJ 6:e4345; DOI 10.7717/peerj.4345**

PeerJ

Subjects Developmental Biology, Evolutionary Studies, Genomics, Molecular Biology, Freshwater Biology
Keywords Divergence, Salmonid, Transcriptome, Evolution, Lake Thingvallavatn, *Salvelinus alpinus*, 3′-bias, RNA sequencing

INTRODUCTION

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

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

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

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

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

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

Sympatric Arctic charr morphs, found in several lakes, most often separate into benthic or limnetic morphotypes varying in many traits (morphology, behavior, color, life history characteristics, habitat use) (*Snorrason & Skúlason, 2004*). A well studied example of polymorphic Arctic charr are the four charr morphs of Lake Thingvallavatn (Fig. 1A). They differ distinctly in various traits, e.g., adult size, age at maturity, head and body morphology, coloration, behavior and habitat use (*Sandlund et al., 1992*). In the lake there are two limnetic morphs, the smaller planktivorous morph (PL, 15–25 cm adult length) that feeds on zooplankton, and the larger piscivorous morph (PI, 25–60 cm adult length) that mainly feeds on threespined stickleback (*Snorrason et al., 1989*; *Malmquist et al., 1992*). The lake harbors two benthic morphs, small benthic charr (SB, 12–20 cm adult length) and large benthic charr (LB, 25–60 cm adult length) both feeding on bottom-dwelling invertebrates in the lava substrate habitat along the shores (*Sandlund et al., 1992*). Rearing experiments showed that morphological and behavioral differences among the morphs arise early in development (*Skúlason et al., 1993*; *Skúlason et al., 1996*), and subsequent studies of developing embryos and juveniles showed significant differences in cartilage and bone formation (*Eiriksson, Skulason & Snorrason, 1999*; *Eiriksson, 1999*). Recently *Ahi et al. (2014)* used geometric morphometrics to capture variation in craniofacial structures among progeny of three of the morphs (PL-, LB- and SB-charr) soon after hatching (280−285 τs, see Materials and Methods for explanation of relative age measured in τs). For the ventral shape of the lower jaw and hyoid arch, distinct differences between the morphs were found at 305 τs, (*Ahi et al., 2014*). Experiments corroborate the contribution of genetic differences, but also demonstrated significant plastic potential of these morphs. The phenotypic plasticity of Arctic charr, and related salmonids is well documented (*Nordeng, 1983*; *Hindar & Jonsson, 1993*; *Skúlason, Snorrasson & Jónsson, 1999*). Studies on developing charr have revealed plastic responses to environmental factors like temperature, water velocity and food type (*Adams & Huntingford, 2004*; *Grünbaum et al., 2007*; *Jonsson & Jonsson, 2014*). Studies of limnetic and benthic charr morphs in Iceland show food type

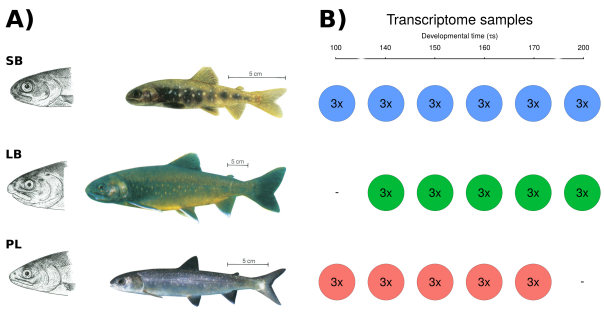


Figure 1 The phenotypically distinct sympatric Arctic charr and the experimental set-up. (A) Four sympatric morphs inhabit Lake Thingvallavatn, three of which are studied and pictured here: small benthic (SB), large benthic (LB) and planktivorous (PL) charr. They differ in size (size bars = 5 cm), the shape of the head and jaws (see drawings) and pigmentation. Adapted from *Sandlund et al. (1992)*, ©Wiley-Blackwell, drawings by Eggert Pétursson. (B) Embryos from pure crosses of the three morphs were sampled at six developmental timepoints prior to hatching, from 100 τs to 200 τs (circles) for RNA sequencing. During this period of development somatogenesis is complete and gill arches, jaws and many other structures are forming (Fig S1). Three biological replicate samples (3×) were taken for each morph and developmental timepoint, each sample being a pool of mRNA from three embryos. Six timepoints were sampled in SB-charr, and five in LB- and PL-charr. In total there were 48 samples, composed of 144 individual charr embryos. The coloring scheme indicating morphs (blue: SB, green: LB, red: PL) will be retained throughout the manuscript.

Full-size DOI: 10.7717/peerj.4345/fig-1

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

Genetic variation in polymorphic and sympatric Arctic charr

The earliest population genetic studies found little genetic separation of the sympatric morphs in Lake Thingvallavatn (*Magnusson & Ferguson, 1987*; *Danzmann et al., 1991*; *Volpe & Ferguson, 1996*). The first microsatellite study detected subtle differences (*Gíslason, 1998*) and the second study with 10 markers estimated overall F_{ST}'s = 0.039, between the LB-, SB- and PL-charr (*Kapralova et al., 2011*). More recently, we detected F_{ST}'s larger than 0.25 between morphs for variants in two immunological genes (*Kapralova et al., 2013*) and a few other loci (*Guðbrandsson et al., 2016*), suggesting substantial genetic separation at specific loci among those sympatric charr morphs. There is a need to study underlying developmental mechanisms, e.g., how differential expression or function of genes promotes

differences in charr development and phenotypes. To date, few studies have addressed these issues. The candidate gene approach illustrates how embryonic morphogenetic mechanisms may influence phenotypic diversity and speciation (*Abzhanov et al., 2004*; *Abouheif et al., 2014*). A qPCR study on muscle tissues in charr, showed that expression of three genes in the *mTOR*-pathway distinguishes five small benthic morphs from two limnetic morphs in Iceland (*Macqueen et al., 2011*). On the other hand the myogenic *paired box protein 7 (Pax7)* gene was not differentially expressed during development in Lake Thingvallavatn morphs (*Sibthorpe et al., 2006*).

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

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

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

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

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

MATERIALS AND METHODS

Sampling, rearing and developmental series

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

Parents were fished in Lake Thingvallvatn with the permissions both from the owner of the land in Mjóanes and from the Thingvellir National Park commission. Ethics committee approval is not needed for regular or scientific fishing in Iceland (The Icelandic law on Animal protection, Law 15/1994, last updated with Law 55/2013).

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

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

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

Sampling of embryos for RNA extraction was performed by Holar University College aquaculture Research Station (HUC-ARC) personnel. Embryos were sampled at designated timepoints, placed in RNAlater (Ambion), stored at +4 °C overnight and frozen at −20 °C. HUC-ARC has an operational license according to Icelandic law on aquaculture (Law 71/2008), which includes clauses of best practices for animal care and experiments.

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

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

Staining of embryos for developmental series

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

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

RNA extraction and transcriptome sequencing

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

Assembly, abundance estimation and annotation

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

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

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

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

Estimation of RNA degradation and 3′-bias

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

Estimating expression differences among morphs

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

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

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

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

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

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

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).

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

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

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

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

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

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

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

qPCR verification of gene expression

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

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

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

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

RESULTS AND DISCUSSION

Transcriptome sequencing, assembly and annotation

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

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

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

Notes.
[a]Longest isoform for each Trinity gene.

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

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

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

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

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

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

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

RNA degradation and 3′-bias in the transcriptome

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

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

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

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

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

Analyses of differential expression (see below) revealed that the estimated 3′-bias was the second most important factor after developmental time with 32,395 significant transcripts ($\alpha = 0.01$, Table 4). Crucially, the results differed considerably if the 3′-bias term was not included; then more transcripts had significant Morph by Time interaction effect ($M \times T$) and fewer significant developmental time effect (Time) (Table 4, Fig S3C and Fig S3D). Many transcripts with significant $M \times T$ interaction effect in a model without a 3′-bias term had significant Time effect after normalizing for 3′-bias (Fig S3A). Thus we concluded that

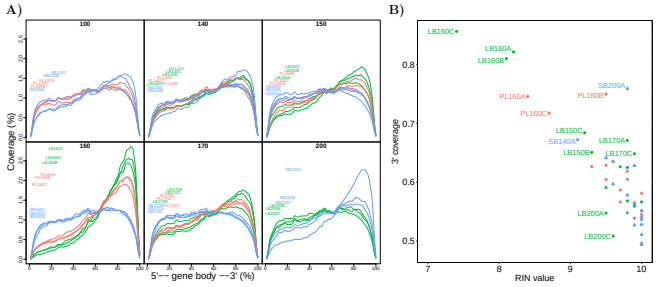


Figure 2 Confounding of 3′-bias with morphs and timepoints in the charr developmental transcriptome data. (A) Average coverage over the length of transcripts for each sample. The coverage was estimated from 381 transcripts that were highly expressed in all samples. The mean coverage for these genes was estimated in 100 windows over the gene body, from the 5′-end to the 3′-end. Samples from different timepoints are graphed separately and colored by morph: LB in green, PL in red and SB in blue. (B) Comparison of RIN-values and 3′ coverage (coverage at the 3′-half divided by total coverage) for each sample, colored by morph.

Full-size DOI: 10.7717/peerj.4345/fig-2

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

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

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

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

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

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

Differential transcript expression between sympatric Arctic charr morphs

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

As transcriptional divergence and genetic divergence tend to be associated (*Whitehead & Crawford, 2006*), this suggests closer relation of the two benthic morphs, with PL-charr as a more distant relative, consistent with one population genetic study

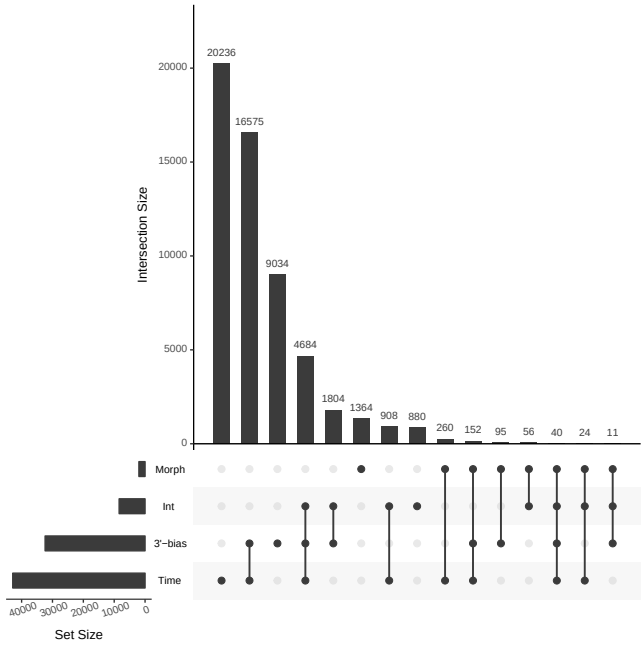


Figure 3 The number of transcripts differentially expressed according to the linear models on developmental timepoint (Time), Morph, 3′-bias and interaction of Morph and Time (int). The set size barplot (sideways) shows cumulated number of transcripts for each of the four main factors, while the intersection size barplot (vertical) shows the number of transcripts significant for each one or a combination of two or more factors. The dots indicate the significant factors or their combinations. For example 20,236 transcripts are only significant for Time effect but 16,575 are significant for both Time and 3′-bias effect.

Full-size DOI: 10.7717/peerj.4345/fig-3

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

We next gauged the functions of the differentially expressed transcripts by Gene Ontology (GO) enrichment analyses, run separately on the 16 co-expression clusters. Note, the GO results should be interpreted cautiously, as mere indications of functional divergence between groups. The analyses were restricted to biological processes and lower level

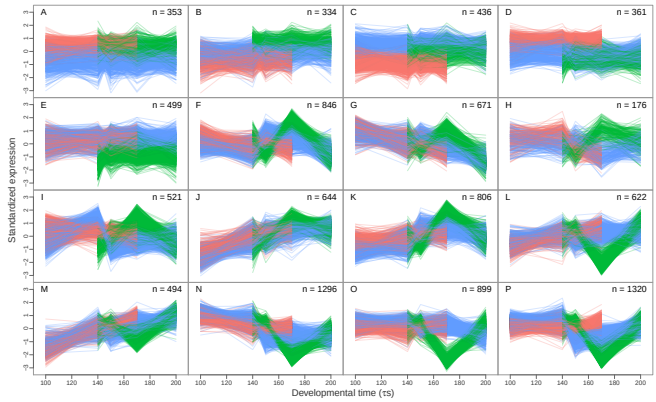


Figure 4 Expression profiles of 16 co-expression clusters. Depicted are transcripts with significant Morph and/or Morph by Time interaction effect, during early charr development (timepoints 100 *τs* to 200 *τs*). Standardized expression normalized by 3′-coverage is plotted against developmental time. Each line is the profile for one transcript. The first five clusters (A–E) capture mainly differences between morphs, while the remaining clusters (F–P) contain almost exclusively transcripts with *MxT* effects (Table 5). The number of transcripts in each cluster is indicated. The morphs are represented by color, SB: blue, LB: green and PL: red.

Full-size DOI: 10.7717/peerj.4345/fig-4

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

Just under 700 GO categories were enriched for clusters of genes with significant Morph by Time interactions (Table S3). While dozens up to a hundred GO categories associated with each co-expression cluster, no general pattern emerged. Many different biological processes were enriched in the co-expression clusters, for instance cluster F was enriched for regulation of growth (e.g., GO:0040008) and antigen processing and presentation (e.g., GO:0048002) and cluster L cartilage condensation (e.g., GO:0001502) and limb bud formation (e.g., GO:0060174). A number of categories showed up in three or more

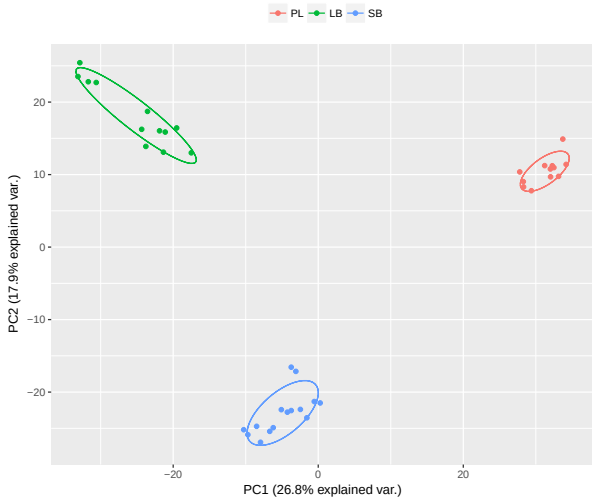


Figure 5 Visualization of the expression differences between the three morphs, with PCA of transcripts in the first five expression clusters (A–E, 1983 transcripts). These clusters were selected as they show an effect of morph, and therefore some separation of morphs in the PCA is expected, but importantly all three morphs are separated from each other. The position of the different morphs in the PCA is informative and indicates that the benthic morphs are more similar in expression as the first axis separates all the morphs with the benthic morphs being close to each other. The second axis separates small benthic from the other two morphs. Standardized expression normalized by 3′-coverage was used as input. Individual samples are graphed (colored by morph, SB: blue, LB: green and PL: red), and overlaid are 68% normal data ellipses for each morph. Figure prepared using the ggbiplot package in R (version 0.55; *Vu, 2011*).

