DOCTOR OF PHILOSOPHY

# Genomic analysis of shifts in life history traits in the Trinidadian guppy, Poecilia reticulata

Perry, Hazel

*Award date:*
2015

*Awarding institution:*
Bangor University

PRIFYSGOL
BANGOR
UNIVERSITY

# Genomic analysis of shifts in life history traits in the Trinidadian guppy, *Poecilia reticulata*

A thesis submitted to Bangor University for the degree of Doctor of Philosophy

By Hazel Perry, B.Sc., MRes
December 2015
Molecular Ecology and Fisheries Genetics Laboratory
School of Biological Sciences
Environment Centre Wales
Bangor University
Bangor, Gwynedd,
LL57 2UW

# Declaration and Consent

**Details of the Work**

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

**Author Name:** Hazel Perry

**Title:** Genomic analysis of shifts in life history traits in the Trinidadian guppy, *Poecilia reticulata*

**Supervisor/Department:** Professor Gary Carvalho, Dr Martin Taylor, Dr Simon Creer, Dr Cock van Oosterhout, Dr Katherine Steele. School of Biological Sciences

**Funding body (if any):** N/A

**Qualification/Degree obtained:** Doctor of Philosophy

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as "the Work". It is identical in content to that deposited in the Library, subject to point 4 below.

**Non-exclusive Rights**

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

**Bangor University Digital Repository**

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library's EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (http://www.llgc.org.uk/index.php?id=4676). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

**Statement 1:**

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed …….. (candidate) Date……..

**Statement 2:**

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

All other sources are acknowledged by footnotes and/or a bibliography.

Signed …….. (candidate) Date……..

**Statement 3:**

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic storage (subject to any constraints as defined in statement 4), and for the title and summary to be made available to outside organisations.

Signed …….. (candidate) Date……..

**Statement 4:**

Choose **one** of the following options

| | |
|---|---|
| a) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material. | Yes |
| b) I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University when the approved **bar on access** has been lifted. | |
| c) I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack | |

| | |
|---|---|
| of permissions for use of third party material. | |

*Options B should only be used if a bar on access has been approved by the University.*

**In addition to the above I also agree to the following:**

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.
3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature:…… …… Date……..……..

To David Perry, wherever you are, I hope I have made you proud.
To Sue Perry, for everything that has been and everything yet to come.
To Adam Pearce, for the life we are building.

# Abstract

It is widely accepted that size selective mortality induced by commercial fishing can and does cause changes in life history traits that include shifts in maturation age, growth rate and body size. However, whether these changes are the result of fisheries induced evolution (genetic change) or phenotypic plasticity is still unclear. Moreover, where evolution is rapid, epigenetic or regulatory change has also been found to drive major shifts in life history change.

To examine the genetic and phenotypic response to size selective harvesting, a previous study (van Wijk 2011a) subjected guppies to divergent size-specific selection. Following selection, a significant difference in both body size and age at maturation was identified as well as signatures of selection at five candidate loci. The project described here utilised these selection lines to examine the genome wide factors contributing to such life history shifts.

To assess the genome wide response to size selective harvesting, RAD sequencing was employed to identify and type large numbers of SNPs in individuals from the selection lines, as well as individuals from the generation prior to selection. Significant and consistent signs of selection were identified at 37 SNPs, the majority of which were located on the sex chromosome. The results showed that, in addition to previously observed genetic change, additional regions of the guppy genome responded to, and were associated with, observed phenotypic shifts.

Variation in the level of predation in wild populations creates variation in life history traits similar to those seen after size selective harvesting. We therefore examined the 37 SNPs identified by the RAD sequencing of the selected lines in 18 populations of wild guppy. No consistent signs of selection were identified in these wild populations, suggesting that the genetic architecture underpinning variation in life history traits in the guppy varies in different populations.

To determine the role of epigenetic change the focus has been on DNA methylation. In order to assess the levels of DNA methylation a technique known as methylation sensitive AFLP has been used. Using this technique, comparisons of the level of DNA methylation between both the selection lines and the before and after selection fish were made as well as comparisons in the levels of DNA methylation between a range of tissue types from the guppy. Results showed that patterns of DNA methylation differ significantly between different tissues in the guppy. Genome wide patterns of genome wide methylation did not differ significantly between the selection lines, however locus-specific variation in DNA methylation was identified.

## Acknowledgements

First and foremost my immense gratitude must go to my supervisors: Gary, Martin, Cock, Si and Katherine. I was often asked whether five supervisors made the project easier or harder, I can honestly say that the opportunity to work with so many great minds has made a huge difference to both my enjoyment and to the work were able to do. That both Martin and Gary have an uncanny ability to make me feel better about both myself and the project is something that kept me and the project going. Si and Katherine have both been able to offer advice on their specialist areas. Cock has so much enthusiasm for science, particularly guppies, which can be infectious.

I must also thank Christine Dreyer, Margarete Hoffman and Axel Künster at the Max Plank institute in Germany for creating and sequencing the RAD libraries which formed the backbone of chapter 2. Additionally my thanks go to Axel for providing me with access to the guppy reference genome and for the many emails answering my queries on it. I am extremely grateful to Jo Cable, Ryan Mohammed and Cock for allowing me to use the samples they collected from the wild populations in Trinidad. From the other side of the world I am grateful to Carlos Rodriguez Lopez for the countless emails of advice he gave me regarding the MSAP technique. For the hours spent teaching me how to look after fish and the help treating them when things went wrong, my thanks go to Gavin and Alix. I am also grateful to Bangor University, for providing me with the funding required to undertake this work.

To Serinde van Wijk, on whose excellent work this project is based. Without her years of hard work my project would not have existed, and without her diligence and care it would have been a lot harder. I am only sorry I couldn't see it further than I did.

The Molecular Ecology and Fisheries and Genetics in Bangor group is a fantastic place to study and work. I thoroughly enjoyed the nearly five and half years I was able to spend there. The atmosphere of the group and the office is made by all of the people there, from those early on to those still working hard today however there are a few I must specifically thank. To Wendy, for the invaluable advice from the simplest and daftest of questions, to the complexities of failed results and everything in between. And to Sarah and Iliana, for the many rants about PhD life and listening when I simply needed to talk a problem out!

Finally, but by no means last thanks to Adam Pearce and Sue Perry. To Sue for the many hours of phone calls about nothing and everything. However most importantly thank you for the many times you pointed out that I couldn't write that like that, and trying to understand enough to write it better! Enjoy your retirement! To Adam, for everything. For preventing me from starving and keeping on top of everything the PhD prevented me from doing.

# Table of contents

# List of Figures

# List of Tables

# Chapter 1: General Introduction

## 1.1 Abstract

Anthropogenic pressures, particularly harvesting, have been found to cause phenotypic trait change in a large number of wild populations. One such pressure which has been widely studied is the mortality imposed by fisheries. It is widely accepted that size selective mortality, induced by commercial fishing, causes changes in life history traits that include shifts in maturation age, growth rate and body size. However, whether these changes are the result of fisheries induced evolution (genetic change) or phenotypic plasticity has been debated. Although there have been several attempts to examine the contribution of genetic change and phenotypic plasticity using modeling, these attempts have been criticised. Recently, a study used the guppy (*Poecilia reticulata*) as a model species to undertake a selection experiment and provide evidence of genetic change associated with harvesting (van Wijk 2011b). Guppies are a model species because they are amenable to experimental manipulation and have been shown to respond to predation by evolving differing life history traits. Furthermore, there exists an extensive genomic toolbox for the species, making them ideal for studying the genomic changes underpinning observed shifts in life history traits. The following chapter discusses fisheries induced evolution and the van Wijk (2011) selection experiment in more detail. Although it is important to examine the genetic basis for shifts in life history traits, it is also necessary to consider whether epigenetic changes have a role to play. DNA methylation is one of the most widely studied epigenetic modifications and its contribution to complex traits is also discussed in the following chapter.

## 1.2 Contemporary evolution

The theory of evolution underpins biology. An understanding of how species evolve and adapt to changing environments, especially those that are rapid, has become increasingly important due to growing anthropogenic pressures. Evolution has historically been considered a very slow process (Darwin 1859, Mayr 1963), however it is now widely recognised that adaptive evolution can take place over contemporary time periods (within tens of generations: Reznick 1987, Khater *et al.* 2014). Although acceptance that anthropogenic impacts may constitute stronger and more direct selection pressures than natural selection is relatively recent (Palumbi 2001, Stockwell *et al.* 2003, Palkovacs *et al.* 2012), some early examples of contemporary evolution provide classic instances of human-induced change (Kettlewell 1958). The escalating impact of anthropogenic pressures such as global warming, habitat destruction, introductions and overexploitation require a better understanding of contemporary evolution in order to enable effective detection, prediction and mitigation of impacts (Carroll *et al.* 2007, Salamin *et al.* 2010, Skelly 2010).

One of the earliest documented examples of both contemporary evolution and human-induced trait change is that of the peppered moth *Biston betularia*, the pale morph of which increased from making up only 0.1% of the population to 98% in less than 50 years as a result of increased air

pollution from industrial revolution (Kettlewell 1958). Although the accuracy of these results has since been questioned (Rudge 1999), there is now an abundance of anthropogenic impacts, such as habitat disturbance (Desrochers 2010, Franssen 2011), pest control (McKenzie and Batterham 1994, Raymond *et al.* 2001), overexploitation (Jachmann *et al.* 1995, Cooke *et al.* 2007, Shackell *et al.* 2009), climate change (Stiling *et al.* 2003) and introductions (Peckarsky and McIntosh 1998, Cousyn *et al.* 2001), which have been shown to induce adaptive changes. While the majority of anthropogenic pressures driving human-induced trait evolution are considered to be negative and often result in maladaptive changes, there are also documented conservation actions that have led to trait changes (Rolshausen *et al.* 2009). Impacts of human-induced trait change are not only limited to the species that the anthropogenic pressure immediately impacts, but can also can lead to trait changes in species with which they interact (Blackstone and Joslyn 1984), as well as other ecological effects (Palkovacs *et al.* 2012).

The main focus of the investigation into human-induced trait change has been the effects of harvesting, with studies having shown that harvesting has the ability to induce trait change 300x faster than natural drivers, and 50x faster than other anthropogenic drivers (Darimont *et al.* 2009). Although humans have always hunted animals, whether for food, clothes or tools, the development of technologies and the growth in industrial-scale mechanised fisheries has resulted in anthropogenic selection pressures that far outstrip most natural selection pressure. However, it was not until the late 1970's (Handford *et al.* 1977) and early 1980's (Ricker 1981) that the effects harvesting can have on phenotypic traits were widely recognised. Today there is a large body of evidence showing changes in phenotypic traits as a direct result of overexploitation (*Ovis canadensis*: Coltman *et al.* 2003; Festa-Bianchet *et al.* 2014; *Cervus elaphus*: Thelen 1991; Rivrud *et al.* 2013; *Vulpes vulpes*: Haldane 1942; *Loxodonta africana*: Jachmann *et al.* 1995; Nuzzo & Traill 2014; Chiyo *et al.* 2015; *Aepyceros melampus*: Muposhi *et al.* 2015; *Ovis gmelini musimon*: Garel *et al.* 2007). Despite its importance there are still many questions surrounding the mechanisms underlying such changes, and few studies have been able to successfully identify the genetic basis underpinning the trait in question (Hendry *et al.* 2008).

## 1.3 Fisheries induced evolution

Although fishing alone would be enough to drive selection for smaller size and age at maturity (Heino *et al.* 2015) the non-random nature of size selective harvesting significantly amplifies such effects. Due to the higher value of larger fish and regulations which prevent the harvesting of fish below a certain size (e.g. Council Regulation [EC] 850/98), the selection imposed on fish populations by fishing is frequently size-specific. Similar selection pressures driven both by size-selective harvesting and natural predation can be seen in other species, particularly in where a predator has been introduced (Strauss *et al.* 2006). However, it has been shown that the number of aquatic species known to have been subjected to size-selective harvesting was almost four and half times larger than among terrestrial species (Fenberg and Roy 2008). For some species of fish

the mortality rate imposed by fisheries can be as high as 400 times their natural mortality (Mertz and Myers 1998), and it is for these reasons that most work into the effects of size-selective harvesting has focused on the impact of fisheries.

The trait or traits human-induced change affect is determined by the phenotype primarily harvested. For example trophy hunting in elephants (*Loxodonta africana*) involves removal of those individuals with the largest tusks, thereby driving selection to favour individuals with smaller tusks (Coltman, O'Donoghue, and Jorgenson 2003). As discussed above, fisheries remove the largest individuals, and life history theory predicts that increased mortality of larger individuals will drive selection to favour individuals which reach sexual maturity earlier and at a smaller size (Roff 1992, Heino and Godø 2002, Olsen *et al.* 2004, Jorgensen *et al.* 2007, Marshall and Browman 2007, Heino and Dieckmann 2008, Conover and Baumann 2009b, Kuparinen *et al.* 2009).

The two main traits which form the focus of fisheries induced evolution are body size and age at maturation. Many commercially important fish species, including cod (Heino *et al.* 2002, Olsen *et al.* 2004, Baulier *et al.* 2006), European and American plaice (Barot *et al.* 2005, Rijnsdorp *et al.* 2005, Grift *et al.* 2007), sole (Mollet *et al.* 2007), haddock (Wright, Gibb, *et al.* 2011), herring (Engelhard and Heino 2004, Enberg and Heino 2007), grayling (Haugen and Vøllestad 2001) and salmon (Ricker 1981, Hard, Gross, Heino, *et al.* 2008), have been shown to reach a smaller size and/or mature at a younger age. One of the difficulties in fully understanding the mechanisms underpinning these changes is the definition of the traits in question. The specific measurement used to determine body size often depends on the species being studied. One of the more commonly used measurements is the standard length however mass has also been used (Rijnsdorp *et al.* 2005). Age at maturation has not often been used in direct measurements of wild fish due to the difficulties in measuring it. It is, however, one of the most commonly studied traits when modelling approaches are used. In these studies age at maturation is calculated as the age at which a fish has a 50% chance of maturing (Marshall *et al.* 2009). It has been argued that the focus of FIE should be growth rate and that using body size and age at maturation as a proxy for growth rate does not fully consider the wide range of mechanisms which can affect growth (Enberg *et al.* 2012). However, until it is possible to directly measure the specific growth rate of wild fish, body size and age at maturation will continue to be used.

Although it is now widely accepted that changes in life history traits do occur as a result of fishing, disentangling how much of the change in life history traits is a result of environmental effects and how much is a result of evolutionary change has proved difficult and controversial (Kuparinen and Merilä 2007, Marshall and Browman 2007, Browman, Law, and Marshall 2008, Enberg *et al.* 2012). Several studies have suggested that observed changes in phenotypic traits for some species are a result of environmental changes and phenotypic plasticity (Morita and Fukuwaka 2006, Kraak 2007, Marshall and Browman 2007, Salmon *et al.* 2008, Daufresne *et al.* 2009). Many fish populations are now at historically low population sizes and densities that reduce intraspecific competition and

potentially alter growth rates (Trippel 1995, Stokes and Law 2000, Sinclair 2002, Roos 2006, Kraak 2007, Thorpe 2007). Moreover, change in water temperature also has the potential to alter growth rates (Stokes and Law 2000, Sinclair *et al.* 2001, Law 2007, Thresher *et al.* 2007, Daufresne *et al.* 2009) and phenotype (Seymour 1959, Hempel and Blaxter 1961). When combined with density effects, these environmental factors have the potential to induce the observed changes in life history traits (early maturation and smaller size: Hutchings 2000; Wootton 1999; Roff 2003).

If evolutionary change as a direct response to fishing pressure is responsible for the observed changes in life history traits, there are four basic conditions which must be met: the trait must (i) show phenotypic variation (Conover and Baumann 2009b, Enberg *et al.* 2012); (ii) some of the phenotypic variation must be heritable (Gjedrem 1983, Kuparinen and Merilä 2007, Carlson and Seamons 2008, Hutchings and Fraser 2008, Conover and Baumann 2009b); (iii) the optimal phenotype under fisheries selection must be different to the optimal phenotype under natural selection (Hendry *et al.* 2011), and finally (iv) the selective force, i.e. fishing, must be sufficiently strong to outweigh natural selection (Edeline *et al.* 2007). All changes in life history traits observed in harvested fish stocks show significant phenotypic variation which has also been found to be heritable (Stokes and Law 2000, Heino and Godø 2002, Law 2007). For many of these traits, the direction of selection imposed by fishing is opposite that of natural selection (Conover 2007, Edeline *et al.* 2007), and the strength of selection imposed is significantly higher than that of natural selection (Mertz and Myers 1998, Edeline *et al.* 2007). However, although there is evidence that fishing fulfils all of the conditions required to drive evolutionary change, irrefutable evidence that genetic change occurs has been challenging to obtain (Heino *et al.* 2015).

There have been several attempts made to disentangle the effects of environmental and genetic change, most of which can be broadly categorised as either indirect approaches, such as modelling and analysis of temporal data sets for exploited fish stocks or, direct approaches, such as selection experiments and use of molecular techniques (Conover and Baumann 2009b). By far the strongest support for fisheries-induced evolution is currently provided by indirect approaches, specifically probabilistic maturation reaction norms (PMRN) (Stearns and Koella 1986, Heino *et al.* 2002, Olsen *et al.* 2004, Dieckmann and Heino 2007, Swain *et al.* 2007, Hard, Gross, and Heino 2008, Heino and Dieckmann 2008, Hutchings and Fraser 2008, Dunlop *et al.* 2009, Sharpe and Hendry 2009). By determining and plotting the probability of an immature fish reaching maturity as a function of age and size, the PMRN aims to control for environmental plasticity (Heino *et al.* 2002). However the reliance the PMRN approach has on maturation being only a function of age and size has led to criticism of the method (Kinnison and Hendry 2001, Kraak 2007, Kuparinen and Merilä 2007, Marshall and McAdam 2007, Morita *et al.* 2009, Uusi-Heikkilä *et al.* 2011), with several studies providing evidence that traits other than age and size can significantly affect the probability of reaching maturity (Morita and Fukuwaka 2006, Grift *et al.* 2007, Kraak 2007, Kuparinen and Merilä 2007, Morita *et al.* 2009). In an attempt to overcome such problems, multidimensional PMRN's have been developed for many species (Baulier *et al.* 2006, Grift *et al.* 2007, Kraak 2007, Mollet *et al.*

2007) and although it is theoretically possible to include all factors influencing body size when producing PMRN's, obtaining relevant data for all possible covariates is extremely difficult. Consequently, most multidimensional PMRN's still only contain a maximum of three dimensions (Vainikka *et al.* 2008, Marshall *et al.* 2009, Pauli and Heino 2013).

In an attempt to overcome the difficulties of fully modelling such complex traits, selection experiments have also been undertaken to investigate FIE. Although an experimental approach was used as early as 1975 (Silliman 1975), the use of selection experiments to distinguish between environmental and genetic effects of harvesting and has only recently begun receiving significant attention (Fuller *et al.* 2005, Conover and Baumann 2009b). One of the most recent experiments into the effect of harvesting on life history traits was undertaken by Conover & Munch (2002), which supported the theory of fisheries-induced evolution. Despite providing insights, selection experiments have been criticised for their simplification of natural environments that has the potential to eradicate or minimise gene X environment interactions (Holloway *et al.* 1990, Hoffmann and Merilä 1999). Studies such as that by Conover & Munch (2002) have been further been criticised for the high level of harvesting imposed, which it has been argued, are significantly higher than most wild harvested populations experience (Hilborn 2006, 2007). Others have argued that such high experimental harvesting rates enable the speed at which traits will change to be calculated for species being harvested at lower rates (Brown *et al.* 2008). Even though selection experiments have provided support for fisheries-induced evolution, direct empirical evidence demonstrating genetic change underpinning the phenotypic shifts was needed before a definitive conclusion could be reached. However, while efforts have being made to apply molecular techniques to wild populations (Nielsen *et al.* 2009, Jakobsdóttir *et al.* 2011), unequivocal genetic evidence is rare.

### 1.3.1 Combining a selection experiment and with a molecular approach

In 2008 a study aiming to provide the desired empirical genetic evidence was initiated (van Wijk 2011a). This study combined the use of a selection experiment on Trinidadian guppies (*Poecilia reticulata*) with molecular genetic techniques in order to examine directly the extent of genetic change associated with controlled contrasting harvesting regimes. The first three generations were left to breed freely with generations being kept separate (by removing all juveniles). From the third generation (F3) 550 males were selected to make the F4 generation. Of those 550, 50 were randomly selected to create the control line and the smallest 20% used to create the small line and the largest 20% used to create the large line. For each line two replicates were created (e.g. small line 1; small line 2; large line 1 and large line 2). In order to create each of the following three generations once 100-150 (male) fish were mature the 50 smallest and largest fish were selected for the next generation, i.e. when 100-150 of the male offspring from the F3 generation in the L1 line (large 1) were mature they were measured and the 50 largest fish became the F4 generation for that line. Once the F6 generation had been created, 50 males were randomly collected and measured for size and age and size at maturation. Throughout generations F3-F6 females were

not subjected to any selection and the females selected to create the next generation were selected randomly. This was primarily due to the fact that females have indeterminate growth whilst males stop growing once they reach maturity, meaning that any size selection of females would have been selecting for age rather than size. (See Figure 1.1 for a schematic representation)

During the course of the study body size, neutral microsatellites and seventeen candidate loci were monitored at every generation. Additionally, measures of maturation size and age were monitored in the generations before and after selection. Findings showed a significant difference in both body size and age and size at maturation following selection (body size showed a change of ±7%, whilst age and size at maturation changed by ±4-6% and ±8-12% respectively). Importantly, this study also showed significant signs of selection at five of the seventeen candidate loci being monitored thereby providing empirical genetic evidence of changes in life history traits as a direct result of harvesting.

By combining a selection experiment with molecular genetic techniques, van Wijk (2011) provided empirical evidence of genetic changes in life history traits as a direct result of harvesting. Now that evidence of genetic change as a direct result of size-selective harvesting is available, it is important to further consider the broader evolutionary processes involved, and the level of genomic change taking place.

