Life in extreme environments: adaptation and evolution of a soda lake cichlid; *Alcolapia alcalica*

Lewis White

PhD

University of York

Department of Biology

February 2021

Abstract

Alcolapia are the only vertebrates to survive the harsh conditions of Lake Natron, this requires multiple adaptations. Due to their evolution under extreme conditions they present themselves as naturally occurring biological experiments and potential model organisms in the study of adaptation to change and evolution under stressful conditions.

Alcolapia are reportedly the only 100% ureotelic teleost, unable to excrete ammonia due to their high pH environment. Lab acclimated *A. alcalica* are not 100% ureotelic as previously reported, although most waste is still excreted as urea. To accommodate this the ornithine urea enzyme, carbamoyl-phosphate synthetase (CPS), has evolved in function and expression. Usually isolated to liver tissue, in *Alcolapia* it is found in muscle and convergently evolved in function to bind ammonia, as done by terrestrial vertebrates, rather than fish CPS which binds glutamate. A proportion of *A. alcalica* nitrogenous waste was excreted as ammonia, with the ammonia transporters rhesus glycoproteins acting to move ammonia in two orthologous systems. These proteins have potentially evolved to move ammonia against a concentration gradient.

Protocols were adapted and good husbandry established for *Alcolapia,* and although no gene editing technology was effective, *A. alcalica* are clearly an interesting study animal. Differences in gene structure and gene expression were notable when comparing *A. alcalica* to model species. The presence of a poly-serine region within the MyoD1 protein of *A. alcalica* and other non-model species shows more information could be gained with more access to non-model species. This region was only present in the MyoD1 of species which retain a *MyoD2* gene. Preliminary analysis suggests this region increases protein stability. *Pax-6, Pax-6-like* and *Pax-10* show spatiotemporal subfunctionalisation of expression across the developing eye and nervous system of *A. alcalica. MyoD1* and *MyoD2* also show subfunctionalisation within muscle tissue showing a likely conserved function of these genes.

Facial view of *in situ* hybridisation showing the expression of *Sonic hedgehog (shh)* in developing *Alcolapia alcalica embryos*

Table of Content:

List of Figures:

List of Tables:

Acknowledgments

This PhD was incredibly stressful and getting through the many years of it (longer than planned thanks to Miss Rona) has left me lots of people that require thanks, and many more who are owed a drink for putting up with me. I have to give my stubbornness some serious credit as it really did push me through, supported by alcohol, great friends and frequent trips to see Sue at Vanbrugh for a coffee and slice of cake.

Thank you to Kanchon for giving me the opportunity to do this PhD and letting me run with any idea that came to mind. I would also like to thank all the people that have supported me in the lab over the years. Caitlin, for being a great friend both inside the lab and out of it, Laura who was a delight and put up with a lot of teasing and Alex for being the supportive fellow fish guy. A big thank you to Gideon who was a big part of keeping me positive throughout the PhD and another to Ali, a great friend, drinking buddy and housemate who was always around to have a good old moan with. A final lab thank you goes to Harv, genuinely a pleasure to have been able to work alongside and always a helpful person to run any idea by.

Throughout this PhD I have met some amazing people who all deserve thanks. James and Nicola were a constant source of gossip and together we lost far too many pub quizzes, the wonder cats live on. Sarah was always a source of entertainment at the other end of the lab and the many nights spent at 13s were always needed, a fantastic friend throughout. Ryan, you were there for me during the final months of write up stress and were always there to motivate me, I feel very lucky to have met you. I'd also like to give thanks to Charlotte, Alice, Carla and Raquel for support and general chats whenever they were needed.

To my parents and sister, thank you for encouraging me and keeping me going across all the degrees, and to James for actually being interested. I would also like to thank my undergraduate supervisor at Liverpool, Lynne Sneddon, who was the main inspiration behind me pursuing a PhD. Finally, I would like to give the biggest thank you to Betsy. Thank you for adopting me into your lab without even knowing what would happen. Your supervision, guidance and friendship are the main reasons I got to the end and I will always be grateful. There will be many summer beers in the future.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Signed  .....

Chapter One – Introduction

1.1 The Central Question

Extremophile organisms are naturally occurring biological experiments; access to a vertebrate species which is adapted to an extreme environment would better allow the study of the limits of evolution and the examination of core biological processes where the body and developing embryo is exposed to stressful conditions. Aquatic ectotherms, such as fish, are especially vulnerable to increased pressure from environmental fluctuations due to their dependence on surrounding water for maintaining homeostasis (Evans, Piermarini et al. 2005, Hwang, Lee et al. 2011). Current predictions of anthropogenic-driven climate change suggest that global temperatures will continue to rise and this is likely to have a large effect on freshwater ecosystems (Visser 2008, Jackson, Betancourt et al. 2009). The increased evaporation associated with rising water temperatures will result in changes to freshwater habitats, including; salt concentrations, water chemistry, availability of oxygen, and water levels (Woodward, Perkins et al. 2010). These changes will result in unprecedented physiological stresses to native species. Understanding the mechanisms by which species adapt to survive such conditions will be crucial for efforts to maintain fish populations. African cichlids are of particular interest due to their distribution across a variety of habitats, from benign to extreme, and the importance of certain species in fisheries, relevant to global food security (Turner 2007, Salzburger, Van Bocxlaer et al. 2014). For example, *Oreochromis niloticus* (Linnaeus 1758) is one of the most farmed fish worldwide, second only to carp (*Cyprinus carpio*, Linnaeus, 1758), with 4.2×10^{6}t produced in 2016 (FAO 2020). The closely related cichlid group, the *Alcolapia* species flock, present themselves as a unique radiation of cichlids which have evolved rapidly from freshwater ancestors to inhabit one of the most extreme aquatic environments supporting fish life. These fish may provide a vertebrate model of

adaptation to extreme conditions. This would be beneficial, not only to the fisheries industry, but for investigating the limits of adaptation in vertebrates and important in the study of how animals may cope under chronic stress.

"For such a large number of problems there will be some animal of choice or a few such animals on which it can be most conveniently studied."

- *Krogh's principle, The progress of physiology, Augustus Krogh 1929*

This PhD project started with the establishment of a colony of the extremophile fish *Alcolapia alcalica* in the Department of Biology at the University of York. With funding from the Fisheries Society of the British Isles (FSBI) and the Genetics Society, the three species of *Alcolapia* endemic to Lake Natron in Northern Tanzania were collected and imported to Bangor University and subsequently transported to York with approval from the departments AWERB and a Home Office licence for Dr ME Pownall (POF 245945).

1.2 Adaptation to Extreme Environments

Evolution by natural selection is the best accepted theory as to how animals change over generations (Darwin 1859). The theory of natural selection states that within a population, specific individuals will be better adapted to their environment and those individuals will either be more likely to survive and reproduce or will have an increased number of offspring. These offspring will in turn carry any selectively advantageous adaptations from their parents and go on to also be successful. For natural selection to occur there must be variation for a given characteristic within a population which selection may act upon. This variation must be heritable and must cause differences in the reproductive success of the individuals. If all of this is true then natural selection may act on organisms and result in changes over multiple generations (Darwin 1859). At the

time, it was believed that parental traits were blended in their offspring which raised doubt into Darwin's theory as it suggested specific traits could not be selected for. The mechanisms behind natural selection were better understood after the work of Gregor Mendel who discovered how traits can be passed on from one generation to the next (Mendel 1866). Named after its discoverer, Mendelian inheritance explains how different alleles of a given gene are allocated to offspring by both parents. In this way it is possible for a favourable genetic adaptation to be present in subsequent generations without the blending of traits (Mendel 1866, Fisher 1930).

Animals are adapted to a given environment with different adaptations being preferential in different environmental conditions. An excellent example of this is the threespine stickleback (*Gasterosteus aculeatus*, Linnaeus, 1758) where most populations have pelvic structures made up of serrated spines that are useful in deterring predation from gape-limited predatory fish (Reimchen 1983, Shapiro, Marks et al. 2004). Positive selection (selection acting on an advantageous characteristic or genetic variant) for this adaptation is evident as individuals with this characteristic would have been more likely to avoid predation and go on to reproduce (Vamosi and Schluter 2004) and over many generations, this structure would have become ubiquitous across the species. However, a number of freshwater populations, which are widely distributed and from independent isolation events, show partial reduction or complete loss of this structure (Bell 1987). There is an absence of gape limited predators in the environments where these populations persist which removes the selective benefit of this structure. Additionally, the changes in their environment mean that there are also active disadvantages to retaining it (Bell, Orti et al. 1993). In some populations it is hypothesised that low levels of dissolved calcium coupled with the absence of predators selects for reduction in this now non-

essential bony formation and lowers the total levels of calcium required for growth (Giles 1983). It has also been shown that the reduction in the external structure of the pelvis correlates with an increase in the presence of grasping insects, the loss of this structure minimises the area of contact and decreases the likelihood of predation (Reimchen 1980, Marchinko 2009).

The stickleback example also illustrates an important mechanism of evolutionary change. Genetic screening has since identified a single gene, *Pituitary homeobox 1 (Pitx1),* which explains the majority of the variance in pelvic size between populations of threespine stickleback (Shapiro, Marks et al. 2004, Coyle, Huntingford et al. 2007). *Pitx1* has a known function in hindlimb formation (Logan and Tabin 1999). Interestingly, the gene itself has little divergence and instead an area upstream of the gene, encoding a tissue specific enhancer, shows signs of positive selection (Chan, Marks et al. 2010). Loss or reduction of the pelvic structure is associated with different mutations in this tissue specific enhancer for *Pitx1* which cause the enhancer to be functionally inactive (Chan, Marks et al. 2010). The changes to the stickleback morphology via selection acting on particular regions of the genome is one of many examples of small changes to environmental conditions incurring advantages for particular characteristics and the evolution of adaptive qualities.

1.2.1 Extreme environments

An animals environment has a major impact on its evolution, as such, organisms which are exposed to extreme conditions are under strong selective pressure which requires rapid adaptation in order for the species to survive (Grant, Grant et al. 2017, Whitehead, Clark et al. 2017). An environment is considered extreme when either one or multiple physical or geochemical conditions is outside of a normal range (Rothschild and

Mancinelli 2001). An organism which is adapted to live in deviations from this range is known as an extremophile. To date, most literature has focused on extremophile microorganisms. This is likely because the vast majority of extremophiles are bacteria and archaea which have a shorter generation time and are able to evolve and adapt to new environmental conditions more rapidly (Cavicchioli 2002, Anitori 2012). The study of extremophile organisms is growing, with the discovery of thousands of novel enzymes now being used in biotechnology and industry (Van Den Burg 2003), as well as information gained about potential extra-terrestrial life (Rothschild and Mancinelli 2001) and the function of important proteins under stressful conditions (Laksanalamai and Robb 2004). Although the majority of this work has been carried out in prokaryotes there are a surprising number of vertebrates which also fit the definition of an extremophile (Riesch, Tobler et al. 2015). With current predictions of anthropogenic driven climate change and suggestions that the world may be facing the next big mass extinction it has never been more crucial to understand how vertebrate species can adapt to extreme and novel stressors (Visser 2008, Barnosky, Matzke et al. 2011). Hence, interest in extremophile vertebrates is growing in diverse areas of biology because they are naturally occurring environmental experiments on how organisms may adapt to change and the potential limits of adaptation (Riesch, Tobler et al. 2015).

Augustus Krogh stated that “For such a large number of problems there will be some animal of choice or a few such animals on which it can be most conveniently studied”, this principle comments on the intense use of model organisms and their selection due to their convenience rather than necessarily being the most appropriate animal for the given study (Krogh 1929, Burggren 2000). Fish species such as zebrafish (*Danio rerio;* Hamilton 1822), medaka (*Oryzias latipes*; Temminck and Schlegel 1846) and fugu (*Takifugu*

rubripes; Temminck & Schlegel, 1850) have been invaluable biological tools as model organisms due to their well annotated genomes and their rapid rates of reproduction (Stevenson, Alward et al. 2018). The large number of people that work on these models has led to the production of databases and tools which has increased the ease of their study and use as simplified models of other systems such as humans (Russell, Theriot et al. 2017). Fantastic biology has been carried out in these organisms with many important discoveries being credited to their study, however, due to their popularity much of our knowledge of specific fields is biased towards single species. For example, zebrafish have been established as a embryological tool and much of our knowledge of fish development is based solely on this species; despite known differences from humans and even other fish species they remain a much used model (Schartl 2014, Meyers 2018). Research using non-model organisms will allow wider species comparisons and a better understanding of adaptive traits (Crawford 2001). Exploring the naturally evolved adaptations in extremophile vertebrates will give insight into core biological functions, which may provide important information for future research in vertebrate biology and may even contribute to our understanding of human disease (Irwin 2010, Garbarino, Orr et al. 2015, Jorge, Borges et al. 2016).

1.2.2 The importance of studying extremophile fish species

Fish account for more than half of extant vertebrates on the planet. They occupy nearly all forms of aquatic environment and many are exposed to conditions considered extreme (Wang and Guo 2019). Aquatic ectotherms, such as fish, are significantly impacted by the changes or stressors in their environment as they rely on the surrounding water for maintaining homeostasis (Evans, Piermarini et al. 2005, Hwang, Lee

et al. 2011), even so fish have colonised and adapted to a variety of extreme habitats (Helfman, Collette et al. 2009). The high diversity of fishes and their presence in a variety of both natural and man-made extreme conditions make them ideal subjects for studying how animals survive under imposed stressors. Fishes have evolved morphological, biochemical, physiological, behavioural and developmental mechanisms allowing them to live in a variety of conditions (Wang and Guo 2019). Although many fish species would be unable to survive the extremes of temperature, pH, salinity, and even environments which seasonally dry out, some specialised animals have adapted to thrive in such conditions (Riesch, Tobler et al. 2015). These kinds of adaptations have made the study of extremophile fish species invaluable in many areas of biology.

1.2.3 Killifish, aging and senescence

Annual fish in the genus *Nothobranchius* are a well-established model in the study of aging and senescence (reviewed in Lucas-Sánchez, Almaida-Pagán et al. 2014). These killifish live in seasonal aquatic environments which dry out on an annual basis, as such members of this genus have evolved fast life-history traits and have desiccation resistant embryos (Cellerino, Valenzano et al. 2016). Due to their fast life-history they are the shortest lived vertebrate species, with *N. furzeri* (Jubb 1971) living for a maximum of only 12 weeks in laboratory conditions (Valdesalici and Cellerino 2003). This means that the entirety of their life can be easily traced and markers of aging and senescence easily monitored (Herrera and Jagadeeswaran 2004). Studies have shown that aging in these fish coincides with a number of age related pathologies (Markofsky and Milstoc 1979, Liu, Liu et al. 2017) and gene expression analysis has confirmed many similarities with aging in humans (Hsu, Chiu et al. 2008). They also exhibit age related deterioration of

dopaminergic neurons and increases in α-synuclein pathologies suggesting that they may be useful in revealing mechanisms associated with Parkinson's disease (Matsui, Kenmochi et al. 2019).

Current genomic research in killifish analyses pathways and genes associated with senescence, making them a valuable tool for the biomedical sciences (Priami, De Michele et al. 2015, Liu, Zheng et al. 2018). In addition to their embryos being desiccation resistant they are also capable of going through multiple stages of diapause, stalling full development until conditions are suitable (Cellerino, Valenzano et al. 2016). Exposure to a multitude of stressors during early development has been linked with a number of diseases (Heindel and Vandenberg 2015). Studying the effects on annual fish which have undergone stressful development may inform the mechanisms of these diseases and how they may be treated or prevented (Cellerino, Valenzano et al. 2016, Philippe, Hautekiet et al. 2018).

1.2.4 Cavefish; mechanisms in adaptation to the dark

Water filled caves are abundant in many parts of the world but are difficult places to live due to often being completely devoid of light, having no primary producers in the water, low levels of dissolved oxygen and a reduced amount and variety of food resources (Rétaux and Casane 2013). Of vertebrate species only salamanders and teleost fish have successfully colonised this type of aquatic habitat (Soares and Niemiller 2013). Currently there are more than 160 described species of cavefish (Riesch, Tobler et al. 2015), which are phylogenetically diverse but share a number of adaptations to their subterranean life (Soares and Niemiller 2013). One species, *Astyanax mexicanus* (De Filippi 1853) has become the model in cavefish research. It has been successfully acclimatised to

laboratory conditions, has a large number of clear embryos and a relatively short generation time of about four to six months (Jeffery 2019). The species has both a surface dwelling form and a number of cave dwelling forms meaning comparisons can be carried out much in the same way as mutant lines are compared to wild-type in other model organisms (Jeffery 2019).

Living in perpetual darkness limits the use of eyesight and the requirements of visual signals in determining whether another individual is both a conspecific and of the opposite sex. As such, cavefish have lost their melanin pigmentation and eye degeneration has caused them to have either reduced or absent eye phenotypes (Borowsky 2008, Borowsky 2018). The mechanisms by which these characteristics were acquired has been studied extensively. Melanin pigmentation plays important roles in protection from ultraviolet light, vision and immunity (Dräger and Balkema 1987, Riley 1997). Surface living species which are depigmented or albino likely show reduced fitness, but in a subterranean environment the selective pressure maintaining pigmentation is relaxed (Riesch, Tobler et al. 2015). A number of genes involved in melanin production, such as *oculocutaneous albinism-2* (*Oca2)* and *melanocortin-1 receptor* (*Mc1r),* are known to have accumulated loss of function mutations in different cavefish species*. Oca2* is the gene encoding a melanosome transmembrane protein which internalises L-tyrosine for conversion to melanin and is the same gene responsible for albinism in humans (King, Pietsch et al. 2003, Jeffery 2009). Mutations resulting in a truncated, loss of function protein, are present in cavefish *Oca2*, meaning melanin synthesis is disrupted at an early stage in the pathway (Protas, Hersey et al. 2006, Klaassen, Wang et al. 2018). *Mc1r* encodes a receptor for the ligand MSHα, upon binding of this ligand downstream target genes are activated which control pigmentation (Jeffery 2009). Mutations in either the

regulatory regions of the gene or those which have produced early stop codons have left this gene non-functional in different populations of *A. mexicanus* (Gross, Borowsky et al. 2009, Stahl and Gross 2015). This has resulted in fewer melanophores and reduced pigmentation, or brown phenotype, in contrast to the *oca2* mutants which completely lack pigment and have a white phenotype (Stahl and Gross 2015). Changes in *Mc1r* are hypothesised to have occurred due to genetic drift over thousands of years (Riesch, Tobler et al. 2015), however the mutations in *oca2* are shared across many populations of *A. mexicanus* as well as multiple other cavefish species suggesting the non-functional gene is under positive selection (Felice, Visconti et al. 2008, Gross and Wilkens 2013). By interrupting the melanin synthesis pathway at an early stage, through preventing the uptake of L-tyrosine, the amino acid is then in surplus. L-tyrosine is also a primary substrate in the catecholamine synthesis pathway (Bilandžija, Ma et al. 2013) and it is hypothesised that this form of albinism in cavefish was selected for because it allowed an increase in catecholamine synthesis. Some studies have linked this to behavioural and physiological adaptations which encourage increased foraging and reduced sleep in cavefish (Duboué, Keene et al. 2011, Elipot, Hinaux et al. 2014).

Cavefish populations of *A. mexicanus* are notable for their reduced or absent eye phenotypes. Eyes begin to form during early development but subsequently stop and degenerate through apoptosis (Jeffery, Strickler et al. 2003). This is in response to the expansion of *sonic hedgehog (shh)* expression in the forebrain (Pottin, Hinaux et al. 2011). It is possible that eye loss has been selected for in cavefish as they are a costly structure to maintain and so in an environment where they are not needed it is better to remove this cost (Riesch, Tobler et al. 2015). Alternatively, eye loss could be an indirect pleiotropic consequence of selection for other structures. *Shh* overexpression is also

associated with an increased number of taste buds and a larger jaw size, traits which are beneficial to living in a darkened environment where food is hard to find (Yamamoto, Byerly et al. 2009, Gross, Meyer et al. 2015). The extremophile cavefish species have proved themselves invaluable in the study of evolution and how gene-environment interactions result in morphological and behavioural adaptation, as discussed, and have become an important tool in multiple biological fields, including; ecology, development and neuroscience (Torres-Paz, Hyacinthe et al. 2018).

In addition to the benefits of studying extremophile fish species discussed, research into how extremophiles can survive under some of the harshest conditions on Earth, overcoming the associated physiological and cellular stressors, may provide medically relevant knowledge for human application in conditions induced by chronic stresses on the body.

1.2.5 Adaptation to temperature

Few animals are capable of surviving below freezing temperatures and it is very rare for any to live at temperatures above 50℃ (McCue 2004). Temperatures below zero may result in a fatally depressed metabolism or the production of ice in the body. High temperatures, on the other hand, can cause the denaturation of enzymes and proteins vital for life plus cause beyond lethal dehydration and may push metabolic processes to their limit (DeVries 1971, Somero 2010). Aquatic ectotherms are particularly susceptible to environmental temperature because they rely completely on external water to regulate their body temperature (Beitinger, Bennett et al. 2000). However, there are a number of fish species that survive below zero temperatures, such as those in the suborder Notothenioidei which are largely found in Antarctic waters (Bargelloni, Ritchie

et al. 1994). In contrast, few fish are able to survive extreme high temperatures with *Alcolapia* species being some of the best described (see section 1.3), however there are species which have adapted to the hot permanent water bodies found in some desert environments.

Fishes that live beneath the ice sheets in the earths Polar Regions are exposed to below freezing water. In other fish species this would be fatal with the cold temperatures causing ice crystals to form within the body. Species adapted to these conditions have convergently evolved a mechanism of preventing their internal structures succumbing to ice formation by using a number of different evolved proteins as antifreeze (Nath, Chaube et al. 2013). A large percentage of the freezing point depression seen in the blood of polar fishes is due to the expression of a variety of these antifreeze glycoproteins (AFGP) (Eastman and DeVries 1986). The convergent molecular evolution of multiple classes of AFGP have allowed phylogenetically and geographically distinct fish populations to survive freezing polar temperatures (Harding, Anderberg et al. 2003). AFGPs have a characteristic motif of repeating Ala-Ala-Thr units which are bound to disaccharides (with some natural variation) and work to lower the freezing temperature of body fluids.

Notothenoids AFGPs likely evolved from a pancreatic trypsinogen gene. The 5' and 3' regions of this ancestral gene were retained and *de novo* amplification of a nine nucleotide region coding for Ala-Ala-Thr joined these two regions and produced the functionally active new gene encoding an AFGP (Chen, DeVries et al. 1997). This constitutes an example of how an old gene may evolve into a new one, with a different function, under pressure imposed by an environmental stressor, in this case freezing temperatures. The AFGPs are produced in the pancreas and released into the rest of the

body. Here, they are capable of binding to the hexagonal ring of ice crystals, which prevents their growth. This ensures that no internal damage is caused by the ice and it can safely be removed from the body (Cheng, Cziko et al. 2006, Riesch, Tobler et al. 2015). Alternatively, the evolution of AFGPs in non-Antarctic fish is hypothesised to have occurred via changes to non-coding DNA (Baalsrud, Tørresen et al. 2018). It is thought that a nine nucleotide repeat encoding for Ala-Ala-Thr was present in the ancestor of these fish, a mutation creating a start codon and open reading frame would have subsequently allowed the region to be both transcribed and translated, producing the functional AFGP genes (Zhuang, Yang et al. 2019). *De novo* genes are more likely to occur in regions with a high GC content as they are more transcriptionally active and less likely to contain a stop codon (McLysaght and Hurst 2016). Indeed, codons for alanine are GC rich which would have increased the likelihood of *de novo* evolution occurring.

Aquatic species are considerably more susceptible to changes in the temperature of their environment than terrestrial species, which accounts for the faster recorded decline in fish populations due to increases in response to climate change (Pinsky, Eikeset et al. 2019). Indeed, few described fish species experience chronically high temperatures with the effects of exposure to heightened environmental temperature being more commonly studied in a number of non-adaptive species, such as exposing zebrafish to increasing temperatures (Schaefer and Ryan 2006). Extremophile pupfish are found in water bodies in desert environments of Central America, here they are regularly exposed to a variety of high temperatures, often exceeding 38℃, and have exhibited some of the highest critical thermal limits of fish species (Lowe and Heath 1969, Brown and Feldmeth 1971). Experimentally exceeding this limit has resulted in reduction of swimming performance, cessation of breathing and eventual death, showing the importance of a species fixed

temperature range for its continued survival (Riesch, Tobler et al. 2015). With increasing temperatures the rate of biological functions is also increased, pushing energy consumption above the speed at which it can be created. The metabolism of an individual living above its thermal tolerance would be heightened to a point where it eventually would be unable to sustain normal processes. An additive problem of temperature for fish is that as its metabolic demand increases it is usually unable to considerably increase oxygen intake for energy production because oxygen solubility is less in warmer waters (Tromans 2000).

1.2.6 Adaptation to pH

The extremes of pH produce environments which are either highly acidic or highly alkaline, these conditions pose problems for fish because most species only survive in neutral waters. In addition to being potentially caustic, the chemical compositions of these waters have implications for ionoregulation and acid-base regulation (Gilmour and Perry 2009). Excess H^+ causes a lowering of a solutions pH, whereas high pH is caused by increased concentration of carbonates (such as HCO_3^-). As well as being naturally occurring in the environment they are also waste products of metabolism and hence are present in the body (Reece, Urry et al. 2014). H^+ and HCO_3^- are known to destabilise biological molecules at high concentration, therefore fish usually excrete these across the gills in exchange for other charged particles (Riesch, Tobler et al. 2015). Problems arise when the external pH becomes acidic or alkaline as it has effects on the gills ability to excrete and absorb different products due to changes in diffusion gradients (Wright and Wood 1985, Wilson, Wood et al. 1999). Subsequently, these external conditions can disrupt internal pH resulting in acidosis or alkalosis. This is known to disrupt a number of

biological functions, including; brain activity, immune system, oxygen delivery and usual cellular function, if allowed to persist any of these may result in the death of the affected individual (Wilkie and Wood 1996, Kwong, Kumai et al. 2014).

Lake Usori (also known as Lake Osorezan) is located on the Japanese island of Honshū and is notable for its strongly acidic waters. The presence of naturally occurring sulphuric and hydrochloric acid have created waters whose pH ranges from 3.4–3.8 (Satake, Oyagi et al. 1995). Few animals can survive there and in fact only one type of fish, a subspecies of Japanese dace (*Tribolodon hakonensis*; Günther 1877), is found in its waters (Masiko 1940). When exposed to acidic conditions fish usually have a marked increase in the expression of H^+-ATPases in the gills which work to actively pump H^+ out of the body and ensure internal pH is maintained (McDonald 2003). Japanese dace use an alternative mechanism to deal with long term exposure to the conditions of Lake Usori. To also allow maintenance of internal Na^+ levels, which decrease in acidic conditions, these fish have a stronger reliance on Na^+/H^+ -exchangers (Hirata, Kaneko et al. 2003). These transporters are located in the mitochondrial rich cells (MRC, previously called chloride cells) of the fish gill. The structure of these cells also differ morphologically to aid in acid-base regulation in Japanese dace. Their MRCs group together and form a follicular structure which increases their apical surface area without imposing limitations on the respiratory capacity of the rest of the gills cell types (Kaneko, Hasegawa et al. 1999). With the amplified surface area of MRCs and the increase in expression of notable transporters, the gill of the Japanese Dace found in the acidic waters of Lake Usori plays a significant role in acid-base regulation and hence acid tolerance.

Alkaline lakes (or soda lakes) are far more common than acidic waters, with high pH environments occurring all over the world. A number of diverse fish species have adapted to these conditions, utilising a variety of mechanisms to survive (Brauner, Gonzalez et al. 2012). The Amur ide (*Leuciscus waleckii;* Dybowski 1869) is a widely dispersed teleost in Northern Asia with one population inhabiting the alkaline lake of Dali Nur (pH 9.6). The availability of the Amur ide genome and the presence of multiple populations in extreme and neutral environments has allowed genomic and transcriptomic comparisons to be made. These have highlighted genes which are differentially expressed between the populations and those which are under positive selection and may be aiding survival under the alkaline conditions of Dali Nur (Xu, Li et al. 2013, Xu, Li et al. 2017). For example, *carbonic anhydrase 4 (CA 4)* was shown to be both under strong positive selection when comparing fish from alkaline and neutral waters as well as having differential expression (Xu, Ji et al. 2013, Xu, Li et al. 2013). CA 4 catalyses the hydration of CO_2 to H^+ and HCO_3^-, these are then associated with ion transport and acid base regulation hence will play an important function in maintaining homeostasis in a high pH environment (Yang, Alvarez et al. 2005). Furthermore, the Amur ides genome has 30 genes encoding for zona pellucida domain-containing proteins (ZP proteins), this is three times the amount found in other teleost fishes, and eight paralogues of vitelline membrane outer layer protein 1 (VMO1) compared to a usual one or two (Xu, Li et al. 2017). These proteins are both found in the outer membrane of developing eggs and, especially ZP proteins, are known to be crucial structural protein which are vital in protecting the egg from environmental stressors (Monné, Han et al. 2006, Chen, Cheng et al. 2008). It is hypothesised that the expansion of these protein types plays an important

role is the survival and development of the Amur ides eggs in an alkaline environment (Xu, Li et al. 2017).

1.2.7 Adaptation to salinity

Hypersaline environments pose difficult challenges for osmoregulation in fish species, especially those usually adapted to freshwater environments. Examples of hypersaline waters include inland salt lakes, which usually have no outlet to the sea (these also tend to be alkaline), and intertidal pools which are separated from the sea during the daily changes in tide (Gonzalez 2012). Osmoregulation in fish is a balance between drinking water, which is filtered through the intestines and kidneys, and the excretion and absorption of salts and water across gill tissue (Reece, Urry et al. 2014). Adaptations to high salt environments have evolved as developments of characteristics present in seawater species, as well as the evolution of novel traits that allow for hypersaline tolerance (Gonzalez 2012, Laverty and Skadhauge 2012). Without appropriate adaptations to allow survival in these conditions, fish would likely go through osmotic stress caused by a change in the solute concentration around tissues and cells. Increased salt would create osmotic potential and water would be drawn from cells via osmosis, putting the cells into a state of 'shock,' preventing usual function and eventually leading to cell death (Fiol and Kültz 2007, Reece, Urry et al. 2014).

To deal with the increase of salt in the body, which occurs in high salt environments, fish will ingest larger amounts of water in an attempt to dilute this (Laverty and Skadhauge 2012). However, to absorb water against an osmotic gradient, from the gut into the body, it is taken up by a variety of co-transporters such as Na^+/K^+-ATPase and Na^+-Cl^-, further increasing internal salt concentration (Grosell 2006, Gonzalez 2012). Therefore, salt

uptake is the limiting factor in how much water can be absorbed, to maximise this, fish exposed to high salt show an increased expression of intestinal transporters (Seidelin, Madsen et al. 2000, Laverty and Skadhauge 2012). In this way, net water absorption can occur but leads to larger requirements of salt excretion. The gills subsequently increase their excretion capacity to remove this extra salt while decreasing water permeability to prevent water loss and salt uptake, negating the cost of drinking and allowing maintenance of internal osmolality (Riesch, Tobler et al. 2015). Reduced permeability is produced by the downregulation of members of the aquaporin protein family in gill tissue (Deane and Woo 2006, Tse, Au et al. 2006), whereas increased salt excretion is often attributed to increased number or size of MRC which have excretory specific salt transporters (Kültz, Jürss et al. 1995, Gonzalez 2012).

1.2.8 Adaptation to fluctuating oxygen levels

In general, fish take in oxygen across the gills where it diffuses into the blood and subsequently is transported to the tissues (Black 1940, Hughes 1984). Here, it is used for a number of metabolic reactions, and importantly, acts as the final electron acceptor in cellular respiration at the mitochondria, during the production of ATP or energy (Senior 1988). All cellular processes in the body ultimately rely in on the utilisation of oxygen for metabolism and so most animals are incapable of surviving in environments where the oxygen levels do not meet their metabolic requirements (Brahimi-Horn and Pouysségur 2007). However, numerous fish species have colonised hypoxic environments by either having an increased ability to uptake oxygen from the water, reducing their metabolic demands and/or selecting for higher dependence on anaerobic metabolism (Childress and Seibel 1998).

During the winter months in Northern Europe many smaller water bodies, including lakes and ponds, become hypoxic and eventually anoxic due to thick ice coverage. The only fish adapted to survive in such waters is the crucian carp (*Carassius carassius;* Linnaeus, 1758, Nilsson and Renshaw 2004). To overcome the low oxygen conditions they have evolved mechanisms to increase oxygen uptake and delivery to tissues when it is scarce. As oxygen levels begin to decrease they remodel their gills by increasing the number of lamellae, which increases the surface area while also reducing diffusion distance (Vornanen, Stecyk et al. 2009). Additionally, its haemoglobin has a very high oxygen affinity allowing it to maintain delivery to its tissues down to very low dissolved oxygen levels in the surrounding water (Sollid, De Angelis et al. 2003).

The crucian carp can also survive for extended periods of time using anaerobic respiration with a very interesting adaptation allowing continued production of energy under anoxic conditions. It produces ethanol as the end product of glycolysis during anaerobic respiration instead of lactate (Johnston and Bernard 1983, Dhillon, Mandic et al. 2018). This has two benefits, firstly it negates the chances of metabolic acidosis as can occur with a build-up of lactate, and secondly ethanol is able to freely diffuse across the gills avoiding any accumulation in the tissues (van Waarde 1991). Pyruvate dehydrogenase complex (PDHc) is made up of three subunits (Enzyme 1, Enzyme 2 and Enzyme 3) and in the presence of oxygen this complex catalyses the conversion of pyruvate at the end of glycolysis to acetyl-CoA which channels into the citric acid cycle. In the absence of oxygen this reaction is not possible (Patel, Nemeria et al. 2014). Members of the genus *Carassius* have a duplicated gene encoding Enzyme 1 of PDHc which has evolved in function allowing the conversion of pyruvate to acetaldehyde during anoxia (Fagernes, Stensløkken et al. 2017). This is then converted to ethanol by alcohol dehydrogenase in

the muscle and subsequently excreted, another adaptation as alcohol dehydrogenase is usually predominantly expressed in hepatic tissues in fish (Nilsson 1988). This process is relatively inefficient and the limiting factor of prolonged anoxia is total available glycogen (Nilsson 1990). In order to maximise survival, the crucian carp prepares itself for winter by increasing its stored glycogen in its large liver, this can be broken down into glucose and transported in the blood to required tissues and allows the prolonged survival of these fish to anoxic conditions (Hyvärinen, Holopainen et al. 1985).

Coral reefs and intertidal zones experience dramatic daily fluctuations and may become isolated during the changing tides. The decreased volume of water or prolonged air exposure reduces the amount of available oxygen to the fish species which inhabit these areas (Mandic and Regan 2018). Unlike the crucian carp, these tidal species are also exposed to high temperature meaning a naturally increased metabolism and potentially higher oxygen demands. The epaulette shark (*Hemiscyllium ocellatum*; Bonnaterre 1788) lives on shallow coral reefs and is known to withstand extended periods of exposure to hypoxic conditions (Wise, Mulvey et al. 1998). In response to low oxygen, adenosine produced from the metabolism of ATP is not recycled. The increased concentration of this in the bloodstream of the sharks causes arterial and venous vasodilation, promoting the circulation of blood to the gills, heart and brain, increasing the delivery of oxygen during exposure to hypoxia (Stensløkken, Sundin et al. 2004). The metabolic requirements of the brain would not be supported by anaerobic respiration and prolonged exposure to anoxic conditions is known to cause brain damage (Vannucci 1990, Yu and Li 2011). In order to further protect the brain from the negative consequences of low oxygen the epaulette shark selectively reduces the metabolism of different regions of the brain that control motor output while maintaining normal levels in sensory regions. This minimises ATP

demand while continuing to support neural function (Mulvey and Renshaw 2000). The combined adaptations of increased oxygen supply to the brain and specifically restricted metabolic depression ensures that oxidative phosphorylation can continue for longer in this vital organ.

In contrast to adaptation to hypoxic and anoxic environments, studies on hyperoxic adaptations are limited. However, it is known that hyperoxia can occur in aquatic systems and may result in an increased production of reactive oxygen species (ROS) (Lushchak and Bagnyukova 2006). If allowed to accumulate, ROS causes oxidative stress and may damage lipids, protein and DNA (Lushchak 2014). The giant goby (*Gobius cobitis*; Pallas 1811) lives in intertidal zones and experiences daily fluctuations in oxygen concentration. When exposed to hyperoxic waters the fish reduce their ventilation rate but maintain heartrate in an attempt to keep oxygen delivery to the tissues at a constant level (Berschick, Bridges et al. 1987). *Hoplerythrinus unitaeniatus* (Spix and Agassiz 1829) is adapted to air breath under certain environmental conditions and as such is often exposed to the higher partial pressures of oxygen present in air. To protect itself from potential damage from ROS it has increased concentration of small antioxidants and the activity of enzymes involved in ROS degradation (Pelster, Giacomin et al. 2016). This is a defence utilised by other fish species also exposed to increased levels of ROS produced by changes to the oxygen concentration of their environments (Lushchak and Bagnyukova 2006).

1.3 Study system

The *Alcolapia* species flock represent a recent, small scale radiation of African cichlids, endemic to the extreme environment of Lakes Natron and Magadi. This lake system is

believed to have gained many of its extreme conditions as recently as 10 thousand years ago (KYA). As such, the *Alcolapia* present themselves as a naturally occurring experiment in the study of adaptation and how animals evolve to survive extreme conditions.

1.3.1 Lake Natron – Magadi basin

Previously connected as part of the paleo-lake Orolonga, the East African soda lakes of Natron and Magadi are considered some of the most extreme environments supporting fish life, with water temperatures of 30-42.8 °C, pH 9-11.5, fluctuating dissolved oxygen levels and high salt concentrations of >20ppt (Ford 2015, Ford, Dasmahapatra et al. 2015). Located across the border of Tanzania and Kenya, the paleo-lake is estimated to have formed ~700 KYA and reached maximal depths of 50-60m (Roberts, Taieb et al. 1993). Lake Natron and Magadi are thought to have reached their maximal levels 10-12 KYA before water levels receded and the lakes were permanently separated. At this time the pH of the lake is estimated to have been ~9.4 with more freshwater conditions than today (Jones, Eugster et al. 1977, Roberts, Taieb et al. 1993). The hyper-salinity and increase in alkalinity of the lakes developed as water levels continued to contract ~9KYA (Butzer, Isaac et al. 1972, Kaufman, Margaritz et al. 1990, Roberts, Taieb et al. 1993). This reduction of the Orolongo basin water level is believed to have coincided with a period of arid climate change and the equatorial extension of the Younger Dryas event (Williamson, Taieb et al. 1993). A negative hydrological balance, with evaporation exceeding rainfall, has produced high levels of evaporites and a concentration of alkaline conditions (Vincens and Casanova 1987). Presently Lakes Natron and Magadi cover an area of 81 - 804km^2 and 75 - 108km^2 respectively, with lake size changing from dry to wet season (Jones, Eugster et al. 1977, Vanden Bossche and Bernacsek 1990, Tebbs, Remedios et al. 2013).

Fossil remains found 12m above the water level of Lake Magadi indicate that the paleo-lake was inhabited by cichlids. The fossils were initially described as *Tilapia (=Oreochromis) nilotica* by H Copley, however later examination by P Whitehead indicated that they were likely *Tilapia (=Alcolapia) grahami* (Whitehead via personal communication in Coe 1966). They appear morphologically similar to the *Alcolapia* but larger in size at 8inches (~200mm) compared to the maximal length of 100mm of modern specimens (Coe 1966, Seegers and Tichy 1999, Ford 2015). The larger size of these fossils is suggestive of these fish inhabiting freshwater conditions; this is further supported by ^{14}C dating of the fossils to ~9KY old, a time when the lake is considered to have been much less hostile than present (Coe 1966, Butzer, Isaac et al. 1972). Although there has been no extensive survey of Lake Natron for fossil remains, a number of fish specimens were found during a field expedition in July 2017 as part of this PhD which are yet to be analysed.

The majority of the African Great lakes tend to be slightly alkaline (pH 7- 9), however, Lakes Natron and Magadi are exceptional in having highly saline (>20ppt) and alkaline (pH ~10.5) conditions (Salzburger, Van Bocxlaer et al. 2014). This is due to the area in which the lakes sit being volcanic, with alkaline hydrothermal springs feeding the lakes with salts and precipitates (Williamson, Taieb et al. 1993). The large amounts of salt produces a thick layer of crystalline trona (sodium sesquicarbonate precipitate) which covers much of the lake; such that the only areas of permanent open water are lagoons at the lakes shore (Eugster 1986, Kaufman, Margaritz et al. 1990). This enclosure to shallow lagoons promotes high levels of avian predation on the fish that inhabit this lake system. The hydrothermal vents also contribute to the high temperature of the springs as well as producing a rich environment for algae and cyanobacteria (Mikhodyuk, Zavarzin et al.

2008). These organisms give the lakes their characteristic red colouration. Together, the high temperatures, algal and cyanobacterial bloom and trona crust cause large fluctuations in dissolved oxygen levels in the lakes (Brock 1970, Grant 2006). The accumulative effects of the extreme conditions in the Lake Natron-Magadi basin have produced an environment that supports no vascular plants and the only fish to survive its waters are the *Alcolapia.* Indeed, it is considered one of the most hostile aquatic environments supporting fish life (Seegers and Tichy 1999, Mikhodyuk, Zavarzin et al. 2008).

1.3.2 *Alcolapia* species flock

The *Alcolapia* species flock is currently represented by four described species; *Alcolapia grahami* present in Lake Magadi, Kenya (Boulenger 1912) and *A. latilabris, A. ndalalani* and *A. alcalica* in Lake Natron, Tanzania (Seegers, Sonnenberg et al. 1999, Ford, Dasmahapatra et al. 2015). They are the only fish to survive the extreme conditions of the Natron-Magadi lake system (Seegers and Tichy 1999). The Lake Natron fish were initially documented as a single species, *Tilapia alcalicus,* with the Magadi species being discovered shortly afterwards, then named *T. grahami* (Hilgendorf 1905, Boulenger 1912). These were later regrouped into the genera *Oreochromis* and noted as subspecies within the subgenus *Alcolapia,* stylised as *Oreochromis Alcolapia alcalicus alcalicus* and *O. A. alcalicus grahami* (Trewavas 1982). Further analysis of Lakes Natron and Magadi specimens noted significant morphological diversity between individuals in Lake Natron as well as variability within populations of the Lake Magadi species. This lead to the re-description of *O. A. alcalicus* and *O. A. grahami* as separate species and is the first appearance of *O. A. ndalalani* and *O. A. latilabris* in the literature based on differences in

head morphology (Seegers and Tichy 1999). The naming of the *Alcolpaia* species flock was once more revised with the subgenus being promoted to genus due to genetic divergence of mtDNA of the *Alcolapia* from other members of the *Oreochromis* (Seegers, Sonnenberg et al. 1999). Most recently, a comprehensive analysis of the molecular phylogeny of the *Oreochromis* genus showed that the *Alcolapia* are nested within the *Oreochromis* and as such the group is paraphyletic (Ford, Bullen et al. 2019). Due to this, it has been suggested that *Alcolapia* be reverted to subgenus as was stylised by Seegers and Tichy (1999).

The four *Alcolapia* species have considerable differences in phenotype, including morphology of oral structures and variation in male breeding colours (Seegers and Tichy 1999). Geological evidence suggests that the species diverged as recently as 7-10 KYA, making the group a very recent example of adaptive radiation in a simplified, closed-system environment (Tichy and Seegers 1999, Ford, Dasmahapatra et al. 2015). Genomic analysis using Restriction site Associated DNA Sequencing (RAD-Seq) showed that the four species are genetically distinct, however there are high levels of interspecific gene flow between species in Lake Natron indicating that reproductive isolation is not yet complete (Ford, Dasmahapatra et al. 2015). It has been suggested that an ancestor of the *Alcolapia* colonised the paleo-lake Orolonga and adapted to the increasingly extreme environment. In the absence of competitors, intraspecific competition led to evolution of different trophic morphologies and the species we see today (Ford, Rüber et al. 2016). This would suggest that species differences today occurred after the adaptation to the extreme conditions of the lake.

1.4 Adaptations of *Alcolapia*

To overcome the extreme conditions of these lakes, the *Alcolapia* must have a series of physiological adaptations to cope with a vast array of stressors. It should be noted that other Oreochromine species are capable of surviving a range of pH levels (Lake Malawi, *Oreochromis spp* pH 7.2-8.6; Salzburger et al., 2014, Lake Manyara, *O. amphimelas* pH 9.8-10; Cohen et al., 1993), variable salinity tolerance (Suresh and Lin 1992) and high temperatures (Upper lethal maximum temperature: *O. niloticus* 39 °C; Baras, Jacobs et al. 2001, *O. mossambicus* 37 °C*;* Stauffer Jr 1986). However, none have been shown to survive the extremes of Lake Natron and Magadi or the variability survived by the *Alcolapia* (Reite, Maloiy et al. 1974). In fact, whereas *A. grahami* can survive acclimation to freshwater conditions, transfer of the closely related *O. niloticus* to Lake Magadi water proved fatal within 1h (Wright, Perry et al. 1990). The tolerance of *Alcolapia* to both extreme and freshwater conditions is likely provisioned by rapid evolution created by a high mutation rate thought to be promoted by the extremes of their environment (Seegers, Sonnenberg et al. 1999, Pörtner, Schulte et al. 2010).

To date, much of the research into the *Alcolapia* species flock has been carried out in *A. grahami* with a few studies on *A. alcalica,* as such many of the noted adaptations of *Alcolapia* to extreme conditions have only been confirmed in one of the four species.

1.4.1 Excretion of nitrogenous waste

The high pH of Lakes Natron and Magadi prevents the passive diffusion of ammonia across the gills of *Alcolapia* into the surrounding water (Wright and Wood 1985). To overcome the toxic build-up of ammonia they instead convert it to urea and are the only noted 100 % ureotelic fish species (Randall, Wood et al. 1989). Production of urea in

Alcolapia is via a fully functioning ornithine-urea cycle (OUC) (Randall, Wood et al. 1989). Enzymes within this cycle are normally isolated to liver tissue but due to the high demands of *Alcolapia* in converting large quantities of ammonia, enzyme activity has also been reported in gill and muscle tissue (Lindley, Scheiderer et al. 1999, Walsh, Grosell et al. 2001). Activity is reportedly greatest in the muscle, which is suggestive of this being the main site of urea production (Lindley, Scheiderer et al. 1999). In addition to this, *Alcolapia* also have other adaptations within this pathway promoting the rapid conversion and excretion of large amounts of ammonia. Glutamine synthetase catalyses the conversion of ammonia to glutamine and is usually restricted to liver tissue. Glutamine is then free to be converted to carbamoyl-phosphate by glutamine dependent carbamoyl-phosphate synthetase III (CPS III) (Anderson 2001, Saha, Datta et al. 2007). However, glutamine synthase is absent in the muscle of *Alcolapia* and instead CPS III is able to preferentially bind ammonia as its primary substrate. In *Alcolapia,* the maximal enzymatic rates of CPS III are greater when using ammonia as a substrate when compared to glutamine, as opposed to in other species where use of ammonia yields enzymatic rates of around 10 % to that of glutamine (Lindley, Scheiderer et al. 1999). The majority of the produced urea is excreted across the gills (80 %) with the remainder removed as urinary urea (Narahara, Maina et al. 1994). Excretion is via the specialised gill transporter Magadi tilapia urea transporter (mtUT), which increases the permeability to urea by 10 times that of passive diffusion through a lipid bilayer and five times greater than that seen in the gulf toadfish (*Opsanus beta;* Goode and Bean 1882)*,* another fish capable of excreting urea (Walsh, Grosell et al. 2001). Together these adaptations allow *Alcolapia* to detoxify and remove their nitrogenous waste in their high pH environment. Unlike facultative ureotelic species (like gulf toadfish), the adaptation of urea production

and excretion is considered fixed, with continued excretion of 100 % of nitrogenous waste as urea even when conditions were considered favourable, after acclimatisation to freshwater, suggesting ureotelism is obligatory in *Alcolapia* (Wood, Wilson et al. 2002).

1.4.2 Respiration in a low oxygen environment

Dissolved oxygen levels fluctuate diurnally in Lakes Natron and Magadi due to high temperatures and the presence of large amounts of algae and cyanobacteria (Brock 1970, Grant 2006). To overcome this, *Alcolapia* are shown to have a number of adaptations to help them respire in such conditions. Field and lab observations have shown surface breathing from *Alcolapia* in instances of hypoxia, with the inspiration of the better oxygenated surface water layer and air (Franklin, Johnston et al. 1995, Narahara, Bergman et al. 1996). Surface breathing has also been witnessed in conditions of normoxia and hyperoxia. In these cases, it has been suggested that *Alcolapia* are voluntarily choosing to surface breath due to the high levels of ROS in the water which act as an irritant and are thought to be the cause of masses in gill lamellae (Johannsson, Bergman et al. 2014). Together, these observations show *Alcolapia* are facultative air breathers, capable of air breathing in instances that are beneficial such as low oxygen or increased levels of potentially negative irritants (Narahara, Bergman et al. 1996, Johannsson, Bergman et al. 2014). To accommodate this, *Alcolapia* have a highly vascularised swim bladder, that is directly connected to their oesophagus, which is capable of gaseous exchange and is possibly being utilised as a primitive respiratory organ (Maina 2000). In addition to supplementary air breathing, the oxygen binding properties of haemoglobin are also adapted to the environmental conditions of Lake Natron and Magadi. *Alcolapia* haemoglobin has a high oxygen affinity, lacks a functional Root or Bohr

effect (low pH sensitivity), is less sensitive to changes in osmolality than other *Oreochromis* species and is highly thermostable (Lykkeboe, Johansen et al. 1975, Franklin, Crockford et al. 1994, Narahara, Bergman et al. 1996). The changes to haemoglobin function allows for maximal oxygen uptake and delivery to tissues in an environment which has a high pH and osmolality, and experiences large fluctuations in oxygen availability (Lykkeboe, Johansen et al. 1975).

Adaptations to allow efficient respiration are particularly vital in the extreme environment of Lake Natron and Magadi, where the *Alcolapia* are exposed to stressful conditions which promote a heightened metabolic rate. Oxygen consumption and metabolic performance recorded in *Alcolapia* is the highest reported in fish, comparable with that of a small mammal, despite living in waters sometimes void of oxygen (Wood, Brix et al. 2016). A large proportion of metabolic output is believed to be used for acid-base regulation, crucial in a high pH environment (Wood, Wilson et al. 2002). Increases in mitochondrial activity can be explained by high temperature, but the sustained heightened metabolic performance is in part due to the thin diffusion area of the gill increasing oxygen uptake. Morphological adaptations to the gills of *Alcolapia* that increase their ability for oxygen uptake also include increased filament length and number (Hughes 1995, Maina, Kisia et al. 1996).

1.4.3 Other adaptations of *Alcolapia*

The fish gill is a complex, multifunctional organ that is in direct contact with the external environment and plays important roles in respiration, excretion, osmoregulation and homeostasis (Laurent and Perry 1991, Hwang, Lee et al. 2011). Due to the extreme conditions they inhabit, the gills of the *Alcolapia* have morphological adaptations

promoting their survival. Some of these have been discussed previously such as the presence of a unique urea transporter (Walsh, Grosell et al. 2001) and the thin diffusion area allowing increased rate of oxygen uptake (Wood, Brix et al. 2016). Other adaptations to the gill include increased number of MRCs compared to *O. niloticus* which may aid in ion movement (Maina 1991). Gill adaptations in *Alcolapia* are hypothesised to be one of the most important for their continued survival in a hostile environment (Maina 1990).

A further adaptation to living in soda lake conditions is seen through alterations to gut morphology, whereby the oesophagus of the *Alcolapia* is trifurcated (Bergman, Laurent et al. 2003). Owing to this, when drinking, the stomach can be bypassed preventing any potentially negative consequences arising from the mixing of the acidic stomach content and alkaline water but allowing large quantities of water to be consumed for osmoregulation purposes (Wood, Wilson et al. 2002). Although complete exclusion of lake water from the stomach may be unavoidable while eating it has been shown that some may be shunted directly into the intestine. This may be through the presence of striated muscle in the pyloric sphincter and the contraction of stomach muscles which could indicate some voluntary control as to what enters the stomach (Bergman, Laurent et al. 2003). The buccal cavity of *Alcolapia* has evolved a joint secretory and protective role under the pressures of hostile waters. The presence of a large number of mucus secretory cells in the epithelium as well as a high concentration of leukocytes supports the protective nature of the buccal cavity which would be beneficial to the fish while feeding, drinking or surface skimming for respiration (Maina 2000).

1.5 My project

1.5.1 Aims and hypothesis

This thesis aimed to establish a breeding colony of *A. alcalica* at the University of York for laboratory experiments, including molecular embryology, to investigate conservation of developmental mechanisms and any notable adaptive changes in this unique fish which aid its survival to extreme conditions. Specifically, this thesis investigates adaptations that allow the excretion of nitrogenous waste, from a high metabolic fish in a very alkaline environment, by examining gene expression profiles and analysing protein function. The conservation of phylotypic gene expression of important developmental regulators, such as *Pax-6* and *MyoD,* are also explored in this non-model species.

To carry out this work this project first had to:

1. Establish husbandry for *A. alcalica,* including discerning best salt and pH conditions.
2. Produce protocols for breeding and obtaining early stage embryos from a mouth-brooding fish.

1.5.2 Field expedition

Fieldwork to Lake Natron, Tanzania, was conducted during June and July of 2017 to collect live specimens of the three endemic species in an attempt to produce stable breeding populations of these fishes in the UK. Live fish were all collected from a single spring containing all three species found in Lake Natron. Small individuals were preferentially selected and more females taken than males to ensure good numbers for maintaining breeding populations. Fish were packaged individually in breather bags which allow the movement of gas across the bag membrane, providing fresh oxygen into the

water and removing excess carbon dioxide. Fish were then shipped to the University of Bangor to increase numbers via breeding. A subset of *A. alcalica* were subsequently moved to the University of York to begin characterising their adaptations. *A. alcalica* were selected for this work due to having the largest brood size of the species collected from Lake Natron. Additional site exploration and environmental readings were also taken during the expedition. An aquatics habitat, stand alone, recirculating aquarium was adapted to house male and female *A. alcalica* in 10 or 30 L tanks at a constant temperature of 30℃, pH 9 and salt concentration of 3800 µS.

Chapter Two – Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish

2.1 Summary

It has been reported that in *Alcolapia* the carbamoyl-phosphate synthetase (CPS) III enzyme, the first enzyme in the ornithine urea cycle, has evolved in function and spatial expression. This chapter aimed to further explore this by examining the expression pattern of *CPS* in *A. alcalica* adult tissues using RT-PCR and in embryos using *in situ* hybridisation methods. These experiments confirmed muscle specific expression of *CPS III*, rather than the characteristic liver expression for this enzyme. This chapter also showed that the *Alcolapia* genome has gained a pair of E-boxes in the promoter region of *CPS III*. Paired E-boxes are a known binding target of MyoD1 which could initiate expression in muscle tissue. Finally, by comparing the coding regions of *CPS III* from teleost species and *CPS I* from terrestrial vertebrates this chapter was able to describe mutations which are likely responsible for changes in CPS function that show convergent evolution between *Alcolapia* CPS III and the terrestrial vertebrate CPS I.

2.2 Introduction

In living organisms, protein metabolism results in the production of nitrogenous wastes which need to be excreted. Most teleosts are ammonotelic, excreting their toxic nitrogenous waste as ammonia across gill tissue by diffusion. As an adaptation to living on land, amphibians and mammals are ureotelic, using liver and kidney tissues to convert waste ammonia into the less toxic and more water soluble urea, which is then excreted in urine. Other terrestrial animals such as insects, birds and reptiles are uricolotic and convert nitrogenous waste into uric acid, which is eliminated as a paste; a process which requires more energy but wastes less water (Wright 1995).

While most adult fish are ammonotelic, the larval stages of some teleosts excrete nitrogenous waste as both ammonia and urea before their gills are fully developed (Zimmer, Wright et al. 2017). Additionally, some adult fish species such as the gulf toad fish (*Opsanus beta;* Barimo, Steele et al. 2004) and the African catfish (*Clarias gariepinus;* Terjesen, Chadwick et al. 2001) also excrete a proportion of their nitrogenous waste as urea. This is usually in response to changes in aquatic conditions, such as high alkalinity. It has been shown experimentally that high external pH prevents diffusion of ammonia across gill tissue (Wright and Wood 1985, Handy and Poxton 1993). Unusually, the cichlid fish species in the subgenus *Alcolapia* (described by some authors as a genus but shown to nest within the genus *Oreochromis*; (Ford, Bullen et al. 2019) inhabit the highly alkaline soda lakes of Natron (Tanzania) and Magadi (Kenya), are reported to be 100 % ureotelic (Randall, Wood et al. 1989, Wilson, Wood et al. 2004).

Once part of a single paleo-lake, Orolonga (Roberts, Taieb et al. 1993), Lakes Natron and Magadi are one of the most extreme environments supporting fish life, with water temperatures up to 42.8 °C, pH ~10.5, fluctuating dissolved oxygen levels, and salt concentrations above 20 parts per thousand (Ford, Dasmahapatra et al. 2015). *Alcolapia* is the only group of fish to survive in these lakes, forming a recent adaptive radiation including the four species: *Alcolapia grahami* (Lake Magadi) and *A. latilabris, A. ndlalani and A. alcalica* (Lake Natron) (Seegers and Tichy 1999, Ford, Dasmahapatra et al. 2015). The harsh environment of the soda lakes presents certain physiological challenges that *Alcolapia* have evolved to overcome, including the basic need to excrete nitrogenous waste. While other species are able to excrete urea in response to extreme conditions, none do so to the level of *Alcolapia* (Walsh, Danulat et al. 1990, Wright, Iwama et al. 1993), and unlike facultative ureotelic species, the adaptation of urea production and

excretion in *Alcolapia* is considered fixed (Wood, Wilson et al. 2002). Moreover, the heightened metabolic rate in *Alcolapia*, a by-product from living in such an extreme environment (Randall, Wood et al. 1989, Kavembe, Franchini et al. 2015), requires an efficient method of detoxification.

Alcolapia and ureotelic tetrapods (including humans) detoxify ammonia using the ornithine urea cycle (OUC) where the mitochondrial enzyme carbamoyl-phosphate synthetase (CPS) is essential for the first and rate limiting step of urea production (Rubio and Cervera 1995). This enzyme, together with the accessory enzyme glutamine synthase, provide an important switch regulating the balance between ammonia removal for detoxification and maintaining a source of ammonia for the biosynthesis of amino acids (de Cima, Polo et al. 2015). CPS has evolved into two biochemically distinct proteins: in terrestrial vertebrates CPS I uses ammonia as its preferential nitrogen donor, while in teleosts CPS III accepts glutamine to produce urea during larval stages (reviewed Zimmer et al 2017). While CPS I / III are mitochondrial enzymes and part of the urea cycle, CPS II is present in the cytosol catalyzing the synthesis of carbamoyl phosphate for pyrimidine nucleotide biosynthesis. *CPS I / III* are syntenic, representing orthologous genes; their somewhat confusing nomenclature is based on the distinct biochemical properties of their proteins. *CPS I /III* genes from different vertebrate species clade together, separate from *CPS II*. For simplicity, we will continue to refer to fish, glutamine binding CPS as CPS III and tetrapod, ammonia binding CPS as CPS I. The teleost CPS III binds glutamine in the glutamine amidotransferase (GAT) domain using two amino acid residues (Hong, Salo et al. 1994), subsequently the nitrogen source provided by the amide group is catalysed by a conserved catalytic triad; Cys-His-Glu (Loong, Chng et al. 2012). In terrestrial vertebrates CPS I lacks a complete catayltic triad and can only generate carbamoyl-phosphate in the

presence of free ammonia (Saeed-Kothe and Powers-Lee 2002). This change in function from glutamine binding CPS III to ammonia binding CPS I is believed to have evolved in the stem lineage of living tetrapods, first appearing in ancestral amphibians (Saeed-Kothe and Powers-Lee 2002).

In tetrapods and most fish, the OUC enzymes are largely localised to the liver (Ip and Chew 2010), the main urogenic organ (Mommsen and Walsh 1989). *Alcolapia* are different, and the primary site for urea production in these extremophile fishes is the skeletal muscle (Lindley, Scheiderer et al. 1999). Notably, glutamine synthase activity is reportedly absent in *Alcolapia* muscle tissue. The kinetic properties of CPS III in *Alcolapia* therefore differ from that of other teleosts in that it preferentially uses ammonia as its primary substrate, having maximal enzymatic rates above that of binding glutamine (although it is still capable of doing so) as opposed to in other species where use of ammonia yields enzymatic rates of around 10 % to that of glutamine (Lindley, Scheiderer et al. 1999). These rates are similar to ureotelic terrestrial species, where CPS I preferentially binds ammonia and is incapable of using glutamine (Loong, Chng et al. 2012).

