Morphological genes in Icelandic Arctic charr

Ragnar Óli Vilmundarson

Líf- og umhverfisvísindadeild
Háskóli Íslands
2011

Morphological genes in Icelandic Arctic charr

Ragnar Óli Vilmundarson

12 eininga ritgerð sem er hluti af
Baccalaureus Scientiarum gráðu í Líffræði

Leiðbeinandi
Arnar Pálsson

Líf- og umhverfisvísindadeild
Verkfræði- og náttúruvísindasvið
Háskóli Íslands
Reykjavík, Maí 2011

Morphological genes in Icelandic Arctic charr
12 eininga ritgerð sem er hluti af *Baccalaureus Scientiarum* gráðu í Líffræði

Líf- og umhverfisvísindadeild
Verkfræði- og náttúruvísindasvið
Háskóli Íslands
Askja, Sturlugötu 7
101 Reykjavík

Sími: 525 4000

Skráningarupplýsingar:
Ragnar Óli Vilmundarson, 2011, *Morphological genes in Icelandic Arctic charr*, BS ritgerð, Líf- og umhverfisvísindadeild, Háskóli Íslands, 47 bls.

Prentun: Háskólaprent
Reykjavík, Maí 2011

Útdráttur

Það eru til fjögur afbrigði af Íslenskri bleikju (*Salvelinus alpinus*) í Þingvallavatni. Tvær þeirra dvergbleikja og murta eru áhugaverðar til rannsóknar. Þær sýna einkennandi breytileika í vexti, æviferli, svæðisvali og fæðuöflun. Í tengslum við þennan vaxtarbreytileika lagði ég fram tilgátuna að vaxtarmunur á milli bleikjuafbrigða í haus, bein og brjóskmyndun hefur undirliggjandi genatengsl sem munu sjást í genum tengdum vaxtarferlum. Kannaður var breytileiki á genum tengdum þessum ferlum. Beitt var raðgreiningu til að kanna breytileika í genum. Af þeim 11 genum sem náðist að raðgreina (*Etbr2, Otx2, Bmp4, Runx1, Pth, Eng2, Fgfr4, Fgop2, Igfbp1, Twist2 og Wnt9*) skáru tvö gen sig úr *Fgop2* og *Pth*. Við frekari athugun kom í ljós að *Pth* genið sem talið vaðr að væri parathyroid hormon geníð reyndist vera peptidyl-tRNA hydrolase sem ekki er þekkt álitsgen fyrir vaxtarferla. Peptidyl-tRNA hydrolase sem hefur sömu skammstöfun var því raðgreint fyrir mistök. Munur reyndist þó á milli bleikjuafbrigðanna í *Pth* geninu þar sem 7 basa úrfelling/innskot í útröð sem fannst í dvergum. Fjöldi raðgreindra einstaklinga (n=12) er svo takmarkaður að það er ekki hægt að fullyrða að þessi breytileiki finnist ekki í murtu. Til þess þarf að raðgreina fleiri einstaklinga. *Fgop2* sýndi líka breytileika fyrir tveggja basa úrfellingu sem var í innröð sem fannst í báðum afbrigðum en einungis murtan var arfhrein fyrir úrfellinguna; en það virtist ekki vera marktækur munur á samsætutíðni á milli afbrigða e.t.v. vegna þess að fjöldinn er lítil (n=39). Breytileiki milli bleikjuafbrigða fannst í tveim genum af 11 sem raðgreind voru. Hugsanleg tengsl við mun í vaxtarfelum eru til staðar en frekari rannsókna er þörf til að sannreyna það.

Abstract

There are four morphs of the Icelandic Arctic charr (*Salvelinus alpinus*) in Thingvallavatn lake. Two of those morphs, dwarf benthic charr and small pelagic charr, are of particular interest and have been selected for study. They show key morphological, life history, niche and foraging difference. The differences pertaining to morphology are of interest and so I propose that these morphological differences have an underlying genetic component and that said variability will be seen within genes that are responsible for head, cartilage and bone development. Candidate genes for these morphological pathways were selected and viable genes were sequenced. Out of the 11 that were sequenced (*Etbr2, Otx2, Bmp4, Runx1, Pth, Eng2, Fgfr4, Fgop2, Igfbp1, Twist2 og Wnt9*) two showed variability, *Fgop2* and *Pth*; although *Pth* was meant to be parathyroid hormone due to unforeseen similarities in gene nomenclature the gene amplified was peptidyl-tRNA hydrolase. *Pth* was found to have a 7 base deletion/insertion in an exon present only in dwarf benthic, although the data is too limited (n=12) to prove that this variation is not in small pelagic. The *Fgop2* gene had a deletion of 2 bases within an intron but there was no significant deviation in allele frequency between the two morphs; although more individuals need to be sequenced (currently n=39). Therefore variation between Arctic charr morphs was present in two of the 11 genes sequenced however to prove that there is a genetic component to morphological differences between morphs further research is required.

Table of contents

Figures

Tables

Acknowledgements

To start with I would like to show my appreciation to Rannís as I would have been unable to get started on this research project without their Nýsköpunarsjóð grant, the knowledge and experience I have accumulated during this project will be a big help in the future and I hope their resources prove to be a wise investment.

I would like to express my gratitude to Arnar Pálsson who has given me good, solid advice throughout my time on this project which always made me bring my goals into sharp focus and I would not have gotten as far as I did without him.

Sigrún Reynisdóttir was there to help me with the day to day workings of the lab and I would like to thank her for always having time to assist me with whatever problems arose and teaching me the techniques that allowed me to accomplish the goals of my project.

Also I would like to recognize the help that Dagmar Ýr Arnardóttir gave me throughout my project, often giving sound advice and taking the time to help me even during the busiest of days.

To all the people I have worked with during my time on this project I would like to express my appreciation for all you have done for me and thank you for making this project so enjoyable.

1 Introduction

1.1 Understanding evolution: general questions

There is a lack of a clear understanding of the connections between the genetic factors of evolution and adaptation and ecological forces. Studies had generally been skewed towards adult forms and concentrated less on developmental changes however this has changed in the last 20 to 30 years with more of a focus on early development. Although traits influenced at those time periods have, to a certain degree, become well known during this time, there is a general lack of information about factors that impact evolution at later stages, such as regulatory systems that are likely to change during adaptation. Although the theory of evolution has itself taken great evolutionary steps in the 152 years since its first appearance in Darwin's "On the Origin of Species" (Brakefield 2005), with the inclusion of the fields of genetics and developmental biology, there is still this lack of a wider perspective on evolution as stated above. To increase the understanding of how adaptation affects species and their developmental systems then a good model species needs to be used. One that has variable forms which share both genetic and life history and so any differences displayed between them are most likely due to selective pressure and by identifying those pressures then it is possible to gain a greater understanding of evolutionary mechanisms (Skulason and Smith 1995). The availability of niches within an occupied environment and a reduction in interspecies competition can influence resource polymorphism and thereby also effect speciation as diversification of a species, in relation to habitat and foraging, can lead to segregation which in turn can result in reproductive isolation and the end result would be divergence into new and separate species (Snorrason et al. 1994). Divergence can occur through either macro- or microevolution, where microevolution is a slow progression through many little changes and macroevolution can occur through a large single leap. Knowledge on the function of morphological evolution, both molecular and genetic, is limited especially with regards to vertebrates.

1.2 Arctic charr: a model for evolutionary study

Several vertebrate species have been put forward as good models for the study of rapid adaptive radiation and the species utilized in this study is the Arctic charr (*Salvelinus alpinus*). This fish species belongs to the Salmonidae family and is found in both fresh and salt water. My focus is on the freshwater Arctic charr that have the most northerly distribution of any freshwater fish today and can be found in lakes and rivers in Canada, Iceland, and Scotland (UNEP/GRID-Arendal Maps 2010). To be more specific I am interested in studying the two differing morphotypes (hereby shortened to morphs) that are named the dwarf benthic charr (ísl. dvergbleikja) and the small pelagic charr (ísl. murta) found in Lake Thingvallavatn (Snorrason et al. 1994). Pelagic refers to the part of the lake where the fish is found, which is in the water column and therefore not near the bottom or the shore. The pelagic charr are also referred to as planktivorous charr, which is descriptive of the plankton they feed on. It is also called limnetic as it resides in the open, well lit part of the water. The word benthic used for the dwarf benthic charr describes the fish that live

at the bottom of the lake, this includes the sedimentary layer and the region of water that lies directly above it.

For studying the Icelandic population of Arctic charr not only is there a readily available set of DNA samples but also the Icelandic Arctic charr populations show an extreme case of morph variability specifically those located within the Icelandic lake Thingvallavatn although the variation certainly isn't unique to that location (Gíslason et al. 1999). The samples are taken from Thingvallavatn and it was chosen because it displays an extreme case of morph variability with 4 morphs present in the lake (see Figure 1.1) and I will be studying the two that show the most deviated phenotypes. The differences between the morphs vary greatly between different Icelandic lakes, with some lakes (e.g. Galtaból) showing large morph variation, in both morphology and phenotype, while other lakes (Vatnshlíðarvatn and Stóra Viðarvatn) show little variation (Gíslason et al. 1999). One of the lakes, called Hólmavatn, only had a single morph and a study of variation at five microsatellite loci (*Ssa85DU, Str60INRA, Sfo18UL, Ssa20.19UCG* and *Omy301UoG* (Gíslason et al. 1999) found no significant deviation from Hardy-Weinberg equilibrium (HWE) in these markers. The Hardy-Weinberg principle states that the allele and genotype frequencies in a population should remain constant from generation to generation unless outside forces interfere; those forces being mutations, selective mating, meiotic drive, a limited population size, overlapping generations, random genetic drift, gene flow and evolutionary selection (Futuyma 2009). Significant deviation from this can be an indication of evolutionary selection and so, in the context of the single morph in Hólmavatn, the lack of deviation from HWE means that within a single morph population little variation is found, at least for these markers, and so this suggests that when comparing HWE between two or more morphs the other data would be showing variation between the actual morphs and not just a variation that was present within the entire Arctic charr population (Gíslason et al. 1999). Also in lakes with a number of different morphs, the variation was of different scales with some morphs showing greater differences between them than with others.

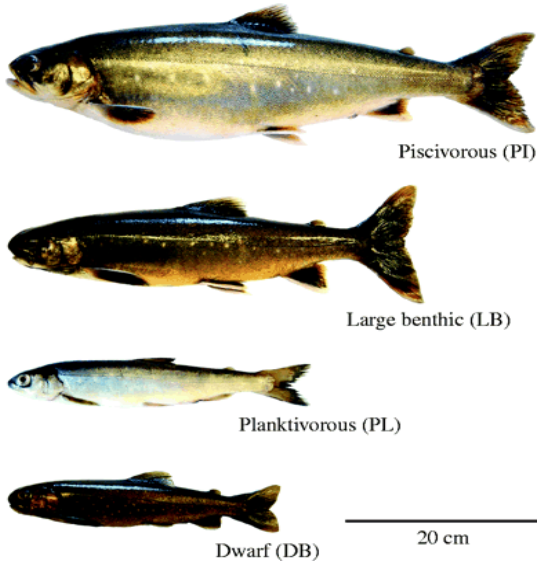

Figure 1.1 The four Arctic charr morphs that are found in Thingvallavatn. Dwarf (DB) and planktivorous (PL) are the morphs I am interested in, with the large benthic (LB) being used as a reference population (reproduced from Johnston et al. 2004).

In their study of the Thingvallavatn charr morphs, Kapralova and colleagues found genetic differentiation between the morphs (taken from five sites, see Figure 1.2), as had been seen in previous studies, however the level of genetic differentiation was low in Thingvallavatn compared to the total level found across Iceland (Fst of 0.030 compared to 0.234) (Kapralova et al. 2011). Although lower, the level of genetic differentiation was still significant between the dwarf benthic and the small pelagic, but the level was similar between all four morphs that are present in Thingvallavatn. Patterns of allele frequencies were found to be morph specific in 8 out of 10 markers, with a single allele only found in the benthic charr within all the populations studied in the lake, however it wasn't a prevalent allele (0,9 to 5,2% prevalence).

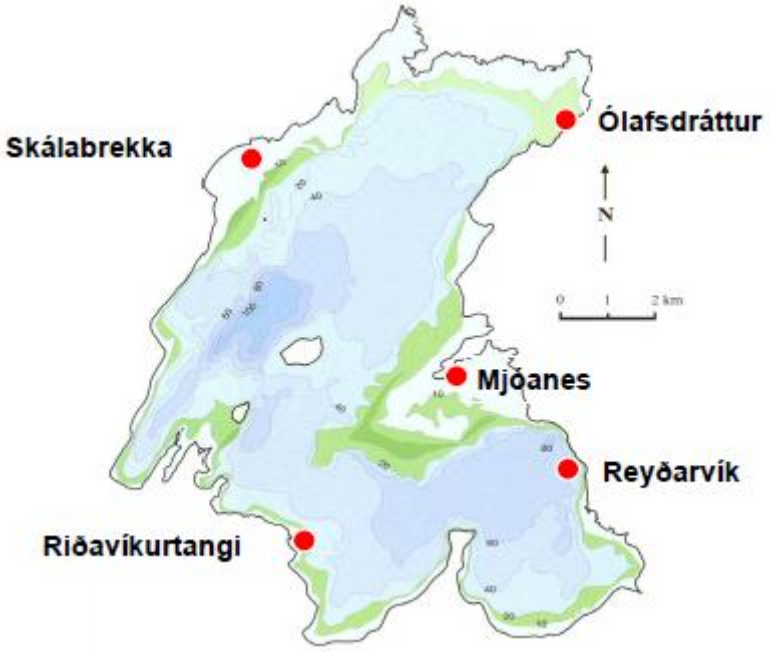

Figure 1.2 The Icelandic lake Thingvallavatn and the Arctic charr sampling locations are marked with red dots (reproduced from Kapralova et al. 2011).

1.3 Morph variation: the benthic and limnetic forms of Arctic charr

This variability in Arctic charr, seen in the different morphs present within the species, has developed since the end of the last ice age which occurred around ten to fifteen thousand years ago (Bernatchez and Landry 2003). It has been shown that fish species invading lake habitats that opened after the end of the last ice age, often occupy two niches, limnetic and benthic (Skúlason and Smith 1995). By invading these post glacial lakes the fish have gained access to an environment with little competition, due to the lack of other fish species, and abundant resources with a variety of different niches open to the fish (Gíslason et al. 1999). My focus on just two of the four Arctic charr morphs that are found within Thingvallavatn is due to influence from a research project by Kalina H. Kapralova (Kapralova et al. 2011) and so the DNA samples were available. The dwarf benthic morph, found in many Icelandic lakes, was selected for analysis because of its specialized feeding apparatus and head that is the most divergent from that which is seen in the other morphs (Kapralova et al. 2011).