## 1.4 Importance of understanding fisheries-induced evolution

In terms of fisheries and their management, the importance of understanding the cause of the observed changes in life history traits lies in their potential for reversibility (Conover and Baumann 2009b, Enberg *et al.* 2009). If the process is a result of environmental changes and phenotypic plasticity, then it is to be expected that the observed changes will be readily reversible, if and when fishing ceases (Dieckmann and Heino 2007, Jorgensen *et al.* 2007, Kuparinen and Merilä 2007). As fishing pressure is removed, environmental conditions such as population density would be expected to return to pre-exploitation levels, indicated by phenotypes reverting to their pre-exploitation optima. If however, the changes in life history traits are a result of genetic change, it is likely that they would require much longer to reverse, if at all (Dieckmann and Heino 2007, Enberg *et al.* 2009). By removing most of the fish whose genes predispose them to grow larger or for later maturation, the alleles that code for these traits would also be reduced in frequency, and may ultimately be lost from populations completely. If such alleles are reduced in frequency, or removed completely, it will take a long time for an exploited population to recover, even when the fishing pressure is reduced or

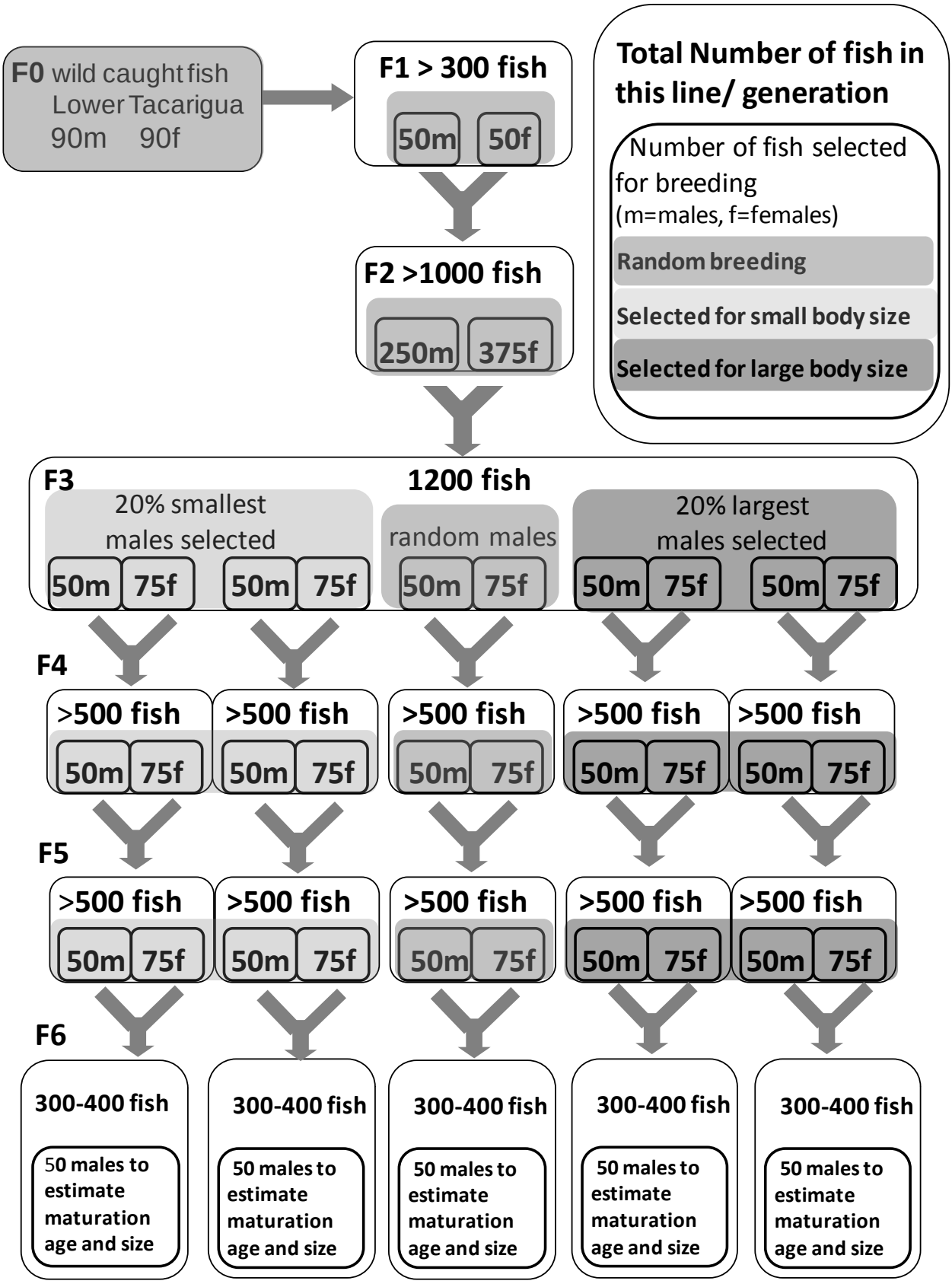



**Figure 1.1:** Schematic representation of selection experiments, showing the number of fish reared in the different generations $F_0$-$F_6$. Numbers of fish selected and used for breeding the next generation are indicated in grey, the different shades indicating the different treatments: light grey for small-selected lines, dark grey for large-selected lines and intermediate shading for random breeding generations. Figure from (van Wijk 2011a).

removed (Roos 2006, Kuparinen and Merilä 2007). Genetic variation underpinning large body size for example, would either have to evolve again through mutation, or via immigration from other populations. Studies have considered the reversibility of fisheries-induced evolution and found that while some traits recovered, others did not (Salinas *et al.* 2012). More information is, however, needed before the reversibility of fisheries-induced evolution can be fully assessed (Kuparinen and Hutchings 2012).

In addition to the importance of understanding whether the causes are primarily genetic or environmental, the genomic nature of genetic changes will also affect the potential rate of recovery (Conover *et al.* 2009). Questions including whether changes are occurring at few small-effect or several large-effect genes (Roesti *et al.* 2012), and whether selection acts on new or standing genetic variation (Akey 2012), will be important in determining rate of recovery. For example if the observed shifts in life history traits are underlain by changes at few large-effect loci, it could be argued that following a moratorium on fishing, a small number of loci involved would facilitate recovery, compared to a scenario where changes had occurred at a large number of small- effect loci.

Size-selective harvesting not only has direct impacts on the species being harvested, but also on wider ecosystem processes, such as primary productivity, decomposition rates and nutrient cycling (Bassar *et al.* 2010). Therefore further knowledge of the rate of recovery will also provide an insight into the likely persistence of such ecosystem-wide effects in exploited communities, as well as where predation is strongly size-selective (Walsh and Reznick 2011, Furness *et al.* 2012, Furness and Reznick 2014, Travis, Reznick, and Bassar 2014, El-Sabaawi, Bassar, *et al.* 2015, El-Sabaawi, Marshall, *et al.* 2015).

In order to answer questions about the genomic architecture underpinning the observed shifts in life history traits it is necessary to examine genome-wide changes following size-selective harvesting. Several studies have examined adaptation of guppies to size-selective predation (see Magurran 2005 and references therin), as well as ongoing work examining the ecosystem wide impacts of changes in life history traits (Travis, Reznick, Bassar, *et al.* 2014, El-Sabaawi, Bassar, *et al.* 2015, El-Sabaawi, Marshall, *et al.* 2015). Collectively, such studies in addition to the existence of size selected lines derived from van Wijk (2011) and van Wijk *et al.* (2013), provide a unique opportunity to examine genome-wide impacts of size-selective harvesting and attempts to address some of the questions outlined above.

## 1.5 The guppy, *Poecilia reticulata*

Despite being native to Venezuela, Guyana, Suriname, Trinidad and Tabago (Farr 1975), the small viviparous guppy, *Poecilia reticulata*, has been introduced to every continent apart from Antarctica, both as a method of controlling mosquitoes (Courtenay *et al.* 1989), and through its huge popularity

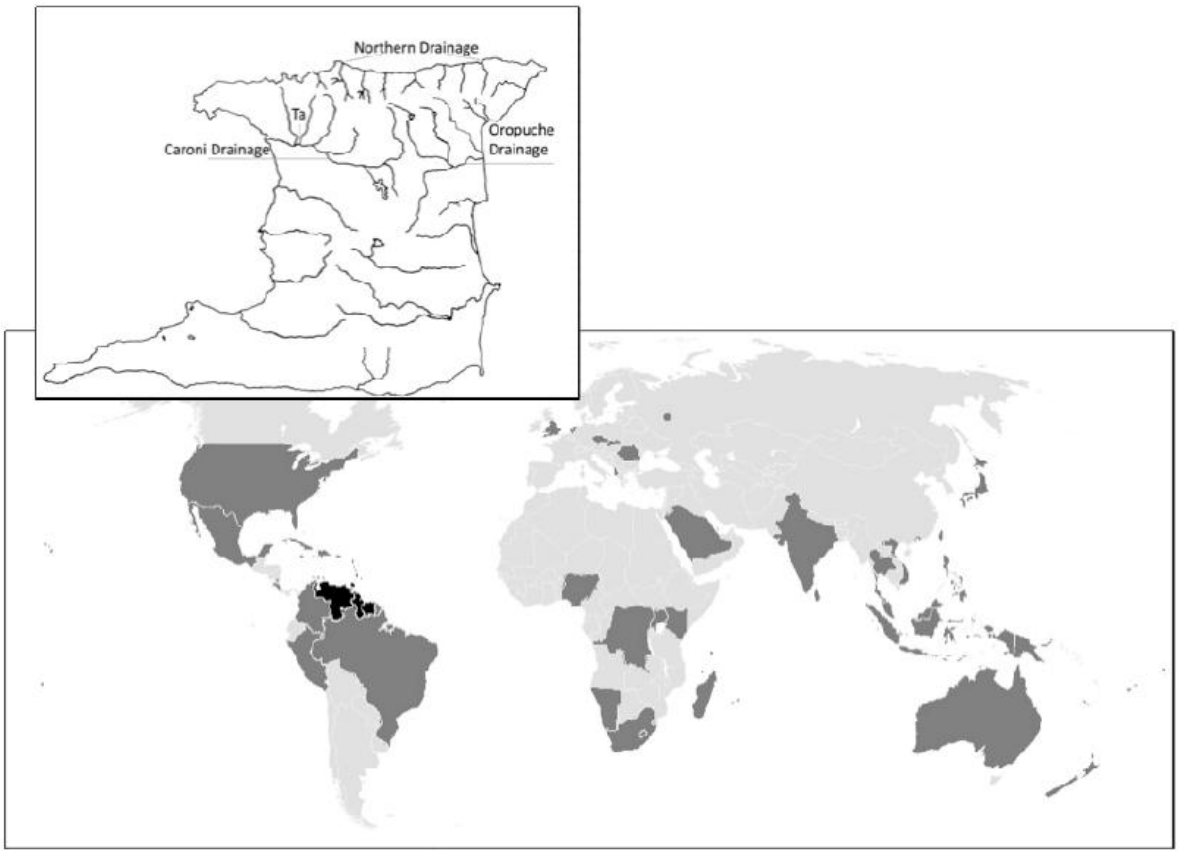



**Figure 1.2:** Worldwide distribution of *Poecilia reticulata* showing native range (black) and invasive distribution (dark grey). Native distributions are taken from Magurran (2005), invasive distributions from: www.fishbase.org (might be incomplete). Inset shows the island of Trinidad with the three drainages. Figure from van Wijk (2011).

in the aquarium trade (Figure 1.2). In Trinidad, guppies can be found widely distributed in freshwaters and can even survive in highly polluted water (Magurran and Phillip 2001). The ability of *P. reticulata* to adapt to, and thrive across, a range of conditions is part of the reason that it is now classed as a model organism in disciplines such as animal behaviour (Magurran *et al.* 1995, Swaney *et al.* 2001, Amundsen 2003, Croft *et al.* 2009, Agrillo *et al.* 2012), population genetics (Carvalho *et al.* 1991, 1996, Shaw *et al.* 1991, Barson *et al.* 2009), evolution (Reznick, Rodd, *et al.* 1996, Reznick *et al.* 1997, Reznick and Ghalambor 2001, 2005), conservation biology (Oosterhout *et al.* 2007), parasitology (Cable and van Oosterhout 2007, Stephenson and Oosterhout 2015, Stephenson *et al.* 2015) and sexual selection (Breden and Stoner 1987, Houde 1988, 1997, Brooks and Endler 2007). Interest in guppies was first sparked in the early 20$^{th}$ century (Schmidt 1919, Winge 1922a, 1937, Haskins and Haskins 1949), and has significantly increased over the last two decades (Magurran 2005).

1As well as being able to thrive in a wide range of conditions guppies are easy to breed, have a short generation time (in natural populations male mature in ~56 days and females mature in ~87 days [Reznick *et al.* 1997]) and respond quickly to manipulation both in the wild and captivity, resulting in the species being coined as ‘fruitflies of fish reproductive behaviour’ (Amundsen 2003). The parallel evolution which has occurred between populations in the wild (see section 1.3.1) has

led not only to variation in life history traits, but also in other traits such as colour patterns (Figure 1.3) and behavioural traits (Endler 1995, Reznick, Rodd, *et al.* 1996, Magurran 2005). Translocation experiments have shown that the response to predation regimes can be easily induced, with fish rapidly evolving phenotypes similar to those in populations with the same predation regime (Endler 1980, Reznick and Bryga 1987a, Reznick 1990, Karim *et al.* 2007). The conspicuous colour patterns exhibited by males to attract females have led to their use in many studies of sexual selection, from a focus on sex-linked genes to trade-offs between sexual selection and predation (Houde 1997, Andersson and Simmons 2006). In addition to mating behaviour, the varying predation regimes have also resulted in variable responses to predation, feeding strategies, schooling and social and learning behaviours (Seghers 1974; Magurran & Seghers 1991, 1994; Magurran *et al.* 1992). More recently, guppies have been used as a model for studying conservation issues, with insights provided from transplantation studies and predator removal (Reznick *et al.* 2008) and estimates on reintroduction success in relation to the impacts of inbreeding and parasitism (Oosterhout *et al.* 2007).

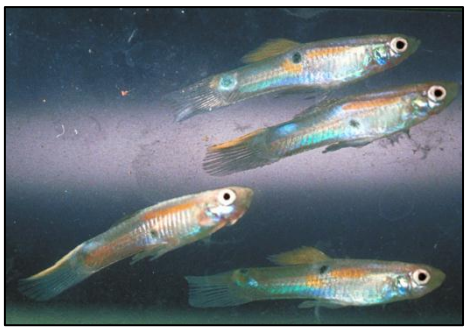


**Figure 1.3:** Variation in colour patterns in guppies from upstream and downstream populations. © Cock Van Oosterhout

### 1.5.1 Variation in life history and morphological traits

The surge of interest in the guppy during the early 1980's was triggered by the discovery of the impact predation has on a guppy phenotypes (Endler 1980, 1988, Reznick and Endler 1982). The geology and the river system of the Northern range of Trinidad have created what has been described as a 'natural laboratory' (Haskins *et al.* 1961). Several systems of parallel rivers offer natural ecological replication. Some drain from the southern slopes of the range, while northern slopes are drained by several rivers that flow into one of two main drainages, the Caroni and the Oropuche, all of which contain guppies. Many rivers are segregated by waterfalls that act as barriers to upward migration by guppies, as well as restricting colonisation by many guppy predators (Magurran 2005). Genetic analysis has found that each of the two drainages on the south side of the range and the seven drainages on the north of the range are made up of three distinct lineages (Alexander *et al.* 2006, Schories *et al.* 2009, Suk and Neff 2009, Willing *et al.* 2010).

It is not by studying the different lineages, but rather by studying populations of guppies above and below the waterfalls which partition the rivers, that the effect of predation on life history traits can be investigated. Although guppies can usually be found in the pools above and below these waterfalls, guppies in the pools below the waterfalls are often subjected to a much higher rate of predation than those above (Haskins and Haskins 1951). Individuals living under a low risk of predation are typically only predated on by the killifish, (*Rivulus hartii*), an occasional predator, and only a

significant threat to smaller and juvenile guppies (Mattingly and Butler 1994). In contrast, individuals living under high risk of predation in lower reaches, are predated typically by several predators such as the pike cichlid (*Crenicichla alta*), the wolf fish (*Hoplias malabaricus*) and the characin (*Astyanax bimaculatus*), many of which will selectively predate on large, mature individuals (Magurran and Phillip 2001, Reznick *et al.* 2012).

As well as the notable difference in predation regimes, populations from upstream and downstream sites also differ in a wide range of other traits (Endler 1995, Rodd, *et al.* 1996, Magurran 2005, Cable and van Oosterhout 2007, Palkovacs *et al.* 2011). One such trait, male colour pattern, was identified by Endler (1980; 1984), whereby the colouration of male guppies varied as a result of a combination of natural and sexual selection. Natural selection for dull colour patterns dominated in highly predated populations, and sexual selection for bright coloured males, dominated in populations with low predation rates. The set of traits which show significant variation, and which render the guppies in Trinidad of particular interest in the context of fisheries-induced evolution, are their life history traits (Reznick and Ghalambor 2005). Fish from low predation, upstream populations mature later, grow larger, allocate less energy to reproduction and produce smaller broods of larger offspring than those in higher predation, downstream counterparts (Reznick and Endler 1982, Reznick, Rodd, *et al.* 1996). Although sites up and downstream do vary in other environmental conditions such as food availability (Grether *et al.* 2001), an on-going transplantation study (Reznick 1997, Reznick and Ghalambor 2005), the results of previous translocation studies (Reznick and Bryga 1987a, Reznick *et al.* 1990) and the results of the selection study undertaken in Bangor (van Wijk 2011a), support the theory of the variation in life history traits between high and low predation populations as being an evolutionary response (a response which is a result of genetic change) to predation.

### 1.5.2 Genomic toolbox for the guppy

The rearing of guppies from different sites under common garden conditions (Reznick 1982; 1990; Reznick & Bryga 1987; 1996) has confirmed a genetic basis to the observed variation in life history traits. There is, however, still very little known about the genes underpinning such traits. Early studies focused on genes involved in colour pattern and their mode of inheritance (Winge 1922a; b; 1927), and identified guppies as the first species known to have Y-linked inheritance of genes (Schmidt 1919). Since then a strong interest in colour pattern genes has continued, with at least 43 sex-linked and autosomal colour pattern genes having been identified (Gordon *et al.* 2012). The development of microsatellite markers for the guppy (Becher *et al.* 2002, Watanabe *et al.* 2003, Olendorf *et al.* 2004, Paterson *et al.* 2005, Shen *et al.* 2007) allowed further development of the population genetic research (Barson *et al.* 2009, Suk and Neff 2009) which had begun early in the 1990's (Carvalho *et al.* 1991, Shaw *et al.* 1994). Such work was later complemented with randomly amplified polymorphic DNA (RFLP's) and amplified fragment length polymorphisms (ALFP's) to create genetic linkage maps of a range of colour pattern genes (Khoo *et al.* 2003, Watanabe *et al.* 2004, Shen *et al.* 2007). The most recent and detailed linkage map of the guppy genome was

produced by Tripathi *et al.* (2009a) using single nucleotide polymorphisms (SNP's) and identified 23 linkage groups corresponding to the 23 chromosomes, the location of the sex locus and quantitative trait loci (QTLs) for sex linked traits including body size, shape and colour (Tripathi *et al.* 2009b).

The genomic toolbox available for guppies in early 2015 consisted of a database of 18,000 expressed sequence tags (EST's) (Dreyer *et al.* 2007), many of which were used in the creation of the Tripathi *et al.* (2009b) linkage map and a full transcriptome (Fraser *et al.* 2011) The transcriptome was sequenced with 454 GS FLX technology (average read depth of 28x) with brain and body tissues from adult male and female guppies (Fraser *et al.* 2011). The EST library was created with SMART cDNA technology using whole embryos, newborn fish and adult liver, testis, brain, retina, and skin (Dreyer *et al.* 2007). Both the transcriptome and the EST library utilised fish from a range of wild populations, and therefore will have included both large and small guppies (Dreyer *et al.* 2007, Fraser *et al.* 2011). The most recent addition to the genomic tools available for the guppy is a reference genome sequence (Fraser *et al.* 2014). The genome sequence which currently exists is that of a female from a high predation site in the Guanapo River in Trinidad. As the guppy exhibits XY sex determination the current genome contains no sequence data from the Y chromosome. Further to the raw sequences, gene models have been produced allowing predictions of the gene functions to be examined (NCBI accession: GCF_000633615.1). The current gene models have predicted that this genome sequence contains a total of 26,058 genes of which 22,982 are thought to be protein coding. The predicted size of the guppy genome is nearly 1 GB (Willing *et al.* 2011) the reference sequence is only 731.62 MB long and is therefore incomplete. Despite being incomplete the reference genome provides an opportunity to examine genome-wide effects of size-selective harvesting, as well as the ability to fully examine genetic variants underpinning quantitative traits of interest.

Despite the now significant genomic resources available for guppies, little is known about the genes underpinning body size and maturation rate in guppies (or indeed any fish). Tripathi *et al.* (2009b) identified QTLs for body size in their mapping population however, the QTLs accounted for only 35-45% (with a major QTL on linkage group 12 accounting for 20-30%) of the variation in body size. These results leave over 50% of the variance unexplained and have yet to be tested in any other population. It is also interesting to note that the major QTL for body size identified on linkage group 12 (also containing the sex locus) in the mapping study undertaken by Tripathi *et al.* (2009b), showed no signs of selection in the selection experiment undertaken by (van Wijk 2011a). Further, a recent study examining selection across the genome in wild guppy populations found only five SNPs which were consistently correlated with predation regime, none of which were linked to body size or maturation timing (Fraser *et al.* 2014).

## 1.6 Molecular Ecology

The use of molecular techniques to identify and characterize the genomic architecture underpinning adaptively important traits such as body size and the timing of maturation is a major goal of molecular ecology (Stapley *et al.* 2010). Such traits can be underpinned not only by variation in the coding DNA sequence but also by regulatory changes. While these regulatory changes can be the result of sequence changes, increasing evidence suggests that epigenetic modifications may be responsible for some of the observed variation (Bossdorf *et al.* 2007).

Today, the term molecular ecology is widely used and encompasses an array of topics from evolutionary genetics and conservation genetics to behavioural ecology and microbial ecology and has been broadly defined as the "application of molecular genetic methods to ecological problems" (Beebee and Rowe 2008). Although the term 'molecular ecology' was not really used before the mid-1980's, the concept of bringing together the fields of ecology and genetics was not new. Ford (1975) pioneered the field of ecological genetics which he described as the "adjustments and adaptations of wild populations to their environment" through the use of both field and laboratory work. Early studies in the field of ecological genetics focused on visible polymorphisms which were easily scored, ecologically important and known to be underpinned by genetic variation, such as melanism in the peppered moth (Kettlewell 1958), polymorphism in the scarlet tiger moth *Panaxia dominula* (Fisher and Ford 1947) and banding patterns in land snails *Cepaea nemoralis* (Cain and Sheppard 1950). One exception to the use of visible polymorphisms was the use of chromosomal inversions to study populations of the fruit fly *Drosophila pseudoobscura* across the West Indies (Ayala *et al.* 1971). These early studies helped to overturn the opinion that natural selection was not capable of bringing about substantial adaptive changes alone, and that any significant genetic change was underpinned by processes not yet understood (Ford 2005). However, the use of phenotypic polymorphism and chromosomal inversions could only provide limited information about the contribution of natural selection and genetic drift in shaping phenotypic diversity due to pleiotropic effects, phenotypic plasticity and potential epigenetic contributions. The use of visible polymorphisms also restricted the species which could be studied to those which had easily quantifiable and ecologically relevant polymorphisms.

## 1.7 Quantitative traits

In order to increase the ubiquity of traits for study, one alternative is to focus on quantitative traits. Although still polygenetic, quantitative traits such as height, mass and growth rate are continuous throughout the population instead of only being discrete polymorphisms (Falconer and Mackay 1966, Mackay 2009). As well as helping to determine the contribution and strength of the processes involved in natural selection, quantitative traits can also be used to help understand how well a population or species will be able to respond to changing environments (Hoffmann and Willi 2008, Björklund *et al.* 2009, Anderson *et al.* 2014). As the majority of traits which will be affected by current and ongoing anthropogenic changes, including life history, behavioural and

morphological traits are quantitative, a thorough understanding of the mechanisms underpinning quantitative traits is key to predicting and mitigating any effects anthropogenic changes may have (Shaw and Etterson 2012, Anderson *et al.* 2014). The area of study which focuses on quantitative traits is known as quantitative trait analysis and has the primary aim of discovering quantitative trait loci (QTL). By determining the genetic loci underpinning quantitative traits researchers hope to be able to answer a wide range of questions such as whether traits are determined by few loci of large effect or many loci of small effect, whether the loci underpinning traits are common across different populations or even species and whether selection is primarily acting on existing standing variation or new mutations (Slate 2005, Hohenlohe *et al.* 2010, Burke 2012, Conte *et al.* 2012, Thurber *et al.* 2013).

### 1.7.1 Mapping a QTL

The principle for mapping a QTL is simple, and is based on linkage disequilibrium between markers and QTLs (Falconer and Mackay 1966). If a marker and QTL are physically close proximity, they would also be in linkage disequilibrium (although not always). Therefore if a marker shows significant association with a trait in a mapping population, it would be expected that either the marker, or a QTL nearby, is exerting a significant effect on the value of the trait (Slate 2005). Mapping of populations is typically created using crosses between lines or populations that differ for the trait of interest. Due to the complete linkage disequilibrium found in the $F_1$'s, the most widely used and powerful cross is that of a cross between two inbred lines, which differ for both the trait in question and the markers being used (Lynch and Walsh 1998, Erickson *et al.* 2004). A wide range of experimental designs can be used with inbred line crosses the most basic of which are the $F_2$ and the backcross design.