2.2.1 Aims of this chapter

A major sequence change in the catalytic region of the glutamine amidotransferase (GAT) domain was reported in *A. grahami* (Lindley, Scheiderer et al. 1999), the validity of this significant difference was queried by Loong et al (2012). As such one aim of this chapter was to sequence the coding regions of *CPS III* in both *A. alcalica* and *A. grahami.* A second was to compare the amino acid sequences of *Alcolapia* CPS to other teleost and terrestrial ureotelic species to extrapolate whether the change in function reported in

Alcolapia could be explained by sequence changes in the substrate binding domain. Finally, this chapter aimed to determine whether the skeletal muscle expression of *CPS III* in *Alcolapia* originates in early development. This was done by using *in situ* hybridisation methods and modelling the mechanism driving this change in gene expression by looking for changes in the *CPS III* promoter region.

2.3 Methods

2.3.1 Expression of *CPS III* in adult tissues

Reverse transcription polymerase chain reaction (RT-PCR) was used to determine the presence of *CPS III* in different tissues (gill, muscle, liver, brain) of three different adult *A. alcalica*. RNA was extracted from dissected tissues with TriReagent (Sigma-Aldrich) to the manufacturers' guidelines. For cDNA synthesis, 1µg of total RNA was reverse transcribed with random hexamers (Thermo Scientific) and superscript IV (Invitrogen). PCR was performed on 2µl of the above cDNA with 10 µl of 2x Promega PCR master mix and 0.5mM of each primer (Forward: CAGTGGGAGGTCAGATTGC, Reverse: CTCACAGCGAAGCACAGGG) in a final volume of 20 µl (PCR conditions; 96°C for 2 mins; 30 cycles of 96°C for 30 s, 56°C for 30 s, 72°C for 30 s [extension time otherwise adjusted as 1 min per kb amplified]; 72°C for 10 mins). Gel electrophoresis of the PCR products determined the presence or absence of *CPS III* RNA.

2.3.2 *In situ* hybridisation for *CPS III*

For the production of antisense probes, complementary to the mRNA of *CPS III* to use in *in situ* hybridisation, the above 399bp PCR product was ligated into PGem-tEasy and transformed into the *E.coli* strain DH5α. This was linearised and in vitro run off transcription was used to incorporate a DIG labelled UTP analogue. To determine the

temporal expression of these proteins in *A. alcalica*, embryos were collected at different stages of development (2, 4 and 7 days post fertilisation [between 15 and 20 for each stage]), fixed for 1 hour in MEMFA (0.1 M MOPS pH 7.4, 2 mM EGTA, 1 mM $MgSO_4$, 3.7% formaldehyde) at room temperature and stored at -20°C in 100% methanol. For in situ hybridisation, embryos were rehydrated and treated with 10 µg/ml proteinase K at room temperature. After post fixation and a 2 hour pre-hybridisation, embryos were hybridised with the probe at 68°C in in hybridisation buffer (50% formamide (Ambion), 1mg/ml total yeast RNA, 5×SSC, 100µg/ml heparin, 1× denharts, 0.1% Tween-20, 0.1% CHAPS, 10mM EDTA. Embryos were extensively washed at 68°C in 2×SSC +0.1% Tween-20, 0.2×SSC +0.1% Tween-20 and maleic acid buffer (MAB; 100mM maleic acid, 150mM NaCl, 0.1% Tween-20, pH7.8). This was replaced with pre-incubation buffer (4× MAB, 10% BMB, 20% heat-treated lamb serum) for 2 hours. Embryos were incubated overnight (rolling at 4°C) with fresh pre-incubation buffer and 1/2000 dilution of anti-DIG coupled with alkaline phosphatase (AP) (Roche). These were then visualised by application of BM purple until staining had occurred.

2.3.3 Sequence analysis and comparisons of CPS III

cDNA was produced from the RNA extracted from whole embryos using the above method for *A alcalica* and *A. grahami*. Multiple primer pairs (Table 1) were used to amplify fragments of *CPS III* from the cDNA via PCR and the products sent for sequencing. The coding region of *CPS III* was then constructed using multiple alignments against the *CPS I* and *III* from other species. The amino acid sequence was then examined for potential changes which could predict the functional differences seen in *Alcolapia*. Phylogenetic analysis was also used to confirm the *Alcolapia* genes analysed here are *CPS*

III (Figure 1). To determine potential changes in promoter region, a 3500bp section of genome (accession number NCBI: MW014910) upstream of the transcriptional site start of *CPS* from *A. alcalica* (unpublished genome)*, Oreochromis niloticus* (Nile tilapia)*, Xenopus tropicalis* (western clawed frog) and *Danio rerio* (zebrafish), genomes were accessed on Ensembl, aligned and examined for binding sites specific to the muscle transcription factor Myod1 (E-boxes) which preferentially binds paired E-boxes in the enhancer regions of myogenic genes with the consensus motif CAG(G/C)TG, as well as E-boxes more broadly (CANNTG). The published genomes of *O. niloticus, D. rerio* and *X. tropicalis* were accessed using Ensembl whereas the *Alcolapia* genome was constructed from whole genome sequences.

Table 1: Primer pairs used to amplify fragments of CPS III by PCR from the embryonic RNA extracted from A. alcalica and A. grahami. These were subsequently sequenced and aligned using BIOEDIT.

Segment of gene amplified (bases)	Forward primer	Reverse primer
1 (1-871)	ATGGCAAAAATCCTCCAAGCTG	CAAATACTGGCTGGGGACGATC
2 (782-1644)	CTAGTGGACCAGGAGATCCATC	CTGCCTGTCCTCTGTTGCC
3 (1567-2436)	CGTGGCATCTTGGACCAGTAC	GGCCTTCTGCATACTCTCC
4 (2344-3255)	CACGGCATGTCTCATGAAA	GTCGATCTGCAGGGGGCTCG
5 (3167-3987)	CAGTGGGAGGTCAGATTGC	CTCACAGCGAAGCACAGGG
6 (3830-4521)	TCAACGTGGCAACCAAAGTG	CTAACCCTGCTGGCTAGACCC

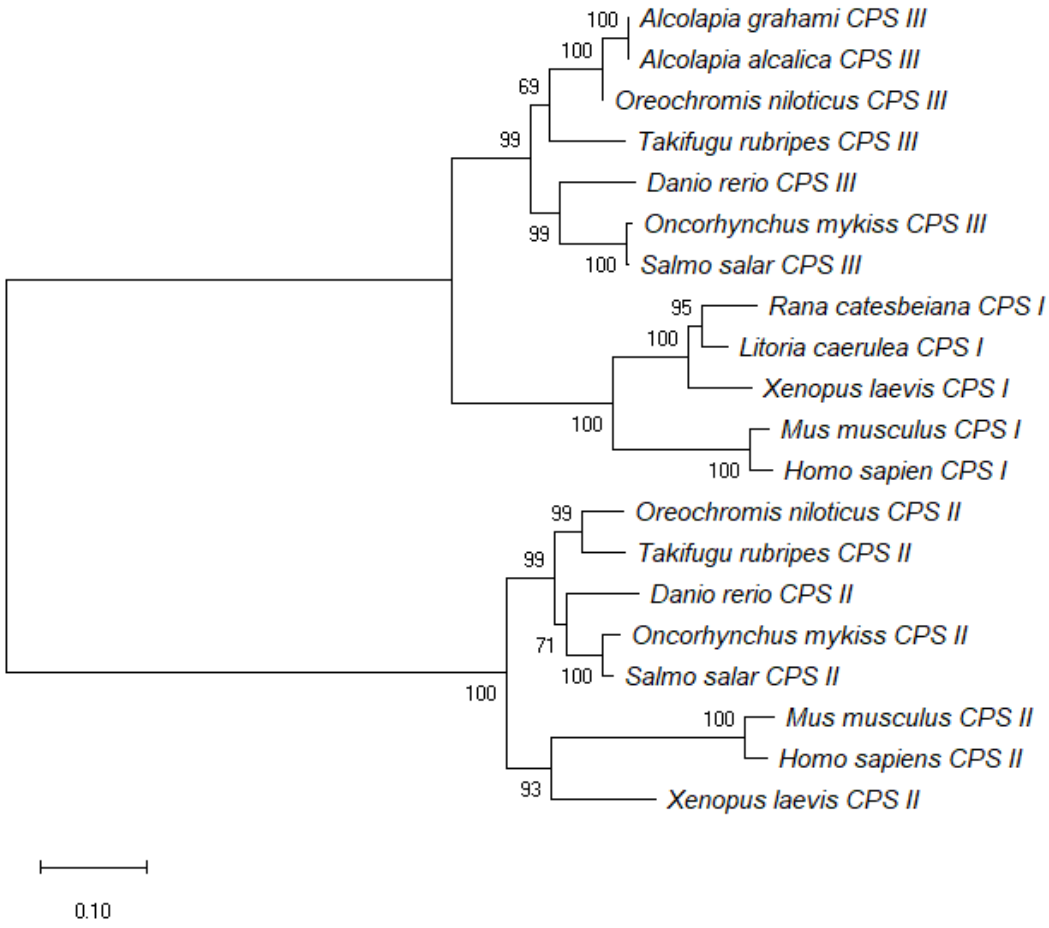


Figure 1: Phylogenetic analysis of the amino acid sequences of CPS proteins from a range of fish and tetrapod species. Analysis was conducted on MEGAX by the maximum likelihood method and the LG model with gamma distribution. All positions containing gaps were excluded and the tree was verified with 100 bootstrap replicates (shown on branches). CPS I/III produce a single clade separate from the CPS II amino acid sequences from a number of the same species. Both Alcolapia CPS III, which were sequenced in this study, clade as expected with CPS III from other fish species.

2.4 Results

2.4.1 ***CPS III* expression is activated early in the skeletal muscle lineage in *A. alcalica***

Analysis of gene expression of *CPS III* in dissected tissues of three adult *A. alcalica* shows that transcripts were only detected in adult muscle (Figure 2A). *In situ* hybridisation methods on *A. alcalica* embryos at different stages were carried out to investigate whether this restricted muscle expression was established during development (Figure 2B-F). Blue colouration indicates hybridisation of the complementary RNA probe and shows strongest expression in the developing somites along the body axis (arrows). Expression was also detected in migratory muscle precursors (MMP; arrowheads), which go on to form the body wall and limb musculature, and in the developing pectoral fin buds (white arrows). All regions of the embryo that show expression of *CPS III* are in the muscle lineage indicating that in *A. alcalica CPS III* expression is restricted to muscle tissues in both adults and the developing embryo.

Many muscle specific genes are activated during development by the muscle specific transcription factor, MyoD. The promoter region of *CPS III* (3.5kb upstream of the transcriptional start site) in *A. alcalica* was compared to that in *O. niloticus, X. tropicalis* and *D. rerio* (Figure 3). Examination of this region revealed a putative paired E-box MyoD binding site 940 to 970 bases (CAGGTGACTGTGATTATATAGTTCACAGGTG) upstream of the transcriptional start site of *CPS III* only in *Alcolapia* species. Intriguingly, while no pair of MyoD E-boxes were found in the upstream region of any other species examined, *O. niloticus* does have a single MyoD E-box motif in the same region upstream of *CPS III,* and within 19 bases of this is a CAGGTT motif which a single point mutation would convert

into a pair of E-boxes (CAGGTGACTGTGATTATATAGTTCACAGGTT). This suggests that it is possible that MyoD could bind and activate transcription of *CPS III* in the muscle of *Alcolapia* species, but not in the closely related *O. niloticus.*

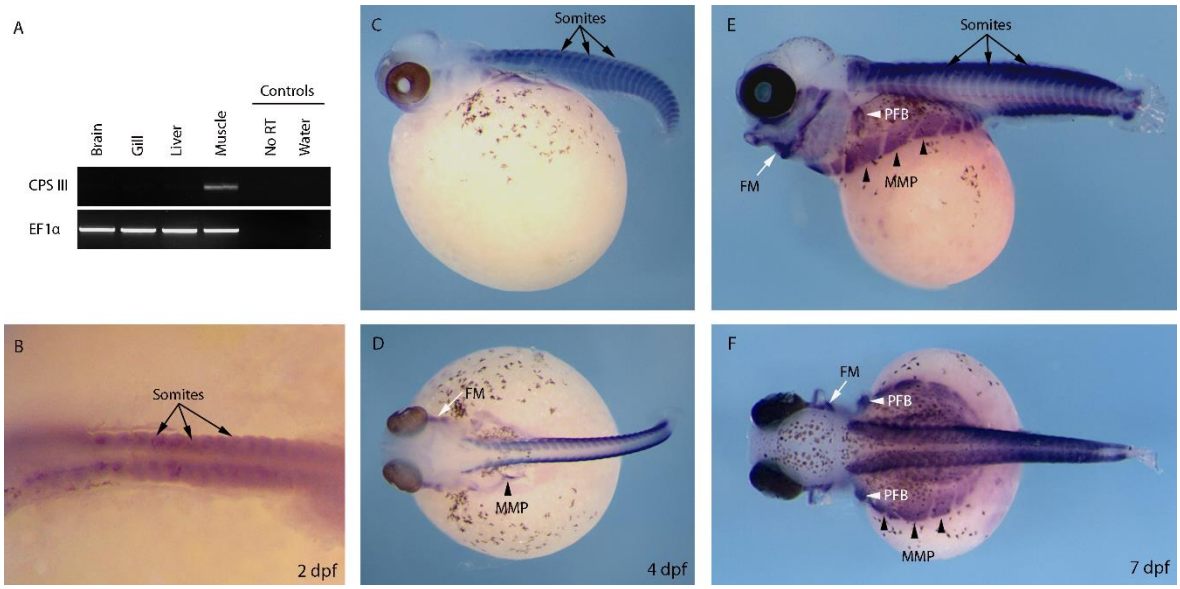


Figure 2: Expression analysis of carbamoyl-phosphate synthetase III (CPS III) from adult tissues and developing embryos of Alcolapia alcalica. A) Reverse transcriptase PCR and gel electrophoresis showing the muscle specific expression of CPS III, EF1α shown as normalisation control. B-F) Lateral (C and E) and dorsal (B, D, and F) views of in situ hybridisation for CPS III in developing A. alcalica embryos at different stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of mRNA. The black/brown is endogenous pigment apparent in the retina and the chromatophores. Black arrows show somites, arrowheads indicate region of migrating muscle progenitors (MMP), white arrows show facial muscle (FM) and white arrowheads indicate developing pectoral fin bud (PFB). Black dots around the yolk and on the body are chromatophores (pigment cells).

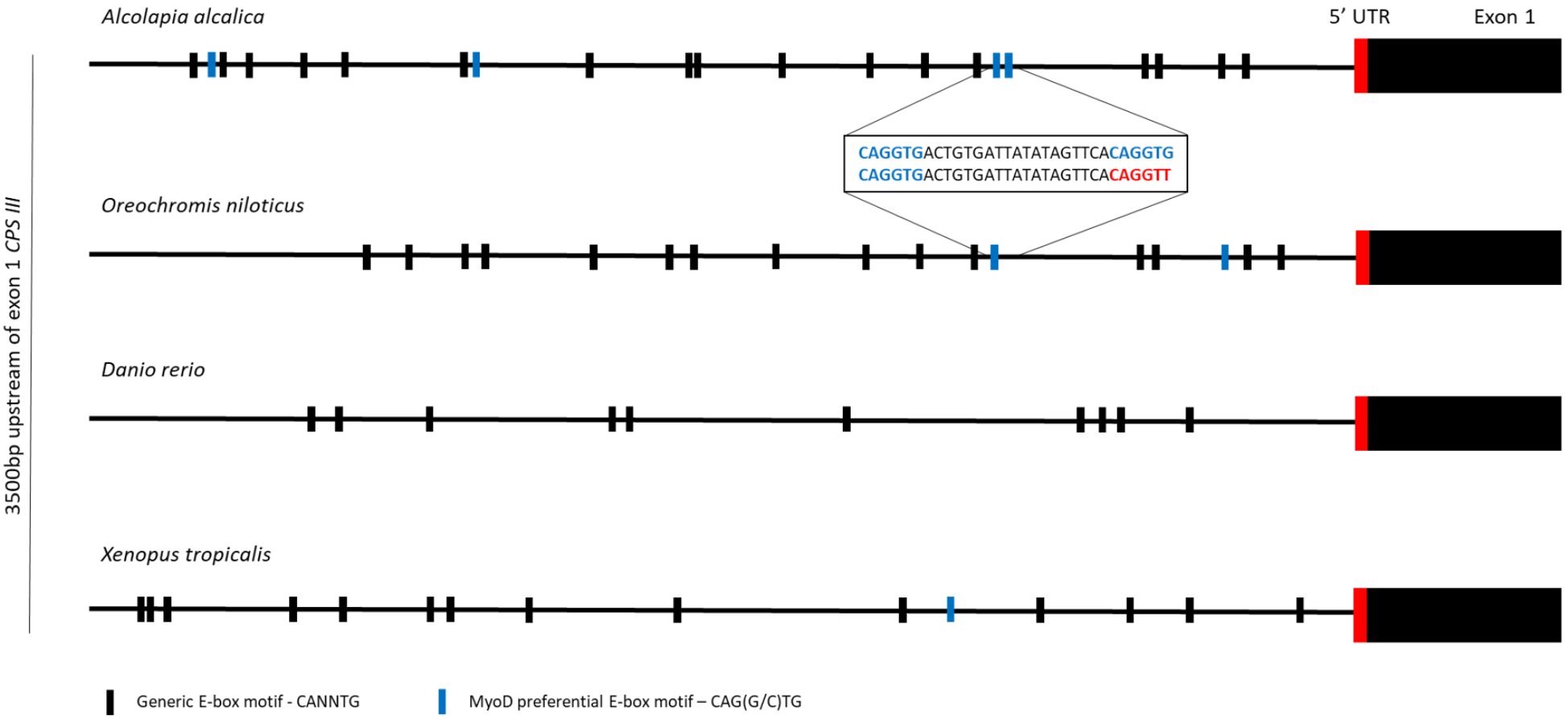


Figure 3: Sequence comparison of a 3500bp region upstream of exon 1 of CPS III. Presence of generic E-box sites with the consensus motif of CANNTG are annotated by black bars and MyoD preferential E-box motifs of CAG(G/C)TG are represented by blue bars. Insert shows sequence of *Alcolapia alcalica* potential paired E-box, MyoD enhancer (Blue nucleotides) compared to *Oreochromis niloticus* which has a single MyoD E-box upstream of a 'presumptive' MyoD E-box (Red nucleotides).

2.4.2 Convergent evolution in adaptive function of CPS III

Sequence analysis of *A. alcalica* and *A. grahami* CPS III revealed a discrepancy in the catalytic triad compared to the published sequence for CPS III in *A. grahami* (accession number NCBI: AF119250). The coding region for *A. alcalica* and *A. grahami* was cloned and sequenced (accession numbers NCBI: MT119353, MT119354). Our data confirmed the error in the published sequence of *A. grahami* CPS III and shows *Alcolapia* species maintain a catalytic triad essential for catalysing the breakdown of glutamine (red boxes in Figure 3). However, similar to terrestrial vertebrate CPS I which lack either one but usually both residues essential for binding glutamine for utilisation by the catalytic triad (arrowheads in Figure 3), *Alcolapia* also lack one of these residues (asterisks Figure 3). This amino acid sequence is consistent with a change in function permitting *Alcolapia* CPS III to bind and catalyse ammonia directly, an activity usually restricted to terrestrial vertebrate CPS I, as elucidated by extensive previous biochemical analyses (Saeed-Kothe and Powers-Lee 2002, Loong, Chng et al. 2012).

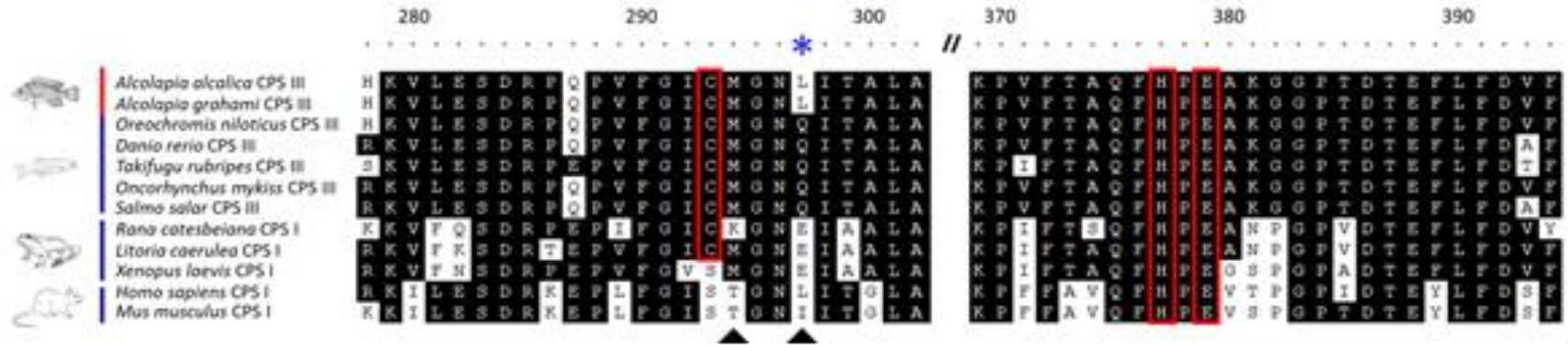


Figure 4: Multiple amino acid alignment of residues 278 to 397 (aligned to Alcolapia alcalica*) of carbamoyl-phosphate synthetase I and III from a number of tetrapod and teleost species respectively.* Conserved amino acids are shaded in black, amino acids in the catalytic triad are boxed in red and arrowheads indicate residues vital for glutamine utilisation of the catalytic triad. The blue asterisk indicates the divergent glutamine binding residue in *Alcolapia* species that likely results in a functional change (inability to bind glutamine).

2.5 Discussion

While most teleosts are ammonotelic, larval fish can convert ammonia to urea for excretion and to do so express the genes coding for the enzymes of the OUC, including CPS III (Zimmer and Wood 2016). Later these genes are silenced in most fish. In the rare cases where urea is produced in adult fish, the OUC enzymes are expressed in the liver (Mommsen and Walsh 1989), however there are some reports of expression in non-hepatic tissues (Korte, Salo et al. 1997, Banerjee, Koner et al. 2020). We report here the expression of *CPS III* in the muscle of adult *A. alcalica*, which is consistent with the detection of CPS III protein and enzyme activity in muscle of *A. grahami* (Lindley, Scheiderer et al. 1999). We also find conserved changes to the amino acid sequence which explains the convergent evolution of *A. alcalica* and *A. grahami* CPS III function with CPS I in terrestrial vertebrates. This conserved change in both *Alcolapia* species suggests that the adaptations in the OUC are likely to have evolved in the ancestral species inhabiting paleolake Orolongo during the period of changing aquatic conditions (over the past ten thousand years) that led to the extreme conditions currently found in Lakes Natron and Magadi.

2.5.1 Activation of CPS III in the myogenic lineage

We find that the expression of *CPS III* is activated in somites and in migratory muscle precursors that will form body wall and limb musculature (indeed expression is seen in developing pectoral fin buds). All skeletal muscle in the vertebrate body is derived from the somites, and these *CPS III* expression patterns are similar to those of muscle specific genes like myosin, actins and troponins (Bucher, Maisonpierre et al. 1988, Pownall, Gustafsson et al. 2002, Talbot, Teets et al. 2019).

Muscle specific expression of *CPS III* in *A. alcalica* embryos is a remarkable finding as most ureotelic species convert nitrogenous waste to urea in the liver (Randall, Wood et al. 1989, Loong, Chng et al. 2012). The expression of *CPS III*, the first enzyme in the OUC, in muscle tissue is likely significant for supporting the high catabolism in a fish species with the highest recorded metabolic rate (Wood, Brix et al. 2016). There are few reports of some OUC gene expression or enzyme activity in non-hepatic tissue including muscle (Korte, Salo et al. 1997, Kong, Edberg et al. 1998, Banerjee, Koner et al. 2020), nonetheless other fish species only evoke the activity of the OUC when exposed to high external pH or during larval stages (Walsh, Danulat et al. 1990, Mommsen and Walsh 1991, Wright, Iwama et al. 1993, Randall and Tsui 2002), and even then, urea production is never to the high level of activity occurring in *Alcolapia* (Lindley, Scheiderer et al. 1999). There is some heterogeneity of the expression patterns of *CPS III* during the development of different species in the teleost lineage; for example *D. rerio* has reported expression in the body (LeMoine and Walsh 2013), *Oncorhynchus mykiss* (rainbow trout) shows expression in the developing body but not in hepatic tissue (Steele, Chadwick et al. 2001) and *C. gariepinus* (African catfish) had *CPS III* expression detected in the dissected muscle from larvae (Terjesen, Chadwick et al. 2001). The early and sustained expression of *CPS III* in the muscle lineage is at this point an observation unique to *Alcolapia*.

Skeletal muscle specific gene expression is activated in cells of the myogenic lineage by a family of bHLH transcription factors, including MyoD (Pownall, Gustafsson et al. 2002). MyoD binds specifically at paired E-boxes in the enhancers of myogenic genes with a preference for the consensus motif of CAG(G/C)TG (Fong, Yao et al. 2012, Conerly, Yao et al. 2016). MyoD is known to require the cooperative binding at two E-boxes in close proximity, to modulate transcription of myogenic genes (Weintraub, Davis et al. 1990).

The presence of a pair of E-boxes in *Alcolapia,* upstream of a gene which has switched to muscle specific expression, is suggestive that MyoD is driving expression early in development. Enhancer modularity is a known mechanism for selectable variation (Shapiro, Marks et al. 2004) and although a single MyoD binding site does not define an enhancer, MyoD is known to interact with pioneer factors and histone deacetylases to open chromatin and activate gene transcription in the muscle lineage (Berkes, Bergstrom et al. 2004, Conerly, Yao et al. 2016). Experimental analysis to determine the activity of any regulatory sequences upstream of OUC genes in different species would shed light on the significance of putative transcription factor binding sites. This approach could also address another intriguing question as to the elements that drive the post-larval silencing of OUC genes in most fish species (LeMoine and Walsh 2013), an area with only minimal research especially when compared to the well characterised promoter region in mammalian species, for instance Christoffels et al 1998. A further instance of an extremophile organism redirecting expression of a hepatic enzyme to muscle tissue occurs in the crucian carp (Fagernes, Stensløkken et al. 2017). Under conditions of anoxia this species switches to anaerobic metabolism, producing ethanol as the end product of glycolysis (Johnston and Bernard 1983, Dhillon, Mandic et al. 2018). This is associated with the expression of *alcohol dehydrogenase* in muscle (Nilsson 1988). Together with our findings, this potentially reveals an example of convergent evolution whereby the muscle becomes the site for detoxifying by-products of metabolism. Elucidating any mechanisms that may include modular enhancers that facilitate the adaptation of gene regulation in response to changing environmental conditions will be of significant interest.

2.5.2 Convergent evolution of adaptive CPS III function

CPS proteins catalyse the production of carbamoyl-phosphate as a first step in nitrogen detoxification by accepting either glutamine or ammonia as a nitrogen donor (Rubio and Cervera 1995). Teleost CPS III binds glutamine: the nitrogen source provided by the amide group of glutamine is catalyzed by the conserved catalytic triad Cys-His-Glu in the glutamine amidotransferase (GAT) domain in the amino terminal part of CPS (Loong, Chng et al. 2012). In terrestrial vertebrates, CPS I lacks the catalytic cysteine residue and only generates carbamoyl-phosphate in the presence of free ammonia (Saeed-Kothe and Powers-Lee 2002). Although CPS in *Alcolapia* shares most sequence identity with fish CPS III (Figure 4), its ammonia binding activity is more similar in function to terrestrial vertebrate CPS I (Lindley, Scheiderer et al. 1999, Loong, Chng et al. 2012). This adaptation to preferentially bind ammonia over glutamine supports efficient waste management in a fish with an exceptionally high metabolic rate (Wood, Brix et al. 2016). CPS I in terrestrial vertebrates have amino acid changes in the catalytic triad which explains their binding ammonia over glutamine; a reduction in glutamine binding capacity drives the use of ammonia (Saeed-Kothe and Powers-Lee 2002). Here we show that *Alcolapia* maintain the catalytic triad, but (similar to mouse and human) lack one of the two residues required for efficient glutamine binding, weakening its affinity to glutamine, and driving the use of ammonia as a primary substrate.

The interesting observation that bullfrog (*Rana catesbeiana*) CPS I retains the catalytic triad, but lacks the two additional conserved amino acids required for glutamine binding, has led to the suggestion that the change from preferential glutamine to ammonia binding originally evolved in the early tetrapod lineage (Saeed-Kothe and Powers-Lee 2002). A further frog species, the tree frog *Litoria caerulea*, retains it catalytic triad and

only one of the two residues required for glutamine binding has been altered, weakening its affinity for glutamine and allowing for direct catabolism of ammonia (Ip, Loong et al. 2012). Much the same as in *Alcolapia, L. caerulea* CPS I is still capable of using glutamine to some extent which lends further support to the notion that the evolutionary transition from CPS III to CPS I occurring in amphibians and the early tetrapod lineage. The changes in protein sequence of *Alcolapia* CPS III represents a convergent evolution in this extremophile fish species, with acquired changes in functionally important domains which likely also evolved in early terrestrial vertebrate CPS I.

2.5.3 Conclusions

Alcolapia have acquired multiple adaptations that allow continued excretion of nitrogenous waste in a high pH environment. Among these is the novel expression of *CPS* in skeletal muscle, as well as acquisition of mutations that change its function. Sequence evidence indicates that like terrestrial vertebrates, and unique among fish, *Alcolapia* CPS III is capable of binding and catalysing the breakdown of ammonia to carbamoyl-phosphate; a convergent evolution of CPS function. The mechanism by which the novel and unique expression of *CPS* in muscle evolved is likely a function of enhancer regions of *A. alcalica* and *A. grahami* that result in its regulation by muscle regulatory factors to direct *CPS* expression in the myogenic lineage during embryonic development. Environmentally driven adaptations have resulted in changes in both the expression and activity of CPS III in *Alcolapia* that underpin its ability to turnover nitrogenous waste in a challenging environment while maintaining a high metabolic rate.

Chapter Three – Characterisation of Rhbg, a functional ammonia transporter in a ureotelic fish species

3.1 Summary

This chapter aimed to explore the function of Rhesus glycoproteins, a family of ammonia transporters, in *A. alcalica;* a species thought to excrete 100 % of its nitrogenous waste as urea. Spatiotemporal expression of *Rhbg, Rhcg1* and *Rhcg2* in adult *A. alcalica* gill tissue and embryos was compared to their homologs in *D. rerio.* Expression patterns within adult and developing gill tissues were conserved between these species. This study also showed that adult and embryonic *A. alcalica* excrete a proportion of their waste as ammonia, contrary to previous studies. Overexpression of *A. alcalica Rhbg* mRNA in *D. rerio* embryos increased the amount of ammonia detected in the holding water of embryos when compared with wildtype controls suggesting *Rhbg* is still functional as an ammonia transporter. Collaboration with the Thomas lab showed similar results when this gene was expressed in yeast.

3.2 Introduction

3.2.1 Excretion of nitrogenous waste in fish

Ammonia is the major nitrogenous waste product in typical fish species (Wood and Evans 1993) and is produced as a result of the catabolism of amino acids via transdeamination (Wright and Fyhn 2001). At high concentrations ammonia is toxic (Handy and Poxton 1993, Oja, Saransaari et al. 2016), it accumulates around glial cells in the nervous system leading to neurotoxicity by disrupting the buffering of potassium (Thrane, Thrane et al. 2013). Due to its toxic effects, ammonia must either be quickly excreted or converted into less toxic compounds such as urea (Mommsen and Walsh 1991). Most fish are ammonotelic and excrete their nitrogenous waste as ammonia, predominantly across gill tissue. This method of excretion avoids the metabolic costs of converting ammonia to less toxic compounds

(Mommsen and Walsh 1991). Additionally, gill permeability to ammonia is at least double that of urea, therefore the majority of marine and freshwater fish utilise this compound for nitrogen excretion (Wright 1995). Under certain environmental conditions, such as heightened pH, ammonia diffusion is blocked and ammonia will accumulate in the body (Wright and Wood 1985). The alkalinity in Lakes Natron and Magadi reaches pH 11 (Ford, Dasmahapatra et al. 2015), the fish that survive there, *Alcolapia spp,* have adapted to these conditions by evolving the capacity to convert all ammonia to urea, making them the only described 100 % ureotelic fish (Randall, Wood et al. 1989). This is due to the high external pH preventing passive ammonia diffusion and an increased need for excretion because of a heightened metabolic rate (Randall, Wood et al. 1989). Some other species are able to utilise urea excretion under such conditions, however not to the level of *Alcolapia* (Walsh, Danulat et al. 1990, Wright, Iwama et al. 1993).

Production of urea in *Alcolapia* is through a fully functioning ornithine-urea cycle (OUC) (Randall, Wood et al. 1989) and, as detailed in chapter 2, there are a number of adaptations in this pathway that promote its efficient production. Upon conversion, the majority of urea is excreted across the gills (80 %) with the remainder being excreted as urinary urea (Narahara, Maina et al. 1994). Excretion across gill tissue in *Alcolapia* is continuous, unlike in other species such as in the gulf toadfish (*Opsanus beta*) which excrete discrete pulses of urea (Wood, Hopkins et al. 1995, Walsh, Grosell et al. 2001). This finding led to the discovery of a specialised urea transporter, the Magadi tilapia urea transporter (mtUT), which increases the permeability of the gills to urea by 10 times that of passive diffusion through a lipid bilayer and 5 times greater than that seen in *O. beta,* which does not have this particular urea transporter (Walsh, Grosell et al. 2001).

3.2.2 Function of Rhesus glycoproteins

The relatively recent discovery that Rhesus glycoproteins (Rh) have ammonia-transporting functions has also changed the way ammonia excretion is viewed in fish (review in Wright and Wood 2009, Zimmer, Wright et al. 2017). A prevalent model is that ammonia is removed through passive diffusion where Rh proteins recruit NH_4^+ for deprotonation before transporting it across the gill, through the transporter, as NH_3 (Khademi, O'connell et al. 2004, Wright and Wood 2012, Abdulnour-Nakhoul, Le et al. 2016, Yeam, Chng et al. 2017). NH_3 is then protonated in the gill acid boundary layer and released into the surrounding water (Weihrauch, Wilkie et al. 2009). However, when fish are exposed to high external ammonia or high pH this mechanism is blocked (Wright and Wood 1985).

There are four teleost Rh proteins with ammonia transport functions and the cellular localisation of these proteins has been determined. Rhag is present in the membrane of red blood cells and Rhbg and Rhcg isoforms are predominantly detected in gill tissue (Nakada, Hoshijima et al. 2007, Wright and Wood 2009). In addition, some expression of different Rh proteins has been reported in skin, kidney and even brain tissue in different fish species (Hung, Tsui et al. 2007, Braun, Steele et al. 2009, Nawata and Wood 2009). Their most important role is believed to be in the gills where the majority of ammonia excretion takes place. One group has used immunohistochemistry to show that Rhbg is present on the basolateral membrane while Rhcg2 is on the apical membrane of pavement cells in *D. rerio* (Figure 5). This arrangement allows the proteins to work together to transport ammonia across gill tissues into the surrounding water (Nakada, Hoshijima et al. 2007, Nakada, Westhoff et al. 2007). Another Rh protein, Rhcg1, is found on the apical membrane of mitochondrial rich cells and is believed to act in conjunction

with a basolaterally positioned Na^+, K^+-ATPase, with ammonia substituting for K^+ allowing excretion across this cell type (Nakada, Westhoff et al. 2007). This isoform was also shown to be apically positioned in *Danio rerio* kidney tissue (Nakada, Hoshijima et al. 2007). While this organisation of Rh proteins has been described in the fish gill, the expression of Rhbg in larval skin during early development, in the absence of Rhcg2, suggests the collaboration of Rh proteins is not always required for ammonia transport (Braun, Steele et al. 2009). Selective knockdown using morpholinos against Rhag, Rhbg and Rhcg1 in developing *D. rerio* resulted in around a 50% reduction in ammonia excretion, regardless of the Rh targeted. This suggests that all Rh proteins are required for efficient ammonia excretion to be maintained which may be due to their differential cellular localisation (Braun, Steele et al. 2009).

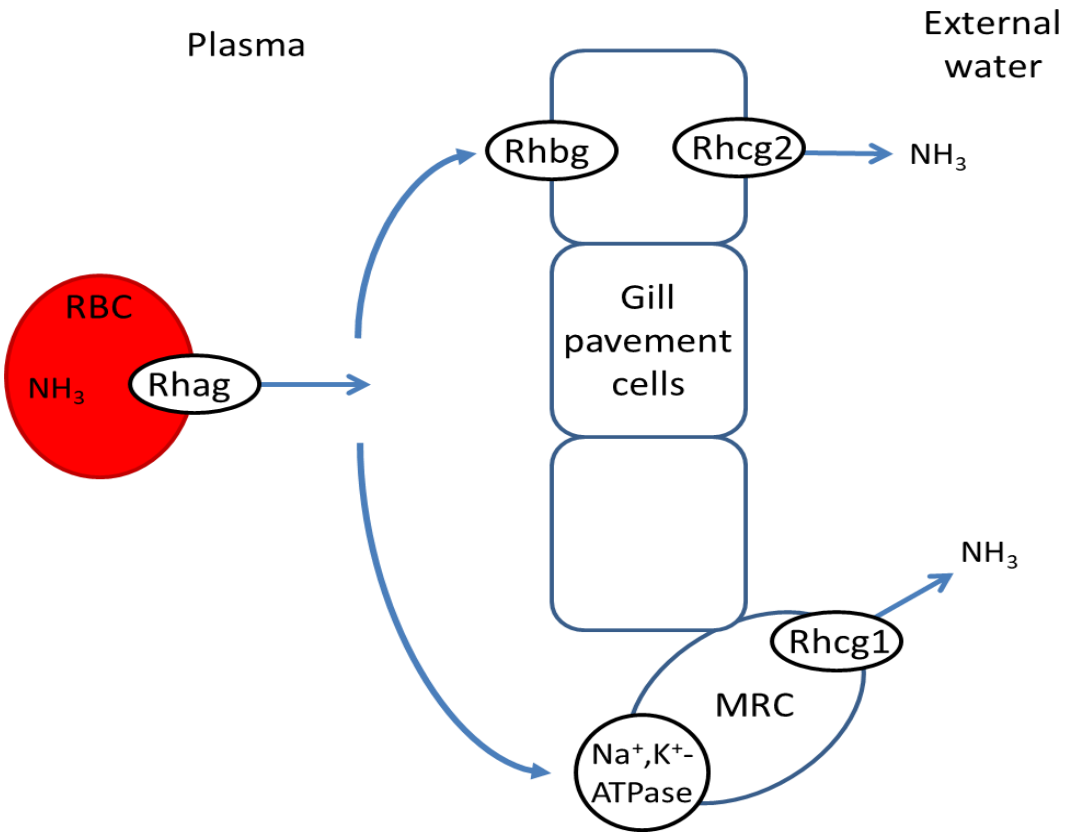


Figure 5: Schematic representation of the hypothesised cellular localisation of the four Rhesus glycoproteins and the ammonia excretion pathway that they facilitate in normal fish gills. Rhag in the erythrocyte membranes, Rhbg in the basolateral membranes of branchial epithelial cells (gill pavement cells which make up the gill filament), Rhcg2 in the apical membranes of branchial epithelial cells, and Rhcg1 in the apical membrane of mitochondrial rich cells (MRC) at the base of the gill filament, transporting ammonia with the aid of Na^+,K^+-ATPase. Ammonia is hypothesised to be transported through the channels as NH_3. Adapted from Wright and Wood 2009 and Nakada et al., 2007.

Investigations into the presence of these ammonia transporters in *Alcolapia* has shown that transcription of Rh proteins persists in gill tissue (Wood, Nawata et al. 2013), despite the fact that these fish do not excrete ammonia (Randall, Wood et al. 1989). Interestingly, the expression of *Rhbg* and *Rhcg2* has been detected in liver, muscle and intestinal tissue (Wood, Nawata et al. 2013), suggesting that Rh proteins could be used to shuttle ammonia into tissues with known OUC activity (Lindley, Scheiderer et al. 1999). The

continued transcription of *Rhbg* and *Rhcg2*, as well as their apparent upregulation when exposed to increased external ammonia, is surprising given that *Alcolapia* are thought to excrete 100 % of nitrogenous waste as urea (Wood, Nawata et al. 2013). It has been suggested that *Alcolapia* Rhbg and Rhcg2 are structurally different from their orthologues in other species, with 10 rather than 12 transmembrane domains (Wood, Nawata et al. 2013), raising the possibility that *Alcolapia* Rh proteins transport another compound or are acting to channel ammonia to tissues for conversion to urea rather than excreting ammonia directly. There is some evidence for this theory as it has been shown that a single point mutation in human Rhag converts it to a cation-exchanger due to a decrease in pore size (Bruce, Guizouarn et al. 2009). Additionally, some Rh proteins have been found to transport other compounds, such as CO_2 or HCO_3^- (Huang and Ye 2010), further supporting the notion that Rh proteins may transport something other than ammonia in *Alcolapia*.

3.2.3 Aims of this chapter

Alcolapia are described as the only noted 100 % ureotelic fish species, yet research has shown they have retained expression of Rh proteins, a family of ammonia transporters. This chapter asks the question; what is the purpose of these ammonia transporters in a fish not believed to excrete ammonia?

This chapter will use sequence analysis, including phylogenetics and tests of positive selection, to determine whether *Alcolapia* Rh proteins significantly differ from those of other fish species. Their RNA expression will be analysed in adult gill tissue and in developing embryos and compared to the model, non-extremophile, fish species *D. rerio,* using *in situ* hybridisation. A main aim of this chapter is to determine whether the

Alcolapia Rh proteins have retained their ability to transport ammonia by overexpressing the *A. alcalica* Rh homologues in developing *D. rerio* embryos via RNA injection and comparing these to wildtype *D. rerio* embryos and those where the gene has been targeted with CRISPR-Cas9 gene editing technology.

3.3 Methods

3.3.1 Phylogenetic analysis and tests of positive selection

Phylogenetic analysis was constructed on MEGAX using the LG method with gamma distribution against the amino acid sequences of the *Alcolapia* Rh proteins (extracted from the unpublished *Alcolapia* genome) along with seven other cichlid species and *D. rerio* and *Takifugu rubripes* as out groups (taken from NCBI). The phylogeny was evaluated by Bootstrap analysis with 100 replicates using all sites. For analysis of occurrences of positive selection within the *Alcolapia* lineage the aligned nucleotide sequences were analysed against the species phylogeny by calculating the dN/dS ratios, for each Rh gene, using codeML on PAML X (Xu and Yang 2013). The composite species phylogeny was constructed with the use of recent literature to infer the currently best accepted hypothesis of phylogenetic relatedness amongst species (Chen, Ortí et al. 2004, Wagner, Harmon et al. 2012, Cao and Xia 2016). Gene synteny analysis was conducted using the NCBI genome viewer on the genomes of *D. rerio* and the close relative of *Alcolapia, O. niloticus,* to further support that the genes discussed here are true homologues. A fully assembled genome is not available for *Alcolapia* species hence the use of *O. niloticus* for this analysis.

3.3.2 Temporal and spatial expression of Rhesus glycoproteins

RT-PCR was used to determine the expression of the different RNA of the Rh proteins in multiple tissues of *A. alcalica* and *D. rerio* as described in section 2.2.1 using the primers shown in Table 2. Gel electrophoresis of the PCR products determined the presence or absence of these RNA in different tissues (n=3). These products were ligated into PGem-tEasy and transformed into the *E.coli* strain DH5α. These were then used to create RNA probes for in situ hybridisation which was carried out on *A. alcalica* and *D. rerio* embryos as described in section 2.2.2.

Table 2: Primers used for amplification of cDNA via RT-PCR of the Rh genes in Alcolapia alcalica and D. rerio, annealing temperatures and predicted product sizes shown.

Gene target	Primer sequence	Annealing temp (˚C)	Product size (bp)
Rhbg – A. alcalica	Forward – TGACTTTCCTCCAGCGTTATG	60	731
	Reverse – GAAGCCCAGAACAGAGATGATAC		
Rhcg1 – A. alcalica	Forward – CCTGGACTACACTGATGGAAAG	58	738
	Reverse – CTCCCACAATGCCACCTATAA		
Rhcg2 – A. alcalica	Forward - GCGTGGGCTTTAACTTTCTAATC	60	707
	Reverse - GTACCCAAAGGTGGAGATGATAC		
Rhbg – D. rerio	Forward – CTACTATCGCTATCCCAGCTTTC	58	662
	Reverse – GTTTGCCCTCTGGATTGACTA		
Rhcg1 – D. rerio	Forward – GTCCTCACTACTGTGGCTATTTC	56	633
	Reverse – GCAAAGTTGGTCTCTCTCATCT		
Rhcg2a – D. rerio	Forward – TCTCACTACCTTTGCCCTTTC	58	602
	Reverse – CATCTGGCGTCCTTTCTACAT		
Rhcg2b – D. rerio	Forward – GTCTACCACTCCGATGTCTTTG	58	688
	Reverse – CATCTTCAGGTACCTCCCAATAAA		

3.3.3 Gene targeting of *Rhbg* in *D. rerio*

CRISPR-Cas9 gene editing technology was used to target *Rhbg* in *D. rerio. Rhbg* was chosen as the focus for gene editing and overexpression experiments because it is the Rh under the strongest positive selection in *Alcolapia* species. sgRNA was designed against the DNA sequence of exon one of the *D. rerio Rhbg* gene using ChopChop (https://chopchop.rc.fas.harvard.edu/) for the target sequence GCATTCTCTTTGGAGTACTGG. Disruption at the 5' part of the gene would result in a truncated protein. sgRNA template was generated using PCR and transcribed using T7 MEGAshortscript (Ambion). Production of sgRNA and Cas9 protein was done as described in McQueen and Pownall, 2017 using the primers in Table 3. *D. rerio* embryos were injected with a 1nl solution containing 300ng/ml sgRNA, 400mM KCl and 1ng of purified Cas9 enzyme. Embryos were raised in petri dishes containing E3

media (5mM NaCl, 0.17mM KCl, 0.33mM CaCl2, 0.33mM MgSO4) with 0.1%

methylene blue before being transferred to a freestanding aquatic through flow system and raised to maturity at 28°C.

Table 3: Name and sequence information of the primers used to create the sgRNA for targeting Rhbg in Danio rerio with CRISPR-Cas9 gene editing technologies. Forward primers include complementary target sequences, shown in italics, as well as the 5' promoter for transcription by T7 RNA Polymerase.

Primer name	Primer sequence 5' – 3'
Rhbg: exon 1- forward primer	GCAGCTAATACGACTCACTATAG GCGATAGTAGAACTCATTCGGTTTTAGAGCTAGAAATA
Universal reverse primer	AAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACG GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

3.3.4 Overexpression of Rh protein RNA in *D. rerio*

Further functional testing of the *A. alcalica* Rh proteins was conducted via the overexpression of Rh protein mRNA in *D. rerio* embryos. RNA was extracted from *A. alcalica* and *D. rerio* embryos and cDNA produced as in section 2.2.1. 1 µl of this cDNA was used to amplify the full transcripts of each of the *Rh* genes from *A. alcalica* and *Rhbg* from *D. rerio* by PCR using Table 4. Forward primers incorporated a EcoR1 digestion site and reverse primers incorporated a Xba1 digestion site to ensure correct orientation in plasmid upon ligation. Products were ligated into the expression plasmid $CS2^{+}$, transformed into the *E.coli* strain DH5α and miniprep DNA prepared as in section 2.2.2. Plasmids were linearised by restriction enzyme digest using NotI and purified by phenolchloroform and isopropanol precipitation. Transcription of synthetic length mRNA was done using mMESSAGEmMACHINE® SP6 kits as per manufacturer's instructions, products were then checked for successful transcription using gel electrophoresis and concentration determined using a NanoDrop.

Table 4: Primers used for the amplification of cDNA via RT-PCR of full length Rhbg from Alcolapia alcalica and Danio rerio. These products were ligated into expression plasmid $CS2^{+}$ for producing mRNA.

Gene target	Primer sequence including digestion site (underlined)
Rhbg – A. alcalica	Forward – AGAGAGGAATTCACCATGGCAAACACAAATACCAACATG
	Reverse – AGAGATCTAGAGTTGAGCTTCTCAGTCTCCTCTGG
Rhbg – D. rerio	Forward – AGAGAGGAATTCACCATGGCTGAGTCAACTAACTTGAGG
	Reverse - AGAGATCTAGAGCTGTTGAGTTTCTCCACCTCGTT

Breeding pairs of adult *D. rerio* were placed in breeding tanks with a divider which was removed when embryos were required. The breeding tanks allow the embryos to be collected at a known time of fertilisation allowing direct access to single cell embryos for injection. 1 nl of 150 ng μl^{-1} mRNA was injected into the yolk cell just under the blastomere of single cell *D. rerio* embryos using a Harvard apparatus gas micro-injector. Overexpression was carried out in embryos from wildtype fish (AB) and compared to CRISPR-Cas9 injected embryos and wildtype controls.

3.3.5 Measurement of ammonia and urea excretion levels

The concentrations of ammonia and urea excreted into surrounding water was measured for adult and embryonic *A. alcalica* and *D. rerio.* Embryos of *D. rerio* included in this study were wildtype, *Rhbg* CRISPR-Cas9 targeted, and embryos where *Rhbg* had been overexpressed. Single adult *A. alcalica* and *D. rerio* were incubated in 600 ml or 400 ml of water from their respective tanks for 3 hours. After which each fish individually weighed before being returned to their home systems. Two water samples were taken for each fish which were immediately frozen at -20℃. Embryos were grouped (5 for *A. alcalica* and 30 for *D. rerio*) in 30 ml petri dishes in 2 ml of liquid, *A. alcalica* in tank water and *D. rerio* in E3 (no methylene blue). Embryos were incubated for 6 hours on a rocker at 28℃ after which chorions were opened under microscope with forceps. This was to ensure all excretory products were accounted for, including those which had not diffused across the chorion. Two water samples were then taken and frozen at -20℃ for future determination of excretory product concentration.

3.3.6 Determination of ammonia concentration

To calculate the concentration of ammonia in sample water 400µl was combined with 100µl of reaction buffer (40 gL^{-1} sodium tetraborate, 40 mgL^{-1} sodium sulphite, and 50 mlL^{-1} of phthalaldehyde dissolved in ethanol, once made the solution is usable for 3 months if stored in the dark) in an Eppendorf tube and vortexed. All samples were then incubated for 3 hours in the dark alongside samples of known ammonia concentration (0 to 100 µM) made using dissolved ammonia sulphate, this was multiplied by two to represent the amount of nitrogen present. After incubation 300 µl of each sample and those of known concentration was moved into a 96 well plate and the absorbance read at 350 nm on an Infinite m200 Pro (Tecan). Concentration in samples was determined using the equation of the straight line produced from a calibration curve of samples of known concentration. Protocol is adapted from Holmes et al. 1999.

3.3.7 Determination of urea concentration

To calculate the concentration of urea in collected water samples 350 µl was combined with 25 µl of DAMO-TSC (8.5 g of diacetylmonoxomine to 240 ml of dH_2O and 10 ml thiosemicarbazide solution [0.95 g in 100 ml) and 80 µl of acid-ferric solution (30 ml of sulphuric acid in 23.5 ml of dH_2O and 50 µl of ferric chloride [0.15 g in 10 ml]) in Eppendorf and vortexed. Reagents were all made fresh on the day of analysis. Samples were incubated at 85˚C for 20 minutes. A calibration curve of known concentration of urea (0 to 100 µM) was analysed in the same way, this was multiplied by two to represent the amount of nitrogen present. 300 µl of each sample and those of known concentration was moved into a 96 well plate and the absorbance read at 585 nm on an Infinite m200 Pro (Tecan). Concentration in samples was determined using the equation of the straight

line produced from a calibration curve of samples of known concentration. Protocol was adapted from Mulvenna and Savidge 1992.

3.3.8 Statistical analysis of the concentration of excretory products

Statistical analysis was carried out on R. To best control for differences in weight, number of embryos used for each sample and amount of water used for experiments, the wildtype adults and embryos of *D. rerio* and *A. alcalica* were analysed as percentage of nitrogen as ammonia (ammonia/[ammonia + urea] X 100) and urea (urea/[ammonia + urea] X 100). T tests were conducted to determine whether ammonia or urea respectively differed between samples. Data for adult *A. alcalica* is also shown as $\mu M\ N\ g^{-1}\ h^{-1}$. For experimental wildtype and injected embryos total values for concentration of nitrogen as ammonia and urea were used for analysis. Two-sample T-tests were conducted to determine whether any of the injection treatments significantly differed from wildtype.

3.4 Results

3.4.1 Phylogenetic analysis and tests of positive selection

Phylogenetic analysis of the amino acid sequence of the Rh protein sequences confirm that *Alcolapia* Rh proteins are orthologous to those in other species included in the analysis and have been annotated correctly (Figure 6). To determine whether any of the *Alcolapia* Rh genes are under positive selection, a PAML analysis was conducted comparing the ratio of non-synonymous to synonymous codon changes (dN/dS). The tests of positive selection using PAML gave varying dN/dS ratios for the four Rh proteins in the *Alcolapia* lineage (Figure 7). *Rhag* and *Rhcg1* show no signs of positive selection, both having dN/dS ratios less than 1. In contrast, results from these tests indicate that both *Rhbg* and *Rhcg2* are under positive selection with heightened dN/dS ratios. Analysis of

gene synteny (Figure 8) for the rhesus glycoproteins in *D. rerio* and *O. niloticus* shows that the genes for *Rhbg, Rhcg1* and *Rhcg2* in *O. niloticus* are located in the same part of the genome as *Rhbg, Rhcg1* and *Rhcg2b* respectively in *D. rerio.* This suggests that *Rhcg2a* is the duplicated gene in *D. rerio* and further confirms that the genes compared here are homologous. Sequence data for the *A. alcalica Rh* was submitted to NCBI (accession numbers: MW448158, MW448159, MW448160).

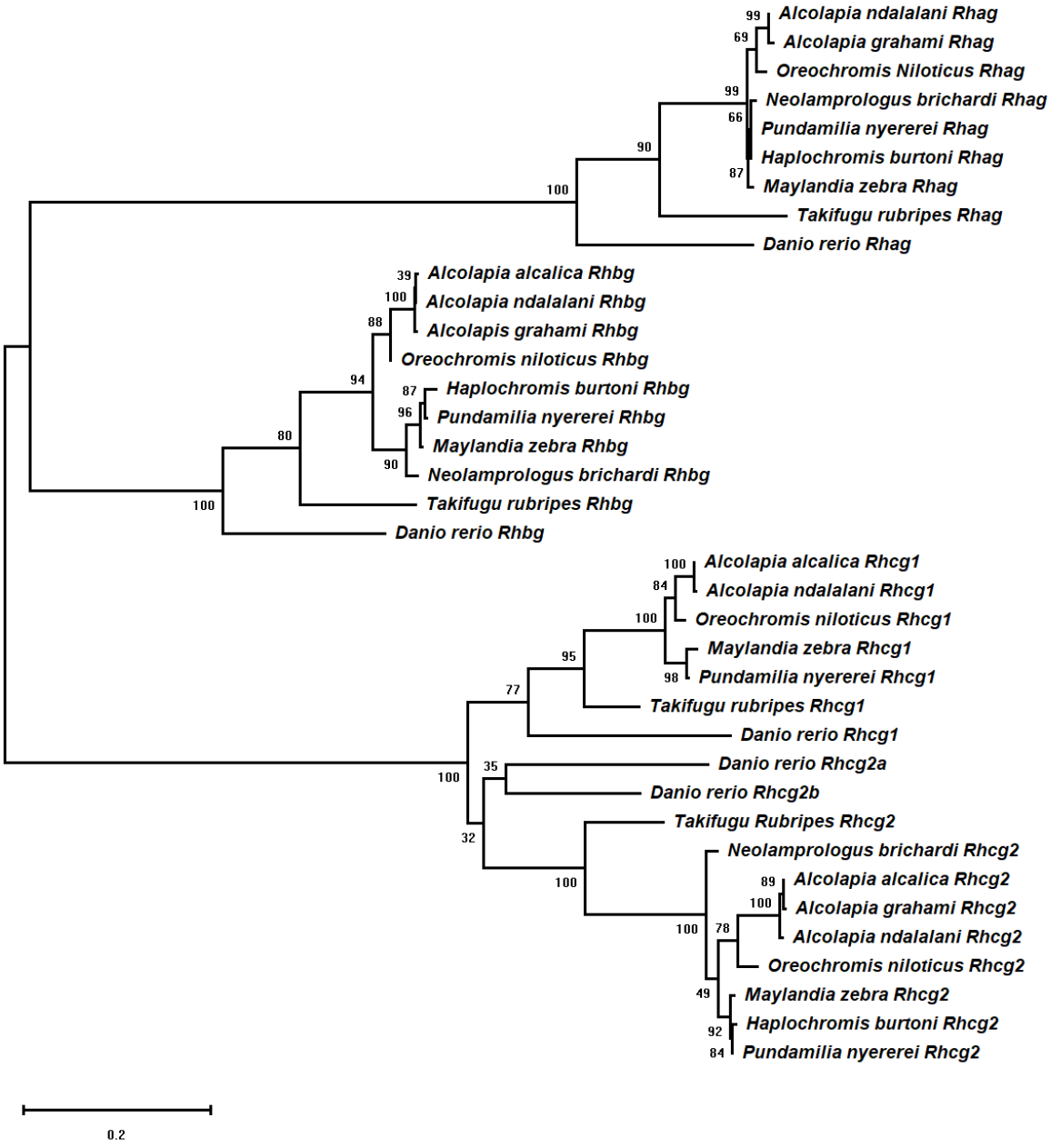


Figure 6: Phylogenetic relationship of Rhesus glycoproteins based on seven cichlid species, with Takifugu rubripes and Danio rerio selected as outgroups. Generated from amino acid alignments using MEGAX and constructed using LG method with gamma distribution with 100 bootstrap replicates (support values indicated below branches). The scale bar represents a genetic distance of 0.2 amino acid substitutions per site.

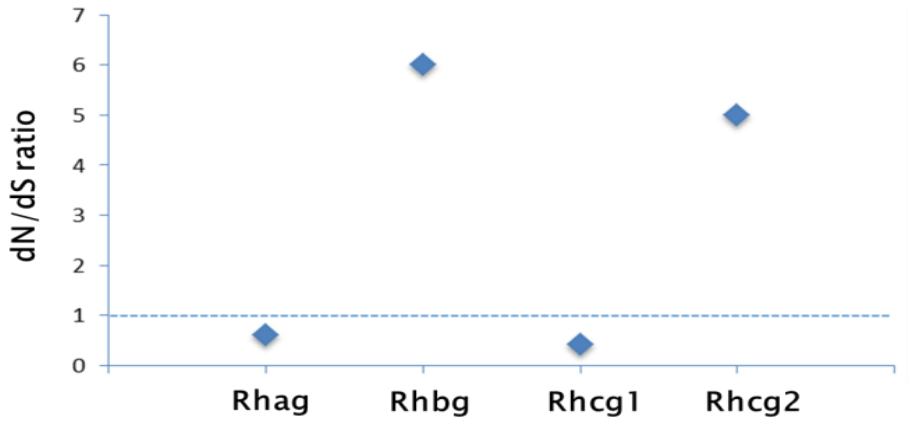


Figure 7: dN/dS values for the Rhesus glycoproteins (Rh) in the Alcolapia lineage. Values greater than one indicate occurrences of positive selection. Synonymous (dS) and non-synonymous (dN) mutations for each Rh gene were compared to find those under positive selection within the Alcolapia lineage (dN/dS ratios) using PAML.

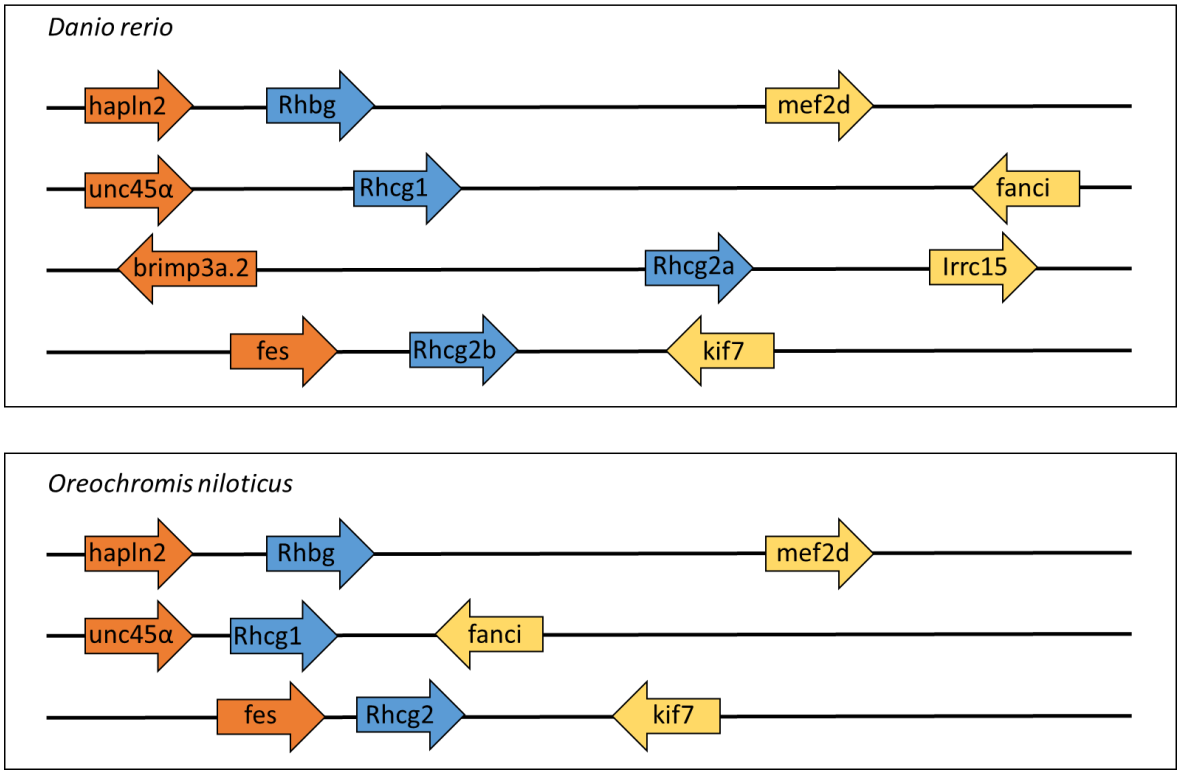


Figure 8: Gene synteny of the Rhesus glycoproteins in Danio rerio and Oreochromis niloticus. The position of the multiple genes for the Rhesus glycoproteins in the genomes of the two fish species is shown confirming that they are syntenic.