The segregation of the limnetic and benthic groups is thought to be sustained by differences in their foraging behaviour, where the limnetic morphs show more dietary variation in relation to prey size than the benthic morphs (Jónsson and Skúlason 2000). This variation in the morphs diet may be related to the difference that is witnessed in the plasticity of foraging behaviour between the two morph types, with the limnetic displaying greater plasticity (Jónsson and Skúlason 2000). The differences in feeding behaviour between morphs appear early in life which gives the impression that this variable is important to the morph segregation (Jónsson and Skúlason 2000). The high morphological plasticity witnessed in the charr morphs may be the result of their variable diet, as has been shown to be the case in the limnetic threespine stickleback (*Gasterosteus aculeatus*) (Day et al. 1994). These two groups, limnetic and benthic are common evolutionary forms that are seen with divergence of lake species and this suggests that the important factors in the evolution of the morphs are niche variation and limiting resources (Gíslason et al. 1999). The variation seen between these forms is not limited to the Arctic charr, with similar patterns seen in threespine stickleback (see Figure 1.3) and such evolutionary divergence, in relation to food resources, has been observed in many other animal groups, most prominently in bird species e.g. the African finch (Smith 1987, Grant and Grant 2002).

Figure 1.3 The benthic and limnetic forms of the threespine stickleback fish. Staining highlights the bone. Benthic fish are larger, with a more stocky body and fewer skeletal lateral plates in comparison to what is seen in the limnetic fish (reproduced from Peichel et al. 2001).

Also, the Arctic charr morphs differ in key facial and cranial characteristics (Sibthorpe et al. 2006) as shown in Figure 1.4 below. The two pelagic morphs have terminal mouths, meaning that their lower and upper jaws are proportionally equal (Snorrason et al. 1994). The form of the fish's mouth is an indicator of its likely feeding behaviour and the fish with terminal mouths are either picking at their prey or chasing them. However the two benthic morphs have sub-terminal mouths that face upwards with the lower jaw being greater in length when compared with the upper jaw. The morphological differences that are listed above are adaptations to their environment with relation to habitat and resources. The benthic morphs have adapted to life at the bottom of the lake, with their sub-terminal jaw being an indicator for that behaviour and allowing for more efficient feeding on the lake floor that's made up of rough lava substrates. The pelagic morphs on the other hand have adapted to the open water with a more streamlined form (Skúlason et al. 1989). The variations seen between the morphs for fin length and body shape are connected to their feeding behaviour, with benthic features allow for selective feeding which is required for feeding at the bottom. The pelagic on the other hand has a more streamlined form with

larger fins that presumably enhances swimming ability and results in a speedier approach to feeding, necessary for feeding in open water and the chasing that such behaviour entails (Skúlason et al. 1989). Also, it has been postulated that the overabundance of one food source in tandem with a lack of competition can result in the differentiation of single species, as can be seen with the charr morphs located in Vatnshlíðarvatni (Jónsson and Skúlason 2000). This could have led to the adaptation of the benthic morphs to become a feeding specialist and the pelagic morphs to have more generalized feeding habits. This differentiation of feeding habits is proposed to have made the coexistence of morphs possible, with the general feeding habits of one morph allowing for the specialized morph to survive solely on one food source (Wilson and Yoshimura 1994).

Figure 1.4 A more detailed picture of the head morphology of the four Arctic charr morphs found in Thingvallavatn. The morphs are as follows, a) planktonivorous (PL), b) piscivorous (PI), c) dwarf benthic (DB) and d) large benthic (LB). The horizontal lines on the bottom left of each morph are the scale, which is 1cm (reproduced from Sibthorpe et al. 2006).

1.4 Morph divergence: sympatric or allopatric?

Two main hypotheses have been put forward to explain how the different Arctic charr morphs originated, with one stating that the evolutionary divergence resulting in morphs occurred within lakes, termed sympatric; and the other hypothesis states that they diverged in different lakes, termed allopatric, and then different morphs invaded into the same lake from different origins (Volpe and Ferguson 1996). The most parsimonious hypothesis, meaning the one that makes the fewest assumptions, is that sympatric divergence resulted in morph creation however the sympatric speciation hypothesis has many theoretical hurdles to overcome and the cases that support it have been called unconvincing by critics (Futuyma 2009). Also a balance must be struck between how parsimonious a hypothesis is and its likelihood. Considering these two hypotheses there was either divergence with secondary or primary contact between the morphs, i.e. contact either after or during divergence (Gíslason et al. 1999).

Gíslason and colleagues used neighbour joining to establish the evolutionary relationships between different morphs in several Icelandic lakes and showed that morphs from different

lakes do not cluster together; it is rather morphs from the same lakes that are clustering together (Gíslason et al. 1999) which supports the sympatric divergence hypothesis i.e. that divergence occurred within the same lake, or at least the same water system (lake and attached streams). Three Icelandic lakes (Vatnshlíðarvatn, Stóra Viðarvatn and Galtaból) have shown results that indicate the occurrence of sympatric speciation, with Galtaból giving the strongest indication with the highest significant Fst score (Gíslason et al. 1999). The fixation index shortened to Fst is a measure of allele frequency variation among populations that ranges from 0, indicating no variation between the populations, to 1, which indicates that populations have different fixed alleles (Futuyma 2009). Further evidence in support of the sympatric divergence hypothesis can be found in the study of the different morphs mitochondrial DNA. Results support primary contact between the morphs however secondary contact cannot be ruled out until it is known that the morphs of different lakes do not have private alleles in common (Gíslason et al. 1999). West-Eberhard also put forward a hypothesis that the polymorphism seen in the Icelandic Arctic charr, which is partially due to phenotypic plasticity, is present because it is the early stage of speciation (West-Eberhard 2003).

1.5 Morph characteristics: key differences

The morphs are highly variable in many characteristics; including morphology, rearing behaviour, life history, growth rate (Noakes 2008), the degree of phenotypic plasticity (Hindar and Jonsson 1993) and their susceptibility to varying parasites (Frandsen, Malmquist and Snorrason 1989); the morphological differences are mainly in cranial development which is what this study will be focusing on. The majority of these differences between the morphs have been shown to be independent of the age, gender, maturity level and/or size of the fish (Jónsson and Skúlason 2000). By inhabiting different areas, or niches, of the same lake, it is thought that environmental factors have influenced their morphological and genetic differences, at least in connection to their variability in size, development and eating habits (Noakes 2008). These differences have been studied in detail in the lake Vatnshlíðarvatn and the two morphs found therein differ in body size and diet, however their diets didn't show any significant variation trends regarding size (Jónsson and Skúlason 2000). So it would appear that their dietary difference is independent from their size difference which develops at an early stage in their lives. This has been confirmed with a separate study that reared wild artic charr morphs progeny (the four found in Thingvallavatn) in a common tank environment (Skúlason et al. 1996). It was shown that the large benthic and piscivorous morphs grew on average to be larger than their planktivorous and dwarf benthic counterparts. This then proves that there is a genetic factor in the body size and sexual maturity variation seen between the morphs. Private alleles are another factor that can be studied to determine whether or not a species is diverging through evolution. These private alleles are confined to a single environment, which in this case are individual lakes, and if those are present, which they have been seen to be in the Icelandic Arctic charr (Gíslason et al. 1999), then it is an indication for divergent evolution. The life history variances are seen in the sexual maturation of the morphs in Vatnshlíðarvatn with the benthic morph maturing younger and at a smaller size, with the pelagic morphs maturing when they are older and thereby larger (Jónsson and Skúlason 2000). Therefore it is possible that natural selection has been, or is, acting on these characteristics and that these differences can affect the allelic variability in genes, specifically those genes connected to development of bone and cartilage and genes that

influence (along with nutrition) the overall body size and developmental genes that control the form of the morph.

The reasons behind this high level of variation within the Arctic charr species has been hotly debated for several decades. A few hypotheses have been put forth to explain this variation but there are still arguments for and against each of them, and no firm conclusion on the matter has been reached. An argument has been made that the Arctic charr species simply has a lot of phenotypic plasticity and it is their environment that determines their phenotype and being raised in a different environment would alter their phenotype accordingly (see Noakes 2008). Another explanation is that these morphs are possibly going through the initial steps of speciation and morph differentiation and there is substantial evidence that this could be the case and that it is more than just the environment that is causing the extreme variability seen between the Arctic charr morphs (Noakes 2008). This was seen by generating hybrids (of juvenile brook charr *Salvenlinus fontinalis)*, the result of crossing two distinct morphs and by studying phenotypes of the F1 and F2 generations it becomes possible to evaluate the role of the environment and genetics (see Noakes 2008). The environment can be taken out of the equation by raising the hybrids in a common tank.

1.6 A model for behavioural research: The brook charr *Salvelinus fontinalis*

Behavioural differences between morphs have been studied in some detail and are related to which niche they are occupying. A study detailing the differing tactics of juvenile brook charr, in connection to foraging for food and socializing, revealed two types of foraging tactic (Noakes 2008). Noakes described the two types as "movers" and "stayers" and they occupied different niches in the water when foraging and so their movement also differed. The movers, as the name indicates, were more active when feeding and took prey from either the water's surface or the substrate on the stream bed. The stayers move a lot less when compared with the movers and there foraging differs in that they rely on the water current to carry their prey to them. It is important to note that both tactics involve foraging in still water pools that are on the edges of streams (McLaughlin 2008). It would seem reasonable that the stayers would have higher fitness because they would use less energy due to the limited movement required for their foraging, comparatively, and they rely on their prey coming to them. However by studying in detail, both in the field and the laboratory, the different grow rates for both tactics along with other factors (e.g. social aspects) it was shown that the fitness for both tactics was comparable. Individuals adopting any other behavioural approach to foraging had lower fitness in comparison. In this case the fitness was judged by measuring the growth rate of individual fish because Salmonids living in streams need a lot of stored energy to be able to swim and feed efficiently enough to live off their first year and so growth rate is a reasonable proxy for fitness (Noakes 2008). Also the argument that the stayers would expend less energy, because they hover in the water more than they move, has not been backed up by field research (McLaughlin and Grant 2001).

The brook charr are closely related to the Icelandic Arctic charr and can be found in freshwater sites (e.g. creeks, lakes and rivers in North America); and although they aren't the focus of this study, they can give an indication of the type of behavioural differences that can lead to individuals of the same species utilizing different niches in the same

environment (Noakes 2008). So the decision made by an individual brook charr, as whether to utilize one tactic or the other and essentially adapt to the constraints in this stream environment, which are the food sources and spatial limitations that apply to them, can be seen as a potential link to the divergence that we are seeing in other salmonid species within larger environments, such as the lakes that the Icelandic Arctic charr inhabit (McLaughlin 2008). However, it is only the juvenile brook charr that display the phenotypic variation and these differences are not brought forth again by isolating the brook charr and breeding them. Therefore it is most possibly an adaptation to their variable environment. The streams that they inhabit are heterogeneous, vary over different seasons and have different currents and conditions, so they may have evolved the ability to adapt to the environmental conditions they find themselves in when they are young and still developing (Noakes 2008).

An interesting continuation of this is the connection between the behaviours mentioned above and the morphological characteristics of the brook charr (in the juveniles of the species). A strong correlation has been shown between behaviour and morphology of the brook charr and the characteristics of the differing niches that the individuals inhabit (McLaughlin and Grant 2001). By noting the fish's behaviour (the position it holds in the water and its tactics) and simultaneously studying in detail its phenotypic characteristics, it is possible to test for correlation between these traits. So the young brook charr that display the stayer tactic are hovering in the water more often than they are moving and they also are more likely to be found in parts of a stream that has faster water flow. This behaviour can then be compared with their phenotype, which in the case of stayers included a narrower caudal peduncle and also larger caudal fins. These two phenotypic changes, that have an effect on how the fish swims, have been shown to be characteristics that are associated with more efficient swimming, which means that the fish uses less energy when swimming with these traits (Noakes 2008).

However these changes in phenotype cannot be completely attributed to phenotypic plasticity because young brook charr reared in a high-velocity control water environment were shown to develop some of these phenotypic traits associated with stayers. They developed the larger caudal fins and more slender bodies, but the phenotype wasn't as dramatic as the one witnessed in the wild juvenile brook charr so it can be said with some confidence that although there is a degree of phenotypic plasticity, it cannot fully explain the differences witnessed in the field between movers or stayers (Imre, McLaughlin, and Noakes 2002). This brings to light the distinct possibility that phenotypic plasticity, where the fish are reacting to environment pressures, could be playing a role in the morphological divergence witnessed in lake populations of Arctic charr, however the extent of its role is still unclear.

The differences between the stream and the lake environments in relation to their endemic charr species have been researched by comparing the Icelandic lake char with the North American brook charr (Noakes 2008). It has been noted how all of the lake charr morphs are displaying the mover tactic, which involves swimming around and actively searching for food. Also it appears that the lake charr are not as easily influenced by the water velocity as their brook charr cousins, because if they are taken out of the lake environment and placed in an artificial environment that resembles the stream conditions, the phenotypic plasticity observed in the brook charr (in their reaction to the speed of the water current) is not displayed (no change in tail or body form). This may be due to the lack of benefit to the stayer tactic when there is no need to stay and defend a certain spot because

food and therefore foraging is no longer confined to a certain location (Noakes 2008). Also the brook charr morphs show no significant change to their foraging tactics or social behaviour (Ferguson, Noakes, and Romani 1983). This gives credence to the hypothesis that they have adapted to best cope with the environmental conditions that are limiting. For the brook charr that would be the stream water velocity, which affects the tactics employed for foraging; but for the lake charr it isn't necessary to adapt to increasing/decreasing water current because lakes do not have currents, therefore the adaptive pressure affecting them would be working on utilizing different niches that would minimize competition and increase fitness.

1.7 Genetic and environmental factors in relation to phenotypic plasticity

The Icelandic Arctic charr have been shown to have a stronger genetic basis for their morphological differences (Snorrason et al. 1994) than the brook charr. Common garden experiments carried out for the four morphs of Thingvallavatn show that genetic differences lie and play a part in the variance of colour, growth, time of sexual maturity and foraging behaviour (Eiriksson, Skulason, and Snorrason 1999). The genetic contribution is small though with two morphometric variables showing 98% and 82% variability that could be explained by the environment (Adams and Huntingford, 2004); and there is not that much that separates the morphs genetically, but it has been shown to have an impact on morphology and other aspects of the Arctic charr. For the Arctic charr, the environment is also playing a role in shaping the differences that the morphs display. A theoretical model has been proposed which shows that behavioural differences can lead to sympatric divergence (evolution of multiple species from a single ancestor) (Kerckhove, Mclaughlin, and Noakes 2006). The behavioural differences have been shown in these models to be an important step to actual morphological differences and they have shown there to be interplay between the two factors; this initial difference can then lead to the development of morphs which are isolated in both their reproduction and their food source as the foraging that they carry out varies, essentially they will occupy different niches (Kerckhove, Mclaughlin, and Noakes 2006).