The $F_2$ design explores associations between phenotype and genotype in the offspring of crosses between the $F_1$s. The backcross design looks for associations in the offspring of a cross between the $F_1$ and one of the parental lines (Lynch and Walsh 1998). Expansions on these basic designs, such as the recombinant inbred lines (RILs) and near isogenic lines (NILs), which are produced by several generations of selfing or backcrossing, as well as double haploid's (DH: created by using chemical treatment to double the number of chromosomes in gametes), have been used to map QTLs in many species of plants (Tanksley *et al.* 1996, Groh *et al.* 1998, Brouwer and St Clair 2004, Wang *et al.* 2006, Szalma *et al.* 2007, Zhang *et al.* 2008, Balasubramanian *et al.* 2009, Huang *et al.* 2011, Naegele *et al.* 2014) and some animals (Young *et al.* 1998, Martinez *et al.* 2002). Although such designs (RIL, NIL, DH and backcross) allow testing for gene x environment effects and for detecting epistasis, only the $F_2$ design allows the level of dominance to be estimated (Lynch & Walsh 1998; Erickson *et al.* 2004), highlighting the need for careful consideration of experimental design.

Despite being the most efficient and powerful method of detecting QTL, creating inbred lines can be difficult, and for many species such manipulation is simply not possible due to logistical (e.g. long

generation times) and/or ethical considerations (e.g. QTL analysis in humans). Another disadvantage of using inbred lines is that only a small sample of the alleles present in the original population will be present in the mapping population making it likely that small effect QTLs will be missed and the effect of QTLs identified overestimated (Erickson *et al.* 2004). The difficulty in detecting QTLs in outbred populations lies in how informative the markers are. To be fully informative with respect to QTL detection, an individual must be heterozygous at both the QTL and marker locus. By crossing individuals from inbred lines, heterozygosity can be guaranteed, though this is not the case when individuals from outbred populations are used. The other key difficulty is that the linkage phase between the marker and the QTL is not known (Gao *et al.* 2009).

Several techniques have been developed for detecting and mapping QTLs in outbred or natural populations, the two most commonly used are that of sibships and pedigrees. Both methods examine variation in the trait values between relatives in conjunction with alleles which are identical by descent (IBD) (Lynch and Walsh 1998, Erickson *et al.* 2004, Slate 2005). Although the use of pedigrees to detect QTLs is time consuming it has been suggested that general pedigree techniques are more powerful than using sibships (Slate *et al.* 1999, George *et al.* 2000, Slate 2005). Despite sibships being used widely for QTL studies in livestock (Olsen *et al.* 2002, Hiendleder 2003, Mizoshita *et al.* 2004, Takasuga *et al.* 2007) and pedigrees in humans (Almasy *et al.* 1999, Deng *et al.* 2002, Liu *et al.* 2004), their use for QTL detection in an evolutionary or ecological context is still rare (Slate *et al.* 2002, Backström *et al.* 2006, Poissant *et al.* 2013). In some cases, methods used traditionally for mapping QTLs from inbred crosses have been adapted for use with outbred populations (Song *et al.* 1999, Martinez *et al.* 2002, Deeb and Lamont 2003). Recent technological advances (Stapley *et al.* 2010) have enabled researchers to identify large numbers of loci and pick only those that are homozygous (and therefore heterozygous in the $F_1$), at the same time as genotyping many individuals (Colosimo *et al.* 2004, Shapiro *et al.* 2004, Baird *et al.* 2008).

Once a mapping population has been created, the type of marker analysis that is going to be used needs to be considered, i.e. single-marker analysis, interval mapping or composite interval mapping (Doerge 2002). Single marker analysis is the simplest, and tests each marker individually for an association between the marker and the trait value (Falconer and Mackay 1966, Lynch and Walsh 1998, Doerge 2002). Although single marker analysis uses simple statistical tests such as t-tests, ANOVA and linear regression, and can therefore be run in most basic statistics packages, it requires large sample sizes to reliably detect a QTL and makes it difficult to differentiate between the position and effect of a QTL (Lynch and Walsh 1998, Doerge 2002). Interval mapping overcomes these difficulties by testing for an association between pairs of adjacent markers and the trait in question. By calculating the logarithm of odds score (LOD) for sections of the genome between each pair of markers (usually every 2cM) using a maximum likelihood estimator, interval mapping is able to determine the probability of a QTL being present in a given interval and therefore differentiate between position and effect (Doerge 2002, Erickson *et al.* 2004). Both single marker

analysis and interval mapping have the disadvantage of not distinguishing multiple linked QTLs. To overcome the latter limitation, composite interval mapping combines interval mapping with multiple regression to consider other markers as well as the adjacent pair (Lynch & Walsh 1998), thus allowing finer resolution QTL location, and multiple linked QTLs to be distinguished from a single large effect QTL.

### 1.7.2 Association mapping

Although a wide range of studies have successfully mapped QTLs in outbred and natural populations using a QTL mapping approach, on-going reduction in the cost of sequencing and genotyping has led to a rise in popularity of association mapping to identify QTLs (Breseghello and Sorrells 2006, Yu and Buckler 2006, Agrama *et al.* 2007, Malosetti *et al.* 2007, Casa *et al.* 2008, Charlier *et al.* 2008, Harjes *et al.* 2008, Sahana *et al.* 2010, Johnston *et al.* 2011, Schielzeth *et al.* 2012). Association mapping involves collecting phenotypic and genotypic data from individuals in a population. Although very similar to QTL mapping, the fundamental difference lies in the level of control the researcher has over recombination. In QTL mapping, the level of recombination is controlled by the researcher and is often very low, while in association mapping the level of recombination is often very high and cannot be controlled by the researcher (Myles *et al.* 2009). Association mapping thereby requires a much larger number of markers and creates the potential for identification of false QTLs if possible relatedness between the selected individuals is not taken into account. However association mapping also allows for high resolution mapping of any QTLs identified.

Like QTL mapping, and perhaps even more so due to the high level of recombination, association mapping relies on the presence of strong linkage disequilibrium (LD) between the markers and the QTL. The power of an association mapping study is determined by how strong the linkage between the marker and QTL is (Long and Langley 1999). Before beginning an association mapping study, it can therefore be extremely useful to consider the speed at which LD decays in the population being studied. In populations for which little genomic information is currently available, such estimates can be difficult, as the degree of LD decay has been shown to vary not only between species and populations, but also between loci (Remington 2001, Tenaillon and Sawkins 2001, Rafalski 2002, Whitt *et al.* 2002, Jung *et al.* 2004). There are two main approaches to identifying a QTL using an association mapping study: the candidate gene approach and genome-wide approach. Both approaches can be, and have been applied to the earlier described QTL mapping techniques, and the same principles of each approach can be applied to both QTL and association mapping.

The candidate gene approach relies on some prior knowledge of the mechanisms underlying the trait of interest (Myles *et al.* 2009). With such knowledge relevant candidate genes can be selected and markers within those genes investigated for an association with the trait. The latter approach has been successfully used to identify genes determining oil content in maize (Zheng *et al.* 2008) and flowering time in *Arabidopsis thaliana* (Werner *et al.* 2005), and was used initially in humans.

However, its failure to identify several disease genes (Altshuler *et al.* 2008, Brown *et al.* 2013) and the reduction in sequencing costs has led to most studies now adopting the genome wide approach.

The genome-wide approach requires no *a priori* knowledge of the mechanisms or genes underlying the trait. As the name suggests, genome wide association studies (GWAS) involve searching the entire genome for markers, typically single nucleotide polymorphisms (SNP's), associating with the trait (Myles *et al.* 2009). The genome-wide approach usually requires a large number of markers, however how many will depend on the size of the genome, the strength of linkage disequilibrium and the density of the map required (Hirschhorn and Daly 2005). The major drawback of association mapping is the high potential for the detection of false positives as a result of population structure or relatedness (Wang and Barratt 2005). It is well known that randomly mating populations rarely exist outside of population genetic theory, and that the non-random mating structure typical of natural populations results in patterns of population structure and relatedness. If these patterns happen to coincide with levels of variation in a particular trait (i.e. plants from one family have larger seeds than plants in another) and remain undetected, association mapping is likely to lead the identification of false positives in which an apparent association between a marker and a trait is actually the result of population structure or relatedness (Lander and Schork 1994). The high levels of false positives found in some studies (Aranzana, Kim, Zhao, and Bakker 2005) has led to the development of several methods which take population structure into account (Spielman and Ewens 1996, Hinds *et al.* 2004, Yu and Buckler 2006, Kang *et al.* 2008, Price *et al.* 2010, Zhou and Stephens 2012).

## 1.8 Advances in technology

Previously, several references to the technological advances in sequencing have been made. We will now briefly discuss these advances and in relation to the identification and mapping of QTLs. Dideoxy sequencing, first described by Sanger *et al.* (1977), was the first widely used sequencing technology and has since seen many improvements such an increase in the length of sequence reads (Varshney *et al.* 2009) at the same time as the cost per reaction has fallen (Mardis 2008). For more than thirty years dideoxy sequencing was the most popular DNA sequencing technique, being key to the success of many significant achievements in genetics such as the completion of the first human genome sequence (Morozova and Marra 2008, Metzker 2010, International Human Genome Consortium 2004) and the initiation of the Barcode of Life project (Shokralla *et al.* 2012). However dideoxy sequencing has its limitations (Morozova and Marra 2008, Ansorge 2009, Varshney *et al.* 2009, Metzker 2010), primarily low throughput. The requirement for electrophoretic separation of DNA fragments in Sanger sequencing increases time and limits the number of reactions that can be run in parallel. When combined with an average read length of 500-600bp, these limitations make Sanger sequencing of large amounts of sequence data both time consuming and expensive. The desire of biologists to overcome these limitations as well as the demand for faster and cheaper DNA sequencing led to the development of what are now known as massively

parallel sequencing (MPS) technologies (Schuster 2008, MacLean *et al.* 2009, Liu *et al.* 2012, Mardis 2013).

The key difference between Sanger sequencing and MPS is the ability of MPS to sequence millions of reads in parallel as opposed to the 96 well capability typical of advanced automated Sanger sequencers (Shokralla *et al.* 2012). By sequencing large numbers of reads in parallel, MPS can produce large quantities of data in a much shorter time and much more cheaply than any Sanger technology. The first of the MPS technologies, the Roche 454 Genome sequencing platform became commercially available in 2005 however it has now been superseded by a wide range of more powerful technologies (e.g. HiSeq 2000; Ion torrent Personal Genome Machine; HeliScope and PacBio RS SMRT, Nanopore). MPS technologies use a wide range of biochemistry for base identification, detailed descriptions of which can be found in several reviews (see Mardis 2008, Morozova and Marra 2008, Shendure and Ji 2008, Shokralla *et al.* 2012).

The advantages of MPS to identifying QTL are simple. By reducing the cost and time required to obtain DNA sequence data, more individuals and markers can be used in a mapping study (both QTL and association mapping). The increase in genomic studies undertaken as a result of the developments in MPS will enable easier and more accurate positioning of a QTL on the genome (Stapley *et al.* 2010). To date, MPS technologies have been used to identify QTLs in both non-model (*Salvelinus namycush:* Goetz *et al.* 2010; *Heliconius melpomene:* Ferguson *et al.* 2010; *Saccharomyces cerevisiae:* Swinnen *et al.* 2012; *Sus spp.*: Ramos *et al.* 2009; *Coregonus spp.*: Renaut *et al.* 2011; *Brassica oleracea*: Lee *et al.* 2015; *Paralichthys olivaceus*: Shao *et al.* 2015; *Oreochromis niloticus*) and model organisms (*Triticum spp.*: Trick *et al.* 2012; *Arabidopsis lyrata:* Turner *et al.* 2010; *Gasterosteus aculeatus:* Hohenlohe *et al.* 2010; *Littorin saxatilis:* Galindo *et al.* 2010; *Drosophila melanogaster*: Najarro *et al.* 2015).

Although the advent of MPS technologies has transformed our ability to sequence and study whole genome data sets, the cost of doing so for the numbers of individuals required to accurately identify QTLs remains out of reach of most small and medium sized projects. Consequently, several reduced representation techniques that identify and genotype large numbers of polymorphisms across many individuals have been developed. Simultaneous identification of polymorphisms and identification of individual genotypes is known as genotyping-by sequencing (GBS), although this term has also been used to refer a specific type of reduced representation sequencing (Poland *et al.* 2012). Regardless of the protocol and terminology being used, most techniques obtain the reduced representation of a genome by digestion with restriction enzymes (Davey *et al.* 2011). Exceptions can be seen when a specific known section of the genome is being targeted, in which case oligonucleotide probes are usually designed to capture the regions before sequencing (Teer and Mullikin 2010, Teer *et al.* 2010, Kiialainen *et al.* 2011).

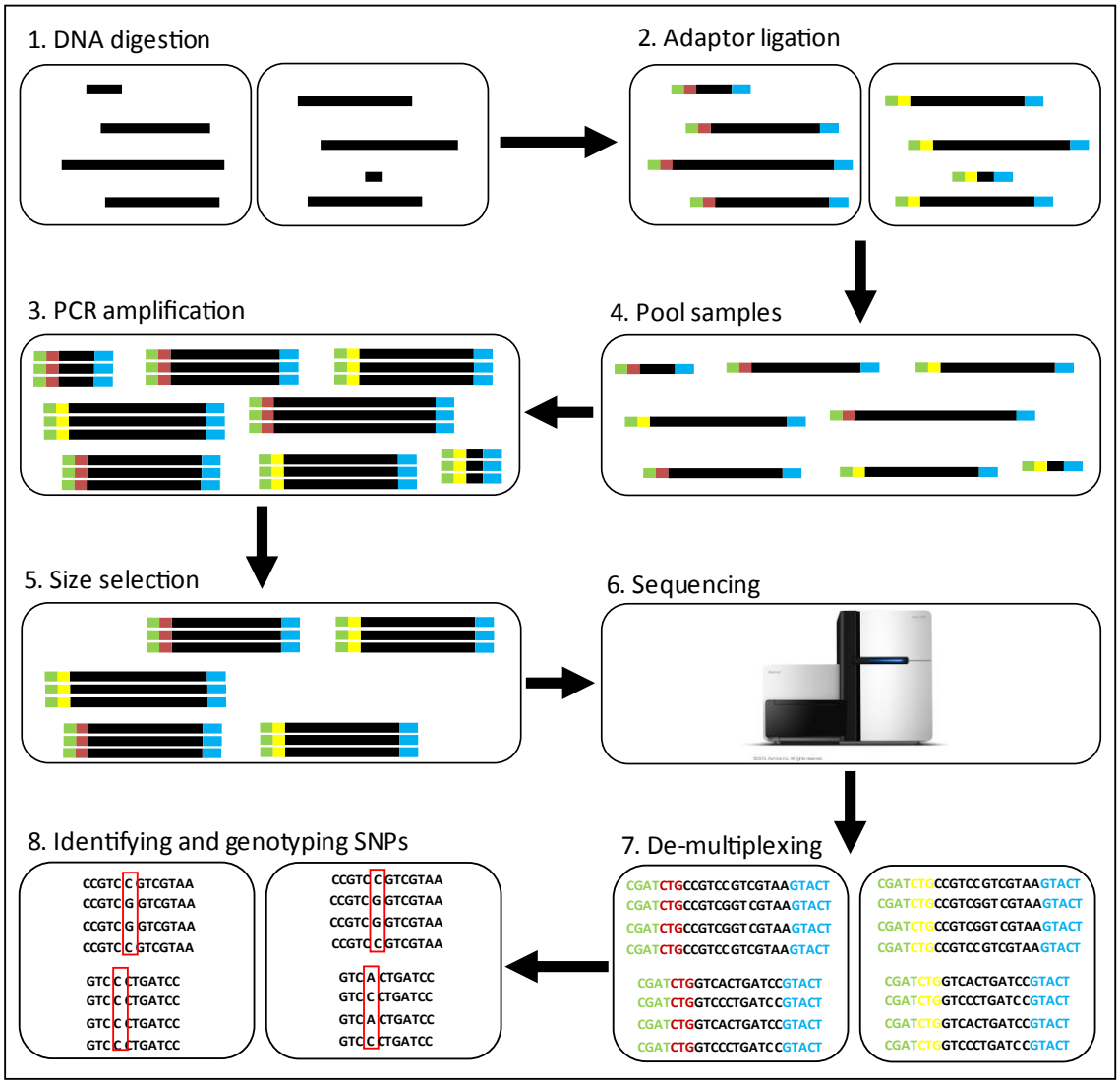



**Figure 1.4:** Schematic showing the RAD procedure from library preparation to SNP calling.

One technique to utilise restriction enzymes is that of restriction-site associated DNA sequencing (RAD-Seq). Although initially used with low-cost microarray genotyping (Miller, Atwood, *et al.* 2007, Miller, Dunham, *et al.* 2007), rapid progress in MPS has led to the combination of MPS with RAD markers and the development of RAD-Seq which has enabled the discovery of thousands of SNPs (Baird *et al.* 2008, Baxter *et al.* 2011). By only sequencing specified restriction sites (according to the enzyme chosen) RAD-seq not only reduces the complexity of the genome, but increases the coverage at each site, thereby allowing more reliable base calling and genotyping, and increases the chances of the same sites being sequenced in each sample (Rowe *et al.* 2011, Houston *et al.* 2012). Before fragments are sequenced, adaptors containing a molecular identification sequence (MID) which are unique to each sample or individual are ligated to each allowing the multiplexing of large numbers of samples (Baird *et al.* 2008, Chutimanitsakun *et al.* 2011, Houston *et al.* 2012). Genomic DNA is first digested with a restriction enzyme before P1 adaptors (containing the Illumina sequencing and amplification primer and a barcode) are ligated. The samples are then pooled, randomly sheared and the fragments size-selected. The P2 adaptor (Y shaped with divergent

ends) is ligated on and the fragments are PCR amplified. Due to the diverged ends of the P2 adaptor, only those fragments to which both the P1 and P2 adaptors have successfully ligated will be amplified, ensuring that all sequenced fragments contain the restriction site (Baird *et al.* 2008). Figure 1.3 shows a schematic of the RAD process.

Since the initial adaptation of RAD sequencing for use on MPS platforms, many developments of the original protocol have occurred (ddRAD: Peterson *et al.* 2012; ezRAD: Toonen *et al.* 2013; 2bRAD: Wang *et al.* 2012 and GBS: Elshire *et al.* 2011, Sonah *et al.* 2013, J. A. Poland *et al.* 2012). As well as discovery of polymorphisms in a plethora of both model and non-model species (Barchi *et al.* 2011, Willing *et al.* 2011, Amish *et al.* 2012, Scaglione *et al.* 2012, Pegadaraju *et al.* 2013, Senn *et al.* 2013, Vandepitte *et al.* 2013), the various RAD techniques have been successfully applied to population genomic applications such as estimation of population structure (Hohenlohe *et al.* 2010, 2012, Andersen, Gerke, *et al.* 2012, Hyma and Fay 2013, Keller *et al.* 2013, Ogden *et al.* 2013), linkage mapping (Amores *et al.* 2011, Baxter *et al.* 2011, Chutimanitsakun *et al.* 2011, Kakioka *et al.* 2013, Gonen *et al.* 2014, Wu *et al.* 2014, Huber *et al.* 2015), phylogenetics and phylogeography (Raymond *et al.* 2001, Emerson *et al.* 2010, Rubin *et al.* 2012, Cariou *et al.* 2013, Catchen *et al.* 2013, Reitzel *et al.* 2013, Roda *et al.* 2013, Jones *et al.* 2013, McCormack *et al.* 2013, Nadeau *et al.* 2013, Hess *et al.* 2014, Hipp *et al.* 2014, Leaché *et al.* 2014) and association studies (Parchman *et al.* 2012; Pfender *et al.* 2011 Chutimanitsakun *et al.* 2011; Palaiokostas *et al.* 2013; Richards *et al.* 2013). Because of the success of RAD-seq in confirming and fine mapping, previously identified QTLs in species for which significant genetic resources already exist (Miller, Atwood, *et al.* 2007, Chutimanitsakun *et al.* 2011) as well its use with species for which no reference genome exists (Willing *et al.* 2011), RAD-seq is a tool likely to be used in many future studies (Stapley *et al.* 2010, Rowe *et al.* 2011).

The study by Hohenlohe *et al.* (2010), which was one of the earliest to use RAD-seq to identify QTLs, utilised parallel populations of Threespine Sticklebacks (*Gasterosteus aculeatus)* which are adapted to freshwater and marine environments. The large number of loci and individuals genotyped using RAD-seq enabled identification of 35 QTLs linked to either osmoregulation or skeletal biology (traits which differ significantly between the populations), many of which had been identified in previous studies. By confirming the presence of previously identified QTLs as well as several previously unknown, the study highlighted the potential of RAD-seq in the identification of QTLs. Since then, RAD-seq has identified QTLs in fish for a range for traits including disease resistance (Houston *et al.* 2012), migration propensity (Hecht *et al.* 2013), climate tolerance (Narum *et al.* 2010), colour dimorphism (Takahashi *et al.* 2013) and life history (Gagnaire *et al.* 2013).

Despite their popularity, genome reduction methodologies utilising restriction enzymes have several inherent biases (Arnold *et al.* 2013, Davey *et al.* 2013, Gautier *et al.* 2013). The most significant of these is variation in the fragment lengths due to incomplete shearing, which can result in low read depth at affected loci (Davey *et al.* 2013). While such effects have the potential to cause significant

problems when the library preparation includes a shearing step, it should not be an issue with protocols utilising two enzymes (Elshire *et al.* 2011, Peterson *et al.* 2012, Poland *et al.* 2012, Sonah *et al.* 2013). Another potential source of bias is that of restriction site heterozygosity (Arnold *et al.* 2013, Davey *et al.* 2013, Gautier *et al.* 2013). Any polymorphism in the restriction site would prevent the restriction enzyme from digesting, resulting potentially in one chromosome for that restriction fragment remaining unsequenced. Such a problem is known as allele dropout, and can result in a heterozygote being incorrectly called as a homozygote, which is similar to null alleles in microsatellite studies. One suggestion for overcoming allele dropout is to consider only SNPs successfully genotyped in all individuals, as it is likely that a polymorphism in the restriction site will result in the SNP not being genotyped in some individuals (Davey *et al.* 2013). However, such a solution assumes that the allele frequency of the non-amplified allele (with the mutation) is sufficiently high so that it can be expected to be found in homozygous state. Only analysing SNPs successfully genotyped in all individuals would, however, result in the loss of large numbers of good quality SNPs and is therefore not an effective solution for most studies. When using reduced representation techniques to identify QTLs their sensitivity to hard and soft selective sweeps should also be taken into account. Loci involved in a hard selective sweep are expected to produce large regions around the causal loci in which diversity is reduced, while following a soft selective sweep the drop in diversity will be less pronounced. A method which only samples at random points across the genome will thus have less power to detect a soft sweep than a hard one (Ferrer-Admetlla *et al.* 2014). Therefore, although reduced representation techniques have successfully identified QTLs in a wide range of species previous knowledge of potential genes selection may be acting on can be useful.

## 1.9 Epigenetics

While a large proportion of the heritable basis of quantitative traits is thought to result from the transmission of variation in DNA sequences at QTL between a parent and its offspring, a growing body of evidence suggests that non-genetic inheritance might also play an important role

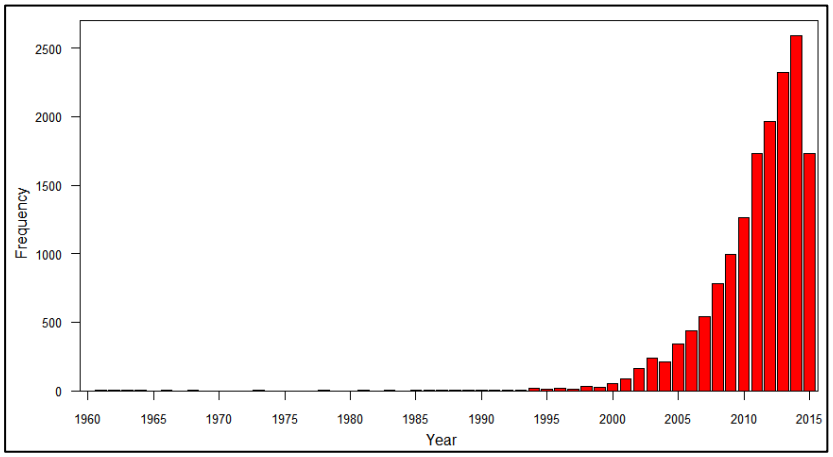



**Figure 1.5:** Number of published papers with "epigenetics" in the topic between 1961 and 2015.

(Bonduriansky *et al.* 2012). The term epigenetics, literally meaning 'above genes', broadly refers to the study of heritable changes which cannot be explained by variation or changes in the DNA sequence (Bird 2007, Tost and Gut 2007), though this simple definition is still debated (Bird 2007, Richards *et al.* 2010, Doğan *et al.* 2013, Felsenfeld 2014, Deans and Maggert 2015). The field of epigenetics has undergone an exponential growth in the last fifteen years, with approximately 15,600 papers published on the subject to date, (based on a search of the Web of Science database for papers with "epigenetics" in the topic, search date 27.09.2015), 75% of which have been published in the last 5 years (figure 1.4).