3.4.2 Expression analysis of *Rh* genes using RT-PCR

Total RNA was extracted from adult tissue from *D. rerio* and *A. alcalica* and the presence of rhesus proteins transcripts was detected using RT-PCR. Figure 9A shows that in *D. rerio*, *rhbg* expression is seen in gill and brain tissue, while the other Rh protein genes are expressed exclusively in the gill. Rhcg2a is expressed lower than the other Rh protein genes. In *A. alcalica*, Figure 9B, the expression of *Rhbg*, *Rhcg1* and *Rhcg2* is restricted to gill tissue (Figure 9). These data contradict a previous report that *Rhbg* and *Rhcg2* are expressed in muscle and liver and failed to show any expression for *Rhcg1* (Wood, Nawata et al. 2013).

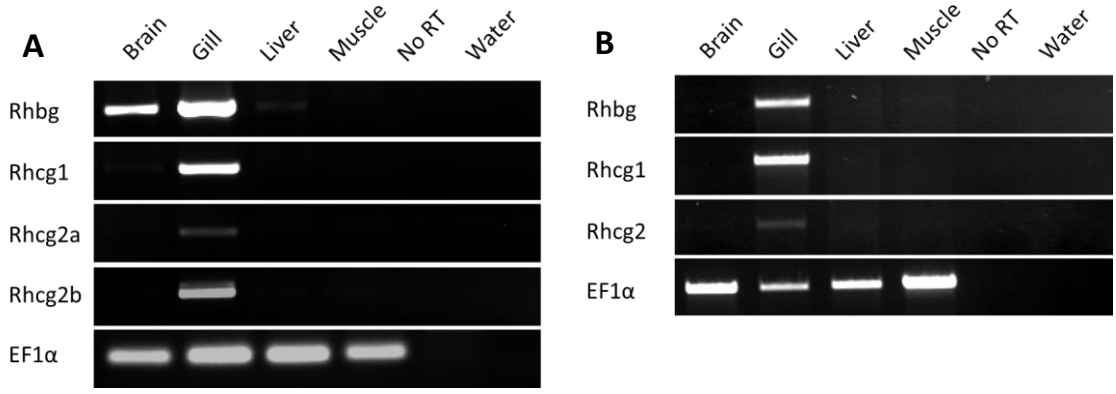


Figure 9: Reverse transcriptase PCR and gel electrophoresis showing the tissue specific expression of members of the Rhesus glycoprotein transporter family in the adult tissues of A) Danio rerio and B) Alcolapia alcalica. EF1α was used as a normalisation control.

3.4.3 Expression pattern of Rh protein genes in *D. rerio* embryos

In situ hybridisation using DIG labelled cRNA probes for Rh protein genes was used to analyse expression in *D. rerio* embryos at 5dpf (Figure 10). Expression for all four Rh proteins shows strongest expression in gill tissue and the expression of *Rhcg2a* and *Rhcg2b* are restricted to this tissue. *Rhcg2b* has not been detected in previous studies (Nakada, Hoshijima et al. 2007), while we demonstrate high, specific expression in the developing (Figure 10J,K) and adult gill (Figure 9A, 10L). The gene duplication of *Rhcg2* in zebrafish appears to have produced subfunctionalisation of the different isoforms, with *Rhcg2a* expression in the tips of gill lamellae and *Rhcg2b* absent in this area but present in the rest of the lamellae (Figure 10I,L). In addition to expression in gill tissue, *Rhbg* is also expressed in the retina (Figure 10A,B; white arrowhead). The white arrows in Figure 10A and D show the expression of *Rhbg* and *Rhcg1* in the distal part of the pronephric tubule. Expression of *Rhcg1* in adult gill tissue is only detected at the base of gill filaments in mitochondrial rich cells.

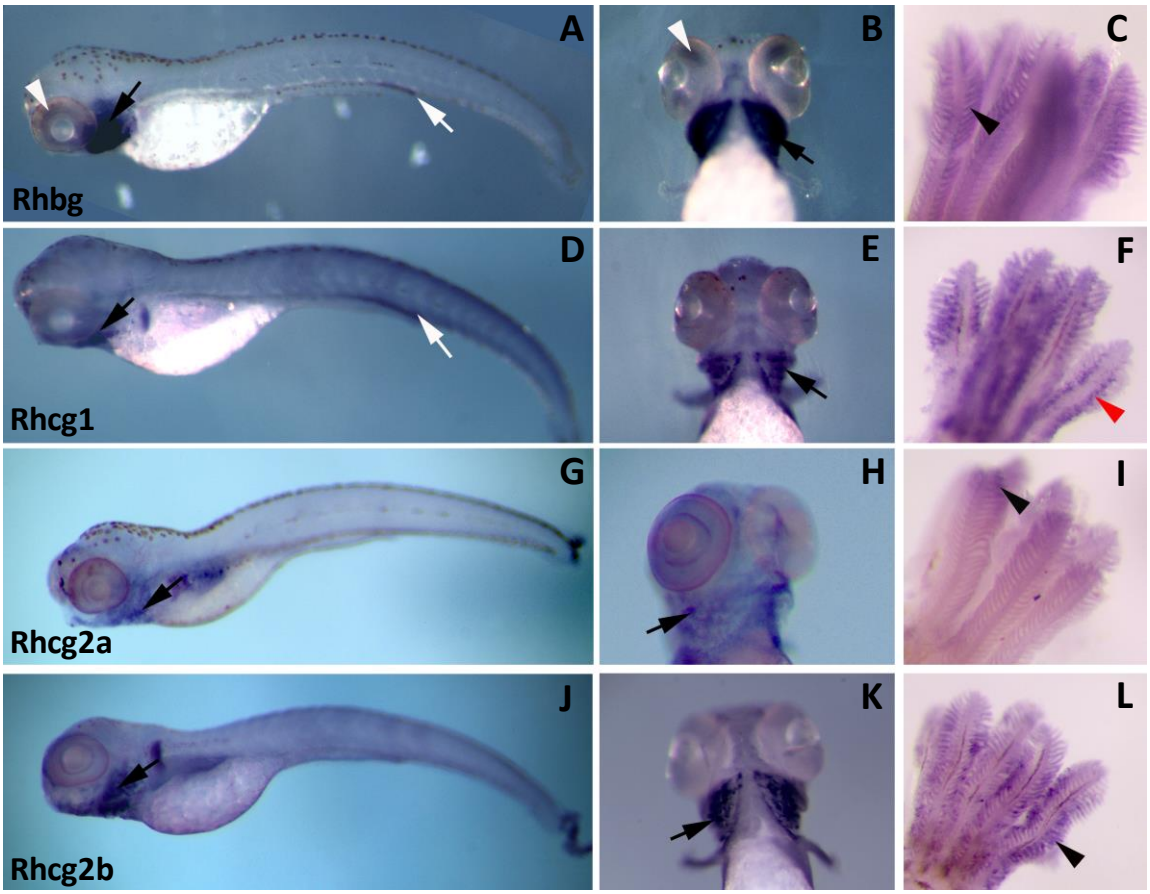


Figure 10: In situ hybridisation of Rhesus glycoproteins (Rh) in Danio rerio embryos at 5dpf and adult gills. A) Rhbg left lateral view B) Rhbg head ventral view C) Rhbg in adult gills D) Rhcg1 left lateral view E) Rhcg1 head ventral view F) Rhcg1 in adult gills G) Rhcg2a left lateral view H) Rhcg2a head ventral view I) Rhcg2a in adult gills J) Rhcg2b left lateral view K) Rhcg2b head ventral view L) Rhcg2b in adult gills. Black arrows indicate expression in gill tissue, white arrows indicate expression in pronephros tissue, black arrowhead indicates expression in gill filament, red arrowhead indicates expression in the chloride cells in gill lamellae and white arrowhead indicates expression in the retina.

3.4.4 Expression pattern of Rh protein genes in *A. alcalica* embryos

In situ hybridisation using DIG labelled cRNA probes for Rh protein genes was used to analyse expression in *A. alcalica* embryos at 5dpf (Figure 11). In comparison to *D. rerio, A. alcalica Rh* have a more restricted expression and are only detected in gill tissue. *Rhbg* has the strongest expression in both adult gill tissue and in the developing embryos (Figure 11A-C). In adults *Rhbg* is detected solely in the gill filament which is made up of pavement cells, much the same as in adult *D. rerio* gill tissue. *Rhcg2* is found in the same region of the gill as *Rhbg* although expressed at a lower level. *Rhcg1,* also restricted to expression in gill tissue, is only detected at the base of gill filaments in mitochondrial rich cells. This is also the case with *D. rerio Rhcg1,* meaning that for all genes analysed cell type expression is conserved within gill tissue.

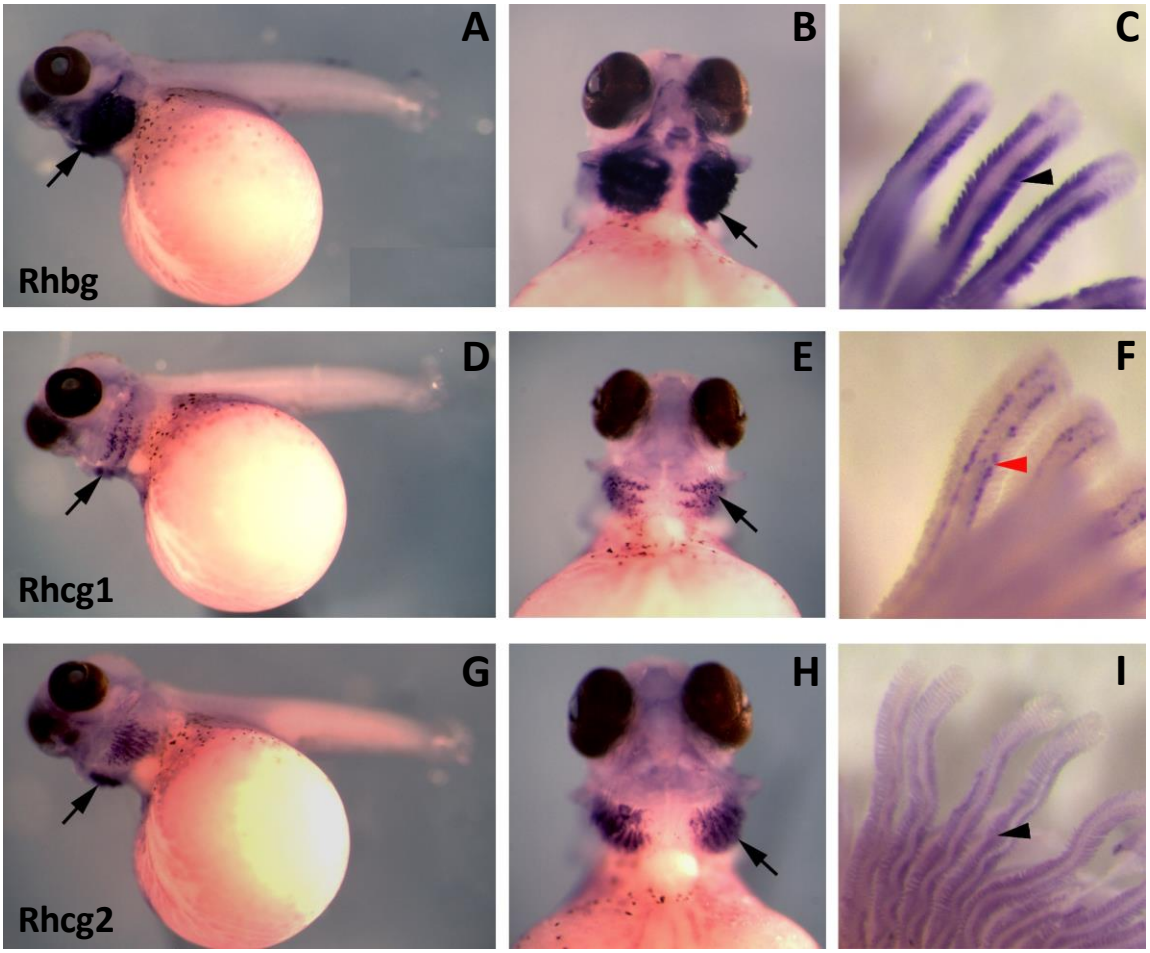


Figure 11: In situ hybridisation of Rhesus glycoproteins (Rh) in Alcolapia alcalica embryos at 5dpf. A) Rhbg left lateral view B) Rhbg head ventral view C) Rhbg in adult gills D) Rhcg1 left lateral view E) Rhcg1 head ventral view F) Rhcg1 in adult gills G) Rhcg2 left lateral view H) Rhcg2 head ventral view I) Rhcg2 in adult gills. Black arrows indicate expression in gill tissue, black arrowhead indicates expression in gill filament and red arrowhead indicates expression in the chloride cells in gill lamellae.

3.4.5 Nitrogen excretion in wildtype *A. alcalica* and *D. rerio*

The percentage of nitrogen excreted as either ammonia or urea was measured in the water holding either adult or embryonic *A. alcalica* or *D. rerio* and was analysed using T test (Figure 12). Both adult and embryonic *A. alcalica* excreted significantly more of their nitrogenous waste as urea than *D. rerio.* Adult *A. alcalica* excreted 64 % (n = 6) of their nitrogenous waste as urea compared to adult *D. rerio* which produced 14 % (n = 14). Embryonic *A. alcalica* and *D. rerio* excreted urea at 79 % (n = 7) and 18 % (n =5) respectively. Table 5 displays the total concentration of ammonia-N and urea-N in µM N $kg^{-1}h^{-1}$ recorded for *A. alcalica* in this study and compares it to published figures for *A. grahami* and *O. niloticus.*

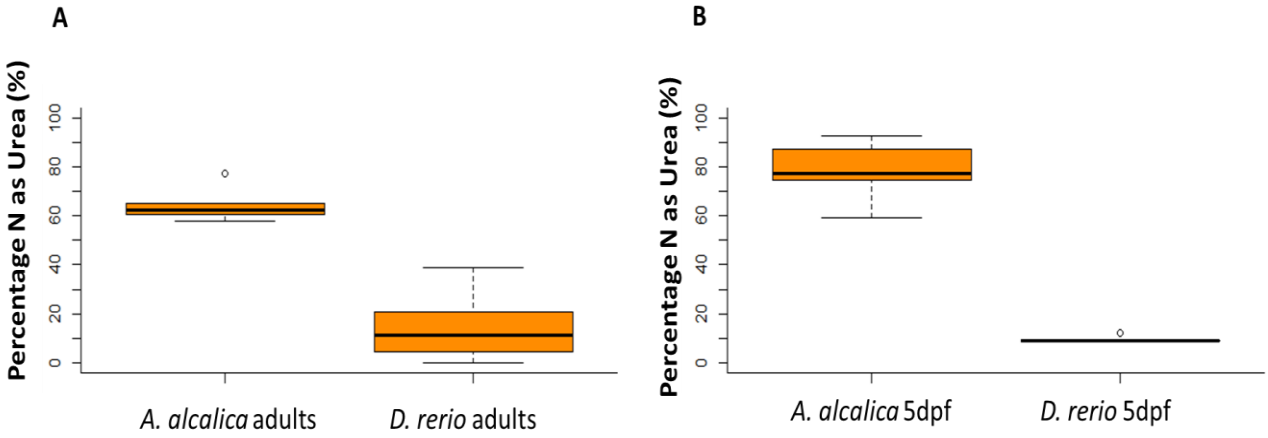


Figure 12: Percentage of nitrogenous waste excreted as urea in wildtype adult (A) and 5dpf embryonic (B) Danio rerio and Alcolapia alcalica. Calculated as a percentage of total nitrogenous waste ([Urea-N/ (Urea-N + Ammonia-N)] x 100). T. tests confirmed differences in amount of excreted ammonia and urea (p-value < 0.0001 for both adult and embryonic comparisons). A. alcalica produce significantly more urea both as adults and embryos when compared to D. rerio.

Table 5: Relative rates of ammonia and urea excretion as products of nitrogenous waste, recorded as Ammonia-N and Urea-N (µM N kg-1h-1) and as percentages of total nitrogen excreted.

Species	pH	Ammonia-N µM N $kg^{-1}h^{-1}$	Urea-N µM N $kg^{-1}h^{-1}$	% N as Ammonia	% N as Urea	Ref
A. alcalica	9	532 (± 145)	957 (± 235)	36	64	This study
A. grahami	10	298 (± 75)	7771 (± 849)	4	96	Wood et al 1989
O. niloticus	8.2	78 (± 5)	18 (± 3)	81	19	Wright 1993
O. niloticus	10	116 (± 26)	146 (± 37)	44	56	Wright 1993

3.4.6 Overexpression and CRISPR targeting of *Rhbg* effects ammonia excretion in *D. rerio* embryos

Over three separate experiments, the effects of overexpressing *D. rerio Rhbg* and *A. alcalica Rhbg* in *D. rerio* and targeting this gene in *D. rerio* with CRISPR-Cas9 were evaluated by measuring the levels of the excretory products, urea and ammonia, in the holding water of 30 pooled experimental embryos. Comparison of these levels to wildtype and analysis using two-sample T-tests showed that there was no effect on urea excretion levels compared to wildtype embryos. Targeting *Rhbg* with CRISPR-Cas9 reduced the levels of ammonia excreted ($t = 4.6559$, $df = 4.7151$, p-value = 0.006432) whereas overexpression of *A. alcalica Rhbg* by mRNA injection increased the amount ammonia detected in holding water ($t = -3.0247$, $df = 3.87$, p-value = 0.04069). There was no effect of overexpressing *D. rerio Rhbg* on ammonia levels detected. Results are shown in figure 13. Generation of a *D. rerio* mutant line lacking *Rhbg* was attempted using CRISPR-Cas9 as previously described. Founders with a 4 base pair deletion were bred to heterozygosity but no homozygous mutant was ever recovered. After many attempts an alternate approach where CRISPR-Cas9 injected embryos were analysed directly for

effects excretory products was used. Sequencing shows the 4 base pair deletion in *Rhbg* proving the sgRNA effective (Figure 14).

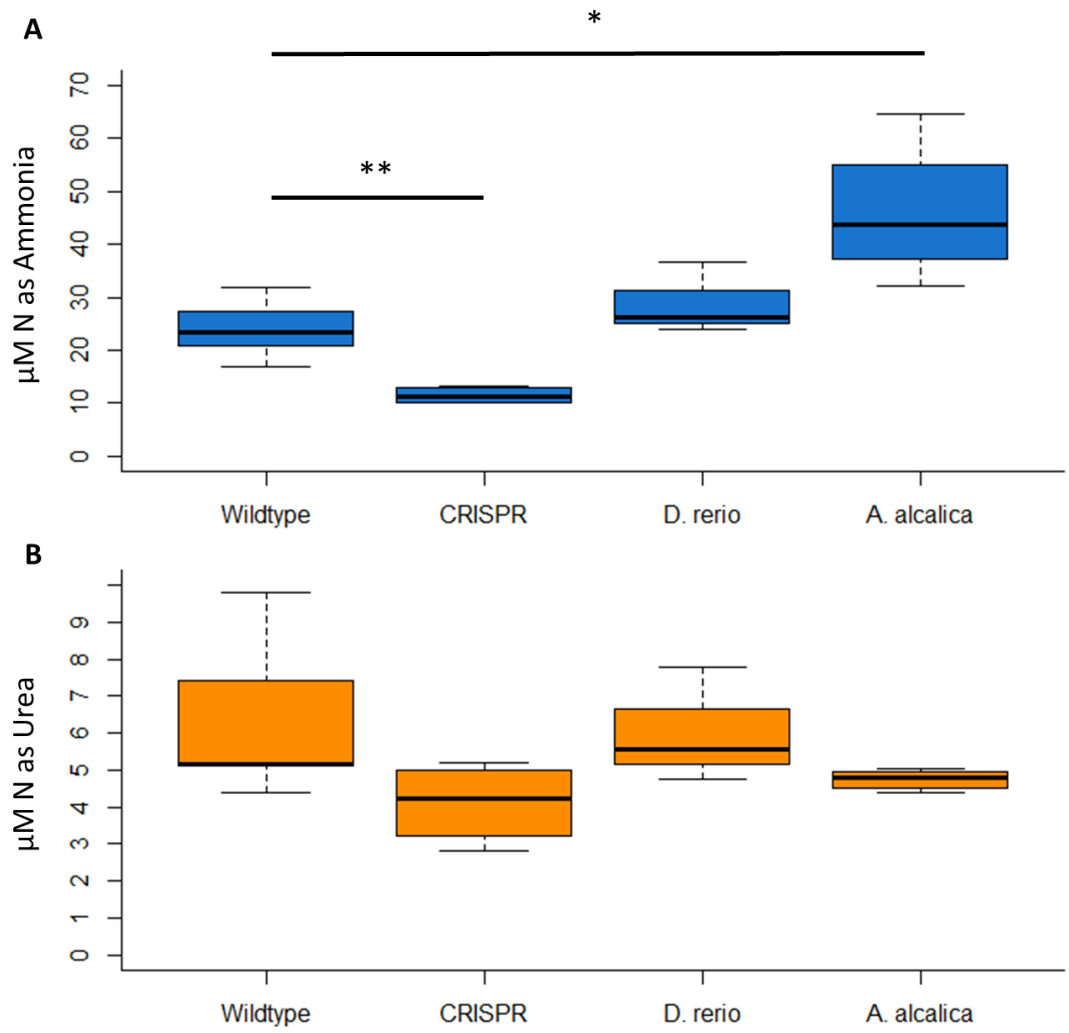


*Figure 13: Effects of CRISPR suppression and overexpression of native D. rerio Rhbg and A. alcalica Rhbg in D. rerio embryos at 1dpf on concertation of A) µM of nitrogen as ammonia and B) µM of nitrogen as urea, compared to wildtype. Groups of 30 embryos per sample. Ammonia and urea concentration represented as µM of nitrogen are shown for wildtype embryos, wildtype injected with CRISPR targeting Rhbg and wildtype injected with either D. rerio Rhbg or A. alcalica Rhbg. Two-sample T-tests were carried out to compare each of the treatments to wildtype, asterisks represent the level of significance compared to wildtype samples (p-values * <0.05, ** < 0.001).*

Wildtype	GAATAAAAACAATTCTACTGATCCTGCCACGAATGAGTTCTACTATCGCTATCCCA
Rhbg Δ4	GAATAAAAACAATTCTACTGATCCTGCCACGA----GTTCTACTATCGCTATCCCA

Figure 14: Sequence alignment of a region of exon one of D. rerio wildtype Rhbg compared to Δ4 deletion produced from CRISPR-Cas9 injection. This mutant line was taken to heterozygosity but no full homozygous individual was ever produced. Underlined region indicates the sgRNA target region. This deletion has incorporated an early stop codon resulting in a protein which would be truncated to just 64 amino acids compared to the 459 amino acids in the wildtype protein.

3.5 Discussion

Alcolapia have been described as a fully ureotelic species and yet retain the genes coding for ammonia transporters. This chapter has shown that *A. alcalica Rhbg* has conserved expression when compared to other fish species and is still functionally active as an ammonia transporter when experimentally examined in model systems.

3.5.1 Phylogenetic analysis of positive selection

Phylogenetic analysis shows that each of the *Alcolapia* Rh proteins cluster within their respective homologous sequences from other cichlid species, corroborating their annotation. This is further supported by gene synteny analysis where the respective *Rh* are shown to be in the same areas of the genome when comparing *O. niloticus* and *D. rerio.* This analysis also shows that *D. rerio Rhcg2b* is syntenic to *O. niloticus Rhcg2* suggesting this is the orthologous *Rhcg2* genes in *D. rerio* rather than *Rhcg2a. Rhag* is found in red blood and usually plays a role in removing ammonia from this cell type, therefore the lack of positive selection shown here likely indicates its conserved function. *Rhcg1* also shows no signs of positive selection. Conversely, the heightened dN/dS ratios for *Rhbg* and *Rhcg2* suggests that within the *Alcolapia* lineage there are a greater number

of mutations than expected resulting in changes to the amino acid sequence. Changes to the amino acid sequence of Rh proteins have been known to change the proteins function. For example, a single point mutation in human Rhag at one of two highly conserved residues leads to a disease known as overhydrated hereditary stomatocytosis. The mutation results in a change to the pore size of the Rhag channel, allowing monovalent cations to leave the red blood cell and causes pathology (Bruce, Guizouarn et al. 2009). Furthermore, changes in dietary potassium levels have been seen to have an effect on levels of *Rhcg* and *Rhbg* in the mouse kidney, suggesting a role for these proteins in potassium homeostasis (Cheval, Duong Van Huyen et al. 2004). The positive selection occurring on *Rhbg* and *Rhcg2* in *Alcolapia* could thus relate to changes in function of these proteins in this lineage, which may be important for their survival in the extreme conditions present in lakes Natron and Magadi.

3.5.2 Conserved expression of *Rh* in the gills of *A. alcalica* and *D. rerio*

In situ hybridisation and PCR methods were used to determine the expression of *Rh* genes in both *D. rerio* and *A. alcalica.* Both analyses show that *A. alcalica* have a more restricted expression when compared to *D. rerio,* with all genes only detected in gill tissues. This gill tissue expression is consistent with not only the *D. rerio* used in this study but also previous *D. rerio* experiments (Nakada, Hoshijima et al. 2007, Zimmer, Wright et al. 2017) as well as the expression seen for the gills of a variety of other species such as rainbow trout (Tsui, Hung et al. 2009), mangrove killifish (Hung, Tsui et al. 2007) and common carp (Sinha, Kapotwe et al. 2016). Further, multiple studies have examined the cellular expression of Rh proteins within gill tissue and have shown that they have a characteristic expression. Rhbg is found on the basolateral membrane while Rhcg2 is on the apical

membrane of pavement cells, whereas Rhcg1 is found on the apical membrane of mitochondrial rich cells (Nakada, Hoshijima et al. 2007, Wright and Wood 2009). For both *A. alcalica* and *D. rerio in situ* hybridisation methods of gill tissue finds that *Rhbg* and *Rhcg2* are expressed in gill filaments and *Rhcg1* at the base of these filaments. Gill filaments are composed of pavement cells and at the base of these filaments are mitochondrial rich cells. This suggests that Rh proteins in *A. alcalica* share the characteristic gill expression seen in other fish species, although this would need to be confirmed with immunohistochemistry with antibodies specific for each of the Rh proteins.

3.5.3 *A. alcalica* excrete a proportion of their nitrogenous waste as ammonia

The proportion of nitrogenous waste excreted as ammonia or urea was measured for adult and embryonic *A. alcalica.* The adults were lab grown and the colony had been acclimated to laboratory conditions for several years. These samples were compared to those from an ammonotelic species, *D. rerio*. This thesis shows that *A. alcalica* excrete the majority of their nitrogenous waste as urea (an average of 64 %) as opposed to the ammonotelic *D. rerio* which excretes the majority (85 %) of its waste as ammonia. Conversely, it has been previously reported that *A. grahami* excrete 100 % of its waste as urea (Randall, Wood et al. 1989). Urea excretion of wild caught *A. grahami* has been shown to be unaffected by short term acclimation to lower pH conditions or to a reduced, N-deficient, diet and the amount of ammonia excreted in these conditions was considered negligible (Randall, Wood et al. 1989, Wood, Perry et al. 1989, De Boeck, Wood et al. 2019). It is possible that the individuals used in other studies were not acclimatised long enough for an effect to be seen on the excretion rates of ammonia or

urea. This study did use a different species of *Alcolapia,* however this is unlikely to be the reason for the discrepancy described here as the adaption to ureotelism occurred prior to the species diverging (see chapter 2). Instead, differences in water conditions and methods of measuring ammonia and urea concentration could possibly explain the observed variation.

Although less than previously reported the amount of nitrogenous waste excreted as urea rather than ammonia by these fish is considerable. Under benign conditions the closely related *O. niloticus* excretes only a small proportion of its nitrogenous waste as urea, around 20 % (Wright 1993). When exposed to elevated pH this is increased to 56 % for the first few hours before again decreasing, suggesting that *O. niloticus* is not capable of sustaining the production of large quantities of urea, which is a more costly method of excretion (Wright 1993, Wright 1995). Even with this increase, overall urea excretion by *O. niloticus* is still lower than the continued urea excretion we see in *A. alcalica.* Maximally *O. niloticus* excreted 146 (± 37) µM N $kg^{-1}h^{-1}$ urea compared to the 957 (± 235) µM N $kg^{-1}h^{-1}$ we see here in *A. alcalica* and the 7771 (± 849) µM N $kg^{-1}h^{-1}$ reported for *A. grahami* (Wood, Perry et al. 1989, Wright 1993). Comparisons between marine and freshwater species has shown that fish rarely use urea as their main excretory product. Combining the data from comparative papers shows that the range of percentage of urea excretion in fish exposed to natural conditions is 1.4 to 24.7 % (Walsh, Wang et al. 2001, Altinok and Grizzle 2004). Overall, cross species comparisons show that the large levels of urea continually excreted by *Alcolapia* is vast compared to other fish although, unlike previously reported for *A. grahami,* it is not the sole excretory product for nitrogenous waste

3.5.4 Expression of *A. alcalica Rhbg* increases movement of ammonia in a fish and yeast model

After cloning the full-length cDNA sequences for Rh protein, a collaboration was established with Prof Gavin Thomas and his PhD student Matthew Rose. The function of *Rhbg* was analysed in two orthologous models: a mutant yeast line missing its ammonia transporter and the embryos of the fish model organism *D. rerio.* In *D. rerio Rhbg* was targeted using CRISPR-Cas9 and the mRNA for *Rhbg* from *D. rerio* and *A. alcalica* was overexpressed. Injection of CRISPR-Cas9 alongside a sgRNA targeting *Rhbg* would create a mosaic effect whereby, within an individual embryo, different cells would be carrying different mutations with some cells not being effected at all (McQueen and Pownall 2017). This mosaicism, rather than causing complete loss of function, creates a knockdown where effected cells produce a truncated mRNA which is broken down through nonsense mediated decay and resulting in overall less of the targeted gene being expressed in the embryo (Billon, Bryant et al. 2017, Hoshijima, Jurynec et al. 2019). Targeting *Rhbg* with CRISPR-Cas9 in *D. rerio* founder embryos causes a reduction in the amount of ammonia excreted. The use of antisense oligonucleotide morpholinos has been shown to successfully knockdown each of the *Rh* genes independently in *D. rerio*. This knockdown reduced the amount of ammonia excreted by ~50 % (Braun, Steele et al. 2009) , similar to the findings of this study. This further supports the role of *Rhbg* in excreting ammonia in fish as well as suggesting that all *Rh* are important for efficient excretion of this waste product (Zimmer, Wright et al. 2017). Overexpression of *D. rerio Rhbg* had no effect on the concentration of excretory products measured however, overexpression of *A. alcalica Rhbg* significantly increased the amount of ammonia excreted by injected embryos. It is possible that the positive selection and the amino acid

mutations in *Alcolapia Rhbg* may have increased its effectiveness at moving ammonia and hence allowing these fish to continue to excrete this waste product even in their high pH environment.

The incorporation of plasmids coding for the different *Rh* from *A. alcalica* and *Rhbg* from *D. rerio* into the yeast *Saccharomyces cerevisiae* triple-mepΔ mutant line, which is lacking its orthologous ammonia transporters, showed interesting results. These data, produced by Matthew Rose, are shown in figure 15. This mutant line has been shown to be unable to grow in low ammonia conditions (Marini, Soussi-Boudekou et al. 1997, Marini, Matassi et al. 2000). Transformation and expression of putative transporters into this strain allows the screening of ammonium transport activity, with an active transporter allowing growth under nitrogen limiting conditions. This experiment showed that *Rhbg* from both species compliment the growth defect in these mutant yeast, indicating they are both functional as ammonia transporters and supporting the result of overexpressing *A. alcalica Rhbg* in *D. rerio* embryos. *Rhcg1* from *A. alcalica* was also able to recover the growth defect suggesting a role in ammonia transportation. Yeast transformed with plasmid for *A. alcalica Rhcg2,* on the other hand, showed similar growth to the triple-mepΔ mutant line suggesting it in not an ammonia transporter. Overexpression of these in *D. rerio* could further confirm that these proteins are not transporting ammonia.

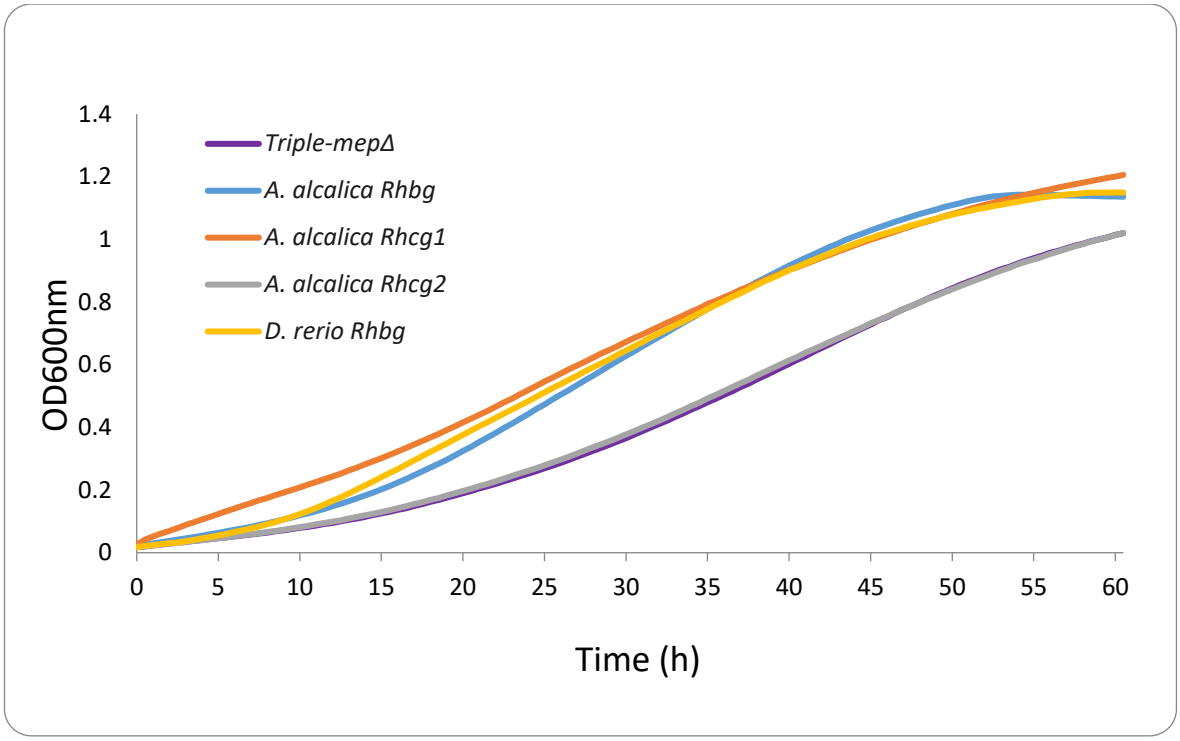


Figure 15: Triple-mepΔ Saccharomyces cerevisiae growth assays on minimal media containing 5mM $(NH_4)_2SO_4$ as a sole nitrogen source. Comparisons of the optical densities (OD600 nm) over time in hours for S. cerevisiae Triple-mepΔ and this mutant line transformed with expression vectors for A. alcalica Rhbg, Rhcg1 and Rhcg2 and D. rerio Rhbg.

3.5.5 Conclusions

Together these results indicate that *Rhbg* in *A. alcalica* is capable of transporting ammonia and potentially the changes to its amino acid sequence have increased its effectiveness at high pH. An efficient method of excretion of nitrogenous waste is crucial due to the toxic effect of high concentrations of ammonia (Mommsen and Walsh 1989). The *A. alcalica* in this study continued to excrete a proportion of its nitrogenous waste as ammonia, unlike previously reported for *A. grahami,* although the majority of its nitrogenous waste was excreted as urea. *Alcolapia* have adaptations to its urea pathway and excrete most of its waste in this way however, it appears that there may also be adaptations to its ammonia transporters to allow some excretion as ammonia in their

high pH environment. Adaptation to allow the movement of ammonia against a concentration gradient is therefore plausible. This study has not fully addressed the function of Rhcg1 and Rhcg2 but results using the triple-mepΔ suggests that Rhcg1 retains ammonia transporter function where as Rhcg2 may not. As *A. alcalica* are capable of excreting ammonia it is likely they share a role in this function. This is further supported by the conserved spatial expression of all Rh protein genes in the gill tissues of *A. alcalica* when compared to other fish species including *D. rerio*.

Chapter Four – Expression analysis of three *Pax* genes involved in eye development in *Alcolapia alcalica* and progress using CRISPR-Cas9 gene editing technology in this non-model organism

4.1 Summary

The main aim of this chapter was to establish CRISPR-Cas9 gene editing technology in *Alcolapia alcalica*. Given a clear prediction of a readily apparent phenotype (loss of eyes), *Pax-6* was chosen as the target gene for experimentation to develop this protocol. These efforts were unsuccessful and did not yield any mutations in *A. alcalica Pax-6* due to the low viability of embryos following microinjection. Nevertheless, this chapter reports an analysis of the wider three genes related to *Pax-6*. Three members of the *Pax-6* gene family were analysed*: Pax-6a, Pax-6b* and *Pax-10.* Phylogenetic analysis and expression patterns are reported for each gene and are consistent with subfunctionalisation.

4.2 Introduction

For any model organism to be a valuable biological tool, an efficient method for gene inhibition is required. No previous attempts at gene editing or inhibition have been made in *Alcolapia* and there are limited molecular studies published in the literature. *A. alcalica* was chosen for this PhD because, of the species of *Alcolapia* available, they produce the largest number of embryos, which would benefit the testing of gene inhibition methods as well as other developmental and molecular biological techniques. However, like many cichlids, *Alcolapia* are a mouthbrooding species and their embryos cannot be artificially fertilised as is done with *Xenopus* nor do they have an environmental cue for mating such as the case with photoresponsive zebrafish. As such, a knowledge of the mating displays of *Alcolapia* was required in order to increase the likelihood of gaining access to recently fertilised embryos. Special mating tanks were created which separated a single male from a group of females via a perforated divider. Removing this divider after a period of separation encourages timely mating and embryo production. The females could then be

physically manipulated to remove embryos from the mouth for use in a variety of experiments.

There are a variety of technologies that have been used to disrupt genes to determine function, these range from post-transcriptional methods to those that target the genome directly. Post-transcriptional methods, such as antisense morpholino oligos, aim to prevent translation or to block splicing by the binding of complementary nucleotide sequences that target RNA and either interrupt translational machinery or bind to splice donor/acceptor sites to prevent splicing of the primary transcript (Summerton and Weller 1997, Draper, Morcos et al. 2001). Alternatively, gene targeting methods act by directly altering the DNA sequence of a gene of choice, resulting in a permanent and heritable effect on the genome rather than a transient one, as is produced with post-transcriptional methods. Gene targeting examples include transcription activator-like effector nuclease (TALEN). TALENs are made up of TALE proteins which are present in bacteria, *Xanthomonas*, where they natively bind DNA and act as a promoter (Wright, Li et al. 2014). When synthetically constructed TALENs are introduced to the cell and attach to a specific target sequence they bind the genomic material. Pairs of TALENs are typically used such that two parts of a catalytic domain are brought together only when both right and left TALENs bind the target sequence , increasing specificity of this technique (Huang, Xiao et al. 2011). Although this technology is very effective at permanently disrupting a gene and is capable of creating full mutant lines in a variety of species it can be costly, time consuming and technically difficult (Khan 2019).

Recent development of an efficient genome editing technology based on the bacteria species *Streptococcus pyogenes* adaptive immune system, clustered regularly interspaced

short palindromic repeats (CRISPR) - associated protein (Cas9), is superseding all other gene inhibition techniques. This is recognised by the 2020 Nobel prize for chemistry being awarded to the researchers behind its discovery; Emmanuelle Charpentier and Jennifer A. Doudna. *S. pyogenes* protects itself from foreign nucleic acid structures, such as bacteriophages, by the production of CRISPR RNA. The bacteria incorporate the foreign phage DNA into its genome after exposure, localising them to a CRISPR region where they can subsequently be transcribed and used to create CRISPR RNA specific to an invading pathogen. CRISPR RNA forms a complex with trans-activating RNA (tracrRNA) and Cas9, the CRISPR RNA in this complex can then bind to the foreign nucleoids, recruiting the Cas9 enzyme. If target sequences contain Proto-spacer Adjacent Motifs (PAMs), the Cas9 is capable of inducing double-stranded breaks in the DNA and inactivating it (Sander and Joung 2014). The study of this system in bacteria led to the development of the CRISPR-Cas9 gene editing tool and protocols have been established for its use in model organisms such as fish and frogs (Hwang, Fu et al. 2013, Nakayama, Blitz et al. 2014). Researchers can target genes of interest via the production of synthetic guide RNA (sgRNA), which incorporates the functions of CRISPR RNA and tracrRNA. Co-injection of this product with Cas9 protein into fish and frog embryos allows for a simple and cheap method of gene editing (McQueen and Pownall 2017, Khan 2019). The sgRNA binds to the target region in the genome, determining where the Cas9 enzyme should cut. The cut in the genome leads to nonhomologous end joining–mediated repair which is error prone and susceptible to the incorporation of insertion/deletion mutations (INDELS). If the change in DNA sequence changes the reading frame it may result in a truncated, non-functional, protein.

This method has been validated in zebrafish, *Danio rerio* (Jao, Wente et al. 2013) and subsequently used to create mutations in the genome of *Oreochromis niloticus* (Li, Yang et al. 2014). Due to this, it is likely that this method should be effective in targeting genes of interest in *Alcolapia* species. A good control to test whether CRISPR-Cas9 conditions are effective is to target tyrosinase, an enzyme essential for pigmentation. If injected embryos are unpigmented then CRISPR conditions are good. *Alcolapia* remain mostly unpigmented during early development with a small pattern of chromatophores showing after 3 dpf therefore an alternative target was required. *Pax-6* was chosen as the target for this chapter as it gives a clear and conserved phenotype when disrupted and CRISPR-Cas9 has been used previously to successfully target this gene in other species (Yasue, Kono et al. 2017, Viet, Reboutier et al. 2019).

Members of the *Pax* gene family are characterised by a 128 amino acid DNA binding region known as the paired box (Bopp, Burri et al. 1986, Neubüser, Koseki et al. 1995, Blake and Ziman 2014). Originally discovered as genes important in segmentation during *Drosophila* development (Bopp, Burri et al. 1986), the *Pax* genes were subsequently shown to have been conserved during the evolution of diverse organisms (*Xenopus;* Heller and Brändli 1999, chicken; Peters, Doll et al. 1995, human; Burri, Tromvoukis et al. 1989, zebrafish; Krauss, Johansen et al. 1991). Transcription factors can either repress or activate gene expression to direct lineage determination during embryogenesis and organogenesis (Chi and Epstein 2002, Blake and Ziman 2014). Known mutations in Pax genes cause defects in development and are associated with a number of congenital diseases in vertebrates (Gruss and Walther 1992, Noll 1993).

4.2.1 Function of *Pax-6* in development

Pax-6 is a highly conserved gene which plays an essential role in eye development throughout the animal kingdom (Gehring 2003). *Pax-6* expression has also been determined as crucial in the development of the central nervous system (CNS), olfactory systems, and pancreas in vertebrates (Simpson and Price 2002), with conserved expression in the eye and CNS in invertebrates (Tomarev, Callaerts et al. 1997). Research into the function of *Pax-6* indicates that this gene regulates common genetic pathways in the development of retinal structures (Pichaud, Treisman et al. 2001) and the overexpression of *Pax-6* from numerous species is capable of inducing the development of ectopic eyes in *Drosophila* (Loosli, Kmita-Cunisse et al. 1996, Glardon, Callaerts et al. 1997) and *Xenopus* (Chow, Altmann et al. 1999). Like other members of the *Pax* gene family, *Pax-6* is a transcriptional regulator known to directly activate a number of structural eye genes, such as keratins and crystallins (Simpson and Price 2002, Cvekl, Yang et al. 2004, Li, Chen et al. 2008). In addition, it also influences the activity of a broad range of other transcription factors that are important for differentiation, cell cycle regulation and cell death (Simpson and Price 2002). Eyes throughout the animal kingdom differ morphologically and developmentally and yet the function of *Pax-6* is still highly conserved across species. Evidence for this is seen by observing the phenotypes of individuals in different species where the *Pax-6* gene has been mutated or deleted. Humans, mice, *Xenopus* and even *Drosophila* with alterations to *Pax-6* show disruption in the formation of the eye and can produce individuals with a reduced or absent eye phenotype (Glaser, Walton et al. 1992, Quinn, West et al. 1996, Kumar and Moses 2001, Nakayama, Fisher et al. 2015). The conserved function seen between animals with such

vastly different eyes like vertebrates and the *Drosophila* is remarkable and is indicative of the ancient origin of this gene.

In the genomes of mammalian species there is a single gene encoding *Pax-6*, however, in many other vertebrates there are multiple genes coding for Pax-6 and Pax-6-like proteins. It is likely that the presence of multiple genes in some species arose from an ancient genome duplication event with specific clades losing the duplicate genes (Ravi, Bhatia et al. 2013). Zebrafish have maintained the duplicated *Pax-6* genes that have subfunctionalised, where each gene shows non-overlapping expression. Zebrafish *Pax-6a* is expressed in the developing forebrain, eye, nasal placode, hindbrain and spinal cord whereas *Pax-6b* has overlapping and non-overlapping expression in the eye, lesser expression in brain, spinal cord as well as expression in the pancreas (Kleinjan, Bancewicz et al. 2008). Interestingly, the combined expression of the paralogues in zebrafish account for total expression of the single *Pax-6* gene in mammals and birds (Nornes, Clarkson et al. 1998). The diverged expression of these genes in zebrafish is thought to be due to changes in the cis-regulatory elements flanking the genes that occurred following the whole genome duplication event (Kleinjan, Bancewicz et al. 2008).

4.2.2 Evolution of *Pax-6* genes in teleost fish

The theorised two rounds of whole genome duplication occurred in the early evolution of the vertebrates, producing thousands of new genes for evolution to act upon and facilitating the evolution of the complex vertebrate body plan (Holland, Garcia-Fernàndez et al. 1994). A further genome duplication event is deduced to have occurred in the ancestry of bony fish and is attributed to making teleosts the most species-rich vertebrate group (Hoegg, Brinkmann et al. 2004). Differential loss, subfunctionalisation or

neofunctionalisation of the genes produced from the teleost genome duplication has produced high diversity amongst species and the raw genetic potential created by this has allowed the teleosts to become successful across many different environments (Volff 2005, Glasauer and Neuhauss 2014). An ancestral gene of Pax6 was duplicated during the two rounds of whole genome duplication which would have produced four related genes in the vertebrate genome. After subsequent subfunctionalisation and loss three genes have remained, Pax-6, Pax-10 and Pax-4 (Ravi, Bhatia et al. 2013, Feiner, Meyer et al. 2014). A further whole genome duplication restricted to teleost fish is responsible for a further Pax-6 gene, and in teleost these are referred to as *Pax-6a* and *Pax-6b* (Feiner, Meyer et al. 2014).

4.2.3 Aims of this chapter

Current interest in extremophile vertebrates is growing because species such as *A. alcalica* are naturally occurring environmental experiments on how organisms may adapt to change and the potential limits of adaptation (Riesch, Tobler et al. 2015). *A. alcalica* has a number of striking adaptations to survive in extreme conditions of temperature, pH, oxygen availability and salinity. In response to the environment, it has been shown that *Alcolapia* have high rates of mutation which have allowed them to adapt rapidly to their changing environment (Wilson, Wood et al. 2004, Pörtner, Schulte et al. 2010). Nevertheless, selection is expected to maintain the function of important developmental genes that are essential for the conserved vertebrate body plan (Slack, Holland et al. 1993). The main aim of this chapter was to determine whether a method of CRISPR-Cas9 gene editing technology could be developed to successfully target genes in *A. alcalica,* as this would support their use as model organism in biological research. *Pax-6* was chosen

as the target gene as it has a clear, conserved phenotype when disrupted. When undertaking the genomic analysis necessary for generating sgRNA, three genes with homology to *Pax-6* were identified. This chapter aims to use sequence analysis and phylogenetics to identify what these genes are before cloning them and characterising their expression patterns by *in situ* hybridisation in *A. alcalica;* a non-model, extremophile vertebrate.

4.3 Methods

4.3.1 Gene annotation and phylogenetic analysis

Three genes orthologous to *Pax-6* were found in the combined genome of the *Alcolapia.* A BLAST search was carried out for each of these genes and a subsequent mass alignment was created on MEGAX (Kumar, Stecher et al. 2018). Several *Pax-4* genes were also included in the alignment. The nucleotide sequences were aligned by codon using the MUSCLE method on MEGAX and phylogenetic analysis conducted on the translated amino acid sequence of the 23 sequences. The phylogenetic relationship was produced by the Jones-Taylor-Thornton (JTT) model and gamma distribution with invariant sites (G+I), determined as the best model for the data through inbuilt model tests. The phylogeny was evaluated by Bootstrap analysis with 100 replicates, sites were excluded if more than 50% of sequences lacked information at a given position. The notations of the different *Pax-6* located in the *Alcolapia* genome were changed to; *Pax-6a, Pax-6b* and *Pax-10* due to the outcome of the phylogenetic analysis and based on the findings by Feiner *et al* 2014 and Ravi *et al* 2019.

4.3.2 Analysis of gene expression using *in situ* hybridisation

RT-PCR was used to determine the expression of the different *Pax-6* RNA in *A. alcalica* embryos as described in section 2.2.1 using the primers shown in Table 6. Gel electrophoresis of the PCR products determined the presence or absence of the different *Pax-6* RNA (n=3). These products were subcloned into pGEM-T Easy and used to create RNA probes for *in situ* hybridisation which was carried out on *A. alcalica* embryos as described in section 2.2.2.

Table 6: Primers used for amplification of cDNA via RT-PCR of the Pax-6 gene family in Alcolapia alcalica, annealing temperatures and predicted product sizes shown.

Gene target	Primer sequence	Annealing temp (˚C)	Product size (bp)
Pax-6a	Forward - CTTGCGACATCTCCAGGATAC	58	719
	Reverse - CCTCCACTTTGCTCTCCTATTT		
Pax-6b	Forward - GGAGCTGAAGGGATGTTTGA	58	689
	Reverse - TGCATGTGTGGAGGAGTATATG		
Pax-10	Forward - TCTAACTGGAGTGGGCCTAA	58	629
	Reverse - GATGGAGGTAACATGCAGGAG		

4.3.3 Gene targeting with CRISPR-Cas9

CRISPR-Cas9 gene editing technology was used to target *Pax-6a* in *A. alcalica*. Pax-6, like all members of the *Pax* gene family, act by binding DNA and activating tissue specific transcription. Therefore, sgRNA were designed to ensure that the DNA binding region, the paired box, was disrupted and the protein functionally deactivated. Two separate sgRNA were designed against the DNA sequence of exon four of the gene using ChopChop

(https://chopchop.rc.fas.harvard.edu/) and the sgRNA template was generated using PCR and transcribed using T7 MEGAshortscript (Ambion). Exon four was chosen as the target due to it being the earliest exon large enough to accommodate multiple sgRNA. Targeting of this exon should lead to a truncated protein with the majority of the Pax-6 functional region, the paired box, being removed (Figure 16). Production of sgRNA and Cas9 protein was done as described in McQueen and Pownall, 2017 using the primers in Table 7.

Figure 16: Schematic representation of the domain structure of Pax-6 with the sequence and location of exon four of the gene shown. The underlined sequence within exon four indicate the guide RNA used for target with CRISPR-Cas9 gene editing technology. The two sets of three nucleotides boxed in blue represent the protospacer adjacent motif (PAM) where the Cas9 enzyme should cut the gene incorporating INDELS (insertion/deletion mutations).

Table 7: Name and sequence information of the primers used to create the sgRNA for targeting Pax-6a in Alcolapia alcalica with CRISPR-Cas9 gene editing technologies. Forward primers include complementary target sequences, shown in italics, as well as the 5' promoter for transcription by T7 RNA Polymerase.

Primer name	Primer sequence 5' – 3'
Pax-6a: exon 4, primer 1 - forward	GCAGCTAATACGACTCACTATAG *GTGTTAATGGCAGACCGCTGC* GTTTTAGAGCTAGAAATA
Pax-6a: exon 4, primer 2 - forward	GCAGCTAATACGACTCACTATAG *GAGTATCCTGGAGATGTCGCA* GTTTTAGAGCTAGAAATA
Universal reverse primer	AAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACG GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

4.3.4 Embryo collection and injection

To gain access to single cell embryos for injection, mixed sex tanks of fish were set up with one dominant male and between four and six sexually mature females. The male was separated from the females by a perforated divider which could be removed when embryos were required. This tank system means that all fish were familiar with each other prior to the divider being removed as well as avoiding the need for any fish to be netted and moved to an unfamiliar tank which can reduce the likelihood of mating occurring. Upon removing the divider fish were monitored for signs of mating. Pre-mating signs include; deep red colouration in the tails of males, the appearance of the females' ovipositor posterior to her anus and coupling of fish which may result in paired movement before egg laying. Upon visual confirmation of mating behaviours being initiated, females were observed for mouthbrooding. After first signs of embryos in the mouth of a female she was left 30 minutes to finish laying and then collected from the main tank by net and moved to a smaller handling tank. To remove embryos, the female

was held in hand and light pressure applied to her mouth while being held in water. Embryos were allowed to fall into the collection tank and the female subsequently returned to her home tank to recover.

Collected embryos were injected into the single blastomere with a 2 nl solution containing the two sgRNA at 300 ng/ml each, 400 mM KCl and 1 ng of purified Cas9 protein. The use of Cas9 protein rather than plasmid or RNA avoided any delay in activity which can occur when relying on translation. A specially designed mould was used to create agar dishes with depressions of varying sizes (ranging from 1.5 mm to 2.5 mm in diameter in 0.25 mm increments) to accommodate different sized embryos. This allowed the embryos to be placed at the correct angle for injection and increased tension so that the microneedle can penetrate the cell more easily (Figure 17). The ovoid embryo was placed with the thin side, where early cell division occurs, facing up for injection. Post injection embryos were individually incubated in 6-well plates in tank water supplemented with methylene blue, kept at 29 °C in a rocking incubator and had their water changed daily.

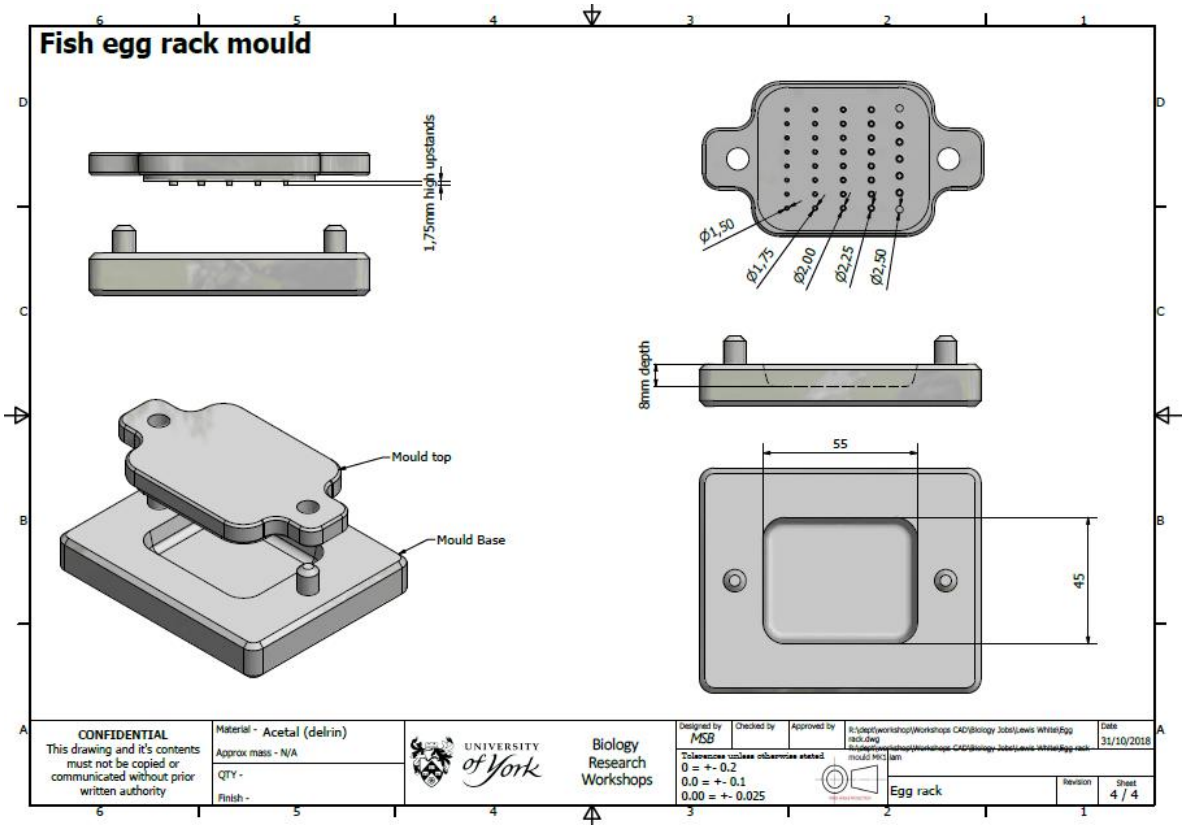


Figure 17: Schematic design of the mould used to make agar dishes, with depressions of varying sizes, to accommodate Alcolapia embryos for microinjection. Mould was 3D printed, designed and made by Mark Bentley, Biology Research Workshop Manager at University of York.

4.3.5 Analysis of injected embryos

Embryos were observed for abnormalities that could be associated with disruption of the *Pax-6a* gene compared to control, non-injected, embryos of the same age and were to be imaged for analysis. To test for successful editing of the gene, genomic DNA was extracted from single embryos and PCR used to amplify the target region. The single embryos were moved into 0.5 ml PCR tubes containing 200 µl of TAD lysis buffer (pH 7, 2.5 ml 1M Tris, 0.5 ml 5 M NaCl, 0.5 ml 0.5 M EDTA, 2.5 ml 10 % SDS made up to 50 ml with dH_2O) containing 0.1 g/ml chelex beads and 10µl 25 µg/µl Protinase K. These were then incubated for 1 hour at 55 °C followed by 15 mins at 95 °C to deactivate the proteinase K. After centrifugation, 1 µl of the supernatant was used for PCR amplification

of the CRISPR-Cas9 target region (Forward primer: CAATCGTGTGAGAGTTTATTGTGAA, Reverse primer: GACTTAAAGCGGTCGTGAAATG). Excess nucleotides and primers were cleaned from the PCR by adding 1 µl of Exonuclease 1 and 1 µl of shrimp alkaline phosphate buffer to 5 µl of PCR and incubated for 15 mins at 37 °C and 15 mins at 55 °C. 5 µl of this was then sent for sequencing along with 5 µl of 5 µM of the aforementioned forward primer. The sequenced regions from all embryos were aligned against wildtype sequence to check for successful disruption.