A point could be made for the phenotypic plasticity being useful when a species was in a varying environment, however if individuals of that species moved to more stable environment then they would retain the phenotypic plasticity of their ancestors (Skúlason et al. 1996). The selective pressure on the ancestors' trait would not be as strong in this new environment and therefore it would become less apparent over time. It has also been argued that the phenotypic plasticity is important, possibly along with genetic polymorphism, in the early stages of phenotypic divergence within the Arctic charr (Skúlason et al. 1996, Gíslason et al. 1999). The divergence however, at a later stage in evolutionary time, would be increased with the reproductive isolation leading to decreased gene flow between the different morphs and greater morph specialization for certain niches (environment and/or food). Therefore the suggestion is that resource polymorphism, the phenotypic plasticity seen in brook charr for example (Noakes 2008), is a crucial step for sympatric divergence to occur among lake fishes; and that phenotypic plasticity itself can be considered a trait in its own right, which can therefore be under evolutionary pressure and selection either for diversity or against it. Phenotypic plasticity has been shown to have played a part in the evolution of variation seen the niches that the Arctic charr occupy (Adams and Huntingford, 2004).

It has been noted that charr spawning sites in lakes are localized to certain places, specifically those near sources of groundwater and these sites have specific qualities that are not found elsewhere within the lake (Noakes 2008). The qualities include a higher concentration of dissolved oxygen, less variation in the temperature and pH of the water (Snucins, Curry, and Gunn 1992). It has been shown that increases in certain elements, such as aluminium, in lake water can have an effect on the development of lake charr (in this case the related *Salvelinus namaycush*). One study, which used lake charr embryos, showed that exposure of high levels of aluminium and a pH of 5 (lower than the pH of their normal lake environment) led to the alevins (newly hatched salmon or trout) being smaller than normal, their bones were less calcified and they were not as effective predators (in regard to a species of lake flee called *Daphnia magna*) (Gunn and Noakes 1987). Therefore the physical environment of the eggs affects the early development and also the time of hatching. This may be one of the reasons for the separation of the Arctic charr into different morphs, because they have site-specific spawning and once that developed in that species, where different individuals spawn at different locations, the differing conditions between those locations could lead to differences in early development and variation in hatching times between different groups; eventually leading to reproductive isolation between the different groups and possibly leading to morph creation (Gíslason et al. 1999).

1.8 Candidate genes: some examples

Reproductive isolation is an indication that distinct morphs could possibly become different species with enough evolutionary time, but for that to occur then the difference between the morphs needs to have a genetic basis. Danzmann et al. (1991) put forward the idea that it was possible to explain the little genetic differentiation between the morphs if the differences lay within a few very important regulatory genes (see also Sibthorpe et al. 2006). In a study carried out by Sibthorpe et al. (2006) *Pax7* was shown to differ between the dwarf and large benthic morphs with expression occurring significantly earlier in the dwarf in comparison to the large benthic (Sibthorpe et al. 2006). *Pax7* was selected as a candidate gene for polymorphism between Arctic charr morphs by Sibthorpe et al. because it plays a critical role during development in other vertebrates; specifically in the morphology of the head, the forming of the skeletal muscle and development of the central nervous system (Mansouri et al. 1996). *Pax7* has also been connected to myogenic stem cell specification (Seale et al. 2000). The gene also undergoes extensive alternative splicing in other species (found to also do so in Arctic charr by Sibthrope et al. 2006) and it has postulated that this alternative splicing could be key to macroevolution (Terai et al. 2003). SNPs and deletions/insertions were discovered in the *Pax7* gene and varied between the morphs although they were not present in the protein-coding region of the gene but rather within intron sequences where splice sites were also located. This presents the possibility of alternative splice sites that could affect morph differentiation, however there was not enough evidence to assert this in their paper, so further research would be required. But due to the aforementioned lack of protein coding variance, this is further evidence that changes in regulatory elements are behind the variation seen between morphs.

Additionally, a study of the major histocompatibility class II alpha (MHC2a) gene, which is fundamental to the immune system, showed distinctive allele frequencies and haplotypes

that distinguished each population of Arctic charr, that are consistent with recent natural selection (Conjeros et. al. 2008).

Due to the limited nature, comparatively, of the previous research on the morphological traits of the Icelandic Arctic charr, or any Arctic charr for that matter, then it becomes vital to draw on available morphological research carried out on other species of fish and also model organisms e.g. *Mus musculus*. The *M. musculus* research has pointed to several genes that are key to morphogenesis of the jaw and the development/arrangement of teeth (Kuratani, Matsuo, and Aizawa 1997), and have also been shown to play a role in East African Cichlid fishes as well (Terai, Morikawa, and Okada 2002). The genes highlighted by the *mus musculus* research were *Dlx1*, *Dlx2*, *Pax9*, *Otx1*, *Otx2*, *Bmp2* and *Bmp4*, Their homologs were found in the Cichlid fishes and so a similar approach can be applied to researching jaw morphogenesis polymorphism in Arctic charr. I will survey sequence variation in those genes, with the aim of finding genetic differences between two Artic charr morphs, dwarf benthic and small pelagic.

1.9 Evolution of fish development: contrasting cichlids and Artic charr

The Cichlids of Eastern Africa have been proposed as an ideal model system for the study of speciation due to the thousands of species within the lakes of Eastern Africa and the wealth of variation that can be found in their morphology (Kocher 2004). Therefore by identifying genes are connected with the variation seen between the Eastern Africa Cichlids, it becomes possible to see which genes are under selection and have led to the speciation seen in the Cichlids.

There are several methods to study evolutionary divergence and differentiation, for instance to i) study the molecular evolution of key genes, ii) population genetic differentiation in specific genes, iii) conduct genetic crosses and mapping in related enough species/morphs and iv) to analyse gene expression, most ideally in specific tissues or developmental time points, in divergent species or morphs. One way to determine which gene/s are most important for the variation seen between the different species of East African Cichlid, the value dN/dS ratio was found for each of the candidate genes mentioned above. The dN/dS ratio (also referred to sometimes as the K_a/K_s ratio) represents the ratio between the non-synonymous (dN) and synonymous (dS) substitutions that have occurred in the protein coding sequence of a certain gene (Yang and Bielawski 2000). Synonymous substitutions are base changes that do not change the amino acid (most often the last base in the codon). An example of this would be a single base change in the phenylalanine (phe) codon (UUU) which could result in either a synonymous substitution (UUC) which would still be a codon for phe or a non-synonymous substitution (UUA) which would change the amino acid in the protein coding sequence from phe to leucine. If this ratio is significantly lower than 1 then it gives an indication of negative selection and a value that significantly exceeds 1 indicates positive selection. Therefore by choosing the gene which shows the highest dN/dS ratio (still well under 1 though, 0,277 +/- 0,0470 to be exact), in this case it was *Bmp4*, then it is possible to concentrate on the morphological gene most likely to be diverging, in this case that gene was *Bmp4*. The *Bmp4* gene was either under negative selection and experienced relaxation of negative selection pressure occurring in tandem with morphological changes or possibly positive

selection in the short term - most possibly occurring at the same point in which the morphological changes occurred (Terai Morikawa, and Okada 2002).

There are certain differences between the Cichlid fishes in question and the Arctic charr, due to the varying time scale of divergence. The Cichlids in Eastern Africa, have diverged over the course of one to two million years (although this varies largely between different species in different lakes); with the Arctic charr morphs (not separate species yet) have emerged after the end of the last ice age, around ten thousand years ago (Bernatchez and Landry 2003). This time difference between the Cichlid and Arctic charr fishes results in the alleles in most, if not all, Cichlid species being fixed. By fixed it is meant that an ancestor polymorphism has gone to fixation in the daughter species. This is due to the fact that the genes have been under selective pressure for such a length of time that the base changes that are relevant to the "desired" phenotype are fixed in the species and other variation that may be dragged along in the short term has slowly been lost.

The Cichlid fishes (with a few exceptions) cannot be researched in the same way as the Arctic charr because the SNP frequency wouldn't be as descriptive of the genes under selective pressure as it would be in the Arctic charr. This is due to the short evolutionary time that the Arctic charr morphs have diverged over and therefore there is insufficient time for fixation of certain alleles with charr. By scanning candidate genes for SNPs we plan to look for signs of recent selection seen as allele frequency differentiation between morphs. When such SNPs are found it is interesting to check for linkage disequilibrium in a larger region and evaluate what change is causing the phenotype or if it is just hitchhiking along with the regions that are actually under selection (elsewhere in the same gene, or perhaps in neighboring genes).

An interesting note is that variation in Cichlids has been shown to be confined to certain regulatory regions, in some instances e.g. *Bmp4*. Therefore because the alleles would become fixed, a scan for variation in other parts of the gene wouldn't show variation and this is deceptive because there is variation within the gene. In the Arctic charr a scan of such a gene would most likely land outside of the region that was having a phenotypic impact however it could possibly show frequency differences due to hitchhiking.

With morphs or closely related species, it is possible to map the genes that correlate with traits of interest. This of course hinges on the trait having a genetic basis, see discussion above. First one creates an experimental cross (termed hybrid), with each of the parents belonging to separate morphs or species. Then the first generation (F1) is crossed with itself (brother-sister cross) and a second generation (F2) is created and with this generation it becomes possible to do a phenotypic assay and test for linkage. Linkage analysis relies on a series of genetic markers, e.g. microsatellites, to create a linkage map that is informative about which sites are likely connected, and therefore by assessing the logarithm of odds (LOD) score it is possible to determine which markers are descriptive for each of the morphs or closely related species (Peichel et al. 2001). To gather genetic marker candidates it's possible to perform large-scale library screening and sequencing to identify microsatellites, and see which ones are most prevalent in the model species genome (Peichel et al. 2001). Genome-wide linkage mapping can be used to see which site in the genome contributes to a particular phenotype (although it may only explain the phenotype to a certain degree). The phenotypic assay involves the detailed analysis of differing traits, e.g. jaw width or the length of certain bones (Albertson, Streelman, and Kocher 2003). This has been carried out in Arctic charr (Noakes 2008), threespine

stickleback (Peichel et al. 2001) and Cichlid fishes (Albertson, Streelman, and Kocher 2003). Genetic analysis can be carried out to find not only the number but also the location of loci that are causing morphological differences between closely related species, in the case of Cichlids or morphs as is the case of Arctic charr and threespine sticklebacks.

Particularly interesting is the threespine stickleback that has also seen rapid evolution and diverse speciation since the last ice age ended (see Peichel et al. 2001). The threespine stickleback bears interesting similarities to the Arctic charr, as it too has gone through the formation of distinct morphs through colonization of new niches (see Peichel et al. 2001). The invasion into new habitats has shown to be connected to adaptations seen within the stickleback species (Peichel et al. 2001). Those adaptations are feeding morphology and other body form changes, which draws a certain parallel to the adaptations seen within the Arctic charr. With a linkage analysis it is possible to explain a large percentage of the phenotypic variance witnessed in the phenotypic assay, although it isn't possible to fully explain the phenotypic variance solely with these markers due to the difference in the importance that the environment plays for some phenotypes. It also depends on how heritable a trait is because the heritability of a phenotype can change depending on the population and the environment; thus estimating heritability accurately can be difficult, even the best estimations have large standard errors (Falconer and Mackay 1996). A problem that can occur with the use of genome-wide linkage mapping is that if the phenotype you are studying is caused by a larger amount of genes that each only have a small phenotypic effect, they won't stand out from the background noise that is present in this methodology. To find such small-effect genes the sample size must be very large, and that will increase the power of the method resulting in identification of genes with a smaller impact on phenotypes of interest (Hatfield 1997).

1.10 Enhancer mutations and their effect on development

Mutations in protein coding regions, regulatory regions, and microRNA can affect the phenotype but it is also possible that phenotypic variation can be the result of a change in an enhancer. It has been shown that *Pitx1* explains two thirds of the variation in threespine stickleback pelvic spikes. The differing amount of skeletal armour is one of the most prominent and obvious morphological differences between different populations of stickleback, and therefore is a strong candidate for the study of rapid adaptive radiation i.e. vertebrate speciation (Peichel et al. 2001). The *Pitx1* gene is connected to the normal development of the hindlimb, which is seen by its expression patterns and the loss of hindlimbs when the gene is inactivated (Lanctot et al. 1999); and strikingly does not have any variation in its protein coding region compared with freshwater and marine fish (Chan et al. 2010). However the expression of *Pitx1* varies greatly between different freshwater populations and Chan et al. (2010) showed that this variation is the result of a change in an enhancer (termed Pel) for *Pitx1*. The possibility of pelvic development failure or changes to unknown *trans*-acting factors being the cause was ruled out with the use of F1 hybrids and so the loss of *Pitx1* expression in the pelvic region was shown to be cause the loss of pelvic spikes. It was also shown that reintroduction of the enhancer restored the formation of pelvic spikes in progeny of individuals belonging to populations thought to be homozygous for a Pel deletion (see figure 1.5).

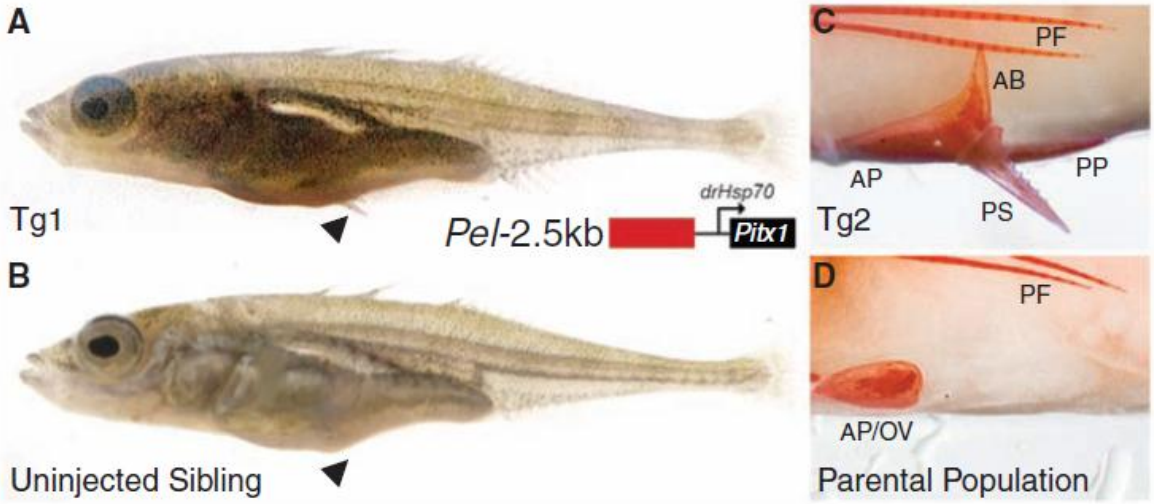

Figure 1.5 Pelvic-reduced juvenile threespine stickleback expressing a *Pitx1* transgene with the Pel enhancer (A) and the resulting formation of external spikes. A sibling that has not been injected (B) (reproduced from Chan et el. 2010).

To be more precise there were 9 separate deletions found in a large genotyping survey (in the enhancer region causing tissue specific expression of *Pitx1* to cease and leads to pelvic reduction. It is thought that several evolution pressures could have contributed to the development of pelvic spine reduction (Chan et al. 2010), with those at the forefront being absence of certain predatory fish where the mutation is prevalent, a limited amount of calcium in those lakes where these threespine stickleback mutants reside, or possibly the predation on the stickleback by certain grasping insects. Another possible explanation for this could be adaptation to the different predators that are found the varying niches within lakes, i.e. the open water vs. the shoreline (Peichel et al. 2001). The pelvic spines increase fitness against certain predators, e.g. birds and fish, but lower fitness when others are present, such as insects (Orr 2001). The region that the enhancer was thought to be located was conserved in zebrafish (*Danio rario*) and other teleosts, suggesting that this is an ancestrally conserved regulatory enhancer with important functions. This example, as Chan et al. (2010) describe it, leads to an alteration of the stickleback skeletal structure (see figure 1.6) and by studying how this change occurred, it is possible to gain insight into the evolution behind such major phenotypic changes in natural populations, in relation particularly to the skeleton.