Epigenetic modifications can be broadly grouped into four main categories: chromatin remodelling; histone modifications; non-coding RNA (ncRNA) and DNA methylation. Chromatin remodelling is often also referred to as nucleosome remodelling and involves the enzymatic disruption, assembly or moving of nucleosomes (Becker and Workman 2013). As the packaging of DNA into nucleosomes creates a barrier to transcription, nucleosome remodelling plays a vital role in the regulation of gene expression (Portela and Esteller 2010). Histone modification involves direct modifications to the tails of the histone proteins which make up the nucleosome. Two of these modifications, acetylation and methylation were among the first epigenetic marks to be linked to transcriptional activity in eukaryote cells (Allfrey *et al.* 1964, Paik and Kim 1971). Since then a wide range of post-transcriptional histone modifications have been described including, but not limited to, phosphorylation, ubiquitination, sumoylation, and biotinylation. The most recently described epigenetic mechanism is that of ncRNA. The discovery that in eukaryotes, the majority of the genome is transcribed, and that these transcribed sections of RNA are not protein coding (Costa 2005) has led to studies of the epigenetic potential of these ncRNAs. These studies have implicated ncRNA in gene regulation both directly and indirectly (see Bernstein & Allis 2005; Costa 2005; Kaikkonen *et al.* 2011 for reviews).

### 1.9.1 DNA Methylation

One of the most widely studied epigenetic marks is DNA methylation in which the cytosine base (C) is converted to 5-methylcytosine ($C^m$) through the covalent addition of a methyl group to the 5 position of the cytosine ring. These methylated cytosine bases, which are sometimes referred to as the "$5^{th}$ base" (Ammerpohl *et al.* 2009, Lister and Ecker 2009, Zhu 2009), can be transmitted through cell division by both DNA strands making them the archetypal epigenetic mark. Methylation in eukaryotes is only found on cytosine (C) bases (Suzuki and Bird 2008) and can be categorised according to the sequence adjacent to the cytosine: CpG; CHG or CHH (where H= A, C or T) (Feng *et al.* 2010). Although conserved throughout most major eukaryote groups, the amount and type of methylation varies significantly between species (Lee *et al.* 2010). In plants methylation occurs symmetrically (on both strands) at CHH, CHG and CpG sites, but also asymmetrically at CHH sites. The model species *Arabidopsis thaliana*, for example, has been found to exhibit methylation levels of 24%, 6.7% and 1.7% at CpG, CHG and CHH sites respectively (Cokus *et al.* 2008, Lister *et al.* 2008). In contrast, vertebrate methylation primarily occurs symmetrically at CpG sites (Bird and

Wolffe 1999), although high levels of non-CpG methylation have been identified in embryonic human stem cells and human and mouse brain cells (Laurent *et al.* 2010, Xie *et al.* 2012, Lister *et al.* 2013, Guo *et al.* 2014). Non-CpG methylation had previously been thought to be established and maintained only by plant specific enzymes (Cao and Jacobsen 2002, Pinney 2014) therefore the presence of non-CpG methylation in humans was somewhat surprising. While studies have examined the role of non-CpG methylation in plants (Saze *et al.* 2012, Colicchio *et al.* 2015) the importance of non-CpG methylation in vertebrates has yet to be clarified.

In addition to variation in the sites methylation occurs at, there is also variation in the pattern of methylation found in vertebrates and invertebrates. Most vertebrates exhibit a global pattern of methylation while invertebrates typically exhibit a mosaic pattern of methylation. Invertebrate genomes exhibit a wide range of methylation levels, from organisms which have almost no detectable methylation (for example *Drosophila melanogaster* (Capuano *et al.* 2014) and *Caenorhabditis elegans* (Bird 2002)), to those with intermediate levels of methylation such as *Apis mellifera* (Lyko *et al.* 2010) and *Crassostrea gigas* (Gavery and Roberts 2010). Where methylation is present in invertebrates it occurs in a mosaic pattern with regions of methylated and unmethylated CpGs interspersed throughout the genome (Suzuki and Bird 2008). The globally methylated genomes found in vertebrates have very high levels of methylation (70-80% of all CpG sites are methylated) (Head 2014). Sections of the genome which have a high frequency of C and G nucleotides, and particularly a high frequency of CpG sites are called CpG islands (Yamakoshi and Shimoda 2003, Illingworth and Bird 2009, Deaton and Bird 2011) These sections of the genome are usually free from methylation occur in the promoter regions of genes.

### 1.9.2 Effect of DNA methylation

DNA methylation is involved in a range of biologically important processes including gene regulation (Turker 2002, Baylin 2005, Klose and Bird 2006, Smith and Meissner 2013, Colicchio *et al.* 2015), silencing of transposable elements (Ikeda and Nishimura 2015), X chromosome inactivation (Sharp *et al.* 2011, Cotton *et al.* 2015), imprinting (Li *et al.* 1993, Reik and Walter 2001, Lucifero *et al.* 2002) and cell differentiation (Huang and Fan 2010, Khavari *et al.* 2010) as well as being associated with disease (Baylin 2005, Robertson 2005, Bakulski *et al.* 2012, Bergman and Cedar 2013) and aging (Horvath 2013, Jones *et al.* 2015, Jung and Pfeifer 2015, Zampieri *et al.* 2015). The location of methylation (intra or intergenic) can determine the effect it has on gene expression. For example, it is widely recognised that methylation of CGIs in promoter regions is associated with long term gene silencing (Jones and Takai 2001, Jaenisch and Bird 2003), although it is thought that the methylation of CGIs is not the initial silencing mechanism but instead acts as a 'lock' to stabilise a previously silenced gene (Bird 2002, Jones 2012). Conversely, intergenic methylation has been linked to genes which are actively transcribed (Hellman and Chess 2007, Zilberman *et al.* 2007), possibly through alternative splicing and transcription from alternative promoters (Maunakea *et al.* 2010, Kulis *et al.* 2013). While the mechanism by which methylation of promotor CGIs represses gene expression is well understood (Illingworth and Bird 2009, Wang *et al.* 2014), much

less is known about the mechanisms underpinning gene body methylation (Hahn *et al.* 2011, Jones 2012).

The majority of research examining the effects of DNA methylation has focused on cancer and other diseases, however an increasing number of studies have examined the contribution of DNA methylation to complex traits. Outside humans, methylation has been linked to phenotypic variation in traits including, but not limited to: coat colour; obesity; tail morphology and regulation of the circadian clock in mice (Wolff *et al.* 1998, Rakyan *et al.* 2003, Azzi *et al.* 2014); flowering time, flower shape, root length and plant height in *Arabidopsis* (Cubas *et al.* 1999, Shindo *et al.* 2006, Johannes *et al.* 2009, Cortijo *et al.* 2014, Hu *et al.* 2015, Yang *et al.* 2015); plate morphology in sticklebacks (Smith, Smith, *et al.* 2014); fruit ripening in tomatoes (Manning *et al.* 2006); mating behaviour in rats (Crews *et al.* 2007, Crews 2008); caste determination in bees (Lyko *et al.* 2010); leaf growth in maize (Candaele *et al.* 2014) and body size in ants (Alvarado *et al.* 2015), tilapia (Zhong *et al.* 2014) and sheep (Cao *et al.* 2015). It is extremely difficult to completely separate the effects of epigenetic and genetic change and the two are often interlinked (Herrera and Bazaga 2010, Smith, Kilaru, *et al.* 2014), however studies where the patterns of epigenetic variation do not correlate with the observed genetic variation, (Bossdorf *et al.* 2007, Vaughn *et al.* 2007) and those where the epigenetic variation is greater than the genetic variation (Keyte *et al.* 2006, Li *et al.* 2008, Lira-Medeiros, Parisod, Fernandes, *et al.* 2010) suggest that it is possible for epigenetic modifications to determine phenotypic traits independently of genetic change.

The results from recent and ongoing research suggest that the epigenetic variation may underpin phenotypic variation in a wide range of complex and adaptive traits (Shindo *et al.* 2006, Cortijo *et al.* 2014, Zhong *et al.* 2014, Alvarado *et al.* 2015, Cao *et al.* 2015, Hu *et al.* 2015, Yang *et al.* 2015). Evidence has also shown that these changes in methylation can be induced by environmental changes such as stress and diet (Wolff *et al.* 1998, Labra *et al.* 2002, Steward *et al.* 2002, Sollars *et al.* 2003, Anway *et al.* 2005, Meaney and Szyf 2005, Ashworth *et al.* 2009, Brown *et al.* 2009, Verhoeven *et al.* 2010, Boyko *et al.* 2010, Chmurzynska 2010, Herrera and Bazaga 2010, Vandegehuchte and Janssen 2011). Methylation does not only create phenotypic variation, but might also act as a source of variation for natural selection (Angers *et al.* 2010, Verhoeven *et al.* 2010, Massicotte *et al.* 2011), thereby allowing environmental change to induce heritable phenotypic change over a short space of time. As a result it is important that the potential for epigenetic mechanisms to influence phenotypes, whether directly or indirectly, is not overlooked when undertaking population genetic studies and using selection experiments to examine phenotypic change.

## 1.10 Aim and objectives of the thesis

The overarching aim of the current thesis is to explore the genomic architecture underpinning shifts in life history traits as a result of size-selective harvesting in *Poecilia reticulata*. Although the study

by van Wijk *et al.* (2013) provided direct empirical evidence that size-selective harvesting can induce genetic change, the extent of genome wide effects was not examined. By utilising the lines produced by the van Wijk *et al.* (2013) study, the recently sequenced guppy genome (Fraser *et al.* 2014) and a reduced representation method of sequencing, here I investigate genomic mechanisms underpinning the phenotypic shifts described previously (van Wijk *et al.* 2013). Furthermore, as well as considering the effects of size-selective harvesting at the level of DNA sequence variation, potential regulatory changes will also be examined by assessing patterns of methylation.

The aim and objectives of each chapter are as follows:

- **Chapter one** aimed to give an overview of the current and relevant ideas and literature. Therefore the objectives of chapter one are:
  - To review the science of and current evidence for fisheries-induced evolution.
  - To introduce the guppy as a model system in ecology and evolution.
  - To consider methods of identifying gene / phenotype correlations.
  - To examine the role of epigenetic modifications in determining phenotypic variability.

- **Chapter two** aims to examine the genome-wide changes associated with the selection experiment described in van Wijk *et al.* (2013). To address this aim the specific objectives of chapter two are:
  - To use RAD sequencing to identify and genotype a large number of SNPs in the selection lines generated by van Wijk *et al.* (2013).
  - To identify SNPs showing signs of selection for body size using a range of $F_{st}$ outlier analyses.
  - To examine the distribution across the guppy genome and the functional significance of SNPs identified as showing signs of selection.

- **Chapter three** aims to identify regions of the genome under selection for body size in wild guppy populations in Trinidad. To address this aim the specific objectives of chapter three are:
  - To determine which rivers of those sampled show significant differences in body size between different sites.
  - To genotype the SNPs showing signs of selection in chapter two in the populations from sites at eight rivers in Trinidad.
  - To use $F_{st}$ outlier analysis to identify SNPs showing signs of selection for body size within each river.
  - To look across all eight rivers for SNPs consistently showing signs of selection.

- **Chapter four** aims to investigate regulatory changes associated with the van Wijk *et al.,* (2013) selection experiment. To address this aim the specific objectives of chapter four are:

- To identify methylation sensitive amplified polymorphisms in fish from the four post-selection lines and the pre-selection line.
- To use the methylation sensitive amplified polymorphisms to examine the level of global methylation in the guppy genome.
- To assess locus specific differences in methylation between the large and small selection lines.
- To examine both the global and locus specific changes in DNA methylation as a result of the selection experiment.

- **Chapter five** aims to bring together the results of the three data chapters in order to discuss their implications both to the study of FIE and the wider field of rapid evolution. To address this aim the specific objectives of chapter five are:
  - To summarise and provide a critical overview of the findings of the project.
  - To consider the current understanding and importance of the guppy sex chromosome.
  - To review the wider ecological impacts of size selective harvesting.
  - To examine what the evidence shows of the potential for fish to recover from size selective harvesting.
  - To discuss future work in the context of the overall thesis aim.

# Chapter 2: Genetic changes underpinning shifts in life history traits

## 2.1 Abstract

Findings from the van Wijk (2011) selection experiment demonstrated not only significant phenotypic shifts in body size and maturation schedules, but also evidence of significant divergent selection for body size at 5 out of 17 putative candidate loci. Utilising tissue samples from size-selected guppies (*Poecilia reticulata*) from the van Wijk (2011) study, the current chapter aimed to identify the genetic factors underpinning the observed life history shifts. RAD sequencing was employed to identify and type a large number of SNPs in each of 151 individuals from both the large and small selection lines, as well as individuals from the generation prior to selection. These SNPs were mapped back to the guppy reference genome and signs of selection were identified using $F_{st}$ outlier analysis and analysis of allele frequency changes. Significant and consistent signs of selection were identified in 37 SNPs, 86% of which were located on the guppy sex chromosome. The results from the current chapter showed that, in addition to previously observed genetic change, additional regions of the guppy genome responded to, and were associated with, observed phenotypic shifts.

## 2.2 Introduction

### 2.2.1 Selection experiments

The use of experimental populations to study the genetic basis of phenotypic traits, from the evolution of mutation rates (Wielgoss *et al.* 2011) and phenotypic plasticity (Garland and Kelly 2006) to the importance of epistatic interactions (Khan *et al.* 2011), is well established, and extends for over a century (Falconer 1992). Currently, a range of terms such as 'experimental evolution' (Kawecki *et al.* 2012) and 'evolve and re-sequence' (Turner *et al.* 2011) all refer to studies examining the evolutionary response of an experimental population to conditions (environmental, genetic, social etc.) imposed by the researcher. Selection experiments can be broadly classified into two categories: natural selection and artifical selection. In natural selection experiments, the founding experimental population is seperated into at least two replicates, before each replicate is exposed to contrasting conditions, and the response observed. These studies can be particularly useful in determining how a species or population might react to environmental change (Borash *et al.* 2000, Zbinden *et al.* 2008, Kolss *et al.* 2009). Artificial selection studies impose selection for a trait on individuals which show specific values of a target trait, to become the progenitors of the next generation (Burke *et al.* 2010, Johansson *et al.* 2010, Parts *et al.* 2011, Turner *et al.* 2011). Imposing artificial selection can help to ascertain the genetic factors underpinning a trait and, by controlling all other conditions, elucidate the environmental and genetic components of observed phenotypic change. Although traditionally, QTLs have been identified using experimental crosses of inbred lines, the increased availability of high throughput sequencing technologies and subsequent use of genome-wide association studies (GWAS: Liu *et al.* 2015; Nishimura *et al.*

2012; Gratten *et al.* 2007; Gutierrez *et al.* 2015; Pardo-Diaz *et al.* 2015) will allow the optimisation of selection experiments designed to identfy loci underpinning complex traits in non-model species.

### 2.2.2 Fisheries-induced evolution

Anthropogenic pressures have been shown to drive rapid phenotypic change in many species (Coltman, O'Donoghue, and Jorgenson 2003, Allendorf and Hard 2009, Darimont *et al.* 2009). One such example is the phenotypic shifts seen in populations of fish as a result of fishing. Data from both, field (Handford *et al.* 1977, Ricker 1981, Olsen *et al.* 2004, Edeline *et al.* 2007, 2009, Swain *et al.* 2007, Mollet *et al.* 2010, Neuheimer and Taggart 2010, Hutchings and Rangeley 2011) and experimental studies (Conover and Munch 2002, Biro and Post 2008, Conover *et al.* 2009, van Wijk *et al.* 2013, Uusi-Heikkilä *et al.* 2015), have shown that harvesting leads to changes in several traits, such as growth rate, size and age at maturation (Conover and Munch 2002, Olsen *et al.* 2004, Hutchings 2005, Conover and Baumann 2009b) and others (Uusi-Heikkilä *et al.* 2008; 2015; Enberg *et al.* 2012).

The potential for fishing to drive evolutionary change has been convincingly demonstrated (Conover and Munch 2002, Smith *et al.* 2007, van Wijk *et al.* 2013, Kendall *et al.* 2014, Heino *et al.* 2015, Marty *et al.* 2015). The basic theory, suggesting that fishing would result in smaller body size, slower growth and earlier maturation, is supported by observations in wild populations (Ricker 1981, Olsen *et al.* 2004, Swain *et al.* 2007, Fenberg and Roy 2008, Neuheimer and Taggart 2010, Hutchings and Rangeley 2011). In unexploited populations, natural selection would be expected to favour individuals which allocate more energy to growth in early life and therefore grow larger and mature later. Regulations in the European Union state that any fish smaller than the minimum size limit when caught (e.g. 35cm for *Gadus morhua*, 6.4 kg for *Thunnus thynnus*) must be returned to the sea (Council Regulation [EC] 850/98). Even before these regulations were brought in the desire for maximum profit resulted in the preferential removal of the largest fish. Today both the regulations and desire for maximum profit mean that the harvesting imposed by fishing is non-random. Following life history theory, the increased adult mortality resulting from the non-random nature of fishing will reverse the direction of selection, favouring individuals which allocate more energy to reproduction in early life and mature earlier and at a smaller size. However, although the potential of fisheries induced evolution (FIE) to drive shifts in life history traits is undisputed, its contribution in relation to phenotypic plasticity and environmental change remains unclear (Marshall and Browman 2007, Browman, Law, and Marshall 2008, Browman, Law, Marshall, *et al.* 2008, Enberg *et al.* 2012, Kuparinen and Hutchings 2012).

The consequences of fisheries induced evolution can be either direct (i.e. on the species being harvested) or indirect (i.e. on the wider ecosystem), both of which are typically undesirable. Direct effects, such as the changes in life history and behavioural traits, may lead to maladapted populations with increased natural mortality and reduced recruitment (Audzijonyte *et al.* 2013, Jørgensen and Holt 2013). In addition, FIE can cause a depletion in population sizes and a low

harvesting yield from these populations (Eikeset *et al.* 2013, Zimmermann *et al.* 2015). From a conservation point of view, the potential consequences on the wider ecosystem are of particular concern, as phenotypic shifts can impact on food webs, nutrient cycling, and overall biodiversity (Palkovacs *et al.* 2012, Audzijonyte *et al.* 2013). Finally, if the shifts in life history traits observed are underpinned by genetic change, there is still debate on how fast, if at all, the changes to phenotypic traits could be reversed (Dunlop *et al.* 2009, Enberg *et al.* 2009, Kuparinen and Hutchings 2012, Marty *et al.* 2015, Uusi-Heikkilä *et al.* 2015).

In order to fully understand the consequences of the observed phenotypic shifts resulting from fishing it is vital that the contribution of genetic changes are determined.

### 2.2.3 Genetic evidence of fisheries induced evolution

Early attempts to disentangle the effects of genetic change and phenotypic plasticity focused on probabilistic maturation reaction norms (PMRN) (Olsen *et al.* 2004). PMRNs attempt plot the probability of a fish maturing as a function of factors such as size and age. However the limited number of factors considered by PMRNs has led to widespread criticism of this technique (Kinnison and Hendry 2001, Kraak 2007, Kuparinen and Merilä 2007, Marshall and McAdam 2007, Morita *et al.* 2009, Uusi-Heikkilä *et al.* 2011). Several studies have attempted to use selection experiments to examine the question of whether size selective harvesting can lead to phenotypic changes which are underpinned by genetic variation.

One of the earliest studies to question whether changes in life history traits were caused by variation in the population density and size class distribution or genetic differentiation utilised the model species *Daphnia magna* (Edley and Law 1988). The study examined effects of harvesting by culling 40% of either the largest or smallest individuals in two replicate populations. The results showed that, following selection, individuals in the large harvested lines grew slower while individuals in the small selected lines grew faster (Edley and Law 1988). Although discussed, the design employed did not distinguish between environmental effect and genetic change.

More recently a study of Atlantic silversides (*Menidia menidia*) used a common garden selection in order to control environmental variation (Conover and Munch 2002). Four generations of size selective harvesting resulted in significant shifts in not only life history traits, but also physiological and behavioural traits (Walsh *et al.* 2006). Having controlled for environmental effects, the authors concluded that the observed phenotypic shifts were a result of genetic change, though evidence was indirect.

Evidence linking genetic change to changes in phenotypic traits as a direct result of size selective harvesting is limited. Considerable past work on adaptation of guppies to size selective predation (see Magurran 2005 and references therin) provided a solid foundation on which to base an experimental study of FIE in *P. reticulata*. The first study to combine a selection experiment utilising

Trinidadian guppies, *Poecilia reticulata*, with molecular genetic techniques, examined the extent of genetic change associated with contrasting harvesting regimes (van Wijk *et al.* 2013- see also section 1.3.1). Initial studies showed that changes in harvesting regime targeted at large and small males results in shifts in life history traits similar to those seen in exploited populations (Reznick et al 1990, 1997). Moreover, findings showed significant signs of selection at five of seventeen candidate loci, thereby providing the first empirical evidence of genetic changes in life history traits as a direct result of harvesting. In the current study, guppy lines derived from the van Wijk (2011) study were utilised to examine genome-wide genomic changes using a RAD sequencing approach.

A study of zebrafish was also able to identify simultaneous genetic and phenotypic change following a size selective harvesting experiment (Uusi-Heikkilä *et al.* 2015). However, although this study considered a much wider range of phenotypic traits, the phenotypic changes observed were not as marked as those seen by van Wijk *et al.* (2013), and the genes in and near where variation was observed were not directly linked to phenotypic change. In the study involving zebrafish, size-selection was imposed for 4/5 generations. However measurements of the phenotypic traits were not taken until generation 6/7, following at least two further generations of no selection. Consequently the life history traits measured may have begun to revert to their original pre-selection values between the final size-selective harvest and when they were actually measured (Conover *et al.* 2009).

### 2.2.4 Genome-wide considerations of phenotypic change associated with FIE

Now that empirical evidence of genetic change correlated with phenotypic change as a direct result of harvesting has been obtained, it is important to elucidate the genetic architecture and mechanisms underpinning such change. The questions asked in most studies of quantitative traits such as how many loci selection is acting on, and whether it is acting on new mutations or standing variation, are relevant to understanding the impacts of fishing. If selection imposed by fisheries induced evolution is having only a small effect on a large number of loci, it is expected that once fishing has ceased, such as during a phase of stock recovery, the trait would quickly recover to its pre-harvesting state. However, if selection is having a large effect on a small number of loci, it would be expected that the observed changes would take longer to revert to an optimum state under natural selection. Under such a scenario, over-exploited populations would therefore take longer to recover (Jain and Stephan 2015).

While a thorough understanding of the genes involved in the traits impacted on by fisheries is important, it may also be the case that selection affects other unstudied traits. Such associations could occur either through direct selection or genetic hitchhiking (Smith and Haigh 1974). Under strong selection, genetic hitchhiking can lead to a so-called selective sweep, which reduces the diversity in the genomic region around the gene under positive selection due to physical linkage (Pennings and Hermisson 2006).