Full-size DOI: 10.7717/peerj.4345/fig-5

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

	Clusters	
Effect	**A–E**	**F–P**
Morph	1,768	234
Time	307	5,761
$M \times T$	296	8,111

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

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

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

Verification of differential expression with qPCR

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

Of those genes, three (*arid4a*, *tsn* and *eif4ebp1*) have known regulatory functions. *Arid4a* encodes a Retinoblastoma binding protein, that has been demonstrated to repress transcription and induce growth arrest in human cell culture (*Lai et al., 1999*). *Tsn* encodes a protein which positively influences the activity of the RISC complex (*Liu et al., 2009*). *Eif4ebp1* encodes a repressor of translation initiation, and is a target of *mTOR* (*Wang et al., 2005*; *Dowling et al., 2010*). The other genes have diverse functions, *prc1* is a cell cycle related gene (*Li, Shridhar & Liu, 2003*), *gpi* a glycolytic enzyme differentially expressed in

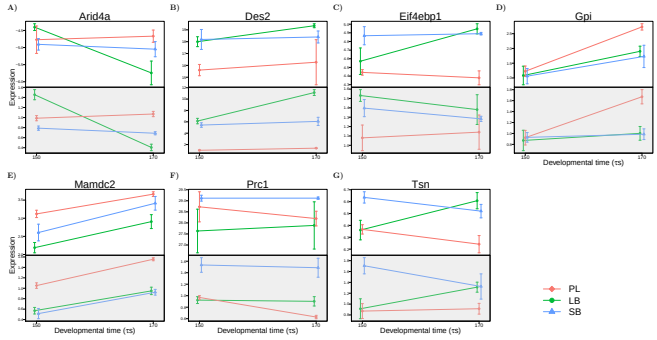


Figure 6 Expression of seven genes that differed between benthic (SB and LB) and limnetic (PL) charr in the transcriptome. (A–G) shows the expression of one gene at developmental timepoints 150 and 170 τs, in the transcriptome (white background) and measured with qPCR (gray background). Colors indicate morph (blue: SB, green: LB, red: PL). The upper panels show expression in transcripts per million (tpm) on log-scale, normalized by the effect of 3′ coverage in the linear model (see Methods). The qPCR expression is normalized to the geometric mean of two reference genes (*actb* and *ub2l3*) and represented relative to one replicate of the PL morph at 150 (τs). Error bars represent 2 standard errors of the mean calculated from three biological replicates each made from a homogenate of three whole embryos.

Full-size DOI: 10.7717/peerj.4345/fig-6

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

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

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

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

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

CONCLUSIONS

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

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

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

ACKNOWLEDGEMENTS

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

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

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

Grant Disclosures

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

Competing Interests

The authors declare there are no competing interests.

Author Contributions

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

Animal Ethics

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

Ethics committee approval is not needed for regular or scientific fishing in Iceland (The Icelandic law on Animal protection, Law 15/1994, last updated with Law 55/2013). Sampling of embryos for RNA extraction was performed by Holar University College Aquaculture Research Station (HUC-ARC) personnel. HUC-ARC has an operational license according to Icelandic law on aquaculture (Law 71/2008), which includes clauses of best practices for animal care and experiments.

Field Study Permissions

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

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

DNA Deposition

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

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

Data Availability

The following information was supplied regarding data availability:

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

Supplemental Information

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. 2009. The MIQE guidelines:

minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55(4)**:611–622 DOI 10.1373/clinchem.2008.112797.

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

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

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

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

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

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

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

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

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

Eiriksson GM, Skulason S, Snorrason SS. 1999. Heterochrony in skeletal development and body size in progeny of two morphs of Arctic charr from Thingvallavatn, Iceland. *Journal of Fish Biology* **55**:175–185 DOI 10.1111/j.1095-8649.1999.tb01054.x.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SUPPLEMENTAL INFORMATION

Table S1. Sequencing effort, quality trimming, mapping and estimated insert size for each sample. DOI: 10.7717/peerj.4345/supp-1

Morph	Relative age	Rep	Index	Lane	Nr of reads	After trim[a]	KaAP[b]	KaFL [c]	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

[a] The proportion of reads retained as pairs after trimming
[b] The proportion of raw reads aligned with kallisto
[c] Estimated fragment length by kallisto

Table S2. Results from differential expression analysis. Multiple testing corrected p-values (q-values) and overall log fold change between morphs from the full model for each transcript are shown (Model FM in methods).

DOI: 10.7717/peerj.4345/supp-2

TransID: Trinity transcript identification code
q_MxT: Q-value for the Morph X Time interaction term from likelihood ratio test between model FM and R1 (see methods)
q_Morph: Q-value for the Morph term from likelihood ratio test between model R1 and R2 (see methods)
q_Time: Q-value for the Time term from likelihood ratio test between model R1 and R3 (see methods)
q_Tprime: Q-value for the 3'-bias terms from likelihood ratio test between model FM and R4 (see methods)
FC_PL_SB: Log fold change between PL and SB. Parameters extract from the full model (FM)
FC_PL_LB: Log fold change between PL and LB. Parameters extract from the full model (FM)
FC_LB_SB: Log fold change between LB and SB. Parameters extract from the full model (FM)
Cluster: Expression cluster for transcripts with significant Morph or Morph X Time interaction
SSncbi_Top_BLASTN_gene_name: Gene name based on top blastn hit in the NCBI *Salmo salar* Annotation
SalmoBase_Top_BLASTN_gene: Id for top blastn gene in SalmoBase *Salmo salar* Annotation

Table S3. The results of GO 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.

DOI: 10.7717/peerj.4345/supp-3

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.

DOI: 10.7717/peerj.4345/supp-4

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 transciptome that the primers bind to and were used for comparison of expression. If there are more than one transcipts for each gene the ids are separated by a semicolon.

Table S5. Tab-delimited text file with detailed annotation of all the filtered transcripts in the Trinity assembly.

On figshare: Annotation_combinded.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

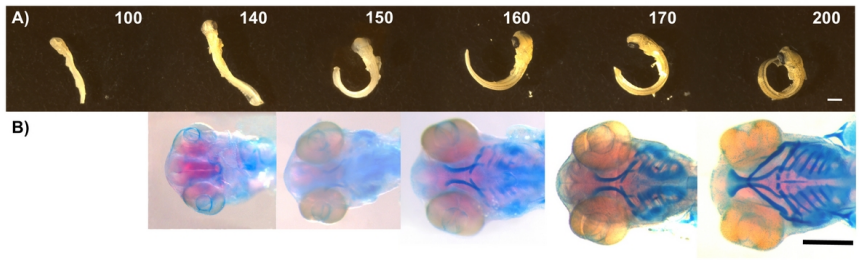


Figure S1. A) Developmental events in the LB-charr at relative ages 100-200 τs (dorsal views of 6 time points). By 100 τs heart contractions have begun and second gill fissures have started to form. By 140 τs all somites are formed and eye pigmentation has started to appear. Between 150-200 τs the upper and lower jaws separate from the yolk, the first melanophores appear and start spreading from the head along the trunk and the operculum covers the first gill arch. Scale bar: 1 mm. **B)** Development and growth of craniofacial cartilage elements at pre-hatching stages at relative stages 140, 150, 160, 170 and 200 τs LB-charr embryos (ventral views of 5 time points): no craniofacial elements are seen at 140 τs; at 150 τs the trabeculae, Meckel's cartilages, and palatoquandrates can be seen clearly; at 160 τs the hyoid arch and the ceratobranchials (cb) 1-3 become visible; at 170 τs: basibranchial (bb) cartilages and cb 1-4 have emerged; at 200 τs the fusing of the ethmoid plate has started and the hypohyal (hh), hypobranchial cartilages (hb) 1–2 and cb 1-5 are visible. Scale bar: 1 mm. DOI: 10.7717/peerj.4345/supp-5

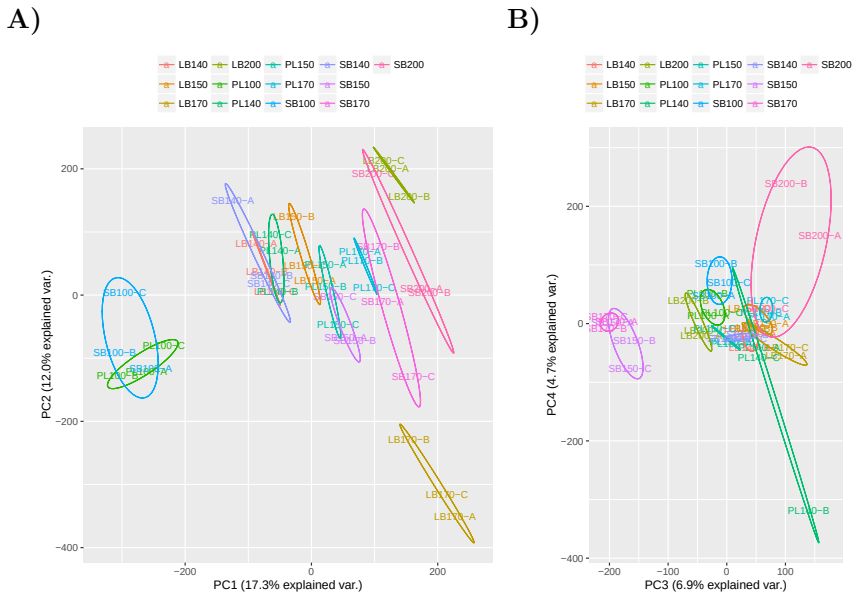


Figure S2. Principal component analysis (PCA) of expression data for all transcripts. **A)** shows the first and second PCA-axis and **B)** the third and fourth PCA-axis. The first PCA-axis correlates with developmental time. Samples from 2011 (SB100, SB140 and PL140) do not deviate largely from other samples for any of the PCA-axis. Standardized expression normalized by 3'-coverage was used as input. Samples are colored according to morph and time, and sample labels are shown for each replicate.
DOI:10.7717/peerj.4345/supp-6

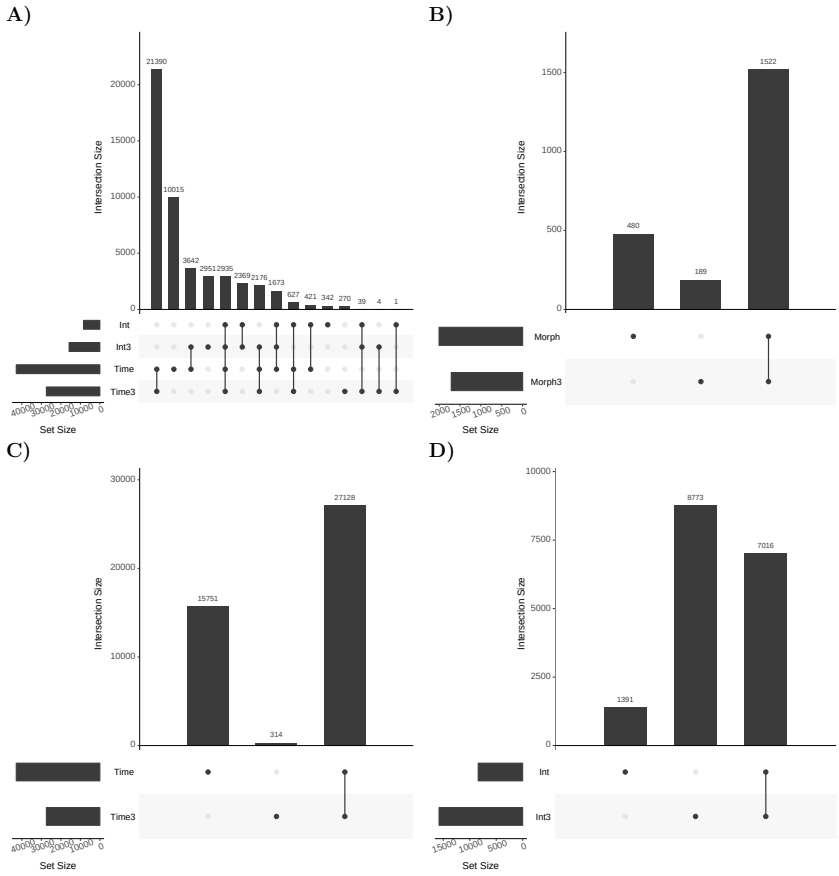


Figure S3. Effect of 3'-bias correction on the number of transcripts differently expressed by developmental timepoint (time), morph and interaction of morph and time (int). Each figure shows the intersection size (upper barplot) - the number of transcripts significant for each one or a combination of 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 *MxT* interaction (int) term. The dots indicate the significant factors or their combinations. DOI:10.7717/peerj.4345/supp-7

Paper V

Extensive genetic divergence between recently evolved sympatric Arctic charr morphs

Jóhannes Guðbrandsson, Kalina H. Kapralova, Sigríður R. Franzdóttir, Völundur Hafstað, Þóra M. Bergsveinsdóttir, Zophonías O. Jónsson, Sigurdur S. Snorrason and Arnar Pálsson

bioRxiv preprint doi:10.1101/489104

JG participated in sampling and designing the study. He performed the variant calling and analyses from the RNA-seq data and participated in variant conformation from population samples. He wrote a first draft and participated in the writing of the paper.

Extensive genetic divergence between recently evolved sympatric Arctic charr morphs

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

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

Abstract

The availability of diverse ecological niches can promote adaptation of trophic specializations and related traits, as has been repeatedly observed in evolutionary radiations of freshwater fish. The role of genetics, environment and history in ecologically driven divergence and adaptation, can be studied on adaptive radiations or populations showing ecological polymorphism. Salmonids, especially the *Salvelinus* genus that includes Arctic charr (*Salvelinus alpinus*), are renowned for both phenotypic diversity and polymorphism. Arctic charr invaded Icelandic streams during the glacial retreat (about 9,000 to 12,000 years ago) and exhibits many instances of sympatric polymorphism. Particularly well studied are the four morphs in Lake Þingvallavatn in Iceland. The small benthic (SB), large benthic (LB), planktivorous (PL) and piscivorous (PI) charr differ in many regards, including size, form and life history traits. To investigate relatedness and genomic differentiation between morphs we extracted variants from developmental transcriptome data from three of those morphs, and verified 22 variants in population samples. The data reveal genetic differences between the morphs, with the two benthic morphs being more similar and the PL-charr more genetically different. The markers with high differentiation map to all linkage groups, suggesting ancient and pervasive genetic separation of these three morphs. No marker associated fully with morph, suggesting polygenic basis of traits separating them. Furthermore, gene ontology analyses suggest differences in collagen metabolism, odontogenesis and sensory systems between PL-charr and the benthic morphs. Genotyping in population samples from all four morphs confirms the genetic separation and indicates that the PI-charr are less genetically distinct than the other three morphs. The genetic separation of the other three morphs indicates certain degree of reproductive isolation. The extent of gene flow between the morphs and the nature of reproductive barriers between them remain to be elucidated.

1 Introduction

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

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

1.1 Genome wide divergence or islands of differentiation?

In the last decade the genomics revolution spawned powerful tools for studying both the connections between genetic and phenotypic variation, e.g. how variations in the form of feeding structures are caused by differential expression of key developmental genes (Abzhanov *et al.* 2006; Guðbrandsson *et al.* 2018), and to determine to what extent these differences are due to genetic differences (and how they are distributed in genomes) (Wolf & Ellegren 2016). These methods have enabled studies of the role of feedback between the organism and its environment, the role of plasticity in generating functional variation and possibly promoting adaptation (Morris & Rogers 2014; Abouheif *et al.* 2014). More central to this study, genomics can reveal differences between populations and ecotypes, and identify genes and pathways that may be under positive selection in natural populations (Malinsky *et al.* 2015; Wolf & Ellegren 2016).