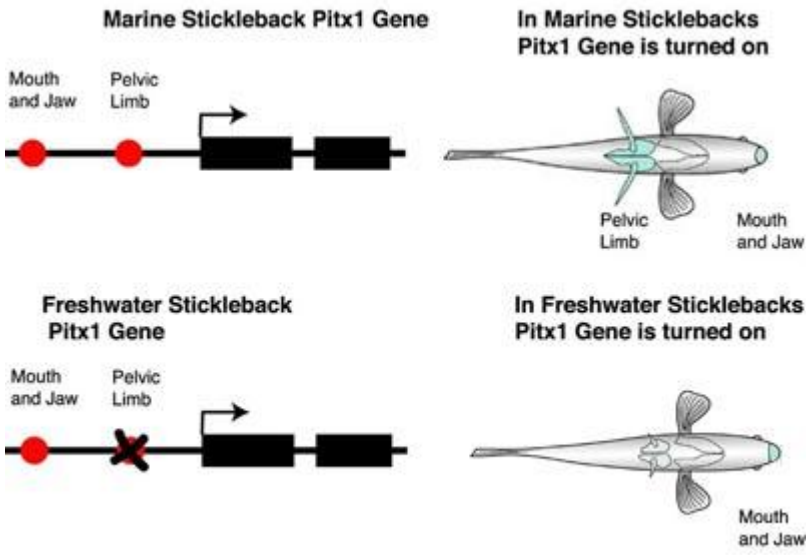

Figure 1.6 The phenotype of the threespine stickleback fish with and without the pelvic limb enhancer (Pel) mutation. The marine stickleback (top right) has the pelvic limb present and also has the pelvic limb enhancer functioning (displayed as a red circle at the top left); the freshwater stickleback (bottom left) has a mutation in the pelvic limb enhancer sequence and therefore it isn't functional and therefore the pelvic limb is not present in the freshwater stickleback (source: http://ncse.com/creationism/analysis/body-plans - accessed 01/03/2011).

This shows that scanning the protein coding part of a gene for variation might not give results if the gene itself is unchanged between the Arctic charr morphs and the actual phenotypic variation was being caused by enhancer changes. Therefore if the allele was not yet fixed, as was discussed earlier (chapter 1.8), and an enhancer was under selection, then a scan for variation in the gene could turn up no variation due to the distance that the enhancer is from the gene (in vertebrates it can be from 1 Kb to 1 Mb). Although variation could be present in the gene if it was hitchhiking along with the mutations in the enhancer, it should be considered when viewing small fragments of a gene for variation that no variation is not sufficient to rule out that gene is under selection or plays a part in morphological differentiation between Arctic charr morphs.

The Pel enhancer of *Pitx1* was shown to be well conserved, with threespine stickleback and an ancestor having the enhancer and so it is most likely present in the Arctic charr. This raises questions about its function in the charr and although the pelvic spines are not present within charr, it may be interesting to study what function the enhancer has and if it the lack of pelvic spines has freed the enhancer to take on another function or if it simply isn't present in Artic charr genome.

The example of pelvic reduction found in various freshwater populations of Stickleback fishes brings with it interesting evidence for macroevolution. The pelvic loss occurring in multiple populations indicates that the mutation is beneficial for the Stickleback fish in the environments where it occurs and being a single deletion in an enhancer sequence means that a large beneficial phenotypic change occurred by a single mutation. This lends credence to the role of macroevolution in evolutionary change but there are obviously other genes which contribute smaller, step by step, changes over a course of thousands or

millions of years and that too leads to beneficial phenotypic changes. The two concepts are by no means mutually exclusive.

1.11 Project aims: hypothesis

With all of this in mind, and in connection with the two charr morphs, I propose the following hypothesis; that the morphological differences seen between dwarf benthic and small pelagic charr have an underlying genetic component and that the differences in said traits are due to, at least in part, a genetic variability in the genes that are responsible for head, cartilage and bone development.

To see if this hypothesis carries any weight I will test for gene variation and have five hundred and three DNA samples from Lake Thingvallavatn available for testing (Kapralova et al 2011). The samples were taken from select populations of both dwarf benthic and small pelagic charr, in five locations within Thingvallavatn and a reference population of fifty large benthic charr were also caught from their principle spawning grounds in Ólafsdráttur and will be used as a control group for the study. The fish were collected and processed by Kalina H. Kapralova et al. in 2005 and a more detailed description of their sampling method can be found their paper (Kapralova et al. 2011).

The central aim of the project is to study genetic differentiation between morphs that may contribute to the morphological variations between two Arctic charr morphs. The focus is on developmental candidate genes with connections to head, bone and cartilage development. This study might provide insight into the genes that could be used in marker assisted selection and this could lead to improvement in Arctic charr morphology. These improvements would lead to improved yield in aquaculture, where the morphological selection would increase body size or musculature. However this study is very much a preliminary step towards this goal, with a lot of analysis and experimentation left to do before this information is of any practical use.

2 Research methods

This chapter contains all of the methodology used in the research project, with detailed explanations of both the procedures and the reactants that are required.

2.1 Creating primers

Primers are required for the PCR due to the fact that you want to amplify a certain part of the genome, in my case developmental genes. These primers, named for their priming of the polymerase, and they are short DNA segments that adhere to certain parts of the genome of interest. Thusly I created primers for these genes using the EST collections from either closely related species (*Salmon salar* and *oncorhynchus mykiss*), or the limited EST collection code available for the Arctic charr itself (http://web.uvic.ca/grasp/ - accessed 01/03/2011). Due to the limited EST collections from Arctic charr I was often required to use sequences of candidate genes from closely related species. I used keyword search, full gene names or abbreviations, on files listing the ESTs from those three species. Using abbreviations can however lead to mistakes, as two or more genes can be represented by the same abbreviation. This was indeed the case for PTH, which represents both parathyroid hormone and peptidyl-tRNA hydrolase, as we found out the hard way. We set out to amplify the former, but retrieved the latter (see results). To identify conserved portions of these genes I BLATed these ESTs to the genome of fugu (*Takifugu rubripes,* genome.ucsc.edu). The logic is that if the gene is preserved in a distant relative then it is most likely located in the Arctic charr as well. Thus I would design my primers to match the conserved parts. This method for primer design is commonly used when dealing with species without a genomic sequence.

2.2 PCR

The purpose of PCR (polymerase chain reaction) is to amplify a certain segment of a genome, in this case certain genes of interest that are connected to development and the immune system of Arctic charr. We used a standard protocol for the PCR reaction (Table 2.1) and PCR cycle (Figure 2.1). The Peltier thermal cycler, DNA engine tetrad 2 system is used for the reaction and the PCR heat cycle is pre-programmed into the system.

Table 2.1 The protocol for a PCR reaction that is used to amplify a gene from a single sample.

DNA	dNTP	TEQ polymerase	10x Buffer	Forward Primer	Reverse Primer	ddH_2O	Total Volume
1 µL	2 µL	0,3 µL	2 µL	0,2 µL	0,2 µL	14,3 µL	20 µL

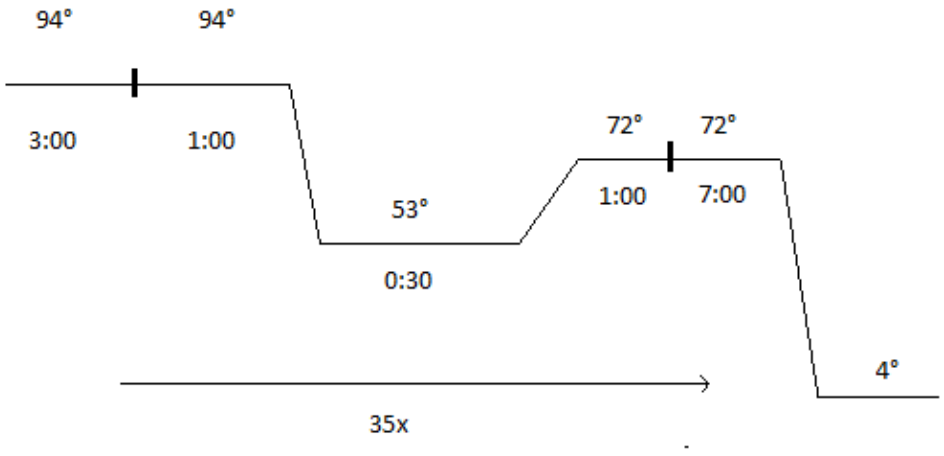

Figure 2.1 The temperature and timing of the PCR reaction. It has six steps and steps two, three and four are repeated thirty five times.

2.3 Allele-specific PCR

Due to the high cost of the sequencing reaction, and also the time consuming steps required prior to sequencing, then an alternative and cheaper method is more suited to genotyping individual charr morphs. This is possible by already having the sequence for a gene of interest that shows variability a one site that is consistent with the variability seen through that gene segment. By which I mean that a single site of polymorphism is an indicator for a certain allele. The method that we use for this en mass genotyping is called Allele specific PCR,

By using the old primers (both forward and reverse) with newly created primers that bind to the site dependent on the polymorphism that the individual has. This works best if the variation between two alleles is an insertion/deletion because variation in a base might cause weak binding of the primer and the PCR would therefore still give a small amount of PCR product. Whereas if the site is missing a base the primer will have a much harder time binding to the sequence due to the "spare" base being pushed out awkwardly to allow for binding of the neighbouring two bases. So therefore two primers are created, and just to be able to give an example let us say that we pair the new reverse primer (which accounts for an insertion) with the old forward and the new forward primer (which would account for a sequencing lacking the insertion) with the old reverse. Therefore, dependent of the two resulting sequences differing in size and the reaction being done on an individual with both alleles (heterozygous), when doing a gel electrophoresis of the PCR product we would see the largest band (the pairing of the old forward and reverse primers), then two smaller bands varying in size and they would give an indication that one allele with the insertion and one without had been magnified. Therefore you could genotype the individuals based on which bands would be present when doing a gel electrophoresis, a homozygous individual would either give the largest and smallest or largest medium size band depending on whether he was homozygous for the insertion or not.

The reaction itself is basically a multiplex PCR reaction, where you are using all the same reactants with two primer pairs instead of the usual one and although you are only using four primers because of the additional pairing that occurs between the "normal" forward and reverse it is possible to get three PCR products when the reaction is carried out on a

heterozygous individual. For further clarification of the exact amount of reactants used then see the protocol in table 2.2 found below. The timing and temperatures of the reaction are the same as those found in the usual PCR seen in figure 2.1, although this can varying depending on the primers use (then the allele temperature, in this case 53 degrees would change depending on the T_m of the primers being used).

Table 2.2 The reactants required for a single allele specific PCR, where the R2 and F2 primers are the additional primers which are allele specific, one binding when the insertion is present and the other binding when there is no insertion present.

DNA	dNTP	TEQ pol	10x Buffer	F primer	R primer	F2 Primer	R2 Primer	ddH_2O	TV*
1 µL	2 µL	0,3 µL	2 µL	0,2 µL	0,2 µL	0,2 µL	0,2 µL	13,9 µL	20 µL

* Total Volume

2.4 EXOSAP reaction

The polymerase chain reaction (PCR) is supposed to amplify a specific gene. However there are a lot of leftover products from that reaction in the PCR solution that can interfere with the sequencing reaction. The purpose of the EXOSAP is to remove the leftover dNTP and primers so that they do not interfere with the sequencing. Table 2.3 shows the EXOSAP reaction mixture. The EXOSAP program executed on the Peltier thermal cycler, DNA engine tetrad 2 system.

Table 2.3 The reaction mixture for an EXOSAP reaction for a single sample.

ddH_2O	Antarctic Phos	Antarctic Phos Buffer	Exo 1	Reactants Volume	PCR product	Total Volume
3,7 µL	0,2 µL	1 µL	0,1 µL	5 µL	5 µL	10 µL

The reaction itself occurs at 38°C for 35 minutes. The Enzymes Antartic Phosphatase (New England Biolabs) and Exo 1 (Fermentas Life Sciences) function optimally at 38°C and then are inactivated and denatured at 80°C (20 minutes).

2.5 DNA sequencing reaction

The DNA sequencing was conducted with the BigDYE (Perkin Elmer Applied Biosystems Division), which contains ddNTP´s, normal dNTP´s and also polymerase. The reaction mixture is shown in Table 2.4.

Table 2.4 The protocol for a DNA sequencing reaction of a single sample.

ddH_2O	BigDYE	VII 5x BigDYE Buffer	Primer	Total Volume
5,25 µL	2,756 µL	0,487 µL	1,5 µL	9,993 µL

The reaction cycle for the sequencing reaction is as follows. 96°C for 10 seconds, 96°C for 10 seconds, 50°C for 5 seconds, 60°C for 2 minutes, 96°C for 10 seconds, repeat steps 2 to

4 twenty-five more times, and then store at 4°C. This program is pre-set into the Peltier thermal cycler, DNA engine tetrad 2 system and said machine is used for the reaction.

2.6 Ethanol precipitation

This method is used to purify and concentrate the DNA after the sequencing reaction (see Table 2.5). The resulting DNA is then run on the sequencing machine. Salt and ethanol together force the DNA (possible also with RNA) to precipitate away from the solution, and you spin down the samples in a Beckman Coulter Allegra 25R Centrifuge, causing the salt/ethanol solution to separate and the DNA forms into a pellet at the bottom of the tube it is in.

Table 2.5 The ethanol precipitation master mix for an entire tray of 96 samples.

ddH_2O	NaOAc	Glycogen (20mg/ µL)
4500 µL	500 µL	25 µL

The first step is to add 45 µL of mixture in table 4 to each sequencing reaction product. After doing so you take 96% ethanol which has been stored at -20°C and add 125 µL to each sample. Next you place aluminium foil lid and mix the solution by turning the tray over 3 or 4 times. Next you spin the samples in a centrifuge at 4000rpm for 30 minutes at 4°C and when this is finished you remove the aluminium lid quickly and pour off the ethanol in one fluid motion to make sure that there is no mixture between samples. Once this is done you spin the trays in the centrifuge upside down on Kimwipes at 300 rpm for 2 minutes and when this is done you add 250 µL 70% ethanol to each sample and spin at 4000rpm for 5 minutes and then pour off the solution again. Finally spin the tray upside down for 5 minutes at 300rpm on Kimwipes. When this is done store the DNA pellets, which shouldn't be visible, in a dark dry place for 20 to 30 minutes and then seal them with aluminium lid and place in fridge or continue straight onto the next step.

The DNA pellets are then resuspended in 9,9 µL of HiDi and then mixed using a vortex for 30 to 60 seconds. The samples are then spun down in the Beckman Coulter Allegra 25R Centrifuge and then run on a 3100 gene analyser.