A selective sweep can either be 'hard' or 'soft', depending on the timing and the effect of the mutation, and either complete or incomplete, according to the degree of fixation observed (Burke 2012). In a complete hard selective sweep, a new mutation will rapidly increase in frequency until it reaches fixation, causing any linked variation to also become fixed and resulting in a region around the causal mutation of significantly reduced diversity (Kaplan *et al.* 1989, Kim and Stephan 2002). A complete soft selective sweep will usually act on several mutations already present in the population (standing variation), increasing their frequency until fixation. As these ancestral mutations are likely to occur on a range of haplotypes in the population, they will not be as strongly associated to the linked genetic variants. As a consequence, soft selection will not result in such a stark reduction in diversity in the DNA surrounding the causal mutation (Hermisson and Pennings 2005, Przeworski *et al.* 2005, Pennings and Hermisson 2006). Incomplete sweeps will occur when the variant on which selection is acting does not reach fixation, either because the population is sampled while the sweep is still ongoing, or because the beneficial effects of the allele are lost through changes in the adaptive landscape or frequency-dependent selection (Schrider *et al.* 2015). Examples of both 'hard' and 'soft' selective sweeps have been seen in a range of species (Daborn *et al.* 2001, Colosimo *et al.* 2005, Przeworski *et al.* 2005, Karasov *et al.* 2010).

If selection imposed by fisheries were to result in a 'hard' selective sweep, it is possible that variation linked to the causal mutation could be deleterious and/or result in changes in other traits. By examining the genetic changes produced in fish during selection experiments in more detail (such as those of van Wijk *et al.* (2013)), it is possible to increase our understanding of the nature of selection imposed on taxa in the wild and the ability of populations to recover once fishing pressure is reduced.

### 2.2.5 Identifying genes under selection

The two most commonly used approaches to identifying the genes under selection are differentiated by the amount of *a priori* information they require. Both approaches are essentially association studies which look for an association between a genetic locus and the phenotypic trait of interest. However, the candidate gene approach requires some *a priori* information about the trait and the genes which may potentially influence it (Piertney and Webster 2010). The candidate genes are then studied in populations or individuals with differing phenotypes for the trait in question to see if an association can be identified. These studies typically examine relatively few loci in a large number of individuals and have been used to identify loci involved in both simple (Nachman *et al.* 2003, Mundy 2005) and more complex (Tabor *et al.* 2002) traits. Genome-wide association (GWA) studies genotype a much larger number of loci spread across the genome in the hope of genotyping markers which are in linkage disequilibrium with the functional allele (Myles *et al.* 2009). As the GWA approach requires no *a priori* information about the mechanisms underpinning the trait it has been used in both model (Visscher *et al.* 2012, Harbison *et al.* 2013) and non-model species (Brachi *et al.* 2011, Campbell *et al.* 2014, Gutierrez *et al.* 2015b).

Advances in sequencing technology have expanded the reach of the GWA approach (Mardis 2008), particularly when combined with methods such as reduced representation sequencing which allow large numbers of loci spread across the genome to be identified and genotyped for relatively little cost (Davey *et al.* 2011). The initial guppy project (van Wijk 2011a) utilised a candidate gene approach to show that size selective harvesting can cause genetic change. In the study described here a GWA approach will be used. By utilising RAD sequencing to identify and genotype a large number of markers, signs of selection at loci across the guppy genome can be detected. Using the current reference sequence and predicted gene models for the guppy it will be possible to identify not only the loci showing signs of selection but also the genes these loci are located in or close to. It will therefore be helpful to consider potential candidate genes for growth and maturity based on current knowledge, in order to assess the proximity of these genes to the loci showing signs of selection.

### 2.2.6 Candidate genes for growth and maturity

In fish, the timing of maturation is controlled by the brain-pituitary-gonad axis (Schulz *et al.* 2010). However, it is widely accepted that growth, and therefore body size, in fish can be influenced by a range of factors such as water temperature (Pauly 1980, Dwyer and Piper 1987, Brander 1995, Pörtner *et al.* 2001), photoperiod (Gross *et al.* 1965, Björnsson *et al.* 1994, Boeuf and Le Bail 1999, Almazan-Rueda *et al.* 2005), salinity (Dendrinos and Thorpe 1985, Fielder and Bardsley 1999, Bœuf and Payan 2001, Denson *et al.* 2003), population density (Wallace *et al.* 1988, Bohlin *et al.* 2002, Lorenzen and Enberg 2002) and food availability (Jones 1986, Graeb *et al.* 2004, Orpwood *et al.* 2006, Andersen, Gerke, *et al.* 2012). The heritabilities observed for growth rates (Martyniuk *et al.* 2003, Perry *et al.* 2005, Kause *et al.* 2007, Dupont-Nivet *et al.* 2008), and the numerous QTLs identified across disparate taxa provide solid evidence for a genetic basis to variation in growth. Due to its importance in agriculture and aquaculture, the genetic architecture underlying growth rate has been studied in a rapidly increasing number of species (Nie *et al.*, Gross and Nilsson 1999, Cheng *et al.* 2000, Tambasco *et al.* 2003, Tao and Boulding 2003, Curi *et al.* 2005, Ma *et al.* 2011).

Possibly the most well-studied genes in relation to growth in fish are those of the somatotropic or growth axis. The key metabolic pathway is primarily composed of the growth hormone (GH), GH regulating factors such as hormones releasing and inhibiting GH (GHRH, GHIH and somatostatin) and insulin-like growth factors (IGFs) and their associated carriers, binding proteins and receptors. Due to its influence on metabolic processes and tissue growth, variation at any gene whose protein product is involved in this pathway can be considered a candidate gene. Indeed, many such genes have been associated with growth in many livestock species, (*GH* in cattle: Tambasco *et al.* 2003, pigs: Cheng *et al.* 2000 and chicken: Nie *et al.* ; *GHR* in cattle: Curi *et al.* 2005 and chicken: Huang *et al.* 1993 and *IGF-I* in cattle: Ge *et al.* 2003 and pigs: Machado *et al.* 2003), as well as fish (*gh* in the smooth tongued sole: Ma *et al.* 2011, Atlantic salmon: Gross & Nilsson 1999 and olive flounder: Kang *et al.* 2002; *igf* in the channel catfish: Peterson & Small 2005 and *ghrh* in arctic charr: Tao & Boulding 2003). Although not direct components of the somatotropic axis, genes such as *leptin*

(Heiman *et al.* 1998), *grelhin* (Kaiya *et al.* 2003, Unniappan and Peter 2005), *pcap* (Gómez-Requeni *et al.* 2012) and *pou1f1* (Parks and Brown 1999) have been linked to the regulation of *gh* and *igfs*.

The other main group of genes known to be linked to growth in teleosts are those belonging to the transforming growth factor (TGF) superfamily. One member of the family is myostatin (Mstn), which, in mammals, is a negative regulator of muscle mass (Joulia-Ekaza and Cabello 2007) and has been linked to increased muscle growth in mice, cattle (McPherron and Lee 1997), dogs (Mosher *et al.* 2007) and, more recently, gilthead seabream (Sánchez-Ramos *et al.* 2012). Although not part of the TGF superfamily, the binding protein follistatin has been found to inhibit myostatin and could therefore be a candidate for growth (Amthor *et al.* 2004). The TGF superfamily also consists of the myogenic regulatory factors, which are sometimes known as the *myod* gene complex. This group of transcriptional factors comprised of *myod*, *myf5*, *mrf4* and *myog,* regulates myogenesis and skeletal muscle growth (Atchley *et al.* 1994).

While numerous studies looking at the genetic architecture of growth have used a candidate gene approach based on prior knowledge of gene function to examine a small number of genes or transcripts, increased sequencing throughput has allowed the identification of new candidate genes for growth based on a wider study of up and down regulated transcripts. These include genes involved in cell cycle control and myoblast proliferation (*DRG1, CEBPD*), muscle fibre differentiation (*SYMD1, RTN1 and HSP90A*), protein degradation pathways (*MURF1, MAFBX and CSTL1*) (Bower and Johnston 2010), muscle structural proteins (*TNC, TNT2 and ACTIN2; FGF6*) (Bower and Johnston 2010), lipid and carbohydrate metabolism (*GAPDH, PFK, APOA1-like* paralogues, *ACBP, FADSD5* and *FADSD6*), nitrogen retention (*GS* and *GDH*), oxygen transport (*FTH, FTM* and *HB*) (Xu *et al.* 2013) and genes involved in the oxidative phosphorylation pathway (*NADH* dehydrogenase, cytochrome b and ATPase) (Salem *et al.* 2012). Additionally, genes such as fgf6 (Campos *et al.* 2013) and *PVALB* (Xu *et al.* 2006) have been suggested as candidate genes for growth but have not yet been widely studied. Several myosin genes have been linked to changes in growth, however the results from these genes are conflicting with studies within the same species which show down regulation during periods of muscle wastage (Salem *et al.* 2006) and periods of growth stimulated by *GH* treatment (Gahr *et al.* 2008).

In addition to body size, here we will consider the timing of maturation, which has been shown to be closely linked to growth rate (Shimada *et al.* 2011), and thus is important in aquaculture. However, despite its significance in aquaculture, the genetic architecture underpinning this trait has been much less well studied. As previously mentioned, the timing of maturation is controlled by the brain-pituitary-gonad axis (Schulz *et al.* 2010). The gonadotropin-releasing hormone stimulates the production of the two gonadotropins, follicle stimulating hormone and the luteinizing hormone which in turn regulate the development of the gonads. Although the genes underpinning this axis have not been widely studied, many candidate genes have been identified (Viñas and Piferrer 2008, Diopere

*et al.* 2013), and include those linked to the gonadotropin-releasing hormone (*GNRH1, GNRH2, GNRH3, GNRHR*) and those linked to the gonadotropins (*FSHB, LHB, FSHR, CYP19*).

### 2.2.7 Aim and objectives

The aim of this chapter is to examine the genetic architecture underpinning the shifts in life history traits resulting from an experiment in size selective harvesting in the guppy, *P.reticulata*. The objectives were to use RAD sequencing to simultaneously identify and genotype SNPs in the guppy, map these to the guppy genome (Fraser *et al.* 2014) and to identify regions showing signs of selection between lines selected for differences in body size. Examining the size, location and putative function of these regions will provide valuable information about the nature of the genetic changes taking place in the target species, with implications that can be extended to exploited populations of other fish species.

## 2.3 Methods

### 2.3.1 Selection experiment as described in van Wijk (2011)

The samples utilised in the RAD sequencing were obtained from a previously conducted selection experiment. Full details of the selection line generation can be found in van Wijk (2011) although for clarity, the essential features are reproduced below. As stated elsewhere, all sampling and experimental testing of fish was undertaken by previous researchers. The current study is focused on the RAD-sequencing, identification and dynamics of SNPs identified in *P. reticulata*.

#### 2.3.1.1 Initial sampling and rearing protocol of fish

A total of 180 *Poecilia reticulata* (90 male and 90 female) were sampled from the Lower Tacurigua River (10°38'49.5"N, 61°22'47.2"W) in Trinidad during March 2008 (figure 3.1). The fish were caught using a fine mesh net (4mm$^2$) and were separated by sex before being transported alive to Bangor University's aquarium facility. Experimental breeding began immediately after arrival in Bangor. During the experiment, fish were maintained in a controlled temperature environment with a 12:12 hour light cycle. Feeding was *ad libitum* each afternoon, primarily with live brine shrimp (*Artermia artemia*). Except where stated fish were kept in large tanks (105-120L) in a continuous flow-through system.

#### 2.3.1.2 Experimental selection

Fish were maintained for five generations of experimental selection in total, with two generations of random selection ($F_1$-$F_2$), followed by three generations of size selective harvesting ($F_3$-$F_5$). The two generations of random selection allowed an increase in population size to 1200 fry, and standardised the breeding environment experienced by $F_3$ fish aimed to minimise maternal effects. Five selection lines were established from the $F_3$: two selecting for large body size (L1 and L2), two selecting for small body size (S1 and S2), and one randomly selected control line (C) with each line consisting of 50 males and 75 (randomly assigned) females. The lines were initially set up with the

largest 20% of the mature $F_3$ fish being assigned to L1 or L2, the smallest 20% being assigned to S1 or S2 and a random 20% being assigned to C. Selection of the $F_4$ and $F_5$ then took place within each line, with the largest/smallest/random 20% of the fish from that line being used to make up the next generation.

Male guppies are known to have determinate growth, such that growth will almost cease once they reach reproductive maturation (Reznick and Endler 1982). Females, on the other hand, continue to grow throughout life. Consequently, selection experiments were conducted on adult male guppies only in order to avoid the confounding effects of variation in size with age between selected lines.

2.3.1.3 Phenotypic measurements

Measurements of standard length (SL) were taken after all fish were mature but before selection in each generation. Additional measurements were taken from most $F_6$ fish which were to be included in the RAD sequencing. SL measurements of each fish were obtained from photographs of fish following analysis in Adobe Photoshop®. As photographs were analysed by two different researchers, a random sample of 30 fish was used to assess repeatability. Each fish was photographed twice and each photograph was analysed by both researchers. Repeatability was then calculated by: $R = \sum_{1\text{-}N} (1-(\sigma_K^2/\sigma_t^2)) / N$, where $\sigma_K^2$ is the standard deviation between all repeated measures of the $K^{th}$ fish, $\sigma_t^2$ is the standard deviation over all measurements and N is the number of fish being measured (Lynch and Walsh 1998). Repeatability between SL measurements taken by the different researchers was high (R=0.973, st. dev. ± 0.033).

Age and size at maturation was measured for a randomly chosen subset of males from the $F_2$ (97 fish) and a randomly chosen subset from within each selection line (50 fish per line). Sexual maturation was determined by daily visual inspection of all fish. Male sexual maturity can be determined when the fleshy hood extends beyond the gonopodium tip (Houde 1997).

### 2.3.2 Restriction associated DNA sequencing

A total of 151 fish from the original van Wijk et al (2013) study were selected for sequencing: 40 from the $F_2$ generation before size selection commenced, 54 from the large selected lines (40 L1 and 14 L2) and 57 from the small selected lines (28 S1 and 29 S2). These fish were chosen to maximise the amount of sequence data which could be obtained within the available budget, and to ensure an equal spread of fish were sampled from both the large and small selection regime. Genetic samples from the $F_2$ fish had been obtained after all $F_2$ males had reached maturity. $F_6$ genetic samples were obtained following termination of the experiment. Due to the low levels of DNA and tissue available, no fish from the $F_3$ were included in the sequencing. Genomic DNA was extracted from tissue samples using the DNeasy blood and tissue kit (Qiagen). DNA was quantified on the Qubit fluorometer prior to library preparation.

Libraries were prepared as described in Poland *et al.* (2012) with addition of a size selection step following PCR amplification. As discussed earlier (Section 1.8), the RAD protocol involves three key steps: (1) a restriction digest, during which the DNA is digested with two different enzymes; (2) adaptor ligation, during which the forward adaptor (made up of the Illumina adaptor and the barcode) and the reverse Y adaptor are ligated onto the restricted DNA and (3) Multiplex and PCR amplification, during which the samples are pooled together and the sequences to which both adaptors have successfully ligated are PCR amplified. The addition of a size selection step in ddRAD protocols reduces the number of reads required to achieve high confidence in the genotyping of a SNP and increases the correlation in read coverage per site between individuals (Peterson *et al.* 2012). Restriction digestion was carried out using *PstI* and *MspI* and samples were identified with 151 unique barcodes of varying lengths (sequences for which can be found in appendix I). All samples were pooled into one library and, in order to ensure sufficient read coverage, the library was sequenced three times (one lane per library replicate) on an Illumina HiSeq 2000. All library preparation and sequencing was undertaken by collaborators in the WeigelWorld group at the Max Plank institute, Germany.

### 2.3.3 Raw data processing

Raw Illumina sequences were de-multiplexed using the "SHORE" pipeline (Ossowski *et al.* 2008) with subsequent quality checking using "FastQC". The de-multiplexed reads were then cleaned using the "*process_radtags*" module of the Stacks pipeline ("Stacks: *process_radtags*") (Catchen *et al.* 2011). Quality checking with "FastQC" identified one position in the first lane of sequencing with very low quality throughout the samples. As a consequence, raw reads from each lane were processed and cleaned separately. As well as removing reads containing uncalled bases and removing reads with an overall low quality score, "Stacks:*process_radtags*" employs a sliding window technique to check the quality of the read. The programme will calculate the quality score of a percentage of the read (known as a window). Once the first window has been checked it will 'slide' one bp along the read and perform the check again. By modifying the size of the sliding window (the fraction of the genome to be checked), reads from lane one with low quality could be successfully removed without the loss of any high quality data. For lane one, the length of the sliding window was set to 5-6% of the read (depending on the length of the sequence) while for lanes two and three the length of the sliding window was set to 10% of the genome. Where quality scores within this window dropped below a 99.8% probability of being correct reads were discarded.

### 2.3.4 Mapping to the reference genome

Once cleaned, the reads from each lane were combined for each individual and mapped to the *Poecilia reticulata* reference genome (genbank genome accession: GCF_000633615.1) using "CLC workbench" (CLC Bio-Qiagen, Aarhus, Denmark). Guppies have an X/Y sex determining system although a large section of the sex chromosome is thought to be pseudoautosomal (see section 2.5.3). Currently only a female genome sequence exists (Fraser *et al.* 2014) and therefore the

guppy genome assembly used will not contain any Y-specific sequences. Reads were mapped to the scaffolds instead of linkage groups in order to enable mapping of reads to unassembled scaffolds to be included in further analysis. The following parameters were used for mapping: mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 1, similarity fraction = 0.9 and random mapping of non-specific matches. The parameters set required the complete read to align at 90% similarity and allowed the location of reads mapping to multiple positions on the reference to be chosen randomly.

### 2.3.5 SNPs calling and inferring genotypes

The "Stacks" pipeline was used to call SNP and infer genotypes for individuals. The aligned reads were assembled into loci within individuals using "Stacks:*pstacks"* with a minimum of five reads required before a set of aligned reads could be classed as a locus (-m). The "Stacks:*pstacks"* module uses an error-bound maximum likelihood model to detect polymorphisms. This model estimates the sequencing error rate for each genotype at each nucleotide within a locus and then uses a likelihood ratio test to determine which genotype is most likely. For this step the upper limit for the estimated error rate (--bound_high) was set to 0.05. These RAD tag loci were then catalogued using "Stacks:*cstacks*" with a maximum of two nucleotide differences allowed between loci from different individuals (-n), before "Stacks:*sstacks*" was used to match individuals against the catalogue. In order to ensure high confidence in all SNPs, the error correction module "Stacks:*rxstacks*" was run. The software applies four corrections: (1) re-evaluate SNPs calls using information from across all individuals, (2) filter loci which have a log likelihood below a user set threshold, (3) remove confounded loci (multiple loci from within one individual which match to one catalogue loci) and (4) prune excess haplotypes from individual loci based on the frequency of haplotypes within the population. For our study, loci with a log likelihood of below -20 were removed. Loci containing more than three SNPs were not included in further analysis.

### 2.3.6 Initial analysis of selection lines

Once SNPs had been called in all individuals, "Stacks:*populations*" was run allowing the calculation of a range of population genetics statistics (including $\pi$, $F_{st}$, $F_{is}$, and observed and expected heterozygosity,). Individuals were assigned to 7 populations: L1, L2, S1, S2, $F_2$, Large (L1 and L2 combined) and Small (S1 and S2 combined), and all statistics were calculated for each population, generating eleven pairwise comparisons: Large vs. Small, Large vs. $F_2$, Small vs. $F_2$, L1 vs. S1, L1 vs. S2, L1 vs. $F_2$, L2 vs. S1, L2 vs. S2, L2 vs. $F_2$, S1 vs. $F_2$ and S2 vs. S2. In order to be included a SNP had to be present in at least 70% of the fish in both of the populations being compared.

### 2.3.7 Identifying SNPs under selection

A range of analyses were used to identify putative SNPs under selection. Initial observations were made by examining the $F_{st}$ values of pairwise comparisons. Outlier analyses were performed with two $F_{st}$ outlier based approaches, "Arlequin" (Excoffier *et al.* 2009, Excoffier and Lischer 2010) and

"BayeScan" (Foll and Gaggiotti 2008). By examining the relationship between $F_{st}$ and heterozygosity, the method implemented in "Arlequin" ('fdist method') (Beaumont and Nichols 1996) simulates a neutral $F_{st}$ distribution and identifies outliers by comparing the observed $F_{st}$ values to those expected under neutrality. Here the 'fdist method' was implemented in "Arlequin" (v3.5.1.3) using 100,000 simulations and 100 demes. A SNP was considered an outlier if the observed $F_{st}$ differed significantly ($p \leq 0.05$) from the $F_{st}$ simulated under neutrality. The second outlier analysis decomposes the genetic differentiation into a population-specific and a locus-specific component and compares the contribution of the two coefficients with the observed pattern of diversity. By defining two models, one in which the locus-specific component is required to explain the diversity, and one where it is not, the posterior probability of each can be calculated, indicating whether the marker is under selection. Here, this method is implemented in "BayeScan" with the default parameters (20 pilot runs of 5,000 iterations and an additional burn in of 50,000 iterations followed by 100,000 iterations). "BayeScan" also calculates a q-value for each locus which is defined as the minimum FDR at which a locus may become significant. Any locus with a q-value ≤ 0.05 was identified as an outlier.

Both outlier analyses were performed on 8 pairwise comparisons (L1 vs. S1; L1 vs. S2; L1 vs. $F_2$; L2 vs. S1; L2 vs. S2; L2 vs. $F_2$; S1 vs. $F_2$ and S2 vs. S2) and a SNP was considered under selection if it was an outlier in either: (i) at least 1 of the 8 pairwise comparisons in the "BayeScan" analysis and at least 4 of the "Arlequin" comparisons or (ii) at least 6 of the 8 "Arlequin" comparisons. Each of the 4 selection lines was included in 3 comparisons (e.g. L1 vs. S1; L1 vs. S2; L1 vs. $F_2$). Requiring any SNP determined to be under selection by "BayeScan" to also be under selection in at least 4 of the "Arlequin" meant that this SNP must be showing signs of selection in comparisons with at least 2 of the selection lines. Similarly, by requiring any SNP only identified as an outlier by "Arlequin" to be an outlier in at least 6 of the 8 comparisons meant that this SNP would be an showing signs of selection in either all comparisons involving two different lines, or in comparisons involving more than two lines.

Although $F_{st}$ outlier analysis is able to detect SNPs showing a significant degree of divergence between two populations being compared, a SNP classified as under selection, as described above, may not be showing concordant patterns of allele frequency change across the selection line replicates. We therefore used a recently described technique (Uusi-Heikkilä *et al.* 2015) to assess allele frequency change across the generations and between the fish before and after selection. Allele frequency within each line (including the $F_2$) was bootstrapped and the resulting values used to calculate 95% confidence intervals. To do this 70% of the individuals from each population were randomly sampled and used to calculate the allele frequency for each SNP. The random sampling was repeated 100 times before the $2.5^{th}$ and $97.5^{th}$ percentile of the 100 calculated allele frequencies was calculated and used as the 95% allele frequency CIs. Although for some SNPs the 95% CI between replicates did not show an overlap, only those SNPs in which the CI values overlapped between large and small lines were removed from further analysis.

### 2.3.8 Identifying the genes under selection

The NCBI annotation pipeline was run on the available unpublished guppy genome (genbank genome accession: GCF_000633615.1). This pipeline takes the reference genome and uses any other species data available (such as expressed sequence tags, transcriptome data and proteins), and in some cases data from closely related species, to predict gene models and protein products from a reference genome (Thibaud-Nissen *et al.* 2013). The location of each SNP showing signs of selection was ascertained and the gene in which it was located, or its nearest gene, and the predicted protein product(s), identified. Using this information, a literature search and relevant databases (NCBI gene database, UniProt knowledgebase) were used to identify putative gene functions.

Since many SNPs showing signs of selection were located across the sex chromosome (see figure 2.3), and as it is likely that regions of this chromosome are linked, all predicted genes and their protein products from this chromosome were examined. Blast2Go searched gene ontology terms for predicted proteins. Genes which have been linked to body size in fish were identified with a literature search. If these were found in the annotated guppy genome, their locations and proximity to the RAD sequenced SNPs were determined.

## 2.4 Results

### 2.4.1 Phenotypic response to selection (van Wijk, *et al*., 2013)

As previously mentioned all sampling and the selection experiment was carried out by previous researchers and a full statistical analysis of the phenotypic shifts observed can be found in van Wijk (2011). For clarity an overview of the phenotypic changes observed follows.