The differentiation and divergence between related groups, populations, sub-species or species is influenced by genomic parameters, population genetic history and evolutionary forces (Seehausen *et al.* 2014; Vijay *et al.* 2017). Depending on multiple factors, e.g. the strength of selection, nature of the adaptive traits, genomic structure, population genetic parameters and history, the genetic separation of populations or subspecies may be rather uniform over the entire genome or localized to "genomic islands of differentiation" (Malinsky *et al.* 2015; Wolf & Ellegren 2016). Studies of closely related species and ecologically distinct populations within species can detect the genetic correlates of adaptive traits (Pease *et al.* 2016), but this is not straightforward. Positive selection can affect alleles, both from older standing variation and more recent variants, private to certain populations (Vijay *et al.* 2017). Adaptive introgression or environmental sorting of an ancestral, possibly balanced, polymorphism, and homoploid speciation can also complicate the analyses (Pardo-Diaz *et al.* 2012; Cruickshank & Hahn 2014; Guerrero & Hahn 2017). Furthermore, distinct populations of a species adapting to different environments, can be viewed as being scattered along a "speciation continuum" (Theis *et al.* 2014; Seehausen *et al.* 2014). Fluctuations in environmental or population genetic parameters can shift populations along this continuum (Grant & Grant 2008). Specifically, changes in the environment or gene flow between populations, can lead to the merging of previously distinct populations, speed up their divergence or produce reproductively isolated species (Hendry *et al.* 2009; Seehausen *et al.* 2014; Lowry & Gould 2016). We are keen to explore the genome-wide differentiation of sympatric polymorphism within a species that recently colonized new habitats, with the aim to study the genetics of ecological specializations.

1.2 Resource polymorphism along a benthic - limnetic axis

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

Salmonids are renowned for their phenotypic diversity, both among and within populations, with multiple examples of resource polymorphism. The most phenotypically diverse and polymorphic species seem to be in the *Salvelinus* genus, including Arctic charr *Salvelinus alpinus* (Klemetsen 2013), Lake charr (also called Lake trout) *Salvelinus namaycush* (Muir *et al.* 2016) and Dolly Varden charr (*Salvelinus malma*) (also called Dolly Varden trout) with as many as seven morphs found in the same lake (Markevich *et al.* 2018). Arctic charr colonized lakes and rivers on the northern hemisphere after the last glaciation period (approx 9,000-12,000 years ago) (Snorrason & Skúlason 2004; Noakes 2008; Klemetsen 2010). In Iceland, multiple lakes harbour polymorphic Arctic charr (Skúlason *et al.* 1992; Snorrason & Skúlason 2004; Wilson *et al.* 2004; Woods *et al.* 2012b) and a unique small benthic morphotype is found in many streams and ponds across the country, especially in cold springs with lava-

rock bottom in the geologically younger parts of the island (Kapralova *et al.* 2011; Kristjánsson *et al.* 2012). One of four sympatric charr morphs found in Lake Þingvallavatn, Iceland's largest lake, is of this type. Population genetics show that different populations of Arctic charr in Iceland are grouped by geography, not morphotype (Gíslason 1998; Kapralova *et al.* 2011), supporting the notion that small benthic morphs and other derived forms have evolved independently in different locations. For instance, the four Lake Þingvallavatn morphs are more closely related to one another than to other charr populations (Kapralova *et al.* 2011). Genetic analysis using allozymes, microsatelites and RADsequencing in Norway, Scotland and Transbaikalia also indicate closely related sympatric morphs (Hindar *et al.* 1986; Gordeeva *et al.* 2015; Jacobs *et al.* 2018), while in other cases sympatric charr morphs seem to have emerged by more than one invasion (Verspoor *et al.* 2010).

1.3 The four sympatric charr morphs in Lake Þingvallavatn

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

The earliest population genetic studies of the Lake Þingvallavatn charr found variation in few genetic markers, and weak separation between morphs (Magnusson & Ferguson 1987; Danzmann *et al.* 1991; Volpe & Ferguson 1996). The estimated relationship between the morphs varied by studies (markers), two datasets placed SB-charr as the most distantly related morph (Magnusson & Ferguson 1987; Gíslason 1998), another placed LB-charr as the most distant morph (Kapralova 2008) while an mtDNA study clustered the benthic morphs (LB- and SB-charr) and placed PL-charr as an outgroup (Volpe & Ferguson 1996). Curiously the PI-charr clustered with both PL- and LB-charr. A study of 9 microsatellite markers in SB- and PL-charr from five spawning sites and LB-charr confirmed morph difference (average F_{ST} = 0.039), but did not resolve their relationship (Kapralova *et al.* 2011). Crucially, those data and coalescence modeling were more consistent with a scenario assuming a brief initial phase in allopatry of PL- and SB-charr with subsequent coexistence, than a sympatric origin of these morphs (Kapralova *et al.* 2011). This observation was supported by data showing high genetic separation in two immunological genes (Kapralova *et al.* 2013) and substantial variation in other loci between morphs (Gudbrandsson *et al.* 2016).

We have studied the developmental and molecular correlates of morph divergence, by analyzing morphology of embryos, miRNA and mRNA expression (Kapralova *et al.* 2014; Ahi *et al.* 2014; Kapralova *et al.* 2015; Gudbrandsson *et al.* 2016; Guðbrandsson *et al.* 2018). Most recently, RNA-sequencing of embryos of three of the four morphs (excluding PI-charr), reared in a common garden set-up, revealed about 2000 transcripts with consistent expression differences between morphs (Guðbrandsson *et al.* 2018). The PL-, LB- and SB-charr differed at the transcriptome level and the results suggested a closer relationship between the two benthic morphs (LB- and SB-charr), with PL-charr more distinct. No enrichment of specific biological pathways (GO analyses) was found for the morph specific transcripts, implying that multiple genes and diverse pathways contribute to the developmental morph differences. However, qPCR analysis of candidate co-expressed genes, including genes related to extra-cellular matrix organization and skeletogenesis as well as their putative regulators, revealed variation along the benthic - limnetic axis (Ahi *et al.* 2013, 2014, 2015). Similarly, Macqueen *et al.* (2011) found differential expression of three genes in the *mTOR*-pathway in muscles of generalist vs. small benthic morphs from multiple locations within Iceland (Macqueen *et al.* 2011). Notably, we found the same increased expression of one of these genes, *Eukaryotic translation initiation factor 4E binding protein 1 (Eif4ebp1)*, in SB-charr from Lake Þingvallavatn (Guðbrandsson *et al.* 2018).

Like other freshwater fishes (Seehausen & Wagner 2014; Bernatchez *et al.* 2010; Berthelot *et al.* 2014), Arctic charr has utility for studies of recent adaptations and the genetics of evolutionary change (Snorrason & Skúlason 2004). Relatively few genomic resources were available for Arctic charr until recently, when a reference genome, that will aid further genomic research of the species, was published (Christensen *et al.* 2018). Two other salmonid

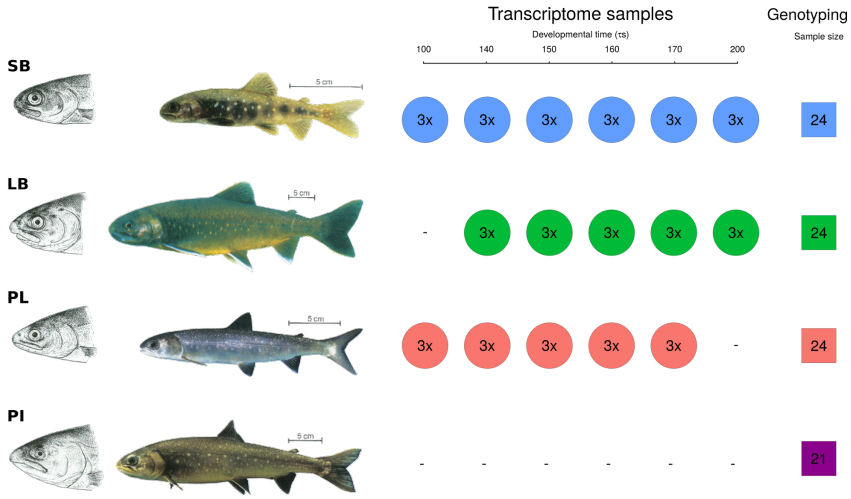


Figure 1: The phenotypically distinct sympatric Arctic charr from Lake Þingvallavatn and the sampling strategy. **A)** The four sympatric morphs are Small benthic (SB), Large benthic (LB), Planktivorous (PL) and Piscivorous (PI) charr. They differ in size (size bars = 5 cm), head and feeding morphology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

genomes are published (Berthelot *et al.* 2014; Lien *et al.* 2016), and a third (Coho salmon, *Oncorhynchus kisutch*) is available on NCBI (Genebank assembly GCA_002021735.1). About 88–103 million years ago an ancestor of Salmonids underwent a whole genome duplication (Ss4R), the fourth on the vertebrate lineage (Allendorf & Thorgaard 1984; Macqueen & Johnston 2014; Berthelot *et al.* 2014). Comparisons of several salmonids established significant synteny of their linkage groups (Danzmann *et al.* 2005; Sutherland *et al.* 2016), despite rearrangements (Timusk *et al.* 2011; Nugent *et al.* 2017). Thus a substantial fraction of salmonid genes are paralogs, and are found in syntenic regions (Nugent *et al.* 2017; Christensen *et al.* 2018). This understandably complicates analyses of differential gene expression (Gudbrandsson *et al.* 2016; Guðbrandsson *et al.* 2018) and genetic variation.

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

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

II Ecological selection on specific traits has lead to genetic separation of morphs, seen as allele frequency dif-

ferences at variants in key genes related to fitness traits, in the face of gene flow between morphs (Seehausen *et al.* 2014; Wolf & Ellegren 2016). Under this model, genomic island of differentiation are expected, with limited genetic differences between morphs in the rest of the genome.

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

These three hypotheses are not mutually exclusive, and some may apply to specific pairs of morphs. The data enabled analyses of the genetic separation of three of the sympatric morphs, the origin of the rare piscivorous charr, the genomic patterns of differentiation between morphs and the genes and molecular systems that may associate with their specializations.

2 Materials and Methods

2.1 Sampling for transcriptome and population genetics

The fishing and sampling of embryos from Lake Þingvallavatn for developmental transcriptome analysis is described in Guðbrandsson *et al.* (2018). Briefly, we caught running SB-, LB- and PL-charr, made crosses and reared embryos in a common garden set-up at the Hólar Aquaculture station, Verið (Sauðárkrókur). Samples of embryos were taken at 5-6 developmental timepoints depending on morph (see Figure 1). For each morph and developmental timepoint three biological replicates were analyzed, each of those isolated from pools of 3 embryos. The total number of samples was 48 (Figure 1). The crosses were made using multiple parents. For most samples the embryos came from crosses created in the 2010 spawning season (SB 150- 200τs, PL 140-170 τs, LB 140-200τs). For SB- and PL-charr ten males and ten females were used for the cross but five of each sex for LB. Due to poor RNA quality of some samples from early developmental stages, we added samples from the 2011 spawning season. For timepoint 100τs in PL we used a cross from the 2011 spawning season with the same setup as before. For SB at timepoints 100τs and 140τs we used two single parent crosses due to difficulties of finding running SB-charr. Samples SB100A and SB100B came from one cross and SB100C and all three samples for timepoint 140τs came from the other one.

For verification of single nucleotide polymorphism's (SNP's) and comparisons between morphs we sampled 93 sexually mature (most of them running) individuals of the four morphs; 24 SB-, 24 LB-, 24 PL- and 21 PI-charr. The fish were caught by gillnet fishing in Ólafsdrattur and Mjóanes in 2015 and 2016.

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

2.2 RNA isolation and sequencing

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

2.3 Transcriptome assembly, Abundance estimation and Annotation

The transcriptome assembly and annotation were described previously (Guðbrandsson *et al.* 2018), but briefly the reads were quality trimmed and adapters removed using Trim Galore! (version 0.3.3, Krueger 2012). The filtered reads from all samples were assembled with Trinity (version v2.1.0, Grabherr *et al.* 2011). We used kallisto (version v0.42.4, Bray *et al.* 2016) to estimate transcripts abundance. To speed up the annotation process transcripts with fewer than 200 mapped reads were not retained for annotation, another reason being that lowly expressed transcripts are unlikely to supply variants with enough coverage to surpass quality filters. The

transcripts were annotated with the Trinotate pipeline (version 2.0.2, Haas 2015). Orthologs of the transcripts in salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*) mRNA and protein sequences were found using `blastn` and `blastx` respectively (Altschul *et al.* 1990). Annotation from NCBI *Salmo salar* Annotation Release 100 (`https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Salmo_salar/100/`, retrieved 2015-12-17) and SalmoBase (Samy *et al.* 2017, `http://salmobase.org/`, version from 2015-09-18) were both used for salmon and annotation for rainbow trout came from Berthelot *et al.* (2014, `http://www.genoscope.cns.fr/trout/data/`, version from 2014-05-19). Only the best match was retained for each reference database. For further details about the annotation process and parameters we refer to Guðbrandsson *et al.* (2018).

2.4 Variant calling and quality filtering

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

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

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

The F_{PT} statistic is analogous to the classical F_{IT}, so low F_{PT} values indicate low variation between samples just as low F_{IT} values indicate low variation among individuals (excess heterozygosity). Filtering on F_{PT} is therefore similar to removing markers that deviate from Hardy-Weinberg equilibrium when genotypes are on individual basis. To determine a reasonable cutoff for F_{PT} we simulated 288 chromosomes with eight different alternative allele frequencies (0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7). Chromosomes were distributed randomly among samples with the same experimental setup as in our study and hierarchical F-statistics were calculated. For each alternative allele frequency 10.000 replicates were simulated. Based on this simulation (see figure S1) we chose F_{PT} = 0.1 as a cutoff and all variants below this value were removed. Samples with coverage of five reads or less were not included in the calculations of F-statistics.

We ran FreeBayes again on filtered variants in order to phase variants into longer haplotypes, when possible. We used a haplotype length window of 70 basepairs (bp) and only asked for biallelic variants (`-use-best-n-allels 2`). Nevertheless some variants in the output were not biallelic and therefore removed. We also removed variants if the new estimate of F_{PT} was the below the cutoff (0.1) in the new run. No further filtering of variants was performed.

2.5 Analysis of variant functionality and distribution among morphs

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

2.6 Genomic distribution of candidate variants

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

2.7 Genotyping of candidate variants in a population sample

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

3 Results

3.1 Genetic variation separating three sympatric Arctic charr morphs

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

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

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

The three morphs separated at the genetic level. The first principal component (calculated for all variants) distinguished the planktivorous morph (PL-charr) from the two benthic morphs (LB- and SB-charr), which in turn separated on the second axis (Figure 2). The first two axes explain 19.5% and 12.7% of the variation, and other axes 4.2% at the most each (data not show). To further investigate the genetic difference between morphs we studied variants with high F_{ST}-values and private alleles in each morph (note that due to the sequencing of pooled samples from sibling embryos, the F_{ST} could be inflated). Using $F_{ST} = 0.2$ as a cutoff yielded 2331 variants (Figure 3). Variants were categorized by the magnitude of the absolute allele frequency difference between one morph and the other two, and called LB-, SB- or PL-charr specific variants. For instance, the SB-specific variants, differed most strongly in allele frequency in SB-charr versus the average allele frequency of LB- and PL-charr. A significant excess (χ^2=140.3, df = 2, p < 0.0001) of variants (1174) belonged to the PL-specific category (separating PL-charr and the two benthic morphs), while 605 and 552 variants associated with SB- and LB-charr, respectively.