2.7 Processing sequence data

After the gene of interest has been sequenced the fluorescent trace from the sequencer needs to be base called, wherein the fluorescent trace is given a value determining its strength and then assigning a base to said value. Phrap is then used to compile all the shotgun sequences and constructs a contig of the sequence. And this contig once edited by using consed to determine if the reads are indeed correct is exported to Genedoc where the final edited sequence can be viewed. Alignment programs can be used to make your sequences of different individuals match up.

2.8 χ^2-square test statistic

Deviations from the Hardy-Weinberg principle can be an indication of selection bias but to be able to say that apparent deviations are significant, statistical analysis needs to be applied to make sure it's not just down to random chance. Therefore the χ^2 test statistic is

used to compare the observed allele frequencies with the expected allele frequencies (without bias, therefore random), to see if a there is a significant deviation between the them. The equation used for this comparison is found below.

Equation 2.1 The χ^2 test equation (Sokal and Rohlf, 1995), used to find the χ^2 value which is then used to find the p-value and that value lets you either refute or accept the null hypothesis.

$$X^2 = \sum \frac{(Observed\ frequency - Expected\ frequency)^2}{Expected\ frequency}$$

The expected frequencies were calculated using the values in the observed frequencies and the following formula, e = (RT/n)*(CT/n)*n, where n is the total amount of alleles, RT is the row total and CT is the column total.

3 Results

3.1 Candidate gene study: an overview

I set out to create primers for around 50 developmental genes but only managed to find twenty seven genes that were viable for study in the EST data available; the other ESTs were either lacking in information because they were too short or they simply were not present in the EST datasets. Thirteen of these twenty seven primers gave positive results and I managed to successfully PCR and sequence twelve of them (see Table 3.1).

Table 3.1 An overview of the sequenced developmental genes.

Gene	# c.a. bases	# of hetz. sites *	# seg. sites ×	# seg. indels $	# individuals		HWE (p-value) €
Etbr2	450	3	1	0	4 D	8 M	No (0,002479)
Otx2	390	0	0	0	9 D	15 M	NA
Bmp4	350	1(1)	1	0	4 D	6 M	No (0,03688)
Runx1	386	4	0	1	18 D	20 M	No (5,603e-09)
Pth £	460	1	0	1	16 D	19 M	No (0,002479)
Eng2	300	11	0	0	4 D	8 M	No (0,002479)
Fgfr4	180	4(1)	1	0	7 D	7 M	No (0,006738)
Fgop2	477	0	0	1	20 D	19 M	No (0,0001282)
Igfbp1	525	2	0	0	2 D	7 M	No (0,01111)
Twist2	369	0	0	1	8 D	8 M	No (3.341e-08)
Wnt9	310	3	0	0	8 D	8 M	No (0,0009119)

* The number of heterozygote sites
× The number of segregating sites
$ The number of segregating insertions or deletions
D for dwarf benthic charr and M for small pelagic charr
€ Whether or not the allele frequencies are in Hardy-Weinberg equilibrium (p-value given)
£ *Pth* is not a developmental gene

The results show that 10 of those 11 genes are indeed represented by two genes in Arctic charr. This is determined by the fact that all the individuals sequenced are heterozygous for 1 and up to 11 different sites. Such a pattern of extreme heterozygosity is best explained by paralogous genes existing in the genome (it is a drastic deviation from Hardy Weinberg proportions as can be seen by the results of χ^2 test analysis performed for each gene in the table 3.1 above).

A subset of genes are of particular interest as they show either varying haplotypes present in both morphs or a haplotype unique to a single morph and not found in the other. Representative examples of these genes are given in the following section, which includes an example of a gene that shows no variation between the morphs and two examples of genes that are potentially different between the morphs.

3.2 *Otx2*

The sequenced portion of one gene, *Otx2* (Orthodenticle homolog 2), showed no variation. *Otx2* is connected to brain and sensory organ development (Kimura-Yoshida et. al. 2004) and there is no variability between small pelagic charr and dwarf benthic charr in the frragment sequence for the *Otx2* gene (see Table 3.1). The data does not indicate multiple copies of *Otx2* as is the case for nine of the other genes that were sequenced. Note that although the individuals sequenced are few, with only a single dwarf benthic charr present in the sample seen in figure 3.1, that I had previously sequenced 16 individuals for the same gene and there was no variability between those sequences either. The example of 8 individuals was simply taken to give an indication of a sequence without variability.

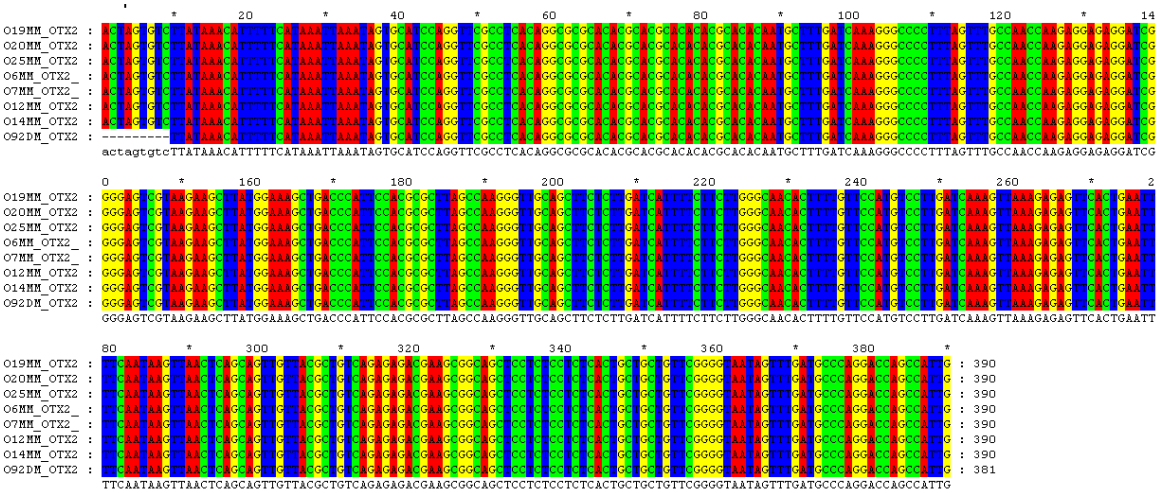

Figure 3.1 The 390 base pair *Otx2* gene fragment which was sequenced for eight small pelagic charr (MM) and a single dwarf charr (DM). Visualized with Genedoc.

3.3 *Pth* (peptidyl-tRNA hydrolase)

The *Pth* gene (Figure 3.2) that encodes a peptidyl-tRNA hydrolase was amplified due to serendipitous events but it turns out to be quite interesting. Within the gene fragment it segregates an insertion or deletion between GGCC in the sequence (base 142), only witnessed in the dwarf benthic charr individuals. Note the sequence becomes unreadable after this insertion, which suggests that one paralog may be fixed for this polymorphism in the dwarf benthic charr. Due to the size of the PCR product, which is around a thousand base pairs, the sequencing of *Pth* in the opposite direction using the reverse primer (to try and get the entire sequence) failed because the sequence went out of phase before the site of the C base insertion, mentioned above. So, further individuals need to be sequenced or genotyped to see if this pattern of variability holds up. This will be attempted in the future using the allele specific method that was also used for *Fgop2* and was described in the methods chapter. However is that fails then the new reverse primer, which was created for the allele specific PCR, can be used in a normal sequencing reaction and due to the size of the PCR product being reduced, the sequencing will be able to cover the site of variability and give the exact sequence for the *Pth* gene. The sequenced for the mRNA of the *Pth* gene was compared with the sequence obtained through genomic PCR and the site of interest was found to be a 7 base deletion within a protein-coding region of the gene.

Future allele specific PCR should therefore be possible with primers that account for the presence and absence of the deletion.

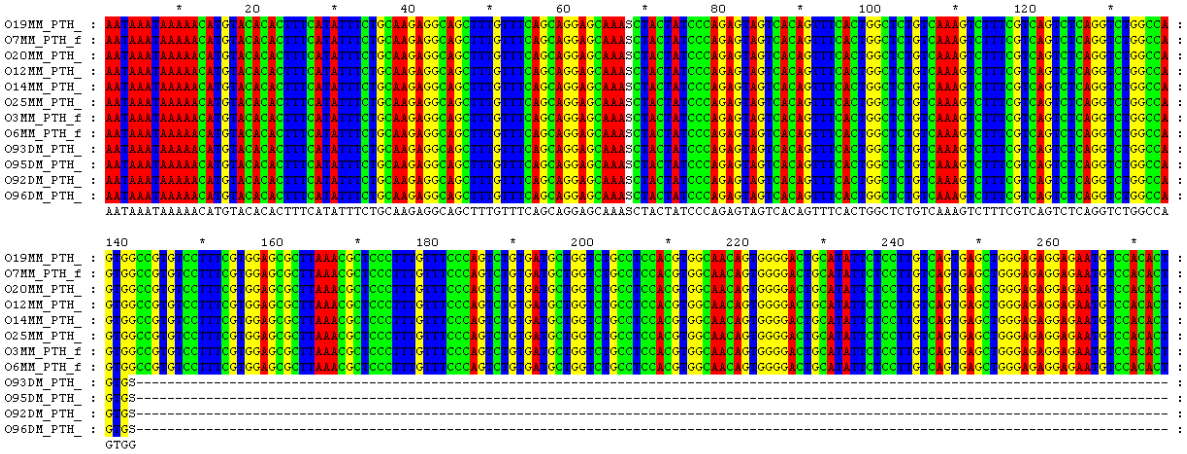

Figure 3.2 The site of interest in the gene *Pth*, a further 200 bases were sequenced however no further variability was witnessed. Sequences for 8 small pelagic (MM) and 4 dwarf benthic charr (DM).Visualized with Genedoc.

A rough estimation of the difference in gene expression is possible with the use of Illumina sequences of RNA from dwarf charr and aquaculture charr (S.R. Franzdóttir, Z.O.Jónsson, and J. Guðbrandsson - unpublished results), because it gives you the amount of mRNA transcripts present in the sample sequenced. Therefore by comparing the amount of transcripts between the dwarf and aquaculture it is possible to get a rough estimate on the difference in gene expression. Also there are four time periods during development that were surveyed for both charr and therefore it is possible to compare the number of transcripts not only between the two morphs as was done above, but also during different developmental periods. The results of this comparison for the gene *Pth* can be seen in figure 3.3 below. A χ^2 test performed on the amount to see which time periods show significant difference between the morphs revealed that the first (p = 0,00723), second (p = 3,745e-06), third (p = 0,03763) and fourth (p = 0,01092) developmental periods show significant difference. It should be stressed that the comparison between the dwarf benthic and aquaculture charr is no more than an indication for gene expression differences between Arctic charr morphs.

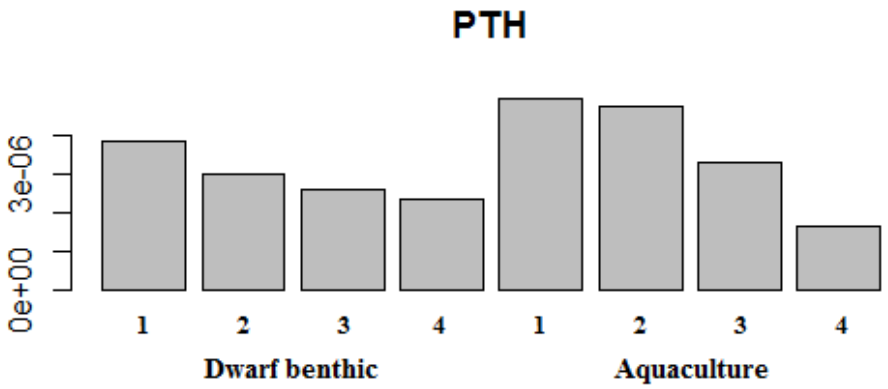

Figure 3.3 A rough comparison of *Pth* expression between the dwarf benthic and aquaculture charr using the amount of mRNA transcripts present in new generation sequencing with Illumina technology. The numbers 1 to 4 represent gene expression at different times during development. On the Y axis is the expression level (relative read counts). Visualized with R.

3.4 *Fgop2*

The *Fgop2* (Fibroblast Growth Factor Receptor 1 oncogene partner 2) fragment shows equal variability between the two haplotypes, one of which goes out of phase prematurely at the heterozygote site (see figure 3.4). This gene was also sequenced using the reverse primer which confirmed the deletion. There is a large variance in the haplotype frequency in the dwarf benthic charr and small pelagic charr, however the small sample size makes assessing the significance of this difficult, although using the 39 individuals sequenced I attempted to see if there were differences in allele frequencies and that can be found in chapter 2.8. The current number of individuals sequenced will have to be increased to be absolutely sure of the results, as the current number sequenced for *Fgop2* is not enough to be fully confident in the results of statistical analysis. It is our focus to try and genotype more individuals with allele-specific PCR, which would lower the cost due to the sequencing stage being avoided. The data also hinted at a difference in frequency of allelic variability between the charr morphs in the *Twist2* and *Runx1* genes, with further sequencing needed to confirm if true polymorphism is at hand.

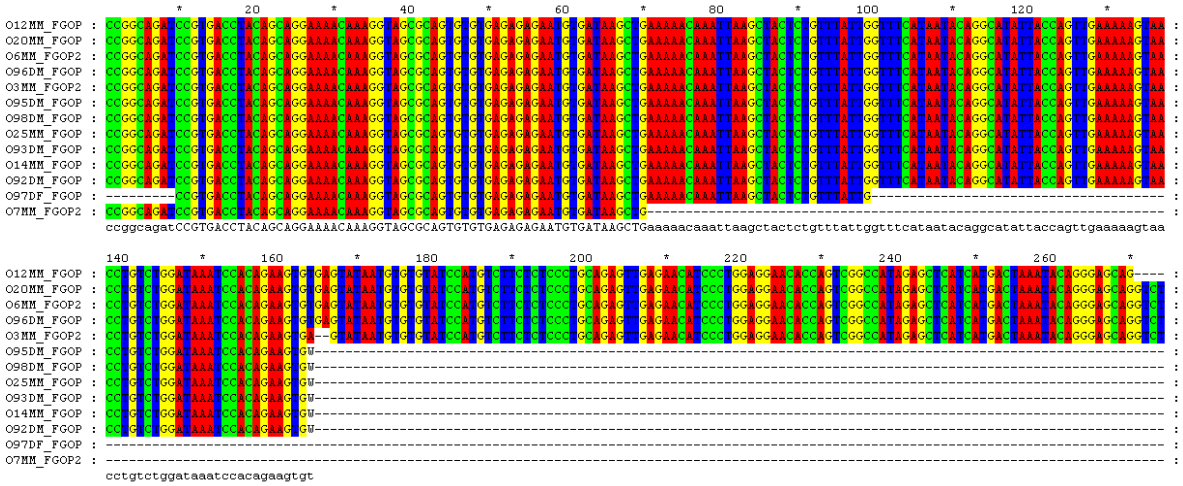

Figure 3.4 The site of interest in the *Fgop2* gene fragment. A further 200 bases were sequenced but did not contain polymorphisms and therefore are not shown. Visualized with Genedoc.