4.4 Results

4.4.1 Gene annotation and phylogenetic analysis

Phylogenetic analysis of the amino acid sequences of the *Pax* genes extracted from the *Alcolapia* genome (genes submitted to NCBI [accession numbers: MW448161, MW448162, MW448163) form three separate clusters when compared to genes from other species (Figure 18). The largest clade has been denoted as *Pax-6* and includes the single *Pax-6* genes from mammalian and avian species, closest to this are *Pax-6-like* genes which have only been found in some teleost species and here are named *Pax-6b*. As first studied by Feiner *et al* 2014, the next clade contains genes here annotated as *Pax-10* and the analysis conducted here supports their findings in the renaming of *Pax-6.2* as *Pax-10.*

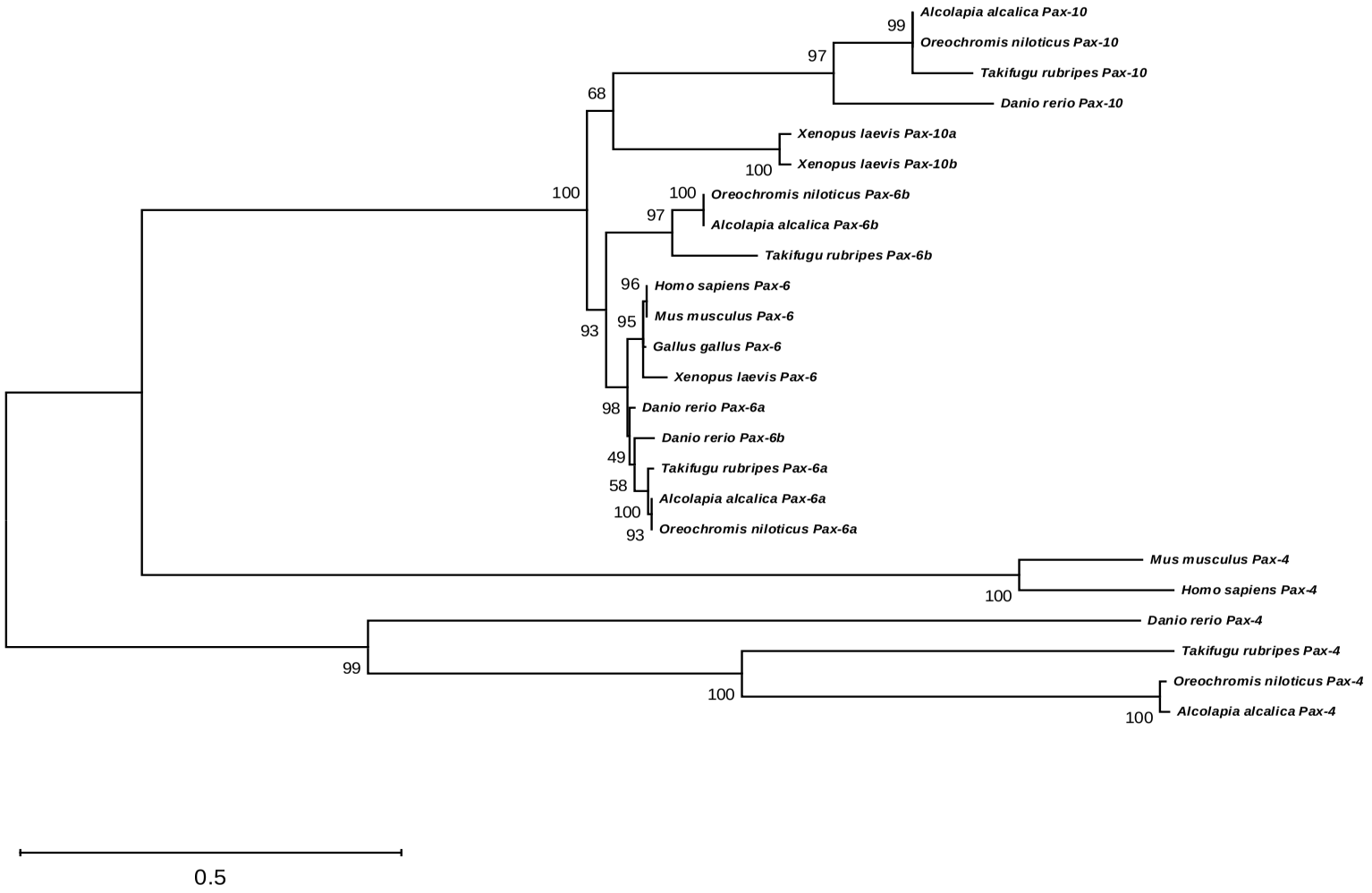


Figure 18: Phylogenetic relationship for the amino acid sequences of the three genes annotated as Pax-6 in the Alcolapia genome. The genes are noted as Pax-6/6a, Pax-6b and Pax-10 depending on which orthologous genes they grouped with. Alignment was carried out on MEGAX using the MUSCLE method and phylogeny constructed using the Jones-Taylor-Thornton (JTT) model and gamma distribution with invariant sites (G+I) with 100 bootstrap replicates (support values indicated beside branches). The scale bar represents a genetic distance of 0.5 amino acid substitutions per site. Phylogenetic trees built from closely related proteins evolving at different rates should not be taken as definitive guides to evolutionary relationships.

4.4.2 Differential expression of *Pax-6a, Pax-6b* and *Pax-10* in *A. alcalica*

Pax-6a, Pax-6b and *Pax-10* were examined by RT-PCR and *in situ* hybridisation methods to determine differences in expression in *A. alcalica* embryos. RT- PCR indicated that all genes were expressed in the developing embryos (Figure 19) which allowed for the production of anti-sense probes used for the *in situ* hybridisation experiments. The three genes examined here had overlapping and distinct spatial expression as well as showing differences in temporal expression, shown by the variation in detected RNA levels which is indicated by the blue colouration in Figures 20, 21 and 22. For both *Pax-6a* (Figure 20) and *Pax-6b* (Figure 21) expression was detected in embryos at 2 and 5 days post fertilisation whereas *Pax-10* (Figure 22) was only detected in the older embryos indicating that it is unlikely to be expressed at earlier stages of development. All three genes have retina expression with *Pax-10* being found exclusively in this tissue and *Pax-6* and *Pax-10* having stronger expression in the retina than *Pax-6b*. *Pax-6a* and *Pax-6b* have overlapping and distinct expression in the developing brain of *A. alcalica,* strongest expression for both is seen in areas of the forebrain with *Pax-6b* also having expression in an area of the hindbrain which could be a region called the locus coeruleus. *Pax-6a* is also detected in the olfactory placode, which gives rise to nasal structures, and in the neural tube.

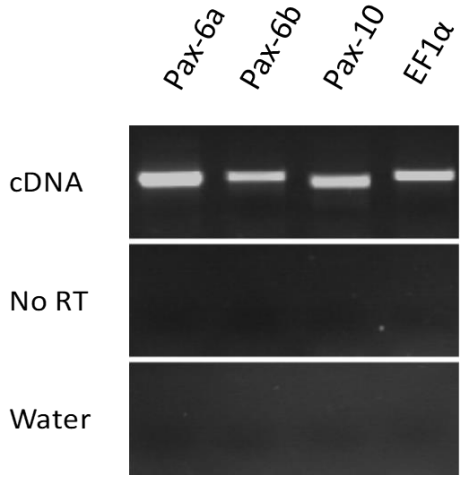


Figure 19: Reverse transcriptase PCR and gel electrophoresis showing the expression of Pax-6a, Pax-6b and Pax-10 in four day old Alcolapia alcalica embryos. EF1α, no RT and water controls shown. Products were used to create RNA probes for in situ hybridisation.

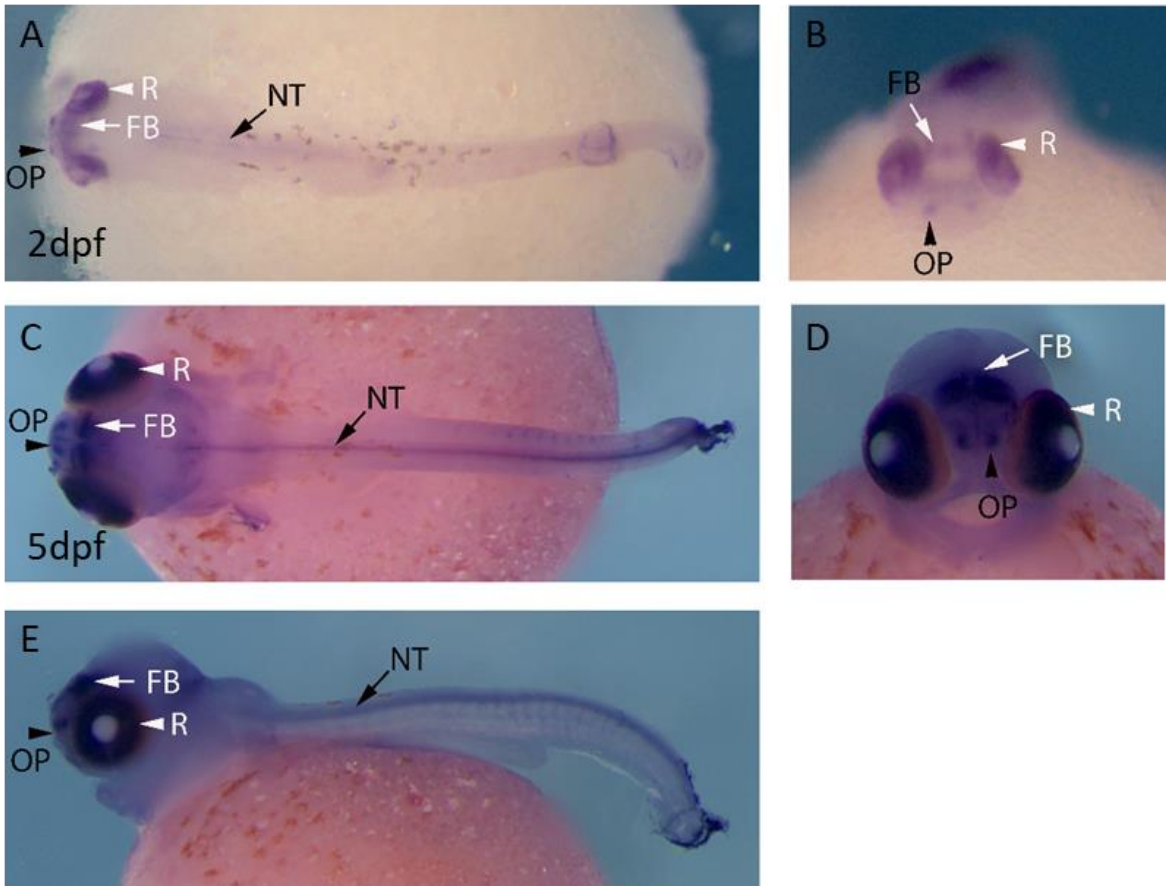


Figure 20: In situ hybridisation showing the expression of Pax-6a in the developing embryos of Alcolapia alcalica at different stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of Pax-6a mRNA. Black dots around the yolk and on the body are chromatophores (pigment cells). A-B) dorsal and face view of 2dpf embryos, C-E) dorsal, face and lateral view of 5dpf embryos. Expression is detected in retina (R), forebrain (FB), olfactory placode (OP) and neural tube (NT).

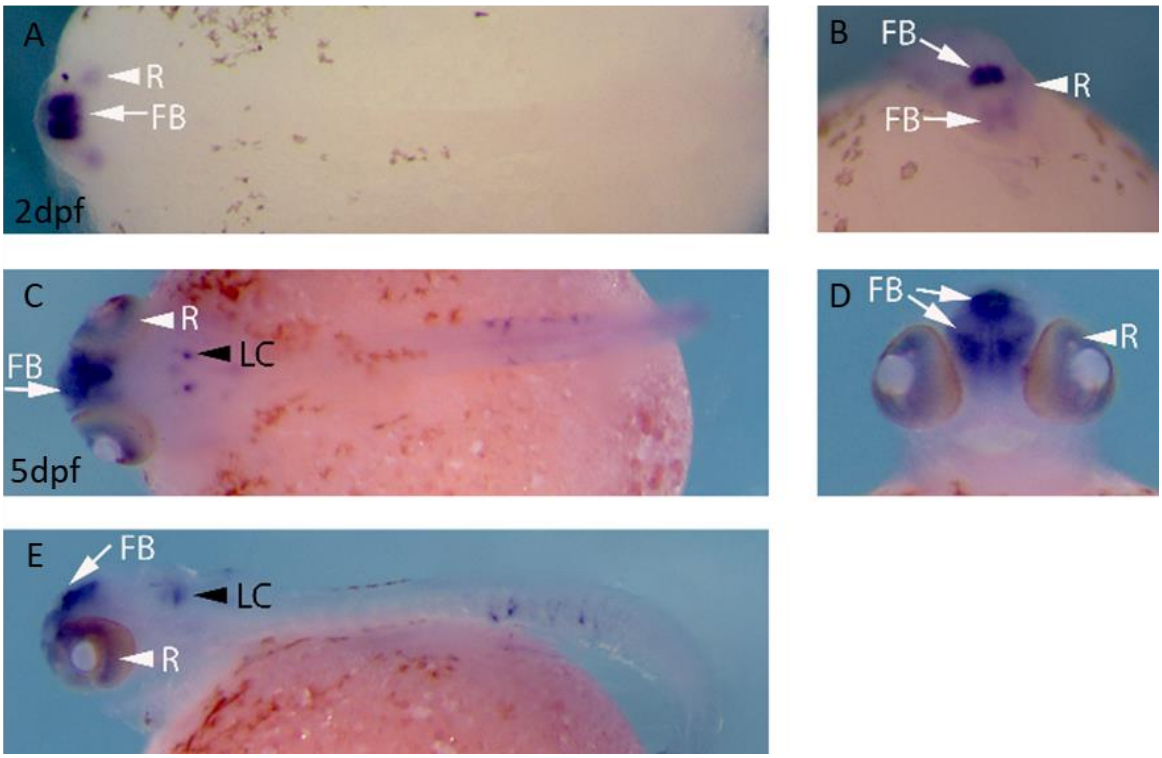


Figure 21: In situ hybridisation showing the expression of Pax-6b in the developing embryos of Alcolapia alcalica at different stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of Pax-6b mRNA. Black dots around the yolk and on the body are chromatophores (pigment cells). A-B) dorsal and face view of 2dpf embryos, C-E) dorsal, face and lateral view of 5dpf embryos. Expression is detected in retina (R), multiple areas of the forebrain (FB) and an area of the hindbrain called the locus coeruleus (LC).

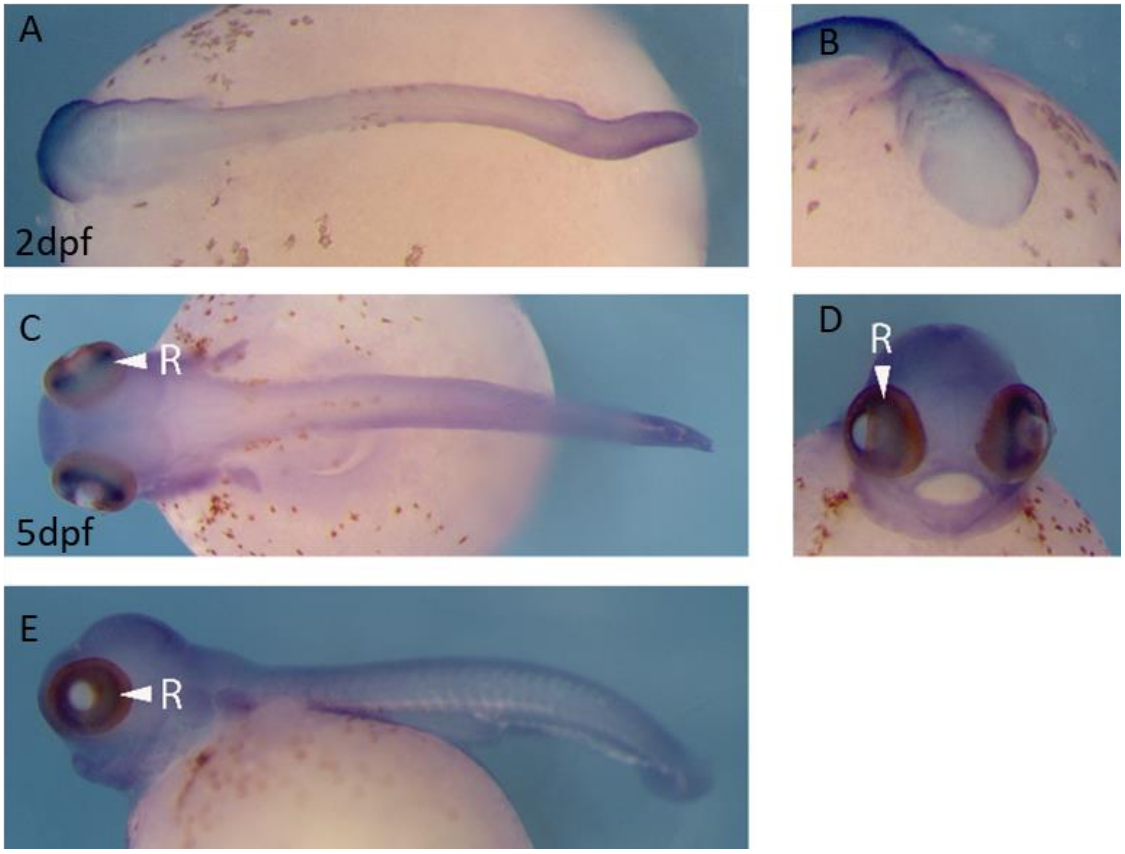


Figure 22: In situ hybridisation showing the expression of Pax-10 in the developing embryos of Alcolapia alcalica at different stages (number of days post fertilisation [dpf] indicated). The blue colour indicates the detection of Pax-10 mRNA. Black dots around the yolk and on the body are chromatophores (pigment cells). A-B) dorsal and face view of 2dpf embryos, C-E) dorsal, face and lateral view of 5dpf embryos. Expression is only detected in the retina (R) and is not detected at all in the embryos at 2dpf.

4.5 Discussion

For a model organism to be valuable in biological research an efficient method for gene editing is required. To begin the development of CRISPR-Cas9 in *A. alcalica, Pax-6a* was chosen as a target gene that would give a clear and predictable phenotype when disrupted. *Pax-6* mutation leads to a loss of eyes across the animal kingdom and targeting this gene with CRISPR-Cas9 technology has been successful in mice and *Xenopus* (Yasue, Kono et al. 2017, Viet, Reboutier et al. 2019). This chapter describes the progress towards identifying, cloning and characterising the expression patterns of *Pax* genes involved in eye development in *A. alcalica* and details the use of CRISPR-Cas9 gene editing technology in this non-model organism.

4.5.1 Differences in the spatiotemporal expression of three genes in the *Pax-6* family in *A. alcalica*

The three genes examined here were all found to be expressed in early stage *A. alcalica* embryos. Phylogenetic analysis show that the genes formed three separate clusters and are here called *Pax-6a, Pax-6b* and *Pax-10.* This is based on the similarity of each gene to genes in other species and in accordance with the nomenclature used by Feiner, Meyer et al. 2014. All genes showed both distinct and overlapping spatial and temporal expression patterns in developing embryos which is suggestive of some subfunctionalisation in their roles (Lynch and Conery 2000).

The *Pax-6b* gene is believed to have resulted from the teleost specific genome duplication event (Feiner, Meyer et al.2004) and shares higher sequence similarity to *Pax-6 and Pax-6a* than does *Pax-10.* The subfunctionalisation of the zebrafish *Pax-6* genes differs from that shown in *A. alcalica. Pax-6a* and *Pax-6b* in *A. alcalica* have similar overall expression

to the combined expression of *Pax-6a* and *Pax-6b* in zebrafish, although the expression of the individual genes differs to achieve this combined expression (Kleinjan, Bancewicz et al. 2008). This combined expression of *Pax-6a* and *Pax-6b* in *A. alcalica* and *Pax-6a* and *Pax-6b* in *D. rerio* matches the expression of the single *Pax-6* gene found in mammals and birds (Nornes, Clarkson et al. 1998). Notable differences are the absence of pancreatic expression in either of the *Pax-6* genes in *A. alcalica* which is seen in other species, including zebrafish, mice and chicken (Ahnfelt-Rønne, Hald et al. 2007, Kleinjan, Bancewicz et al. 2008, Mitchell, Nguyen-Tu et al. 2017), as well as the novel expression of *A. alcalica Pax-6b* in an area of the hindbrain in the region of the locus coeruleus. A complete absence of *Pax-6* expression in the pancreas results in a decrease in the number of islets, abnormal pancreatic morphology and causes early onset diabetes (Sander, Neubüser et al. 1997, Ashery-Padan, Zhou et al. 2004). With this in mind, it is unlikely that the role *Pax-6* plays in pancreas development is not being completed in some way in the *Alcolapia.* Pancreatic expression in this species may be lower than elsewhere and as such *in situ* hybridisation is not sensitive enough to detect it, other techniques such as qPCR may be better placed to determine whether in fact *Alcolapia* have retained *Pax-6* expression in pancreatic tissues. Alternatively, the timing of pancreatic development in *A. alcalica* may not have been captured in the stages examined in this study.

4.5.2 Novel expression of a *Pax-6* gene in the locus coeruleus

A. alcalica Pax-6b was detected in an area of the hindbrain likely to be the locus coeruleus. *Pax-6* expression has not previously been reported in the locus coeruleus across a broad range species examined for the genes activity (Kleinjan, Bancewicz et al. 2008, Duan, Fu et al. 2013, López, Morona et al. 2020). The locus coeruleus is a small

cluster of noradrenergic cells in the hindbrain that controls the noradrenaline released into the brain and as such plays a major role in alertness (including that of the major senses), amongst other functions (Sara 2009, Schwarz and Luo 2015, Benarroch 2018). Although *Pax-6* expression has not previously been reported in the locus coeruleus it has been detected in this region. The locus coeruleus expresses considerable levels of tyrosine hydroxylase (TH) which converts tyrosine into the neurotransmitter dopamine that is subsequently converted to noradrenaline and adrenaline. *Pax-6,* although not expressed in TH positive cells in the locus coeruleus, has been shown to induce the expression of TH in neighbouring cells. Its inductive ability is supported by its close spatiotemporal but not overlapping expression pattern with TH positive cells in this region (Vitalis, Cases et al. 2000, Wullimann and Rink 2001, Duan, Fu et al. 2013). To determine whether *Pax-6b* RNA is in the locus coeruleus of *A. alcalica* embryos an *in situ* hybridisation double labelling for *TH* and *Pax-6b* would be helpful or the use of immunostaining with antibodies specific to these proteins could give better resolution. It is possible that the teleost specific genome duplication event which produced the additional *Pax-6* in some fish species allowed relaxation of the duplicates function and lead to its expression in a novel cell type.

4.5.3 Pax-10 expression occurs later in development

Pax-10 has the most restricted expression both spatially and temporally. It was only detected in the retina of the developing *A. alcalica* embryo and only after 4dpf. This is in accordance with the restricted expression reported in zebrafish, *Xenopus* and anole lizard (Feiner, Meyer et al. 2014). It is believed that *Pax-10* has a specific role in eye growth, after the significant roles of eye specification is carried out by *Pax-6a* and *Pax-6b;* this

could explain the late onset *Pax-10* expression in the retina (Ravi, Bhatia et al. 2013, Feiner, Meyer et al. 2014). Its presence in the genomes of both fish and tetrapods suggests that it is a product of an early whole genome duplication event in the base of the vertebrates and its restricted and similar expression to *Pax-4* could be the cause of its subsequent loss in many lineages (Feiner, Meyer et al. 2014).

4.5.4 Progress in the use of CRISPR-Cas9 gene editing

Multiple attempts at injecting numerous different broods of *A. alcalica* embryos resulted in most embryos not surviving long enough to determine whether targeting of *Pax-6a* with CRISPR-Cas9 was effective. Of the few embryos which survived long enough to have DNA extracted and sequenced around the target region all were found to be the same as wildtype. Observation suggests that early stage *A. alcalica* embryos do not survive well out of the mothers mouth with the majority of uninjected control embryos also perishing. Other control injected embryos such as water injected and those injected with mRNA for GFP also did not survive. Additionally, the embryos appear to be very fragile with the act of injection alone potentially being the cause of loss. If the microneedle at any point punctured the yolk, the embryo would lose integrity. Early removal of the embryo from the chorion was also largely fatal and gentle handling with the use of a hair-loop preferred to forceps.

For a model organism to be a valuable tool in biological research an effective method of gene editing is considered essential. This study has been unable to determine whether one such tool, CRISPR-Cas9, is capable of editing the genome of *A. alcalica.* CRISPR-Cas9 has been shown to be effective in editing the genomes of bacteria (Hegde, Nilyanimit et al. 2019), plants (Bortesi and Fischer 2015), mammals (Hunker, Soden et al. 2020) and fish

(Datsomor, Olsen et al. 2019, Chaudhary, Singh et al. 2020); including a close relative of the *Alcolapia, O. niloticus* (Li, Yang et al. 2014). With this in mind, it is unlikely that CRISPR would not be successful in editing the genome of *A. alcalica* if an effective method of administration could be found. From this study it would appear that the embryos of these fish are fragile at early life stages when removed from the mouth of the mother and are often vulnerable after microinjection. Electroporation has been used to introduce DNA, RNA, Cas9 and sgRNA into the developing embryos of zebrafish, with the use of CRISPR-Cas9 successfully creating a non-pigmented phenotype by targeting a solute carrier which aids in the activity of tyrosinase in melanosomes (Zhang, Ren et al. 2020). The authors of the paper state how they hope the evidence they provide in zebrafish in the use of electroporation for application of CRISPR-Cas9 could be used to develop techniques in other fish species whose embryos do not lend themselves to microinjection. Although this chapter was unable to determine whether CRISPR-Cas9 is effective in *A. alcalica* embryos by microinjection, techniques were developed to culture embryos externally for other experiments, including *in situ* hybridisation, and animal husbandry was improved to allow access to early stage embryos.

4.5.5 Conclusions

This chapter aimed to target the *Pax-6a* gene in *A. alcalica* embryos using CRISPR-Cas9 gene editing technology. Embryos were found to lose integrity upon microinjection and none survived long enough to determine whether the use of CRISPR-Cas9 resulted in mutations in the target gene. Although this is the case, this work improved our understanding of *Alcolapia* husbandry and allowed access to early stage embryos for other biological techniques. Additionally, three members of the *Pax-6* gene family, *Pax-*

6a, Pax-6b and *Pax-10*, were successfully cloned and their spatiotemporal expression patterns determined using *in situ* hybridisation. *Pax-6* and *Pax-6-like* show subfunctionalisation of their roles in the development of the CNS and *Pax-6b* is shown to be expressed in the locus coeruleus, potentially unusual for a *Pax-6* gene. Further, *Pax-10* is shown to have restricted expression to the retina in older embryos, confirming previous studies on the gene in other species.

Chapter Five – Muscle development in an extremophile fish species; *Alcolapia alcalica*

5.1 Summary

Many teleost fish retain two *MyoD* genes in their genome; *MyoD1* and *MyoD2.* Mass sequence alignments showed that a number of teleost species also contain a poly-serine region in their MyoD1 which has not previously been studied. This chapter aimed to explore the evolutionary history of these *MyoD* genes in teleost fish using phylogenetics. Analysis suggested that there was a link between retention of *MyoD2* and the incorporation of a poly-serine region in MyoD1. Further, this chapter examined the expression patterns of these genes in *A. alcalica* showing evidence of subfunctionalisation. Preliminary functional analysis suggests that the poly-serine region may increase the protein stability of MyoD1.

5.2 Introduction

5.2.1 Genome duplication and the evolution of myogenic regulatory factors

Although genome duplication events are considered rare in animal lineages, they produce a substantial amount of genetic material for evolution to act upon. The ancestral vertebrate genome is believed to have been shaped by two rounds of whole genome duplication. Subsequent evolution on these duplicated genes led to great genetic novelty and contributed to the development of the complex vertebrate body plan (Holland, Garcia-Fernàndez et al. 1994, Meyer and Schartl 1999, Onai, Irie et al. 2014, MacKintosh and Ferrier 2017). A third whole genome duplication event occurred at the base of the teleost lineage; this is cited as a cause behind the high diversity seen amongst these fishes (Hoegg, Brinkmann et al. 2004).

Gene duplicates resulting from a whole genome duplication event are referred to as ohnologs after Susumu Ohno who first proposed multiple genome duplications in vertebrates (Ohno 1970, Singh and Isambert 2020). Mutations in either the coding region or regulatory regions of ohnologs can lead to nonfunctionalisation, subfunctionalisation or neofunctionalisation of the new genes (Wagner 1998, MacKintosh and Ferrier 2017). For example, the vertebrate whole genome duplications created eight myogenic regulatory factor (MRF) ohnologs from a single ancestral pair. Subsequent loss of some genes and subfunctionalisation of others produced the differing number of MRFs required for skeletal muscle development in vertebrate species (Aase-Remedios, Coll-Lladó et al. 2020).

MRFs are a family of bHLH (basic helix-loop-helix) transcription factors with overlapping yet essential roles in myogenesis. Conserved across vertebrates, these important regulatory genes fall into two different functional groups (Aase-Remedios, Coll-Lladó et al. 2020) and their expression during embryonic development leads to the specification and differentiation of muscle cells (Pownall, Gustafsson et al. 2002). MyoD and Myf5 act early in development and have function in muscle cell determination. These early expressed genes subsequently activate the later onset MRFs, MyoG and Myf6, which are more involved in differentiation (Pownall, Gustafsson et al. 2002, Asfour, Allouh et al. 2018). These genes, while restricted to the skeletal lineage, exhibit expression patterns suggestive of subfunctionalisation having occurred following their duplication (Aase-Remedios, Coll-Lladó et al. 2020). The third whole genome duplication event, at the base of the teleost lineage, was responsible for a further MRF. Although lost in some lineages, including the branch leading to *D. rerio,* MyoD2 is present in the genomes of a large number of teleost fish (Macqueen and Johnston 2008). Some analysis has been carried

out on the two MyoD paralogues in teleost species however, similar to the other MRF duplicates, it has been shown that they have both overlapping and distinct expression (Tan and Du 2002, Andersen, Dahle et al. 2009). These studies showed that in both Atlantic halibut (*Hippoglossus hippoglossus*) and gilthead seabream (*Sparus aurata*) *MyoD2* had a more restricted expression than *MyoD1* with RNA being detected in fewer somites. In addition to this, the Atlantic halibut, a flatfish, showed asymmetrical expression for *MyoD2* along the body (Tan and Du 2002, Andersen, Dahle et al. 2009).

5.2.2 Aims of this Chapter

The genomes of *Alcolapia* species have retained both MyoD paralogues, as would be expected by their phylogenetic positioning (Macqueen and Johnston 2008). With the limited research on the differences between *MyoD1* and *MyoD2* genes, this chapter aimed to determine their expression patterns in *A. alcalica* using *in situ* hybridisation. Upon comparison of the MyoD1 and MyoD2 protein sequences from a number of teleost species, it was noted that *Alcolapia,* along with some other teleosts, have incorporated a poly-serine repeat between the amino terminal transactivation domains (TAD) and the cysteine-histidine rich region (H/C) in MyoD1. As such, an additional aim of this chapter is to examine the evolutionary history of *MyoD1* and *MyoD2* and compare that to the presence of the here-named poly-serine region using phylogenetics. Further, preliminary studies were carried out in an attempt to discern any functional impact of this poly-serine region to MyoD1.

5.3 Methods

5.3.1 Sequence and phylogenetic analysis of MyoD genes in teleosts

The Myod1 sequence for *A. alcalica* was extracted from the constructed genome and aligned to the Myod1 sequences of *O. niloticus* and *D. rerio*. Visual examination of the amino acid alignment identified a poly-serine region present in *A. alcalica* and *O. niloticus* but absent in *D. rerio.* A subsequent mass alignment identified this region in a large number of teleost fish. The genomes of some teleost fish also contain an additional MyoD gene, MyoD2, and this was also included in the sequence analysis. A gene tree was constructed to determine the relationship of teleost MyoD genes to one another. A species phylogeny was used to infer when the evolution of the poly-serine insertion, found in some species MyoD1, may have occurred and which clades have gained or retained the MyoD2 gene. *A. alcalica MyoD1* and *MyoD2* were submitted to NCBI (accession numbers: MW448164, MW448165).

An amino acid sequence alignment containing 97 genes (59 MyoD1 [34 with a poly-serine insertion] and 38 MyoD2) from 54 teleost species was constructed. All sequences were downloaded from ENSEMBL. The phylogeny was produced from this alignment using MEGAX (Kumar, Stecher et al. 2018) and the model Jones-Taylor-Thornton w/freq and gamma distribution, which was determined to be the best fit for the data based on the programmes model testing software. Bootstrap analysis using 100 repeats was used to evaluate the resulting phylogeny.

The published teleost phylogeny (Betancur-R, Wiley et al. 2017) was used to construct a best accepted species phylogeny for all fish species which had sequence data available for MyoD and whose position in the mass teleost phylogeny could be inferred. The R

packages phytools, ape, and geiger were used to prune the phylogeny so only these species were included in the final tree. The final tree contained 43 fish species and outgroups of the frog, *Xenopus tropicalis,* and coelacanth, *Latimeria chalumnae*. Comparison of this tree with the sequence data for MyoD1 and MyoD2 in teleost fish revealed in which clades the poly-serine insertion in MyoD1 had occurred and which clades had both a MyoD1 and Myod2 gene.

5.3.2 Spatial expression of *MyoD* genes in *A. alcalica*

To determine the spatial expression of genes important in the development of muscle in *A. alcalica*, *in situ* hybridisation was carried out against the RNA of *MyoD1* and *MyoD2* with antisense probes designed against the mRNA of these genes. RNA was extracted from whole embryos and used to make cDNA, which was then amplified by PCR as in section 2.2.1 using the primers in table 8 and subsequently used to produce probes for *in situ* as in section 2.2.2. *In situ* hybridisation was carried out as in section 2.2.2 on *A. alcalica* embryos at multiple stages for both *MyoD1* and *MyoD2* to determine any differences in spatial expression over development.

Table 8: Primers used for amplification of cDNA via RT-PCR of MyoD1 and MyoD2 in Alcolapia alcalica, annealing temperatures and predicted product sizes shown. Products used to make antisense RNA probes for in situ hybridisation.

Gene target	Primer sequence	Annealing temp (℃)	Product size (bp)
MyoD1	Forward – CCCACCGCTGATGATTTCTAT	58	603
	Reverse – GCTTCGTCTTCTGGTTGTCT		
MyoD2	Forward – AGAGAGGAATTCATGGATCTGTCAGACTTTCCCTTCGTT	58	792 (full length)
	Reverse – AGAGAGTCAGAAGCATAATCTGGAACATCATATGGATAACAAA TGTTGCTGGACTCGGCAGACAG		

5.3.3 **Protein expression in *Xenopus***

Full length cDNA of *MyoD1* from *D. rerio* and *A. alcalica* was amplified by PCR from cDNA produced from the RNA extracted from fish embryos as in section 2.2.1. Primers used to produce these constructs are shown in table 9 and include sequence to incorporate HA-tags to the 3' tail of each amplicon. *MyoD2* construct was produced in the same way using the primers from table 8 but was not used for this study. PCR products were ligated into the plasmid $CS2^{+}$ and transformed into *E.coli* strain DH5α. Transformations were plated onto LB agar plates containing ampicillin (100 μg ml^{-1}) and grown overnight at 37°C. A single colony was cultured in 3ml of LB medium at 37°C overnight at 180rpm and Qiagen miniprep kits used to extract the cultured plasmids. Plasmids were then sequenced using the $CS2^{+}$ primer SP6 to ensure the nucleotide sequence was correct and in the forward orientation. These plasmids were then linearised by restriction enzyme digest and

cleaned of excess enzyme and nucleotides by phenol chloroform extraction and ethanol precipitation. mRNA was produced from all plasmids using SP6 mMESSAGE mMACHINE® kits (Invitrogen) to manufacturers guidelines, purified by adding 1 µl of Turbo DNase and incubating at 37 ℃ for 15 minutes and cleaned by phenol chloroform extraction and isopropanol precipitation with the addition of ammonium acetate. After desiccation mRNA was re-suspended in 20 µl of dH_2O and concentration measured using a spectrophotometer (nanodrop).

Table 9: Primers used for the amplification of cDNA via RT-PCR of full length MyoD1 from Alcolapia alcalica and Danio rerio. These products were ligated into expression plasmid $CS2^+$ for producing mRNA.

Gene target	Primer sequence with incorporated HA tag (underlined)
MyoD1 – A. alcalica	Forward – ATGGAGTTGTCGGATATCTC
	Reverse – AGCATAATCTGGAACATCATATGGATATAGGACTTGGTAAATCAGG
MyoD1 – D. rerio	Forward – ATGGAGTTGTCGGATATCC
	Reverse - AGCATAATCTGGAACATCATATGGATAAAGCACTTGATAAATGGTTTCCTG

All mRNA was diluted to 80 ng μl^{-1} with dH_2O and 1.6 ng injected into the embryos of *Xenopus laevis* at the one cell stage using a micro-injector. Embryos were fertilised and then cultured in NAM/10 (Normal Amphibian Medium) at 14-15 ˚C in Petri dishes lined with 1.5 % agarose. 2.5 % L-cysteine hydrochloride monohydrate (Sigma) pH 7.8 was used to clean jelly from embryos prior to injection. Post-injection embryos were cultured again

in Petri dish until they reached stage 8-9, at which point the 'animal cap' was removed from the animal pole using mounted tungsten needles. The caps were rested in NAM/2 for an hour after the dissections and a subset of samples collected to account for time zero while the rest were moved into a solution of 10 µg/ml cycloheximide in NAM/2 to inhibit further translation of the injected RNA. Control animal caps from un-injected individuals were also collected at time zero. Samples were then collected at one and three hour time points to test for the stability of the different MyoD proteins over time. Samples consisted of 10 caps which were transferred directly to 1X volume 2X sample buffer, pipette homogenised and heated to 90℃ for 2 minutes before being flash-frozen in liquid nitrogen. All collected samples were stored at -80℃ for future analysis. A number of whole embryos were also collected prior to dissection for each sample, these were fixed in MEMFA and stored in methanol at -20 °C for use in immunohistochemistry.

5.3.4 Immunohistochemistry of injected *Xenopus* embryos

Immunostaining was carried out on whole *X. laevis* embryos overexpressed with the different MyoD constructs. Embryos were rehydrated and incubated for 40 mins in potassium dichromate (7.35 g potassium chromate, 12.5 ml acetic acid, made up to 250 ml with dH_2O). After extensive washing embryos were permeabilised in 5 % H_2O_2 in PBS for 45 mins followed by further washes. Embryos were then placed in 1.5 ml BBT (20 ml 10x PBS, 2 g BSA, 1 ml 20 % Triton X-100, made up to 200 ml with dH_2O) for an hour and then solution replaced with more BBT for a further hour. Block was then carried out for 1 hour in 1 ml BBT and 5 % horse serum before being placed in fresh BBT and 5 % horse serum with anti-HA antibody at a concentration of 1 in 1000 and rolled overnight at 4 °C. Embryos were then washed extensively in BBT and re-blocked for an hour in BBT and 5 %

horse serum before being incubated overnight in BBT with 5 % horse serum and the secondary antibody alexaflur488 at a concentration of 1 in 1000. The following day embryos were washed extensively in PBS and then examined for fluorescence labelling and imaged.

5.3.5 Testing the stability of MyoD proteins in *X. laevis* embryos using western blot

To test for protein stability over time the above samples of *X. laevis* animal caps were defrosted, vortexed and boiled at 95℃ for 5 minutes. These samples were loaded onto 12 % acrylamide/SDS-PAGE gels along with 10 µl PageRuler™ Prestained Protein Ladder 26616 (ThermoFisher). 20 ml of resolving gel consisted of 8 ml 30 % acrylamide, 5 ml 1.5 M Tris pH 8.8, 7 ml dH_2O, 200 µl 10 % SDS, 100 µl 10 % APS and 50 µl TEMED and 10 ml of stacking gel consisted of 1.3 ml 30 % acrylamide, 2.5 ml 0.5 M Tris pH 6.8, 6.1 ml dH_2O, 100 µl 10 % SDS, 50 µl 10 % APS and 50 µl TEMED. All gels were run at 180V for 2 hours in Tris-Glycine gel running buffer (3 g/l Tris, 14.4 g/l Glycine, 10 ml/l 10 % SDS).

Gels were transferred at 100V for 2 hours in gel Tris-Glycine transfer buffer (3 g/l Tris, 14.4 g/l Glycine, 10 ml/l 10 % SDS, 100 ml/l methanol) onto PVDF membrane pre-soaked in methanol prior to transfer. Post-transfer, membranes were blocked in 5 % Marvel milk powder/PBST for 1 hour before transfer to fresh 5 % Marvel milk powder/PBST with the primary antibody of anti-HA (Sigma) at a concentration of 1 in 4000 overnight at 4℃. Following this the membrane was washed five times in PBST for 30 minutes, blocked as before in 5 % Marvel milk powder/PBST for 1 hour before a 1 hour incubation in secondary antibody, 1 in 4000 anti-mouse conjugated with HRP. Membranes were washed as previous, developed with BM Chemiluminescence kit (Roche) and exposed to

hyperfilm ECL until visualised. Following this, membranes were washed in PBST and blocked again in 5 % Marvel milk powder/PBST. It was then exposed to a second primary antibody, β-tubulin at a concentration of 1 in 10,000 followed by an anti-rabbit secondary at a concentration of 1 in 2000. Membranes were washed, treated with secondary and washed again as previous before being developed and visualised for a second time using an Enhanced chemiluminescence kit exposed to film.

5.4 Results

5.4.1 Schematic representation of *A. alcalica* and *D. rerio* MyoD1

Comparison of the MyoD1 sequences from teleost species indicates that the poly-serine insertion is present in a large number of the genes analysed. The insertion ranges from 16 to 27 amino acids that are mostly serine but also includes proline and leucine residues, with some heterogeneity between species. The here-named poly-serine region is located between the transactivation domains (TAD) and the cysteine-histidine rich region (H/C) of MyoD1 (Figure 23). A literature search was unable to find any research into the presence of this region or its function although its existence is acknowledged.

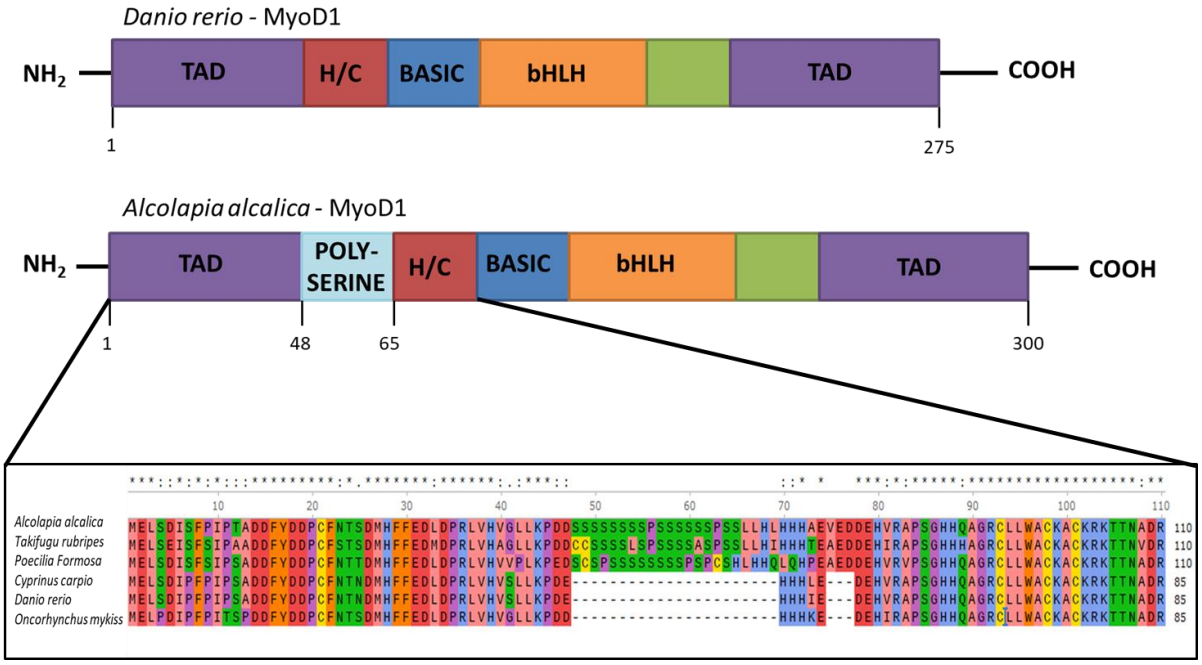


*Figure 23: Analysis of the Myod1 poly-serine insertion. Schematic representation of the Myod1 amino acid sequences of Danio rerio and Alcolapia alcalica showing the position of the poly-serine insertion. Adapted from Bharathy & Taneja 2012. Enlarged area of multiple species alignment indicating examples of individuals with and without the poly-serine insertion. * represents exact match and : represents residues where the majority of species show conservation.*

5.4.2 Phylogenetic analysis reveals relationship between MyoD genes in teleost fish

A mass alignment of the amino acid sequences of the multiple MyoD's found in the genomes of teleost fish was produced and compared via phylogenetics. The analysis produced two separate clusters with strong bootstrap support indicating that the multiple genes in these fish are comprised of MyoD1 and MyoD2 (Figure 24), with some species retaining multiple MyoD1 genes. Interestingly, the MyoD1 sequences with and without the poly-serine insertion clustered together in the analysis.

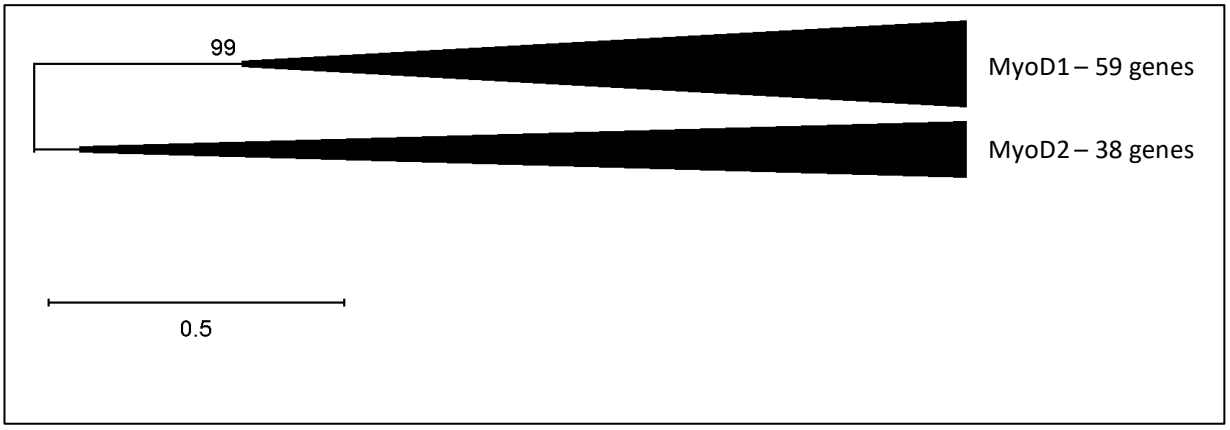


Figure 24: Collapsed phylogenetic tree analysing the amino acid sequences of 97 MyoD proteins (59 MyoD1 [34 with a poly-serine insertion] and 38 MyoD2) from 54 teleost species. Analysis was conducted on MEGAX using the JTT w/freq method and gamma distribution. MyoD1 and MyoD2 proteins claded separately with both MyoD1 containing and not containing a poly-serine insertion cladding as MyoD1. Clades are supported by a bootstrap value of 99, scale bar represents a genetic distance of 0.5 amino acid substitutions per site. Non-collapsed tree shown in full in appendix 1.

The teleost phylogeny published by Betancur-R et al (2017) was used to investigate the evolution of the different MyoD genes in these fish species (Figure 25). The presence of the poly-serine insertion in some species MyoD1 is strongly linked to the presence of a MyoD2 gene. In fact, amongst the species in this study, the poly-serine is only found in species with a MyoD2 and the majority of those without a poly-serine lack a MyoD2. This analysis indicates that the poly-serine insertion had a single evolutionary event and was incorporated and retained by the majority of species in this lineage. However, some species appear to have subsequently lost the poly-serine insertion whilst retaining MyoD2. Further, some species which have neither a poly-serine insertion in their MyoD1 or a MyoD2 gene have multiple MyoD1.

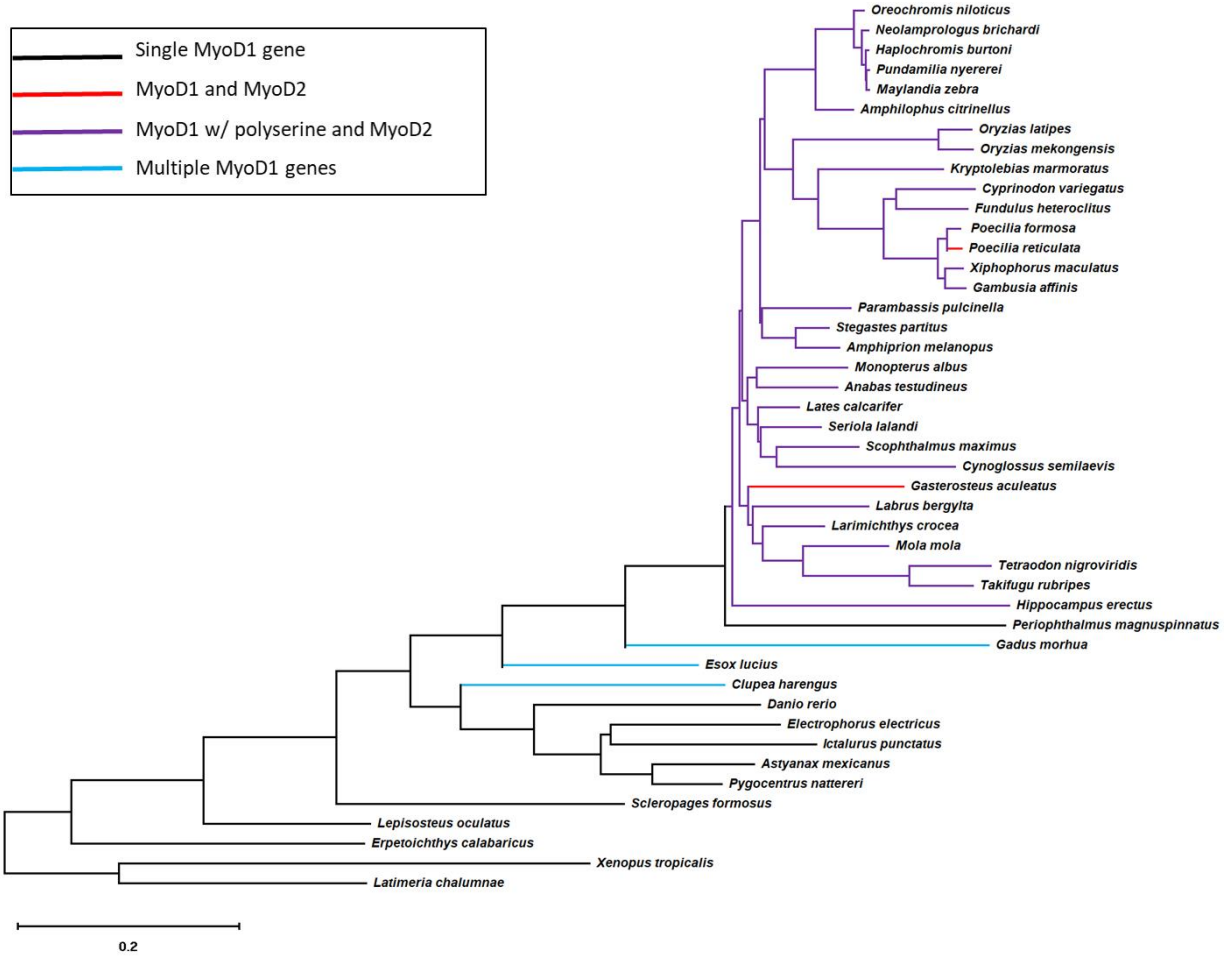


Figure 25: Teleost species phylogeny for all species where complete MyoD sequence data was available and which could be assigned to the published phylogeny by Betancur-R, Wiley et al. 2017. Branches are coloured to represent the species differences in MyoD genes found in the genome and the presence or absence of the MyoD1 poly-serine insertion. The phylogeny indicates that the poly-serine insertion is only found in species which have retained a MyoD2.

5.4.3 Expression patterns of *MyoD1* and *MyoD2* in developing *A. alcalica* embryos

In situ hybridisation using DIG labelled cRNA probes for the two *MyoD* genes found in *A. alcalica* was used to determine their spatial expression patterns (Figure 26). In the two stages examined in this study the staining for *MyoD1* RNA was always stronger when compared to *MyoD2* suggesting a greater expression in the developing embryo. Both genes show expression in the developing somites along the body axis (black arrows), in the pectoral fin buds (white arrows) and in facial muscle tissue (white arrowheads), typical of *MyoD* genes. *MyoD1* expression can also be seen in adaxial cells connecting the somites, this expression cannot be seen in the embryos probed for *MyoD2.*

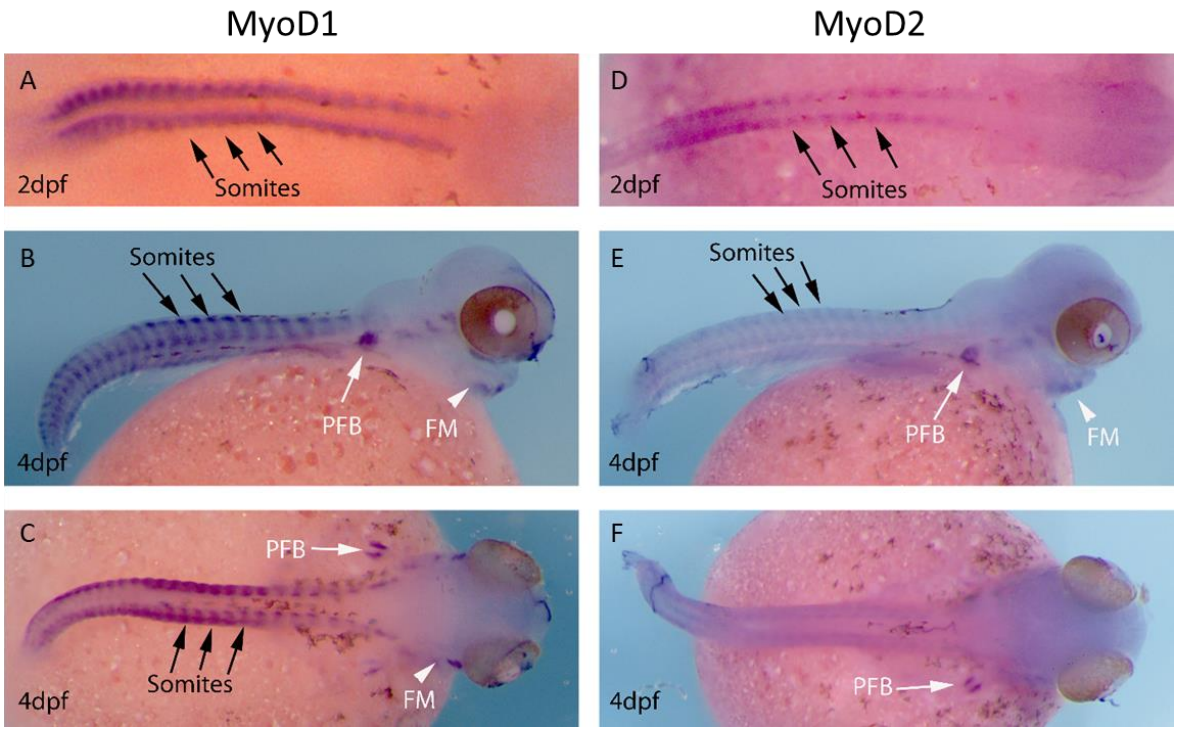


Figure 26: Expression analysis of Myod1 and Myod2 in the developing embryos of Alcolapia alcalica. A-C) Lateral and dorsal views of in situ hybridisation for MyoD1 and D-F) for MyoD2, in developing A. alcalica embryos at different stages (number of days post fertilisation [dpf] indicated). Arrows show somites, white arrowheads indicate facial muscle (FM) and white arrows indicate developing pectoral fin bud (PFB). Black dots around the yolk and on the body are chromatophores (pigment cells). Staining for Myod1 was always stronger than for MyoD2.

5.4.4 *MyoD1* from *A. alcalica* and *D. rerio* expressed in *Xenopus laevis* embryos show nuclear localisation

HA-tagged mRNA constructs representing *MyoD1* from both *A. alcalica* and *D. rerio* were expressed in *Xenopus* embryos. Immunostaining using a anti-HA antibody with alexaflur488 secondary was subsequently conducted to determine the localisation of the translated proteins. Embryos were imaged for fluorescence showing that, for both *A. alcalica* and *D. rerio MyoD1,* HA was shown to be found in punctate regions across the embryos (Figure 27). This suggests that MyoD1 protein from these species localises to the nucleus of the cell.

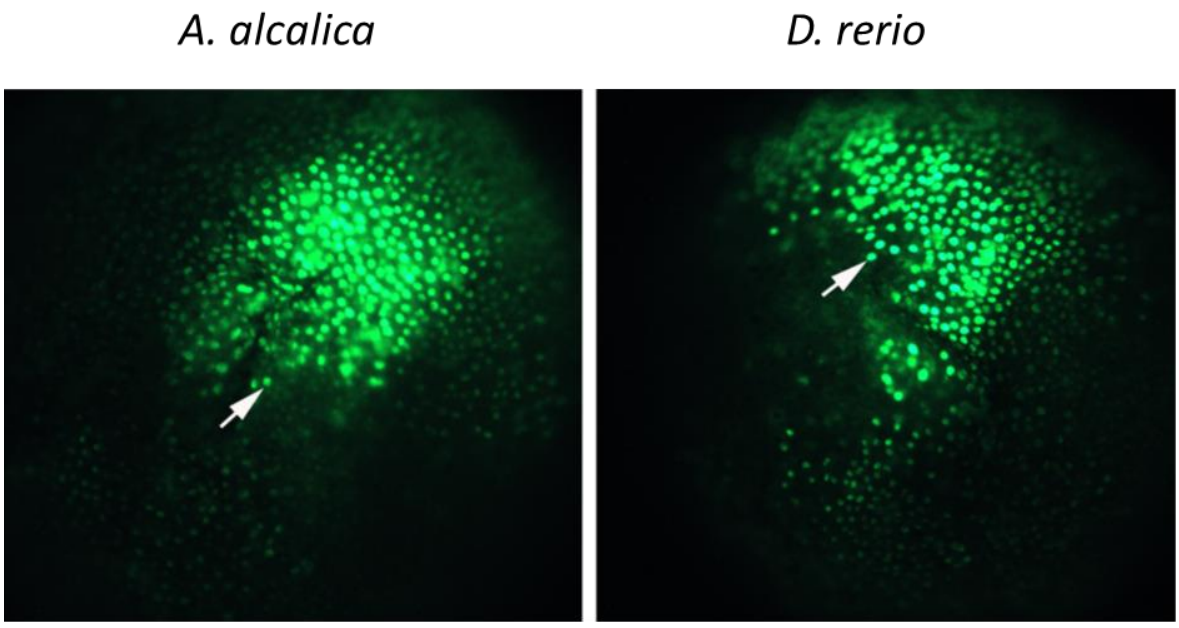


Figure 27: Immunostaining for HA-tagged MyoD1 proteins representing MyoD1 from both A. alcalica and D. rerio shows nuclear localisation when mRNA constructs are expressed in Xenopus laevis embryos. White arrows show examples of the punctate expression representative of nuclear localisation.

5.4.5 MyoD1 protein from *A. alcalica* is more stable than MyoD1 from *D. rerio*

To compare the relative stability of the MyoD1 protein with poly-serine insertion found in *A. alcalica* to the poly-serine lacking MyoD1 protein found in *D. rerio* RNA for each was injected into *Xenopus laevis* embryos at the one cell stage. Dissected animal caps were incubated in cyclohexamide, a translational inhibitor, and collected at different time points and loaded onto a 12 % acrylamide/SDS-PAGE gel which was subsequently probed with an anti-HA antibody (Figure 28). After three hours *A. alcalica* MyoD1 protein was consistently still detectable whereas, at the same time point, *D. rerio* protein was absent and broken down. This analysis suggests that *A. alcalica* MyoD1 is more stable than *D. rerio* MyoD1, a feature that could be due to the poly-serine insertion.

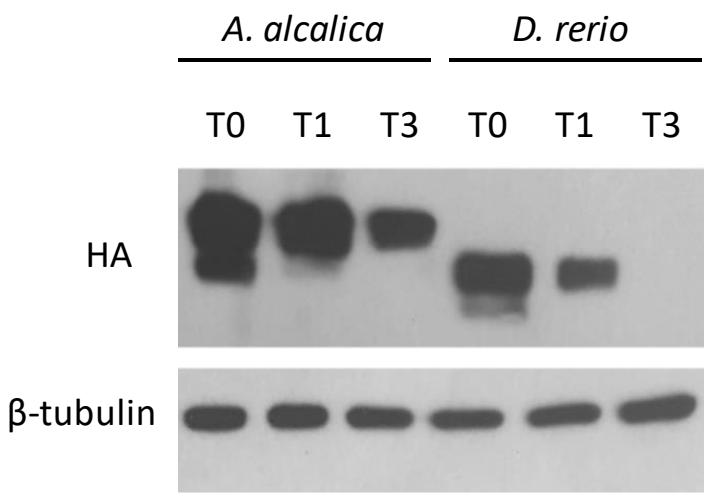


Figure 28: Western blot analysis conducted on Xenopus laevis animal caps overexpressed with constructed MyoD1 RNA [with HA tag] from either Alcolapia alcalica or Danio rerio. Animal caps were incubated for either 0, 1 or three hours in the translational inhibitor, cyclohexamide. The membranes were probed with anti-HA and anti-β-tubulin antibodies and show the stability of the overexpressed constructs over time. D. rerio constructs are consistently less stable than A. alcalica constructs with the former being absent by the three hour time point and the latter persisting.