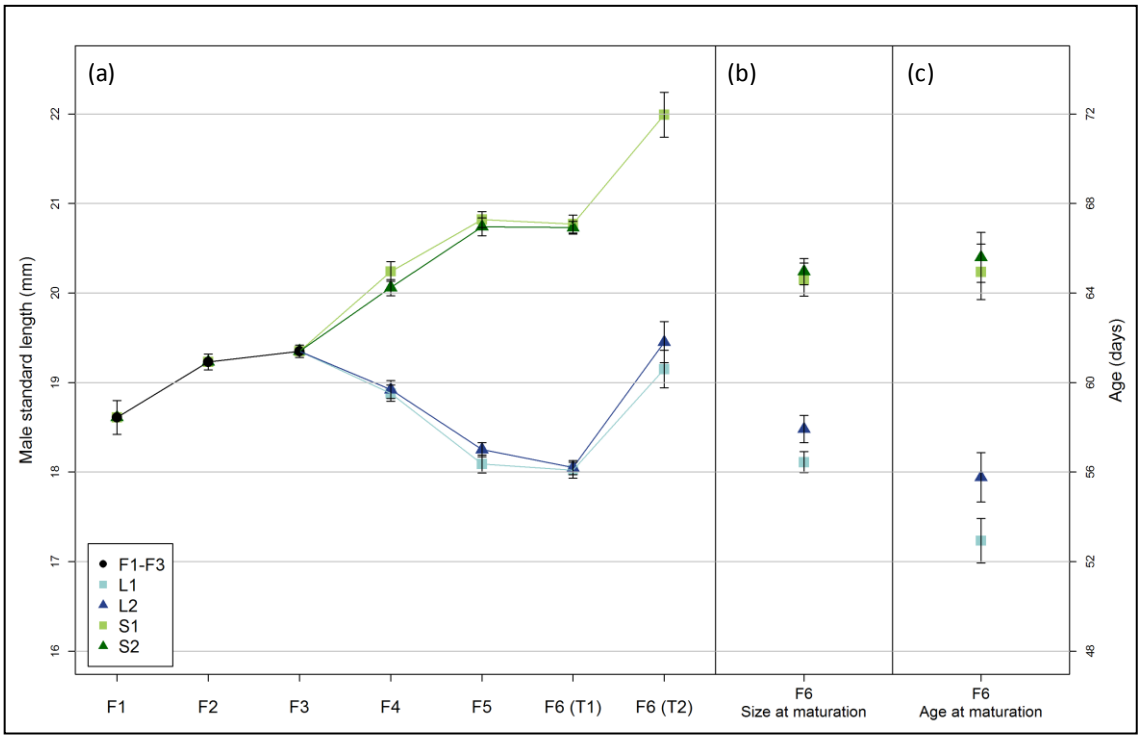



**Figure 2.1:** Phenotypic response to size selective harvesting. (a) change in male standard length in random breeding generations ($F_1$-$F_3$) and in generations selected for size ($F_4$-$F_6$), (b) male size at maturation in the four selected lines in the $F_6$ generation and (c) male age at maturation in the four selected lines in the $F_6$ generation. Blue triangles and squares represent the large selected lines and green triangles and squares represent the small selected lines.

Although male SL significantly increased in the $F_1$ generation, there was no observed change between the $F_1$-$F_3$. Between the $F_3$-$F_6$ male, SL increased from 19.30mm to 20.75mm in the large lines and decreased to 18.05mm in the small lines (figure 2.1. a). No significant difference was observed between the replicates, and no significant change in SL occurred in the C line between the $F_3$-$F_6$. Both age and size at maturation also showed significant differences between the selection regimes in the $F_6$ (age at maturation P=0.038 and SL at maturation P= 0.022), although no significant difference was observed between the C line and the small lines (figure 2.1, b). Female SL showed no response to selection and showed no significant difference between selection regimes in the $F_6$.

Where possible, $F_6$ fish used in the RAD sequencing were measured again immediately prior to sequencing, however measurements were obtained only for a total of 72 of the 111 $F_6$ fish that were RAD sequenced (23 S1; 20 S2; 27 L1; 2 L2). Due to low sample size in L2 and uncertainty surrounding the age of these fish, size measurements from this line were not included in further analyses. When the size measurements of the $F_6$ fish used by van Wijk (2011) were taken the fish were approximately 50-70 days old (time point 1), however, the measurements taken from the fish used in the RAD sequencing were taken when the fish were approximately 18 months old (time point 2). The age difference between the two time points meant that male SL was 1.13-1.4mm (5.8-7.7%; figure 2.1, a) larger at time point 2, than it was at time point 1. The increase in SL was similar

across the selection lines (t= -0.816, df=1, p=0.538) and the difference in SL between the lines at time point 2 was still significant (L1 vs. S1 U=34 p ≤ 0.001; L1 vs. S2 U= 36.5 p ≤ 0.001).

### 2.4.2 Restriction associated DNA sequencing

The three lanes of sequencing produced a total of 455,606,946 reads from 151 individuals. Three individuals with less than 100,000 reads each from the three lanes combined were removed from further analysis. Initial processing and cleaning removed 62,156,281 reads leaving a total of 393,345,137 which were then mapped to the guppy reference genome. 73.76% (290,125,694) of the reads were successfully mapped to the genome.

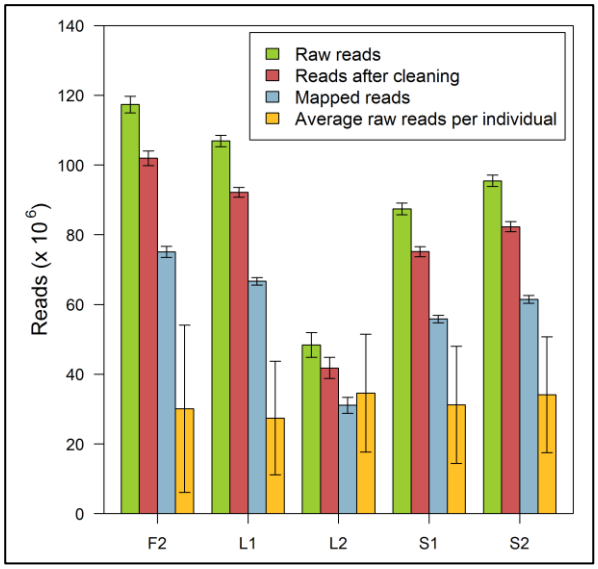



**Figure 2.2:** Number of raw reads, reads remaining after cleaning and reads successfully mapped to the reference genome.

After removal of individuals with low read counts, the numbers of individuals sequenced in each line were as follows: 39 $F_2$; 39 L1; 14 L2; 28 S1 and 28 S2. Although significant variation in the number of reads sequenced was observed in each of the lines (figure 2.2), such variance arose from variation in the number of individuals sequenced rather than DNA quality or quantity.

**Table 2.1:** The number of SNPs showing an elevated Fst (≥0.05) and the number of SNPs identified as an outlier in each of the pairwise comparisons.

| | Fst ≥ 0.05 | Arlequin outlier | BayeScan outlier |
|---|---|---|---|
| L1 / S1 | 2655 | 2200 | 17 |
| L1 / S2 | 2591 | 2323 | 35 |
| L2 / S1 | 3258 | 3697 | 7 |
| L2 / S2 | 3124 | 4774 | 3 |
| L1 / F2 | 802 | 3304 | 23 |
| L2 / F2 | 1214 | 2419 | 0 |
| S1 / F2 | 1056 | 1424 | 1 |
| S2 / F2 | 996 | 1625 | 2 |

A total of 339,338 putative RAD loci were catalogued which had a mean depth per tag and per individual of 6.29. Of these, a total of 193,741 loci contained between 1 and 3 SNPs. The frequency of tags across the genome can be seen in figure 2.3. Applying the threshold that loci had to be present in at least 70% of individuals in both populations being compared, the number of SNPs analysed in the 11 pairwise comparisons ranged from 18,365 to 31,310. The mean number of SNPs in the pairwise comparisons used in the outlier analysis was 23,196.

### 2.4.3 Identifying SNPs under selection

#### 2.4.3.1 $F_{st}$ analysis

Although global $F_{st}$ values between the lines were low (table 2.1), 7,624 SNPs showed elevated $F_{st}$ values (≥ 0.05) in at least one of the pairwise comparisons. When the samples from the two large

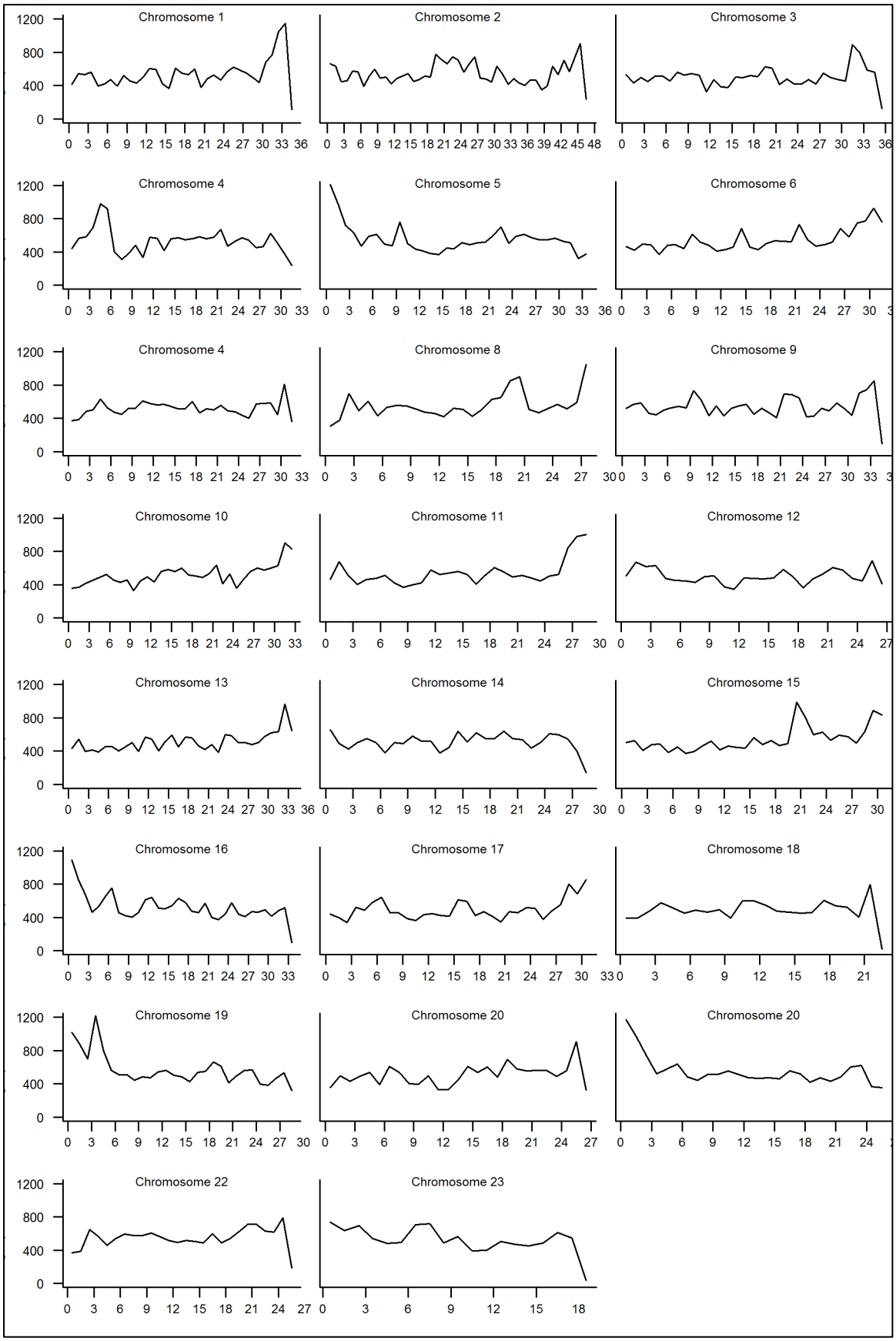



**Figure 2.3:** Plots showing the number of RAD tags identified per MB across the guppy genome.

and two small selection lines were pooled together into L and S pools, 1591 SNPs showed a significantly elevated $F_{st}$ ($p \leq 0.05$) in at least one of the pairwise comparisons (L vs S, L vs. $F_2$ and S vs. S2), 13% of which were located on the sex chromosome (chromosome 12) (figure 2.4). The $F_{st}$ values observed show the significant genetic divergence which exists between the large and small selection lines, particularly on the sex chromosome.

It has been suggested that elevated $F_{st}$ values in regions of reduced recombination (such as sex chromosomes) may derive from the reduced levels of diversity often found in these regions due to sensitivity of the $F_{st}$ statistic to variance in heterozygosity (Cruickshank and Hahn 2014). To ensure that the elevated $F_{st}$ values observed were not an artefact of reduced diversity on the sex chromosome, we compared the observed heterozygosity of a subset of autosomal SNPs to SNPs located on the sex chromosome. Comparisons of 5 random subsets for each population (25 in total) found a significant difference in diversity between the sex chromosome and the autosomes only twice (out of 25) (Appendix II), indicating that the elevated $F_{st}$ values found on chromosome 12 were not an artefact of reduced diversity.

2.4.3.2 Fdist method

Across the 8 comparisons (L1 vs S1, L1 vs S2, L1 vs $F_2$, L2 vs S1, L2 vs S2, L2 vs $F_2$, S1 vs $F_2$ and S2 vs $F_2$) the mean number of outliers identified using the 'fdist method' as implemented in "Arlequin" was 2720 (SD ± 1128), (table 2.1). A total of 28 outliers were present in at least 6 of the

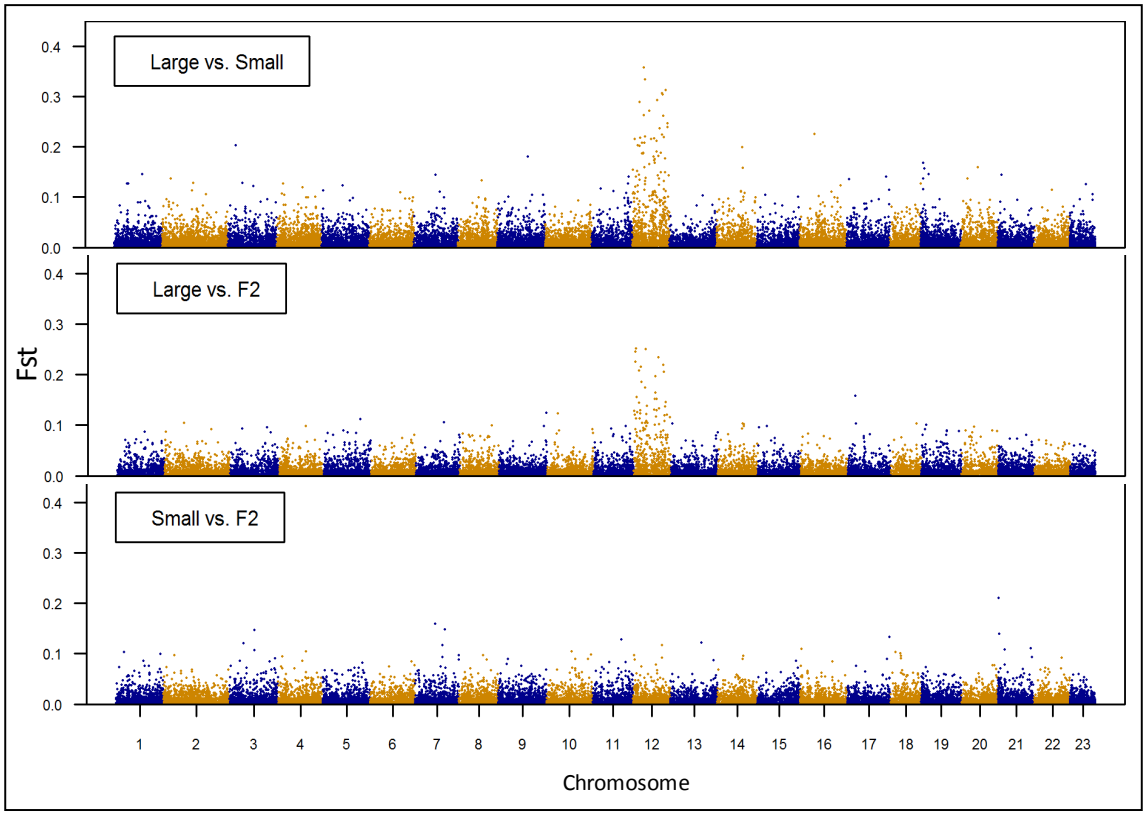



**Figure 2.4:** Fst values across the genome from three pairwise comparisons.

pairwise comparisons, 19 of which were located on the sex chromosome. "Arlequin" has previously been shown to exhibit high type I error rates (Narum and Hess 2011) and so, by requiring a SNP to be an outlier in 6 of the 8 comparisons, we reduced the false positive rate and erroneously identifying a SNP as being under selection.

2.4.3.3 Bayesian method

The Bayesian method as implemented in "BayeScan" identified on average 2700 fewer outliers than 'fdist method' as implemented in "Arlequin" (table 2.1), though as previously shown (Pérez-Figueroa *et al.* 2010a, Narum and Hess 2011, Vilas *et al.* 2012), "BayeScan" performs better for detecting SNPs under selection. It is thus likely that a large number of putative outlier SNPs identified by "Arlequin" are false positives. Despite this, we only considered a SNP identified by "BayeScan" to be showing signs of selection if it also exhibited signs of selection in 4 of the 8 pairwise comparisons analysed in "Arlequin".

The number of SNPs identified as outliers in the 8 pairwise comparisons varied from 0 to 35 ($\bar{x}$ = 11, SD ± 12.75). Of the 88 SNPs which were identified as an outlier in at least one pairwise comparison in "BayeScan", 36 were also outliers in at least 4 of the pairwise comparisons in "Arlequin". Of these, 31 were located on the sex chromosome (chromosome 12).

2.4.3.4 Outlier analysis combined

As previously described, a SNP was considered under selection for body size or maturation if it was identified as an outlier in either (i) at least 6 pairwise comparisons analysed in "Arlequin" or (ii) 1 of the pairwise comparisons in "BayeScan" and 4 of the pairwise comparisons in "Arlequin". Under these criteria, a total of 53 SNPs were considered to be under selection (table 2.2). Of these SNPs 17% (9 SNPs) are located on autosomes, 75% (40 SNPs) on the sex chromosome and 7% (4 SNPs) on unassembled scaffolds (figure 2.5). A Fisher's exact test to assess the difference in allele frequency between the $F_2$ and the small and large replicates combined, showed that of the 37 SNPs showing consistent signs of selection, allele frequency had changed significantly only in the large lines in 29 SNPs, while 4 had only changed significantly in only the small lines (table 2.3).

### 2.4.4 Allele frequency change

In order to assess the direction and consistency of change in the SNPs under selection the allele frequency of each SNP within each line was bootstrapped and used to calculate 95% confidence intervals. The results show that for 37 of the 53 SNPs under selection, the direction of allele frequency change was consistent between replicates, with no overlap in CI between large or small

**Table 2.2:** Fst values and the number of times each SNP was identified as an outlier for each of the 53 SNPs identified as being under selection. Comparisons with an Fst ≥ 0.05 are highlighted. Un denotes a SNP located on an unassembled scaffold. * denotes a SNP which was identified as showing inconsistent patterns of allele frequency change in the bootstrap analysis. N/A denotes a SNP which was not present in enough individuals to be included in that pairwise comparison.