The gene expression comparison between morphs and different developmental periods, as was described above in chapter 3.3, can be seen in figure 3.5 for the gene *Fgop2*. A χ^2 test, to compare the number of *Fgop2* transcripts for the first and developmental period between the two morphs, shows that the variation is significant (p = 2.2e-16). The second (p = 2.2e-16) and third (p = 0,002524) developmental periods also showed significant difference between the dwarf benthic and aquaculture charr; However the fourth period did not show a significant difference (p = 0,2679).

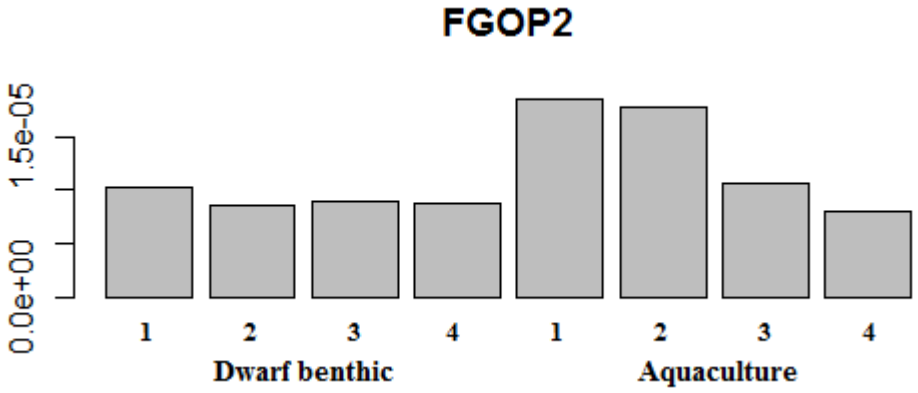

Figure 3.5 A rough comparison of *Fgop2* expression between the dwarf benthic and aquaculture charr using the amount of mRNA transcripts present in new generation sequencing with Illumina technology. The numbers 1 to 4 represent gene expression at different times during development. On the Y axis is the expression level (relative read count). Visualized with R.

3.5 Statistical analysis of allele frequencies

I put forth two hypotheses in connection with the allele frequencies seen in table 3.2. The null hypothesis (H_0) is that there is no significant difference between the observed and random expectation frequencies. The other hypothesis (H_1) is that there is a significant difference between the observed and expected frequencies. The α value for these hypotheses are set at 95%.

Table 3.2 The differing genotypes for the *Fgop2* gene fragment.

Morph	Genotypes			
	AC/AC	AC/del	del/del	Total
Dwarf Benthic charr	10	10	0	20
Small Pelagic charr	11	4	4	19

The observed allele frequencies in the two morphs are p_{AC} = 30/40 for Dwarf and p_{AC} = 26/38 for small pelagic (Table 3.3).

Table 3.3 The observed allele frequencies for the *Fgop2* gene fragment.

Allele	Dwarf Benthic charr	Small Pelagic charr	RT
AC	30	26	56
-	10	12	22
CT	40	38	n=78

RT stands for row total and CT stands for column total.

The expected frequencies for the *Fgop2* gene fragment were calculated using the method described in chapter 2.8.

Table 3.4 The expected allele frequencies for the *Fgop2* gene fragment.

Allele	Dwarf Benthic charr	Small Pelagic charr	RT
AC	29	27	56
del	11	11	22
CT	40	38	n=78

With both the observed and expected allele frequencies known it is possible to use the χ^2 test equation (equation 2.1.) to work out the χ^2 test value to see if the null hypothesis is refuted or accepted. So the χ^2-test statistic is worked out for *Fgop2*, as can be seen below.

$$X^2 = \frac{(30-29)^2}{29} + \frac{(10-11)^2}{11} + \frac{(26-27)^2}{27} + \frac{(12-11)^2}{11} = 0{,}2533$$

The degree of freedom for this χ^2 test is 2 and so the p-value is 0,881042 and so the null hypothesis is not rejected and therefore there is not a significant statistical difference between the allele frequencies of the *Fgop2* gene.

The data for the other gene of interest, *Pth*, is limited and as such a comparison of allele frequencies would have little merit. The limited nature of the data can be seen in table 3.5 below.

Table 3.5 The different genotypes for the *Pth* gene in the two Arctic charr morphs.

Arctic charr morph	G/G	G/C	C/C	Total
Dwarf Benthic charr (Dverg Bleikja)	0	4	0	4
Small Pelagic charr (Murta)	8	0	0	8

4 Discussion

My original hypothesis that the differences between the two morphs could have a genetic factor seems to be likely viewing the evidence of the limited selection of genes detailed above. It would however be premature to verify the hypothesis, as further study of the candidate developmental genes is required. This study should however be a good indication of the developmental genes that could show variability and therefore explain, at least in part, morphological variation seen in the Arctic charr.

Several of the developmental genes were present in multiple copies, confirmed by our HWE analysis. These multiple copies give the false appearance of polymorphism and thus pose difficulties for the reading of sequences, so for further study of the certain genes it would be recommended to isolate one of the paralogs through DNA cloning using bacteria and then sequence that DNA to get a clear indication of the actual sequence without false positives and increased background signal obscuring the sequence. For several of the developmental genes, for example *Runx1* and *Twist2*, the results were unclear due to the paralogs present and one course of action would be to isolate one of the paralogs and then check for polymorphism. However due to the limited variability present within *Runx1* and *Twist2* it may be wiser to simply move on to other candidates instead of chasing genes that show very little, if any, variability.

The developmental gene that showed the strongest indication of variability between morphs was the Fibroblast growth factor receptor 1 oncogene partner 2 or *Fgop2* (also referred to as *Fgfr1op2* in other literature (Lin et al. 2010)). It is a small cytoskeleton-associated molecule that has, in rats, been associated with the stimulation of wound closure, through facilitating fibroblasts which in turn were responsible for the actual wound closure, and it has also been connected with cell motility (Lin et al. 2010). This was shown to be the case both *in vitro* and *in vivo* and this was carried out in rats which doesn't give us the protein function in Arctic charr; however in chickens the FGF and FGFR signalling pathways have been shown to be vital to the normal formation of mandibular bone and cartilage of the mandibular arch (Mina, Havens and Velonis, 2007) and therefore it could be interesting to see how exactly the Fgop2 protein connects to these signalling pathways. The product of the gene is connected to *Fgfr1* (Fibroblast growth factor receptor 1) in that it takes part in its signalling pathway and it has also been shown to be an oncogene, which means that certain mutations in *Fgop2* have been linked to an increased chance of cancer (Gu et al. 2006).

The variation that was found in the *Fgop2* gene fragment, which was a two base deletion as mentioned in the results chapter, was present in an intron of the *Fgop2* gene. This was seen by comparing my sequences with the mRNA dataset that other members of the research group had compiled and the comparison revealed that the part of the gene that the deletion occurred in was not present in the mRNA sequences. The mRNA data covered the majority of the sequence that I had obtained and therefore the exclusion of the deletion site in the mRNA dataset can be said with confidence to reveal that the site of variation present between the morphs is not present in mRNA and therefore occurs in an intron. This means that this deletion is not affecting the actual protein structure but may be an

indication of divergence on the *Fgop2* gene itself in the evolution of the morphs; it could be a by-product of selection on another site in the gene (linkage disequilibrium). Therefore the mutational difference seen between the two morphs could simply be hitchhiking along with a mutation that is undergoing selection. The next step regarding *Fgop2* study would be to study the transcription during development in Arctic charr and also further study of the variation within the gene itself with possible identification of the site that the deletion detailed above is hitchhiking along with. It would also be possible to identify the location and time of the proteins translation and study the function that the protein has in Arctic charr. Although first we would need to carry out a verification of the divergence signal, to see if it is actually strong or is just chance and this would be carried out by looking at more fishes.

The other gene that showed potential allele frequency difference between the two Arctic charr morphs was *Pth*, which I thought at first was the parathyroid hormone gene but when the sequence was BLATed turned out to be the gene encoding peptidyl-tRNA hydrolase. The gene *Pth* is not connected with development but rather is involved in protein synthesis and therefore you wouldn't expect to see differences between such closely related morphs, at least not in relation to vital protein function. However by comparing the mRNA of the *Pth* gene with the sequence obtained in this study, it was found that not only was the site of interest present within the protein-coding region of the gene, but it was also a 7 base deletion. It is very interesting that the morphs have such a large variation within an enzyme connected to protein synthesis and this needs further study. Allele PCR would be possible in the future as a genotyping method, as the 7 base deletion present in the gene fragment can be accounted for or excluded in the primers which creates a very reliable amplification.

Although only a small number of individuals were sequenced there is no reason to examine this part of *Otx2* further as a sample of sixteen should suffice to see if variability is present or not for this gene fragment. However this does not mean that the gene should be excluded from further research as the variability could be located within another part of the gene such as a regulatory element of the gene, as was shown with *Bmp4* in cichlids and *Pitx1* in threespine stickleback (Terai, Morikawa and Okada, 2002, Chan et al. 2010). The *Otx2* gene fragment was a very clean and background free sequence, which suggests that paralogs of the gene are not present in the genome, or that they are very similar. However the majority for the other genes sequenced and discussed above there is a high fraction of gene duplicates and this is due to the whole genome duplication event that has been proven to have occurred in Salmonidae (Jaillon et al. 2004).

There is a known phenotypic variation between sexes of Arctic charr, as Jónsson and Skúlason (2000) have shown with regard to two morphs that are found in Vatnslíðarvatn. The variations between those morphs are statistically significant and show themselves in the size of certain body parts. However, differences between sexes of Icelandic Arctic charr are minimal when compared to those seen between morphs and are therefore unlikely to influence morph comparison (Snorrason et al. 1994). The focus at present is on genotypic differences between morphs however in future research it would be prudent to see if genotypes for development genes show significant variation between sexes. This would only increase confidence in the results and would remove the potential for gender variation influencing data.

In conclusion the potential for variation in developmental genes between the two morphs has been witnessed with certainty in two genes (*Pth* and *Fgop2*) and possibly in two other

genes (*Runx1* and *Twist2*). This is quite a high fraction and so it could be worthwhile to study more genes involved in these processes and perhaps a set of random genes with functions not assumed to affect ecological specification or developmental differences. However with the sequenced genes what lies ahead is a more in depth study of these genes with a larger number of individuals sequenced. Further research will give an idea of the haplotype frequencies for these developmental genes, and then differentiation between the two morphs, dwarf benthic and small pelagic, will give an idea of possible development genes that could lie behind the witnessed morphological differences between said morphs.

Bibliography

Adams, C.E. and Huntingford, F.A. 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

Albertson, R.C., Streelman, J.T. and Kocher, T.D. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences* 100, no. 9 (4): 5252-5257. doi:10.1073/pnas.0930235100

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, no. 3 (May): 363-377. doi: 10.1046/j.1420-9101.2003.00531.x

Brakefield, P.M. 2005. Bringing evo devo to life. *PLoS Biol* 3(10): e340. doi:10.1371/journal.pbio.0030340

Chan, Y. F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M. et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science (New York, N.Y.)* 327, no. 5963 (January 15): 302-305. doi:10.1126/science.1182213

Conejeros, P. et al., 2008. MH class IIalpha polymorphism in local and global adaptation of Arctic charr (Salvelinus alpinus L.). *Immunogenetics*, 60(6), 325-337. doi:10.1007/s00251-008-0290-6

Danzmann, R.G., Ferguson, M.M., Skúlason, S., Snorrason, S.S., Noakes, D.L.G. 1991. Mitochondrial DNA diversity among four sympatric morphs of Arctic charr, Salvelinus alpinus L., from Thingvallavatn Iceland. *J. Fish Biol.* 39, 649-659. doi: 10.1111/j.1095-8649.1991.tb04395.x

Day, T., Pritchard, J. and Schluter, D. 1994. A comparison of two sticklebacks. *Evolution* **48**: 1723–1724. doi:10.1016/S0169-5347(00)89135-1

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, no. sa (12): 175-185. doi:10.1111/j.1095-8649.1999.tb01054.x

Falconer, D.S. and Mackay, T.F.C. 1996. *Introduction to quantitative genetics. Fourth edition*. Longman group limited. Essex, England.

Ferguson, M.M., Noakes, D.L.G. and Romani, D. 1983. Restricted behavioural plasticity of juvenile lake charr, Salvelinus namaycush. *Environmental Biology of Fishes* 8, no. 2 (2): 151-156. doi:10.1007/BF00005181

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, no. 2 (2): 281-297. doi:10.1111/j.1095-8649.1989.tb03309.x

Futayma, D.J. 2009. *Evolution*. Sinauer Associates Inc. 2nd Edition. Sunderland, USA.

Gíslason, D, Ferguson, M.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, no. 12: 2229-2234. doi:10.1139/cjfas-56-12-2229

Genedoc. (http://www.nrbsc.org/gfx/genedoc/).

Grant, P.R. and B.R. Grant. 2002. Unpredictable Evolution in a 30-Year Study of Darwin's Finches. *Science* 296, no. 5568 (4): 707-711. doi:10.1126/science.1070315

Gu, T.L., Goss, V.L., Reeves, C., Popova, L., Nardone, J., MacNeill, J., Walters, D.K. et al. 2006. Phosphotyrosine profiling identifies the KG-1 cell line as a model for the study of FGFR1 fusions in acute myeloid leukemia. *Blood* 108, no. 13 (12): 4202-4204. doi:10.1182/blood-2006-06-026666

Gunn, J.M. and Noakes, D.L.G. 1987. Latent Effects of Pulse Exposure to Aluminum and Low pH on Size, Ionic Composition, and Feeding Efficiency of Lake Trout (Salvelinus namaycush) Alevins. *Canadian Journal of Fisheries and Aquatic Sciences* 44, no. 8: 1418-1424. doi:10.1139/f87-170

Hatfield, T. 1997. Genetic Divergence in Adaptive Characters Between Sympatric Species of Stickleback. *The American Naturalist* 149, no. 6 (6): 1009-1029. doi:10.1086/286036

Hindar, K. and Jonsson, B. 1993. Ecological polymorphism in Arctic charr. *Biological Journal of the Linnean Society* 48, no. 1 (1): 63-74. doi:10.1006/bijl.1993.1006

Imre, I., McLaughlin, R.L. and Noakes, D.L.G. 2002. Phenotypic plasticity in brook charr: changes in caudal fin induced by water flow. *Journal of Fish Biology* 61, no. 5 (11): 1171-1181. doi:10.1111/j.1095-8649.2002.tb02463.x

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature 431: 946-957.* doi:10.1038/nature03025

Johnston, I.A., Abercromby, M., Vieira, V.L.A., Sigursteindóttir, R.J., Kristjansson, B.K., Sibthorpe, D. and Skúlason, S. 2004. Rapid evolution of muscle fibre number in post-glacial populations of Arctic charr Salvelinus alpinus. *Journal of Experimental Biology* 207, no. 25 (12): 4343-4360. doi:10.1242/jeb.01292

Jónsson, B. and Skúlason, S. 2000. Polymorphic segregation in Arctic charr Salvelinus alpinus (L.) from Vatnshlidarvatn, a shallow Icelandic lake. *Biological Journal of the Linnean Society* 69, no. 1 (1): 55-74. doi:10.1111/j.1095-8312.2000.tb01669.x

Kapralova, K.H, Morrissey, M.B., Kristjánsson, B.K., Ólafsdóttir, G.Á., Snorrason, S.S. and Ferguson, M.M. 2011. Evolution of adaptive diversity and genetic connectivity in Arctic charr (Salvelinus alpinus) in Iceland. *Heredity* 106, no. 3 (1): 472-487. doi:10.1038/hdy.2010.161

Kerckhove, D.D., Mclaughlin, R.L. and Noakes, D.L.G. 2006. Ecological mechanisms favouring behavioural diversification in the absence of morphological diversification: a theoretical examination using brook charr (Salvelinus fontinalis). *Journal of Animal Ecology* 75, no. 2 (3): 506-517. doi:10.1111/j.1365-2656.2006.01071.x

Kocher, T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews. Genetics* 5, no. 4 (April): 288-298. doi:10.1038/nrg1316

Kuratani, S., Matsuo, I. and Aizawa, S. 1997. Developmental patterning and evolution of the mammalian viscerocranium: genetic insights into comparative morphology. *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 209, no. 2 (June): 139-155. doi:10.1002/(SICI)1097-0177(199706)209:2<139::AID-AJA1>3.0.CO;2-J

Lanctot, C., Moreau, A., Chamberland, M., Tremblay, M.L. and Drouin, J. 1999. Hindlimb patterning and mandible development require the Ptx1 gene. *Development* 126, 1805-1810.