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

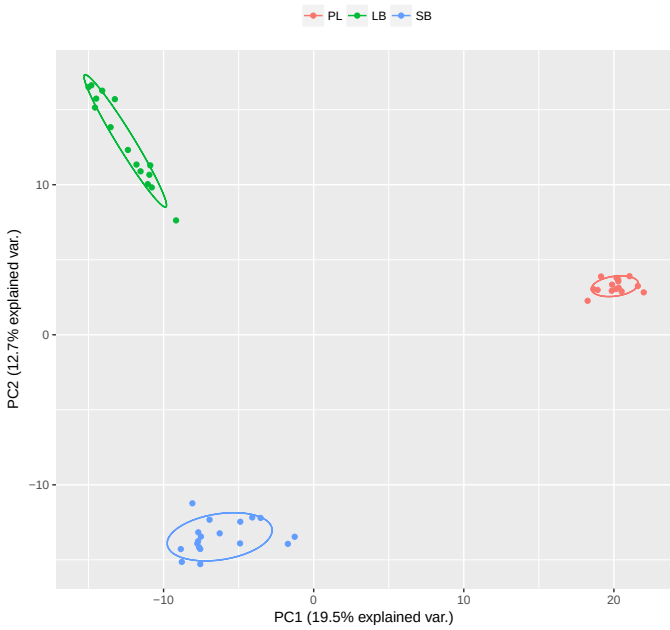


Figure 2: Genetic separation of samples from three sympatric Arctic charr morphs, based on principal component analysis of 19,252 transcriptome variants. The first and second principal components are shown with the proportion of variance explained. Individual samples are graphed (colored by color, SB: blue, LB: green and PL: red), and overlaid are 68% normal data ellipses for each morph.

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

In order to gauge if certain biological systems differed between the morphs we tested for GO-enrichment in variants associating with specific morphs. As some variants were annotated to different transcript isoforms of the same gene, testing was done both on transcripts and gene level using annotation to salmon genes. No GO categories were significant for SB-specific and LB-specific variants. However, PL-specific variants were significantly enriched in 10 GO-categories (Table 2). Nine categories were found on the transcipts level and seven on the gene level. Only one of the categories significant on gene level was not at the transcript level (*collagen fibril organization*) but it was close to being significant ($FDR = 0.015$). Also, variants were enriched in one other category related to the extracellular matrix *collagen catabolic process*. That may have led to two higher level categories to be significant at the transcript level but not at the gene level (*multicellular organismal catabolic process* and *macromolecule metabolic process*). Interestingly, four of the categories relate to environmental sensing and responses, *i.e.* light and sound, and showed a strong signal both on transcript and gene level, *e.g. inner ear morphogenesis* and *visual perception*. The final two categories were *tooth mineralization* and *odontogenesis*, both related to tooth development.

GO-enrichment tests on private variants only revealed three GO-categories related to tRNA aminoacylation significant at the transcript level for variants private to PL-charr. The signal is most likely because three transcripts are annotated to the same gene rather than being a real biological pattern (Table S4). Together, the GO analyses

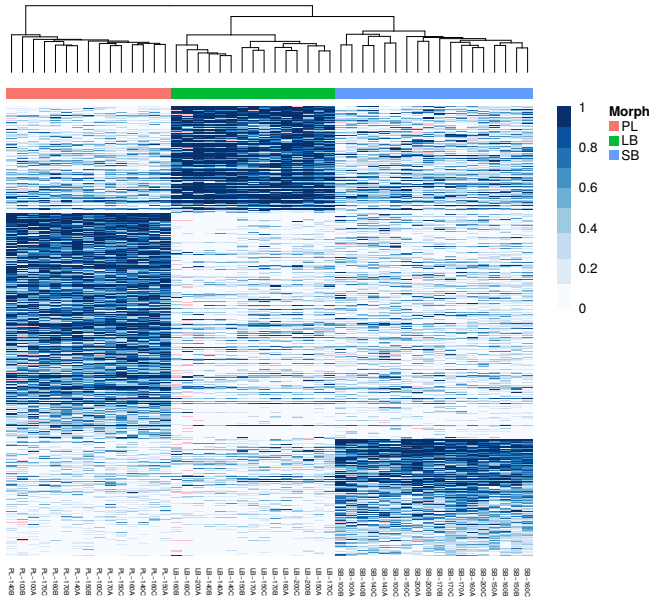


Figure 3: Separation of morphs based on the 2331 variants with F_{ST} >0.2. Each column represents a sample, name below indicates morph, developmental timepoint and biological replicate. The white-blue scale depicts allele frequency, higher frequency of the alternative allele with darker blue. Hierarchical clustering grouped samples and variants. This separated the morphs (abscissa) and similar variants by allele frequency (ordinate). A total of 1174 variants had higher frequency in PL-charr, 552 in LB- and 605 in SB-charr. Missing values are indicated by pink. Coloring of individuals by morph, SB: blue, LB: green and PL: red.

of biological functions point to genetic differences in sensing, collagen metabolism and mineralization, between benthic and limnetic morphs. This could be linked to distinct differences in the main feeding habitats and principal prey species of PL- versus LB- and SB-charr and to differences in craniofacial features of these morphs.

3.3 Genome-wide divergence of sympatric charr morphs

To explore if the genetic differences between morphs were restricted to specific chromosomal regions or more widely distributed, we mapped the variants to the Arctic charr genome using `blastn` (see methods). A large fraction (93%, 17933 of 19,252) mapped to the genome. Of the variants mapped to the genome 10,956 or 61% mapped to chromosomes but the rest mapped to unplaced scaffolds and contigs. The number of variants found for each linkage group was, as expected, related to the linkage group size (Figure S2). Markers with F_{ST} > 0.2 were found on all linkage groups, and peaks of more pronounced differentiation were found on many chromosomes. Similarly, the three types of high F_{ST} markers (PL-, SB- and LB-charr specific) were found on almost all linkage groups (Figure 4, B).

There were indications of enrichment of high F_{ST} variants of particular type in particular linkage groups, most dramatically on the mitochondrial chromosome. Previous study of genetic differentiation in the transcriptomes of SB-charr and Aquaculture charr implicated morph differences in the mtDNA (Gudbrandsson *et al.* 2016), and

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

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

genotyping of population samples revealed SNPs with high frequency derived alleles in LB- and SB-charr. In this transcriptome, we found 6 variants within the mtDNA (Table S1), and all but one had high F_{ST} (Figure 4). Consistent with previous results, these variants were LB- or SB-specific. Three markers (m1829G>A, m3211T>C and m3411C>T) within the 12s and 16s rRNA gene were observed previously (Gudbrandsson *et al.* 2016). Smaller fraction of markers had high F_{ST}'s on other linkage groups. The distribution of markers indicated distinct chromsomal regions with high differentiation associating with particular morphs. For instance a high peak of PL-specific variants was on LG18, and a peak of SB-specific markers on LG10 (Figure 4). On LG1 high F_{ST} variants of all categories (PL, SB, LB specific) were found. In sum, the genetic separation between morphs was found on all linkage groups, including strong differences in the mitochondrial DNA.

3.4 Verification of variants in population sample confirms morph separation

As the candidate variants were derived from sequenced pools of embryos from multi-parent families, we wanted to verify them in population samples from the wild to address three questions: First, do estimates of allele frequencies in the transcriptome and in wild populations correspond? Second, do the three morphs differ genetically in population samples? Third, does the understudied Piscivorous (PI) charr differ genetically from the other three morphs or could it be an ontogenetic morph of PL-charr that learn to utilize threespine sticklebacks as prey (Snorrason *et al.* 1989)?

Candidate variants (23 in total, Table S5) were chosen based on high F_{ST} values in the transcriptome sequencing, chromosomal location, biological functions and/or differential expression (Guðbrandsson *et al.* 2018). Sexually mature (some running) individuals of all four morphs in Lake Þingvallavatn were genotyped. All markers but one amplified successfully and behaved as true single nucleotide polymorphisms (SNPs, Table S5). The allele frequencies in the transcriptome and population sample were highly correlated (Kendall's $\tau = 0.71$, $p < 0.0001$) (Figure S5A). The same applied to F_{ST} values (Kendall's $\tau = 0.60$,$p < 0.001$, Figure S5B). Allele frequencies seemed to be underestimated in the transcriptome, particularly lower frequencies (< 0.3) (Figure S5A). The overall F_{ST}'s for individual markers ranged from 0.01 (*gnl3l_G795T*) to 0.66 (*gas1l_A3641C* and *wee1_T305A*) (Table S7). The correspondence of allele frequencies in the transcriptome and the population sample, suggests patterns in the former are real. Majority of markers (18/22) deviated significantly from Hardy-Weinberg equilibrium when tested on the 93 samples (all four morphs), but almost all were in HWE within each morph (Table S7). Furthermore, majority of the variance in genotype frequency associated with morph (Figure 5, first two principal components (PCs) explain 59.6% of the variance). PC1 (45.2% of variance explained) separated benthic-limnetic morphs while PC2 (14.4%) distinguished the three transcriptome morphs (SB, LB and PL-charr). The HWE tests and PC analyses both indicated genetic differentiation of the three morphs. Note, the higher proportion of variance explained by the first two PC's for the genotyped SNPs (compared to transcriptome) likely reflects the non-random choice of variants with large frequency differences for genotyping.

The phyletic relationship of the rare PI-charr to the other morphs is unknown (Gíslason 1998; Volpe & Ferguson 1996). No transcriptome data were generated for PI-charr, but 21 individuals were genotyped. Contrary to

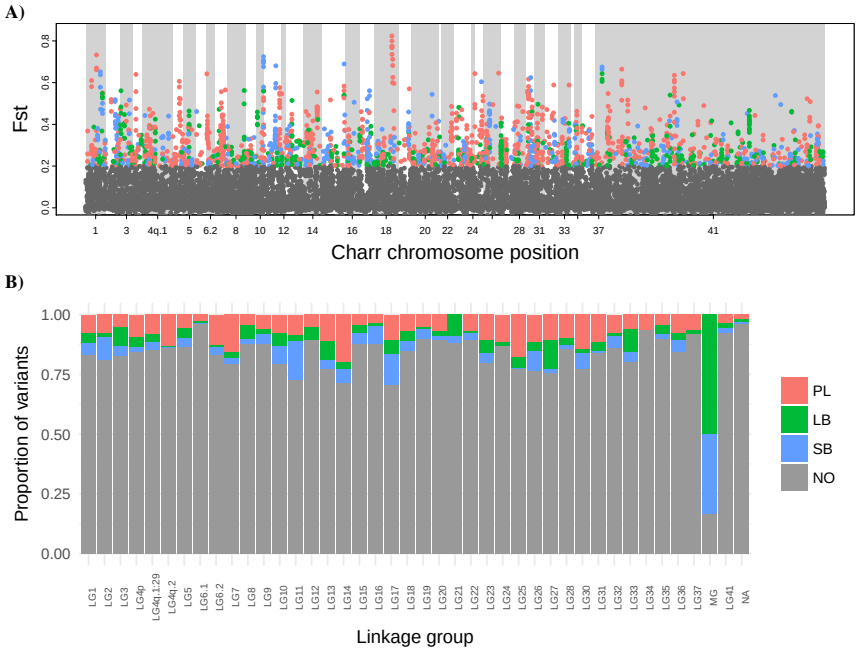


Figure 4: **A)** F_{ST} values plotted by position of variants on the charr genome from (Christensen *et al.* 2018). The colors indicate which morph differs most strongly in allele frequency from the other two for variants with F_{ST} above 0.2. Red PI, blue SB, green LB and gray represents variants with F_{ST} below 0.2. **B)** Proportion of each variant group on each linkage group (MG: mitochondrial chromosome). Unplaced scaffolds and contigs are represented by "linkage" group 41 and in **B)** NA refers to unmapped markers.

the other three morphs, PI individuals did not group in the PC plot, suggesting genetic heterogeneity. Half of the PI-charr were in the PL-cluster but the rest (except one) fell inbetween the PL- and LB-clusters. Thus while the other three morphs separated clearly genetically, the genetic status and nature of the piscivorous charr remains in doubt.

3.5 Linkage disequilibrium in the population samples and peaks of differentiation

Finally, we asked how linkage disequilibrium (LD) between markers in the populations samples associated with peaks of differentiation. We analyzed variants from the transcriptome and population samples. Variants were chosen for genotyping to i) tag distinct linkage groups and ii) study segregation of linked variants on specific linkage groups.

The genotyped variants mapped to 12 separate linkage groups and 3 unplaced scaffolds (Supplemental table S5). We picked eleven markers that separate benthic-limnetic morphs (Figure 6), but despite strong associations, no marker coupled fully to a morph or morphotype. Several putatively linked markers segregated together, for instance three markers mapping to LG18. All separated PL-charr from the benthic morphs (Figure 7 A), and while *lrrc1* and *cox11* are 2.8 Mb apart the markers showed tight LD ($r^2 > 0.72$) (Table 3). The third marker (*adk_C1075T*), located 3.4 Mb downstream of *cox11*, had weak or no LD with the other two (Table 3), indicating two differentiated regions. Note, these LD-estimates are imprecise because only 21 or 24 fish were sampled for each morph (Yan *et al.* 2009).

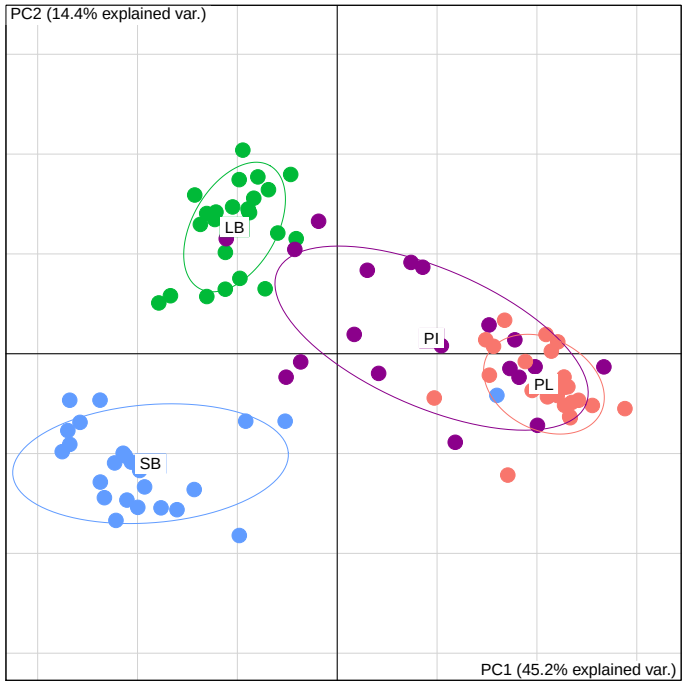


Figure 5: Genetic separation of the four morphs sympatric charr morphs depicted with principal component analyses, based on KASP genotyping of population samples. Individuals are graphed according to scores of the first two PC's on genotypes from 22 markers, along with 68% normal data ellipses for each morph (SB: blue, LB: green, PL: red and PI: purple).The PL, LB and SB-charr form distinct clusters, except one SB groups with PL.