5.5 Discussion

5.5.1 Phylogenetic analysis reveals single evolutionary event for poly-serine domain in MyoD1

Analysis of the amino acid sequences of MyoD1 and MyoD2 from a broad range of teleost species confirms that a large number of fish retain two separate MyoD genes which likely arose from the teleost specific genome duplication event (Aase-Remedios, Coll-Lladó et al. 2020). These genes are in different areas of the genome and are distinct from the multiple *MyoD1* found in the genomes of a few species of fish such as those in the Salmonidae which arose from a lineage specific genome duplication (Macqueen and Johnston 2006). The investigation into the evolution of MyoD2 carried out in this study, using the published teleost phylogeny (Betancur-R, Wiley et al. 2017), suggests an alternate single evolutionary event not encompassing all teleost and hence having not occurred at the base of the teleost lineage. However, recent findings from other labs have shown that teleost species which lack a *MyoD2* gene retain the ghost loci in the same region (Macqueen and Johnston 2008, Aase-Remedios, Coll-Lladó et al. 2020). This provides evidence for the whole genome duplication event resulting in a second MyoD in teleost, which was subsequently lost in a number of lineages. Interestingly, the presence of a poly-serine domain in the sequence of MyoD1 is strongly linked to the presence of MyoD2. The phylogenetic analysis comparing MyoD1 sequences with and without this poly-serine region suggests that the region had a single evolutionary event and was retained in the vast majority of evolved species following its inclusion. The observation that acquiring a poly-serine insertion in MyoD1 is closely tied to the retention of MyoD2 in nearly all the Actinopterygii suggests a functional significance.

5.5.2 Preliminary functional analysis of MyoD1 poly-serine region and future directions

The incorporation of a single amino acid repeat regions into a protein sequence is common in animals and plants. These regions are produced by slippage during replication whereby the DNA is denatured and displacement of complementary strands results in mispairing and the incorporation of additional bases. These are then assimilated during recombination, in this way a growing string of repeat amino acids may emerge (Levinson and Gutman 1987, Pâques, Leung et al. 1998, Persi, Wolf et al. 2016). The theory to the evolution of repeat regions via replication slippage has led to the hypothesis that repeat amino acids are more likely to occur in proteins with relaxed selection pressures, such as duplicated proteins (Haerty and Golding 2010). This could explain the link between the presence of a poly-serine region in the MyoD1 of species with the retention of a MyoD2, with reduced selection pressures on MyoD1 because a second MyoD protein was capable of sharing function. Repeat regions within proteins present new genetic material for evolution to act upon and have been associated with change in protein function, increased protein flexibility or stability and a number of diseases (Siwach, Pophaly et al. 2006, Kumar, Sowpati et al. 2016). Although the repeat amino acid region found in teleost MyoD1 is variable, it is mostly made up of serine residues, at least 11. The variability is likely a consequence of the increased mutation rates which occur in short repeat regions of DNA (Kashi and King 2006, Haerty and Golding 2010).

The fact that the polys-serine region evolved and was retained suggests it may have provided a selective advantage enhancing the function of MyoD1. Amino acid repeat regions in proteins have been shown to have many different roles. One is to change protein stability (Kumar, Sowpati et al. 2016). To examine whether the inclusion of a poly-

serine region affected the stability of MyoD1, the mRNA coding for the protein from a species with and a species without the insertion, *A. alcalica* and *D. rerio* respectively, were injected into the embryos of *X. laevis* and the breakdown of the protein measured over time. MyoD activity is regulated by phosphorylation which can trigger ubiquitination and degradation (Song, Wang et al. 1998). Analysis by western blot showed that the poly-serine containing MyoD1 from *A. alcalica* was stable for at least three hours, whereas *D. rerio* MyoD1 was no longer detected by this time point. Similar to *D. rerio,* mouse MyoD1 does not contain a poly-serine region and has also been shown to be unstable with a half-life of about 45 mins (Sadeh, Breitschopf et al. 2008). Further analysis with chimeric proteins where the poly-serine region is removed from the *A. alcalica* sequence and inserted in the *D. rerio* sequence would be needed to confirm that this region is responsible for the difference in stability.

Another function of amino acid repeats in protein sequences is that they are capable of effecting DNA binding and either repressing gene expression or increasing transcriptional ability (Gerber, Seipel et al. 1994, Galant and Carroll 2002). MyoD is a master regulator of muscle differentiation and a potent transcription factor (Pownall, Gustafsson et al. 2002). If the incorporated poly-serine region is affecting its DNA binding capability it would be interesting to examine any differential activity of *Alcolapia* MyoD1 on known downstream targets.

5.5.3 Differences in the expression patterns of *MyoD1* and *MyoD2* in *A. alcalica*

In situ hybridisation was used to determine the expression patterns of *MyoD1* and *MyoD2* in the developing embryos of *A. alcalica.* Both genes show strongest expression in

somites along the body axis, developing pectoral fin buds and facial muscle tissues, as would be expected for genes involved in muscle cell determination and differentiation (Pownall, Gustafsson et al. 2002). *Myod1* expression was found to be more expansive than that of *MyoD2,* with RNA detected in most somites, as well as in the adaxial cells which give rise to slow muscle tissue (Devoto, Melançon et al. 1996). Somatic *MyoD2* expression was more restricted being expressed in fewer somites and not in adaxial cells. This suggests that there may be some subfunctionalisation of the two genes. *MyoD1* expression in somites and adaxial cells suggests it is important in both fast and slow muscle development, while *MyoD2* in not expressed in the slow muscle lineage so may only having a role in fast muscle development. Expression analysis carried out in Atlantic halibut (*Hippoglossus hippoglossus*) and gilthead seabream (*Sparus aurata*) showed similar findings to those seen here in *A. alcalica,* with *MyoD2* showing expression in a reduced number somites and being absent in adaxial cells. This suggests the subfunctionalisation of the genes occurred early in the teleost lineage (Tan and Du 2002, Andersen, Dahle et al. 2009). The Atlantic halibut, a flatfish, showed further differential expression of the *MyoD* isoforms with left-right asymmetry of *MyoD2* expression (Andersen, Dahle et al. 2009).

5.5.4 Conclusions

Considering the analysis here with the published work of other authors, it is clear that the two *MyoD* isoforms found in the genomes of a large number of teleost fish arose from the teleost specific whole genome duplication event and was subsequently lost in certain lineages (Macqueen and Johnston 2008, Aase-Remedios, Coll-Lladó et al. 2020). The MyoD1 protein in those lineages which retained a MyoD2 have evolved a poly-serine

region between the transactivation domains (TAD) and the cysteine-histidine rich region (H/C). This region has been retained by the vast majority of species included in this study which would suggest a functional role. Preliminary analysis on the function of this poly-serine region suggests it may have a role in increasing protein stability however, this would need to be elucidated with further research. Additionally, incorporated repeat amino acids in protein sequences are known to have a number of different functions, including changes to transcriptional activity, which could provide interesting future avenues to explore. Finally, the expression patterns of *MyoD1* and *MyoD2* in developing embryos of *A. alcalica* show both overlapping and distinct expression, suggesting some subfunctionalisation, consistent with findings in other fish species (Tan and Du 2002, Andersen, Dahle et al. 2009).

Chapter Six – Discussion

This PhD initially aimed to better understand how *Alcolapia* species survive the harsh conditions seen in Lake Natron and Lake Magadi. Two of the data chapters (Chapter 2 and 3) presented here examine adaptations to the excretion of nitrogenous waste that plays a role in allowing these fish to persist in the extreme environment they inhabit. All four data chapters document the progress in studying *A. alcalica* which had not previously been adapted to laboratory conditions. It was necessary to learn successful husbandry, understand how these animals breed, find ways of accessing embryos for experiments and adapting existing biological techniques to best study this species. Chapter 4 then goes further into developing *A. alcalica* as a potential future model species which could be a valuable tool for research due to their rapid evolution under extreme conditions. To do this, this chapter also documents the progress in using the gene editing technology CRISPR-Cas9, which has not previously been carried out in *Alcolapia*. Finally, chapter 5 was initially intended to examine muscle and muscle development in *A. alcalica* but instead broadened into the study of the evolution of muscle genes in fish. This final data chapter is somewhat a commentary on the need for studying a large range of species rather than relying on few model organisms to inform our understanding of biological processes.

6.1 Adaptation and evolution of *Alcolapia* to extreme conditions

To overcome the extreme conditions they experience in Lake Natron, *A. alcalica,* as well as the entire *Alcolapia* species flock, have a series of adaptations crucial to their survival. The *Alcolapia* have evolved to survive their high pH environment via adaptations to the excretion pathway (Randall, Wood et al. 1989, Lindley, Scheiderer et al. 1999). It has been reported that *A. grahami* convert 100 % of their nitrogenous waste to urea due to

difficulties excreting the toxic compound ammonia in water with an increased pH (Randall, Wood et al. 1989). To accommodate the conversion of large quantities of ammonia *Alcolapia* have adaptions to enzymes in the OUC (Lindley, Scheiderer et al. 1999), changes in expression patterns of genes crucial to excretion (Wood, Nawata et al. 2013) and have a novel transporter for efficient removal of urea (Walsh, Grosell et al. 2001). These adaptations have occurred due to natural selection acting on favourable mutations in different genes, allowing *Alcolapia* to survive their environment by utilising urea as an excretory product.

Mutations may occur in any region of the genome of an individual organism, over many generations any given mutation may become fixed. Mutations can be lethal or have negative consequences on an individual and may be selected against until alleles of this kind are no longer present, they may have no effect at all and may persist in a population through genetic drift, or a mutation may concur a selective advantage and will spread through a population until it becomes ubiquitous (Kosiol and Anisimova 2019). Mutations which occur within the coding region of a given gene may change the amino acid sequence of its protein, in turn effecting this proteins function. Alternatively, changes elsewhere in the genome may incorporate new, or destabilise existing, enhancer/promoter regions potentially causing changes in spatiotemporal gene expression (Haldane 1933, Holland 2013).

This thesis has shown that mutations have occurred in numerous genes within the *Alcolapia* lineage that have changed the function and expression of proteins vital for the excretion of nitrogenous waste. CPS III has accumulated both such mutations; changes in the coding region of the gene have affected function and mutations upstream of the gene

have incorporated a new promoter region which induces expression in tissues outside its usual expression range. CPS III is the first enzyme in the OUC and in fish usually binds glutamate which is produced by GS from ammonia whereas terrestrial vertebrate CPS I is capable of binding ammonia directly (Saeed-Kothe and Powers-Lee 2002, Loong, Chng et al. 2012). *Alcolapia CPS III* shows convergent evolution with terrestrial vertebrate *CPS I* with mutations changing amino acid residues altering binding capabilities. A further adaptation to this enzyme in *Alcolapia* is seen in its expression pattern. Enzymes of the OUC are predominantly restricted to hepatic tissues whereas *Alcolapia* have gained a pair of E-boxes which likely facilitates expression in muscle. The adaptation to bind ammonia directly favours a more efficient management of nitrogenous waste and by utilising a larger area such as muscle tissue *Alcolapia* facilitate the conversion of large quantities of ammonia into urea.

Converting large quantities of ammonia to urea requires a significant amount of energy compared with excreting ammonia directly (Wright 1995). It has been suggested that *Alcolapia* have become 100 % ureotelic in order to survive their high pH environment and that they have since lost the ability to excrete ammonia, becoming obligate ureotelic species (Randall, Wood et al. 1989, Wood, Wilson et al. 2002). Ammonia is a major nitrogenous waste product and is produced as a result of the catabolism of amino acids via transdeamination (Wright and Fyhn 2001) meaning for obligatory ureotelism to be true all ammonia must be rapidly converted to urea otherwise under suitable conditions ammonia excretion would have been recorded. With this in mind, obligate ureotelism would remove a cheap and easy method of excretion which would appear to be counter intuitive as, if conditions allowed, it makes sense to excrete ammonia over urea. This thesis has found that, although *Alcolapia* rely on urea to excrete the majority of their

nitrogenous waste, *A. alcalica* in laboratory conditions continue to excrete a proportion of their waste as ammonia. This finding is contrary to previous reports (Wood, Wilson et al. 2002).

Genes in the ammonia transporter family, Rh, are shown to be expressed in gill tissue of *A. alcalica*, as is the case with other fish species (Tsui, Hung et al. 2009, Sinha, Kapotwe et al. 2016, Zimmer, Wright et al. 2017). Analysis of adult gill tissue using *in situ* hybridisation methods against the different Rh proteins indicate that *Alcolapia* have also likely retained the characteristic expression patterns within the gill. In addition to conserved expression, experiments in *D. rerio* embryos and a triple-mepΔ mutant yeast strain confirmed that *A. alcalica Rhbg* was capable of moving ammonia in these orthologous systems, suggesting the Rh are still functional as ammonia transporters. These data show that *A. alcalica* is in fact not an obligate ureotelic species and uses the same ammonia excretion method as other teleost species by using members of the ammonia transporter family, Rh, to move ammonia across gill tissue. The function of *A. alcalica* Rhcg1 and Rhcg2 was also examined in the triple-mepΔ mutant yeast strain, here it was established that Rhcg1 could initiate movement of ammonia, although to a lesser extent than Rhbg, and Rhcg2 showed no sign of being an ammonia transporter. Interestingly, *A. alcalica Rhbg* and *Rhcg2* both showed sign of positive selection. It is plausible that there are changes to these genes which mean they are still capable of moving ammonia at high pH, which would be a beneficial adaptation. Alternatively, Rhcg2 may indeed no longer be a functional ammonia transporter, however this would require further study.

Excretion of nitrogenous waste is clearly an important adaptation for the survival of *Alcolapia* to their high pH environment. This is evident by the multiple adaptations to

excretion shown both here and in previous studies (Randall, Wood et al. 1989, Lindley, Scheiderer et al. 1999). The pH of freshwater ecosystems across the world is expected to change due to climate change and one of the effects this will have on fish populations is their ability to excrete nitrogenous waste (Handy and Poxton 1993, Hasler, Butman et al. 2016, Spyra 2017). This thesis has highlighted that if species are to survive these rapid changes, like the *Alcolapia,* they will require an efficient method of excretion suitable to the pH of their environment. Facilitative ureotelism is reasonably widespread amongst teleost species, although to varying degrees. This is because fish retain the genes coding for the enzymes of the OUC and many are shown to express them during early development (Korte, Salo et al. 1997, LeMoine and Walsh 2013) as well as when exposed to high pH environments (Kong, Edberg et al. 1998, Banerjee, Koner et al. 2020). This means, in theory, evolution could act to express these genes to higher levels if needed rather than having to evolve an entire mechanism of excretion. These experiments have shown that *A. alcalica* presents itself as an interesting organism and exemplifies the benefits of having an extremophile model available for research.

6.2 Other adaptations of interest

During this PhD other adaptations presented themselves as interesting avenues for exploration. Preliminary analysis was conducted on the morphology of *A. alcalica* sperm showing they differed from the published images of *A. grahami* (Papah, Kisia et al. 2013), which in turn appeared to differ from the sperm of other cichlid species. As external fertilisers, the gametes of *Alcolapia* come into direct contact with hostile water conditions and hence must have adaptations allowing them to successfully form a zygote. The genome of another fish adapted to a high pH environment, the Amur ide, was shown to

contain more genes coding for zona pellucida domain-containing proteins or vitelline envelope proteins than is usual (Xu, Li et al. 2017). These proteins coat egg cells and, in addition to playing a role in sperm recognition, they protect the cell in the external environment (Litscher and Wassarman 2018). Further examination of *A. alcalica* gametes may show similar findings and access to a fully sequenced and annotated *Alcolapia* genome would make it possible to search for these and other adaptations more easily.

Access to a fully annotated genome would have allowed widescale genetic screening which could have pinpointed genes under positive selection within the *Alcolapia* lineage. These genes may have important implications for the local adaptation seen in these species (Manel, Perrier et al. 2016). This information, in conjunction with direct access to these fish and their embryos, would allow experimental functional analysis of a variety of adaptations. *Alcolapia* survive water temperatures of 30-42.8 °C, high pH of 9-11.5, fluctuating dissolved oxygen levels and high salt concentrations of >20ppt (Ford 2015, Ford, Dasmahapatra et al. 2015). To overcome this, they must have a series of physiological adaptations to cope with a vast array of stressors and for maintaining homeostasis, in addition to the adaptations to nitrogen excretion examined here. These adaptations will have a genetic basis, whether this be changes in gene function or expression, and analysis of the *Alcolapia* genome would allow faster identification of these genes.

6.3 Progress in developing *A. alcalica* as a model organism

Model organisms are non-human species which can be easily studied to test biological theory with hopes that the data collected is then applicable to a wider range of species (Leonelli and Ankeny 2013). Model organisms tend to share a number of features which

makes them easy study systems to represent a wider group. For instance, they are usually small in size, cheap and easy to house and have a short reproductive cycle with a large number of offspring (Hedges 2002, Leonelli and Ankeny 2013). These features allow large numbers of individuals to be kept in a single facility and examination of the full life cycle of an individual can be experienced in a short period of time. These characteristics lend these species as genetic tools for broad biological study. For example, *D. rerio* or zebrafish were common house pets prior to their use in research, as such methods of husbandry are very well established and they are relatively easy and cheap to care for (Briggs 2002, Westerfield 2007). These fish reach sexual maturity in around three months and a single breeding pair is capable of producing over a thousand embryos. In addition, *D. rerio* are photoresponsive breeders meaning that a large number of single cell embryos can be accessed on demand for use in experiments (Westerfield 2007, Adatto, Lawrence et al. 2011).

A. alcalica are not particularly small in size but can be kept at higher concentration than other model organisms such as mice and rats. Their system conditions are not as simple as zebrafish however, the water chemistry best suited for them has now been established and once set up is easily maintained. A recurring problem with the use of *A. alcalica* as a model organism is access to a large number of single cells embryos. Breeding tanks with dividers were made to separate male and female fish which was shown to be successful in encouraging individuals to breed when the divider was removed. This was not as effective as with zebrafish with the majority of tanks not producing embryos on any given day. To overcome this, flexibility with experiments or a large number of breeding tanks is required. Further, although *A. alcalica* produced the largest number of embryos of the *Alcolapia* species available a single female would usually carry fewer that 60 (this PhD and

Seegers and Tichy 1999). Difficulties were also seen in using this species as a genetic tool and attempts at using CRISPR-Cas9 to target *Pax-6* was unsuccessful (chapter 4). Also detailed in chapter 4 were the issues with young embryos, like many cichlids the *Alcolapia* species are mouth brooders (Balshine-Earn and Earn 1998). It was possible to culture young embryos in petri dish however survival rate was often very low which poses problems with the study of developmental biology.

The number of animals considered model organisms has grown rapidly with the increase in genome-sequencing projects (Hedges 2002). This is because access to an organisms genome significantly increases the quality and amount of research which can be conducted. The genome information available for *A. alcalica* is not fully annotated, not easily searchable (a knowledge of the coding language unix is required) nor is it publicly available. A further issues arises in the current stock of *A. alcalica* available for research. Currently, three universities house population of *A. alcalica,* York, Bangor and Hull. All of these are descended from the original population collected from Lake Natron in 2017 and although there has been no extensive breeding if further research groups wished to utilise this species inbreeding would quickly become a problem.

With further advancements it may be possible for *A. alcalica* to be used as a model by the wider biological community, but at present they do not suit research as a model organism because of the reasons detailed here. This is not to say that *A. alcalica* is not an interesting experimental organism just that it does not lend itself as an easy model for the study of general biological phenomena. All chapters in this thesis have established techniques in the molecular, developmental and physiological study of these fish which could be used for future research into their biology.

6.4 Importance of studying non-model species

Model species present themselves as simplified systems for biological research. The large number of people that work on them has led to the creation of expansive databases and well documented tools and protocols which increase the ease of their study (Russell, Theriot et al. 2017). These species have been invaluable in progressing our understanding of different scientific principles; however, it must be noted that for many areas of biology our knowledge is greatly restricted to the information gained from a single, or few, species. For example, *D. rerio* has become a well established embryological tool for developmental biology despite known differences between this species and humans as well as between *D. rerio* and other fish species (Schartl 2014, Meyers 2018). The study of non-model species would allow wider comparisons and better understanding of biological function and adaptive traits (Crawford 2001). Extremophile organisms present themselves as naturally occurring biological experiments which could prove valuable in research of core biological function and the constraints of evolution. As an extremophile vertebrate, *Alcolapia* could provide important information in the study of vertebrate species, potentially contributing to our understanding of developmental and molecular biology, physiology and even human disease (Irwin 2010, Jorge, Borges et al. 2016).

The teleost specific genome duplication created new material for evolution to act upon. Subsequent losses in some lineages has resulted in much variation between species and wider study of these genes could uncover important or even novel functions (Hoegg, Brinkmann et al. 2004, Volff 2005, Glasauer and Neuhauss 2014). This thesis has studied the *Pax-6* and the *MyoD* gene families in *A. alcalica* and found that the genome of *D. rerio*

has lost some the gene duplicates, *Pax-6-like* and *MyoD2,* which are present in the *Alcolapia* genome.

The *Pax-6* gene family in *A. alcalica* show spatiotemporal differences in expression suggestive of subfunctionalisation of the genes. *Pax-10* is detected later in development and only in retina which matches studies in zebrafish, *Xenopus* and anole lizard (Feiner, Meyer et al. 2014). *Pax-6* and *Pax-6-like* contribute to the combined expression of *Pax-6a* and *Pax-6b* in zebrafish and *Pax-6* in mammals and birds (Nornes, Clarkson et al. 1998, Kleinjan, Bancewicz et al. 2008). *A. alcalica Pax-6-like* is also detected in an area of the hindbrain in the region of the locus coeruleus which, if accurate, would represent an area of the brain a *Pax-6* gene had not previously been reported to show expression. Further species comparisons would be required to determine whether this is an adaption specific to *Alcolapia,* potentially important for their survival to extreme conditions, or whether *Pax-6-like* expression in the locus coeruleus is more widespread. Only by the study of non-model species will this type of evidence be gained.

MyoD1 and *MyoD2* also show differential expression in *A. alcalica.* Both are expressed in the developing somites, facial muscle and pectoral fin bud, however *MyoD2* is detected in fewer somites and at lower levels. Additionally, *MyoD1* shows expression in adaxial cells whereas *MyoD2* does not. This *MyoD2* expression seen here matches that seen in Atlantic halibut and gilthead seabream (Tan and Du 2002, Andersen, Dahle et al. 2009). There are only a limited number of studies that have examined *MyoD2* which could be due to its absence from the *D. rerio* genome. These studies have not examined the role this additional *MyoD* plays in development. Interestingly, the species which have retained a *MyoD2* have gained a poly-serine insertion in their *MyoD1.* Again, the absence of this

from the *MyoD1* of common model species *D. rerio* means that research into the function of this region is lacking. Preliminary analysis conducted in this thesis comparing *MyoD1* from *A. alcalica* and *D. rerio* would suggest that this region may increase protein stability. Of the species examined it would appear that those which have retained *MyoD2* are also those with a *Pax-6-like,* this could indicate a large scale loss of areas of the genome having occurred outside of this lineage or an ancient pressure resulting in it being retained by their shared ancestry. The data presented in this thesis examining genes in both the *Pax-6* and *MyoD* gene families gives support for the wider study of non-model species. It gives evidence of subfunctionalisation and potentially differences in gene function that could not have been established from using *D. rerio* as the model species for fish development.

A. alcalica are an interesting experimental animal and although they do not present themselves as a model species their interesting evolution under extreme conditions means that they may be useful in studying adaptation. These fish inhabit some of the most extreme conditions survived by vertebrate species. This survival must be supported by interesting and important adaptations which means the *Alcolapia* may join other extremophile fish species as a study organism in broad areas of biology; such as *A. mexicanus* in development (Jeffery 2019) and *N. furzeri* in biomedical research (Priami, De Michele et al. 2015).

6.5 Overall conclusions

This discussion shows the progress made in not only our knowledge of the adaptations found in the *Alcolapia* lineage but also the advancements made in the study of this group of species. Evidence is presented on *Alcolapia* species adaptation to excretion of

nitrogenous waste, showing that multiple genes are involved in allowing continued, efficient excretion in their high pH environment. Many of the genes involved have evolved new function or have different spatiotemporal expression to allow the *Alcolapia* to convert large quantities of ammonia to urea and to potentially excrete any remaining ammonia across gill tissue and against a concentration gradient. To conduct this work, this thesis shows that correct husbandry for *A. alcalica* was established and a number of protocols were adapted to best study this species. Although *A. alcalica* do not present themselves as a model organism, due to the difficulties in studying them discussed here, the advancements made show that research can be conducted and that they are an interesting experimental animal.

Further, this discussion comments on the study of non-model organisms and the benefits of wider species comparisons. Members of the *Pax-6* and *MyoD* gene family were studied as part of this thesis and evidence presented of differences between the genes found in *A. alcalica* when compared to common fish model organism, *D. reiro.* Phylogenetic analysis and gene comparisons show that these differences are likely not specific to *A. alcalica* and lends support to the notion of the study of a wider range of non-model organisms. The *Alcolapia* species flock are an interesting system of extremophile vertebrates. They must have many undocumented adaptations to survive their environment and further study of them may improve our understanding of many biological processes. Their acclimation to laboratory conditions means that future study of this species, including experimental research, is now much more feasible.

References

Aase-Remedios, M. E., C. Coll-Lladó and D. E. Ferrier (2020). "More than one-to-four via 2R: evidence of an independent amphioxus expansion and two-gene ancestral vertebrate state for MyoD-related Muscle Regulatory Factors (MRFs)." Molecular Biology and Evolution.

Abdulnour-Nakhoul, S., T. Le, E. Rabon, L. L. Hamm and N. L. Nakhoul (2016). "Structural determinants of NH 3 and NH 4+ transport by mouse Rhbg, a renal Rh glycoprotein." American Journal of Physiology-Renal Physiology **311**(6): F1280-F1293.

Adatto, I., C. Lawrence, M. Thompson and L. I. Zon (2011). "A new system for the rapid collection of large numbers of developmentally staged zebrafish embryos." PloS one **6**(6): e21715.

Ahnfelt-Rønne, J., J. Hald, A. Bødker, H. Yassin, P. Serup and J. Hecksher-Sørensen (2007). "Preservation of proliferating pancreatic progenitor cells by Delta-Notch signaling in the embryonic chicken pancreas." BMC developmental biology **7**(1): 63.

Altinok, I. and J. M. Grizzle (2004). "Excretion of ammonia and urea by phylogenetically diverse fish species in low salinities." Aquaculture **238**(1-4): 499-507.

Andersen, Ø., S. W. Dahle, S. van Nes, T. Bardal, A. Tooming-Klunderud, E. Kjørsvik and T. F. Galloway (2009). "Differential spatio-temporal expression and functional diversification of the myogenic regulatory factors MyoD1 and MyoD2 in Atlantic halibut (*Hippoglossus hippoglossus*)." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **154**(1): 93-101.

Anderson, P. M. (2001). "Urea and glutamine synthesis: environmental influences on nitrogen excretion." Fish Physiology **20**: 239-277.

Anitori, R. P. (2012). Extremophiles: microbiology and biotechnology. Norfolk, UK, Horizon Scientific Press.

Asfour, H. A., M. Z. Allouh and R. S. Said (2018). "Myogenic regulatory factors: The orchestrators of myogenesis after 30 years of discovery." Experimental Biology and Medicine **243**(2): 118-128.

Ashery-Padan, R., X. Zhou, T. Marquardt, P. Herrera, L. Toube, A. Berry and P. Gruss (2004). "Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes." Developmental biology **269**(2): 479-488.

Baalsrud, H. T., O. K. Tørresen, M. H. Solbakken, W. Salzburger, R. Hanel, K. S. Jakobsen and S. Jentoft (2018). "De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data." Molecular Biology and Evolution **35**(3): 593-606.

Balshine-Earn, S. and D. J. Earn (1998). "On the evolutionary pathway of parental care in mouth–brooding cichlid fishes." Proceedings of the Royal Society of London. Series B: Biological Sciences **265**(1411): 2217-2222.

Banerjee, B., D. Koner, R. Hasan and N. Saha (2020). "Molecular characterization and ornithine-urea cycle genes expression in air-breathing magur catfish (*Clarias magur*) during exposure to high external ammonia." Genomics **112**: 2247-2260.

Baras, E., B. Jacobs and C. Mélard (2001). "Effect of water temperature on survival, growth and phenotypic sex of mixed (XX–XY) progenies of Nile tilapia *Oreochromis niloticus*." Aquaculture **192**(2-4): 187-199.

Bargelloni, L., P. A. Ritchie, T. Patarnello, B. Battaglia, D. M. Lambert and A. Meyer (1994). "Molecular evolution at subzero temperatures: mitochondrial and nuclear phylogenies of

fishes from Antarctica (suborder Notothenioidei), and the evolution of antifreeze glycopeptides." Molecular Biology and Evolution **11**(6): 854-863.

Barimo, J. F., S. L. Steele, P. A. Wright and P. J. Walsh (2004). "Dogmas and controversies in the handling of nitrogenous wastes: ureotely and ammonia tolerance in early life stages of the gulf toadfish, *Opsanus beta*." Journal of Experimental Biology **207**(12): 2011-2020.

Barnosky, A. D., N. Matzke, S. Tomiya, G. O. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey and K. C. Maguire (2011). "Has the Earth's sixth mass extinction already arrived?" Nature **471**(7336): 51.

Beitinger, T. L., W. A. Bennett and R. W. McCauley (2000). "Temperature tolerances of North American freshwater fishes exposed to dynamic changes in temperature." Environmental Biology of Fishes **58**(3): 237-275.

Bell, M. A. (1987). "Interacting evolutionary constraints in pelvic reduction of threespine sticklebacks, *Gasterosteus aculeatus* (Pisces, Gasterosteidae)." Biological Journal of the Linnean Society **31**(4): 347-382.

Bell, M. A., G. Orti, J. A. Walker and J. P. Koenings (1993). "Evolution of pelvic reduction in threespine stickleback fish: a test of competing hypotheses." Evolution **47**(3): 906-914.

Benarroch, E. E. (2018). "Locus coeruleus." Cell and tissue research **373**(1): 221-232.

Bergman, A. N., P. Laurent, G. Otiang'a-Owiti, H. L. Bergman, P. J. Walsh, P. Wilson and C. M. Wood (2003). "Physiological adaptations of the gut in the Lake Magadi tilapia, *Alcolapia grahami*, an alkaline-and saline-adapted teleost fish." Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology **136**(3): 701-715.

Berkes, C. A., D. A. Bergstrom, B. H. Penn, K. J. Seaver, P. S. Knoepfler and S. J. Tapscott (2004). "Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential." Molecular Cell **14**(4): 465-477.

Berschick, P., C. Bridges and M. Grieshaber (1987). "The influence of hyperoxia, hypoxia and temperature on the respiratory physiology of the intertidal rockpool fish *Gobius cobitis* Pallas." Journal of Experimental Biology **130**(1): 368-387.

Betancur-R, R., E. O. Wiley, G. Arratia, A. Acero, N. Bailly, M. Miya, G. Lecointre and G. Orti (2017). "Phylogenetic classification of bony fishes." BMC Evolutionary Biology **17**(1): 162.

Bilandžija, H., L. Ma, A. Parkhurst and W. R. Jeffery (2013). "A potential benefit of albinism in Astyanax cavefish: downregulation of the oca2 gene increases tyrosine and catecholamine levels as an alternative to melanin synthesis." PLoS One **8**(11): e80823.

Billon, P., E. E. Bryant, S. A. Joseph, T. S. Nambiar, S. B. Hayward, R. Rothstein and A. Ciccia (2017). "CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons." Molecular cell **67**(6): 1068-1079. e1064.

Black, E. C. (1940). "The transport of oxygen by the blood of freshwater fish." The Biological Bulletin **79**(1): 215-229.

Blake, J. A. and M. R. Ziman (2014). "Pax genes: regulators of lineage specification and progenitor cell maintenance." Development **141**(4): 737-751.

Bonnaterre, P. J. (1788). Tableau encyclopédique et méthodique des trois règnes de la nature...: Icthyologie, chez Panckoucke libraire.

Bopp, D., M. Burri, S. Baumgartner, G. Frigerio and M. Noll (1986). "Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of *Drosophila*." Cell **47**(6): 1033-1040.

Borowsky, R. (2008). "*Astyanax mexicanus*, the blind Mexican cave fish: a model for studies in development and morphology." Cold Spring Harbor Protocols **3**(11): 1001-1005.

Borowsky, R. (2018). "Cavefishes." Current Biology **28**(2): R60-R64.

Bortesi, L. and R. Fischer (2015). "The CRISPR/Cas9 system for plant genome editing and beyond." Biotechnology advances **33**(1): 41-52.

Boulenger, G. (1912). "LVI.—Description of a new fish from British East Africa." Journal of Natural History **9**(53): 519-521.

Brahimi-Horn, M. C. and J. Pouysségur (2007). "Oxygen, a source of life and stress." FEBS letters **581**(19): 3582-3591.

Braun, M. H., S. L. Steele, M. Ekker and S. F. Perry (2009). "Nitrogen excretion in developing zebrafish (*Danio rerio*): a role for Rh proteins and urea transporters." American Journal of Physiology-Renal Physiology **296**(5): F994-F1005.

Brauner, C. J., R. J. Gonzalez and J. M. Wilson (2012). Extreme environments: hypersaline, alkaline, and ion-poor waters. Fish Physiology, Elsevier. **32:** 435-476.

Briggs, J. P. (2002). "The zebrafish: a new model organism for integrative physiology." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **282**(1): R3-R9.

Brock, T. D. (1970). "High temperature systems." Annual Review of Ecology and Systematics **1**(1): 191-220.

Brown, J. H. and C. R. Feldmeth (1971). "Evolution in constant and fluctuating environments: thermal tolerances of desert pupfish (Cyprinodon)." Evolution: 390-398.

Bruce, L., H. Guizouarn, N. Burton, N. Gabillat, J. Poole, J. Flatt, R. Brady, F. Borgese, J. Delaunay and G. Stewart (2009). "The monovalent cation leak in overhydrated stomatocytic red blood cells results from amino acid substitutions in the Rh-associated glycoprotein." Blood **113**(6): 1350-1357.

Bucher, E. A., P. Maisonpierre, S. Konieczny and C. Emerson (1988). "Expression of the troponin complex genes: transcriptional coactivation during myoblast differentiation and

independent control in heart and skeletal muscles." Molecular and Cellular Biology **8**(10): 4134-4142.

Burggren, W. W. (2000). "Developmental physiology, animal models, and the August Krogh principle." Zoology **102**: 148-156.

Burri, M., Y. Tromvoukis, D. Bopp, G. Frigerio and M. Noll (1989). "Conservation of the paired domain in metazoans and its structure in three isolated human genes." The EMBO Journal **8**(4): 1183-1190.

Butzer, K. W., G. L. Isaac, J. L. Richardson and C. Washbourn-Kamau (1972). "Radiocarbon dating of East African lake levels." Science **175**(4026): 1069-1076.

Cao, W. and Y. Xia (2016). "The first complete mitochondrial genome of the *Pundamilia nyererei*." Mitochondrial DNA Part A **27**(6): 4292-4293.

Cavicchioli, R. (2002). "Extremophiles and the search for extraterrestrial life." Astrobiology **2**(3): 281.

Cellerino, A., D. R. Valenzano and M. Reichard (2016). "From the bush to the bench: the annual Nothobranchius fishes as a new model system in biology." Biological Reviews **91**(2): 511-533.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood and J. Schmutz (2010). "Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer." Science **327**(5963): 302-305.

Chaudhary, D. K., S. K. Singh, N. Gohil and G. Bhattacharjee (2020). Recent progress of CRISPR-Cas9 in zebra fish. Genome Engineering via CRISPR-Cas9 System, Elsevier**:** 251-261.

Chen, L., A. L. DeVries and C.-H. C. Cheng (1997). "Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish." Proceedings of the National Academy of Sciences **94**(8): 3811-3816.

Chen, W.-J., G. Ortí and A. Meyer (2004). "Novel evolutionary relationship among four fish model systems." TRENDS in Genetics **20**(9): 424-431.

Chen, Z., C.-H. C. Cheng, J. Zhang, L. Cao, L. Chen, L. Zhou, Y. Jin, H. Ye, C. Deng and Z. Dai (2008). "Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish." Proceedings of the National Academy of Sciences **105**(35): 12944-12949.

Cheng, C.-H. C., P. A. Cziko and C. W. Evans (2006). "Nonhepatic origin of notothenioid antifreeze reveals pancreatic synthesis as common mechanism in polar fish freezing avoidance." Proceedings of the National Academy of Sciences **103**(27): 10491-10496.

Cheval, L., J. P. Duong Van Huyen, P. Bruneval, J.-M. Verbavatz, J.-M. Elalouf and A. Doucet (2004). "Plasticity of mouse renal collecting duct in response to potassium depletion." Physiological genomics **19**(1): 61-73.

Chi, N. and J. A. Epstein (2002). "Getting your Pax straight: Pax proteins in development and disease." TRENDS in Genetics **18**(1): 41-47.

Childress, J. J. and B. A. Seibel (1998). "Life at stable low oxygen levels: adaptations of animals to oceanic oxygen minimum layers." Journal of Experimental Biology **201**(8): 1223-1232.

Chow, R. L., C. R. Altmann, R. A. Lang and A. Hemmati-Brivanlou (1999). "Pax6 induces ectopic eyes in a vertebrate." Development **126**(19): 4213-4222.

Christoffels, V. M., T. Grange, K. H. Kaestner, T. J. Cole, G. J. Darlington, C. M. Croniger and W. H. Lamers (1998). "Glucocorticoid receptor, C/EBP, HNF3, and protein kinase A

coordinately activate the glucocorticoid response unit of the carbamoylphosphate synthetase I gene." Molecular and Cellular Biology **18**(11): 6305-6315.

Coe, M. J. (1966). "The biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya." Acta trop **23**: 146-177.

Cohen, A. S., J. Halfpenny, M. Lockley and E. Michel (1993). "Modern vertebrate tracks from Lake Manyara, Tanzania and their paleobiological implications." Paleobiology **19**(4): 433-458.

Conerly, M. L., Z. Yao, J. W. Zhong, M. Groudine and S. J. Tapscott (2016). "Distinct activities of Myf5 and MyoD indicate separate roles in skeletal muscle lineage specification and differentiation." Developmental Cell **36**(4): 375-385.

Coyle, S. M., F. A. Huntingford and C. L. Peichel (2007). "Parallel evolution of Pitx1 underlies pelvic reduction in Scottish threespine stickleback (*Gasterosteus aculeatus*)." Journal of Heredity **98**(6): 581-586.

Crawford, D. L. (2001). "Functional genomics does not have to be limited to a few select organisms." Genome Biology **2**(1): interactions1001. 1001.

Cvekl, A., Y. Yang, B. K. Chauhan and K. Cveklova (2004). "Regulation of gene expression by Pax6 in ocular cells: a case of tissue-preferred expression of crystallins in lens." The International journal of developmental biology **48**(8-9): 829.

Darwin, C. (1859). "On the origin of species by means of natural selection." Murray, London.

Datsomor, A. K., R. E. Olsen, N. Zic, A. Madaro, A. M. Bones, R. B. Edvardsen, A. Wargelius and P. Winge (2019). "CRISPR/Cas9-mediated editing of Δ5 and Δ6 desaturases impairs Δ8-desaturation and docosahexaenoic acid synthesis in Atlantic salmon (Salmo salar L.)." Scientific reports **9**(1): 1-13.

De Boeck, G., C. M. Wood, K. V. Brix, A. K. Sinha, V. Matey, O. E. Johannsson, A. Bianchini, L. F. Bianchini, J. N. Maina and G. D. Kavembe (2019). "Fasting in the ureotelic Lake Magadi tilapia, Alcolapia grahami, does not reduce its high metabolic demand, increasing its vulnerability to siltation events." Conservation Physiology **7**(1): coz060.

de Cima, S., L. M. Polo, C. Díez-Fernández, A. I. Martínez, J. Cervera, I. Fita and V. Rubio (2015). "Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis." Scientific Reports **5**: 16950.

De Filippi, F. (1853). Nouvelles espèces de poissons.

Deane, E. E. and N. Y. Woo (2006). "Tissue distribution, effects of salinity acclimation, and ontogeny of aquaporin 3 in the marine teleost, silver sea bream (*Sparus sarba*)." Marine Biotechnology **8**(6): 663-671.

Devoto, S. H., E. Melançon, J. S. Eisen and M. Westerfield (1996). "Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation." Development **122**(11): 3371-3380.

DeVries, A. L. (1971). Freezing resistance in fishes. Fish physiology. New York and London, Academic Press. **6:** 157-190.

Dhillon, R. S., M. Mandic, L. Yao, Z.-D. Cao, S.-J. Fu, C. J. Brauner, Y. S. Wang and J. G. Richards (2018). "Ethanol metabolism varies with hypoxia tolerance in ten cyprinid species." Journal of Comparative Physiology B **188**(2): 283-293.

Dräger, U. C. and G. W. Balkema (1987). "Does melanin do more than protect from light?" Neuroscience Research Supplements **6**: S75-S86.

Draper, B. W., P. A. Morcos and C. B. Kimmel (2001). "Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown." Genesis **30**(3): 154-156.

Duan, D., Y. Fu, G. Paxinos and C. Watson (2013). "Spatiotemporal expression patterns of Pax6 in the brain of embryonic, newborn, and adult mice." Brain Structure and Function **218**(2): 353-372.

Duboué, E. R., A. C. Keene and R. L. Borowsky (2011). "Evolutionary convergence on sleep loss in cavefish populations." Current Biology **21**(8): 671-676.

Dybowski, B. (1869). "Fishes of the Baikal system." Izvestiya Sibirskogo Otdeleniya Russkogo Geografi cheskogo Obshchtva **7**: 1-25.

Eastman, J. T. and A. L. DeVries (1986). "Antarctic fishes." Scientific American **255**(5): 106-114A.

Elipot, Y., H. Hinaux, J. Callebert, J.-M. Launay, M. Blin and S. Rétaux (2014). "A mutation in the enzyme monoamine oxidase explains part of the Astyanax cavefish behavioural syndrome." Nature communications **5**: 3647.

Eugster, H. (1986). "Lake Magadi, Kenya: a model for rift valley hydrochemistry and sedimentation?" Geological Society, London, Special Publications **25**(1): 177-189.

Evans, D. H., P. M. Piermarini and K. P. Choe (2005). "The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste." Physiological Reviews **85**(1): 97-177.

Fagernes, C. E., K.-O. Stensløkken, Å. K. Røhr, M. Berenbrink, S. Ellefsen and G. E. Nilsson (2017). "Extreme anoxia tolerance in crucian carp and goldfish through neofunctionalization of duplicated genes creating a new ethanol-producing pyruvate decarboxylase pathway." Scientific Reports **7**(1): 7884.

FAO. (2020). "Fisheries and Aquaculture Department, species fact sheet: *Oreochromis niloticus*." Retrieved 16/04/2020, 2020, from http://www.fao.org/fishery/species/3217/en.

Feiner, N., A. Meyer and S. Kuraku (2014). "Evolution of the vertebrate Pax4/6 class of genes with focus on its novel member, the Pax10 gene." Genome biology and evolution **6**(7): 1635-1651.

Felice, V., M. A. Visconti and E. Trajano (2008). "Mechanisms of pigmentation loss in subterranean fishes." Neotropical Ichthyology **6**(4): 657-662.

Fiol, D. F. and D. Kültz (2007). "Osmotic stress sensing and signaling in fishes." The FEBS journal **274**(22): 5790-5798.

Fisher, R. A. (1930). The genetical theory of natural selection, Oxford: Clarendon Press.

Fong, A. P., Z. Yao, J. W. Zhong, Y. Cao, W. L. Ruzzo, R. C. Gentleman and S. J. Tapscott (2012). "Genetic and epigenetic determinants of neurogenesis and myogenesis." Developmental Cell **22**(4): 721-735.

Ford, A. (2015). Evolutionary relationships of East African soda lake cichlid fish, UCL (University College London).

Ford, A. G. P., T. R. Bullen, L. Pang, M. J. Genner, R. Bills, T. Flouri, B. P. Ngatunga, L. Rüber, U. K. Schliewen, O. Seehausen, A. Shechonge, M. L. J. Stiassny, G. F. Turner and J. J. Day (2019). "Molecular phylogeny of Oreochromis (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments." Molecular Phylogenetics and Evolution **136**: 215-226.

Ford, A. G. P., K. K. Dasmahapatra, L. Rüber, K. Gharbi, T. Cezard and J. J. Day (2015). "High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment." Molecular Ecology **24**(13): 3421-3440.

Ford, A. G. P., L. Rüber, J. Newton, K. K. Dasmahapatra, J. D. Balarin, K. Bruun and J. J. Day (2016). "Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation." Evolution **70**(12): 2718-2735.

Franklin, C., I. Johnston, T. Crockford and C. Kamunde (1995). "Scaling of oxygen consumption of Lake Magadi tilapia, a fish living at 37 C." Journal of Fish Biology **46**(5): 829-834.

Franklin, C. E., T. Crockford, I. A. Johnston and C. Kamunde (1994). "The thermostability of haemoglobins from the hot-spring fish, *Oreochromis alcalicus grahami*: Comparisons with antarctic and temperate species." Journal of thermal biology **19**(4): 277-280.

Galant, R. and S. B. Carroll (2002). "Evolution of a transcriptional repression domain in an insect Hox protein." Nature **415**(6874): 910-913.

Garbarino, V. R., M. E. Orr, K. A. Rodriguez and R. Buffenstein (2015). "Mechanisms of oxidative stress resistance in the brain: lessons learned from hypoxia tolerant extremophilic vertebrates." Archives of biochemistry and biophysics **576**: 8-16.

Gehring, W. J. (2003). "The genetic control of eye development and its implications for the evolution of the various eye-types." International Journal of Developmental Biology **46**(1): 65-73.

Gerber, H.-P., K. Seipel, O. Georgiev, M. Hofferer, M. Hug, S. Rusconi and W. Schaffner (1994). "Transcriptional activation modulated by homopolymeric glutamine and proline stretches." Science **263**(5148): 808-811.

Giles, N. (1983). "The possible role of environmental calcum levels during the evolution of phenotypic diversity in Outer Hebridean populations of the Three-spined stickleback, *Gasterosteus aculeatus*." Journal of zoology **199**(4): 535-544.

Gilmour, K. and S. Perry (2009). "Carbonic anhydrase and acid–base regulation in fish." Journal of Experimental Biology **212**(11): 1647-1661.

Glardon, S., P. Callaerts, G. Halder and W. J. Gehring (1997). "Conservation of Pax-6 in a lower chordate, the ascidian *Phallusia mammillata*." Development **124**(4): 817-825.

Glasauer, S. M. and S. C. Neuhauss (2014). "Whole-genome duplication in teleost fishes and its evolutionary consequences." Molecular genetics and genomics **289**(6): 1045-1060.

Glaser, T., D. S. Walton and R. L. Maas (1992). "Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene." Nature genetics **2**(3): 232.

Gonzalez, R. (2012). "The physiology of hyper-salinity tolerance in teleost fish: a review." Journal of Comparative Physiology B **182**(3): 321-329.

Goode, G. B. and T. H. Bean (1882). "A list of the species of fishes recorded as occurring in the Gulf of Mexico." Proceedings of the United States National Museum.

Grant, P. R., B. R. Grant, R. B. Huey, M. T. Johnson, A. H. Knoll and J. Schmitt (2017). "Evolution caused by extreme events." Philosophical Transactions of the Royal Society B: Biological Sciences **372**(1723): 20160146.

Grant, W. (2006). Alkaline environments and biodiversity, in extremophiles. Encyclopaedia of life support systems (EOLSS). G. Charles and G. Nicolas, Oxford, UK, Eolss Publishers.

Grosell, M. (2006). "Intestinal anion exchange in marine fish osmoregulation." Journal of Experimental Biology **209**(15): 2813-2827.

Gross, J. B., R. Borowsky and C. J. Tabin (2009). "A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*." PLoS genetics **5**(1): e1000326.

Gross, J. B., B. Meyer and M. Perkins (2015). "The rise of Astyanax cavefish." Developmental Dynamics **244**(9): 1031-1038.

Gross, J. B. and H. Wilkens (2013). "Albinism in phylogenetically and geographically distinct populations of Astyanax cavefish arises through the same loss-of-function Oca2 allele." Heredity **111**(2): 122.

Gruss, P. and C. Walther (1992). "Pax in development." Cell **69**(5): 719-722.

Günther, A. (1877). "LVI.—Preliminary notes on new fishes collected in Japan during the expedition of HMS 'Challenger'." Journal of Natural History **20**(119): 433-446.

Haerty, W. and G. B. Golding (2010). "Genome-wide evidence for selection acting on single amino acid repeats." Genome research **20**(6): 755-760.

Haldane, J. (1933). "The part played by recurrent mutation in evolution." The American Naturalist **67**(708): 5-19.

Hamilton, F. (1822). An account of the fishes found in the river Ganges and its branches, Bishen Singh Mahendra Pal Singh.

Handy, R. and M. Poxton (1993). "Nitrogen pollution in mariculture: toxicity and excretion of nitrogenous compounds by marine fish." Reviews in Fish Biology and Fisheries **3**(3): 205-241.

Harding, M. M., P. I. Anderberg and A. Haymet (2003). "'Antifreeze'glycoproteins from polar fish." European journal of biochemistry **270**(7): 1381-1392.

Hasler, C. T., D. Butman, J. D. Jeffrey and C. D. Suski (2016). "Freshwater biota and rising pCO 2?" Ecology Letters **19**(1): 98-108.

Hedges, S. B. (2002). "The origin and evolution of model organisms." Nature Reviews Genetics **3**(11): 838-849.

Hegde, S., P. Nilyanimit, E. Kozlova, E. R. Anderson, H. P. Narra, S. K. Sahni, E. Heinz and G. L. Hughes (2019). "CRISPR/Cas9-mediated gene deletion of the ompA gene in symbiotic Cedecea neteri impairs biofilm formation and reduces gut colonization of Aedes aegypti mosquitoes." PLoS neglected tropical diseases **13**(12): e0007883.

Heindel, J. J. and L. N. Vandenberg (2015). "Developmental origins of health and disease: a paradigm for understanding disease etiology and prevention." Current opinion in pediatrics **27**(2): 248.

Helfman, G., B. B. Collette, D. E. Facey and B. W. Bowen (2009). The diversity of fishes: biology, evolution, and ecology, John Wiley & Sons.

Heller, N. and A. W. Brändli (1999). "*Xenopus* Pax-2/5/8 orthologues: novel insights into Pax gene evolution and identification of Pax-8 as the earliest marker for otic and pronephric cell lineages." Developmental genetics **24**(3-4): 208-219.

Herrera, M. and P. Jagadeeswaran (2004). "Annual fish as a genetic model for aging." The Journals of Gerontology Series A: Biological Sciences

Medical Sciences **59**(2): B101-B107.

Hilgendorf, F. (1905). "Fische von Deutsch und Englisch Ost-Africa." Zoologische Jahrbücher, Abteilung für Systematik, Geographie und Biologie der Tiere (Jena) **22**: 405-420.

Hirata, T., T. Kaneko, T. Ono, T. Nakazato, N. Furukawa, S. Hasegawa, S. Wakabayashi, M. Shigekawa, M.-H. Chang and M. F. Romero (2003). "Mechanism of acid adaptation of a fish living in a pH 3.5 lake." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **284**(5): R1199-R1212.

Hoegg, S., H. Brinkmann, J. S. Taylor and A. Meyer (2004). "Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish." Journal of Molecular Evolution **59**(2): 190-203.

Holland, P. W. (2013). "Evolution of homeobox genes." Wiley Interdisciplinary Reviews: Developmental Biology **2**(1): 31-45.

Holland, P. W., J. Garcia-Fernàndez, N. A. Williams and A. Sidow (1994). "Gene duplications and the origins of vertebrate development." Development **1994**(Supplement): 125-133.

Holmes, R. M., A. Aminot, R. Kérouel, B. A. Hooker and B. J. Peterson (1999). "A simple and precise method for measuring ammonium in marine and freshwater ecosystems." Canadian Journal of Fisheries and Aquatic Sciences **56**(10): 1801-1808.

Hong, J., W. L. Salo, C. J. Lusty and P. M. Anderson (1994). "Carbamyl phosphate synthetase III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl phosphate synthetases." Molecular Biology **243**: 131-140.

Hoshijima, K., M. J. Jurynec, D. K. Shaw, A. M. Jacobi, M. A. Behlke and D. J. Grunwald (2019). "Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish." Developmental cell **51**(5): 645-657. e644.

Hsu, C. Y., Y. C. Chiu, W. L. Hsu and Y. P. Chan (2008). "Age-related markers assayed at different developmental stages of the annual fish *Nothobranchius rachovii*." The Journals of Gerontology Series A: Biological Sciences **63**(12): 1267-1276.

Huang, C.-H. and M. Ye (2010). "The Rh protein family: gene evolution, membrane biology, and disease association." Cell. Mol. Life Sci **67**: 1203-1218.

Huang, P., A. Xiao, M. Zhou, Z. Zhu, S. Lin and B. Zhang (2011). "Heritable gene targeting in zebrafish using customized TALENs." Nature biotechnology **29**(8): 699-700.

Hughes, G. (1984). General anatomy of the gills. Fish Physiology, Elsevier. **10:** 1-72.

Hughes, G. (1995). "Preliminary morphometric study of the gills of *Oreochromis alcalicus grahami* from Lake Magadi and a comparison with *O. niloticus*." Journal of fish biology **47**(6): 1102-1105.

Hung, C., K. Tsui, J. Wilson, C. Nawata, C. Wood and P. Wright (2007). "Rhesus glycoprotein gene expression in the mangrove killifish *Kryptolebias marmoratus* exposed to elevated environmental ammonia levels and air." Journal of Experimental Biology **210**(14): 2419-2429.

Hunker, A. C., M. E. Soden, D. Krayushkina, G. Heymann, R. Awatramani and L. S. Zweifel (2020). "Conditional single vector CRISPR/SaCas9 viruses for efficient mutagenesis in the adult mouse nervous system." Cell reports **30**(12): 4303-4316. e4306.

Hwang, P.-P., T.-H. Lee and L.-Y. Lin (2011). "Ion regulation in fish gills: recent progress in the cellular and molecular mechanisms." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **301**(1): R28-R47.

Hwang, W. Y., Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. J. Yeh and J. K. Joung (2013). "Efficient genome editing in zebrafish using a CRISPR-Cas system." Nature biotechnology **31**(3): 227-229.

Hyvärinen, H., I. Holopainen and J. Piironen (1985). "Anaerobic wintering of crucian carp (*Carassius carassius* L.)-I. Annual dynamics of glycogen reserves in nature." Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology **82**(4): 797-803.

Ip, A. Y. and S. F. Chew (2010). "Ammonia production, excretion, toxicity, and defense in fish: a review." Frontiers in Physiology **1**: 134.

Ip, Y. K., A. M. Loong, Y. R. Chng, K. C. Hiong and S. F. Chew (2012). "Hepatic carbamoyl phosphate synthetase (CPS) I and urea contents in the hylid tree frog, *Litoria caerulea:* transition from CPS III to CPS I." Journal of Comparative Physiology B **182**(8): 1081-1094.

Irwin, J. A. (2010). "Extremophiles and their application to veterinary medicine." Environmental technology **31**(8-9): 857-869.

Jackson, S. T., J. L. Betancourt, R. K. Booth and S. T. Gray (2009). "Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions." Proceedings of the National Academy of Sciences **106**(Supplement 2): 19685-19692.

Jao, L.-E., S. R. Wente and W. Chen (2013). "Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system." Proceedings of the National Academy of Sciences **110**(34): 13904-13909.

Jeffery, W. R. (2009). "Regressive evolution in Astyanax cavefish." Annual review of genetics **43**: 25-47.

Jeffery, W. R. (2019). *Astyanax mexicanus*: A vertebrate model for evolution, adaptation, and development in caves. Encyclopedia of Caves, Elsevier**:** 85-93.

Jeffery, W. R., A. G. Strickler and Y. Yamamoto (2003). "To see or not to see: evolution of eye degeneration in Mexican blind cavefish." Integrative and Comparative Biology **43**(4): 531-541.

Johannsson, O., H. Bergman, C. Wood, P. Laurent, D. Kavembe, A. Bianchini, J. Maina, C. Chevalier, L. Bianchini and M. Papah (2014). "Air breathing in Magadi tilapia *Alcolapia grahami*, under normoxic and hyperoxic conditions, and the association with sunlight and reactive oxygen species." Journal of Fish Biology **84**(3): 844-863.

Johnston, I. A. and L. M. Bernard (1983). "Utilization of the ethanol pathway in carp following exposure to anoxia." Journal of Experimental Biology **104**(1): 73-78.

Jones, B. F., H. P. Eugster and S. L. Rettig (1977). "Hydrochemistry of the lake Magadi basin, Kenya." Geochimica et Cosmochimica Acta **41**(1): 53-72.

Jorge, C. D., N. Borges, I. Bagyan, A. Bilstein and H. Santos (2016). "Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare." Extremophiles **20**(3): 251-259.

Jubb, R. A. (1971). "A new Nothobranchius (Pisces, Cyprinodontidae) from Southeastern Rhodesia." Journal of the American Killifish Association **8**(12): 314-321.

Kaneko, T., S. Hasegawa, K. Uchida, T. Ogasawara, A. Oyagi and T. Hirano (1999). "Acid tolerance of Japanese dace (a cyprinid teleost) in Lake Osorezan, a remarkable acid lake." Zoological Science **16**(6): 871-877.

Kashi, Y. and D. G. King (2006). "Simple sequence repeats as advantageous mutators in evolution." TRENDS in Genetics **22**(5): 253-259.

Kaufman, A., M. Margaritz, M. Paul, C. Hillaire-Marcel, G. Hollos, E. Boaretto and M. Taieb (1990). "The 36C1 ages of the brines in the Magadi-Natron basin, East Africa." Geochimica et Cosmochimica Acta **54**(10): 2827-2833.

Kavembe, G. D., P. Franchini, I. Irisarri, G. Machado-Schiaffino and A. Meyer (2015). "Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one of earth's most extreme environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa." Journal of Molecular Evolution **81**(3-4): 90-109.

Khademi, S., J. O'connell, J. Remis, Y. Robles-Colmenares, L. J. Miercke and R. M. Stroud (2004). "Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å." Science **305**(5690): 1587-1594.

Khan, S. H. (2019). "Genome-editing technologies: concept, pros, and cons of various genome-editing techniques and bioethical concerns for clinical application." Molecular therapy. Nucleic acids **16**: 326.

King, R. A., J. Pietsch, J. P. Fryer, S. Savage, M. J. Brott, I. Russell-Eggitt, C. G. Summers and W. S. Oetting (2003). "Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype." Human genetics **113**(6): 502-513.

Klaassen, H., Y. Wang, K. Adamski, N. Rohner and J. E. Kowalko (2018). "CRISPR mutagenesis confirms the role of oca2 in melanin pigmentation in Astyanax mexicanus." Developmental Biology **441**(2): 313-318.

Kleinjan, D. A., R. M. Bancewicz, P. Gautier, R. Dahm, H. B. Schonthaler, G. Damante, A. Seawright, A. M. Hever, P. L. Yeyati and V. van Heyningen (2008). "Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence." PLoS genetics **4**(2).

Kong, H., D. D. Edberg, J. J. Korte, W. L. Salo, P. A. Wright and P. M. Anderson (1998). "Nitrogen excretion and expression of carbamoyl-phosphate synthetase III activity and mRNA in extrahepatic tissues of largemouth bass (*Micropterus salmoides*)." Archives of Biochemistry and Biophysics **350**(2): 157-168.

Korte, J. J., W. L. Salo, V. M. Cabrera, P. A. Wright, A. K. Felskie and P. M. Anderson (1997). "Expression of carbamoyl-phosphate synthetase III mRNA during the early stages of development and in muscle of adult rainbow trout (*Oncorhynchus mykiss*)." Journal of Biological Chemistry **272**(10): 6270-6277.

Kosiol, C. and M. Anisimova (2019). Selection acting on genomes. Evolutionary Genomics, Springer**:** 373-397.

Krauss, S., T. Johansen, V. Korzh and A. Fjose (1991). "Expression pattern of zebrafish pax genes suggests a role in early brain regionalization." Nature **353**(6341): 267.

Krogh, A. (1929). "The progress of physiology." American Journal of Physiology-Legacy Content **90**(2): 243-251.

Kültz, D., K. Jürss and L. Jonas (1995). "Cellular and epithelial adjustments to altered salinity in the gill and opercular epithelium of a cichlid fish (*Oreochromis mossambicus*)." Cell and Tissue Research **279**(1): 65-73.

Kumar, A. S., D. T. Sowpati and R. K. Mishra (2016). "Single amino acid repeats in the proteome world: structural, functional, and evolutionary insights." PLoS One **11**(11): e0166854.

Kumar, J. P. and K. Moses (2001). Eye specification in *Drosophila*: perspectives and implications. Seminars in cell & developmental biology, Elsevier.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura (2018). "MEGA X: molecular evolutionary genetics analysis across computing platforms." Molecular Biology and Evolution **35**(6): 1547-1549.

Kwong, R. W., Y. Kumai and S. F. Perry (2014). "The physiology of fish at low pH: the zebrafish as a model system." Journal of Experimental Biology **217**(5): 651-662.