Lin, A., Hokugo, A., Choi, J. and Nishimura, I. 2010. Small cytoskeleton-associated molecule, fibroblast growth factor receptor 1 oncogene partner 2/wound inducible transcript-3.0 (FGFR1OP2/wit3.0), facilitates fibroblast-driven wound closure. *The American Journal of Pathology* 176, no. 1 (January): 108-121. doi:10.2353/ajpath.2010.090256

Mansouri, A, Stoykova, A., Torres, M. and Gruss, P. 1996. Dysgenesis of cephalic neural crest derivatives in Pax7-/- mutant mice. *Development (Cambridge, England)* 122, no. 3 (March): 831-838.

McLaughlin, R.L. 2008. Behavioural diversification in brook charr: adaptive responses to local conditions. *Journal of Animal Ecology* 70, no. 2 (7): 325-337. doi:10.1111/j.1365-2656.2001.00491.x

McLaughlin, R.L., and Grant, J.W.A. 2001. Field examination of perceptual and energetic bases for intermittent locomotion by recently-emerged brook charr in still-water pools. *Behaviour* 138, no. 5 (5): 559-574. doi:10.1163/156853901316924476

Mina, M., Havens, B. and Velonis D.A. 2007. FGF signaling in mandibular skeletogenesis. *Orthodontics & Craniofacial Research*, 10(2), 59-66. doi:10.1111/j.1601-6343.2007.00385.x

Noakes, D.L.G. 2008. Charr truth: sympatric differentiation in Salvelinus species. *Environmental Biology of Fishes* 83, no. 1 (7): 7-15. doi:10.1007/s10641-008-9379-x

Orr, A. H. 2001. The genetics of species differences. *Trends in Ecology & Evolution* 16, no. 7 (7): 343-350. doi:10.1016/S0169-5347(01)02167-X

Phred, Phrap and Consed. (http://www.phrap.org/phredphrapconsed.html).

Peichel, C.L. et al. 2001. The genetic architecture of divergence between threespine stickleback species. *Nature*, 414(6866), 901-905. doi:10.1038/414901a

Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P. and Rudnicki, M.A. 2000. Pax7 is required for the specification of myogenic satellite cells. *Cell* 102, no. 6 (September 15): 777-786. doi: 10.1016/S0092-8674(00)00066-0

Sibthorpe, D., Sturlaugsdóttir, R., Kristjansson, B.K., Thorarensen, H., Skúlason, S. and Johnston, I.A. 2006. Characterisation and expression of the paired box protein 7 (Pax7) gene in polymorphic Arctic charr (Salvelinus alpinus). *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology* 145, no. 3-4 (December): 371-383. doi:10.1016/j.cbpb.2006.08.013

Skulason, S. and Smith, T.B. 1995. Resource polymorphisms in vertebrates. *Trends in Ecology & Evolution (Personal Edition)* 10, no. 9 (September): 366-370. doi:10.1016/S0169-5347(00)89135-1

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 (12): 225-232. doi:10.1111/j.1095-8649.1989.tb03065.x

Skúlason, S., Snorrason, S.S., Noakes, D.L.G. and Ferguson, M.M. 1996. Genetic basis of life history variations among sympatric morphs of Arctic char, Salvelinus alpinus. *Canadian Journal of Fisheries and Aquatic Sciences* 53, no. 8 (8): 1807-1813. doi:10.1139/f96-098

Smith, T.B. 1987. Bill size polymorphism and intraspecific niche utilization in an African finch. *Nature* 329, no. 6141 (10): 717-719. doi:10.1038/329717a0

Snorrason, S.S., Skúlason, S., Jonsson, B., Malmquist, H.J., Jónasson, P.M., Sandlund, O.T. and Lindem, T. 1994. Trophic specialization in Arctic charr Salvelinus alpinus (Pisces; Salmonidae): morphological divergence and ontogenetic niche shifts. *Biological Journal of the Linnean Society* 52, no. 1 (5): 1-18. doi:10.1111/j.1095-8312.1994.tb00975.x

Snucins, E.J., Curry, R.A. and Gunn, J.M. 1992. Brook trout (Salvelinus fontinalis) embryo habitat and timing of alevin emergence in a lake and a stream. *Canadian Journal of Zoology* 70, no. 3: 423-427. doi:10.1139/z92-064

Sokal, R.R. and Rohlf, F.J. (1994). *Biometry: the principles and practice of statistics in biological research.*, 3rd edition. New York: Freeman. ISBN 0-7167-2411-1

Terai, Y., Morikawa, N. and Okada, N. 2002. The evolution of the pro-domain of bone morphogenetic protein 4 (Bmp4) in an explosively speciated lineage of East African cichlid fishes. *Molecular Biology and Evolution* 19, no. 9(sept):1628-1632.

Terai, Y., Morikawa, N., Kawakami, K. and Okada, N. 2003. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. *Proceedings of the National Academy of Sciences*, 100(22), 12798-12803. doi:10.1073/pnas.2132833100

Volpe, J.P. and Ferguson, M.M. 1996. Molecular genetic examination of the polymorphic Arctic charr Salvelinus alpinus of Thingvallavatn, Iceland. *Molecular Ecology* 5, no. 6 (December): 763-772. doi: 10.1111/j.1365-294X.1996.tb00372.x

West-Eberhard, M. 2003. *Developmental plasticity and evolution*. Oxford University Press.

Wilson, D. and Yoshimura, J. 1994. On the Coexistence of Specialists and Generalists. *The American Naturalist* 144, no. 4 (October 1): 692-707.

Yang, and Bielawski. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution (Personal Edition)* 15, no. 12 (December 1): 496-503.

UNEP/GRID-Arendal, Arctic char species complex, distribution map, UNEP/GRID-Arendal Maps and Graphics Library, http://maps.grida.no/go/graphic/arctic-char-species-complex-distribution-map (Accessed 15 September 2010)

Appendix A

Confirmation of amplified genes by BLAT on genome.ucsc.edu, using the Fugu assembly (version: Oct. 2004 (JGI 4.0/fr2) and ncbi BLAST on blast.ncbi.nlm.nih.gov/Blast.cgi using the nucleotide collection database.

BMP4 (*Bone morphogenetic protein 4*) – Fugu blat

Identity = 87,2%

```
00000005 caccaggagcccggggaggactgggagcagctacgccccctgctggtcac 00000054
>>>>>>>> |||||||| || || ||||||||||||||| |||||||||| || ||||| >>>>>>>>
37931147 caccaggaacctggagaggactgggagcagatacgccccctccttgtcac 37931196

00000055 cttcggcca 00000063
>>>>>>>> |||| |||| >>>>>>>>
37931197 cttcagcca 37931205


00000160 gtggacttcagcgatgtgggctggaatgactggatagtggcgcccccagg 00000209
>>>>>>>> |||||||| ||||| |||||||||||||||||||||||||| ||||| || >>>>>>>>
37931303 gtggactttagcgacgtgggctggaatgactggatagtggcacccccctgg 37931352

00000210 gtaccaggcatactactgccatggggagtgccccttccccctggcagacc 00000259
>>>>>>>>  || || || || |||||||| ||||| || || |||||||||||| || | >>>>>>>>
37931353 ttatcaagcctattactgccacggggattgtccgttccccctggcggatc 37931402

00000260 acctgaactctaccaaccacgccatcgttcagacgttggtgaactcggtg 00000309
>>>>>>>> | |||||||| |||||||||||||| ||||||||  ||||||| || ||| >>>>>>>>
37931403 atctgaactcgaccaaccacgccattgttcagacactggtgaattccgtg 37931452

00000310 aacaccaacattcccaaggcctgctgcgtgcccacggagctcagtgccat 00000359
>>>>>>>> |||| |||||||||||||||||||||||||| || |||||||||||||| >>>>>>>>
37931453 aacagcaacattcccaaggcctgctgcgtgccaaccgagctcagtgccat 37931502

00000360 ctccatgctctac 00000372
>>>>>>>>  || ||||||||| >>>>>>>>
37931503 ttcaatgctctac 37931515
```

ETBR2 (endothelin B receptor-like protein 2)– Fugu blat

Identity = 91,7%

```
000000025 cactcctcacagcagcagc.gcagcagtccatgaa 000000058
>>>>>>>>> ||||||| |||||||||| |||||||||||||| >>>>>>>>>
119390799 cactcctgacagcagcagcagcagcagtccatgaa 119390833


000000084 cacagcagcagcactggcgtcaccgacgacttgaagaacaa 000000124
>>>>>>>>> |||||||||||||| || ||||||||||||||||||||||| >>>>>>>>>
119390970 cacagcagcagcaccggggtcaccgacgacttgaagaacaa 119391010


000000366 ggccagctggcacagcagggtgaagacca.cgggcaggcagaagtagcag 000000414
>>>>>>>>> |||||| |||||||| | ||||||||| | |||| |||||||||||||| >>>>>>>>>
119391260 ggccagttggcacaggatggtgaagacgaccggg.aggcagaagtagcag 119391308

000000415 ccaaagtaccaccacatgcg 000000434
>>>>>>>>> || ||| |||||||||||| >>>>>>>>>
```

```
119391309 ccgaagcaccaccacatgcg 119391328
```

OTX2 (orthodenticle homolog 2)

GENE ID: 30501 otx2 | orthodenticle homolog 2 [*Danio rerio*]

```
Score = 60.8 bits (66), Expect = 6e-06
Identities = 65/81 (81%), Gaps = 5/81 (6%)
Strand=Plus/Minus

Query  9     CTTATAAACAT-TTTTCATAAATTAAATAGTGCATCCAGGTTCGCCTCACAGGCGcgcac  67
             ||||||||||  ||||| ||||| || |||||||||||| ||||  ||  || |||  |
Sbjct  1728  CTTATAAACACCTTTTCTTAAATAAAGTAGTGCATCCAGG-TCGC--CATGGGGGCGTTC  1672

Query  68    acgcacgcacacacgcacaca  88
             | |||| |||||||||| ||
Sbjct  1671  A-GCACACACACACGCACGCA  1652
```

Pth (peptidyl-tRNAhydrolase 1 homolog)

```
Length=168064

Score = 64.4 bits (70), Expect = 9e-07
Identities = 82/112 (74%), Gaps = 1/112 (0%)
Strand=Plus/Plus

Query  296     AACCTTCTGCTCCTCCTGAGAGAAGCGGCCCAGAACATGCCGGTCCACTGATGTTTTACC  355
               ||| || || ||||| |  || || || |||| |||||  ||||| |  |  |||||| |
Sbjct  154165  AACGTTTTGTTCCTCTTTTGAAAATCGTCCCAAAACATAGCGGTCGATGGGCGTTTTATC
154224

Query  356     TGATGGTCTGCCAATCCCAACACGCAGTCTGGGCATCACCTGGGGGGCACAA  407
                || |||| |||||| |||| |||| | |||||||||| | ||||||||
Sbjct  154225  CGACGGTCGGCCAATGCCAATGCGCAACCGCGGCATCACCT-GTGGGCACAA  154275
```

FGFR4 (fibroblast growth factor receptor 4)

GENE ID: 100000160 fgfr4 | fibroblast growth factor receptor 4 [*Danio rerio*]

```
Score = 161 bits (178), Expect = 1e-36
Identities = 132/162 (82%), Gaps = 0/162 (0%)
Strand=Plus/Plus

Query  22    CTCCCTGTCCAGAAACTCTCCAAGTTCCCCCTCCGCAGACAGTACTCTGTGGAKTCTAAC  81
             ||||| || |||||||||||||| || ||  | |||||||||| || ||||| || ||
Sbjct  1781  CTCCCAGTACAGAAACTCTCCAAATTTCCTTTGCGCAGACAGTATTCAGTGGAGTCCAAT  1840

Query  82    TCRTCAGGGAAGTCCAGTGCKTCTCTGATGAGAGTGGCCCGSCTGTCTTCCAGCTGCTCC  141
             || || || || || ||||| || |||||||| ||||| || || || ||||| || |||
Sbjct  1841  TCTTCTGGAAAATCAAGTGCGTCACTGATGAGGGTGGCTCGTCTTTCCTCCAGTTGTTCC  1900

Query  142   CCCATGCTGGCCGGAGTCATGGAGTTTGAACTGCCTTACGAC  183
             || |||||||| ||||| ||||||||||||||||||||| |||
Sbjct  1901  CCAATGCTGGCTGGAGTTATGGAGTTTGAACTGCCTTATGAC  1942
```

FGOP2(fibroblast growth factor oncogene 2) – Fugu blat

```
Identity = 87,8%

000000237 tagagctcatcatgactaaatacagggagcaggtcttcagactcctcatg 000000286
>>>>>>>>> |||||||||| ||||| |||||||||||||||| ||||| || |||||| >>>>>>>>>
338630187 tagagctcattatgaccaaatacagggagcaggttttcaggctgctcatg 338630236

000000287 gccagtaagaaggatgacccagccattgtcacccaattaaaggagcagca 000000336
>>>>>>>>> ||||| |||||||| ||||| |||||||| | |||  | |||||||||| >>>>>>>>>
338630237 gccagcaagaaggacgaccctgccattgtgagccagctgaaggagcagca 338630286
```

```
000000337 caccac 000000342
>>>>>>>>> |||||| >>>>>>>>>
338630287 caccac 338630292
```