A pair of PL-specific markers (*Kiaa1324_TC393AA* and *eif4g2b_G652A*) mapped to LG26 (Figure 7 B). As the markers are 16.8 Mb apart, very low or no association was observed between them (Table 3). The same holds for two markers in (*msi1* and *tcf15*) on LG11 that are 11.9 Mb apart and had low LD (Figure 7 C, Table 3). The region around *calm1* had numerous variants with high differentiation between PL and the two benthic morphs (Figure 7 D), but it is uncertain whether that reflects a single or many differentiated regions. Finally, three SB-specific variants (*gas1l_A3641C*, *wee1_T305A* and *dennd5a_A2555T*) were in strong LD ($r^2 > 0.45$) in all morphs, indicating long haplotypes. The variants in *dennd5a* and *wee1* mapped 118 kb apart on LG10, but on the other hand the variant in *gas1l* had the best blast hit on LG15 (The second best hit was on LG10 close to the other two, Figure 7 E). Notably, the *gas1l* variant was in linkage equilibrium with another variant on LG15 (in *tmem9b*, Figure 7 F, Table 3). The three SB-specific variants all mapped to the same chromosome in *S. salar* (ssa26) and *O. mykiss* (omy06) genome. LG10 and LG15 are ohnologous and still retain over 90% sequence similarity. This discrepancy may be due to assembly errors in the Arctic charr draft genome by Christensen *et al.* (2018), as homologous region can be tricky to assemble. Other pairs of markers mapping to different linkage groups had low LD (Table S8), the curious exception being (*cdk2ap1_A257G* and *gnl3l_G795T*) which had strong LD ($r^2 >$ 0.85) in three morphs but not PL-charr ($r^2 = 0.09$) (Table 3). These two markers map to non-homologous linkage groups (LG20 and LG1). The reason for these apparent differences in LD between morphs could be chance, sampling effects, or differences in chromosomal structure between morphs. The validated variants were often in close proximity to other variants not taken for validation with strong allele differences between morphs (Figure S6). In summary, the results suggest no single variant associated fully with specific morph, but indicate that many

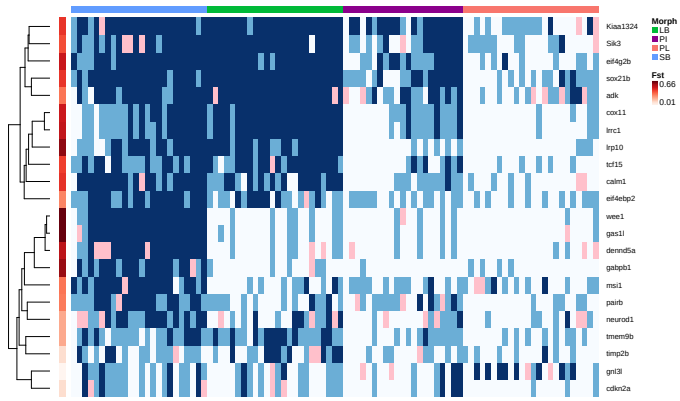


Figure 6: Heatmap of genetic variation in population samples of four charr morphs, depicting association of variants with morphology and linkage of markers. Genotypes of 22 loci were clustered by genes. F_{ST} values (shades of red) for morphs are graphed for each marker. The morphs are color coded, SB: blue, LB: green, PL: red and PI: purple, and genotypes, homozygous reference allele (white), heterozygous (light-blue) and homozygous alternate allele (blue), with pink indicating missing data.

distinct chromosomal regions harbor variants separating the morphs.

4 Discussion

The four sympatric morphs of Arctic charr in Lake Þingvallavatn differ in a host of phenotypic attributes, e.g. in adult morphology, habitat choice, feeding preferences, growth pattern, size and age at maturity (see references in Sandlund *et al.* 1992). Common garden rearing experiments indicate these differences are based on genotype but also influenced by the environment, and emerge during embryonic and juvenile development (Skúlason *et al.* 1989a; Parsons *et al.* 2010; Kapralova *et al.* 2015) While population genetics indicated genetic separation of PL-, LB- and SB-charr (Magnusson & Ferguson 1987; Volpe & Ferguson 1996), the degree of relatedness and phyletic relationships of these morphs have remained unresolved. Here we utilized SNPs from RNAseq data derived from embryos of pure crosses of three morphs (Guðbrandsson *et al.* 2018) to assess their genetic separation. For each variant we asked if one of the morphs was more diverged, and if so, how morph specific variants were distributed in the genome and whether they related to particular functional systems. Although the experimental design, pools of individuals from bulk crosses of each morph, was not optimal for these purposes, and despite the fact that the transcriptome suffered from 3'bias due to mRNA degradation (Guðbrandsson *et al.* 2018), the data yielded many informative variants. Consistently, 22 of the 23 selected variants genotyped in independent population samples were real variants. According to the data, estimates of F_{ST} in the transcriptome may have been inflated, relative to the population samples (16 out of 22 markers had lower F_{ST} in the latter). Notably, the frequencies of rare alleles were underestimated in the transcriptome as expected (Konczal *et al.* 2014), but might be exaggerated due to the sequencing of related individuals, pooling of three embryos in individual samples, or Beavis effects (Beavis 1994) because high F_{ST} SNPs were chosen for verification. These facts limit the interpretation of this dataset, precluding for example estimation of the site frequency spectrum, but do not negate the observed widespread genetic differentiation between morphs. Previously, we (Gudbrandsson *et al.* 2016) extracted variants from RNAseq data from SB-charr of Þingvallavatn and an aquaculture breeding stock derived from several populations (mainly from the north of Iceland (Svavarsson 2007) - an outgroup for Þingvallavatn charr). While that study yielded a similar number of candidate variants ($\sim$ 20,000) as were found here, the proportion of variants with high F_{ST} was larger, probably reflecting more divergence between the aquculture charr and the SB-charr than among the three Þingvallavatn morphs studied here (and consistent with monophyletic status

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

Var1	Var2	Pattern of divergence[a]	Linkage group	Genomic distance (bp)	LB	SB	PL	PI
lrrc1	*cox11*	PL	LG18	2,762,681	1.00	0.91	0.73	0.90
cox11	*adk*	PL	LG18	3,370,151	NA	0.12	0.05	0.23
lrrc1	*adk*	PL	LG18	6,132,832	NA	0.10	0.00	0.32
eif4g2b	*Kiaa1324*	PL	LG26	16,793,974	0.00	0.03	0.01	0.06
msi1	*tcf15*	SB, PL	LG11	11,914,739	0.06	0.01	0.04	0.00
dennd5a	*wee1*	SB	LG10	118,070	0.92	0.82	0.53	1.00
wee1	*gas1l*	SB	LG10[b]	[b]210,891	0.85	0.77	0.51	0.95
dennd5a	*gas1l*	SB	LG10[b]	[b]328,961	0.79	0.63	0.46	0.95
gas1l	*tmem9b*	SB, PL	LG15	400,866	0.00	0.01	0.00	0.26
cdk2ap1[c]	*gnl3l*	SB, SB	LG20 / LG1[c]		0.85	0.97	0.09	0.88

[a] In which morph did the variant differ most in frequency, first description for Var1 and second for Var2.
[b] Second best blast hit for *gas1l* was on LG10, the best was on LG15.
[c] The LD for this pair was reported because of the strong association, though the markers do not map to the same chromosome.
NA: Not available, because variant *adk* was not polymorphic in LB.

of the Lake Þingvallavatn morphs compared to other Icelandic populations (Kapralova *et al.* 2011)). It remains to to be determined how many of the variants observed in these RNAseq data are shared with other populations in Iceland and across the species range (Brunner *et al.* 2001).

4.1 Genetic separation of recently evolved sympatric morphs

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

Studying the relatedness of the morphs can shed some light on the history of colonization and divergence in Lake Þingvallavatn. The morphs (PL-, LB- and SB-charr) are more closely related to one another than to other Icelandic populations (Kapralova *et al.* 2011). The data presented here firmly reject hypothesis I, that the PL-, LB- and SB-charr morphs are formed by environmental influence on individuals of a shared gene pool (the PI-charr may be an exception, see below). The data also suggest that the benthic morphs (LB- and SB-charr) are more closely related. This is congruent with patterns of differential expression in this transcriptome (Guðbrandsson *et al.* 2018) and one previous study (Volpe & Ferguson 1996) but not others (Magnusson & Ferguson 1987; Danzmann *et al.* 1991; Gíslason 1998). This incongruence and the observed distribution of the high F_{ST} variants into PL-, LB- and SB-specific categories, may reflect limited resolution due to few markers in earlier studies, incomplete sorting of alleles or differential effects of selection on distinct traits and variants. Coalescence simulations based on microsatellite data were more consistent with a brief phase of allopatric separation of the ancestors of PL- and SB-charr, rather than sympatric evolution (Kapralova *et al.* 2011). This study considered two demographic scenarios and two parameters (migration and population size). Analyses of other scenarios, *i.e.* with changing migration rates or introgression would be most interesting (see for example Jacobs *et al.* 2018). Considering the episodic nature of ice retreat and the high geological activity in the Þingvallavatn area during the formation of the lake (Saemundsson 1992; Norðdahl *et al.* 2008) charr may have invaded the waters more than once, or experienced intromittent isolation of populations in rifts. The current study lacks a natural outgroup, and future population genomic surveys of these morphs should include populations from the neighboring geographic area, including the Þingvallavatn outflow river Sog and its downstream Lake Úlfljótsvatn. That could provide better estimates of relatedness, gene flow and history of these populations and morphs. Lake Úlfljótsvatn is particularly interesting as it also hosts landlocked morphs of charr, including forms similar to the PL- and SB-charr (Jóhannsson *et al.* 1994;

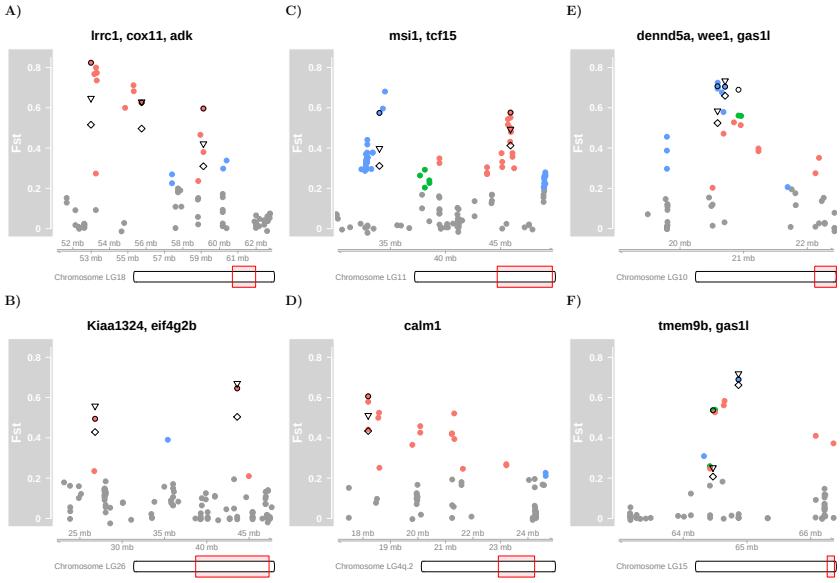


Figure 7: Detailed view of chromosomal regions and variants showing high divergence in the KASP-assay. **A)** A region on LG18 that includes *lrcc1*, *cox11* and *adk*. **B)** A large region on LG26 that includes *Kiaa1324* and *eif4g2b*. **C)** A region on LG11 that includes *msi1* and *tcf15*.**D)** Region on that includes *calm1* and many other variants with high F_{ST} values that were not validated. **E)** A region on LG10 that includes *dennd5a*, *wee1*. The F_{ST}-value from the transcriptome for the variant in *gas1* is shown at the location of the second best blast hit with a black ring. **F)** A region on LG15 that includes *tmem9b* and *gas1l*. The regions on LG10 and LG15 shown in **E)** and **F)** are homologous and some of the markers might belong better in the opposite linkage group as we suggest is the case for *gas1l*. The colored dots mark the values from the transcriptome as in Fig. 4 and the variants taken for validation are marked with a black ring (○). The triangles (▽) show the F_{ST} value from the KASP-assay for the three transcriptome morphs (PL, SB and LB) and the diamonds (◇) F_{ST} for all morphs (including PI).

Woods *et al.* 2012a).

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

the morphs.

The ambiguous genetic status of the large, rare and piscivorous morph is quite intriguing. PI-charr was only studied by genotyping 22 markers, but curiously PI individuals did not cluster as tightly as the other morphs. PI individuals grouped with LB, PL-charr and a few in between. Previous studies found PI individuals affiliated with either or both morphs (Magnusson & Ferguson 1987; Volpe & Ferguson 1996). Several models can account for this pattern. First, the classification of the morph based on phenotype may not be stringent enough and thus some LB-charr might have been misidentified as PI-charr. As the LB- and PI-charr utilize different prey and differ in spawning periods (the former in August and latter in October (Skúlason *et al.* 1989b)) this misidentification is unlikely. However, it is possible that some LB-charr males might be erroneously classified as PI because of an extended jaw-hook, if they are still running in October (some are). Secondly, the PI-charr may be more heterogeneous at the genetic level. Snorrason *et al.* (1989) postulated that PI-charr emerge as sizable PL-charr learn to eat fish. PI-charr would then be an environmentally induced morph, genetically identical to PL-charr. Thirdly, it is possible that PI-charr are genetically distinct and have either recently evolved or cross hybridize with other morphs. In this last scenario, there might be directional gene-flow from, for instance from LB-charr into PI-charr. Currently, we can not distinguish between these possibilities.

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

4.2 Genome-wide separation of sympatric morphs

How extensive is the genetic separation of morphs and how is the genetic differentiation distributed in the genome? For instance, are differentiating variants localized to few islands (Nadeau *et al.* 2012; Andrew & Rieseberg 2013; Malinsky *et al.* 2015) as stated in hypothesis II or are there multiple signals on many chromosomes (Hohenlohe *et al.* 2010; Jones *et al.* 2012) according hypothesis III?

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

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

Genomic differentiation reflects the history and evolution of groups, degree and age of separation of groups, extent of gene flow and past hybridization events (Seehausen *et al.* 2014; Shapiro *et al.* 2016). But intrinsic factors, such as the nature of the genes and their biological actions, also shape the rates of differentiation and divergence, and genomic features, such as centromeres, gene density, recombination and GC content influence estimates of

sequence divergence and allele frequency differences (Seehausen *et al.* 2014; Burri *et al.* 2015; Vijay *et al.* 2017). More population genomic data and analyses are needed to disentangle the role of positive selection and intrinsic factors on patterns of differentiation in this system. More broadly, the confirmed synteny of large genomic regions (Nugent *et al.* 2017), the range of species, subspecies and ecologically distinct populations (Klemetsen 2013; Jacobs *et al.* 2018) sets the stage for future studies of the genomic and ecological correlates of divergence and polymorphism in salmonids.

4.3 Potential mechanisms of phenotypic and developmental differences between sympatric morphs

Which genes influence differences in size, head shape, feeding apparatus, coloration, dietary preferences, parasite load, etc. between the morphs? Genetic variation in the transcriptome tags large fraction of the potentially functional regions of the genome. The genome-wide differentiation between morphs implies polygenic basis of their differences, but major genes influencing specific traits may be segregating. The GO-analyses of variants with high F_{ST}s pointed to enrichment of several categories between PL-charr and the benthic morphs. This further supports polygenic roots of morph differentiation. The genomic resolution of the current data is low, and while we can describe frequency differences in variants in specific genes, it is more likely that other linked polymorphisms/loci are actually contributing to phenotypic differences.