Laksanalamai, P. and F. T. Robb (2004). "Small heat shock proteins from extremophiles: a review." Extremophiles **8**(1): 1-11.

Laurent, P. and S. F. Perry (1991). "Environmental effects on fish gill morphology." Physiological Zoology **64**(1): 4-25.

Laverty, G. and E. Skadhauge (2012). "Adaptation of teleosts to very high salinity." Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology **163**(1): 1-6.

LeMoine, C. M. and P. J. Walsh (2013). "Ontogeny of ornithine-urea cycle gene expression in zebrafish (*Danio rerio*)." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **304**(11): R991-R1000.

Leonelli, S. and R. Ankeny (2013). "What makes a model organism?" Endeavour **37**(4): 209.

Levinson, G. and G. A. Gutman (1987). "Slipped-strand mispairing: a major mechanism for DNA sequence evolution." Molecular biology and evolution **4**(3): 203-221.

Li, M., H. Yang, J. Zhao, L. Fang, H. Shi, M. Li, Y. Sun, X. Zhang, D. Jiang and L. Zhou (2014). "Efficient and heritable gene targeting in tilapia by CRISPR/Cas9." Genetics **197**(2): 591-599.

Li, W., Y. T. Chen, Y. Hayashida, G. Blanco, A. Kheirkah, H. He, S. Y. Chen, C. Y. Liu and S. Tseng (2008). "Down-regulation of Pax6 is associated with abnormal differentiation of corneal epithelial cells in severe ocular surface diseases." The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland **214**(1): 114-122.

Lindley, T. E., C. L. Scheiderer, P. J. Walsh, C. M. Wood, H. L. Bergman, A. L. Bergman, P. Laurent, P. Wilson and P. M. Anderson (1999). "Muscle as the primary site of urea cycle enzyme activity in an alkaline lake-adapted tilapia, *Oreochromis alcalicus grahami*." Journal of Biological Chemistry **274**(42): 29858-29861.

Linnaeus, C. V. (1758). "Systema Naturae, edition X, vol. 1 (Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata)." Holmiae Salvii **824**.

Litscher, E. S. and P. M. Wassarman (2018). The fish Egg's zona Pellucida. Current topics in developmental biology, Elsevier. **130:** 275-305.

Liu, S., Z. Zheng, S. Ji, T. Liu, Y. Hou, S. Li and G. Li (2018). "Resveratrol reduces senescence-associated secretory phenotype by SIRT1/NF-κB pathway in gut of the annual fish *Nothobranchius guentheri*." Fish shellfish immunology **80**: 473-479.

Liu, T., S. Liu, L. Ma, F. Li, Z. Zheng, R. Chai, Y. Hou, Y. Xie and G. Li (2017). "Oogenesis, vitellogenin-mediated ovarian degeneration and immune response in the annual fish *Nothobranchius guentheri*." Fish shellfish immunology **66**: 86-92.

Logan, M. and C. J. Tabin (1999). "Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity." Science **283**(5408): 1736-1739.

Loong, A., Y. Chng, S. Chew, W. Wong and Y. Ip (2012). "Molecular characterization and mRNA expression of carbamoyl phosphate synthetase III in the liver of the African lungfish, *Protopterus annectens*, during aestivation or exposure to ammonia." Journal of Comparative Physiology B **182**(3): 367-379.

Loosli, F., M. Kmita-Cunisse and W. J. Gehring (1996). "Isolation of a Pax-6 homolog from the ribbonworm *Lineus sanguineus*." Proceedings of the National Academy of Sciences **93**(7): 2658-2663.

López, J. M., R. Morona, N. Moreno, D. Lozano, S. Jiménez and A. González (2020). "Pax6 expression highlights regional organization in the adult brain of lungfishes, the closest living relatives of land vertebrates." Journal of Comparative Neurology **528**(1): 139-163.

Lowe, C. H. and W. G. Heath (1969). "Behavioral and physiological responses to temperature in the desert pupfish *Cyprinodon macularius*." Physiological Zoology **42**(1): 53-59.

Lucas-Sánchez, A., P. F. Almaida-Pagán, P. Mendiola and J. de Costa (2014). "*Nothobranchius* as a model for aging studies. A review." Aging disease **5**(4): 281.

Lushchak, V. I. (2014). "Free radicals, reactive oxygen species, oxidative stress and its classification." Chemico-Biological Interactions **224**: 164-175.

Lushchak, V. I. and T. V. Bagnyukova (2006). "Effects of different environmental oxygen levels on free radical processes in fish." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **144**(3): 283-289.

Lykkeboe, G., K. Johansen and G. Maloiy (1975). "Functional properties of hemoglobins in the teleost *Tilapia grahami*." Journal of comparative physiology **104**(1): 1-11.

Lynch, M. and J. S. Conery (2000). "The evolutionary fate and consequences of duplicate genes." science **290**(5494): 1151-1155.

MacKintosh, C. and D. E. Ferrier (2017). "Recent advances in understanding the roles of whole genome duplications in evolution." F1000Research **6**.

Macqueen, D. J. and I. A. Johnston (2006). "A novel salmonid myoD gene is distinctly regulated during development and probably arose by duplication after the genome tetraploidization." FEBS letters **580**(21): 4996-5002.

Macqueen, D. J. and I. A. Johnston (2008). "An update on MyoD evolution in teleosts and a proposed consensus nomenclature to accommodate the tetraploidization of different vertebrate genomes." PLoS One **3**(2): e1567.

Maina, J. (1990). "A study of the morphology of the gills of an extreme alkalinity and hyperosmotic adapted teleost *Oreochromis alcalicus grahami* (Boulenger) with particular emphasis on the ultrastructure of the chloride cells and their modifications with water dilution." Anatomy and Embryology **181**(1): 83-98.

Maina, J. (1991). "A morphometric analysis of chloride cells in the gills of the teleosts *Oreochromis alcalicus* and *Oreochromis niloticu*s and a description of presumptive urea-excreting cells in *O. alcalicus*." Journal of anatomy **175**: 131.

Maina, J. (2000). "Functional morphology of the gas-gland cells of the air-bladder of *Oreochromis alcalicus grahami* (Teleostei: Cichlidae): an ultrastructural study on a fish adapted to a severe, highly alkaline environment." Tissue and Cell **32**(2): 117-132.

Maina, J. (2000). "The highly specialized secretory epithelium in the buccal cavity of the alkalinity adapted Lake Magadi cichlid, *Oreochromis alcalicus grahami* (Teleostei: Cichlidae): a scanning and transmission electron microscope study." Journal of Zoology **251**(4): 427-438.

Maina, J., S. Kisia, C. Wood, A. Narahara, H. Bergman, P. Laurent and P. Walsh (1996). "A comparative allometric study of the morphometry of the gills of an alkalinity adapted cichlid fish, *Oreochromis alcalicus grahami,* of Lake Magadi, Kenya." International Journal of Salt Lake Research **5**(2): 131.

Mandic, M. and M. D. Regan (2018). "Can variation among hypoxic environments explain why different fish species use different hypoxic survival strategies?" Journal of Experimental Biology **221**(21): jeb161349.

Manel, S., C. Perrier, M. Pratlong, L. Abi-Rached, J. Paganini, P. Pontarotti and D. Aurelle (2016). "Genomic resources and their influence on the detection of the signal of positive selection in genome scans." Molecular ecology **25**(1): 170-184.

Marchinko, K. B. (2009). "Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback." Evolution: International Journal of Organic Evolution **63**(1): 127-138.

Marini, A.-M., G. Matassi, V. Raynal, B. André, J.-P. Cartron and B. Chérif-Zahar (2000). "The human Rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast." Nature genetics **26**(3): 341-344.

Marini, A.-M., S. Soussi-Boudekou, S. Vissers and B. Andre (1997). "A family of ammonium transporters in Saccharomyces cerevisiae." Molecular and cellular biology **17**(8): 4282-4293.

Markofsky, J. and M. Milstoc (1979). "Aging changes in the liver of the male annual cyprinodont fish, *Nothobranchius guentheri*." Experimental gerontology **14**(1): 11-IN16.

Masiko, K. (1940). "Limnological study of Lake Osorezan-ko, a remarkable acidotrophic lake in Japan." Sci. Rep. Tohoku Imp. Univ, 4th Ser, Biology **15**: 331-356.

Matsui, H., N. Kenmochi and K. Namikawa (2019). "Age-and α-Synuclein-Dependent Degeneration of Dopamine and Noradrenaline Neurons in the Annual Killifish *Nothobranchius furzeri*." Cell reports **26**(7): 1727-1733. e1726.

McCue, M. D. (2004). General effects of temperature on animal biology. Temperature dependent sex determination. Washington DC, Smithsonian books**:** 71-78.

McDonald, M. D. (2003). "Fish at home in acid lake." Journal of Experimental Biology **206**(11): 1769-1769.

McLysaght, A. and L. D. Hurst (2016). "Open questions in the study of de novo genes: what, how and why." Nature Reviews Genetics **17**(9): 567.

McQueen, C. and M. E. Pownall (2017). "An analysis of MyoD-dependent transcription using CRISPR/Cas9 gene targeting in *Xenopus tropicalis* embryos." Mechanisms of Development **146**: 1-9.

Mendel, G. (1866). "Versuche über Pflanzenhybriden." Verhandlungen des naturforschenden Vereines in Brunn 4: 3 **44**.

Meyer, A. and M. Schartl (1999). "Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions." Current opinion in cell biology **11**(6): 699-704.

Meyers, J. R. (2018). "Zebrafish: development of a vertebrate model organism." Current Protocols Essential Laboratory Techniques **16**(1): e19.

Mikhodyuk, O., G. Zavarzin and R. Ivanovsky (2008). "Transport systems for carbonate in the extremely natronophilic cyanobacterium *Euhalothece sp*." Microbiology **77**(4): 412-418.

Mitchell, R. K., M.-S. Nguyen-Tu, P. Chabosseau, R. M. Callingham, T. J. Pullen, R. Cheung, I. Leclerc, D. J. Hodson and G. A. Rutter (2017). "The transcription factor Pax6 is required for pancreatic β cell identity, glucose-regulated ATP synthesis, and Ca2+ dynamics in adult mice." Journal of Biological Chemistry **292**(21): 8892-8906.

Mommsen, T. P. and P. J. Walsh (1989). "Evolution of urea synthesis in vertebrates: the piscine connection." Science **243**(4887): 72-75.

Mommsen, T. P. and P. J. Walsh (1991). Urea synthesis in fishes: evolutionary and biochemical perspectives, Elsevier.

Monné, M., L. Han and L. Jovine (2006). Tracking down the ZP domain: from the mammalian zona pellucida to the molluscan vitelline envelope. Seminars in reproductive medicine, Copyright© 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New ….

Mulvenna, P. F. and G. Savidge (1992). "A modified manual method for the determination of urea in seawater using diacetylmonoxime reagent." Estuarine, Coastal and Shelf Science **34**(5): 429-438.

Mulvey, J. M. and G. M. J. J. o. E. B. Renshaw (2000). "Neuronal oxidative hypometabolism in the brainstem of the epaulette shark (*Hemiscyllium ocellatum*) in response to hypoxic pre-conditioning." Neuroscience Letters **290**(1): 1-4.

Nakada, T., K. Hoshijima, M. Esaki, S. Nagayoshi, K. Kawakami and S. Hirose (2007). "Localization of ammonia transporter Rhcg1 in mitochondrion-rich cells of yolk sac, gill, and kidney of zebrafish and its ionic strength-dependent expression." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **293**(4): R1743-R1753.

Nakada, T., C. M. Westhoff, A. Kato and S. Hirose (2007). "Ammonia secretion from fish gill depends on a set of Rh glycoproteins." The FASEB Journal **21**(4): 1067-1074.

Nakayama, T., I. L. Blitz, M. B. Fish, A. O. Odeleye, S. Manohar, K. W. Cho and R. M. Grainger (2014). Cas9-based genome editing in Xenopus tropicalis. Methods in enzymology, Elsevier. **546:** 355-375.

Nakayama, T., M. Fisher, K. Nakajima, A. O. Odeleye, K. B. Zimmerman, M. B. Fish, Y. Yaoita, J. L. Chojnowski, J. D. Lauderdale and P. A. Netland (2015). "*Xenopus* pax6 mutants affect eye development and other organ systems, and have phenotypic similarities to human aniridia patients." Developmental Biology **408**(2): 328-344.

Narahara, A., H. L. Bergman, P. Laurent, J. Maina, P. J. Walsh and C. M. Wood (1996). "Respiratory physiology of the Lake Magadi tilapia (*Oreochromis alcalicus grahami*), a fish adapted to a hot, alkaline, and frequently hypoxic environment." Physiological Zoology: 1114-1136.

Narahara, A., J. Maina, C. Wood, P. Walsh, H. Bergman and P. Laurent (1994). "Urea production, acid-base regulation and their interactions in the Lake Magadi tilapia, a unique teleost adapted to a highly alkaline environment." Journal of Experimental Biology **189**: 13-36.

Nath, A., R. Chaube and K. Subbiah (2013). "An insight into the molecular basis for convergent evolution in fish antifreeze Proteins." Computers in Biology and Medicine **43**(7): 817-821.

Nawata, C. M. and C. M. Wood (2009). "mRNA expression analysis of the physiological responses to ammonia infusion in rainbow trout." Journal of Comparative Physiology B **179**(7): 799-810.

Neubüser, A., H. Koseki and R. Balling (1995). "Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1." Developmental Biology **170**(2): 701-716.

Nilsson, G. E. (1988). "A comparative study of aldehyde dehydrogenase and alcohol dehydrogenase activities in crucian carp and three other vertebrates: apparent adaptations to ethanol production." Journal of Comparative Physiology B **158**(4): 479-485.

Nilsson, G. E. (1990). "Long-term anoxia in crucian carp: changes in the levels of amino acid and monoamine neurotransmitters in the brain, catecholamines in chromaffin tissue, and liver glycogen." Journal of Experimental Biology **150**(1): 295-320.

Nilsson, G. E. and G. M. Renshaw (2004). "Hypoxic survival strategies in two fishes: extreme anoxia tolerance in the North European crucian carp and natural hypoxic preconditioning in a coral-reef shark." Journal of Experimental Biology **207**(18): 3131-3139.

Noll, M. (1993). "Evolution and role of Pax genes." Current opinion in Genetics & Development **3**(4): 595-605.

Nornes, S., M. Clarkson, I. Mikkola, M. Pedersen, A. Bardsley, J. P. Martinez, S. Krauss and T. Johansen (1998). "Zebrafish contains two Pax6 genes involved in eye development1." Mechanisms of Development **77**(2): 185-196.

Ohno, S. (1970). Evolution by gene duplication, Springer Science & Business Media.

Oja, S. S., P. Saransaari and E. R. Korpi (2016). "Neurotoxicity of Ammonia." Neurochemical Research: 1-8.

Onai, T., N. Irie and S. Kuratani (2014). "The evolutionary origin of the vertebrate body plan: the problem of head segmentation." Annual Review of Genomics and Human Genetics **15**: 443-459.

Pallas, P. S. (1811). Zoographia Rosso-Asiatica: sistens omnium animalium in extenso Imperio Rossico et adiacentibus maribus observatorum recensionem, domicilia, mores et descriptiones anatomen atque icones plurimorum., Caes.

Papah, M., S. Kisia, R. Ojoo, A. Makanya, C. Wood, G. Kavembe, J. Maina, O. Johannsson, H. Bergman and P. Laurent (2013). "Morphological evaluation of spermatogenesis in Lake Magadi tilapia (Alcolapia grahami): a fish living on the edge." Tissue and cell **45**(6): 371-382.

Pâques, F., W.-Y. Leung and J. E. Haber (1998). "Expansions and contractions in a tandem repeat induced by double-strand break repair." Molecular and cellular biology **18**(4): 2045-2054.

Patel, M. S., N. S. Nemeria, W. Furey and F. Jordan (2014). "The pyruvate dehydrogenase complexes: structure-based function and regulation." Journal of Biological Chemistry **289**(24): 16615-16623.

Pelster, B., M. Giacomin, C. M. Wood and A. L. Val (2016). "Improved ROS defense in the swimbladder of a facultative air-breathing erythrinid fish, jeju, compared to a non-air-breathing close relative, traira." Journal of Comparative Physiology B **186**(5): 615-624.

Persi, E., Y. I. Wolf and E. V. Koonin (2016). "Positive and strongly relaxed purifying selection drive the evolution of repeats in proteins." Nature communications **7**(1): 1-11.

Peters, H., U. Doll and J. Niessing (1995). "Differential expression of the chicken Pax-1 and Pax-9 gene: in situ hybridization and immunohistochemical analysis." Developmental Dynamics **203**(1): 1-16.

Philippe, C., P. Hautekiet, A. F. Grégoir, E. S. Thoré, T. Pinceel, R. Stoks, L. Brendonck and G. D. Boeck (2018). "Combined effects of cadmium exposure and temperature on the annual killifish (*Nothobranchius furzeri*)." Environmental Toxicology Chemistry **37**(9): 2361-2371.

Pichaud, F., J. Treisman and C. Desplan (2001). "Reinventing a common strategy for patterning the eye." Cell **105**(1): 9-12.

Pinsky, M. L., A. M. Eikeset, D. J. McCauley, J. L. Payne and J. M. Sunday (2019). "Greater vulnerability to warming of marine versus terrestrial ectotherms." Nature **569**(7754): 108-111.

Pörtner, H.-O., P. M. Schulte, C. Wood and F. Schiemer (2010). "Niche dimensions in fishes: an integrative view." Physiological and Biochemical Zoology **83**(5): 808-826.

Pottin, K., H. Hinaux and S. Rétaux (2011). "Restoring eye size in *Astyanax mexicanus* blind cavefish embryos through modulation of the Shh and Fgf8 forebrain organising centres." Development **138**(12): 2467-2476.

Pownall, M. E., M. K. Gustafsson and C. P. Emerson Jr (2002). "Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos." Annual review of Cell and Developmental Biology **18**(1): 747-783.

Priami, C., G. De Michele, F. Cotelli, A. Cellerino, M. Giorgio, P. G. Pelicci and E. Migliaccio (2015). "Modelling the p53/p66Shc aging pathway in the shortest living vertebrate *Nothobranchius furzeri*." Aging Disease **6**(2): 95.

Protas, M. E., C. Hersey, D. Kochanek, Y. Zhou, H. Wilkens, W. R. Jeffery, L. I. Zon, R. Borowsky and C. J. Tabin (2006). "Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism." Nature Genetics **38**(1): 107.

Quinn, J. C., J. D. West and R. E. Hill (1996). "Multiple functions for Pax6 in mouse eye and nasal development." Genes and Development **10**(4): 435-446.

Randall, D., C. Wood, S. Perry, H. Bergman, G. Maloiy, T. Mommsen and P. Wright (1989). "Urea excretion as a strategy for survival in a fish living in a very alkaline environment." Nature **337**: 165-166.

Randall, D. J. and T. Tsui (2002). "Ammonia toxicity in fish." Marine Pollution Bulletin **45**(1-12): 17-23.

Ravi, V., S. Bhatia, P. Gautier, F. Loosli, B.-H. Tay, A. Tay, E. Murdoch, P. Coutinho, V. van Heyningen and S. Brenner (2013). "Sequencing of Pax6 loci from the elephant shark reveals a family of Pax6 genes in vertebrate genomes, forged by ancient duplications and divergences." PLoS Genetics **9**(1): e1003177.

Ravi, V., S. Bhatia, P. Shingate, B.-H. Tay, B. Venkatesh and D. A. Kleinjan (2019). "Lampreys, the jawless vertebrates, contain three Pax6 genes with distinct expression in eye, brain and pancreas." Scientific Reports **9**(1): 1-12.

Reece, J. B., L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky and R. B. Jackson (2014). Campbell biology, Pearson Boston, MA.

Reimchen, T. (1980). "Spine deficiency and polymorphism in a population of *Gasterosteus aculeatus*-an adaptation to predators." Canadian Journal of Zoology **58**: 1232-1244.

Reimchen, T. (1983). "Structural relationships between spines and lateral plates in threespine stickleback (*Gasterosteus aculeatus*)." Evolution **37**(5): 931-946.

Reite, O., G. Maloiy and B. Aasehaug (1974). "pH, salinity and temperature tolerance of Lake Magadi tilapia." Nature **247**(5439): 315.

Rétaux, S. and D. Casane (2013). "Evolution of eye development in the darkness of caves: adaptation, drift, or both?" EvoDevo **4**(1): 26.

Riesch, R., M. Tobler and M. Plath (2015). Extremophile fishes: ecology, evolution, and physiology of teleosts in extreme environments. Switzerland, Springer.

Riley, P. (1997). "Melanin." The international journal of biochemistry cell biology **29**(11): 1235-1239.

Roberts, N., M. Taieb, P. Barker, B. Damnati, M. Icole and D. Williamson (1993). "Timing of the Younger Dryas event in East Africa from lake-level changes." Nature **366**(6451): 146-148.

Rothschild, L. J. and R. L. Mancinelli (2001). "Life in extreme environments." Nature **409**(6823): 1092.

Rubio, V. and J. Cervera (1995). "The carbamoyl-phosphate synthase family and carbamate kinase: structure-function studies." Biochemical Society Transactions **23**(4): 879–883.

Russell, J. J., J. A. Theriot, P. Sood, W. F. Marshall, L. F. Landweber, L. Fritz-Laylin, J. K. Polka, S. Oliferenko, T. Gerbich and A. Gladfelter (2017). "Non-model model organisms." BMC Biology **15**(1): 55.

Sadeh, R., K. Breitschopf, B. Bercovich, M. Zoabi, Y. Kravtsova-Ivantsiv, D. Kornitzer, A. Schwartz and A. Ciechanover (2008). "The N-terminal domain of MyoD is necessary and sufficient for its nuclear localization-dependent degradation by the ubiquitin system." Proceedings of the National Academy of Sciences **105**(41): 15690-15695.

Saeed-Kothe, A. and S. G. Powers-Lee (2002). "Specificity determining residues in ammonia-and glutamine-dependent carbamoyl phosphate synthetases." Journal of Biological Chemistry **277**(9): 7231-7238.

Saha, N., S. Datta, Z. Y. Kharbuli, K. Biswas and A. Bhattacharjee (2007). "Air-breathing catfish, *Clarias batrachus* upregulates glutamine synthetase and carbamyl phosphate

synthetase III during exposure to high external ammonia." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **147**(3): 520-530.

Salzburger, W., B. Van Bocxlaer and A. S. Cohen (2014). "Ecology and Evolution of the African Great Lakes and Their Faunas." Annu. Rev. Ecol. Evol. Syst **45**: 519-545.

Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." Nature biotechnology **32**(4): 347.

Sander, M., A. Neubüser, J. Kalamaras, H. C. Ee, G. R. Martin and M. S. German (1997). "Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development." Genes & development **11**(13): 1662-1673.

Sara, S. J. (2009). "The locus coeruleus and noradrenergic modulation of cognition." Nature reviews neuroscience **10**(3): 211-223.

Satake, K., A. Oyagi and Y. Iwao (1995). "Natural acidification of lakes and rivers in Japan: the ecosystem of Lake Usoriko (pH 3.4–3.8)." Water, Air, and Soil Pollution **85**(2): 511-516.

Schaefer, J. and A. Ryan (2006). "Developmental plasticity in the thermal tolerance of zebrafish *Danio rerio*." Journal of Fish Biology **69**(3): 722-734.

Schartl, M. (2014). "Beyond the zebrafish: diverse fish species for modeling human disease." Disease models & mechanisms **7**(2): 181-192.

Schwarz, L. A. and L. Luo (2015). "Organization of the locus coeruleus-norepinephrine system." Current Biology **25**(21): R1051-R1056.

Seegers, L., R. Sonnenberg and R. Yamamoto (1999). "Molecular analysis of the Alcolapia flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution." Ichthyological Exploration of Freshwaters **10**: 175-199.

Seegers, L. and H. Tichy (1999). "The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya, with descriptions of two new species." Ichthyological Exploration of Freshwaters **10**: 97-146.

Seidelin, M., S. S. Madsen, H. Blenstrup and C. K. Tipsmark (2000). "Time-course changes in the expression of Na+, K+-ATPase in gills and pyloric caeca of brown trout (*Salmo trutta*) during acclimation to seawater." Physiological and Biochemical Zoology **73**(4): 446-453.

Senior, A. E. (1988). "ATP synthesis by oxidative phosphorylation." Physiological Reviews **68**(1): 177-231.

Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter and D. M. Kingsley (2004). "Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks." Nature **428**(6984): 717.

Simpson, T. I. and D. J. Price (2002). "Pax6; a pleiotropic player in development." Bioessays **24**(11): 1041-1051.

Singh, P. P. and H. Isambert (2020). "OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates." Nucleic acids research **48**(D1): D724-D730.

Sinha, A., M. Kapotwe, S. Dabi, C. Montes, J. Shrivastava, R. Blust and G. Boeck (2016). "Differential modulation of ammonia excretion, Rhesus glycoproteins and ion-regulation in common carp (Cyprinus carpio) following individual and combined exposure to waterborne copper and ammonia." Aquatic toxicology (Amsterdam, Netherlands) **170**: 129-141.

Siwach, P., S. D. Pophaly and S. Ganesh (2006). "Genomic and evolutionary insights into genes encoding proteins with single amino acid repeats." Molecular biology and evolution **23**(7): 1357-1369.

Slack, J. M., P. W. Holland and C. F. Graham (1993). "The zootype and the phylotypic stage." Nature **361**(6412): 490-492.

Soares, D. and M. L. Niemiller (2013). "Sensory adaptations of fishes to subterranean environments." BioScience **63**(4): 274-283.

Sollid, J., P. De Angelis, K. Gundersen and G. E. Nilsson (2003). "Hypoxia induces adaptive and reversible gross morphological changes in crucian carp gills." Journal of Experimental Biology **206**(20): 3667-3673.

Somero, G. (2010). "The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'." Journal of Experimental Biology **213**(6): 912-920.

Song, A., Q. Wang, M. G. Goebl and M. A. Harrington (1998). "Phosphorylation of nuclear MyoD is required for its rapid degradation." Molecular and Cellular Biology **18**(9): 4994-4999.

Spix, J. v. and L. Agassiz (1829). Selecta Genera et Species *Piscium Brasilien* sium, Typis C. Wolf, Monachii (München).

Spyra, A. (2017). "Acidic, neutral and alkaline forest ponds as a landscape element affecting the biodiversity of freshwater snails." The Science of Nature **104**(9-10): 73.

Stahl, B. A. and J. B. Gross (2015). "Alterations in Mc1r gene expression are associated with regressive pigmentation in Astyanax cavefish." Development Genes and Evolution **225**(6): 367-375.

Stauffer Jr, J. R. (1986). "Effects of salinity on preffered and lethal temperatures of the Mozambique Tilapia, *Oreochromis mossambicus*." JAWRA Journal of the American Water Resources Association **22**(2): 205-208.

Steele, S. L., T. D. Chadwick and P. A. Wright (2001). "Ammonia detoxification and localization of urea cycle enzyme activity in embryos of the rainbow trout (*Oncorhynchus mykiss*) in relation to early tolerance to high environmental ammonia levels." Journal of Experimental Biology **204**(12): 2145-2154.

Stensløkken, K.-O., L. Sundin, G. M. Renshaw and G. E. Nilsson (2004). "Adenosinergic and cholinergic control mechanisms during hypoxia in the epaulette shark (*Hemiscyllium ocellatum*), with emphasis on branchial circulation." Journal of Experimental Biology **207**(25): 4451-4461.

Stevenson, T. J., B. A. Alward, F. J. Ebling, R. D. Fernald, A. Kelly and A. G. Ophir (2018). "The value of comparative animal research: Krogh's principle facilitates scientific discoveries." Policy Insights from the Behavioral and Brain Sciences **5**(1): 118-125.

Summerton, J. and D. Weller (1997). "Morpholino antisense oligomers: design, preparation, and properties." Antisense and Nucleic Acid Drug Development **7**(3): 187-195.

Suresh, A. V. and C. K. Lin (1992). "Tilapia culture in saline waters: a review." Aquaculture **106**(3-4): 201-226.

Talbot, J. C., E. M. Teets, D. Ratnayake, P. Q. Duy, P. D. Currie and S. L. Amacher (2019). "Muscle precursor cell movements in zebrafish are dynamic and require Six family genes." Development **146**(10): dev171421.

Tan, X. and S. Du (2002). "Differential expression of two MyoD genes in fast and slow muscles of gilthead seabream (*Sparus aurata*)." Development genes and evolution **212**(5): 207-217.

Tebbs, E., J. Remedios, S. Avery and D. Harper (2013). "Remote sensing the hydrological variability of Tanzania's Lake Natron, a vital Lesser Flamingo breeding site under threat." Ecohydrology & Hydrobiology **13**(2): 148-158.

Temminck, C. and H. Schlegel (1846). "Pisces in Siebold's Fauna Japonica." Lugduni Batavorum, Leiden.

Temminck, C. and H. Schlegel (1850). "Fauna Japonica (Seibold), Pisces."

Terjesen, B. F., T. D. Chadwick, J. A. Verreth, I. Rønnestad and P. A. Wright (2001). "Pathways for urea production during early life of an air-breathing teleost, the African catfish *Clarias gariepinus* Burchell." Journal of Experimental Biology **204**(12): 2155-2165.

Thrane, V. R., A. S. Thrane, F. Wang, M. L. Cotrina, N. A. Smith, M. Chen, Q. Xu, N. Kang, T. Fujita and E. A. Nagelhus (2013). "Ammonia triggers neuronal disinhibition and seizures by impairing astrocyte potassium buffering." Nature medicine **19**(12): 1643-1648.

Tichy, H. and L. Seegers (1999). "The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: a model for the evolution of" new" species flocks in historical times?" Ichthyological Exploration of Freshwaters **10**: 147-174.

Tomarev, S. I., P. Callaerts, L. Kos, R. Zinovieva, G. Halder, W. Gehring and J. Piatigorsky (1997). "Squid Pax-6 and eye development." Proceedings of the National Academy of Sciences **94**(6): 2421-2426.

Torres-Paz, J., C. Hyacinthe, C. Pierre and S. Rétaux (2018). "Towards an integrated approach to understand Mexican cavefish evolution." Biology Letters **14**(8): 20180101.

Trewavas, E. (1982). "Genetic groupings of Tilapiini used in aquaculture." Aquaculture **27**(1): 79-81.

Tromans, D. (2000). "Modeling oxygen solubility in water and electrolyte solutions." Industrial & engineering chemistry research **39**(3): 805-812.

Tse, W., D. Au and C. Wong (2006). "Characterization of ion channel and transporter mRNA expressions in isolated gill chloride and pavement cells of seawater acclimating eels." Biochemical and Biophysical Research Communications **346**(4): 1181-1190.

Tsui, T., C. Hung, C. Nawata, J. Wilson, P. Wright and C. Wood (2009). "Ammonia transport in cultured gill epithelium of freshwater rainbow trout: The importance of Rhesus glycoproteins and the presence of an apical Na+/NH4+ exchange complex." Journal of Experimental Biology **212**(6): 878-892.

Turner, G. F. (2007). "Adaptive radiation of cichlid fish." Current Biology **17**(19): R827-R831.

Valdesalici, S. and A. Cellerino (2003). "Extremely short lifespan in the annual fish *Nothobranchius furzeri*." Proceedings of the Royal Society of London. Series B: Biological Sciences **270**(suppl_2): S189-S191.

Vamosi, S. M. and D. Schluter (2004). "Character shifts in the defensive armor of sympatric sticklebacks." Evolution **58**(2): 376-385.

Van Den Burg, B. (2003). "Extremophiles as a source for novel enzymes." Current Opinion in Microbiology **6**(3): 213-218.

van Waarde, A. (1991). "Alcoholic fermentation in multicellular organisms." Physiological zoology **64**(4): 895-920.

Vanden Bossche, J.-P. and G. M. Bernacsek (1990). Source book for the inland fishery resources of Africa, Food & Agriculture Org.

Vannucci, R. C. (1990). "Experimental biology of cerebral hypoxia-ischemia: relation to perinatal brain damage." Pediatric Research **27**(4): 317.

Viet, J., D. Reboutier, S. Hardy, S. A. Lachke, L. Paillard and C. Gautier-Courteille (2019). "Modeling ocular lens disease in *Xenopus*." Developmental Dynamics: 1-12.

Vincens, A. and J. Casanova (1987). "Modern background of Natron-Magadi basin (Tanzania-Kenya): physiography, climate, hydrology and vegetation. Contexte actuel du bassin Natron-Magadi (Tanzanie-Kenya): physiographie, climat, hydrologie et végétation." Sciences Géologiques, bulletins et mémoires **40**(1): 9-21.

Visser, M. E. (2008). "Keeping up with a warming world; assessing the rate of adaptation to climate change." Proceedings of the Royal Society of London B: Biological Sciences **275**(1635): 649-659.

Vitalis, T., O. Cases, D. Engelkamp, C. Verney and D. J. Price (2000). "Defects of tyrosine hydroxylase-immunoreactive neurons in the brains of mice lacking the transcription factor Pax6." Journal of Neuroscience **20**(17): 6501-6516.

Volff, J. (2005). "Genome evolution and biodiversity in teleost fish." Heredity **94**(3): 280.

Vornanen, M., J. A. W. Stecyk and G. E. Nilsson (2009). The anoxia-tolerant crucian carp (*Carassius carassius* L.). Fish Physiology, Elsevier. **27:** 397-441.

Wagner, A. (1998). "The fate of duplicated genes: loss or new function?" Bioessays **20**(10): 785-788.

Wagner, C. E., L. J. Harmon and O. Seehausen (2012). "Ecological opportunity and sexual selection together predict adaptive radiation." Nature **487**(7407): 366-369.

Walsh, P., E. Danulat and T. Mommsen (1990). "Variation in urea excretion in the gulf toadfish *Opsanus beta*." Marine Biology **106**(3): 323-328.

Walsh, P., M. Grosell, G. G. Goss, H. L. Bergman, A. N. Bergman, P. Wilson, P. Laurent, S. L. Alper, C. P. Smith and C. Kamunde (2001). "Physiological and molecular characterization of urea transport by the gills of the Lake Magadi tilapia (*Alcolapia grahami*)." Journal of Experimental Biology **204**(3): 509-520.

Walsh, P., Y. Wang, C. Campbell, G. De Boeck and C. Wood (2001). "Patterns of nitrogenous waste excretion and gill urea transporter mRNA expression in several species of marine fish." Marine Biology **139**: 839-844.

Wang, Y. and B. Guo (2019). "Adaption to extreme environments: a perspective from fish genomics." Reviews in Fish Biology and Fisheries **29**: 1-13.

Weihrauch, D., M. P. Wilkie and P. J. Walsh (2009). "Ammonia and urea transporters in gills of fish and aquatic crustaceans." Journal of Experimental Biology **212**(11): 1716-1730.

Weintraub, H., R. Davis, D. Lockshon and A. Lassar (1990). "MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation." Proceedings of the National Academy of Sciences **87**(15): 5623-5627.

Westerfield, M. (2007). The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish Danio (" Brachydanio Rerio"), University of Oregon.

Whitehead, A., B. W. Clark, N. M. Reid, M. E. Hahn and D. Nacci (2017). "When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations." Evolutionary Applications **10**(8): 762-783.

Wilkie, M. P. and C. M. Wood (1996). "The adaptations of fish to extremely alkaline environments." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **113**(4): 665-673.

Williamson, D., M. Taieb, B. Damnati, M. Icole and N. Thouveny (1993). "Equatorial extension of the Younger Dryas event: rock magnetic evidence from Lake Magadi (Kenya)." Global and Planetary Change **7**(1-3): 235-242.

Wilson, P. J., C. M. Wood, P. J. Walsh, A. N. Bergman, H. L. Bergman, P. Laurent and B. N. White (2004). "Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia." Physiological and Biochemical Zoology **77**(4): 537-555.

Wilson, R. W., C. M. Wood, R. J. Gonzalez, M. L. Patrick, H. L. Bergman, A. Narahara and A. L. Val (1999). "Ion and acid-base balance in three species of Amazonian fish during gradual acidification of extremely soft water." Physiological and Biochemical Zoology **72**(3): 277-285.

Wise, G., J. M. Mulvey and G. M. Renshaw (1998). "Hypoxia tolerance in the epaulette shark (*Hemiscyllium ocellatum*)." Journal of Experimental Zoology **281**(1): 1-5.

Wood, C., S. Perry, P. Wright, H. Bergman and D. Randall (1989). "Ammonia and urea dynamics in the Lake Magadi tilapia, a ureotelic teleost fish adapted to an extremely alkaline environment." Respiration physiology **77**(1): 1-20.

Wood, C. M., K. V. Brix, G. De Boeck, H. L. Bergman, A. Bianchini, L. F. Bianchini, J. N. Maina, O. E. Johannsson, G. D. Kavembe and M. B. Papah (2016). "Mammalian metabolic rates in the hottest fish on earth." Scientific Reports **6**: 26990.

Wood, C. M. and D. Evans (1993). "Ammonia and urea metabolism and excretion [in fish]." CRC Mar Sci Ser: 379-426.

Wood, C. M., T. Hopkins, C. Hogstrand and P. Walsh (1995). "Pulsatile urea excretion in the ureagenic toadfish *Opsanus beta:* an analysis of rates and routes." Journal of Experimental Biology **198**(8): 1729-1741.

Wood, C. M., C. M. Nawata, J. M. Wilson, P. Laurent, C. Chevalier, H. L. Bergman, A. Bianchini, J. N. Maina, O. E. Johannsson and L. F. Bianchini (2013). "Rh proteins and NH4+-activated Na+-ATPase in the Magadi tilapia (*Alcolapia grahami*), a 100% ureotelic teleost fish." Journal of Experimental Biology **216**(16): 2998-3007.

Wood, C. M., P. Wilson, H. L. Bergman, A. N. Bergman, P. Laurent, G. Otiang'a-Owiti and P. J. Walsh (2002). "Ionoregulatory strategies and the role of urea in the Magadi tilapia (*Alcolapia grahami*)." Canadian Journal of Zoology **80**(3): 503-515.

Wood, C. M., P. Wilson, H. L. Bergman, A. N. Bergman, P. Laurent, G. Otiang'a-Owiti and P. J. Walsh (2002). "Obligatory urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity and alkalinity." Physiological Biochemical Zoology **75**(2): 111-122.

Woodward, G., D. M. Perkins and L. E. Brown (2010). "Climate change and freshwater ecosystems: impacts across multiple levels of organization." Philosophical Transactions of the Royal Society of London B: Biological Sciences **365**(1549): 2093-2106.

Wright, D. A., T. Li, B. Yang and M. H. Spalding (2014). "TALEN-mediated genome editing: prospects and perspectives." Biochemical Journal **462**(1): 15-24.

Wright, P. and H. Fyhn (2001). "Ontogeny of nitrogen metabolism and excretion." Fish Physiology **20**: 149-200.

Wright, P., S. Perry, D. Randall, C. Wood and H. Bergman (1990). "The effects of reducing water pH and total CO2 on a teleost fish adapted to an extremely alkaline environment." Journal of Experimental Biology **151**(1): 361-369.

Wright, P. A. (1993). "Nitrogen excretion and enzyme pathways for ureagenesis in freshwater tilapia (*Oreochromis niloticus*)." Physiological Zoology **66**(6): 881-901.

Wright, P. A. (1995). "Nitrogen excretion: three end products, many physiological roles." Journal of Experimental Biology **198**(2): 273-281.

Wright, P. A., G. K. Iwama and C. M. Wood (1993). "Ammonia and urea excretion in Lahontan cutthroat trout (*Oncorhynchus clarki* henshawi) adapted to the highly alkaline Pyramid Lake (pH 9.4)." Journal of Experimental Biology **175**(1): 153-172.

Wright, P. A. and C. Wood (1985). "An analysis of branchial ammonia excretion in the freshwater rainbow trout: effects of environmental pH change and sodium uptake blockade." Journal of Experimental Biology **114**(1): 329-353.

Wright, P. A. and C. M. Wood (2009). "A new paradigm for ammonia excretion in aquatic animals: role of Rhesus (Rh) glycoproteins." Journal of Experimental Biology **212**(15): 2303-2312.

Wright, P. A. and C. M. Wood (2012). "Seven things fish know about ammonia and we don't." Respiratory physiology & neurobiology **184**(3): 231-240.

Wullimann, M. F. and E. Rink (2001). "Detailed immunohistology of Pax6 protein and tyrosine hydroxylase in the early zebrafish brain suggests role of Pax6 gene in development of dopaminergic diencephalic neurons." Developmental brain research **131**(1-2): 173-191.

Xu, B. and Z. Yang (2013). "PAMLX: a graphical user interface for PAML." Molecular Biology and Evolution **30**(12): 2723-2724.

Xu, J., P. Ji, B. Wang, L. Zhao, J. Wang, Z. Zhao, Y. Zhang, J. Li, P. Xu and X. Sun (2013). "Transcriptome sequencing and analysis of wild Amur Ide (*Leuciscus waleckii*) inhabiting an extreme alkaline-saline lake reveals insights into stress adaptation." PLoS One **8**(4).

Xu, J., J.-T. Li, Y. Jiang, W. Peng, Z. Yao, B. Chen, L. Jiang, J. Feng, P. Ji and G. Liu (2017). "Genomic Basis of Adaptive Evolution: The Survival of Amur Ide (*Leuciscus waleckii*) in an Extremely Alkaline Environment." Molecular Biology and Evolution **34**(1): 145-159.

Xu, J., Q. Li, L. Xu, S. Wang, Y. Jiang, Z. Zhao, Y. Zhang, J. Li, C. Dong and P. Xu (2013). "Gene expression changes leading extreme alkaline tolerance in Amur ide (*Leuciscus waleckii*) inhabiting soda lake." BMC Genomics **14**(1): 682.

Yamamoto, Y., M. S. Byerly, W. R. Jackman and W. R. Jeffery (2009). "Pleiotropic functions of embryonic sonic hedgehog expression link jaw and taste bud amplification with eye loss during cavefish evolution." Developmental Biology **330**(1): 200-211.

Yang, Z., B. V. Alvarez, C. Chakarova, L. Jiang, G. Karan, J. M. Frederick, Y. Zhao, Y. Sauvé, X. Li and E. Zrenner (2005). "Mutant carbonic anhydrase 4 impairs pH regulation and causes retinal photoreceptor degeneration." Human Molecular Genetics **14**(2): 255-265.

Yasue, A., H. Kono, M. Habuta, T. Bando, K. Sato, J. Inoue, S. Oyadomari, S. Noji, E. Tanaka and H. Ohuchi (2017). "Relationship between somatic mosaicism of Pax6 mutation and variable developmental eye abnormalities—an analysis of CRISPR genome-edited mouse embryos." Scientific Reports **7**(1): 1-10.

Yeam, C. T., Y. R. Chng, J. L. Ong, W. P. Wong, S. F. Chew and Y. K. Ip (2017). "Molecular characterization of two Rhesus glycoproteins from the euryhaline freshwater white-rimmed stingray, *Himantura signifer*, and changes in their transcript levels and protein abundance in the gills, kidney, and liver during brackish water acclimation." Journal of Comparative Physiology B: 1-19.

Yu, X. and Y. V. Li (2011). "Zebrafish as an alternative model for hypoxic-ischemic brain damage." International Journal of Physiology, Pathophysiology, Pharmacology **3**(2): 88.

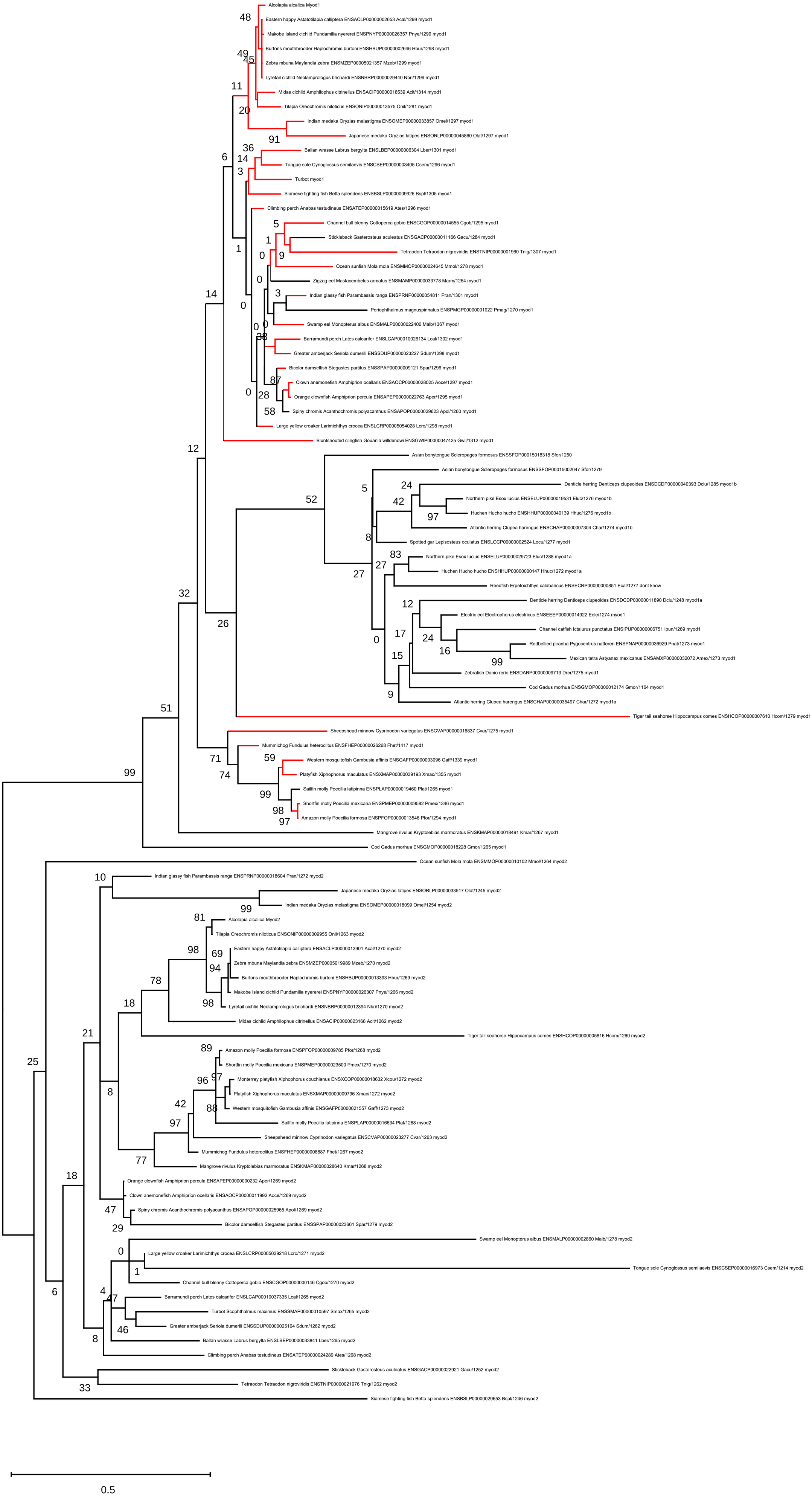
Zhang, C., Z. Ren and Z. Gong (2020). "Transgenic Expression and Genome Editing by Electroporation of Zebrafish Embryos." Marine Biotechnology **22**(5): 644-650.

Zhuang, X., C. Yang, K. R. Murphy and C.-H. C. Cheng (2019). "Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids." Proceedings of the National Academy of Sciences **116**(10): 4400-4405.

Zimmer, A. M. and C. M. Wood (2016). "Physiological and molecular ontogeny of branchial and extra-branchial urea excretion in posthatch rainbow trout (*Oncorhynchus mykiss*)." American Journal of Physiology-Regulatory, Integrative and Comparative Physiology **310**(3): R305-R312.

Zimmer, A. M., P. A. Wright and C. M. Wood (2017). "Ammonia and urea handling by early life stages of fishes." Journal of Experimental Biology **220**(21): 3843-3855.

Appendix 1 – Expanded phylogeny of teleost MyoD1 and MyoD2


Alcolapia alcalica Myod1
Eastern happy Astatotilapia calliptera ENSACLP00000002653 Acal/1299 myod1
Makobe Island cichlid Pundamilia nyererei ENSPNYP00000026357 Pnye/1299 myod1
Burtons mouthbrooder Haplochromis burtoni ENSHBUP00000002646 Hbur/1298 myod1
Zebra mbuna Maylandia zebra ENSMZEP00005021357 Mzeb/1299 myod1
Lyretail cichlid Neolamprologus brichardi ENSNBRP00000029440 Nbri/1299 myod1
Midas cichlid Amphilophus citrinellus ENSACIP00000018539 Acit/1314 myod1
Tilapia Oreochromis niloticus ENSONIP00000013575 Onil/1281 myod1
Indian medaka Oryzias melastigma ENSOMEP00000033857 Omel/1297 myod1
Japanese medaka Oryzias latipes ENSORLP00000045860 Olat/1297 myod1
Ballan wrasse Labrus bergylta ENSLBEP00000006304 Lber/1301 myod1
Tongue sole Cynoglossus semilaevis ENSCSEP00000003405 Csem/1296 myod1
Turbot myod1
Siamese fighting fish Betta splendens ENSBSLP00000009926 Bspl/1305 myod1
Climbing perch Anabas testudineus ENSATEP00000015619 Ates/1296 myod1
Channel bull blenny Cottoperca gobio ENSCGOP00000014555 Cgob/1295 myod1
Stickleback Gasterosteus aculeatus ENSGACP00000011166 Gacu/1284 myod1
Tetraodon Tetraodon nigroviridis ENSTNIP00000001960 Tnig/1307 myod1
Ocean sunfish Mola mola ENSMMOP00000024645 Mmol/1278 myod1
Zigzag eel Mastacembelus armatus ENSMAMP00000033778 Marm/1264 myod1
Indian glassy fish Parambassis ranga ENSPRNP00000054811 Pran/1301 myod1
Periophthalmus magnuspinnatus ENSPMGP00000001022 Pmag/1270 myod1
Swamp eel Monopterus albus ENSMALP00000022400 Malb/1367 myod1
Barramundi perch Lates calcarifer ENSLCAP00010026134 Lcal/1302 myod1
Greater amberjack Seriola dumerili ENSSDUP00000023227 Sdum/1298 myod1
Bicolor damselfish Stegastes partitus ENSSPAP00000009121 Spar/1296 myod1
Clown anemonefish Amphiprion ocellaris ENSAOCP00000028025 Aoce/1297 myod1
Orange clownfish Amphiprion percula ENSAPEP00000022763 Aper/1295 myod1
Spiny chromis Acanthochromis polyacanthus ENSAPOP00000029623 Apol/1260 myod1
Large yellow croaker Larimichthys crocea ENSLCRP00005054028 Lcro/1298 myod1
Bluntsnouted clingfish Gouania willdenowi ENSGWIP00000047425 Gwil/1312 myod1
Asian bonytongue Scleropages formosus ENSSFOP00015018318 Sfor/1250
Asian bonytongue Scleropages formosus ENSSFOP00015002047 Sfor/1279
Denticle herring Denticeps clupeoides ENSDCDP00000040393 Dclu/1285 myod1b
Northern pike Esox lucius ENSELUP00000019531 Eluc/1276 myod1b
Huchen Hucho hucho ENSHHUP00000040139 Hhuc/1276 myod1b
Atlantic herring Clupea harengus ENSCHAP00000007304 Char/1274 myod1b
Spotted gar Lepisosteus oculatus ENSLOCP00000002524 Locu/1277 myod1
Northern pike Esox lucius ENSELUP00000029723 Eluc/1288 myod1a
Huchen Hucho hucho ENSHHUP00000000147 Hhuc/1272 myod1a
Reedfish Erpetoichthys calabaricus ENSECRP00000000851 Ecal/1277 dont know
Denticle herring Denticeps clupeoides ENSDCDP00000011890 Dclu/1248 myod1a
Electric eel Electrophorus electricus ENSEEEP00000014922 Eele/1274 myod1
Channel catfish Ictalurus punctatus ENSIPUP00000006751 Ipun/1269 myod1
Redbellied piranha Pygocentrus nattereri ENSPNAP00000036929 Pnat/1273 myod1
Mexican tetra Astyanax mexicanus ENSAMXP00000032072 Amex/1273 myod1
Zebrafish Danio rerio ENSDARP00000009713 Drer/1275 myod1
Cod Gadus morhua ENSGMOP00000012174 Gmor/1164 myod1
Atlantic herring Clupea harengus ENSCHAP00000035497 Char/1272 myod1a
Tiger tail seahorse Hippocampus comes ENSHCOP00000007610 Hcom/1279 myod1
Sheepshead minnow Cyprinodon variegatus ENSCVAP00000016837 Cvar/1275 myod1
Mummichog Fundulus heteroclitus ENSFHEP00000026268 Fhet/1417 myod1
Western mosquitofish Gambusia affinis ENSGAFP00000003096 Gaff/1339 myod1
Platyfish Xiphophorus maculatus ENSXMAP00000039193 Xmac/1355 myod1
Sailfin molly Poecilia latipinna ENSPLAP00000019460 Plat/1265 myod1
Shortfin molly Poecilia mexicana ENSPMEP00000009582 Pmex/1346 myod1
Amazon molly Poecilia formosa ENSPFOP00000013546 Pfor/1294 myod1
Mangrove rivulus Kryptolebias marmoratus ENSKMAP00000018491 Kmar/1267 myod1
Cod Gadus morhua ENSGMOP00000018228 Gmor/1265 myod1
Ocean sunfish Mola mola ENSMMOP00000010102 Mmol/1264 myod2
Indian glassy fish Parambassis ranga ENSPRNP00000018604 Pran/1272 myod2
Japanese medaka Oryzias latipes ENSORLP00000033517 Olat/1245 myod2
Indian medaka Oryzias melastigma ENSOMEP00000018099 Omel/1254 myod2
Alcolapia alcalica Myod2
Tilapia Oreochromis niloticus ENSONIP00000009955 Onil/1263 myod2
Eastern happy Astatotilapia calliptera ENSACLP00000013901 Acal/1270 myod2
Zebra mbuna Maylandia zebra ENSMZEP00005019989 Mzeb/1270 myod2
Burtons mouthbrooder Haplochromis burtoni ENSHBUP00000013393 Hbur/1269 myod2
Makobe Island cichlid Pundamilia nyererei ENSPNYP00000026307 Pnye/1266 myod2
Lyretail cichlid Neolamprologus brichardi ENSNBRP00000012394 Nbri/1270 myod2
Midas cichlid Amphilophus citrinellus ENSACIP00000023168 Acit/1262 myod2
Tiger tail seahorse Hippocampus comes ENSHCOP00000005816 Hcom/1260 myod2
Amazon molly Poecilia formosa ENSPFOP00000009785 Pfor/1268 myod2
Shortfin molly Poecilia mexicana ENSPMEP00000023500 Pmex/1270 myod2
Monterrey platyfish Xiphophorus couchianus ENSXCOP00000018632 Xcou/1272 myod2
Platyfish Xiphophorus maculatus ENSXMAP00000009796 Xmac/1272 myod2
Western mosquitofish Gambusia affinis ENSGAFP00000021557 Gaff/1273 myod2
Sailfin molly Poecilia latipinna ENSPLAP00000016634 Plat/1268 myod2
Sheepshead minnow Cyprinodon variegatus ENSCVAP00000023277 Cvar/1263 myod2
Mummichog Fundulus heteroclitus ENSFHEP00000008887 Fhet/1267 myod2
Mangrove rivulus Kryptolebias marmoratus ENSKMAP00000028640 Kmar/1268 myod2
Orange clownfish Amphiprion percula ENSAPEP00000000232 Aper/1269 myod2
Clown anemonefish Amphiprion ocellaris ENSAOCP00000011992 Aoce/1269 myod2
Spiny chromis Acanthochromis polyacanthus ENSAPOP00000025965 Apol/1269 myod2
Bicolor damselfish Stegastes partitus ENSSPAP00000023661 Spar/1279 myod2
Swamp eel Monopterus albus ENSMALP00000002860 Malb/1278 myod2
Large yellow croaker Larimichthys crocea ENSLCRP00005039218 Lcro/1271 myod2
Tongue sole Cynoglossus semilaevis ENSCSEP00000016973 Csem/1214 myod2
Channel bull blenny Cottoperca gobio ENSCGOP00000000146 Cgob/1270 myod2
Barramundi perch Lates calcarifer ENSLCAP00010037335 Lcal/1265 myod2
Turbot Scophthalmus maximus ENSSMAP00000010597 Smax/1265 myod2
Greater amberjack Seriola dumerili ENSSDUP00000025164 Sdum/1262 myod2
Ballan wrasse Labrus bergylta ENSLBEP00000033841 Lber/1265 myod2
Climbing perch Anabas testudineus ENSATEP00000024289 Ates/1268 myod2
Stickleback Gasterosteus aculeatus ENSGACP00000022921 Gacu/1252 myod2
Tetraodon Tetraodon nigroviridis ENSTNIP00000021976 Tnig/1262 myod2
Siamese fighting fish Betta splendens ENSBSLP00000029653 Bspl/1246 myod2
0.5

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: White LJ, Sutton G, Shechonge A, Day JJ, Dasmahapatra KK, Pownall ME. 2020 Adaptation of the carbamoyl-phosphate synthetase enzyme in an extremophile fish. *R. Soc. Open Sci.* **7**: 201200.
http://dx.doi.org/10.1098/rsos.201200

Received: 8 July 2020
Accepted: 21 September 2020

Subject Category:
Organismal and evolutionary biology

Subject Areas:
evolution/developmental biology/biochemistry

Keywords:
urea, carbamoyl-phosphate synthetase, *Alcolapia*, extremophile, fish

Author for correspondence:
Lewis J. White
e-mail: ljw569@york.ac.uk

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5172175.

THE ROYAL SOCIETY PUBLISHING

Adaptation of the carbamoyl-phosphate synthetase enzyme in an extremophile fish

Lewis J. White[1], Gemma Sutton[1], Asilatu Shechonge[2], Julia J. Day[3], Kanchon K. Dasmahapatra[1] and Mary E. Pownall[1]

[1]Biology Department, University of York, York YO10 5DD, UK
[2]Tanzania Fisheries Research Institute, PO BOX 98, Kyela, Mbeya, Tanzania
[3]Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

LJW, 0000-0002-8764-2051; GS, 0000-0002-0459-0912

Tetrapods and fish have adapted distinct carbamoyl-phosphate synthase (CPS) enzymes to initiate the ornithine urea cycle during the detoxification of nitrogenous wastes. We report evidence that in the ureotelic subgenus of extremophile fish *Oreochromis Alcolapia*, CPS III has undergone convergent evolution and adapted its substrate affinity to ammonia, which is typical of terrestrial vertebrate CPS I. Unusually, unlike in other vertebrates, the expression of CPS III in *Alcolapia* is localized to the skeletal muscle and is activated in the myogenic lineage during early embryonic development with expression remaining in mature fish. We propose that adaptation in *Alcolapia* included both convergent evolution of CPS function to that of terrestrial vertebrates, as well as changes in development mechanisms redirecting *CPS III* gene expression to the skeletal muscle.

1. Introduction

In living organisms, protein metabolism results in the production of nitrogenous wastes which need to be excreted. Most teleosts are ammonotelic, excreting their toxic nitrogenous waste as ammonia across gill tissue by diffusion. As an adaptation to living on land, amphibians and mammals are ureotelic, using liver and kidney tissues to convert waste ammonia into the less toxic and more water-soluble urea, which is then excreted in urine. Other terrestrial animals such as insects, birds and reptiles are

uricolotic and convert nitrogenous waste into uric acid, which is eliminated as a paste; a process which requires more energy but wastes less water [1].

While most adult fish are ammonotelic, the larval stages of some teleosts excrete nitrogenous waste as both ammonia and urea before their gills are fully developed [2]. Additionally, some adult fish species such as the gulf toad fish (*Opsanus beta*; [3]) and the African catfish (*Clarias gariepinus*; [4]) also excrete a proportion of their nitrogenous waste as urea. This is usually in response to changes in aquatic conditions, such as high alkalinity. It has been shown experimentally that high external pH prevents diffusion of ammonia across gill tissue [5,6]. Unusually, the cichlid fish species in the subgenus *Alcolapia* (described by some authors as a genus but shown to nest within the genus *Oreochromis*) [7], which inhabit the highly alkaline soda lakes of Natron (Tanzania) and Magadi (Kenya), are reported to be 100% ureotelic [8,9].

Once part of a single palaeolake, Orolonga [10], Lakes Natron and Magadi are one of the most extreme environments supporting fish life, with water temperatures up to 42.8°C, pH approximately 10.5, fluctuating dissolved oxygen levels, and salt concentrations above 20 parts per thousand [11]. *Alcolapia* is the only group of fish to survive in these lakes, forming a recent adaptive radiation including the four species: *Alcolapia grahami* (Lake Magadi) and *A. latilabris*, *A. ndlalani* and *A. alcalica* (Lake Natron) [11,12]. The harsh environment of the soda lakes presents certain physiological challenges that *Alcolapia* have evolved to overcome, including the basic need to excrete nitrogenous waste. While other species are able to excrete urea in response to extreme conditions, none do so to the level of *Alcolapia* [13,14], and unlike facultative ureotelic species, the adaptation of urea production and excretion in *Alcolapia* is considered fixed [15]. Moreover, the heightened metabolic rate in *Alcolapia*, a by-product from living in such an extreme environment [8,16], requires an efficient method of detoxification.