IGFP1 (insulin-like growth factor binding protein-1) – Fugu blat

Identity = 84,5%

```
000000008 tgatatgtttgtgcctggccaccagatgcatacttgt 000000044
<<<<<<<<< ||||||||||||||  |||||| |   |||||||||| <<<<<<<<<
180746187 tgatatgtttgtgcgaggccacaatgcgcatacttgt 180746151


000000051 tttaagtatttcccc 000000065
<<<<<<<<< ||||||||||||||| <<<<<<<<<
180746145 tttaagtatttcccc 180746131


000000070 cccattgaggaaa 000000082
<<<<<<<<< |||||||||||| <<<<<<<<<
180746125 cccattgaggaaa 180746113


000000108 tgtcaacatgt....tgaatga 000000125
<<<<<<<<< ||||||||||    ||||||| <<<<<<<<<
180746091 tgtcaacatgtcaaatgaatga 180746070


000000132 ggctgtg 000000138
<<<<<<<<< ||||||| <<<<<<<<<
180746067 ggctgtg 180746061


000000247 gagttatttagtcagaaagacaaatgcatac 000000277
<<<<<<<<< |||||||||||||  || ||||||| |||| <<<<<<<<<
180745963 gagttatttagtca..aacacaaatgtatac 180745935


000000279 tgttgccctatttac 000000293
<<<<<<<<< ||||||||||||||| <<<<<<<<<
180745932 tgttgccctatttac 180745918


000000300 aaaataggaaac 000000311
<<<<<<<<< |||||||||||| <<<<<<<<<
180745909 aaaataggaaac 180745898


000000318 aatctaaa 000000325
<<<<<<<<< |||||||| <<<<<<<<<
180745885 aatctaaa 180745878
```

RUNX1 (runt-related transcription factor 1)

GENE ID: 58126 runx1 | runt-related transcription factor 1 [Danio rerio]
(Over 10 PubMed links)

Score = 235 bits (260), Expect = 1e-58
Identities = 261/346 (76%), Gaps = 21/346 (6%)
Strand=Plus/Plus

```
Query  10    caccctgctgggcgctaccacacctacctccccccaccctaccccacaaacgcccccca  69
             || |||||| || |||||||||| ||||| || ||  | |||||| |   | || | ||
Sbjct  1058  CAGCCCTGCGGGACGCTACCACACATACCTGCCGCCGGCGTACCCCGCGGGCTCCTCGCA  1117
```

```
Query  70    GGGCCAAGGTGGACCMTTCCAGGCCGGTTCCTCTCCCTACCACCTGTAYTATAGCACGGC  129
             || ||| | |||| | ||||||||  | || || || || ||||||| || ||||  ||
Sbjct  1118  GGCCCAGGCTGGAGCCTTCCAGGCGAGCTCGTCCCCATATCACCTGTACTACAGCAGCGC  1177

Query  130   CGCTGGGTCYTACCAATTCTCCATGATG----GCGG-GGGGAGGC----GGGGAGCGGTC  180
             ||| || || ||||| ||||||||||||    |||| |||| |||    |||||||| ||
Sbjct  1178  CGCCGGCTCCTACCAGTTCTCCATGATGCCCAGCGGAGGGGCGGCAGCAGGGGAGCGCTC  1237

Query  181   TCCTCCGAGGATATTCCCTCCCTGCACCAACGCATCCACAGGCTCCTCCCTYCTGCACCC  240
              || ||  |   | |||  |||||||||||||| |||||||||||| |||| ||||||||
Sbjct  1238  GCCGCCCCG---AATCCTACCCTGCACCAACGCCTCCACAGGCTCCGCCCTCCTGCACCC  1294

Query  241   GTCTCTGCCAAATCAGAGCGAAGGGGTGGGGGTGGAAGCGGAGGGAAGTCACAGTAGCTC  300
              || || || |||||||||||||| ||   |||||| |||||||||||| ||||| || ||
Sbjct  1295  CTCGCTCCCCAATCAGAGCGAAGGAGT---GGTGGAGGCGGAGGGAAGCCACAGCAGTTC  1351

Query  301   CCCTACCAATATG---GTACCCGAGGCCGTATGGCGGCCATATTGA  343
              || ||||  |||   |||   |||||||| ||| ||||||| |||
Sbjct  1352  GCCAACCAGCATGTCTGTA---GAGGCCGTCTGGAGGCCATACTGA  1394
```

WNT9 (wingless-type MMTV integration site family) – Fugu blat

Identity = 89,6%

```
000000001 cttcccaagttccgcgagctcggctacattctcaaagagaagtatgccca 000000050
<<<<<<<<< ||||||||||||||||||||||| |||||||||||||||||||||| ||| <<<<<<<<<
236775372 cttcccaagttccgcgagctcggttacattctcaaagagaagtatgtcca 236775323

000000051 tgctgtgaccgtggaaccggtcaaagccagccgcaacaagcggcccaaa. 000000099
<<<<<<<<<  || |||  ||||||| | |||||||||||||||||||||| ||||||  <<<<<<<<<
236775322 cgcggtgcacgtggaatcagtcaaagccagccgcaacaagcg.cccaaaa 236775274

000000100 ttcctcaagatcaagaag.cctactcctaccggaagcccatggacacaga 000000148
<<<<<<<<< ||||||||| | |||||| ||||||| |||||||||||||||| || || <<<<<<<<<
236775273 ttcctcaagctgaagaagccctactcgtaccggaagcccatggatacgga 236775224

000000149 ccttgtgttcattgagaagtcccctaactactgcgaggcggaccc 000000193
<<<<<<<<< ||| |||| ||| || ||||| || |||||||||||||||||| <<<<<<<<<
236775223 cctggtgtacatcgacaagtctcccaactactgcgaggcggaccc 236775179
```

TWIST2 (Twist-related protein 2)

Identity = 89,2%

```
000000021 gaagaggagacacagtaaaaagtcgagcgaggactgcag 000000059
<<<<<<<<< ||||||||| |||||||||| || ||||||||| |||| <<<<<<<<<
246594762 gaagaggaggcacagtaaaaaatccagcgaggacagcag 246594724


000000178 aacgaagccttctcgtctttacgcaaaatcatccccactctaccctcgga 000000227
<<<<<<<<< ||||| |||||| |||||||||||||||||||||||| || || ||||| <<<<<<<<<
246594596 aacgaggccttcgcgtctttacgcaaaatcatccccacgctgccgtcgga 246594547

000000228 taaactgagcaagatccagacactaaaactggcctccagatacatagact 000000277
<<<<<<<<<  |||||||||||||| ||||| || || ||||| ||||| ||||| |||| <<<<<<<<<
246594546 caaactgagcaagatacagacgctgaagctggcgtccaggtacattgact 246594497

000000278 tcctctatcaggtgctgcaaagcgacgagatggacaacaagatgtcgagc 000000327
<<<<<<<<< | |||| |||||||||||| |||||||||||||||||||||||| ||| <<<<<<<<<
246594496 ttctctgtcaggtgctgcagagcgacgagatggacaacaagatgtccagc 246594447

000000328 tgcagctacgttgcgcacgagagactcagttatgctttctc 000000368
<<<<<<<<< |||||||||||| |||||||| |||||||||||| |||||||| <<<<<<<<<
246594446 tgcagctacgtcgcgcacgaaagactcagttacgctttctc 246594406
```

Eng2 (engrailed 2)

GENE ID: 30243 eng2a | engrailed 2a [*Danio rerio*]

Score = 158 bits (174), Expect = 2e-35
Identities = 209/294 (72%), Gaps = 0/294 (0%)
Strand=Plus/Minus

```
Query  9    TGAGAACTATCCGAGTCCGAGCTTAGGCACTGATCTCCATTCTCTCCGCGGGGTTTCAGG  68
            || |||||||||||  || ||  |  | |||||||||  |||||||  ||||||||  ||||||
Sbjct  423  TGTGAACTATCCGATTCTGAACCCAAGCACTGATCCACATTCTCCCCGCGGGGCTTCAGG  364

Query  69   GGCTCTTCKGTRRCTATTTCAGCCTTATTGGCCACGCCGCCTGGATGAGGAGTGGAAGTT  128
            || |||||  |  |||| || ||||  ||  |  ||| ||  | ||| |  ||||||| |
Sbjct  363  GGTTCTTCGCTCTCTATCTCTGCCTCCTTTCCTCCGCTGCTCGTATGCGTTGTGGAAGCT  304

Query  129  CCTTCCGTTGGCACAGTACTCCCCACTTGCTCCGYTCTTGGAACAGCCGGGCTATGGTTC  188
             ||||||  ||| ||||||||  |||||  ||  |||  ||||||  | |  ||| | || |||
Sbjct  303  TCTTCCGCTGGTACAGTACTTCCCACCTGTCCCGTGCTTGGACCCGTAGGGTTGTGATTC  244

Query  189  TCTCTGGTARCGTGACTATTTTCTTCMTGMTKTRTGTTRKCTTCTTTTTTCCGCCCAAAG  248
            ||||  |   |||||  | |  || |   |     | ||  |||||||||| || |||||
Sbjct  243  TCTCGTGCGCCGTGATTGTCCTCATAGCGCGTGATATTCGCTTCTTTTTTACGGCCAAAA  184

Query  249  TCCGGCCGTAAGATATTATCGATGAAGAAATTGGTGACTCGATGCGGGATCTGC  302
            || || || | ||| || |||||||||||  || ||||  |||  ||  ||  | ||||
Sbjct  183  TCTGGTCGCAGGATGTTGTCGATGAAGAAGTTTGTGATTCGGTGTGGCAGCTGC  130
```

Appendix B

Appendix B Table. An overview of all the developmental candidate genes and the success in creating primers, getting the primers to work in PCR and then sequencing said PCR products. If the PCR has worked but it hasn't been sequenced then that is due to the size of the PCR product not being viable, and that would mean it was too large (often over 1000bp), or that the PCR was unstable and therefore didn't always work for all individuals. The sequences for the primers are included.

Genes	Primers	Forward Primer	Reverse Primer	**PCR**	**Sequenced**
BMP2	Yes	F-GCACAATGGCATGATTGGTA	R-CCATCGCATCAACATCTACG	No	No
BMP4	Yes	F-CAAGTCCTGCTGGGAGAGAG	R-GGGCTTCAGAACCTCGTACA	Yes	Yes
BMP4	Yes	F2-GTcCTGCACCTCAACCAgAC	R2-TCcAgGTAGAGCATgGAgATG	No	No
brachyury (BRAC)	Yes	F-CGAGAAAGGAGATGCCTCAG	R-TGAGTCTCGGGGAAAGACTG	No	No
col1a1	Yes	F-CCCTGACTCAGAAGGTCGAG	R-TGTAGGCGATGCTGTTCTTG	No	No
col2a1	Yes	F-TCAGCCAGAAGGTGGAGAAC	R-ACGTCCATAGGAGCGATGTC	No	No
DLX1	Yes	F-CAAATGTCGCCTTCCTCAAT	R-GTTCCACCACCTTGCTTCAT	No	No
DLX5	Yes	F-TCACTGGAATTGTCCGTTCA	R-TCCGAGAACCAAAGAGTGCT	No	No
Engrailed 2	Yes	F-GTCGGCGGATGAGTCAAATA	R-ATCTGGTGCAGTACACCCAAG	Yes	Yes
ETBR2	Yes	F-GAGCTGTCCTTGGCTTTGTC	R- ACGCCCTGGTCATCAACTAC	Yes	Yes
ETBR2	Yes	F2-gctcttgagtttggcctcag	R2-GTgGTCTTCACCcTgCTGTG	No	No
FGFR 2	Yes	F-GATGCTGGGGAGTACACCTG	R-TGCTTGCTGAGTTTGTGGAC	No	No
FGFR 4	Yes	F-TATCTACGCGTCTGGCTTCC	R-GGGTCGTAAGGCAGTTCAAA	Yes	Yes
FGOP2	Yes	F-GGAGATTGAATCGCTGAACC	R-TCTATGTGCGCTTGCATTTC	Yes	Yes

Appendix B Table – continued.

IGF-1	Yes	F-TCCCCAGAACTCTCAACACC	R-AAGGAGGAAAGGGGTTGCT	No	No
IGF-1	Yes	F2-tccccagaactctcaacacc	R2-AAGGAGGAAAGGggTTGCT	No	No
IGFBP-1	Yes	F-CAGAAGACCTGCACAGACCA	R-TGCCCTCAAACCTACACACA	Yes	Yes
IGFBP-1	Yes	F2-tcccagaactctcaacacc	R2-AAGGAGGAAAGGGGTTGCT	No	No
OTX2	Yes	F-GCCAAAACACTTAGGGGACA	R-AAAGGCATCGTTTTCCAATG	Yes	Yes
OTX2	Yes	F2-ttagccaaggattgcagctt	R2-gggactgagcttcagtggtc	No	No
PAX9	Yes	F-TGGGACCAAAGTACCGACTC	R-TTCTCTCAATCTGCCCGTCT	No	No
PAX9	Yes	F2-tgggaccaaagtaccgactc	R2-ttctctcaatctgcccgtct	No	No
PTH	Yes	F-CGTGGTGATAGGCATGTGTC	R-CAGTGGGGACTGCATATCCT	No	No
PTH	Yes	F2-gtgtgcttactcggcattca	R2-gcatcaagcctgagcacata	Yes	Yes
PTH2	Yes	F-TGGACGGTAGATGAAGACGA	R-CACAAAATCAGCTGCTCCAG	No	No
retinoic acid (RA)	Yes	F-GATGTGCCGAAGGTGTTTCT	R-CTTCCCTCTTCATGCCCATA	No	No
RORA	No			No	No
RORB	Yes	F-TGGCGTAATCACCTGTGAAG	R-CGAGGATGATGATGGTGATG	No	No
RUNX1	Yes	F-CCTACTGGGGCCTTCACATA	R-GCCAGCAAGTCTGTCAATCA	Yes	Yes
RUNX1	Yes	F2-cctactggggccttcacata	R2-gccagcaagtctgtcaatca	No	No
sox21	Yes	F-GAGCCGAGTGGAAACTTCTG	R-GCCAGCCCGTACAGTTACAT	No	No
sox9	Yes	F-GTCATACTGCGCTCTGGTCA	R-GACTTGGTCTCCAGCAAAGC	No	No
TWIST2	Yes	F-AGTTCTCCCATCTCCCCTGT	F-TCCTCCACACAGAGAAAGCA	Yes	Yes
BRAF	No			No	No
CHL1	No			No	No
CHL12	No			No	No
cyo26b	No			No	No
Ednra 1	No			No	No

Appendix B Table – continued.

Ednra 2	No			No	No
Endothelin 1	No			No	No
Follistatin-like 1b	No			No	No
Gpc4	No			No	No
IGFBP-3	No			No	No
KRAS	No			No	No
Lazarus (pbx4)	No			No	No
MSX	No			No	No
NF-Y	No			No	No
Noggin 1	No			No	No
osteocalcin (bgp)	No			No	No
PTHrp	No			No	No
RUNX2	No			No	No
SCPP	No			No	No
SNAIL	No			No	No
TBX22	No			No	No
Tbx5	No			No	No
TWIST1	No			No	No
wnt2	Yes	F-CATGAGGTTCAGCCAGTCCT	R-GTCCATCTCCTGGTTGCAGT	Yes	No
wnt9	Yes	F-TGGAATGCAAGTGTCATGGT	R-GACGTAGCAGCACCACAAGA	Yes	Yes