The developmental roots of the sympatric polymorphism are unknown, but the current data can inform future studies by bringing attention to particular chromosomal regions, genes and systems. The GO results suggest the limnetic and benthic morphs differ genetically in three systems (collagen metabolism, tooth mineralization and sensory functions). The enrichment of variants in collagen organzation/catabolism and extracellular matrix categories is consistent with observed differential expression of ECM related and cartilage remodeling genes in benthic and limnetic morphs (Ahi *et al.* 2014, 2015). Curiously, one genomic region with strong differentiation contained a variant in *calm1*. Calmodulins are broadly expressed in vertebrate tissues, including bone and articular cartilage in humans (Mototani *et al.* 2005). Notably, a microarray screen identified higher expression of *Calm1* in beak primordia of developing finches with longer beaks (Abzhanov *et al.* 2006). Perhaps most intruiging, three SB-specific variants on linkage group 10 (or 15) had high F_{ST}'s. This might represent long haplotypes with strong differentiation between morphs, and that the small benthic phenotype may be influenced by allelic variation in this region. Curiously one gene, *Wee1* is a key regulator of the timing of mitosis and cell size, first identified in fission yeast (Nurse & Thuriaux 1980) and another *gas1* (*Growth arrest-specific protein 1*) associates with embryonic cranial skeleton morphogenesis and palate development (Seppala *et al.* 2007). Furthermore, in light of altered expression of *eif4ebp1* in SB-charr, both in muscles of adult fish (Macqueen *et al.* 2011) and developing embryos (this trancriptome (Guðbrandsson *et al.* 2018)) it is curious that the frequency of a variant in this gene differed between morphs (F_{ST} = 0.28). *eif4ebp1* is involved in the insulin receptor signaling pathway and mTOR signalling (Bidinosti *et al.* 2010; Gkogkas *et al.* 2012; Banko *et al.* 2007). However, here this variant associated with benthic morphotype not small benthic charr, which is at odds with the postulation that reduced growth of the SB populations is mediated through changes in the mTOR activity. We stress the general caveat, that while markers with strong frequency differences between morphs may reside in the causative genes, its more likely that they are linked to other causative variants. Few traits have been mapped to a gene in salmonids. The exception being *vgll3* which associates with age of maturity in wild populations, with very curious sex dependent dominance of alleles (Barson *et al.* 2015; Ayllon *et al.* 2015). Unfortunately the *vgll3* gene was not transcribed in our data, but SB-specific variants with F_{ST} around 0.4 are found $\sim$1 Mb from the location of *vgll3* on LG2 (Figure S6 K). Further work is needed to check if variants in *vgll3*, *tulp4* (see Larson *et al.* 2017) or the genes highlighted here contribute to phenotypic differences in charr populations. QTL and/or fine scale association mapping (Zimmerman *et al.* 2000; Palsson & Gibson 2004; Dworkin *et al.* 2005) are needed to find variants that associate with body size, growth rate and trophic traits in these sympatric charr.

4.4 Conclusion and future perspective

We mined genetic variation of the sections of the genome more related to function, from three of the four morphs in Lake Þingvallavatn to address questions about genetic separation, genome-wide differentiation of morphs and potential functionality of loci separating them. We formulated three hypothesis about the putative causes of morph differences: Hypothesis I stated that the morphs are environmentally induced. According to hypothesis II, ecological specialization led to few genomic island of strong genetic differentiation between morphs, with low background separation across the genome. Hypotheses III postulated substantial differentiation across the genome

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

5 Acknowledgments

Special thanks to present and past members of the Arctic charr group at the University of Iceland for support, comments and suggestions. Thanks to Ben Sutherland and Eric Normandeau for excellent advice and open communication about MapComp and methods for placing variants on linkage maps. This project was supported by The Icelandic Center for Research (RANNIS #100204011) to SSS and coworkers, The University of Iceland Doctoral Fund to JG and University of Iceland research fund to AP, SSS and ZOJ.

6 Author contributions

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

References

Abouheif E, Favé MJ, Ibarrarán-Viniegra AS, Lesoway MP, Rafiqi AM, Rajakumar R (2014) Eco-Evo-Devo: The Time Has Come. In *Ecological Genomics - Ecology and the Evolution of Genes and Genomes* (edited by CR Landry, N Aubin-Horth), chap. 6, pp. 107–125. Springer Netherlands, Dordrecht. 2

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, **442**, 563–567. 2, 18

Adams CE, Wilson AJ, Ferguson MM (2008) Parallel divergence of sympatric genetic and body size forms of Arctic charr, *Salvelinus alpinus*, from two Scottish lakes. *Biological Journal of the Linnean Society*, **95**, 748–757. 16

Ahi EP, Gudbrandsson J, Kapralova KH, *et al.* (2013) Validation of Reference Genes for Expression Studies during Craniofacial Development in Arctic Charr. *PloS ONE*, **8**, e66389. 3

Ahi EP, Kapralova KH, Pálsson A, *et al.* (2014) Transcriptional dynamics of a conserved gene expression network associated with craniofacial divergence in Arctic charr. *EvoDevo*, **5**, 40. 3, 18

Ahi EP, Steinhäuser SS, Pálsson A, *et al.* (2015) Differential expression of the aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr. *EvoDevo*, **6**, 27. 3, 18

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the Evolution of Salmonid Fishes. In *Evolutionary Genetics of Fishes* (edited by BJ Turner), pp. 1–53. Springer US. 4, 6, 8

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology*, **215**, 403–10. 6

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–2482. 17

Ayllon F, Kjærner-Semb E, Furmanek T, *et al.* (2015) The *vgll3* Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (*Salmo salar* L.) Males. *PLOS Genetics*, **11**, e1005628. 18

Banko JL, Merhav M, Stern E, Sonenberg N, Rosenblum K, Klann E (2007) Behavioral alterations in mice lacking the translation repressor 4E-BP2. *Neurobiology of Learning and Memory*, **87**, 248–256. 18

Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT (2011) BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, **27**, 1691–1692. 6

Barson NJ, Aykanat T, Hindar K, *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**, 405–408. 18

Beavis WD (1994) The power and deceit of QTL experiments: Lessons from comparitive QTL studies. In *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference* (edited by DB Wilkinson), pp. 250–266. American Seed Trade Association, Washington DC. 14

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

Bernatchez L, Renaut S, Whiteley AR, *et al.* (2010) On the origin of species: Insights from the ecological genomics of lake whitefish. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 1783–800. 2, 3

Berthelot C, Brunet F, Chalopin D, *et al.* (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications*, **5**, 3657. 3, 4, 6

Bidinosti M, Ran I, Sanchez-Carbente MR, *et al.* (2010) Postnatal Deamidation of 4E-BP2 in Brain Enhances Its Association with Raptor and Alters Kinetics of Excitatory Synaptic Transmission. *Molecular Cell*, **37**, 797–808. 18

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

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001) Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **55**, 573–586. 15

Burri R, Nater A, Kawakami T, *et al.* (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, **25**, 1656–1665. 18

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

Cauwelier E, Gilbey J, Sampayo J, Stradmeyer L, Middlemas SJ (2017) Identification of a single genomic region associated with seasonal river return timing in adult Scottish Atlantic salmon (*Salmo salar* L.) identified using a genome-wide association study. *Canadian Journal of Fisheries and Aquatic Sciences*, pp. cjfas–2017–0293. 17

Christensen KA, Rondeau EB, Minkley DR, *et al.* (2018) The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLOS ONE*, **13**, e0204076. 3, 4, 7, 12, 13

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157. 2

Danzmann RG, Cairney M, Davidson WS, *et al.* (2005) A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). *Genome*, **48**, 1037–51. 4

Danzmann RG, Ferguson MM, Skulason S, Snorrason SS, Noakes DLG (1991) Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, *Salvelinus alpinus* L., from Thingvallavatn, Iceland. *Journal of Fish Biology*, **39**, 649–659. 3, 15

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

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

Dworkin I, Palsson A, Gibson G (2005) Replication of an *Egfr*-wing shape association in a wild-caught cohort of *Drosophila melanogaster*. *Genetics*, **169**, 2115–2125. 18

Foote AD, Vijay N, Ávila-Arcos MC, *et al.* (2016) Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications*, **7**, 11693. 2

Frandsen F, Malmquist HJ, Snorrason SS (1989) Ecological parasitology of polymorphic Arctic charr, *Salvelinus alpinus* (L.), in Thingvallavatn, Iceland. *Journal of Fish Biology*, **34**, 281–297. 3

Fruciano C, Franchini P, Kovacova V, *et al.* (2016) Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nature Communications*, **7**, 12736. 2

Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483–2497. 17

Garduño-Paz MV, Adams CE, Verspoor E, Knox D, Harrod C (2012) Convergent evolutionary processes driven by foraging opportunity in two sympatric morph pairs of Arctic charr with contrasting post-glacial origins. *Biological Journal of the Linnean Society*, **106**, 794–806. 16

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint*, **7**. 6

Gíslason D (1998) *Genetic and Morphological Variation in Polymorphic Arctic Charr* Salvelinus Alpinus*, from Icelandic Lakes*. Master's thesis, The University of Guelph. 3, 11, 15

Gkogkas CG, Khoutorsky A, Ran I, *et al.* (2012) Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature*, **493**, 371–377. 18

Gordeeva NV, Alekseyev SS, Matveev AN, Samusenok VP, Taylor E (2015) Parallel evolutionary divergence in Arctic char *Salvelinus alpinus* complex from Transbaikalia: Variation in differentiation degree and segregation of genetic diversity among sympatric forms. *Canadian Journal of Fisheries and Aquatic Sciences*, **72**, 96–115. 3, 16

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

Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, **29**, 644–52. 5

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

Grant PR, Grant BR (2002) Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science*, **296**, 707–711. 1

Gudbrandsson J, Ahi EP, Franzdottir SR, *et al.* (2016) The developmental transcriptome of contrasting Arctic charr (*Salvelinus alpinus*) morphs [version 3; referees: 2 approved, 1 approved with reservations]. *F1000Research*, **4**. 3, 4, 10, 11, 14, 17

Guðbrandsson J, Franzdóttir SR, Kristjánsson BK, *et al.* (2018) Differential gene expression during early development in recently evolved and sympatric Arctic charr morphs. *PeerJ*, **6**, e4345. 2, 3, 4, 5, 6, 7, 8, 11, 14, 15, 18

Gudmundsdóttir R, Kornobis E, Kristjánsson BK, Pálsson S (2017) Genetic analysis of ciliates living on the groundwater amphipod *Crangonyx islandicus* (Amphipoda: Crangonyctidae). *Acta Zoologica*, **99**, 188–198. 15

Guerrero RF, Hahn MW (2017) Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, pp. 1–7. 2

Haas BJ (2015) Trinotate: Transcriptome Functional Annotation and Analysis. 6

Haas BJ, Papanicolaou A (2015) TransDecoder (Find Coding Regions Within Transcripts). 7

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

Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM (2013) Evaluating Adaptive Divergence Between Migratory and Nonmigratory Ecotypes of a Salmonid Fish, *Oncorhynchus mykiss*. *G3; Genes|Genomes|Genetics*, **3**, 1273–1285. 17

Harris LN, Chavarie L, Bajno R, *et al.* (2015) Evolution and origin of sympatric shallow-water morphotypes of Lake Trout, *Salvelinus namaycush*, in Canada's Great Bear Lake. *Heredity*, **114**, 94–106. 17

Hartl DL, Clark AG (2006) *Principles of Population Genetics, Fourth Edition*. Sinauer Associates, Inc., Sunderland, Massachusetts, USA. 6

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

Hendry AP, Bolnick DI, Berner D, Peichel CL (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, **75**, 2000–2036. 2

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

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**. 17

Jacobs A, Carruthers M, Yurchenko A, *et al.* (2018) Convergence in form and function overcomes non-parallel evolutionary histories in Arctic charr. *bioRxiv*, p. 265272. 3, 15, 16, 17, 18

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

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

Jones FC, Chan YF, Schmutz J, *et al.* (2012) A Genome-wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-Pair Divergence in Sticklebacks. *Current Biology*, **22**, 83–90. 2, 17

Jonsson B, Skúlason S, Snorrason SS, *et al.* (1988) Life History Variation of Polymorphic Arctic Charr (*Salvelinus alpinus*) in Thingvallavatn, Iceland. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**, 1537–1547. 3

Kapralova KH (2008) *Evolutionary Origins of Widespread Small Benthic Charr (*Salvelinus Alpinus*) over Different Geographic Scales in Iceland*. MSc., University of Iceland, Reykjavík, Iceland. 3

Kapralova KH (2014) *Study of Morphogenesis and miRNA Expression Associated with Craniofacial Diversity in Arctic Charr (*Salvelinus Alpinus*) Morphs*. Ph.D. thesis, University of Iceland, Reykjavík, Iceland. 19

Kapralova KH, Franzdóttir SR, Jónsson H, Snorrason SS, Jónsson ZO (2014) Patterns of MiRNA Expression in Arctic Charr Development. *PLoS ONE*, **9**, e106084. 3

Kapralova KH, Gudbrandsson J, Reynisdottir S, *et al.* (2013) Differentiation at the *MHCII*α and *Cath2* Loci in Sympatric *Salvelinus alpinus* Resource Morphs in Lake Thingvallavatn. *PLoS ONE*, **8**, e69402. 3, 19

Kapralova KH, Jónsson ZO, Palsson A, *et al.* (2015) Bones in motion: Ontogeny of craniofacial development in sympatric arctic charr morphs. *Developmental Dynamics*, **244**, 1168–1178. 3, 14

Kapralova KH, Morrissey MB, Kristjánsson BK, Ólafsdóttir GA, Snorrason SS, Ferguson MM (2011) Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity*, **106**, 472–87. 3, 5, 15, 19

Klemetsen A (2010) The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes. *Freshwater Reviews*, **3**, 49–74. 2, 16

Klemetsen A (2013) The Most Variable Vertebrate on Earth. *Journal of Ichthyology*, **53**, 781–791. 2, 18

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

Kornobis E, Pálsson S, Kristjánsson BK, Svavarsson J (2010) Molecular evidence of the survival of subterranean amphipods (Arthropoda) during Ice Age underneath glaciers in Iceland. *Molecular ecology*, **19**, 2516–30. 15

Kristjánsson BK, Skúlason S, Snorrason SS, Noakes DLG (2012) Fine-scale parallel patterns in diversity of small benthic Arctic charr (*Salvelinus alpinus*) in relation to the ecology of lava/groundwater habitats. *Ecology and evolution*, **2**, 1099–112. 3, 15

Krueger F (2012) Trim Galore! 5

Larson WA, Limborg MT, McKinney GJ, Schindler DE, Seeb JE, Seeb LW (2017) Genomic islands of divergence linked to ecotypic variation in sockeye salmon. *Molecular Ecology*, **26**, 554–570. 17, 18

Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079. 6

Lien S, Koop BF, Sandve SR, *et al.* (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature*, **533**, 200–205. 4

Limborg MT, Larson WA, Shedd K, Seeb LW, Seeb JE (2018) Novel RAD sequence data reveal a lack of genomic divergence between dietary ecotypes in a landlocked salmonid population. *Conservation Genetics Resources*, **10**, 169–171. 17

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature*, **457**, 830–836. 1

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

Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132881. 4

Macqueen DJ, Kristjánsson BK, Paxton CGM, Vieira VLA, Johnston IA (2011) The parallel evolution of dwarfism in Arctic charr is accompanied by adaptive divergence in mTOR-pathway gene expression. *Molecular ecology*, **20**, 3167–84. 3, 18

Magnusson KP, Ferguson MM (1987) Genetic analysis of four sympatric morphs of Arctic charr, *Salvelinus alpinus*, from Thingvallavatn, Iceland. *Environmental Biology of Fishes*, **20**, 67–73. 3, 14, 15, 17

Malinsky M, Challis RJ, Tyers AM, *et al.* (2015) Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, **350**, 1493–1498. 2, 17

Malmquist HJ, Snorrason SS, Skulason S, Jonsson B, Sandlund OT, Jonasson PM (1992) Diet Differentiation in Polymorphic Arctic Charr in Thingvallavatn, Iceland. *The Journal of Animal Ecology*, **61**, 21–35. 3