| SNP name | Chromosome | Fst values | | | | | | | | Outlier analysis | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | L1 / S1 | L1 / S2 | L1 / F2 | L2 / S1 | L2 / S2 | L2 / F2 | S1 / F2 | S2 / F2 | Arelquin | Bayescan |
| 272384_31* | 2 | 0.144 | 0.105 | 0.062 | 0.073 | 0.015 | 0.000 | 0.041 | 0.010 | 7 | 0 |
| 255284_22 | 3 | 0.121 | 0.101 | 0.000 | 0.168 | 0.146 | 0.004 | 0.136 | 0.114 | 6 | 0 |
| 122944_80* | 7 | 0.025 | 0.077 | 0.030 | 0.021 | 0.100 | 0.030 | 0.097 | 0.168 | 4 | 1 |
| 63328_58 | 11 | 0.138 | 0.101 | 0.078 | 0.218 | 0.135 | 0.085 | 0.030 | 0.005 | 6 | 0 |
| 63292_77 * | 11 | 0.020 | 0.092 | 0.002 | 0.116 | 0.225 | 0.058 | 0.009 | 0.067 | 6 | 0 |
| 142285_30* | 11 | 0.315 | 0.048 | 0.059 | 0.149 | 0.002 | 0.004 | 0.114 | 0.001 | 4 | 1 |
| 23724_34 | 12 | 0.336 | 0.437 | 0.242 | 0.160 | 0.296 | 0.059 | 0.021 | 0.076 | 7 | 3 |
| 23986_62 | 12 | 0.229 | 0.151 | 0.097 | 0.297 | 0.179 | 0.103 | 0.044 | 0.009 | 7 | 0 |
| 58352_33 | 12 | 0.303 | 0.061 | 0.075 | 0.447 | 0.090 | 0.100 | 0.114 | 0.001 | 6 | 2 |
| 23547_69 | 12 | 0.235 | 0.124 | 0.099 | 0.293 | 0.128 | 0.094 | 0.052 | 0.002 | 6 | 0 |
| 59179_32 | 12 | 0.224 | 0.207 | 0.053 | 0.123 | 0.109 | 0.010 | 0.066 | 0.057 | 6 | 0 |
| 108125_34 | 12 | 0.284 | 0.333 | 0.279 | 0.238 | 0.299 | 0.214 | 0.002 | 0.017 | 6 | 3 |
| 108291_87 | 12 | 0.279 | 0.318 | 0.256 | 0.236 | 0.293 | 0.179 | 0.006 | 0.023 | 6 | 3 |
| 22272_28 | 12 | 0.311 | 0.264 | 0.249 | 0.296 | 0.229 | 0.188 | 0.013 | 0.002 | 6 | 3 |
| 59448_87 | 12 | 0.280 | 0.325 | 0.308 | 0.206 | 0.263 | 0.223 | 0.000 | 0.008 | 6 | 3 |
| 58232_79 | 12 | 0.294 | 0.267 | 0.271 | 0.095 | 0.075 | 0.072 | 0.001 | 0.000 | 6 | 2 |
| 58111_52 | 12 | 0.263 | 0.117 | 0.172 | 0.266 | 0.085 | 0.123 | 0.023 | 0.005 | 6 | 1 |
| 60122_83 | 12 | 0.188 | 0.286 | 0.099 | 0.148 | 0.272 | 0.050 | 0.022 | 0.077 | 6 | 1 |
| 21765_69 | 12 | 0.157 | 0.229 | 0.122 | 0.157 | 0.242 | 0.109 | 0.004 | 0.025 | 6 | 0 |
| 24329_40 | 12 | 0.110 | 0.092 | 0.056 | 0.245 | 0.215 | 0.140 | 0.011 | 0.005 | 6 | 0 |
| 24329_73 | 12 | 0.110 | 0.092 | 0.056 | 0.245 | 0.215 | 0.140 | 0.011 | 0.005 | 6 | 0 |
| 59480_84 | 12 | 0.118 | 0.076 | 0.115 | 0.126 | 0.083 | 0.102 | 0.000 | 0.006 | 6 | 0 |
| 59832_25 | 12 | 0.137 | 0.231 | 0.085 | 0.118 | 0.222 | 0.062 | 0.008 | 0.048 | 6 | 0 |
| 59969_19 | 12 | 0.091 | 0.166 | 0.067 | 0.115 | 0.173 | 0.084 | 0.002 | 0.025 | 6 | 0 |
| 21588_42 | 12 | 0.113 | 0.216 | 0.157 | 0.033 | 0.093 | 0.057 | 0.009 | 0.004 | 6 | 1 |
| 22486_62 | 12 | 0.163 | 0.331 | 0.247 | 0.068 | 0.252 | 0.136 | 0.013 | 0.023 | 5 | 2 |
| 23814_45 | 12 | 0.266 | 0.387 | 0.318 | 0.078 | 0.160 | 0.104 | 0.003 | 0.005 | 5 | 2 |
| 108025_75 | 12 | 0.207 | 0.269 | 0.122 | 0.115 | 0.178 | 0.036 | 0.020 | 0.048 | 5 | 1 |
| 20087_46 * | 12 | 0.085 | 0.034 | 0.238 | 0.023 | 0.002 | 0.108 | 0.054 | 0.112 | 5 | 1 |
| 21588_40 * | 12 | 0.113 | 0.216 | 0.235 | 0.016 | 0.067 | 0.080 | 0.039 | 0.002 | 5 | 1 |
| 24163_79 | 12 | 0.192 | 0.323 | 0.155 | 0.076 | 0.185 | 0.047 | 0.003 | 0.042 | 5 | 1 |
| 338159_30* | 12 | 0.306 | 0.100 | 0.075 | 0.116 | 0.005 | 0.001 | 0.089 | 0.002 | 5 | 1 |
| 338323_59* | 12 | 0.200 | 0.426 | 0.205 | 0.029 | 0.157 | 0.032 | 0.000 | 0.051 | 4 | 2 |
| 20521_62 | 12 | 0.137 | 0.171 | 0.198 | 0.039 | 0.058 | 0.068 | 0.006 | 0.001 | 4 | 1 |
| 23539_11 | 12 | 0.278 | 0.283 | 0.175 | 0.051 | 0.053 | 0.012 | 0.014 | 0.017 | 4 | 1 |
| 21655_33 * | 12 | 0.061 | 0.354 | 0.155 | 0.009 | 0.192 | 0.051 | 0.025 | 0.050 | 4 | 1 |
| 58137_23 * | 12 | 0.215 | 0.091 | 0.208 | 0.068 | 0.010 | 0.058 | 0.000 | 0.029 | 4 | 1 |
| 58581_51 * | 12 | 0.407 | 0.149 | 0.234 | 0.119 | 0.003 | 0.023 | 0.034 | 0.012 | 4 | 2 |
| 24213_82 | 12 | 0.390 | 0.335 | N/A | 0.232 | 0.155 | N/A | N/A | N/A | 4 | 2 |
| 337414_45 | 12 | 0.400 | 0.361 | N/A | 0.170 | 0.128 | N/A | N/A | N/A | 4 | 2 |
| 20087_64 * | 12 | 0.115 | 0.151 | 0.280 | 0.028 | 0.044 | 0.118 | 0.047 | 0.028 | 4 | 1 |
| 58413_79 | 12 | 0.140 | 0.282 | 0.152 | 0.044 | 0.137 | 0.048 | 0.000 | 0.025 | 4 | 1 |
| 58592_35* | 12 | 0.026 | 0.348 | 0.082 | 0.001 | 0.202 | 0.006 | 0.016 | 0.120 | 4 | 1 |
| 106700_83 | 12 | 0.212 | 0.381 | N/A | 0.095 | 0.224 | N/A | N/A | N/A | 4 | 1 |
| 57318_25 | 12 | 0.210 | 0.121 | N/A | 0.422 | 0.274 | N/A | N/A | N/A | 4 | 1 |
| 24234_71 * | 12 | 0.214 | 0.153 | 0.207 | 0.042 | 0.014 | 0.033 | 0.000 | 0.005 | 4 | 1 |
| 202503_31 | 14 | 0.174 | 0.258 | 0.044 | 0.096 | 0.181 | 0.005 | 0.053 | 0.112 | 5 | 1 |
| 97286_34 | 17 | 0.093 | 0.105 | 0.008 | 0.133 | 0.143 | 0.001 | 0.149 | 0.163 | 6 | 0 |
| 278569_65 | 23 | 0.220 | 0.078 | 0.064 | 0.101 | 0.018 | 0.012 | 0.058 | 0.000 | 6 | 0 |
| 277772_45* | Un | 0.071 | 0.077 | 0.081 | 0.031 | 0.033 | 0.034 | N/A | N/A | 6 | 0 |
| 160333_89 | Un | 0.156 | 0.114 | 0.008 | 0.100 | 0.071 | 0.005 | 0.127 | 0.086 | 6 | 0 |
| 139122_7 | Un | 0.196 | 0.229 | 0.163 | 0.313 | 0.369 | 0.230 | 0.010 | 0.027 | 7 | 2 |
| 155586_53* | Un | 0.024 | 0.108 | 0.118 | 0.014 | 0.064 | 0.069 | 0.057 | 0.001 | 5 | 1 |

**Table 2.3:** Results of fishers exact test comparing the allele frequencies of F2 with the large and small selected lines.

| Locus | Line | P value |
|---|---|---|
| 106700_83 | Large | 0.000 |
| | Small | 0.165 |
| 108025_75 | Large | 0.000 |
| | Small | 0.020 |
| 108125_34 | Large | 0.000 |
| | Small | 0.291 |
| 108291_87 | Large | 0.000 |
| | Small | 0.147 |
| 139122_7 | Large | 0.000 |
| | Small | 0.073 |
| 160333_89 | Large | 0.415 |
| | Small | 0.000 |
| 202503_31 | Large | 0.051 |
| | Small | 0.000 |
| 20521_62 | Large | 0.000 |
| | Small | 0.517 |
| 21588_42 | Large | 0.000 |
| | Small | 1.000 |
| 21765_69 | Large | 0.000 |
| | Small | 0.181 |
| 22272_28 | Large | 0.000 |
| | Small | 0.388 |
| 22486_62 | Large | 0.000 |
| | Small | 1.000 |
| 23539_11 | Large | 0.000 |
| | Small | 0.179 |
| 23547_69 | Large | 0.000 |
| | Small | 0.150 |
| 23724_34 | Large | 0.000 |
| | Small | 0.006 |
| 23814_45 | Large | 0.000 |
| | Small | 1.000 |
| 23986_62 | Large | 0.000 |
| | Small | 0.064 |
| 24163_79 | Large | 0.000 |
| | Small | 0.122 |
| 24213_82 | Large | 0.000 |
| | Small | 0.165 |

| Locus | Line | P value |
|---|---|---|
| 24329_40 | Large | 0.000 |
| | Small | 0.247 |
| 24329_73 | Large | 0.000 |
| | Small | 0.245 |
| 255284_22 | Large | 1.000 |
| | Small | 0.000 |
| 278569_65 | Large | 0.007 |
| | Small | 0.132 |
| 337414_45 | Large | 0.000 |
| | Small | 0.064 |
| 57318_25 | Large | 0.000 |
| | Small | 0.099 |
| 58111_52 | Large | 0.000 |
| | Small | 0.832 |
| 58232_79 | Large | 0.000 |
| | Small | 0.858 |
| 58352_33 | Large | 0.000 |
| | Small | 0.119 |
| 58413_79 | Large | 0.000 |
| | Small | 0.402 |
| 59179_32 | Large | 0.019 |
| | Small | 0.003 |
| 59448_87 | Large | 0.000 |
| | Small | 1.000 |
| 59480_84 | Large | 0.000 |
| | Small | 0.628 |
| 59832_25 | Large | 0.000 |
| | Small | 0.085 |
| 59969_19 | Large | 0.001 |
| | Small | 0.281 |
| 60122_83 | Large | 0.001 |
| | Small | 0.007 |
| 63328_58 | Large | 0.000 |
| | Small | 0.127 |
| 97286_34 | Large | 0.648 |
| | Small | 0.000 |

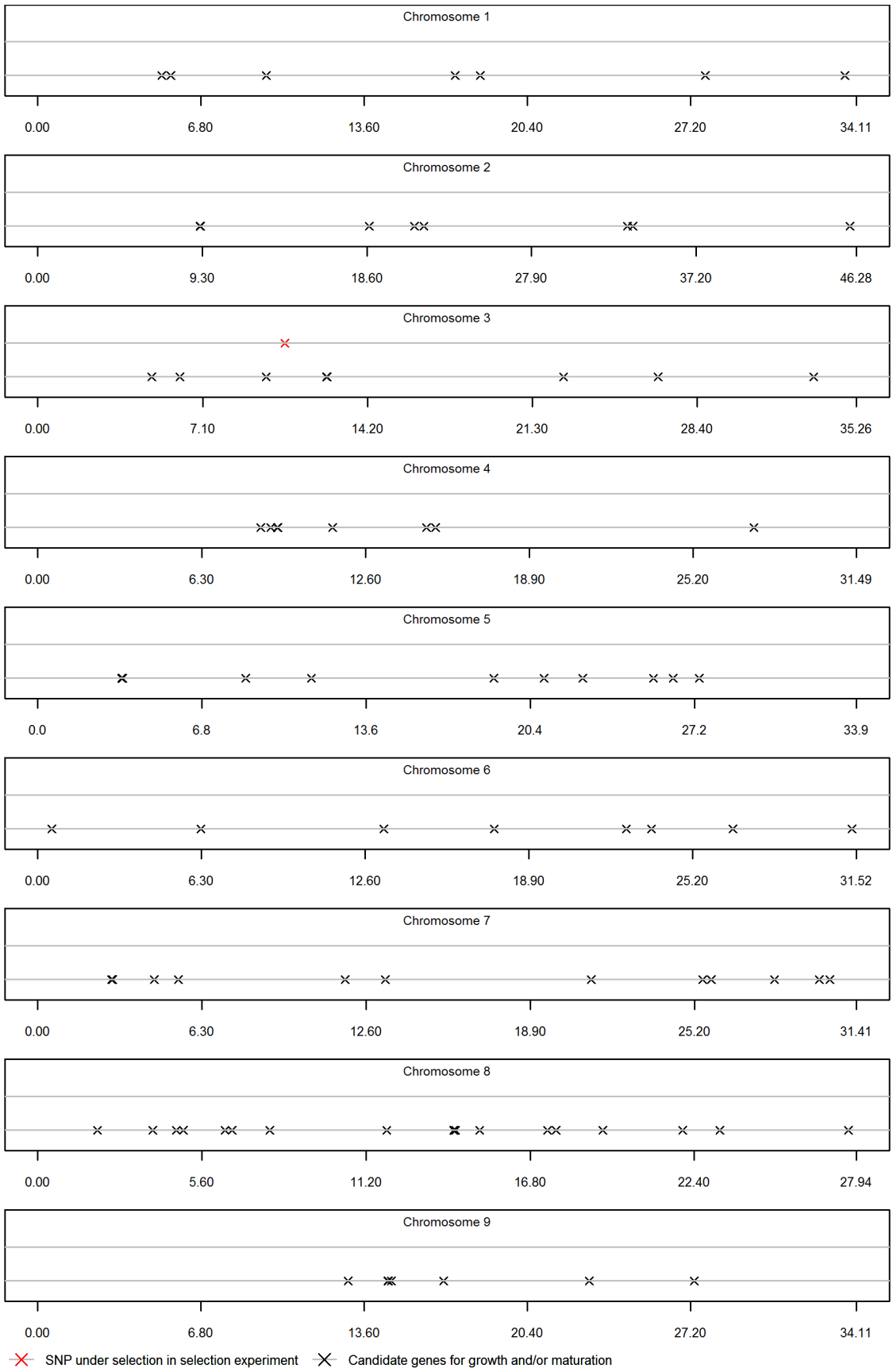



**Figure 2.5:** Location of SNPs showing signs of selection following experimental selection and candidate genes for growth and/or maturation across the guppy genome. For candidate genes a black cross marks the start point of the gene.

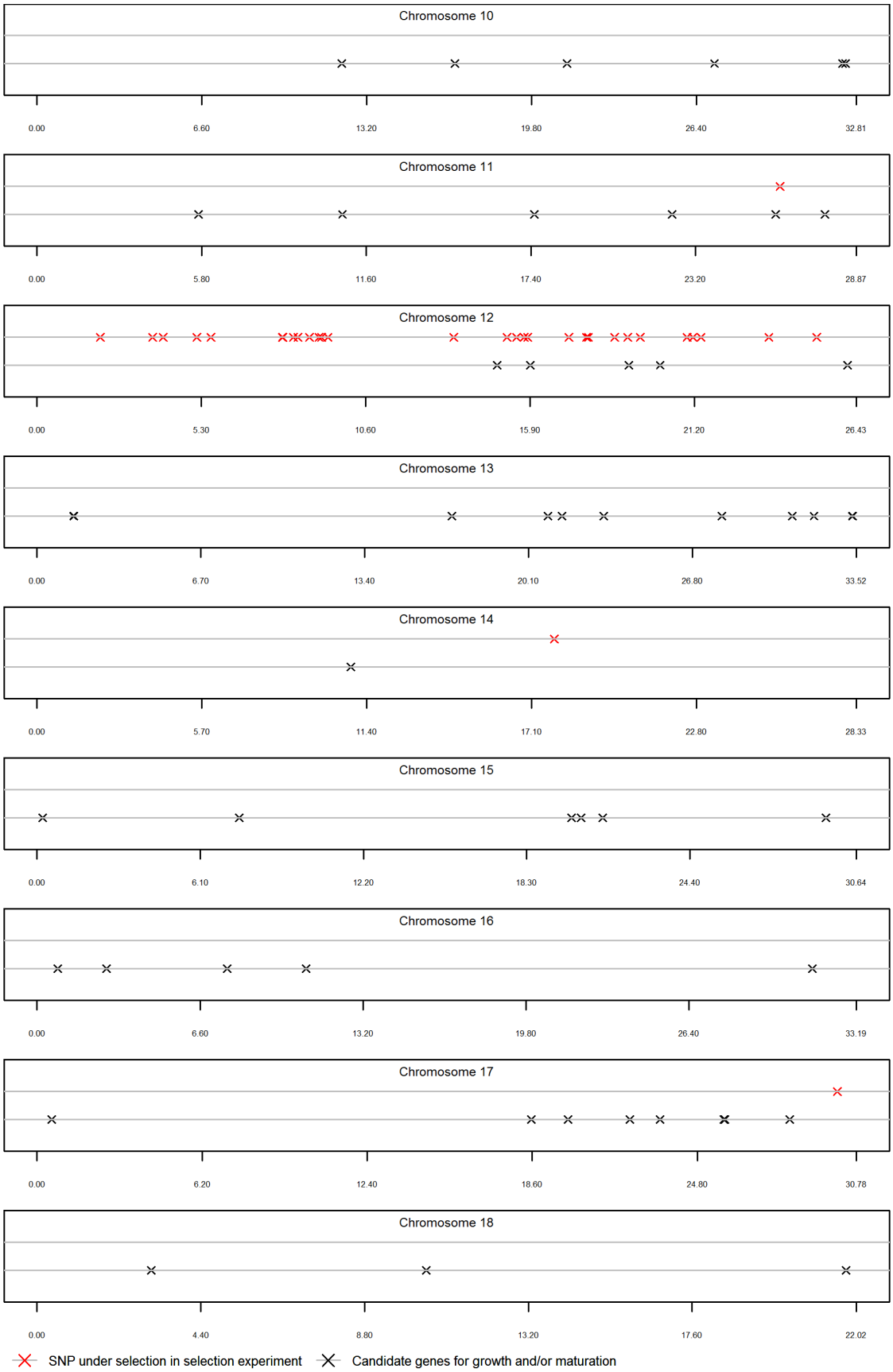



**Figure 2.5:** Continued

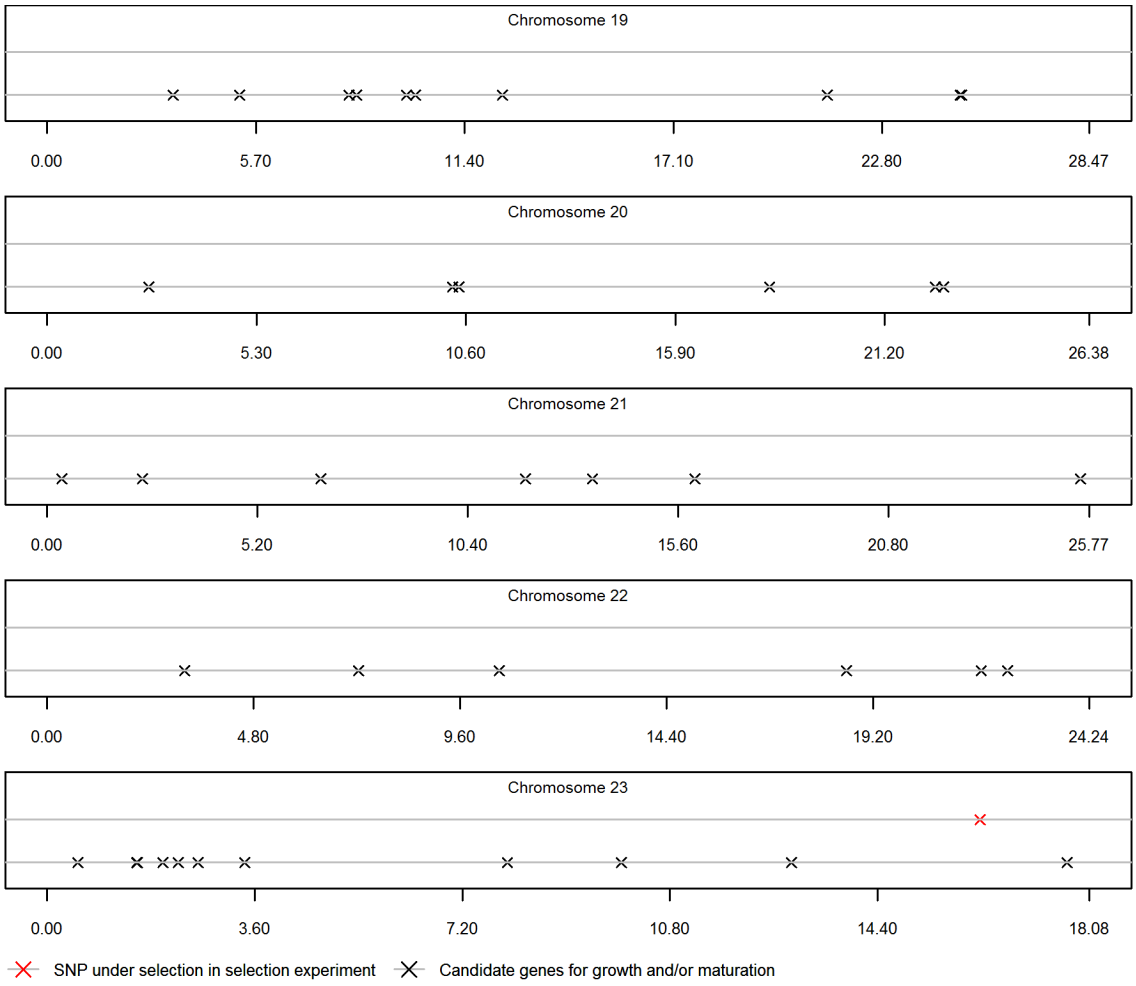



**Figure 2.5:** Continued

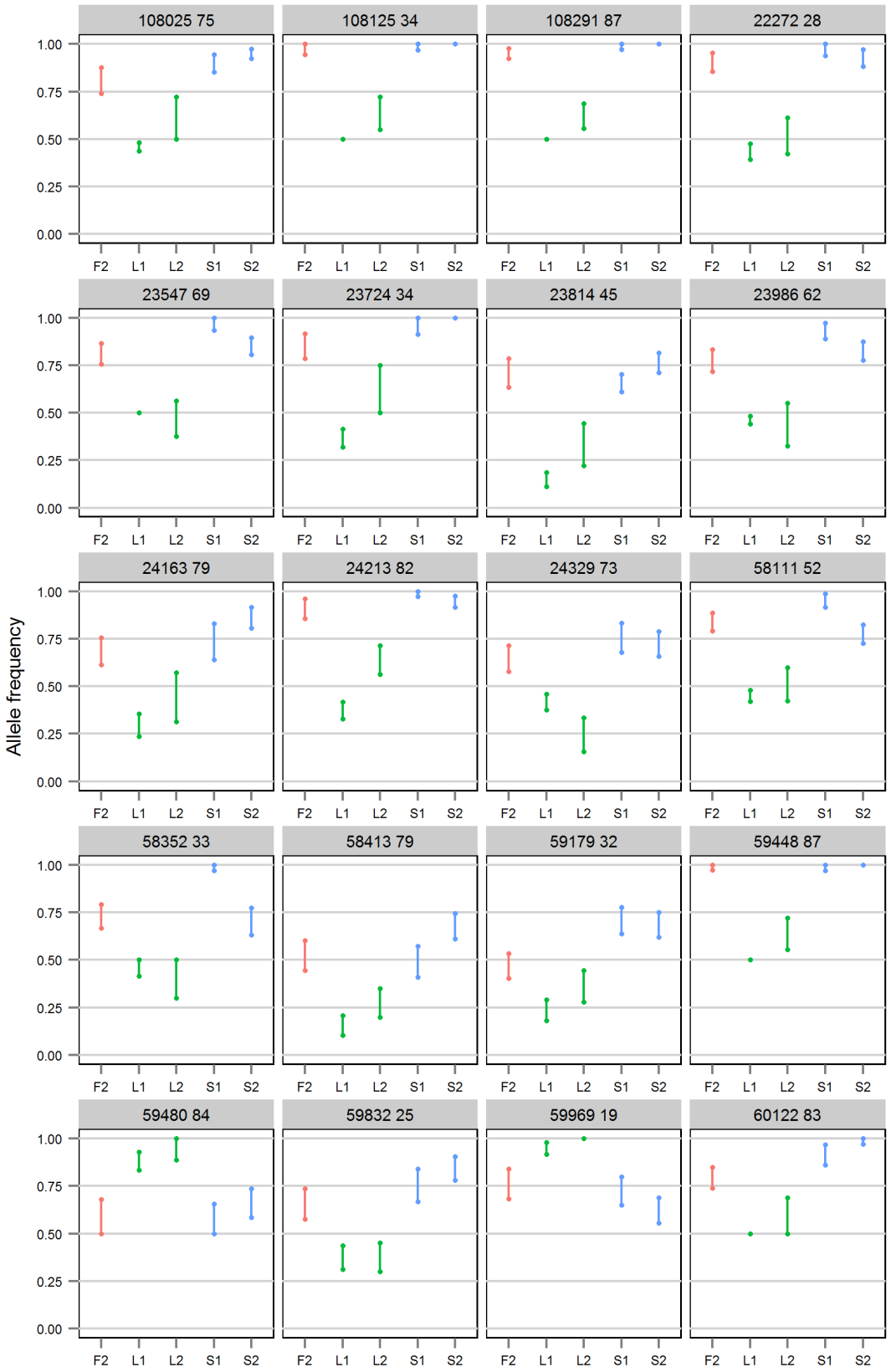



**Figure 2.6:** Allele frequency confidence intervals in SNPs showing consistent changes between replicates and no overlap in values between the small and large selection regimes. The generation before selection can be seen in red, the large selected lines in green and the small selected lines in blue.