Alcolapia and ureotelic tetrapods (including humans) detoxify ammonia using the ornithine urea cycle (OUC) where the mitochondrial enzyme carbamoyl-phosphate synthetase (CPS) is essential for the first and rate-limiting step of urea production [17]. This enzyme, together with the accessory enzyme glutamine synthase, provide an important switch regulating the balance between ammonia removal for detoxification and maintaining a source of ammonia for the biosynthesis of amino acids [18]. CPS has evolved into two biochemically distinct proteins: in terrestrial vertebrates CPS I uses ammonia as its preferential nitrogen donor, while in teleosts CPS III accepts glutamine to produce urea during larval stages (reviewed Zimmer *et al.* [2]). While CPS I/III are mitochondrial enzymes and part of the urea cycle, CPS II is present in the cytosol catalysing the synthesis of carbamoyl phosphate for pyrimidine nucleotide biosynthesis. *CPS I/III* are syntenic, representing orthologous genes; their somewhat confusing nomenclature is based on the distinct biochemical properties of their proteins. *CPS I/III* genes from different vertebrate species clade together, separate from *CPS II* (electronic supplementary material, figure). For simplicity, we will continue to refer to fish, glutamine binding CPS as CPS III and tetrapod, ammonia binding CPS as CPS I. The teleost CPS III binds glutamine in the glutamine amidotransferase (GAT) domain using two amino acid residues [19], subsequently, the nitrogen source provided by the amide group is catalysed by a conserved catalytic triad; Cys-His-Glu [20]. In terrestrial vertebrates CPS I lacks a complete catalytic triad and can only generate carbamoyl-phosphate in the presence of free ammonia [21]. This change in function from glutamine binding CPS III to ammonia binding CPS I is believed to have evolved in the stem lineage of living tetrapods, first appearing in ancestral amphibians [21].

In tetrapods and most fish, the OUC enzymes are largely localized to the liver [22], the main urogenic organ [23]. *Alcolapia* are different, and the primary site for urea production in these extremophile fishes is the skeletal muscle [24]. Notably, glutamine synthase activity is reportedly absent in *Alcolapia* muscle tissue. The kinetic properties of CPS III in *Alcolapia*, therefore, differ from that of other teleosts in that it preferentially uses ammonia as its primary substrate, having maximal enzymatic rates above that of binding glutamine (although it is still capable of doing so) as opposed to in other species where the use of ammonia yields enzymatic rates of around 10% to that of glutamine [24]. These rates are similar to ureotelic terrestrial species, where CPS I preferentially binds ammonia and is incapable of using glutamine [20].

Here, we report the amino acid sequence of two *Alcolapia* species (*A. alcalica* and *A. grahami*) that reveals a change in CPS III substrate binding site. In addition, we show that the expression of *Alcolapia CPS III* in skeletal muscle arises early in embryonic development where transcripts are restricted to the somites, the source of skeletal muscle in all vertebrates, and migrating myogenic precursors. We discuss changes to the structure of functional domains and modular gene enhancers that probably underpin evolutionary changes in *Alcolapia* CPS III substrate binding and the redirection

of gene expression from the hepatogenic to myogenic lineage [25]. Our findings point to adaptation in *Alcolapia* including both convergent evolution of CPS function to that of terrestrial vertebrates, as well as changes in development mechanisms redirecting *CPS III* gene expression to the skeletal muscle.

2. Methods

2.1. Experimental animals

Fieldwork at Lake Natron, Tanzania, was conducted during June and July of 2017 to collect live specimens of the three endemic species in an attempt to produce stable breeding populations of these fishes in the UK. Live fish were all collected from a single spring (site 5 [11,26]) containing all three species found in Lake Natron and identified using morphology as described in Seegers & Tichy [12]. A stand-alone, recirculating aquarium was adapted to house male and female *A. alcalica* in 10 or 30 l tanks at a constant temperature of 30°C, pH 9 and salt concentration of 3800 µS at the University of York.

2.2. Expression of *CPS III* in adult tissues

Reverse transcription–polymerase chain reaction (RT-PCR) was used to determine the presence of *CPS III* in different tissues (gill, muscle, liver, brain) of three different adult *A. alcalica*. RNA was extracted from dissected tissues with TriReagent (Sigma-Aldrich) to the manufacturers' guidelines. For cDNA synthesis, 1 µg of total RNA was reverse transcribed with random hexamers (Thermo Scientific) and superscript IV (Invitrogen). PCR was performed on 2 µl of the above cDNA with Promega PCR master mix and 0.5 mM of each primer (forward: CAGTGGGAGGTCAGATTGC, reverse: CTCACAGCGAAGCACAGGG). Gel electrophoresis of the PCR products determined the presence or absence of *CPS III* RNA.

2.3. *In situ* hybridization

For the production of antisense probes, complementary to the mRNA of *CPS III* to use in *in situ* hybridization, the above 399 bp PCR product was ligated into PGem-tEasy and transformed into the *Escherichia coli* strain DH5*α*. This was linearized and *in vitro* run-off transcription was used to incorporate a DIG-labelled UTP analogue. To determine the temporal expression of these proteins in *A. alcalica*, embryos were collected at different stages of development (2, 4 and 7 days post fertilization (between 15 and 20 for each stage)), fixed for 1 hour in MEMFA (0.1 M MOPS pH 7.4, 2 mM EGTA, 1 mM $MgSO_4$, 3.7% formaldehyde) at room temperature and stored at −20°C in 100% methanol. For *in situ* hybridization, embryos were rehydrated and treated with 10 µg mg^{-1} proteinase *K* at room temperature. After post-fixation and a 2 h pre-hybridization, embryos were hybridized with the probe at 68°C in hybridization buffer (50% formamide (Ambion), 1 mg ml^{-1} total yeast RNA, 5×SSC, 100 µg ml^{-1} heparin, 1× denharts, 0.1% Tween-20, 0.1% CHAPS, 10 mM EDTA. Embryos were extensively washed at 68°C in 2×SSC + 0.1% Tween-20, 0.2×SSC + 0.1% Tween-20 and maleic acid buffer (MAB; 100 mM maleic acid, 150 mM NaCl, 0.1% Tween-20, pH 7.8). This was replaced with pre-incubation buffer (4× MAB, 10% BMB, 20% heat-treated lamb serum) for 2 h. Embryos were incubated overnight (rolling at 4°C) with fresh pre-incubation buffer and 1/2000 dilution of anti-DIG coupled with alkaline phosphatase (AP) (Roche). These were then visualized by the application of BM purple until staining had occurred.

2.4. Sequence analysis of CPS III

cDNA was produced from the RNA extracted from whole embryos using the above method for *A. alcalica* and *A. grahami*. Multiple primer pairs (electronic supplementary material, table S1) were used to amplify fragments of *CPS III* from the cDNA via PCR and the products sent for sequencing. The coding region of *CPS III* was then constructed using multiple alignments against the *CPS I* and *III* from other species. The amino acid sequence was then examined for potential changes which could predict the functional differences seen in *Alcolapia*. Phylogenetic analysis was also used to confirm the *Alcolapia* genes analysed here are *CPS III* (electronic supplementary material, figure S1). To determine potential changes in the promoter region, a 3500 bp section of genome (accession number NCBI: MW014910) upstream of the transcriptional site start of *CPS* from *A. alcalica* (unpublished genome), *Oreochromis*

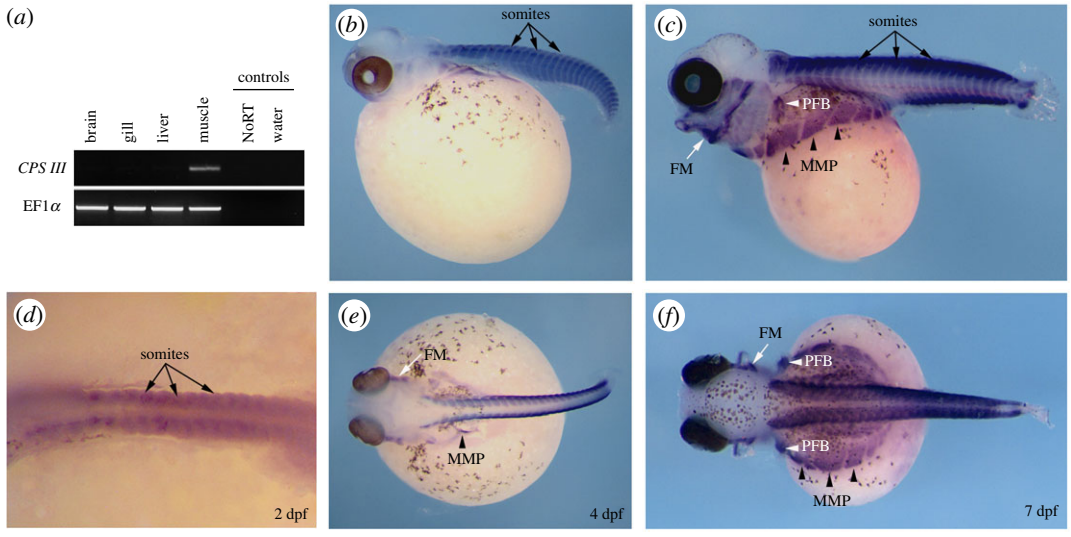


Figure 1. Expression analysis of carbamoyl-phosphate synthetase III (*CPS III*) from adult tissues and developing embryos of *Alcolapia alcalica*. (*a*) Reverse transcriptase PCR and gel electrophoresis showing the muscle-specific expression of *CPS III*, EF1α shown as normalization control. (*b–f*) Lateral (*c* and *e*) and dorsal (*b*, *d* and *f*) views of *in situ* hybridization for *CPS III* in developing *A. alcalica* embryos at different stages (number of days post fertilization [dpf] indicated). The blue colour indicates the detection of mRNA. The black/brown is endogenous pigment apparent in the retina and the chromatophores. Black arrows show somites, black arrowheads indicate region of migrating muscle progenitors (MMP), white arrows show facial muscle (FM) and white arrowheads indicate developing pectoral fin bud (PFB). Black dots around the yolk and on the body are chromatophores (pigment cells).

niloticus (Nile tilapia), *Xenopus tropicalis* (western clawed frog) and *Danio rerio* (zebrafish) genomes was accessed on Ensembl, aligned and examined for binding sites specific to the muscle transcription factor Myod1 (E-boxes) which preferentially binds paired E-boxes in the enhancer regions of myogenic genes with the consensus motif CAG(G/C)TG, as well as E-boxes more broadly (CANNTG). The published genomes of *O. niloticus*, *D. rerio* and *X. tropicalis* were accessed using Ensembl, whereas the *Alcolapia* genome was constructed from whole-genome sequences.

3. Results

3.1. *CPS III* expression is activated early in the skeletal muscle lineage in *A. alcalica*

Analysis of gene expression of *CPS III* in dissected tissues of three adult *A. alcalica* shows that transcripts were only detected in adult muscle (figure 1*a*). *In situ* hybridization methods on *A. alcalica* embryos at different stages were carried out to investigate whether this restricted muscle expression was established during development (figure 1*b–f*). Blue coloration indicates hybridization of the complementary RNA probe and shows the strongest expression in the developing somites along the body axis (black arrows). Expression was also detected in migratory muscle precursors (MMP; black arrowheads), which go on to form the body wall and limb musculature, and in the developing pectoral fin buds (white arrows). All regions of the embryo that show expression of *CPS III* are in the muscle lineage indicating that in *A. alcalica CPS III* expression is restricted to muscle tissues in both adults and the developing embryo.

Many muscle-specific genes are activated during development by the muscle-specific transcription factor, MyoD. The promoter region of *CPS III* (3.5 kb upstream of the transcriptional start site) in *A. alcalica* was compared to that in *O. niloticus*, *X. tropicalis* and *D. rerio* (figure 2). Examination of this region revealed a putative paired E-box MyoD binding site 940–970 bases (CAGGTGACTGTGATTATA-TAGTTCACAGGTG) upstream of the transcriptional start site of *CPS III* only in *Alcolapia* species. Intriguingly, while no pair of MyoD E-boxes were found in the upstream region of any other species examined, *O. niloticus* does have a single MyoD E-box motif in the same region upstream of *CPS III*, and within 19 bases of this is a CAGGTT motif which a single point mutation would convert into a pair of E-boxes (CAGGTGACTGTGATTATATAGTTCACAGGTT). This suggests that it is possible that

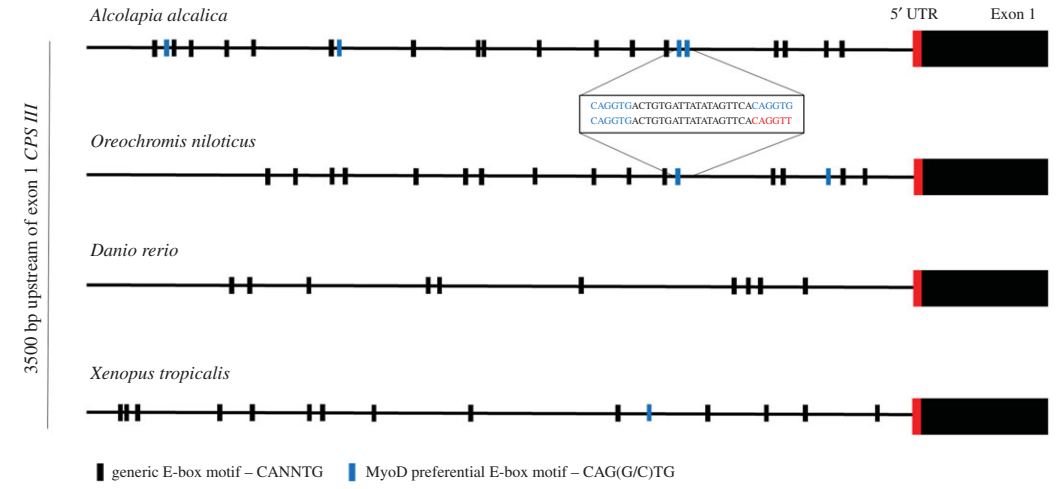


Figure 2. Sequence comparison of a 3500 bp region upstream of exon 1 of *CPS III*. Presence of generic E-box sites with the consensus motif of CANNTG are annotated by black bars and MyoD preferential E-box motifs of CAG(G/C)TG are represented by blue bars. Insert shows sequence of *Alcolapia alcalica* potential paired E-box, MyoD enhancer (blue nucleotides) compared to *Oreochromis niloticus* which has a single MyoD E-box upstream of a 'presumptive' MyoD E-box (red nucleotides).

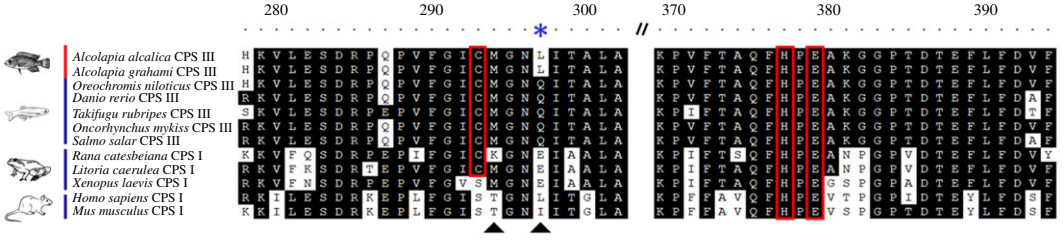


Figure 3. Multiple amino acid alignment of residues 278–397 (aligned to *Alcolapia alcalica*) of carbamoyl-phosphate synthetase I and III from a number of tetrapod and teleost species, respectively. Conserved amino acids are shaded in black, amino acids in the catalytic triad are boxed in red and arrowheads indicate residues vital for glutamine utilization of the catalytic triad. The blue asterisk indicates the divergent glutamine binding residue in *Alcolapia* species that probably results in a functional change (inability to bind glutamine).

MyoD could bind and activate transcription of *CPS III* in the muscle of *Alcolapia* species, but not in the closely related *O. niloticus*.

3.2. Convergent evolution in the adaptive function of CPS III

Sequence analysis of *A. alcalica* and *A. grahami* CPS III revealed a discrepancy in the catalytic triad compared to the published sequence for CPS III in *A. grahami* (accession number NCBI: AF119250). The coding region for *A. alcalica* and *A. grahami* was cloned and sequenced (accession numbers NCBI: MT119353, MT119354). Our data confirmed the error in the published sequence of *A. grahami* CPS III and shows *Alcolapia* species maintain a catalytic triad essential for catalysing the breakdown of glutamine (red boxes in figure 3). However, similar to terrestrial vertebrate CPS I which lack either one but usually both residues essential for binding glutamine for utilization by the catalytic triad (arrowheads in figure 3), *Alcolapia* also lack one of these residues (asterisk in figure 3). This amino acid sequence is consistent with a change in function permitting *Alcolapia* CPS III to bind and catalyse ammonia directly, an activity usually restricted to terrestrial vertebrate CPS I, as elucidated by extensive previous biochemical analyses [20,21].

4. Discussion

While most teleosts are ammonotelic, larval fish can convert ammonia to urea for excretion and to do so express the genes coding for the enzymes of the OUC, including CPS III [27]. Later these genes are silenced in most fish. In the rare cases where urea is produced in adult fish, the OUC enzymes

are expressed in the liver [23]; however, there are some reports of expression in non-hepatic tissues [28,29]. We report here the expression of *CPS III* in the muscle of adult *A. alcalica*, which is consistent with the detection of CPS III protein and enzyme activity in muscle of *A. grahami* [24]. We also find conserved changes to the amino acid sequence which explains the convergent evolution of *A. alcalica* and *A. grahami* CPS III function with CPS I in terrestrial vertebrates. This conserved change in both *Alcolapia* species suggests that the adaptations in the OUC are likely to have evolved in the ancestral species inhabiting palaeolake Orolongo during the period of changing aquatic conditions (over the past 10 000 years) that led to the extreme conditions currently found in Lakes Natron and Magadi.

4.1. Activation of *CPS III* in the myogenic lineage

We find that the expression of *CPS III* is activated in somites and in migratory muscle precursors that will form body wall and limb musculature (indeed expression is seen in developing limb buds). All skeletal muscle in the vertebrate body is derived from the somites, and these *CPS III* expression patterns are similar to those of muscle-specific genes like myosin, actins and troponins [30–32].

Muscle-specific expression of *CPS III* in *A. alcalica* embryos is a remarkable finding as most ureotelic species convert nitrogenous waste to urea in the liver [8,20]. The expression of *CPS III*, the first enzyme in the OUC, in muscle tissue is probably significant for supporting the high catabolism in a fish species with the highest recorded metabolic rate [33]. There are few reports of some OUC gene expression or enzyme activity in non-hepatic tissue including muscle [28,29,34]; nonetheless, other fish species only evoke the activity of the OUC when exposed to high external pH or during larval stages [13,14,35,36], and even then, urea production is never to the high level of activity occurring in *Alcolapia* [24]. There is some heterogeneity of the expression patterns of *CPS III* during the development of different species in the teleost lineage; for example, *D. rerio* has reported expression in the body [37], *Oncorhynchus mykiss* (rainbow trout) shows expression in the developing body but not in hepatic tissue [38] and *C. gariepinus* (African catfish) had *CPS III* expression detected in the dissected muscle from larvae [4]. The early and sustained expression of *CPS III* in the muscle lineage is at this point an observation unique to *Alcolapia*.

Skeletal muscle-specific gene expression is activated in cells of the myogenic lineage by a family of bHLH transcription factors, including MyoD [30]. MyoD binds specifically at paired E-boxes in the enhancers of myogenic genes with a preference for the consensus motif of CAG(G/C)TG [39,40]. MyoD is known to require the cooperative binding at two E-boxes in close proximity, to modulate transcription of myogenic genes [41]. The presence of a pair of E-boxes in *Alcolapia*, upstream of a gene which has switched to muscle-specific expression, is suggestive that MyoD is driving expression early in development. Enhancer modularity is a known mechanism for selectable variation [42] and although a single MyoD binding site does not define an enhancer, MyoD is known to interact with pioneer factors and histone deacetylases to open chromatin and activate gene transcription in the muscle lineage [40,43]. Experimental analysis to determine the activity of any regulatory sequences upstream of OUC genes in different species would shed light on the significance of putative transcription factor binding sites. This approach could also address another intriguing question as to the elements that drive the post-larval silencing of OUC genes in most fish species [37], an area with only minimal research, especially when compared to the well-characterized promoter region in mammalian species, for instance, Christoffels *et al.* [44]. A further instance of an extremophile organism redirecting the expression of a hepatic enzyme to muscle tissue occurs in the crucian carp [45]. Under conditions of anoxia, this species switches to anaerobic metabolism, producing ethanol as the end product of glycolysis [46,47]. This is associated with the expression of *alcohol dehydrogenase* in muscle [48]. Together with our findings, this potentially reveals an example of convergent evolution whereby the muscle becomes the site for detoxifying by-products of metabolism. Elucidating any mechanisms that may include modular enhancers that facilitate the adaptation of gene regulation in response to changing environmental conditions will be of significant interest.

4.2. Convergent evolution of adaptive CPS III function

CPS proteins catalyse the production of carbamoyl-phosphate as the first step in nitrogen detoxification by accepting either glutamine or ammonia as a nitrogen donor [17]. Teleost CPS III binds glutamine: the nitrogen source provided by the amide group of glutamine is catalysed by the conserved catalytic triad Cys-His-Glu in the glutamine amidotransferase (GAT) domain in the amino terminal part of CPS [20]. In

terrestrial vertebrates, CPS I lacks the catalytic cysteine residue and only generates carbamoyl-phosphate in the presence of free ammonia [21]. Although CPS in *Alcolapia* shares the most sequence identity with fish CPS III (figure 3), its ammonia binding activity is more similar in function to terrestrial vertebrate CPS I [20,24]. This adaptation to preferentially bind ammonia over glutamine supports efficient waste management in a fish with an exceptionally high metabolic rate [33]. CPS I in terrestrial vertebrates have amino acid changes in the catalytic triad which explains their binding ammonia over glutamine; a reduction in glutamine binding capacity drives the use of ammonia [21]. Here, we show that *Alcolapia* maintain the catalytic triad, but (similar to mouse and human) lack one of the two residues required for efficient glutamine binding, weakening its affinity to glutamine and driving the use of ammonia as a primary substrate.

The interesting observation that bullfrog (*Rana catesbeiana*) CPS I retains the catalytic triad, but lacks the two additional conserved amino acids required for glutamine binding, has led to the suggestion that the change from preferential glutamine to ammonia binding originally evolved in the early tetrapod lineage [21]. A further frog species, the tree frog *Litoria caerulea*, retains its catalytic triad and only one of the two residues required for glutamine binding has been altered, weakening its affinity for glutamine and allowing for direct catabolism of ammonia [49]. Much the same as in *Alcolapia*, *L. caerulea* CPS I is still capable of using glutamine to some extent which lends further support to the notion that the evolutionary transition from CPS III to CPS I occurring in amphibians and the early tetrapod lineage. The changes in the protein sequence of *Alcolapia* CPS III represents a convergent evolution in this extremophile fish species, with acquired changes in functionally important domains which probably also evolved in early terrestrial vertebrate CPS I.

5. Conclusion

Alcolapia have acquired multiple adaptations that allow continued excretion of nitrogenous waste in a high pH environment. Among these is the novel expression of *CPS* in skeletal muscle, as well as the acquisition of mutations that change its function. Sequence evidence indicates that like terrestrial vertebrates, and unique among fish, *Alcolapia* CPS III is capable of binding and catalysing the breakdown of ammonia to carbamoyl-phosphate; a convergent evolution of *CPS* function. The mechanism by which the novel and unique expression of CPS in muscle evolved is probably a function of enhancer regions of *A. alcalica* and *A. grahami* that result in its regulation by muscle regulatory factors to direct CPS expression in the myogenic lineage during embryonic development. Environmentally driven adaptations have resulted in changes in both the expression and activity of CPS III in *Alcolapia* that underpin its ability to turn over nitrogenous waste in a challenging environment while maintaining a high metabolic rate.

Ethics. The research was approved by the University of York AWERB and under the Home Office licence for Dr. M.E. Pownall (POF 245945).

Data accessibility. Sequence data has been made available on NCBI (accession numbers: MT119353, MT119354) and in the electronic supplementary material, files.

Authors' contributions. Experiments were designed by L.J.W. and M.E.P., work was carried out by L.J.W. and G.S. and the manuscript written and edited by all authors. L.J.W., J.J.D. and A.S. collected the fish from Lake Natron.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the BBSRC as a studentship to L.J.W. (BB/M011151/1). Additional support for fieldwork was provided by the Fisheries Society of the British Isles (small research grant) and the Genetics Society (Heredity fieldwork grant).

Acknowledgements. The authors are grateful to Antonia Ford and George Turner for their help in collecting the fish from Lake Natron (COSTECH permit no. 2017-259-NA-2011-182) and their continued support in this project.

References

1. Wright PA. 1995 Nitrogen excretion: three end products, many physiological roles. *J. Exp. Biol.* **198**, 273–281.
2. Zimmer AM, Wright PA, Wood CM. 2017 Ammonia and urea handling by early life stages of fishes. *J. Exp. Biol.* **220**, 3843–3855. (doi:10.1242/jeb.140210)
3. Barimo JF, Steele SL, Wright PA, Walsh PJ. 2004 Dogmas and controversies in the handling of nitrogenous wastes: ureotely and ammonia tolerance in early life stages of the gulf toadfish, *Opsanus beta*. *J. Exp. Biol.* **207**, 2011–2020. (doi:10.1242/jeb.00956)
4. Terjesen BF, Chadwick TD, Verreth JA, Rønnestad I, Wright PA. 2001 Pathways for urea production during early life of an air-breathing teleost, the African catfish *Clarias gariepinus* Burchell. *J. Exp. Biol.* **204**, 2155–2165.
5. Handy R, Poxton M. 1993 Nitrogen pollution in mariculture: toxicity and excretion of

nitrogenous compounds by marine fish. *Rev. Fish Biol. Fish.* **3**, 205–241. (doi:10.1007/BF00043929)
6. Wright PA, Wood C. 1985 An analysis of branchial ammonia excretion in the freshwater rainbow trout: effects of environmental pH change and sodium uptake blockade. *J. Exp. Biol.* **114**, 329–353.
7. Ford AGP *et al.* 2019 Molecular phylogeny of Oreochromis (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments. *Mol. Phylogenet. Evol.* **136**, 215–226. (doi:10.1016/j.ympev.2019.04.008)
8. Randall D, Wood C, Perry S, Bergman H, Maloiy G, Mommsen T, Wright PA. 1989 Urea excretion as a strategy for survival in a fish living in a very alkaline environment. *Nature* **337**, 165–166. (doi:10.1038/337165a0)
9. Wilson PJ, Wood CM, Walsh PJ, Bergman AN, Bergman HL, Laurent P, White BN. 2004 Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiol. Biochem. Zool.* **77**, 537–555. (doi:10.1086/422054)
10. Roberts N, Taieb M, Barker P, Damnati B, Icole M, Williamson D. 1993 Timing of the Younger Dryas event in East Africa from lake-level changes. *Nature* **366**, 146–148. (doi:10.1038/366146a0)
11. Ford AGP, Dasmahapatra KK, Rüber L, Gharbi K, Cezard T, Day JJ. 2015 High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Mol. Ecol.* **24**, 3421–3440. (doi:10.1111/mec.13247)
12. Seegers L, Tichy H. 1999 The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya, with descriptions of two new species. *Ichthyol. Explor. Freshw.* **10**, 97–146.
13. Walsh P, Danulat E, Mommsen T. 1990 Variation in urea excretion in the gulf toadfish *Opsanus beta*. *Mar. Biol.* **106**, 323–328. (doi:10.1007/BF01344308)
14. Wright PA, Iwama GK, Wood CM. 1993 Ammonia and urea excretion in Lahontan cutthroat trout (*Oncorhynchus clarki* henshawi) adapted to the highly alkaline Pyramid Lake (pH 9.4). *J. Exp. Biol.* **175**, 153–172.
15. Wood CM, Wilson P, Bergman HL, Bergman AN, Laurent P, Otiang'a-Owiti G, Walsh PJ. 2002 Obligatory urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity and alkalinity. *Physiol. Biochem. Zool.* **75**, 111–122. (doi:10.1086/340626)
16. Kavembe GD, Franchini P, Irisarri I, Machado-Schiaffino G, Meyer A. 2015 Genomics of adaptation to multiple concurrent stresses: insights from comparative transcriptomics of a cichlid fish from one of earth's most extreme environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa. *J. Mol. Evol.* **81**, 90–109. (doi:10.1007/s00239-015-9696-6)
17. Rubio V, Cervera J. 1995 The carbamoyl-phosphate synthase family and carbamate kinase: structure-function studies. *Biochem. Soc. Trans.* **23**, 879–883. (doi:10.1042/bst0230879)
18. de Cima S, Polo LM, Díez-Fernández C, Martínez AI, Cervera J, Fita I, Rubio V. 2015 Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. *Sci. Rep.* **5**, 16950. (doi:10.1038/srep16950)
19. Hong J, Salo WL, Lusty CJ, Anderson PM. 1994 Carbamyl phosphate synthetase III, an evolutionary intermediate in the transition between glutamine-dependent and ammonia-dependent carbamyl phosphate synthetases. *Mol. Biol.* **243**, 131–140. (doi:10.1006/jmbi.1994.1638)
20. Loong A, Chng Y, Chew S, Wong W, Ip Y. 2012 Molecular characterization and mRNA expression of carbamoyl phosphate synthetase III in the liver of the African lungfish, *Protopterus annectens*, during aestivation or exposure to ammonia. *J. Comp. Physiol. B* **182**, 367–379. (doi:10.1007/s00360-011-0626-7)
21. Saeed-Kothe A, Powers-Lee SG. 2002 Specificity determining residues in ammonia- and glutamine-dependent carbamoyl phosphate synthetases. *J. Biol. Chem.* **277**, 7231–7238. (doi:10.1074/jbc.M110926200)
22. Ip AY, Chew SF. 2010 Ammonia production, excretion, toxicity, and defense in fish: a review. *Front. Physiol.* **1**, 134.
23. Mommsen TP, Walsh PJ. 1989 Evolution of urea synthesis in vertebrates: the piscine connection. *Science* **243**, 72–75. (doi:10.1126/science.2563172)
24. Lindley TE, Scheiderer CL, Walsh PJ, Wood CM, Bergman HL, Bergman AL, Laurent P, Wilson P, Anderson PM. 1999 Muscle as the primary site of urea cycle enzyme activity in an alkaline lake-adapted tilapia, *Oreochromis alcalicus grahami*. *J. Biol. Chem.* **274**, 29 858–29 861. (doi:10.1074/jbc.274.42.29858)
25. Rebeiz M, Tsiantis M. 2017 Enhancer evolution and the origins of morphological novelty. *Curr. Opin Genet. Dev.* **45**, 115–123. (doi:10.1016/j.gde.2017.04.006)
26. Ford AGP, Rüber L, Newton J, Dasmahapatra KK, Balarin JD, Bruun K, Day JJ. 2016 Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution* **70**, 2718–2735. (doi:10.1111/evo.13072)
27. Zimmer AM, Wood CM. 2016 Physiological and molecular ontogeny of branchial and extra-branchial urea excretion in posthatch rainbow trout (*Oncorhynchus mykiss*). *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **310**, R305–RR12. (doi:10.1152/ajpregu.00403.2015)
28. Korte JJ, Salo WL, Cabrera VM, Wright PA, Felskie AK, Anderson PM. 1997 Expression of carbamoyl-phosphate synthetase III mRNA during the early stages of development and in muscle of adult rainbow trout (*Oncorhynchus mykiss*). *J. Biol. Chem.* **272**, 6270–6277. (doi:10.1074/jbc.272.10.6270)
29. Banerjee B, Koner D, Hasan R, Saha N. 2020 Molecular characterization and ornithine-urea cycle genes expression in air-breathing magur catfish (*Clarias magur*) during exposure to high external ammonia. *Genomics* **112**, 2247–2260. (doi:10.1016/j.ygeno.2019.12.021)
30. Pownall ME, Gustafsson MK, Emerson Jr CP. 2002 Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **18**, 747–783. (doi:10.1146/annurev.cellbio.18.012502.105758)
31. Talbot JC, Teets EM, Ratnayake D, Duy PQ, Currie PD, Amacher SL. 2019 Muscle precursor cell movements in zebrafish are dynamic and require six family genes. *Development* **146**, dev171421. (doi:10.1242/dev.171421)
32. Bucher EA, Maisonpierre P, Konieczny S, Emerson C. 1988 Expression of the troponin complex genes: transcriptional coactivation during myoblast differentiation and independent control in heart and skeletal muscles. *Mol. Cell. Biol.* **8**, 4134–4142. (doi:10.1128/MCB.8.10.4134)
33. Wood CM *et al.* 2016 Mammalian metabolic rates in the hottest fish on earth. *Sci. Rep.* **6**, 26990. (doi:10.1038/srep26990)
34. Kong H, Edberg DD, Korte JJ, Salo WL, Wright PA, Anderson PM. 1998 Nitrogen excretion and expression of carbamoyl-phosphate synthetase III activity and mRNA in extrahepatic tissues of largemouth bass (*Micropterus salmoides*). *Arch. Biochem. Biophys.* **350**, 157–168. (doi:10.1006/abbi.1997.0522)
35. Randall DJ, Tsui T. 2002 Ammonia toxicity in fish. *Mar. Pollut. Bull.* **45**, 17–23. (doi:10.1016/S0025-326X(02)00227-8)
36. Mommsen TP, Walsh PJ. 1991 Urea synthesis in fishes: evolutionary and biochemical perspectives. *Biochem. Mol. Biol. Fish.* **1**, 137–163
37. LeMoine CM, Walsh PJ. 2013 Ontogeny of ornithine-urea cycle gene expression in zebrafish (*Danio rerio*). *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **304**, R991–R1000. (doi:10.1152/ajpregu.00411.2012)
38. Steele SL, Chadwick TD, Wright PA. 2001 Ammonia detoxification and localization of urea cycle enzyme activity in embryos of the rainbow trout (*Oncorhynchus mykiss*) in relation to early tolerance to high environmental ammonia levels. *J. Exp. Biol.* **204**, 2145–2154.
39. Fong AP, Yao Z, Zhong JW, Cao Y, Ruzzo WL, Gentleman RC, Tapscott SJ. 2012 Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev. Cell* **22**, 721–735. (doi:10.1016/j.devcel.2012.01.015)
40. Conerly ML, Yao Z, Zhong JW, Groudine M, Tapscott SJ. 2016 Distinct activities of Myf5 and MyoD indicate separate roles in skeletal muscle lineage specification and differentiation. *Dev. Cell* **36**, 375–385. (doi:10.1016/j.devcel.2016.01.021)
41. Weintraub H, Davis R, Lockshon D, Lassar A. 1990 MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation. *Proc. Natl Acad. Sci. USA* **87**, 5623–5627. (doi:10.1073/pnas.87.15.5623)
42. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004 Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717. (doi:10.1038/nature02415)

43. Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. 2004 Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol. Cell* **14**, 465–477. (doi:10.1016/S1097-2765(04)00260-6)
44. Christoffels VM, Grange T, Kaestner KH, Cole TJ, Darlington GJ, Croniger CM, Lamers WH. 1998 Glucocorticoid receptor, C/EBP, HNF3, and protein kinase A coordinately activate the glucocorticoid response unit of the carbamoylphosphate synthetase I gene. *Mol. Cell. Biol.* **18**, 6305–6315. (doi:10.1128/MCB.18.11.6305)
45. Fagernes CE, Stensløkken K-O, Røhr ÅK, Berenbrink M, Ellefsen S, Nilsson GE. 2017 Extreme anoxia tolerance in crucian carp and goldfish through neofunctionalization of duplicated genes creating a new ethanol-producing pyruvate decarboxylase pathway. *Sci. Rep.* **7**, 7884. (doi:10.1038/s41598-017-07385-4)
46. Johnston IA, Bernard LM. 1983 Utilization of the ethanol pathway in carp following exposure to anoxia. *J. Exp. Biol.* **104**, 73–78.
47. Dhillon RS, Mandic M, Yao L, Cao Z-D, Fu S-J, Brauner CJ, Wang YS, Richards JG. 2018 Ethanol metabolism varies with hypoxia tolerance in ten cyprinid species. *J. Comp. Physiol. B* **188**, 283–293. (doi:10.1007/s00360-017-1131-4)
48. Nilsson GE. 1988 A comparative study of aldehyde dehydrogenase and alcohol dehydrogenase activities in crucian carp and three other vertebrates: apparent adaptations to ethanol production. *J. Comp. Physiol. B* **158**, 479–485. (doi:10.1007/BF00691145)
49. Ip YK, Loong AM, Chng YR, Hiong KC, Chew SF. 2012 Hepatic carbamoyl phosphate synthetase (CPS) I and urea contents in the hylid tree frog, *Litoria caerulea:* transition from CPS III to CPS I. *J. Comp. Physiol. B* **182**, 1081–1094. (doi:10.1007/s00360-012-0682-7)

Journal of *Developmental Biology*

MDPI

Article

Exploring the Expression of Cardiac Regulators in a Vertebrate Extremophile: The Cichlid Fish *Oreochromis (Alcolapia) alcalica*

Gemma Sutton [1], Lewis J. White [1], Antonia G.P. Ford [2], Asilatu Shechonge [3], Julia J. Day [4], Kanchon K. Dasmahapatra [1] and Mary E. Pownall [1,*]

1 Biology Department, University of York, York YO10 5DD, UK; gs569@exeter.ac.uk (G.S.); ljw569@york.ac.uk (L.J.W.); kanchon.dasmahapatra@york.ac.uk (K.K.D.)
2 Department of Life Sciences, Centre for Research in Ecology, Whitelands College, University of Roehampton, Holybourne Avenue, London SW15 4JD, UK; Antonia.Ford@roehampton.ac.uk
3 Tanzania Fisheries Research Institute, P.O.BOX 98 Kyela, Mbeya, Tanzania; ashechonge@yahoo.com
4 Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK; j.day@ucl.ac.uk
* Correspondence: betsy.pownall@york.ac.uk; Tel.: +44-1904-328692

Received: 2 September 2020; Accepted: 30 September 2020; Published: 4 October 2020

Abstract: Although it is widely accepted that the cellular and molecular mechanisms of vertebrate cardiac development are evolutionarily conserved, this is on the basis of data from only a few model organisms suited to laboratory studies. Here, we investigate gene expression during cardiac development in the extremophile, non-model fish species, *Oreochromis (Alcolapia) alcalica*. We first characterise the early development of *O. alcalica* and observe extensive vascularisation across the yolk prior to hatching. We further investigate heart development by identifying and cloning *O. alcalica* orthologues of conserved cardiac transcription factors *gata4*, *tbx5*, and *mef2c* for analysis by in situ hybridisation. Expression of these three key cardiac developmental regulators also reveals other aspects of *O. alcalica* development, as these genes are expressed in developing blood, limb, eyes, and muscle, as well as the heart. Our data support the notion that *O. alcalica* is a direct-developing vertebrate that shares the highly conserved molecular regulation of the vertebrate body plan. However, the expression of *gata4* in *O. alcalica* reveals interesting differences in the development of the circulatory system distinct from that of the well-studied zebrafish. Understanding the development of *O. alcalica* embryos is an important step towards providing a model for future research into the adaptation to extreme conditions; this is particularly relevant given that anthropogenic-driven climate change will likely result in more freshwater organisms being exposed to less favourable conditions.

Keywords: cichlid fish; extremophile; environmental adaptation; cardiac myogenesis

1. Introduction

Aquatic ectotherms, such as fish, are significantly impacted by the changes or stressors in their environment, as they rely on the surrounding water for maintaining homeostasis [1,2]. Despite this restriction, fish have evolved morphological, biochemical, physiological, behavioural, and developmental mechanisms allowing them to colonise nearly all aquatic environments, including a variety of extreme habitats [3,4]. Although most fish species are unable to survive extremes of temperature, pH, salinity, and environments which seasonally dry out, some specialised species have adapted to thrive in such conditions, making them useful subjects for studying the molecular and developmental mechanisms underpinning adaptation to imposed stressors [5]. Exploring the naturally evolved adaptations in extremophile vertebrates will give insight into core biological

functions, which may provide important information for future research in vertebrate biology, as well as the diversity in animal development, and it may even contribute to our understanding of human disease [6–8].

Knowledge of development is key for understanding mechanisms of evolutionary variation, as changes in developmental processes can result in novel phenotypes [9]. Currently, the majority of knowledge of early ontogeny of fishes stems from studies in the model organisms zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*) [10,11]. These species were selected as suitable model organisms for developmental biology due to their size, short development periods, accessible genetic tools, and ability to be bred and reared easily in laboratory conditions. However, these selected teleosts may not be representative of early development in other fish species. The study of development in non-model species has the potential to provide additional insights into the molecular evolution of biological diversity in teleosts which are by far the most species-rich vertebrate clade. The Teleostei comprise ca. 34,000 species, accounting for nearly 98% of all actinopterygian (ray-finned) fishes, i.e., half of all extant vertebrates [12]. Research using non-model organisms will allow wider species comparisons and a better understanding of adaptive traits [13].

With their high species richness and recent adaptive radiations, cichlid fishes are a model system in evolutionary biology, particularly for studying speciation [14–16]. One subgenus, *Alcolapia* (nested within the genus *Oreochromis* [17]), presents a unique radiation of extremophile cichlids, inhabiting the extreme waters of the East African soda lakes of Natron and Magadi. This adaptive radiation comprises four described species, *Oreochromis (Alcolapia) alcalica*, *O. (A.) latilabris*, *O. (A.) ndalalani* (Lake Natron, Tanzania), and *O. (A.) grahami* (Lake Magadi, Kenya), and diverged as recently as 10,000 years ago from freshwater ancestors [18–20]. The *Alcolapia* fish experience water temperatures of 30–42 °C, pH 9–11.5, fluctuating dissolved oxygen levels (0.08–6.46 mg/L), and high salt concentrations (>20 ppt) [21]. Living in this environment has led to a number of key adaptations including 100% ureotelism, facultative airbreathing, a specialised gut morphology, and maintaining a heightened metabolic rate [22–27]

An aspect of development that shows striking conservation across animals with very different lifestyles, from insects to fish to tetrapods, is that of heart function to pump fluid around the body, transporting nutrients, metabolites, and oxygen to tissues [28–30]. How this remarkable adaptation develops during embryogenesis has been studied in several experimental model organisms, providing an understanding of conserved mechanisms of cardiac development in vertebrates [28]. As the primary teleost model organism, zebrafish has emerged as an important vertebrate model for studying cardiovascular development and disease [31].

Cardiac myogenesis in zebrafish begins with the emergence of cardiac progenitor cells at 5 h post fertilisation (hpf) in the lateral marginal zone of the blastula, with ventricular progenitors situated more dorsally and closer to the margin than atrial progenitors [32,33]. During gastrulation, the cardiac progenitors migrate to the anterior lateral plate mesoderm where bilateral progenitors fuse into the cardiac disc with endocardial cells in the centre, surrounded by ventricular cardiomyocytes, surrounded by atrial cardiomyocytes that elongate into the linear heart tube. By 24 hpf, heart tube contraction initiates [34,35] and the linear heart tube undergoes cardiac jogging, whereby left-right symmetry is broken as the heart tube migrates leftwards and begins looping [36]. At 48 hpf, the two-chambered heart is clearly distinguishable by the constriction of the atrioventricular canal [37].

The network of cardiac transcription factors that regulate embryonic heart development displays a high degree of evolutionary conservation across vertebrates; here, we identify and describe the expression of a subset of these regulators, GATA-binding protein 4 (GATA4), T-box 5 (Tbx5), and Myocyte enhancer factor 2c (Mef2c) in *Oreochromis (Alcolapia) alcalica*. Members of the GATA family of zinc-finger transcription factors are involved in the early specification of cardiac progenitors. The GATA factors are crucial during haematopoiesis and cardiac myogenesis in vertebrates and *Drosophila* [38,39]. Tbx5 is a member of the T-box family of transcription factors and is expressed during development of heart, eyes, and forelimbs. Mutations in *TBX5* in humans cause Holt–Oram syndrome (HOS), an autosomal-dominant disorder characterised by forelimb malformations and

cardiac defects [40–42]. Cardiac abnormalities of HOS patients include septation defects and conduction disease [43]. Myocyte enhancer factor 2 (Mef2) proteins are MADs (MCM1, Agamous, Deficiens, Serum response factor)-box transcription factors. There are four vertebrate *mef2* genes, *mef2a–d*, expressed in precursors of heart, skeletal, and smooth muscle lineages [44]. Loss of function of the single *mef2* gene in *Drosophila* results in abnormal development of all muscle types, including cardiomyocytes [45]. In mice, *Mef2a*, *c*, and *d* are expressed in cardiac mesoderm and *Mef2a*- or *Mef2c*-null mice display cardiac developmental defects [46].

We report here methods to acquire and culture *Oreochromis (Alcolapia) alcalica* embryos and to interrogate gene expression patterns using whole-mount in situ hybridisation. We present the first developmental stage series of *O. alcalica* and the expression patterns of the *O. alcalica* orthologues of the cardiac regulatory genes *gata4*, *tbx5*, and *mef2c* as assessed by in situ hybridisation.

2. Methods

Live specimens of *O. alcalica* were collected from a single spring (site 5) [20,47] at Lake Natron in Tanzania (permit 2017-259- NA-2011-182) in June 2017. Fish were packaged individually in breather bags and transported to Bangor University to establish a breeding colony. Some of the *O. alcalica* specimens were subsequently moved to the University of York and housed in a recirculating aquarium (Aquatics Habitat) with constant water conditions kept at a temperature of 30 °C, pH 9, and conductivity of 3800 µS. A separate zebrafish system was maintained at 27 °C, pH 7.4, and conductivity 800 µS. This study was carried out using procedures authorised by the United Kingdom (UK) Home Office in accordance with the Animals Scientific Procedures Act (1986) and approved by the Animal Welfare and Ethical Review Body at the University of York and the UK Home Office project licence to MEP (POF245295).

2.1. Fish maintenance and Embryo Collection

To control *O. alcalica* breeding, dividers were placed in tanks to separate a single male from a group of 3–4 females. After acclimatising for several days, dividers were removed and mating behaviour proceeded. *O. alcalica* is a mouth-brooding cichlid, in which females carrying fertilised eggs are identified by an enlarged buccal cavity. To obtain embryos, the females were removed from the water, and embryos were released by applying light pressure to the buccal cavity. After collection, *O. alcalica* embryos were incubated at 29–30 °C gently shaking in system water with fungicide methylene blue (0.0003 mg/mL) (Sigma). Zebrafish (AB) embryos were collected as per standard procedures. At specific stages, *O. alcalica* and zebrafish embryos were dechorionated and fixed for 1 h at room temperature in MEMFA (0.1 M 3-(*N*-Morpholino)propanesulfonic acid (MOPS), pH 7.4, 2 mM Ethylene glycol tetraacetic acid (EGTA), 1 mM $MgSO_4$, 3.7% formaldehyde) and stored at −20 °C in 100% methanol.

2.2. Phylogeny

Zebrafish *gata4*, *tbx5*, and *mef2c* coding sequences were identified on ZFIN (https://zfin.org/). BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to identify these sequences in the published genome of the freshwater cichlid *Oreochromis niloticus*. *Oreochromis niloticus* sequences were aligned to the unpublished *O. alcalica* genome (Dashamapatra, in preparation) to identify *O. alcalica* coding sequences of the cardiac regulators; we found these putative orthologues to have 60% (Tbx5), 63% (Mef2c), and 62% (GATA4) amino-acid sequence identity with the zebrafish proteins. The *gata4*, *tbx5*, and *mef2c* sequences were retrieved from *Drosophila melanogaster*, *Takifugu rubripes*, *Maylandia zebra*, *Nothobranchus furzeri*, *Oryzias latipes*, *Danio rerio*, *Callorhinchus milii*, *Xenopus tropicalis*, *Gallus gallus*, *Mus musculus*, *Pan troglodytes*, and *Homo sapiens* using BLAST. The predicted full-length amino-acid sequence was deduced from the coding sequence of each gene, and multiple sequence alignment was conducted using MEGA-X by MUSCLE (Robert C. Edgar, 2004; codon alignment). Phylogenetic tree reconstruction was conducted in MEGA-X on individual genes using maximum likelihood [48]. The tree with the highest log likelihood is presented in Figures 2–4. The percentage of trees in which

the associated taxa clustered together is shown next to the branches. Initial trees for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a Jones-Taylor-Thornton (JTT) model, and then selecting the topology with a superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites, and a tree was drawn to scale, with branch lengths measured in the number of substitutions per site.

2.3. RNA Extraction, Complementary DNA (cDNA) Synthesis, and Reverse-Transcription PCR (RT-PCR)

RNA was extracted from *O. alcalica* and zebrafish embryos using TRI Reagent (Sigma-Aldrich) using phase separation and isopropanol precipitation. Here, 1 µg of total RNA was used for cDNA synthesis with Superscript IV (Invitrogen) and random hexamers (Thermo Scientific). Primers (Sigma) for the amplification of *O. alcalica* cardiac regulatory genes were designed against sequences identified in the *O. alcalica* genome using PrimerSelect (DNASTAR) (Table 1) to amplify regions of 400–600 bp in length; a similar approach was used for zebrafish orthologues using sequences in NCBI. The cDNAs were amplified by RT-PCR, cloned into pGEM T-Easy (Promega), and sequenced. The following sequences were submitted to Genbank: *Oa tbx5* (MT904199), *Oa mef2c* (MT904200), and *Oa gata4* (MT904201). For generating antisense RNA probes, SalI (Promega) linearised plasmids were used as templates for T7 transcription (Ambion), and NcoI (Promega) linearised plasmids were used as templates for SP6 transcription (Ambion), depending on orientation of the insert. In vitro run-off transcription at 37 °C was used to incorporate digoxigenin (DIG)-labelled UTP analogue (Roche).

Table 1. *Oreochromis (Alcolapia) alcalica* and zebrafish primers for RT-PCR and probe synthesis.

Gene	Forward Primer 5′–3′	Reverse Primer 5′–3′
O. alcalica gata4	TGTCTCCGCGCTTCACCTTCTCCA	GACCGGCTCTCCCTCTGCGTTCC
O. alcalica mef2c	ATGGGGCGAAAGAAGAT	AGCCCACCCTGATTACTG
O. alcalica tbx5	AGAGGCAGCGACGACAATGAGC	GGGGGATAGGAGGAGGGGTGATAG
Zebrafish *gata4*	CCTACAGGCACCCCAGCAGAGCAG	CCCGCCGCCACAGAGGAGTC
Zebrafish *mef2ca*	TTGCGCGGATAATGGACGAACG	GGGGGCCGGTGGGTGACTC
Zebrafish *tbx5b*	GCTCCCCGCCACTACAAACTCAG	GATATGTCCGAAAGGGTCCAGGTG

2.4. Whole-Mount In Situ Hybridisation

Fixed embryos were rehydrated with washes of decreasing levels of methanol/phosphate-buffered saline-Tween-20 (PBSAT) (75% methanol/PBSAT, 50% methanol/PBSAT, 100% PBSAT) and treated with 10 µg/mg proteinase K (Roche) at room temperature. Zebrafish embryos were treated with proteinase K for 10 min per day post fertilisation (dpf). *O. alcalica* embryos of 2dpf, 3dpf, and 7dpf were treated with proteinase K for 4, 15, and 30 min, respectively. Embryos were treated with 0.1 M triethanolamine/acetic anhydride and washed in PBSAT. Following post-fixation with 10% formalin and 2 h incubation at 68 °C in hybridisation buffer (50% formamide (Ambion), 1 mg/mL total yeast RNA, 5× saline-sodium citrate (SSC), 100 µg/mL heparin, 1× Denharts, 0.1% Tween-20, 0.1% CHAPS, 10 mM EDTA), embryos were incubated in 1 mL of hybridisation buffer with 3–6 µL of DIG-labelled antisense probe at 68 °C overnight.

Embryos were extensively washed at 68 °C in 2× SSC + 0.1% Tween-20, 0.2× SSC + 0.1% Tween-20, and maleic acid buffer (MAB; 100 mM maleic acid, 150 mM NaCl, 0.1% Tween-20, pH 7.8). This was replaced with pre-incubation buffer (4× MAB, 10% BMB, 20% heat-treated lamb serum) for 2 h. Embryos were incubated overnight (rolling at 4 °C) with fresh pre-incubation buffer and 1/2000 dilution of anti-DIG coupled with alkaline phosphatase (AP) (Roche). Embryos were washed in 1× MAB and AP buffer (100 mM Tris, 50 mM MgCl, 100 mM NaCl, 0.1% Tween-20). BM purple (Roche), the substrate for AP, was applied to embryos and left at room temperature until colour developed. Embryos were fixed in MEMFA and photographed. The 7dpf *gata4* embryos were washed in PBSAT and bleached in 5% H_2O_2 (Merck) under bright light to remove pigmentation. Once bleached, embryos were

washed in PBSAT and stored in MEMFA. Stack images of the in situ hybridisations were taken using a SPOT Camera (14.2 Colour Mosaic, Diagnostic Instruments). Focus stacking and image editing was undertaken on Adobe Photoshop 2020.

3. Results

3.1. Early Development of O. alcalica

The early development of *O. alcalica* during the first 10 dpf at 28 °C is documented in Figure 1. Embryonic development could be subdivided into six periods: zygote (Figure 1A), cleavage (Figure 1B), blastula-gastrula (Figure 1C,D), segmentation (Figure 1E), pharyngula (Figure 1F–H), and hatching (Figure 1G). *Oreochromis (Alcolapia) alcalica*'s embryonic development is similar to that of other mouth-brooding cichlids [49–51]. The embryo was surrounded by the chorion, a transparent membrane that stuck closely to the egg which persisted until hatching (Figure 1A–H). The yolk was opaque yellow and homogeneous in appearance, making observations of the embryonic anlage difficult. At early embryonic stages, there was almost no perivitelline space between the chorion and yolk, and dechorionation was difficult without puncturing the yolk (Figure 1A–D).

Newly fertilised eggs of *O. alcalica* had an ovoid shape, with the animal pole narrower than the vegetal pole (Figure 1A). The blastodisc sat at the animal pole. The first cleavage furrow was meridional and occurred at 2 hpf, resulting in two blastomeres (Figure 2B). The further mitotic divisions that occur during the cleavage period resulting in many blastomeres were not distinguishable under a dissecting microscope. Furthermore, the specific timing and cellular dynamics of *O. alcalica* gastrulation were difficult to observe due to the yolk (Figure 1C,D). Epiboly was previously described in Nile tilapia, *Oreochromis niloticus*, another mouth-brooding cichlid that is closely related to *O. alcalica* [17], by sequential fixation of blastula-gastrula stage embryos [49]. As *O. alcalica* embryos dissociated rapidly when dechorionated and fixed before the segmentation stage, we were unable to document gastrulation in *O. alcalica* in this study.

By 42 hpf, the *O. alcalica* embryos completed gastrulation and entered segmentation period (Figure 1E). At 66 hpf, the embryos entered the pharyngula period. Melanocytes emerged from the embryonic axis and began migrating across the yolk. Embryos had a clearly formed head with unpigmented eyes (Figure 1F). At this stage, a linear heart tube formed, and rhythmic contractions could be observed (Supplementary Movie A). At 3 dpf, the head developed upwards from the yolk. Early vasculature development was observed in patches on the yolk and around the heart (Figure 1G). This early vasculature was pumped across the yolk, through the heart tube and embryo (Supplementary Movie B). By 4 dpf, the eyes developed pigmentation, the heart started to loop, and extensive vasculature developed across the yolk (Figure 1H, Supplementary Movie C). At 5 dpf, embryos hatched from the chorion membrane (Figure 1I). At 5–6 dpf, the heart looping was complete and the two-chambered heart formed (Figure 1I–J, Supplementary Movie D). After hatching, many aspects of the adult body plan were apparent, and embryos could swim actively by 7 dpf (Figure 1J–M).

O. alcalica rapidly develops adult morphology without a prolonged, free-feeding larval stage [49,51] and, therefore, fits the definition of a direct-developing fish species which have large, yolk-rich eggs and complete their development while living off the maternally deposited yolk supply, until transforming directly into a free-feeding juvenile [52,53].

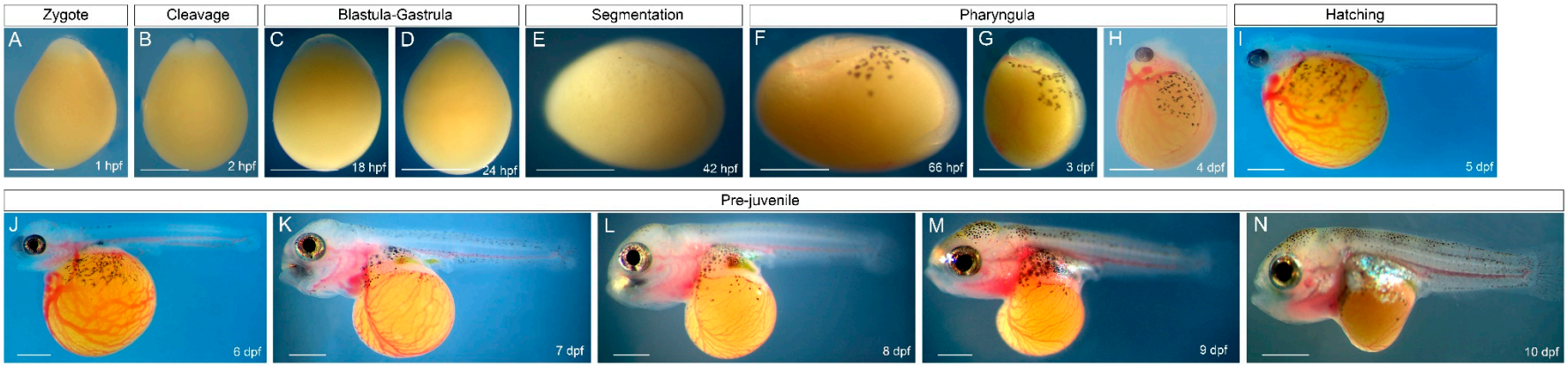


Figure 1. Embryonic and pre-juvenile development of *Oreochromis (Alcolapia) alcalica*. (**A**) At 1 h post fertilisation (hpf), *O. alcalica* embryos were ovoid in shape with the embryonic region located at the animal pole, on top of the large yellow yolk. (**B**) At 2 hpf, the first cell division occurred and two blastomeres were observed. (**C**,**D**) The embryo underwent gastrulation (18 hpf–24 hpf). The cellular dynamics of *O. alcalica* gastrulation were not easily observed due to the large yellow yolk. (**E**) Gastrulation was completed by 42 hpf. The antero-posterior axis could be distinguished, and the neural tube formed. (**F**) At 66 hpf, the antero-posterior axis elongated further. Pigmented melanocytes of the neural crest lineage emerged from the embryonic axis. A clear head and unpigmented eye could be distinguished. (**G**) By 3 days post fertilisation (dpf), there was an increased number of melanocytes migrating on the yolk. The embryo began to vascularise in patches on the yolk and in the area of the beating heart. (**H**) At 4 dpf, the vascular system extended across the yolk and a clearly beating heart was visible with strong blood circulation. The eye developed pigmentation. (**A**–**H**) Throughout these stages, the embryo was surrounded by the chorion membrane. (**I**) At 5 dpf, embryos hatched from the chorion. (**J**–**N**) Following hatching at 6–10 dpf, the yolk decreased in size and embryos rapidly developed adult morphology. (**A**–**D**) Animal pole to the top. (**E**,**F**,**I**–**N**) Lateral views; anterior to the left. (**G**,**H**) Lateral views; anterior to the top. Scale bars correspond to 1 mm. Heart contraction was observed and recorded at 66 hpf, 3 dpf, 4 dpf, and 5 dpf (see Supplementary Movies Supplementary A–D).