Markevich G, Esin E, Anisimova L (2018) Basic description and some notes on the evolution of seven sympatric morphs of Dolly Varden *Salvelinus malma* from the Lake Kronotskoe Basin. *Ecology and Evolution*, **8**, 2554–2567. 2, 16

Marteinsson VT, Rúnarsson A, Stefánsson A, *et al.* (2013) Microbial communities in the subglacial waters of the Vatnajökull ice cap, Iceland. *ISME Journal*, **7**, 427–437. 15

May-Mcnally SL, Quinn TP, Woods PJ, Taylor EB (2015) Evidence for genetic distinction among sympatric ecotypes of Arctic char (*Salvelinus alpinus*) in south-western Alaskan lakes. *Ecology of Freshwater Fish*, **24**, 562–574. 16

Moghadam HK, Ferguson MM, Danzmann RG (2011) Whole genome duplication: Challenges and considerations associated with sequence orthology assignment in Salmoninae. *Journal of Fish Biology*, **79**, 561–574. 6, 8

Morris M, Rogers SM (2014) Integrating Phenotypic Plasticity Within an Ecological Genomics Framework: Recent Insights from the Genomics, Evolution, Ecology, and Fitness of Plasticity. In *Ecological Genomics - Ecology and the Evolution of Genes and Genomes* (edited by CR Landry, N Aubin-Horth), chap. 5, pp. 73–105. Springer Netherlands, Dordrecht. 2

Mototani H, Mabuchi A, Saito S, *et al.* (2005) A functional single nucleotide polymorphism in the core promoter region of *CALM1* is associated with hip osteoarthritis in Japanese. *Human molecular genetics*, **14**, 1009–17. 18

Muir AM, Hansen MJ, Bronte CR, Krueger CC (2016) If Arctic charr *Salvelinus alpinus* is 'the most diverse vertebrate', what is the lake charr *Salvelinus namaycush* ? *Fish and Fisheries*, **17**, 1194–1207. 2, 16

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

Nadeau NJ, Whibley A, Jones RT, *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 343–353. 17

Noakes DLG (2008) Charr truth: Sympatric differentiation in *Salvelinus* species. *Environmental Biology of Fishes*, **83**, 7–15. 2, 16

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

Nosil P (2012) *Ecological Speciation*. Oxford series in ecology and evolution. Oxford University Press, Oxford ; New York. 1

Nugent CM, Easton AA, Norman JD, Ferguson MM, Danzmann RG (2017) A SNP Based Linkage Map of the Arctic Charr (*Salvelinus alpinus*) Genome Provides Insights into the Diploidization Process After Whole Genome Duplication. *G3; Genes|Genomes|Genetics*, **7**, 543–556. 4, 18

Nurse P, Thuriaux P (1980) Regulatory genes controlling mitosis in the fission yeast *Schizosaccharomyces pombe*. *Genetics*, **96**, 627–637. 18

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

Palsson A, Gibson G (2004) Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*. *Genetics*, **167**, 1187–1198. 18

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

Pardo-Diaz C, Salazar C, Baxter SW, *et al.* (2012) Adaptive Introgression across Species Boundaries in *Heliconius* Butterflies. *PLoS Genetics*, **8**, e1002752. 2

Parsons KJ, Sheets HD, Skúlason S, Ferguson MM (2011) Phenotypic plasticity, heterochrony and ontogenetic repatterning during juvenile development of divergent Arctic charr (*Salvelinus alpinus*). *Journal of Evolutionary Biology*, **24**, 1640–1652. 3

Parsons KJ, Skúlason S, Ferguson M (2010) Morphological variation over ontogeny and environments in resource polymorphic arctic charr (*Salvelinus alpinus*). *Evolution and Development*, **12**, 246–257. 3, 14

Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLOS Biology*, **14**, e1002379. 2

Perreault-Payette A, Muir AM, Goetz F, *et al.* (2017) Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior. *Molecular Ecology*, **26**, 1477–1497. 17

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

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

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

Samy JKA, Mulugeta TD, Nome T, *et al.* (2017) SalmoBase: An integrated molecular data resource for Salmonid species. *BMC Genomics*, **18**, 482. 6

Sandlund OT, Gunnarsson K, Jónasson PM, *et al.* (1992) The arctic charr *Salvelinus alpinus* in Thingvallavatn. *Oikos*, **64**, 305–351. 4, 14, 16

Sandlund OT, Jonsson B, Malmquist HJ, *et al.* (1987) Habitat use of arctic charr *Salvelinus alpinus* in Thingvallavatn, Iceland. *Environmental Biology of Fishes*, **20**, 263–274. 3

Seehausen O, Butlin RK, Keller I, *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192. 2, 5, 17, 18

Seehausen O, Wagner CE (2014) Speciation in Freshwater Fishes. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 621–651. 1, 3

Seppala M, Depew MJ, Martinelli DC, Fan CM, Sharpe PT, Cobourne MT (2007) *Gas1* is a modifier for holoprosencephaly and genetically interacts with sonic hedgehog. *Journal of Clinical Investigation*, **117**, 1575–1584. 18

Shapiro BJ, Leducq JB, Mallet J (2016) What Is Speciation? *PLOS Genetics*, **12**, e1005860. 17

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

Skúlason S, Noakes DL, Snorrason SS (1989a) Ontogeny of trophic morphology in four sympatric morphs of arctic charr *Salvelinus alpinus* in Thingvallavatn, Iceland. *Biological Journal of the Linnean Society*, **38**, 281–301. 3, 14

Skúlason S, Smith TB (1995) Resource polymorphisms in vertebrates. *Trends in Ecology & Evolution*, **10**, 366–370. 2, 3

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

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

Skúlason S, Snorrason SS, Noakes DLG, Ferguson MM, Malmquist HJ (1989b) Segregation in spawning and early life history among polymorphic Arctic charr, *Salvelinus alpinus*, in Thingvallavatn, Iceland. *Journal of Fish Biology*, **35**, 225–232. 3, 5, 16, 17

Smith TB, Skúlason S (1996) Evolutionary Significance of Resource Polymorphisms in Fishes, Amphibians, and Birds. *Annual Review of Ecology and Systematics*, **27**, 111–133. 2, 16

Snorrason SS, Skúlason S (2004) Adaptive Speciation in Northern Freshwater Fishes. In *Adaptive Speciation* (edited by U Dieckmann, M Doebeli, JAJ Metz, D Tautz), chap. 10, pp. 210–228. Cambridge University Press, Cambridge. 2, 3

Snorrason SS, Skúlason S, Sandlund OT, Malmquist HJ, Jonsson B, Jonasson PM (1989) Shape polymorphism in Arctic charr, *Salvelinus alpinus*. *Physiology and Ecology Japan*, **1**, 393–404. 11, 17, 19

Sutherland BJG, Gosselin T, Normandeau E, *et al.* (2016) Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps. *Genome Biology and Evolution*, **8**, evw262. 4

Svavarsson E (2007) árangur í kynbótum á bleikju og næstu skref. In *Fræðaþing Landbúnaðarsins 4*, pp. 121–125. 14

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

Timusk ER, Ferguson MM, Moghadam HK, Norman JD, Wilson CC, Danzmann RG (2011) Genome evolution in the fish family salmonidae: Generation of a brook charr genetic map and comparisons among charrs (Arctic charr and brook charr) with rainbow trout. *BMC genetics*, **12**, 68. 4

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

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

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

Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JB (2017) Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology*, **26**, 4284–4295. 2, 18

Vincent B, Dionne M, Kent MP, Lien S, Bernatchez L (2013) Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation. *Evolution*, **67**, 3469–3487. 17

Volpe JP, Ferguson MM (1996) Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. *Molecular ecology*, **5**, 763–72. 3, 11, 14, 15, 17, 19

Wilson AJ, Gíslason D, Skúlason S, *et al.* (2004) Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular ecology*, **13**, 1129–42. 2, 16

Wolf JBW, Ellegren H (2016) Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, **18**, 87–100. 2, 5

Woods PJ, Skúlason S, Snorrason SS, Kristjánsson BK, Malmquist HJ, Quinn TP (2012a) Intraspecific diversity in Arctic charr, *Salvelinus alpinus*, in Iceland: I. Detection using mixture models. *Evolutionary Ecology Research*, **14**, 973–992. 16

Woods PJ, Skúlason S, Snorrason SS, Kristjánsson BK, Malmquist HJ, Quinn TP (2012b) Intraspecific diversity in Arctic charr, *Salvelinus alpinus*, in Iceland: II. Which environmental factors influence resource polymorphism in lakes? *Evolutionary Ecology Research*, **14**, 993–1013. 2

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic Characterization and Linkage Disequilibrium Estimation of a Global Maize Collection Using SNP Markers. *PLoS ONE*, **4**, e8451. 12

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome biology*, **11**, R14. 7

Zimmerman E, Palsson A, Gibson G (2000) Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics*, **155**, 671–83. 18

7 Supplement

7.1 Supp-Tables and files

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

In Dropbox: variants.tsv

id: Identifier for each variant from 1 to 19,252
transcript_id: Name of the transcript from the Trinity assembly.
start: Start position of polymorphism within contig.
end: End position of polymorphism within contig.
ref: Base(s) of the reference allele.
alt: Base(s) of the alternative allele.
var_pos: Position of polymorphism within contig.
NCBI_id: Sequence identifier for chromosome or scaffold in genome.
Chr: Name of chromosome.
Chr_pos: Position mapped to in chromosome.
Freq_LB: Frequency of alternative allele within the LB morph.
Freq_PL: Frequency of alternative allele within the PL morph.
Freq_SB: Frequency of alternative allele within the SB morph.
Fst: Estimate of F_{ST} between morphs.
Fpt: Estimate of F_{PT} or variation among samples (pools). See methods for further explanation.
effect: Which effect does the mutation have on the gene, 3'- and 5' UTR indicate mutations in those regions, synonymous do not change the reading frame but mutations that do so are coded as transitions from one amino acid to another (single letter a.a. code)
gene_name: Gene name based on SalmonBase annotation.

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

In Dropbox: var_freebayes_final.vcf

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

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

Continued on Next Page...

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

Transcript id	Var pos	Ref	Alt	Chr	Chr pos	Freq	F_{ST}	F_{PT}	Effect	Gene name
T165162c16g1i1	975;984	A(N8)A	T(N8)C	LG1	17,578,870	0.33	0.13	0.38	UTR	TOX high mobility group box family member 2-like, isoform X4
T168092c2g1i5	2724;2726	TTA	ATT	LG7	26,670,845	0.35	0.23	0.51	UTR	45 kDa calcium-binding protein
T168398c0g1i4	1759	G	T	LG32	20,522,655	0.33	0.28	0.35	P to H	Kinesin-like protein
T168651c0g1i1	863	A	G	LG1	17,539,258	0.36	0.30	0.40		alpha-1-syntrophin-like, isoform X1
T168962c1g1i2	994	G	C	LG14	7,560,825	0.35	0.24	0.38	S to T	translocating chain-associated membrane protein 2-like
T169544c12g4i1	442	A	C	LG41	294,246	0.48	0.34	0.53	UTR	CIGSSA_056585.t2
T170480c1g1i1	2202	A	T	LG4q.2	9,848,462	0.34	0.16	0.49		CIGSSA_044851.t3
T170529c17g1i1	176	A	T	LG36	28,810,075	0.40	0.31	0.49		integral membrane protein 2B-like, isoform X1
T171302c3g1i1	521	C	T	LG17	8,116,418	0.42	0.24	0.39		Glutamine–tRNA ligase
T171326c3g1i2	1206	A	T	LG11	21,367,985	0.33	0.15	0.48	UTR	Peptidylarginine deiminase hypothetical
T171686c7g2i1	193	G	C	LG17	8,312,053	0.49	0.39	0.50		Zgc:110709
T171686c7g2i1	665	G	C	LG17	8,311,589	0.53	0.42	0.58		Zgc:110709
T171686c7g2i1	982	T	A	LG17	8,311,272	0.50	0.43	0.50		Zgc:110709
T171686c7g2i1	1424	T	G	LG17	8,310,831	0.42	0.33	0.46		Zgc:110709
T172162c0g1i1	425	T	A	LG7	16,736,407	0.47	0.35	0.54		TAR DNA binding protein
T172217c4g4i1	1238	T	A	LG32	6,053,989	0.53	0.36	0.57	UTR	WW domain-containing adapter protein with coiled-coil
T172217c4g4i1	2113	A	G	LG32	6,054,862	0.58	0.47	0.58	UTR	WW domain-containing adapter protein with coiled-coil
T172357c4g1i1	2197	A	C			0.42	0.41	0.38	Syn,UTR	coronin-7-like
T172967c3g1i1	199	G	C	LG35	8,330,659	0.51	0.35	0.51		RAB11a member RAS oncogene family
T172992c11g4i1	448	T	C	LG14	14,078,902	0.57	0.44	0.60	Syn	Beta-taxilin|hypothetical
T172933c13g1i1	2089	C	T	LG7	16,959,001	0.48	0.33	0.47		calsyntenin 1, isoform X4
T173769c10g1i2	742;743;748	TATTTTT	ATTTTTG	Scf1321	226,642	0.34	0.18	0.49		Inosine-5-monophosphate dehydrogenase 1
T174499c6g3i3	1150	A	G	LG7	16,917,732	0.50	0.38	0.53	S to G	ubiquitination factor E4B, UFD2 homolog (S. cerevisiae), isoform X1
T174814c3g1i1	411	T	A	LG4q.1:29	37,790,310	0.33	0.24	0.47		CMP-N-acetylneuraminate-poly-alpha-28-sialyltransferase
T174821c6g1i1	967	C	G	LG36	31,081,180	0.41	0.35	0.40		unconventional myosin-Ib-like, isoform X1
T174821c6g1i1	1199	G	A	LG36	31,080,948	0.37	0.23	0.36		unconventional myosin-Ib-like, isoform X1
T175009c8g1i2	2905	C	T	LG30	8,090,556	0.47	0.34	0.45	UTR	ATP-binding cassette sub-family F member 1-like
T175236c1g3i4	5881	T	G	LG31	24,323,360	0.61	0.45	0.56	UTR	collagen alpha-1(XVI) chain-like, isoform X1
T175348c1g10i1	258	A	G	LG20	21,264,718	0.32	0.18	0.40		ubiquitin domain containing 2
T175403c90g7i1	1245	A	C	Scf396	147,836	0.73	0.55	0.83	UTR	peptidyl-prolyl cis-trans isomerase FKBP3-like
T175479c1g2i5	407	A	C	LG35	8,345,621	0.50	0.38	0.50	V to G	Ubiquilin-4
T176050c1g1i1	5019	C	A	LG35	8,316,952	0.52	0.44	0.51	UTR	RNA-binding protein MEX3A
T210338c0g1i1	1772	G	A	LG17	8,125,541	0.50	0.36	0.48	Syn	glutaminyl-tRNA synthetase
Private to SB										
T145511c0g1i1	2002	T	C	Scf4114	11,937	0.26	0.26	0.25	Syn	Myelin expression factor 2|hypothetical
T160725c0g1i2	919	A	G	LG17	22,513,650	0.40	0.41	0.37	Syn	transcription factor 15-like
T162598c0g1i1	729;742	G(N12)C	A(N12)T	Scf1046	1,457	0.44	0.23	0.44		chromobox homolog 6
T165516c2g3i4	382	C	A	Scf4168	5,044	0.27	0.19	0.46	UTR	
T165516c2g3i4	873;876;912;924;927	CCAC(N35)A-(N10)AAAT	TCAT(N35)T-(N10)GAAC	LG18	24,320,993	0.43	0.38	0.65	Syn	
T168835c3g2i2	1142	T	A	LG4q.1:29	21,965,763	0.33	0.33	0.37	UTR	Phosphoglycerate mutase 2
T169028c43g1i1	3476	A	C	LG32	2,199,330	0.31	0.29	0.40	UTR	Muscle-related coiled-coil protein

Continued on Next Page...