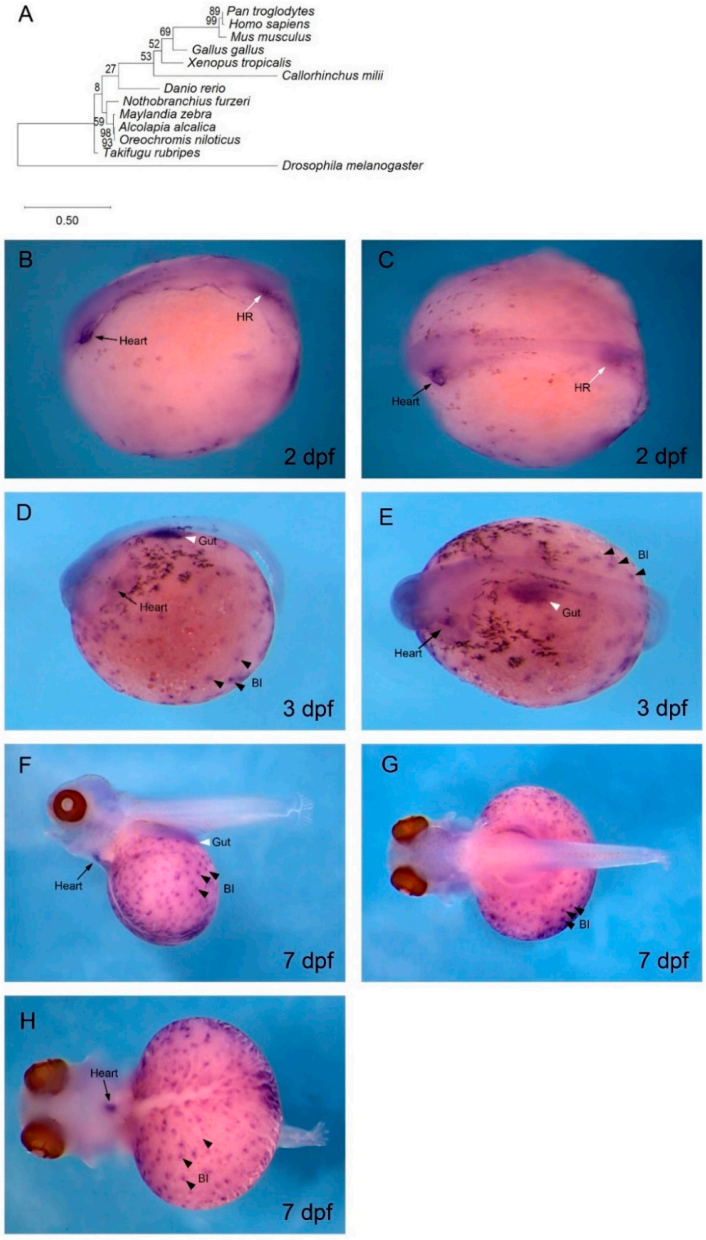


Figure 2. *gata4* is expressed in cardiac and vascular regions of *O. alcalica* embryos. (**A**) Phylogenetic analysis of GATA4. The evolutionary history was inferred using the maximum likelihood method and Dayhoff with frequency model and gamma distribution [58]. (**B**–**H**) In situ hybridisation of *O. alcalica gata4* at 2 days post fertilisation (dpf) (**B**,**C**), 3 dpf (**D**,**E**), and 7 dpf (**F**–**H**). At 2 dpf, *Oa gata4* was expressed in the heart and haematopoietic region (HR) (**B**,**C**). At 3 dpf and 7 dpf, *Oa gata4* was expressed in the heart, gut, and blood islands (BI) (**D**–**H**). At 7 dpf, embryos were bleached to remove melanocyte pigmentation (**F**–**H**). Black arrows denote the heart, black arrowheads mark blood islands (BI), white arrows denote the haematopoietic region (HR), and white arrowheads mark the gut. (**B**,**D**,**F**) Lateral views; anterior to the left. (**C**,**E**,**G**) Dorsal views; anterior to the left. (**H**) Anterior view; dorsal to the left.

One of the most striking aspects of *O. alcalica* development was the emergence of an extensive vascular system which began in patches on the yolk and subsequently branched and formed a network of vessels across the yolk between 3 and 4 dpf (Figure 1G,H). In the closely related freshwater Nile tilapia (*O. niloticus*), the earliest heartbeat and blood circulation were reported at 40–42 hpf, and yolk vascularisation was also observed at 4 dpf [49,54]. When cultured at about the same temperature

used in these previous studies (28–30 °C), we found and recorded heart contractions at 66 hpf, 3 dpf, 4 dpf, and 5 dpf, as shown in movies in the Supplementary Materials. The early vascularisation seen in these cichlids is very unlike that of *Danio rerio* and shows more similarity to vascularisation seen in aminotes such as chick embryos. To investigate this further, the genes for the conserved cardiac transcription factors, *gata4*, *tbx5*, and *mef2c*, were cloned from extracted RNA and used for in situ hybridisation studies in *O. alcalica*. The expression of these genes was previously described in zebrafish [55–57]; nevertheless, they are included here for comparison. To do this, *gata4*, *tbx5*, and *mef2c* were cloned from zebrafish and analysed by in situ hybridisation, as presented in Figures S1–S3 (Supplementary Materials).

3.2. *gata4* is Expressed in Cardiac and Haematopoietic Regions of *O. alcalica* Embryos

To confirm that the cDNA isolated from *O. alcalica* was indeed *gata4*, a phylogenetic analysis was undertaken and revealed high conservation of our sequence with known GATA4 proteins (Figure 2A). In situ hybridisation analysis of embryos fixed at 2 dpf showed that the expression of *Oa gata4* in the linear heart tube was localised at the anterior left side of the embryo (Figure 2B,C), consistent with our observations of the early pharyngula stage, where the contracting linear heart tube forms (Figure 1F, Supplementary Movie A). The *gata4* expression in the heart tube was more restricted by 3 dpf (Figure 2E,F), and, by 7 dpf, *gata4* transcripts were still detected in a limited region of the heart. The *gata4* in situ hybridisation of zebrafish at 1 dpf identified *gata4*-expressing cardiomyocytes in the anterior lateral plate mesoderm (Figure S1A,B), and, by 2 dpf and 3 dpf, *gata4* transcripts were expressed in the anterior midline (Figure S1C–F). Other regions of *gata4* expression were identified at the posterior of the embryo in the haematopoietic region (Figure 2B,C).

Later in development, at 3 dpf and 7 dpf, *gata4* transcripts localised to numerous puncta across the yolk (Figure 2D–H). The embryos in Figure 2F–H were bleached to remove pigmentation from melanocytes, allowing better visualisation of these puncta. The *gata4* expression on the yolk, overlapping with the extensive vasculature, was first distinguishable at 3 dpf (Figure 1F). These extraembryonic *gata4*-expressing puncta were likely blood islands (Figure 2D–H), a feature of haematopoiesis that is absent in zebrafish, but present during amniote development.

As *O. alcalica* are direct-developers, they are more dependent on the yolk as an energy source than indirect-developers that have a free-feeding larval stage such as zebrafish. The development of this early vasculature from blood islands likely provides a vital, early energy supply in *O. alcalica* embryos. Interestingly, it is not *GATA4*, but *GATA1* and *GATA2* that are expressed in the blood islands of chicks and mice [59,60].

3.3. *tbx5* is Expressed in the Developing Pectoral Fin, Eyes, and Heart of *O. alcalica*

The expression of *Oa tbx5* was found in the heart, limb, and eye by in situ hybridisation, consistent with its expression in other vertebrate embryos. Phylogenetic analysis of Tbx5 protein in multiple vertebrates confirmed the identity of the *tbx5* cDNA isolated from *O. alcalica* (Figure 3A). At 2 dpf and 3 dpf, *Oa tbx5* transcripts were detected and localised specifically to the dorsal region of the eyes (Figure 3B,C). This expression in the eye was easily visualised prior to pigmentation of the retina in *O. alcalica*; however, by 7dpf, *Oa tbx5* expression in the eye was not visible in the retinal pigment epithelium (RPE) due to the black pigmentation (Figure 3F,G). This expression of *Oa tbx5* in the dorsal retina is conserved in vertebrates [61–63], and, in zebrafish, *tbx5b* was expressed in the dorsal retina from 1–3 dpf (Figure S2).

At 3 dpf, there was transient expression of *Oa tbx5* in the heart (Figure 3D), as also seen in zebrafish (Figure S2). The expression of *Oa tbx5* in *O. alcalica* hearts at 3 dpf, during the linear heart tube stage, is consistent with the known role of Tbx5 in regulating cardiac looping and jogging in zebrafish. At 3 dpf, *Oa tbx5* transcripts localised to the *O. alcalica* pectoral limb bud (LB) (Figure 3D,E). *Oa tbx5* was still expressed in the LB at 7 dpf (Figure 3F, G). Similarly, in zebrafish, *tbx5b* was expressed in the LB from 1–3 dpf (Figure S2).

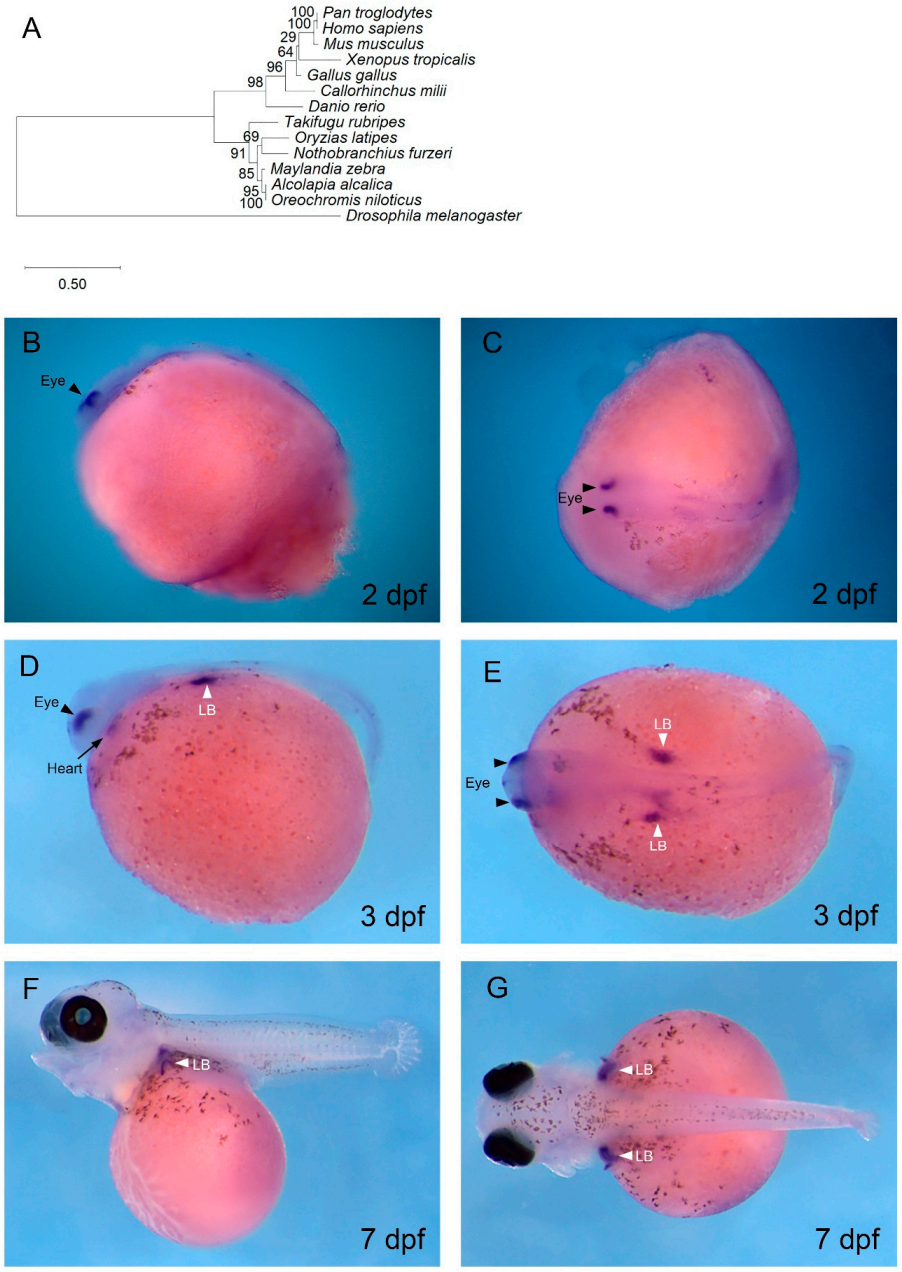


Figure 3. *tbx5* expression in *O. alcalica* dorsal retina, heart, and pectoral limb buds. (**A**) Phylogenetic analysis of Tbx5. The evolutionary history was inferred using the maximum likelihood method and Jones with frequency model [64]. (**B**–**G**) In situ hybridisation of *O. alcalica tbx5* at 2 days post fertilisation (dpf) (**B**,**C**), 3 dpf (**D**,**E**), and 7 dpf (**F**,**G**). At 2 dpf, *tbx5* was expressed in the dorsal region of the eye (**B**,**C**). At 3 dpf, *tbx5* was expressed in the eye, heart, and limb bud (LB) (**D**,**E**). At 7 dpf, *tbx5* was expressed in the LB (**F**,**G**). Black arrows denote the heart, black arrowheads mark the eyes, and white arrowheads denote the limb buds (LB). (**B**,**D**,**F**) Lateral views; anterior to the left. (**C**,**E**,**G**) Dorsal views; anterior to the left.

3.4. *mef2c is Expressed in Developing Muscle in O. alcalica*

The expression of the highly conserved cardiomyogenic regulator, *mef2c*, was assessed in *O. alcalica* by in situ hybridisation. In *Drosophila*, *D-mef2* is essential for the formation of cardiac muscle, and loss of *D-mef2* results in failure of cardiomyocyte differentiation [45,65]. In zebrafish, Mef2c is also required for cardiac development; combinatorial loss-of-function of *mef2ca* and *mef2cb* causes developmental arrest of cardiomyocytes. These *mef2c* paralogues control the expression of myocardial sarcomeric

genes [66]. Phylogenetic analysis of Mef2c protein in multiple vertebrates confirmed the identity of the *Oa mef2c* cDNA isolated from *O. alcalica* (Figure 4A), and it was distinct from the highly related Mef2d sequences (Supplementary Materials). Consistent with previous studies, *Dr mef2ca* transcripts were detected by in situ hybridisation in early cardiac regions in zebrafish (Figure S3); however, we found no expression of *Oa mef2c* in cardiac regions of *O. alcalica*. The expression shown in Figure 4B–E is similar to that described in *Xenopus laevis* [67], and we expect that the stages we analysed were too late to detect *Oa mef2c* in the heart because it is an early regulator of cardiac differentiation. Nonetheless, our expression analysis provides insight into the process of myogenesis in *O. alcalica;* at 2 dpf and 3 dpf, *Oa mef2c* was robustly expressed in the somites (Figure 4B–E), the embryonic source of all skeletal muscle in the vertebrate body [68]. This expression is conserved in zebrafish where *mef2ca* transcripts were also detected in somites from 1–3 dpf (Figure S3).

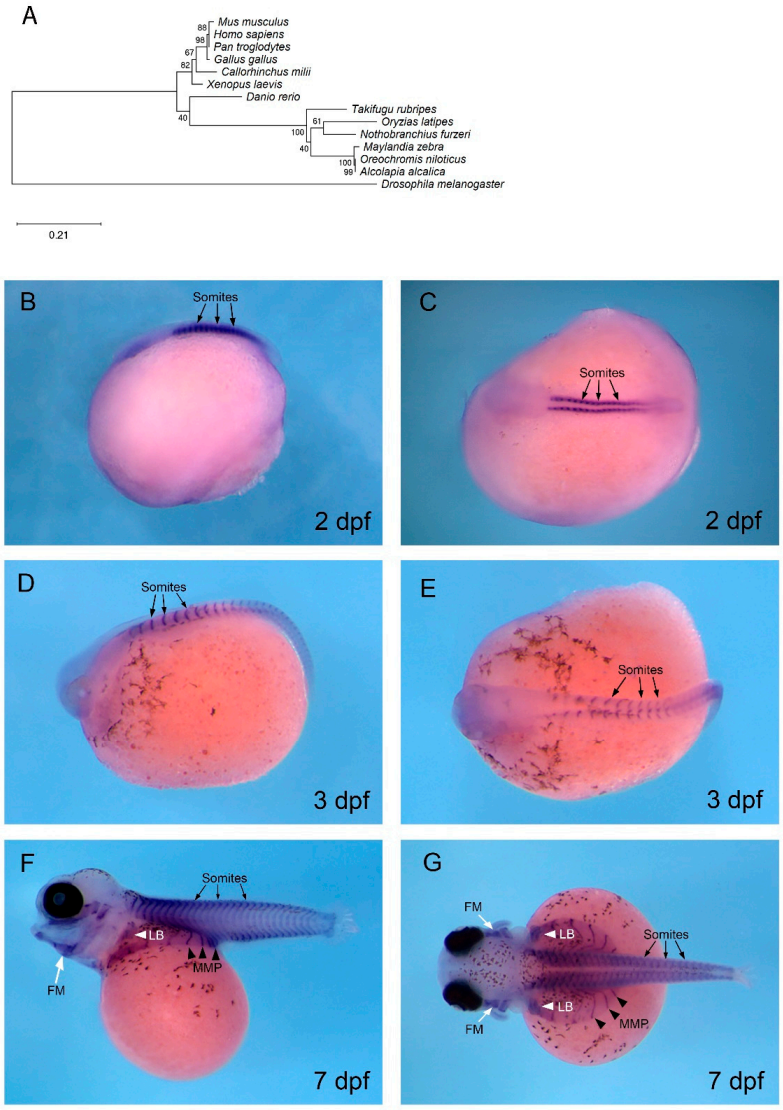


Figure 4. *mef2c* is expressed in developing muscle in *O. alcalica*. (**A**) Phylogenetic analysis of Mef2c. The evolutionary history was inferred by the maximum likelihood method and Jones with frequency model [64]. (**B**–**G**) In situ hybridisation of *O. alcalica mef2c* at 2 days post fertilisation (dpf) (**B**,**C**), 3 dpf (**D**,**E**), and 7 dpf (**F**,**G**). At 2 dpf and 3 dpf, *mef2c* was expressed in the somites (**B**–**E**). At 7 dpf, *mef2c* was expressed in somites, limb buds (LB), migrating muscle precursors (MMP), and facial muscle (FM). Black arrows denote the somites, black arrowheads mark migrating muscle precursors (MMP), white arrows denote facial muscle (FM), and white arrowheads mark limb buds (LB). (**B**,**D**,**F**) Lateral views; anterior to the left. (**C**,**E**,**G**) Dorsal views; anterior to the left.

At 7 dpf, *Oa mef2c* continued to be expressed in the distinctive chevron structure of the somites and in hypaxial muscle. These *Oa mef2c*-expressing streams of migratory muscle precursors (MMPs) migrated to form the muscles of the ventral body wall (arrowheads in Figure 4F,G). In *O. alcalica*, *Oa mef2c* expression revealed that the MMPs spread out as direct extensions of anterior somites that extended ventrally across the yolk; these extensions were not seen in zebrafish embryos.

4. Discussion

We report here the expression of three evolutionarily conserved regulators of cardiogenesis in the context of a direct-developing extremophile fish, *Oreochromis (Alcolapia) alcalica*. We include a developmental stage series and a set of videos to illustrate the onset of heart development and contraction.

4.1. gata4

Bilateral cardiomyocytes in the anterior lateral plate mesoderm can be identified through expression of *gata4/5/6* [69,70]. GATA4, in particular, is a potent driver of cardiac myogenesis; ectopically expressed *gata4* induces pluripotent *Xenopus* animal cap organoids to form contracting cardiomyocyte tissue [71]. Furthermore, human congenital heart defects are linked to mutations in *GATA4*, including valve and septal defects [72,73]. Morpholino oligonucleotide (MO) knockdown of *gata4/5/6* in *Xenopus* and zebrafish embryos attests to their roles in cardiac myogenesis. *Xenopus* morphants of individual GATA factors display cardia bifida, whereas knockdown of all three GATA factors eliminated expression of markers of cardiac differentiation [74]. In zebrafish, however, MO knockdown of *gata5* and *gata6* resulted in cardia bifida and substantial reduction in expression of contractile protein genes, whereas *gata4* morphant cardiomyocytes migrated to the midline normally [74]. The difference in phenotype severity of *gata5* morphants in *Xenopus* and zebrafish suggests a change in the activity of these GATA factors during evolution [75].

Our finding that *Oa gata4* is expressed in blood islands, the bipotential precursors of both vascular endothelia and blood cells, is interesting. In zebrafish and mammals, haematopoiesis occurs in shifting anatomical regions of the developing embryo [76]. In zebrafish, haematopoietic stem cells (HSCs) arise from the hemogenic endothelium lining the ventral wall of the dorsal aorta (VDA) [77]. HSCs originating from the VDA seed three haematopoietic organs: the caudal haematopoietic tissue (CHT), the thymus, and the kidney, the adult site of haematopoiesis in zebrafish [78]. In mammals, the VDA is functionally equivalent to the aorta-gonad-mesonephros (AGM) [79,80]. Mammalian HSCs emerge from the AGM, and transiently expand in the foetal liver before being maintained in the bone marrow, the adult site of haematopoiesis [81–83]. The mammalian foetal liver is equivalent to the zebrafish CHT, a vascularised region in the tail [84]. It was shown that primary cell lines generated from the zebrafish CHT at 3 dpf support HSC proliferation and differentiation [85]. The posterior region expressing *Oa gata4* in *O. alcalica*, labelled as the haematopoietic region (Figure 2B,C), may be similar to the CHT in zebrafish.

Blood islands in mice and chick form as clusters of mesodermal cells in the yolk sac that accumulate haemoglobin and are surrounded by outer endodermal cells that flatten and later form the endothelium of blood vessels [86]. The mesodermal and endodermal lineages that form the blood islands derive from common precursors known as hemangioblasts [87]. Zebrafish do not form yolk sac blood islands during embryogenesis. However, in some fish, such as killifish (*Fundulus spp.*), angelfish (*Pterophylum scalare*), and chondrichthyans, blood development predominantly occurs in the yolk sac [88]. This suggests that there is diversity in the location of embryonic blood development in teleosts.

The sites of haematopoiesis and vasculogenesis in *O. alcalica* embryos are more aligned with blood development in amniotes than with zebrafish. In vertebrates, the conserved GATA factors are divided into two subfamilies; *gata1/2/3* are expressed in developing blood cells, and *gata4/5/6* are expressed in mesodermal and endodermal tissues such as heart, liver, pancreas, and gut [89]. In the murine yolk sac, *GATA1* and *GATA2* transcripts were detected in blood islands by in situ hybridisation. Furthermore,

GATA1- and *GATA2*-null mouse embryos are deficient in erythroid cells [60,90–92]. This suggests that, in mice, *GATA1* and *GATA2* are crucial for haematopoietic development in the yolk sac. Our novel finding that *Oa gata4* transcripts localise to blood islands of *O. alcalica* embryos raises the possibility of the neofunctionalisation of *Alcolapia* GATA factors. Therefore, it would be valuable to undertake further research with the aim of characterising the expression and function of other GATA factors in *O. alcalica* development.

4.2. *tbx5*

Due to its clinical relevance, *tbx5* loss-of-function vertebrate models have been developed [93,94]. Septation is the process whereby the looped heart tube transitions into a multi-chambered heart; this occurs via the addition of cardiomyocytes that contribute to the atrium and the inflow tract [95]. Conditional knockout mice, whereby Tbx5 is haplo-insufficient only in these cardiomyocytes, showed that *tbx5* expression is required for the proliferation of atrial septum progenitors [96]. Furthermore, tissue-specific deletion of *tbx5* in ventricular cardiomyocytes of mice resulted in cardiac contraction dysfunction, which reflects HOS patient symptoms [97]. A *tbx5* mutant strain has been developed in zebrafish, known as the *heartstrings* mutant. In *heartstrings* mutants, the heart tube at 1 dpf is indistinguishable from wild type. Similar to HOS patients and the conditional knockout mice, they display a reduction in rate of contraction. *Heartstrings* hearts do not undergo looping but remain as linear heart tubes. By 3–4 dpf, they stretch to a string-like morphology [94]. Furthermore, double MO knockdowns of both *tbx5* paralogues display defects in cardiac jogging and looping; linear heart tubes remain at the midline or jog to the right, and looping is either abolished or reversed [98].

The role of *tbx5* in eye development has been examined in chick embryos by misexpressing this gene in the ventral retina. This resulted in upregulation of dorsal markers and downregulation of ventral markers, as well as altered projections of retinal ganglion cells [99]. In zebrafish, MO knockdown of both *tbx5* paralogues, *tbx5a* and *tbx5b*, significantly reduces the expression domains of dorsal markers in the eye [98]. Furthermore, transcriptomic analysis revealed that 54 genes were differentially expressed in the eye in these double-morphant embryos [100]. This suggests that *tbx5* is critical in dorsal–ventral patterning of the developing retina in vertebrates, which is consistent with this specific expression observed in *O. alcalica*.

Fish pectoral fins are homologous to tetrapod forelimbs, and genetic loss of function of *tbx5* in mice results in absence of forelimbs [101–103]. Similarly, in zebrafish, *tbx5b* morphants develop smaller limbs than controls [98]. The LB is established at a specific antero-posterior position of the lateral plate mesoderm. Limb mesenchyme precursors protrude from the trunk to form the LB. At the distal tip of the LB is the apical ectodermal ridge which controls LB growth [104]. *tbx5* is the earliest gene to be expressed in the presumptive pectoral fin mesenchyme, and the regulation of *tbx5* expression in vertebrates has been crucial to understanding the evolutionary origin of paired appendages. In situ hybridisation analysis of *tbx5* during development of the vertebrate sea lamprey (*Petromyzon marinus*), which does not have paired appendages, found *tbx5* expression to be exclusive to the heart field. It was, therefore, hypothesised that *tbx5* expression in the lateral plate mesoderm posterior to the heart is associated with the evolution of pectoral fins [105]. A study aimed at identifying a limb-specific enhancer of *tbx5* identified a potential *cis*-regulatory element downstream from *tbx5* conserved in vertebrates, known as cns12sh. Injection of *tbx5* alongside cns12sh into *heartstrings* embryos, which do not develop pectoral limbs, rescued LB development [105]. However, CRISPR/Cas9 knockout of endogenous cns12sh regulatory elements in mice and zebrafish exhibited normal *tbx5* expression and LB initiation, suggesting there may be other DNA elements that share redundant function or that cns12sh is a pseudo-enhancer that was once required for LB activation but was lost during evolution [106]. Therefore, noncoding DNA elements that regulate *tbx5* expression in the vertebrate LB remain elusive. Further limb enhancers could be identified by comparative analyses of the genomic landscape of vertebrate *tbx5* including *O. alcalica*, which shares this specific expression in the LB.

4.3. *mef2c*

An essential role for Mef2c in heart development was revealed when *Mef2c*-null mice were found to be embryonic lethal due to abnormalities in inflow tract and outflow tract formation, and insufficient cardiac looping results in complete absence of the right ventricle [107–109]. More recently, this phenotype was recapitulated in mice with cardiomyocyte-specific *Mef2c* deletions [110]. Similarly, zebrafish Mef2c is crucial in early heart and skeletal muscle development [66,111,112]. Two *mef2c* paralogues exist in zebrafish: *mef2ca* and *mef2cb* [113]. *Mef2cb* is expressed in cardiomyocytes, and MO knockdown of *mef2cb* eliminates cardiomyocytes in the outflow tract, reminiscent of *Mef2c*-null mice [112]. Combinatorial loss-of-function analysis of both paralogues revealed an essential role of *mef2ca* and *b* in controlling the expression of myocardial sarcomeric genes in cardiomyocytes [66].

Although we found no cardiac expression of *O. alcalica mef2c* at the stages we investigated (Figure 4B–G), we conclude that this is because we did not look early enough. In a study of the Mef2 family in Atlantic cod (*Gadus morhua*), *mef2c* was robustly expressed at the cardiac ring stage, during cardiac ring extension at the arterial and venous poles, and after completion of heart formation [114]. Although invertebrates only have a single *mef2* gene, amniotes have four genes (*Mef2a-d*) [115]. The teleost-specific gene duplication has led to six *mef2* genes in zebrafish that undergo extensive alternative splicing [66,116].

We did find extensive expression of *Oa mef2c* in the skeletal muscle cell lineage (in the somites, limb buds, facial muscle, and MMPs) which is also seen in *Xenopus* [67] and zebrafish ([66] (Figure S3C–F). Cells become committed to the myogenic lineage via the cooperation of myogenic regulatory factors (MRFs) Myod, Myf5, Myogenin, and Mrf4, [117] and members of the MEF2 family. *mef2* genes are activated by MRFs and expressed during the terminal differentiation of muscle cells [44,118]. It has been shown that *mef2* genes are required in myogenesis in zebrafish, as *mef2* knockdown by MOs causes downregulation of messenger RNAs (mRNAs) encoding thick filament proteins (myosins), resulting in disrupted sarcomere assembly [111].

Two distinct mechanisms of MMP migration have been described for the body wall and appendage musculature. There is the primitive mechanism of direct epithelial somitic extension, whereby differentiated muscle of the somite extends directly to provide musculature of the body wall. Alternatively, there is a mechanism of long-range migration, whereby MMPs delaminate from somites after undertaking epithelial-to-mesenchymal transition and undergo long-range migration [119]. In zebrafish fin development, some hypaxial muscle precursors are specified in the somites and undergo long-range migration. This was established by assessing expression of *lbx1*, a gene that specifically labels limb muscle precursors [120,121]. In contrast, the body wall musculature of amniotes and the development of chondrichthyan fin muscles are formed via direct extension from the somites [121–123]. Therefore, this expression analysis of *mef2c* in *O. alcalica* shows body wall musculature extending from somites, consistent with findings in amniotes [122]. In contrast, it appears that *Oa mef2c*-expressing myocytes in the LB are separated from somites, indicating that long-range migration is the likely mechanism via which muscle precursors move to the LB, similar to terrestrial vertebrates. To conclusively determine the mechanism of hypaxial migration, further expression analyses of genes regulating MMP migration, such as *lbx1* [120,121], should be undertaken.

5. Perspectives

Understanding acquired adaptations of *O. alcalica* that allow it to thrive in extreme conditions could be useful when projecting resilience of other freshwater vertebrates living in ecosystems affected by climate change. The developmental stage series presented here reveals *O. alcalica* as a direct-developing teleost and, together with the described methods for undertaking expression analyses, this forms the basis of comparative developmental work with other cichlids. A notable limitation of this study was the difficulty in analysing *O. alcalica* blastula–gastrula stages of development, and, for future studies, it would be valuable to develop a method of observing gene expression in these early embryos.

Overall, our findings show that *O. alcalica* shares the highly conserved molecular regulation of the vertebrate body plan. *gata4* and *tbx5* are expressed in the *O. alcalica* embryonic heart, consistent with these genes having a conserved role in cardiac development in this species. Surprisingly, we also found *Oa gata4* localised to haematopoietic regions of the developing embryo, including yolk sac blood islands, which are characteristic of amniote blood development. The prominence of *Oa gata4* expression in haematopoietic regions sets it apart from other vertebrates; thus, it would be informative to characterise the phylogeny and expression of other members of the GATA family in *O. alcalica*.

Supplementary Materials: The following are available online at http://www.mdpi.com/2221-3759/8/4/22/s1: Figure S1. Expression of cardiac regulator *gata4* in zebrafish; Figure S2. Expression of cardiac regulator *tbx5b* in zebrafish; Figure S3. Expression of cardiac regulator *mef2ca* in zebrafish; Movies: Contracting heart during *O. alcalica* development.

Author Contributions: Data curation, G.S. and L.J.W.; investigation, A.G.P.F. and A.S.; methodology, L.J.W.; supervision, K.K.D. and M.E.P.; writing—original draft, G.S.; writing—review and editing, A.G.P.F. and J.J.D. All authors have read and agreed to the published version of the manuscript.

Funding: L.J.W. was supported by a BBSRC PhD studentship (BB/M011151/1) which funded the work.

Acknowledgments: G.S. carried out this work as part of an undergraduate project at the University of York. The fish were collected from Lake Natron by L.J.W., J.J.D., A.G.P.F. and A.S. with the Tanzania Fishery Research Institute (TAFIRI) (COSTECH permit no. 2017-259- NA-2011-182).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Evans, D.H.; Piermarini, P.M.; Choe, K.P. The Multifunctional Fish Gill: Dominant Site of Gas Exchange, Osmoregulation, Acid-Base Regulation, and Excretion of Nitrogenous Waste. *Physiol. Rev.* **2005**, *85*, 97–177. [CrossRef]
2. Hwang, P.-P.; Lee, T.-H.; Lin, L.-Y. Ion regulation in fish gills: Recent progress in the cellular and molecular mechanisms. *Am. J. Physiol. Integr. Comp. Physiol.* **2011**, *301*, R28–R47. [CrossRef]
3. Helfman, G.; Collette, B.B.; Facey, D.E.; Bowen, B.W. *The Diversity of Fishes: Biology, Evolution, and Ecology*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
4. Wang, Y.; Guo, B. Adaption to extreme environments: A perspective from fish genomics. *Rev. Fish. Biol. Fish.* **2019**, *29*, 735–747. [CrossRef]
5. Riesch, R.; Tobler, M.; Plath, M. *Extremophile Fishes: Ecology, Evolution, and Physiology of Teleosts in Extreme Environments*; Springer: New York, NY, USA, 2015.
6. Garbarino, V.R.; Orr, M.E.; Rodriguez, K.A.; Buffenstein, R. Mechanisms of oxidative stress resistance in the brain: Lessons learned from hypoxia tolerant extremophilic vertebrates. *Arch. Biochem. Biophys.* **2015**, *576*, 8–16. [CrossRef] [PubMed]
7. Irwin, J.A. Extremophiles and their application to veterinary medicine. *Environ. Technol.* **2010**, *31*, 857–869. [CrossRef] [PubMed]
8. Jorge, C.D.; Borges, N.; Bagyan, I.; Bilstein, A.; Santos, H. Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare. *Extremophiles* **2016**, *20*, 251–259. [CrossRef]
9. Jeffery, W.R. Chapter 8 Evolution and Development in the Cavefish *Astyanax*. *Curr. Top. Dev. Biol.* **2009**, *86*, 191–221. [CrossRef]
10. Kimmel, C.B.; Ballard, W.W.; Kimmel, S.R.; Ullmann, B.; Schilling, T.F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **1995**, *203*, 253–310. [CrossRef]
11. Iwamatsu, T. Stages of normal development in the medaka *Oryzias latipes*. *Mech. Dev.* **2004**, *121*, 605–618. [CrossRef]
12. Froese, R.; Pauly, D.; FishBase. World Wide Web electronic publication. Available online: https://www.scienceopen.com/document?vid=dc419213-0ca3-48cc-901c-2934ecf4441e (accessed on 2 January 2014).
13. Crawford, D.L. Functional genomics does not have to be limited to a few select organisms. *Genome Biol.* **2001**, *2*, 1001.
14. Kocher, T.D. Adaptive evolution and explosive speciation: The cichlid fish model. *Nat. Rev. Genet.* **2004**, *5*, 288–298. [CrossRef] [PubMed]
15. Salzburger, W.; Meyer, A. The species flocks of East African cichlid fishes: Recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* **2004**, *91*, 277–290. [CrossRef] [PubMed]

16. McGee, M.D.; Faircloth, B.C.; Borstein, S.R.; Zheng, J.; Hulsey, C.D.; Wainwright, P.C.; Alfaro, M.E. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. *Proc. R. Soc.* **2016**, *283*, 20151413. [CrossRef]
17. Ford, A.G.P.; Bullen, T.R.; Pang, L.; Genner, M.J.; Bills, R.; Flouri, T.; Ngatunga, B.P.; Ruber, L.; Schliewen, U.K.; Seehausen, O.; et al. Molecular phylogeny of *Oreochromis* (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments. *Mol. Phylogenetics Evol.* **2019**, *136*, 215–226. [CrossRef]
18. Roberts, N.; Taieb, M.; Barker, P.; Damnati, B.; Icole, M.; Williamson, D. Timing of the Younger Dryas event in East Africa from lake-level changes. *Nature* **1993**, *366*, 146–148. [CrossRef]
19. Tichy, H.; Seegers, L. The *Oreochromis alcalicus* flock (Teleostei Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: A model for the evolution of new species flocks in historical times. *Ichthyol. Explor. Freshw.* **1999**, *10*, 147–174.
20. Ford, A.G.P.; Dasmahapatra, K.K.; Rüber, L.; Gharbi, K.; Cezard, T.; Day, J.J. High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Mol. Ecol.* **2015**, *24*, 3421–3440. [CrossRef]
21. Clarisse, L.; Van Damme, M.; Gardner, W.; Coheur, P.-F.; Clerbaux, C.; Whitburn, S.; Hadji-Lazaro, J.; Hurtmans, D. Atmospheric ammonia (NH_3) emanations from Lake Natron's saline mudflats. *Sci. Rep.* **2019**, *9*, 4441. [CrossRef]
22. Randall, D.J.; Wood, C.M.; Perry, S.F.; Bergman, H.; Maloiy, G.M.O.; Mommsen, T.P.; Wright, P.A. Urea excretion as a strategy for survival in a fish living in a very alkaline environment. *Nature* **1989**, *337*, 165–166. [CrossRef]
23. Narahara, A.; Bergman, H.L.; Laurent, P.; Maina, J.N.; Walsh, P.J.; Wood, C.M. Respiratory physiology of the Lake Magadi Tilapia (*Oreochromis alcalicus grahami*), a fish adapted to a hot, alkaline, and frequently hypoxic environment. *Physiol. Zool.* **1996**, *69*, 1114–1136. [CrossRef]
24. Bergman, A.N.; Laurent, P.; Otiang'a-Owiti, G.; Bergman, H.L.; Walsh, P.J.; Wilson, P.; Wood, C.M. Physiological adaptations of the gut in the Lake Magadi tilapia, *Alcolapia grahami*, an alkaline- and saline-adapted teleost fish. *Comp. Biochem. Physiol. Part. A Mol. Integr. Physiol.* **2003**, *136*, 701–715. [CrossRef]
25. Kavembe, G.D.; Franchini, P.; Irisarri, I.; Machado-Schiaffino, G.; Meyer, A. Genomics of Adaptation to Multiple Concurrent Stresses: Insights from Comparative Transcriptomics of a Cichlid Fish from One of Earth's Most Extreme Environments, the Hypersaline Soda Lake Magadi in Kenya, East Africa. *J. Mol. Evol.* **2015**, *81*, 90–109. [CrossRef] [PubMed]
26. Wood, C.M.; Brix, K.V.; De Boeck, G.; Bergman, H.L.; Bianchini, A.; Bianchini, L.F.; Maina, J.N.; Johannsson, O.E.; Kavembe, G.D.; Papah, M.B.; et al. Mammalian metabolic rates in the hottest fish on earth. *Sci. Rep.* **2016**, *6*, 26990. [CrossRef] [PubMed]
27. White, L.J.; Sutton, G.; Shechonge, A.; Day, J.J.; Dasmahapatra, K.K.; Pownall, M.E. Adaptation of the carbamoyl-phosphate synthetase (CPS) enzyme in an extremophile fish. *R. Soc. Open Sci.* **2020**, in press.
28. Bishopric, N.H. Evolution of the Heart from Bacteria to Man. *Ann. New York Acad. Sci.* **2005**, *1047*, 13–29. [CrossRef]
29. Solc, D. The heart and heart conducting system in the kingdom of animals: A comparative approach to its evolution. *Exp. Clin. Cardiol.* **2007**, *12*, 113–118.
30. Meijler, F.L.; Meijler, T.D. Archetype, adaptation and the mammalian heart. *Neth. Hear. J.* **2011**, *19*, 142–148. [CrossRef]
31. Asnani, A.; Peterson, R.T. The zebrafish as a tool to identify novel therapies for human cardiovascular disease. *Dis. Model. Mech.* **2014**, *7*, 763–767. [CrossRef]
32. Stainier, D.Y.; Lee, R.K.; Fishman, M.C. Cardiovascular development in zebrafish. I. Myocardial fate map and heart tube formation. *Development* **1993**, *119*, 31–40.
33. Keegan, B.R.; Meyer, D.; Yelon, D. Organization of cardiac chamber progenitors in the zebrafish blastula. *Development* **2004**, *131*, 3081–3091. [CrossRef]
34. Yelon, D.; Horne, S.A.; Stainier, D.Y.R. Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish. *Dev. Biol.* **1999**, *214*, 23–37. [CrossRef] [PubMed]
35. Rohr, S.; Otten, C.; Abdelilah-Seyfried, S. Asymmetric involution of the myocardial field drives heart tube formation in zebrafish. *Circ. Res.* **2008**, *102*, 12–19. [CrossRef]

36. Chen, J.N.; van Eeden, F.J.; Warren, K.S.; Chin, A.; Nusslein-Volhard, C.; Haffter, P.; Fishman, M.C. Left-right pattern of cardiac BMP4 may drive asymmetry of the heart in zebrafish. *Development* **1997**, *124*, 4373–4382.
37. Beis, D.; Bartman, T.; Jin, S.-W.; Scott, I.C.; D'Amico, L.A.; Ober, E.A.; Verkade, H.; Frantsve, J.; Field, H.A.; Wehman, A.; et al. Genetic and cellular analyses of zebrafish atrioventricular cushion and valve development. *Development* **2005**, *132*, 4193–4204. [CrossRef]
38. Fossett, N.; Schulz, R.A. Functional conservation of hematopoietic factors in *Drosophila* and vertebrates. *Differentiation* **2001**, *69*, 83–90. [CrossRef] [PubMed]
39. Nemer, G.M.; Nemer, M. Regulation of heart development and function through combinatorial interactions of transcription factors. *Ann. Med.* **2001**, *33*, 604–610. [CrossRef] [PubMed]
40. Holt, M.; Oram, S. Familial heart disease with skeletal malformations. *Br. Heart J.* **1960**, *22*, 236–242. [CrossRef]
41. Basson, C.T.; Cowley, G.S.; Solomon, S.D.; Weissman, B.; Poznanski, A.K.; Traill, T.A.; Seidman, J.G.; Seidman, C.E. The Clinical and Genetic Spectrum of the Holt-Oram Syndrome (Heart-Hand Syndrome). *N. Engl. J. Med.* **1994**, *330*, 885–891. [CrossRef]
42. Newbury-Ecob, R.A.; Leanage, R.; Raeburn, J.A.; Young, I.D. Holt-Oram syndrome: A clinical genetic study. *J. Med. Genet.* **1996**, *33*, 300–307. [CrossRef]
43. Basson, C.T.; Bachinsky, D.R.; Lin, R.C.; Levi, T.; Elkins, J.A.; Soults, J.; Grayzel, D.; Kroumpouzou, E.; Traill, T.A.; Leblanc-Straceski, J.; et al. Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* **1997**, *15*, 30–35. [CrossRef]
44. Black, B.L.; Olson, E.N. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu. Rev. Cell Dev. Bio.* **1998**, *14*, 167–196. [CrossRef] [PubMed]
45. Lilly, B.; Zhao, B.; Ranganayakulu, G.; Paterson, B.M.; Schulz, R.A.; Olson, E.N. Requirement of MADS domain transcription factor D-MEF2 for muscle formation in *Drosophila*. *Science* **1995**, *267*, 688–693. [CrossRef] [PubMed]
46. Edmondson, D.G.; Lyons, G.E.; Martin, J.F.; Olson, E.N. Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development* **1994**, *120*, 1251–1263. [PubMed]
47. Ford, A.G.P.; Rüber, L.; Newton, J.; Dasmahapatra, K.K.; Balarin, J.D.; Bruun, K.; Day, J.J. Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution* **2016**, *70*, 2718–2735. [CrossRef] [PubMed]
48. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
49. Fujimura, K.; Okada, N. Development of the embryo, larva and early juvenile of Nile tilapia *Oreochromis niloticus* (Pisces: Cichlidae). Developmental staging system. *Dev. Growth. Differ.* **2007**, *49*, 301–324. [CrossRef]
50. De Jong, I.M.L.; Witte, F.; Richardson, M.K. Developmental stages until hatching of the Lake Victoria cichlid Haplochromis piceatus (Teleostei: Cichlidae). *J. Morphol.* **2009**, *270*, 519–535. [CrossRef]
51. Woltering, J.M.; Holzem, M.; Schneider, R.F.; Nanos, V.; Meyer, A. The skeletal ontogeny of *Astatotilapia burtoni*—a direct-developing model system for the evolution and development of the teleost body plan. *BMC Dev. Biol.* **2018**, *18*. [CrossRef]
52. Flegler-Balon, C. *Direct and Indirect Development in Fishes—Examples of Alternative Life-History Styles*; Springer Science and Business Media LLC: Dordrecht, The Netherland, 1989; Volume 6, pp. 71–100.
53. Balon, E.K. Alternative Ways to Become a Juvenile or a Definitive Phenotype (and on Some Persisting Linguistic Offenses). *Environ. Boil. Fishes* **1999**, *56*, 17–38. [CrossRef]
54. Le Pabic, P.; Stellwag, E.J.; Scemama, J.-L. Embryonic Development and Skeletogenesis of the Pharyngeal Jaw Apparatus in the Cichlid Nile Tilapia (*Oreochromis niloticus*). *Anat. Rec.* **2009**, *292*, 1780–1800. [CrossRef]
55. Holtzinger, A.; Evans, T. Gata4 regulates the formation of multiple organs. *Development* **2005**, *132*, 4005–4014. [CrossRef]
56. Albalat, R.; Baquero, M.; Minguillon, C. Identification and characterisation of the developmental expression pattern of tbx5b, a novel tbx5 gene in zebrafish. *Gene Expr. Patterns* **2010**, *10*, 24–30. [CrossRef] [PubMed]
57. Kudoh, T.; Tsang, M.; Hukriede, N.A.; Chen, X.F.; Dedekian, M.; Clarke, C.J.; Kiang, A.; Schultz, S.; Epstein, J.A.; Toyama, R.; et al. A gene expression screen in zebrafish embryogenesis. *Genome Res.* **2001**, *11*, 1979–1987. [CrossRef] [PubMed]
58. Schwartz, R.; Dayhoff, M.O. Matrices for Detecting Distant Relationships. In *Atlas of Protein Sequences*; Dayhoff, M., Ed.; National Biomedical Research Foundation: Washington, DC, USA, 1979; pp. 353–358.
59. Minko, K.; Bollerot, K.; Drevon, C.; Hallais, M.F.; Jaffredo, T. From mesoderm to blood islands: Patterns of key molecules during yolk sac erythropoiesis. *Gene Expr. Patterns* **2003**, *3*, 261–272. [CrossRef]

60. Silver, L.; Palis, J. Initiation of murine embryonic erythropoiesis: A spatial analysis. *Blood* **1997**, *89*, 1154–1164. [CrossRef] [PubMed]
61. Chapman, D.L.; Garvey, N.; Hancock, S.; Alexiou, M.; Agulnik, S.I.; Gibson-Brown, J.J.; Cebra-Thomas, J.; Bollag, R.J.; Silver, L.M.; Papaioannou, V.E. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev. Dyn.* **1996**, *206*, 379–390. [CrossRef]
62. Gibson-Brown, J.J.; Agulnik, S.I.; Silver, L.M.; Papaioannou, V.E. Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. *Mech. Dev.* **1998**, *74*, 165–169. [CrossRef]
63. Gruenauer-Kloevekorn, C.; Reichel, M.B.; Duncker, G.I.W.; Froster, U.G. Molecular Genetic and Ocular Findings in Patients with Holt-Oram Syndrome. *Ophthalmic Genet.* **2005**, *26*, 1–8. [CrossRef]
64. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **1992**, *8*, 275–282. [CrossRef]
65. Ranganayakulu, G.; Zhao, B.; Dokidis, A.; Molkentin, J.D.; Olson, E.N.; Schulz, R.A. A Series of Mutations in the D-MEF2 Transcription Factor Reveal Multiple Functions in Larval and Adult Myogenesis in *Drosophila*. *Dev. Biol.* **1995**, *171*, 169–181. [CrossRef]
66. Hinits, Y.; Pan, L.; Walker, C.; Dowd, J.; Moens, C.B.; Hughes, S.M. Zebrafish Mef2ca and Mef2cb are essential for both first and second heart field cardiomyocyte differentiation. *Dev. Biol.* **2012**, *369*, 199–210. [CrossRef] [PubMed]
67. Della Gaspera, B.; Armand, A.S.; Sequeira, I.; Lecolle, S.; Gallien, C.L.; Charbonnier, F.; Chanoine, C. The *Xenopus* MEF2 gene family: Evidence of a role for XMEF2C in larval tendon development. *Dev. Biol.* **2009**, *328*, 392–402. [CrossRef] [PubMed]
68. Brand-Saberi, B.; Christ, B. Evolution and development of distinct cell lineages derived from somites. *Curr. Top. Dev. Biol.* **2000**, *48*, 1–42. [CrossRef]
69. Reiter, J.F.; Alexander, J.; Rodaway, A.; Yelon, D.; Patient, R.; Holder, N.; Stainier, D.Y.R. Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes Dev.* **1999**, *13*, 2983–2995. [CrossRef] [PubMed]
70. Schoenebeck, J.J.; Keegan, B.R.; Yelon, D. Vessel and blood specification override cardiac potential in anterior mesoderm. *Dev. Cell* **2007**, *13*, 254–267. [CrossRef]
71. Latinkic, B.V.; Kotecha, S.; Mohun, T.J. Induction of cardiomyocytes by GATA4 in *Xenopus* ectodermal explants. *Development* **2003**, *130*, 3865–3876. [CrossRef]
72. Garg, V.; Kathiriya, I.S.; Barnes, R.; Schluterman, M.K.; King, I.N.; Butler, C.A.; Rothrock, C.R.; Eapen, R.S.; Hirayama-Yamada, K.; Joo, K.; et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **2003**, *424*, 443–447. [CrossRef]
73. Rajagopal, S.K.; Ma, Q.; Obler, D.; Shen, J.; Manichaikul, A.; Tomita-Mitchell, A.; Boardman, K.; Briggs, C.; Garg, V.; Srivastava, D.; et al. Spectrum of heart disease associated with murine and human GATA4 mutation. *J. Mol. Cell. Cardiol.* **2007**, *43*, 677–685. [CrossRef]
74. Peterkin, T.; Gibson, A.; Patient, R. Redundancy and evolution of GATA factor requirements in development of the myocardium. *Dev. Biol.* **2007**, *311*, 623–635. [CrossRef]
75. Dobrzycki, T.; Lalwani, M.; Telfer, C.; Monteiro, R.; Patient, R. The roles and controls of GATA factors in blood and cardiac development. *IUBMB Life* **2019**, *72*, 39–44. [CrossRef]
76. Ciau-Uitz, A.; Monteiro, R.; Kirmizitas, A.; Patient, R. Developmental hematopoiesis: Ontogeny, genetic programming and conservation. *Exp. Hematol.* **2014**, *42*, 669–683. [CrossRef] [PubMed]
77. Bertrand, J.Y.; Chi, N.C.; Santoso, B.; Teng, S.; Stainier, D.Y.R.; Traver, D. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **2010**, *464*, 108–111. [CrossRef] [PubMed]
78. Gore, A.V.; Pillay, L.M.; Galanternik, M.V.; Weinstein, B.M. The zebrafish: A fintastic model for hematopoietic development and disease. *Wiley Interdiscip. Rev. Dev. Biol.* **2018**, *7*, e312. [CrossRef] [PubMed]
79. Dzierzak, E. The emergence of definitive hematopoietic stem cells in the mammal. *Curr. Opin. Hematol.* **2005**, *12*, 197–202. [CrossRef]
80. Cumano, A.; Godin, I. Ontogeny of the hematopoietic system. *Annu. Rev. Immunol.* **2007**, *25*, 745–785. [CrossRef]
81. Johnson, G.R.; Moore, M.A.S. Role of stem cell migration in initiation of mouse foetal liver haemopoiesis. *Nature* **1975**, *258*, 726–728. [CrossRef]
82. Houssaint, E. Differentiation of the mouse hepatic primordium. II. Extrinsic origin of the haemopoietic cell line. *Cell Differ.* **1981**, *10*, 243–252. [CrossRef]

83. Keller, G.; Lacaud, G.; Robertson, S. Development of the hematopoietic system in the mouse. *Exp. Hematol.* **1999**, *27*, 777–787. [CrossRef]
84. Murayama, E.; Kissa, K.; Zapata, A.; Mordelet, E.; Briolat, V.; Lin, H.-F.; Handin, R.I.; Herbomel, P. Tracing hematopoietic precursor migration to successive hematopoietic organs during zebrafish development. *Immunity* **2006**, *25*, 963–975. [CrossRef]
85. Wolf, A.; Aggio, J.; Campbell, C.; Wright, F.; Marquez, G.; Traver, D.; Stachura, D.L. Zebrafish Caudal Haematopoietic Embryonic Stromal Tissue (CHEST) Cells Support Haematopoiesis. *Sci. Rep.* **2017**, *7*, 44644. [CrossRef]
86. Ueno, H.; Weissman, I.L. The origin and fate of yolk sac hematopoiesis: Application of chimera analyses to developmental studies. *Int. J. Dev. Biol.* **2010**, *54*, 1019–1031. [CrossRef] [PubMed]
87. Ferkowicz, M.J.; Yoder, M.C. Blood island formation: Longstanding observations and modern interpretations. *Exp. Hematol.* **2005**, *33*, 1041–1047. [CrossRef]
88. Detrich, H.W.; Kieran, M.W.; Chan, F.Y.; Barone, L.M.; Yee, K.; Rundstadler, J.A.; Pratt, S.; Ransom, D.; Zon, L.I. Intraembryonic hematopoietic cell migration during vertebrate development. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10713–10717. [CrossRef] [PubMed]
89. Aronson, B.E.; Stapleton, K.A.; Krasinski, S.D. Role of GATA factors in development, differentiation, and homeostasis of the small intestinal epithelium. *Am. J. Physiol. Liver Physiol. Liver Physiol.* **2014**, *306*, G474–G490. [CrossRef] [PubMed]
90. Pevny, L.; Simon, M.C.; Robertson, E.; Klein, W.H.; Tsai, S.-F.; D'Agati, V.; Orkin, S.H.; Costantini, F. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **1991**, *349*, 257–260. [CrossRef]
91. Tsai, F.-Y.; Keller, G.; Kuo, F.C.; Weiss, M.; Chen, J.; Rosenblatt, M.; Alt, F.W.; Orkin, S.H. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **1994**, *371*, 221–226. [CrossRef]
92. Weiss, M.J.; Keller, G.; Orkin, S.H. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **1994**, *8*, 1184–1197. [CrossRef]
93. Bruneau, B.G.; Nemer, G.; Schmitt, J.P.; Charron, F.; Robitaille, L.; Caron, S.; Conner, D.A.; Gessler, M.; Nemer, M.; Seidman, C.E.; et al. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **2001**, *106*, 709–721. [CrossRef]
94. Garrity, D.M.; Childs, S.; Fishman, M.C. The heartstrings mutation in zebrafish causes heart/fin Tbx5 deficiency syndrome. *Development* **2002**, *129*, 4635–4645.
95. Mommersteeg, M.T.M.; Soufan, A.T.; de Lange, F.J.; van den Hoff, M.J.B.; Anderson, R.H.; Christoffels, V.M.; Moorman, A.F.M. Two distinct pools of mesenchyme contribute to the development of the atrial septum. *Circ. Res.* **2006**, *99*, 351–353. [CrossRef]
96. Xie, L.; Hoffmann, A.D.; Burnicka-Turek, O.; Friedland-Little, J.M.; Zhang, K.; Moskowitz, I.P. Tbx5-Hedgehog Molecular Networks Are Essential in the Second Heart Field for Atrial Septation. *Dev. Cell* **2012**, *23*, 280–291. [CrossRef] [PubMed]
97. Zhu, Y.; Gramolini, A.O.; Walsh, M.A.; Zhou, Y.-Q.; Slorach, C.; Friedberg, M.K.; Takeuchi, J.K.; Sun, H.; Henkelman, R.M.; Backx, P.H.; et al. Tbx5-dependent pathway regulating diastolic function in congenital heart disease. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5519–5524. [CrossRef] [PubMed]
98. Pi-Roig, A.; Martin-Blanco, E.; Minguillon, C. Distinct tissue-specific requirements for the zebrafish tbx5 genes during heart, retina and pectoral fin development. *Open Biol.* **2014**, *4*, 140014. [CrossRef] [PubMed]
99. Koshiba-Takeuchi, K.; Takeuchi, J.K.; Matsumoto, K.; Momose, T.; Uno, K.; Hoepker, V.; Ogura, K.; Takahashi, N.; Nakamura, H.; Yasuda, K.; et al. Tbx5 and the retinotectum projection. *Science* **2000**, *287*, 134–137. [CrossRef]
100. Anderson, E.A.T.B.; Ho, R.K. A transcriptomics analysis of the Tbx5 paralogues in zebrafish. *PLoS ONE* **2018**, *13*, e0208766. [CrossRef]
101. Tamura, K.; Yonei-Tamura, S.; Belmonte, J.C.I. Differential expression of Tbx4 and Tbx5 in Zebrafish fin buds. *Mech. Dev.* **1999**, *87*, 181–184. [CrossRef]
102. Begemann, G.; Ingham, P.W. Developmental regulation of Tbx5 in zebrafish embryogenesis. *Mech. Dev.* **2000**, *90*, 299–304. [CrossRef]
103. Agarwal, P.; Wylie, J.N.; Galceran, J.; Arkhitko, O.; Li, C.; Deng, C.; Grosschedl, R.; Bruneau, B.G. Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. *Development* **2003**, *130*, 623–633. [CrossRef]

104. Mercader, N. Early steps of paired fin development in zebrafish compared with tetrapod limb development. *Dev. Growth Differ.* **2007**, *49*, 421–437. [CrossRef]
105. Adachi, N.; Robinson, M.; Goolsbee, A.; Shubin, N.H. Regulatory evolution of Tbx5 and the origin of paired appendages. *Proc.S Natl. Acad. Sci. USA* **2016**, *113*, 10115–10120. [CrossRef]
106. Cunningham, T.J.; Lancman, J.J.; Berenguer, M.; Dong, P.D.S.; Duester, G. Genomic Knockout of Two Presumed Forelimb Tbx5 Enhancers Reveals They Are Nonessential for Limb Development. *Cell Rep.* **2018**, *23*, 3146–3151. [CrossRef] [PubMed]
107. Lin, Q.; Schwarz, J.; Bucana, C.; Olson, E.N. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **1997**, *276*, 1404–1407. [CrossRef] [PubMed]
108. Verzi, M.P.; McCulley, D.J.; De Val, S.; Dodou, E.; Black, B.L. The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev. Biol.* **2005**, *287*, 134–145. [CrossRef] [PubMed]
109. Vong, L.; Bi, W.; O'Connor-Halligan, K.E.; Li, C.; Cserjesi, P.; Schwarz, J.J. MEF2C is required for the normal allocation of cells between the ventricular and sinoatrial precursors of the primary heart field. *Dev. Dyn.* **2006**, *235*, 1809–1821. [CrossRef]
110. Materna, S.C.; Sinha, T.; Barnes, R.M.; van Bueren, K.L.; Black, B.L. Cardiovascular development and survival require Mef2c function in the myocardial but not the endothelial lineage. *Dev. Biol.* **2019**, *445*, 170–177. [CrossRef]
111. Hinits, Y.; Hughes, S.M. Mef2s are required for thick filament formation in nascent muscle fibres. *Development* **2007**, *134*, 2511–2519. [CrossRef]
112. Lazic, S.; Scott, I.C. Mef2cb regulates late myocardial cell addition from a second heart field-like population of progenitors in zebrafish. *Dev. Biol.* **2011**, *354*, 123–133. [CrossRef]
113. Ticho, B.S.; Stainier, D.Y.R.; Fishman, M.C.; Breitbart, R.E. Three zebrafish MEF2 genes delineate somitic and cardiac muscle development in wild-type and mutant embryos. *Mech. Dev.* **1996**, *59*, 205–218. [CrossRef]
114. Torgersen, J.S.; Takle, H.; Andersen, Ø. Differential spatial expression of mef2 paralogs during cardiac development in Atlantic cod (*Gadus morhua*). *Comp. Biochem. Physiol. Part. B Biochem. Mol. Biol.* **2011**, *158*, 181–187. [CrossRef]
115. Breitbart, R.E.; Liang, C.-s.; Smoot, L.B.; Laheru, D.A.; Mahdavi, V.; Nadal-Ginard, B. A fourth human MEF2 transcription factor, hMEF2D, is an early marker of the myogenic lineage. *Development* **1993**, *118*, 1095–1106.
116. Ganassi, M.; Badodi, S.; Polacchini, A.; Baruffaldi, F.; Hughes, S.M.; Hinits, Y.; Molinari, S. Distinct functions of alternatively spliced isoforms encoded by zebrafish mef2ca and mef2cb. *Biochim. Et Biophy. Acta (BBA)* **2014**, *1839*, 559–570. [CrossRef]
117. Pownall, M.E.; Gustafsson, M.K.; Emerson, C.P. Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **2002**, *18*, 747–783. [CrossRef] [PubMed]
118. Hinits, Y.; Osborn, D.P.S.; Hughes, S.M. Differential requirements for myogenic regulatory factors distinguish medial and lateral somitic, cranial and fin muscle fibre populations. *Development* **2009**, *136*, 403–414. [CrossRef] [PubMed]
119. Cole, N.J.; Currie, P.D. Insights from sharks: Evolutionary and developmental models of fin development. *Dev. Dyn.* **2007**, *236*, 2421–2431. [CrossRef] [PubMed]
120. Gross, M.K.; Moran-Rivard, L.; Velasquez, T.; Nakatsu, M.N.; Jagla, K.; Goulding, M. Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development* **2000**, *127*, 413–424.
121. Neyt, C.; Jagla, K.; Thisse, C.; Thisse, B.; Haines, L.; Currie, P.D. Evolutionary origins of vertebrate appendicular muscle. *Nature* **2000**, *408*, 82–86. [CrossRef]
122. Cinnamon, Y.; Kahane, N.; Kalcheim, C. Characterization of the early development of specific hypaxial muscles from the ventrolateral myotome. *Development* **1999**, *126*, 4305–4315.
123. Okamoto, E.; Kusakabe, R.; Kuraku, S.; Hyodo, S.; Robert-Moreno, A.; Onimaru, K.; Sharpe, J.; Kuratani, S.; Tanaka, M. Migratory appendicular muscles precursor cells in the common ancestor to all vertebrates. *Nat. Ecol. Evol.* **2017**, *1*, 1731–1736. [CrossRef]