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

Transcript id	Var pos	Ref	Alt	Chr	Chr pos	Freq	F_{ST}	F_{PT}	Effect	Gene name
T169667c0g1i4	2617	A	G	LG18	12,429,904	0.28	0.30	0.32	F to S	DNA topoisomerase 2-alpha
T169667c0g1i4	2835	G	A	LG18	12,430,772	0.28	0.31	0.32	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3054	G	T	LG18	12,431,454	0.28	0.30	0.30	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3198	A	G	LG18	12,431,726	0.26	0.30	0.30	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3375	A	G	LG18	12,432,251	0.29	0.31	0.31	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3447	G	A	LG18	12,432,602	0.29	0.31	0.36	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	3672	T	C	LG18	12,433,037	0.28	0.32	0.32	Syn	DNA topoisomerase 2-alpha
T169667c0g1i4	4308	C	A	LG18	12,435,579	0.27	0.29	0.32	Syn	DNA topoisomerase 2-alpha
T170294c11g5i4	1744	T	C	LG17	22,247,612	0.33	0.31	0.32	Syn	ras-related protein Rap-1A-like, isoform X1
T173826c6g1i1	342	A	G	LG28	18,758,190	0.36	0.32	0.43		Si:dkey-77p13.2
T210331c21g1i1	4709	G	A	MG	3,411	0.74	0.68	0.79	UTR	

*Validated in KASP-assay

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

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

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

In Dropbox: Supplement_KaspMarkersSequencesv2.ods

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

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

In Dropbox: Datafile_ResultSameinad7.csv

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

In Dropbox: Supplement_KaspFstHW.ods

Gene_SNP: Name of variant
Total (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests, on the entire dataset
LB (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for LB
PI (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for PI
PL (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for PL
SB (P): Significance of the test for Hardy Weinberg proportions as estimated from Fisher's exact tests for SB
Fst: F-statistics for variation between morphs
Fit: F-statistics for variation between individuals
Fis: F-statistics for variation within population

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

In Dropbox: SuppLDAll.ods

7.2 Supp-Figures

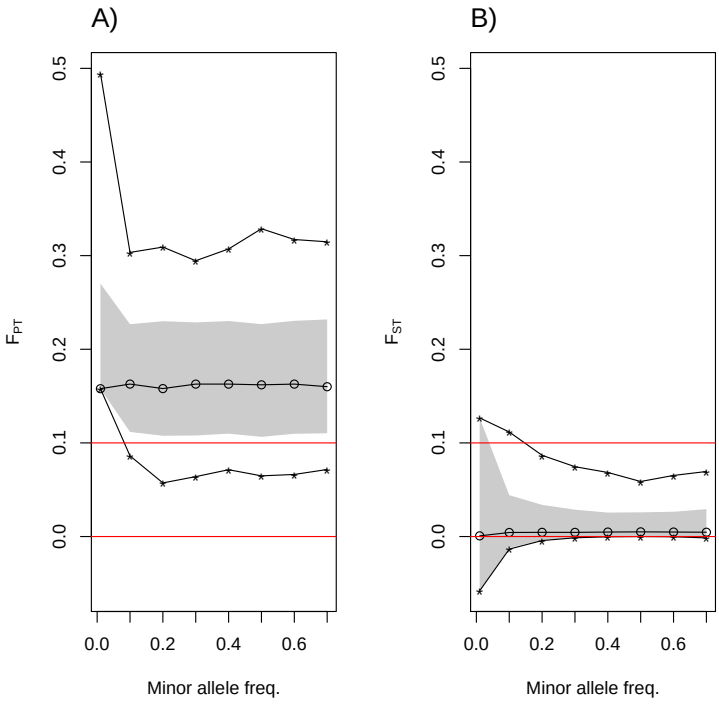


Figure S1: Results from simulations on F-statistics for a biallelic variant with the same allele frequency among morphs. The gray area indicate the 95% confidence area. The median, minimum and maximum values are shown by dots and lines. A) Shows F_{PT} values and B) F_{ST} values. As expected F_{PT}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

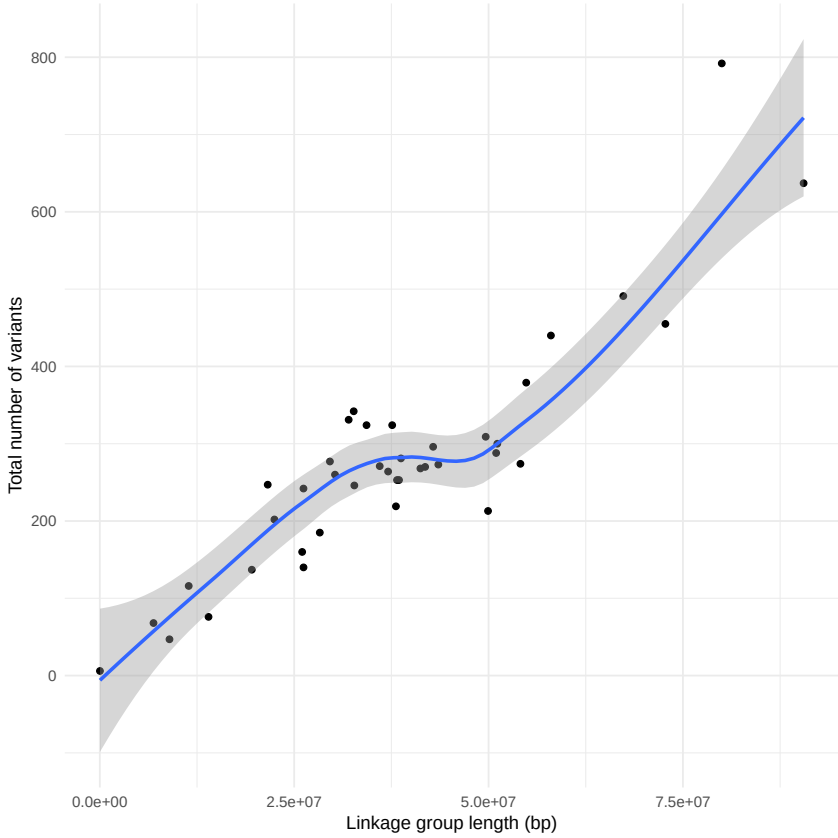


Figure S2: Total number of transcriptome variants by linkage group length. A loess smooth curve (blue) with 95% confident interval (gray) is also graphed.

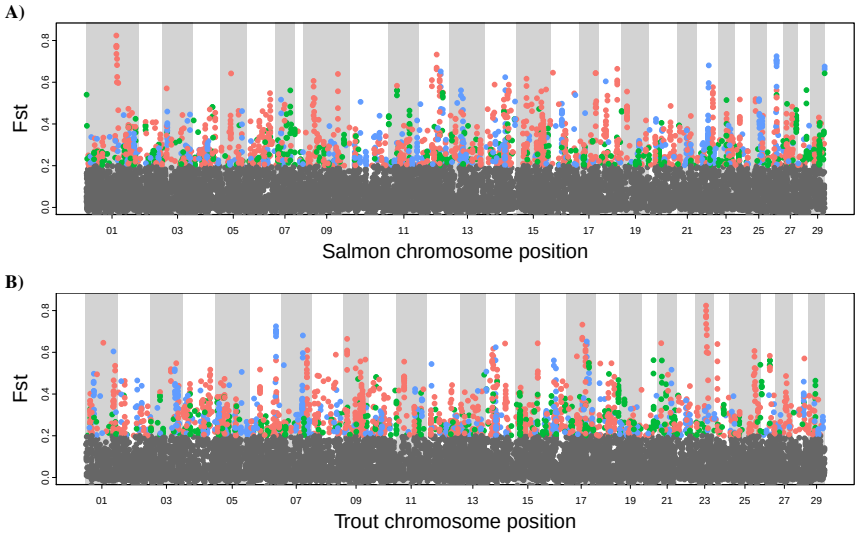


Figure S3: F_{ST} values plotted by position of variants on the salmon **A)** and rainbow trout **B)** genome. The colors indicate which morphs differs most strongly in allele frequency from the other two for variants with F_{ST} >0.2. Red PL, blue SB and green LB

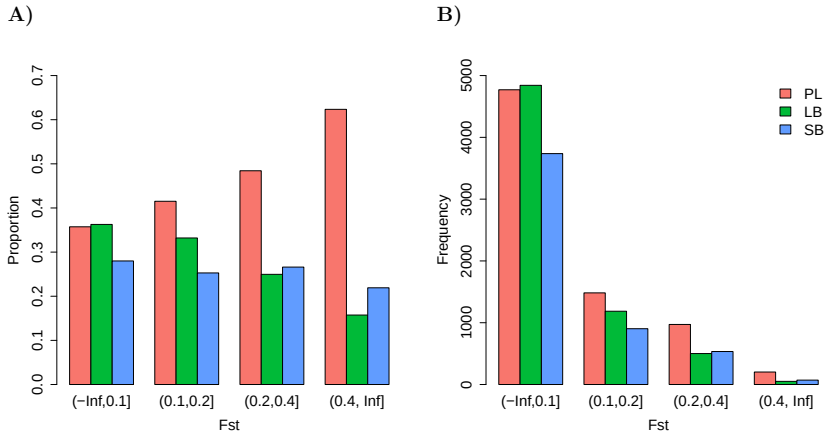


Figure S4: **A)** Proportion of variants (within F_{ST} groups) for different F_{ST}-values grouped by the morph with highest deviation in allele frequency from the other two. **B)** The number of variants in each F_{ST} category grouped by the morph with highest deviation in allele frequency from the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

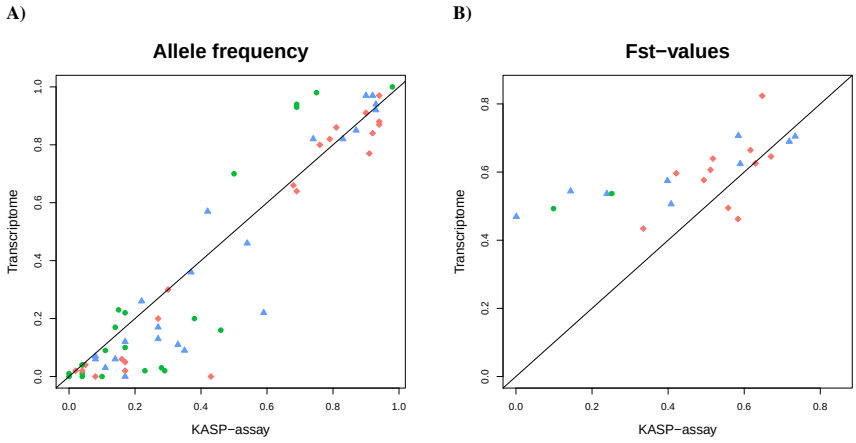


Figure S5: Tight relationship was found between the allele frequencies and F_{ST} values estimated from the transcriptome and the population genetic sample (Kasp assay). **A)** The allele frequencies for the 22 markers were estimated for each morph (color coded) and the estimates from the two methods show high positive correlation (Kendall's $\tau = 0.71$, Pearson's $r = 0.94$, $p < 0.0001$). The diagonal line represents the 1:1 relationship. **B)** F_{ST}-values calculated for the transcriptome and the population sample had comparatively weaker association (Kendall's $\tau = 0.60$, Pearsons's $r = 0.67$, $p < 0.001$). Notably, five markers deviated from the 100% relationship (diagonal line), with higher F_{ST} in the transcriptome compared to the population sample (due to underestimation of some rarer allele frequencies in the transcriptome). The colors indicate which morphs shows the highest deviation from the other two in mean allele frequency for each marker in the transcriptome.

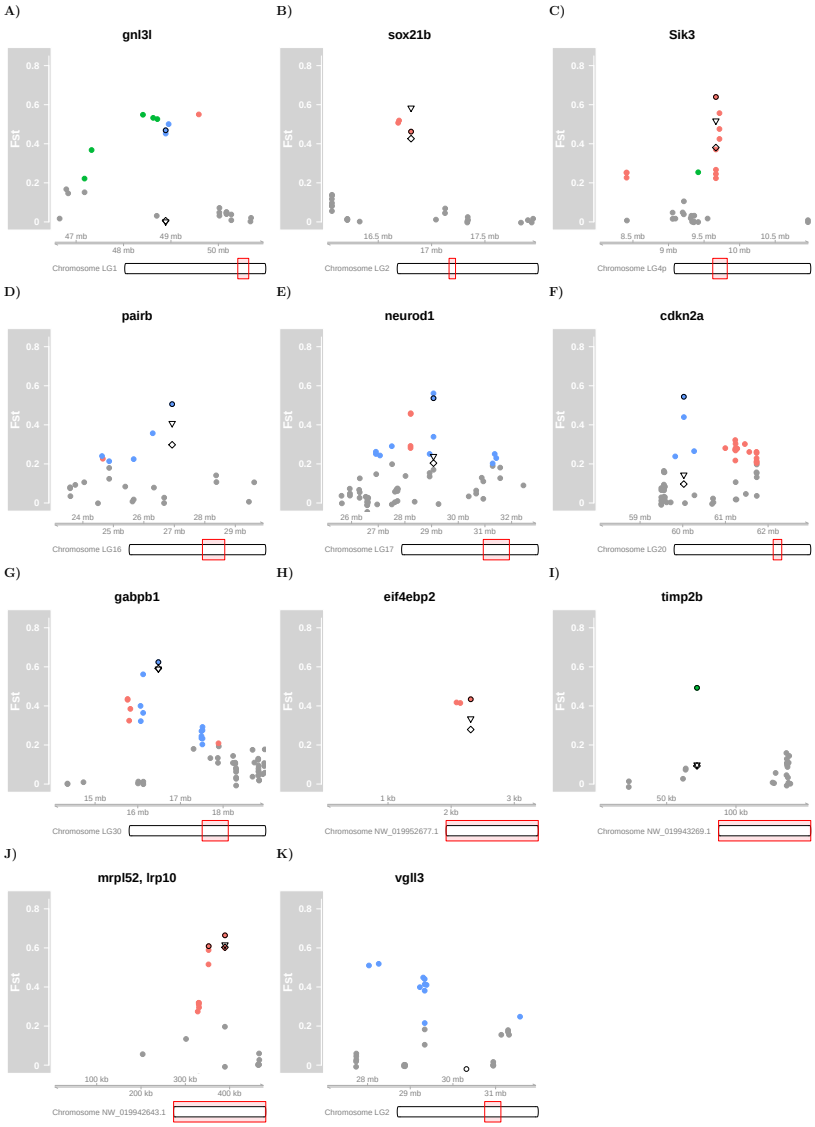


Figure S6: **A-J)** Detailed view of chromosomal regions of variants in KASP-assay and nearby transcriptome variants (not-validated). The colored dots mark F_{ST} values from the transcriptome as in Fig. 4 and the variants taken for validation are marked with a black circl (○). The triangles (▽) show the F_{ST} value from the KASP-assay for the three transcriptome morphs (PL, SB and LB) and the diamonds (◇) F_{ST} for all morphs (including PI). **K)** Detailed view of variants nearby the *vgll3* locus. The circle (○) represents the location of the *vgll3* gene, not transcribed and thus no variants detected, using the same color code for variants.