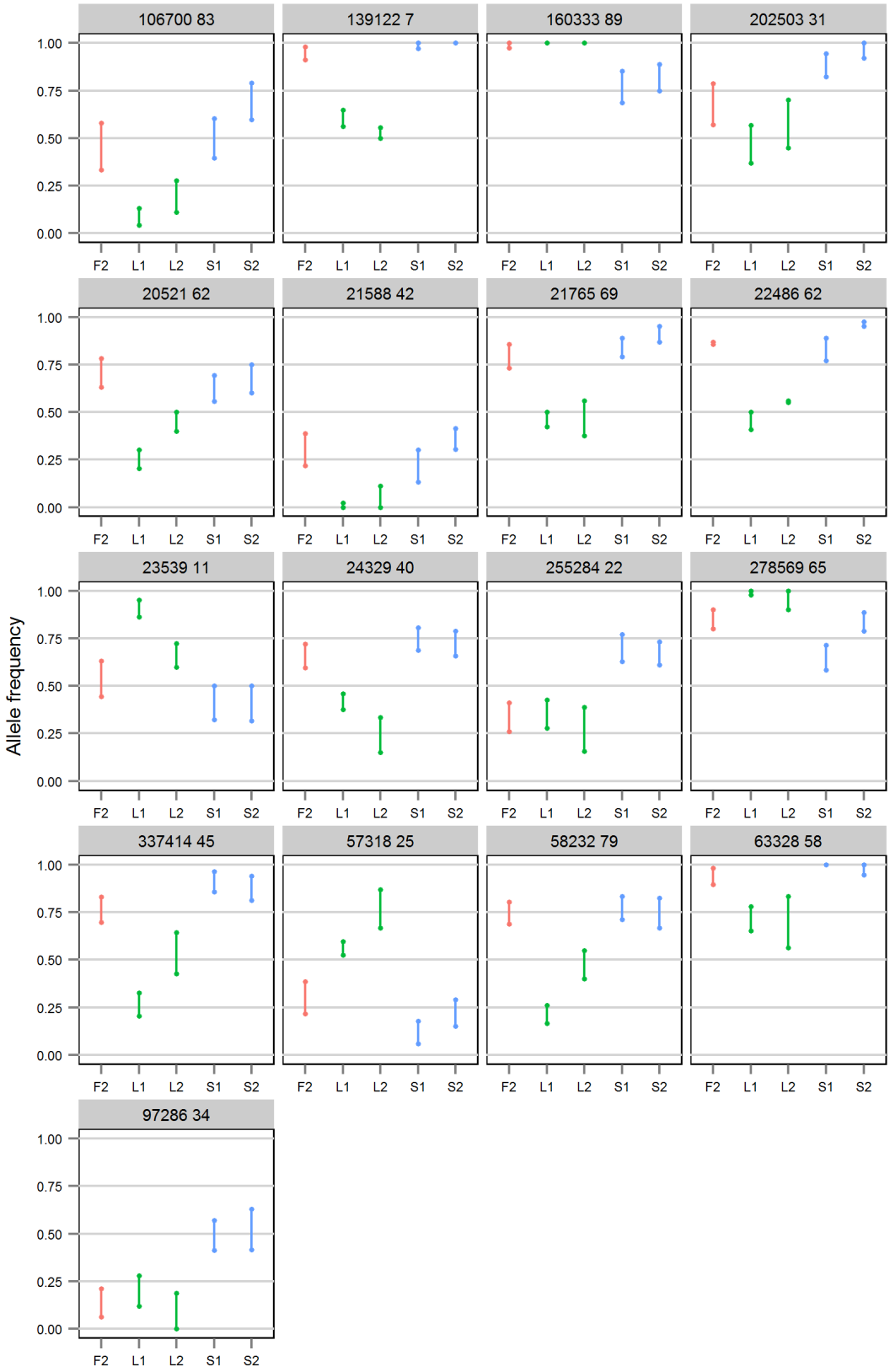



**Figure 2.6:** Continued

lines (figure 2.6). For a further 4 SNPs, trends in allele frequencies were similar for both replicates within the two selection regimes but with an overlap of > 0.05 between a large and small line (figure 2.7, A) and these 4 SNPs were removed from any further analysis. For the remaining 12 SNPs, patterns of allele frequency change were either similar across all four lines or different between replicates within the same regime as well as having an overlap between a large and small line (figure 2.7, B). These 12 SNPs were also removed from any further analysis. The results of the bootstrap analysis for the 37 SNPs where the direction of allele frequency change is the same between both replicates support the hypothesis that these SNPs are under selection. However, it is unlikely the 12 SNPs showing similar allele frequency changes across all four lines or between replicates from different selection regimes (e.g. similar pattern in both L1 and S1) are under selection in our experiment.

Of the 37 SNPs consistently showing signs of selection, 15 were located on a tag which also contained one other SNP not showing signs of selection. A further 2 SNPs were located on a tag which also contained two other SNPs not showing signs of selection. Such apparent discrepancy can be explained by variation in the degree of linkage between the causal variant and the SNP, which in turn is explained by the proportion of alleles of each SNP that are fall on the same haplotype as the causal gene. Figure 2.8 provides a schematic overview of the SNPs filtered out at each stage of the analysis.

### 2.4.5 Functional analysis of SNPs under selection

Using the predicted gene models produced by the NCBI annotation pipeline, the genomic location (intra/intergenic) of the 37 SNPs which had been identified by the outlier analysis as under selection and exhibiting concordant shifts in allele frequencies between the lines, and across generations, were determined. Of 37 SNPs, 29 were found within 26 genes with 13 located in introns and 16 located in exons (table 2.4). The majority (14) of the 16 SNPs located within exons were at the located third codon position with only one each at the first and second codon positions. For SNPs not located within a gene, the closest gene was identified. Predicted proteins were identified in genes associated with a SNP showing signs of selection. A search of the literature and relevant databases (NCBI gene database, UniProt knowledgebase) provided putative functional information for 28 of the predicted proteins (table 2.4).

Across the guppy genome, 184 candidate genes for growth and/or maturation were identified (appendix III). Of these, only six were located on the sex chromosome. Gene ontology (GO) terms were obtained for 17% of the 1623 predicted proteins from genes located on chromosome 12. GO terms were obtained for 3 of the SNPs showing signs of selection (table 2.5). GO terms are split into one of three categories: biological process, molecular function and cellular compartment. A biological process states which broad biological process the gene is associated with; a molecular function term describes the fundamental activity of the gene products at the molecular level and a

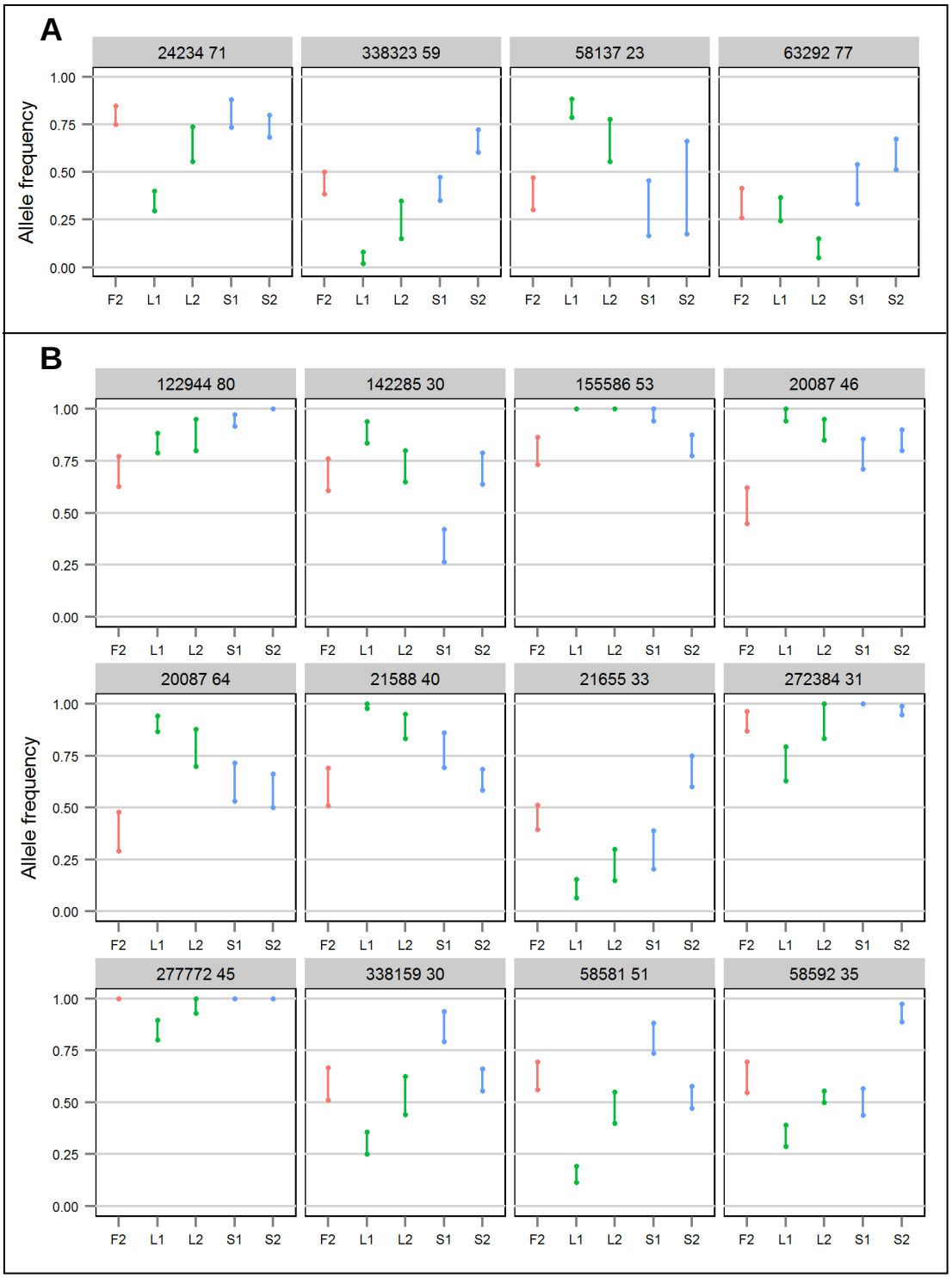



**Figure 2.7:** Allele frequency confidence intervals in (A) SNPs showing similar patterns of allele frequency between replicates but an overlap in values between a small and large line and (B) SNPs showing either variation in the patterns of allele frequency changes between replicates and an overlap in values between a small and large line or a similar pattern of variation across all 4 $F_6$ lines. The generation before selection can be seen in red, the large selected lines in green and the small selected lines in blue.

**Table 2.4:** SNPs showing signs of selection and the genes they are located in/near. * for SNPs located on unassembled scaffolds the SNP position is the position on the scaffold.

| SNP name | Scaffold | Chr | SNP position* | Codon position | Genomic location | Gene ID | Gene Start | Protein ID | Protein description | Further information | Reference | Species function described in |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 255284_22 | 4 | 3 | 10639458 | 3 | Ex | 103462010 | 10633222 | XP_008402685.1 | PREDICTED: bifunctional purine biosynthesis protein PURH | Catalyses steps in de-novo purine synthesis | Greasley *et al.* 2001 | N/A |
| 63328_58 | 142 | 11 | 26175324 | 2 | Itn | 103472981 | 26027974 | XP_008421087.1 | PREDICTED: microtubule-actin cross-linking factor 1 | Regulates cell polarity in early oogenesis | Langdon and Mullins 2011 | Zebrafish |
| 20521_62 | 10 | 12 | 2038395 | 3 | Ex | 103473176 | 2034465 | XP_008421403.1 | PREDICTED: arrestin domain-containing protein 3-like | Regulates signal for transduction | Moore *et al.* 2007 | Zebrafish |
| 21588_42 | 10 | 12 | 3730395 | 1 | Itn | 103473261 | 3717238 | XP_008421581.1 | PREDICTED: kinesin-like protein KIF2A | ATP binding; microtubule binding | Uniprot entry | Human |
| 21765_69 | 10 | 12 | 4068940 | 2 | Itn | 103473261 | 3717238 | XP_008421581.1 | | | | |
| 22272_34 | 10 | 12 | 5158033 | 3 | Ex | 103473322 | 5124625 | XP_008421676.1; XP_008421677.1; XP_008421675.1; XP_008421679.1; | PREDICTED: tenascin-like isoform X1, X1, X2, X3, X4, X5 | Extracellular matrix glycoproteins | Hsia and Schwarzbauer 2005 | Range |
| 22486_62 | 10 | 12 | 5603862 | 1 | Ex | 103473340 | 5599140 | XP_008421721.1 | PREDICTED: netrin receptor UNC5D-like | Morphogenesis of the vascular system | Lu *et al.* 2004 | Zebrafish |
| 23539_11 | 10 | 12 | 7907396 | 2 | Itn | 103473396 | 7882790 | XP_008421844.1 | PREDICTED: laminin subunit gamma-3-like | Non available | - | - |
| 23547_69 | 10 | 12 | 7935019 | 2 | Itn | 103473396 | 7882790 | XP_008421844.1 | | | | |
| 23724_34 | 10 | 12 | 8276234 | 3 | Intra | 103473408 | NA | NA | Uncharacterized gene | Non available | - | - |
| 23814_45 | 10 | 12 | 8410913 | 2 | Ex | 103473416 | 8407054 | XP_008421871.1 | PREDICTED: kynurenine--oxoglutarate transaminase 1 | Catalysis of amino acid degradation | Uniprot entry | Human |
| 23986_62 | 10 | 12 | 8782596 | 3 | Ex | 103473504 | 8779256 | XP_008422027.1 | PREDICTED: CDK5 regulatory subunit-associated protein 2 | Regulation of CDK5. Cell dependant kinase 5 (CDK5) is a regulator of neuronal migration | Dhavan and Tsai 2001 | Human/Mouse |
| 24163_79 | 10 | 12 | 9095304 | 3 | Ex | 103473441 | 9093049 | XP_008421920.1; XP_008421919.1; XP_008421921.1 | PREDICTED: protein transport protein Sec16A isoform X1, X2 and X3 | Defines endoplasmic reticulum sites | Uniprot entry | Human |

**Table 2.4:** Continued

| SNP name | Scaffold | Chr | SNP position* | Codon position | Genomic location | Gene ID | Gene Start | Protein ID | Protein description | Further information | Reference | Species function described in |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24213_82 | 10 | 12 | 9178923 | 1 | Itn | 103473449 | 9165302 | XP_008421934.1 | PREDICTED: lipoxygenase homology domain-containing protein 1 | Involved in function of inner ear hair cells | Uniprot entry | Human |
| 24329_40 | 10 | 12 | 9379545 | 3 | Ex | 103473507 | 9368915 | XP_008422030.1 | PREDICTED: proprotein convertase subtilisin/kexin type 5-like | Involved in the development of the brain and sensory organs, particularly the lateral line | Chitramuthu *et al.* 2010 | Zebrafish |
| 24329_73 | 10 | 12 | 9379578 | 3 | Ex | 103473507 | 9368915 | XP_008422030.1 | | | | |
| 57318_25 | 13 | 12 | 13442128 | 1 | Intra | 103473593 | 13463595 | XP_008422157.1 | PREDICTED: transmembrane protein 132C-like | Non available | - | - |
| 58111_52 | 13 | 12 | 15157586 | 3 | Intra | 103473662 | 15146763 | XP_008422281.1 | PREDICTED: lysosome membrane protein 2-like | Biogenesis of lysosomes and endosomes | Gonzalez *et al.* 2014 | Zebrafish |
| 58232_79 | 13 | 12 | 15466340 | 3 | Ex | 103473677 | 15465735 | XP_008422307.1; XP_008422306.1 | PREDICTED: small conductance calcium-activated potassium channel protein 2 isoform X1 and X2 | Formation of small conductance calcium-activated potassium channels | Adelman *et al.* 2012 | Human/Mouse |
| 58352_33 | 13 | 12 | 15714984 | 3 | Ex | 103473698 | 15714681 | XP_008422342.1; XP_008422343.1 | PREDICTED: NF-kappa-B inhibitor-like protein 1 isoform X1 and X2 | Regulation of innate immune response | Uniprot | Human |
| 58413_79 | 13 | 12 | 15824981 | 2 | Itn | 103473709 | 15820285 | XP_008422365.1 | PREDICTED: anthrax toxin receptor 1-like | Homeostasis of the extracellular matrix | Cingolani *et al.* 2011 | Zebrafish |
| 59179_32 | 13 | 12 | 17157630 | 3 | Ex | 103473787 | 17151774 | XP_008422503.1 | PREDICTED: endoplasmic reticulum metallopeptidase 1 | Development of follicular structures | Garcia-Rudaz *et al.* 2007 | Mice |
| 59448_87 | 13 | 12 | 17727953 | 3 | Ex | 103473803 | 17724942 | XP_008422525.1; XP_008422526.1 | PREDICTED: ADP-ribosylation factor-like protein 3 isoform X1 and X2 | Transportation of myristoylated proteins | Wright *et al.* 2011 | Cell lines |
| 59480_84 | 13 | 12 | 17778728 | 2 | Itn | 103473807 | 17769344 | XP_008422538.1; XP_008422537.1; XP_008422540.1 | PREDICTED: PH and SEC7 domain-containing protein 2-like | Promotes ADP-ribosylation factor 6 (ARF6-required for intracellular transport) activation | Derrien *et al.* 2002 | Cell lines |

**Table 2.4:** Continued

| SNP name | Scaffold | Chr | SNP position* | Codon position | Genomic location | Gene ID | Gene Start | Protein ID | Protein description | Further information | Reference | Species function described in |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 59832_25 | 13 | 12 | 18635986 | 2 | Itn | 103473835 | 18420426 | XP_008422607.1; XP_008422606.1 | PREDICTED: astrotactin-2 isoform X1, X2 and X3 | Regulation of protein localisation | Uniprot | Human |
| 59969_19 | 13 | 12 | 19052144 | 2 | Intra | 103473840 | 19087903 | XP_008422612.1 | PREDICTED: transforming growth factor beta receptor type 3-like | Receptor for a range of beta transforming growth factors | Dalla *et al.* 2005 | Human |
| 60122_83 | 13 | 12 | 19452559 | 2 | Itn | 103473852 | 19440034 | XP_008422649.1 | PREDICTED: trimeric intracellular cation channel type B | Maintenance of intracellular calcium release | Volodarsky *et al.* 2013 | Human |
| 337414_45 | 80 | 12 | 23599887 | 2 | Itn | 103474006 | 23459593 | XP_008422931.1 | PREDICTED: guanine nucleotide exchange factor VAV2 | Involved in neuron development | Moon and Gomez 2010 | Xenopus |
| 106700_83 | 185 | 12 | 25154930 | 2 | Intra | 103474038 | 25177790 | XP_008422977.1 | PREDICTED: uncharacterized protein LOC103474038 | Non available | - | - |
| 108025_75 | 188 | 12 | 20962700 | 3 | Ex | 103473909 | 20941585 | XP_008422780.1; XP_008422781.1 | PREDICTED: WD repeat-containing protein 52 isoform X1 andX2 | Non available | - | - |
| 108125_34 | 188 | 12 | 21151202 | 3 | Ex | 103473914 | 21151025 | XP_008422787.1 | PREDICTED: one cut domain family member 2 | Development of and cell differentiation in the liver | Matthews *et al.* 2008 | Zebrafish |
| 108291_87 | 188 | 12 | 21412323 | 2 | Itn | 103473938 | 21380147 | XP_008422818.1 | PREDICTED: protein Shroom3-like | Linked to lateral line development | Ernst *et al.* 2012 | Zebrafish |
| 202503_31 | 34 | 14 | 17878299 | 3 | Intra | 103476104 | 17831378 | XP_008426423.1; XP_008426426.1; XP_008426424.1; XP_008426425.1; XP_008426422.1; | PREDICTED: voltage-dependent T-type calcium channel subunit alpha-1H-like isoform X1, X1, X1, X1, X2 and X3 | Subunit of a voltage dependant calcuim channel | Lory *et al.* 2006 | Human |
| 97286_34 | 173 | 17 | 30057250 | 3 | Itn | 103480045 | 29967725 | XP_008433010.1, XP_008433009.1, XP_008433011.1 | PREDICTED: unconventional myosin-X isoform X1, X2 and X4 | Involved in neuron development | Sittaramane and Chandrasekhar 2008 | Zebrafish |
| 278569_65 | 58 | 23 | 16180629 | 3 | Ex | 103459689 | 16178201 | XP_008399689.1 | PREDICTED: ATP-dependent RNA helicase DDX51 | Biogenesis of 60S ribosomal subunits | Uniprot entry | Zebrafish |
| 139122_7 | 219 | NA | 116372 | 2 | Intra | 103460262 | 2354 | XP_008400591.1 | PREDICTED: netrin receptor DCC, partial | Axon guidance. | Fricke and Chien 2005 | Zebrafish |
| 160333_89 | 259 | NA | 102618 | 3 | Intra | 103460576 | 130665 | XP_008401031.1 | Gene description: shisa family member 6 | Transmembrane adaptor | Pei and Grishin 2012 | Zebrafish |

**Table 2.5:** GO terms from SNPs consistently showing signs of selection

| SNP | Protein id | Biological process | Molecular function | Cellular compartment |
|---|---|---|---|---|
| 24329_40 | XP_008422030.1 | • Anterior lateral line development | • Peptidase activity | |
| 21588_42 | XP_008421581.1 | • Microtubule-based movement | • Microtubule motor activity | • Kinesin complex<br>• Microtubule |
| 59448_87 | XP_008422525.1; XP_008422526.1 | • Small GTPase mediated signal transduction<br>• Protein transport phosphatidylinositol biosynthetic process<br>• Obsolete GTP catabolic process<br>• Phospholipid metabolic process<br>• Vesicle-mediated transport | • GTP binding/ GTPase activity<br>• Nucleotide binding | • Golgi membrane<br>• Perinuclear region of cytoplasm<br>• Extracellular exosome<br>• Intracellular |

cellular compartment term describes the subcellular and extracellular location of gene products (Ashburner *et al.* 2000).

## 2.5 Discussion

The aim of the experiment outlined here was to use RAD sequencing to study the genetic mechanisms and elucidate the genetic architecture underpinning the shift in life history traits as a result of experimental selection for body size. Using RAD sequencing, the current study identified and genotyped numerous SNPs in four lines of guppies which had previously been selected for body size, alongside fish taken from the generation before selection began. A range of outlier analyses, as well as examination of the allele frequency change between fish from before and after selection identified thirty-seven SNPs showing consistent signs of selection. The results show that in addition to previously observed genetic change (van Wijk *et al.* , 2013), additional regions of the guppy genome responded to, and were associated with, observed phenotypic shifts. However, it should be pointed out at the outset that the detected genetic changes are likely to represent a conservative estimate of genome-wide genetic change due to the strict constraints employed in SNP outlier identification and subsequent analyses.

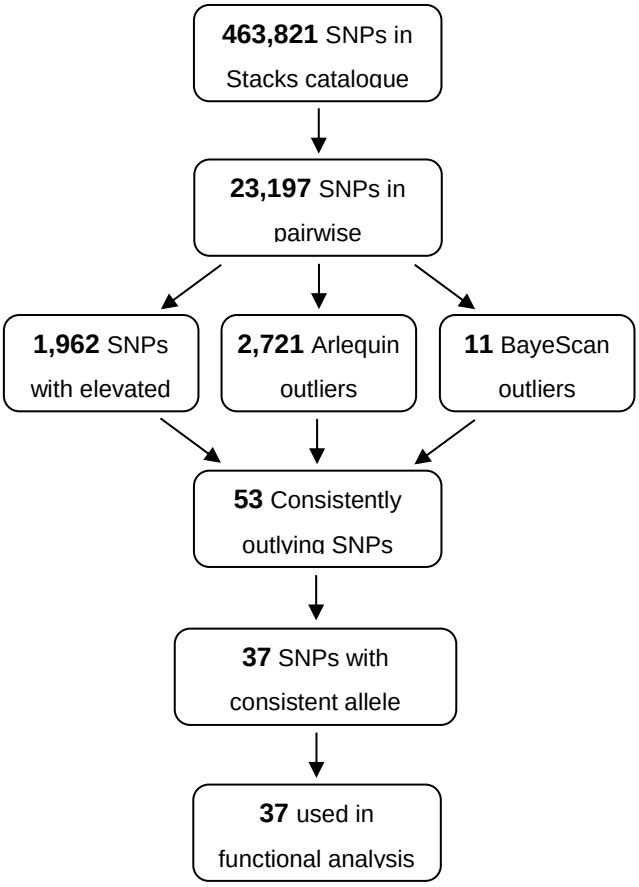



**Figure 2.8:** Schematic showing signs SNPs identified at each stage of analysis

Based on an *in slico* digestion of the guppy genome, the number of RAD loci expected was 93,194. The number of RAD tags in the Stacks catalogue was higher than this at 399,338. The reason for the discrepancy between the number of tags identified and the number predicted is likely to be two-fold: firstly the reference genome used for the *in silco* digestion is incomplete, resulting in the estimated number of tags being an underestimation; secondly) a number of the tags in the Stacks catalogue have low read depth and may have been classified as unique tags due to sequencing error. The stringent conditions which were applied to the SNP calling and genotyping will have prevented any of these erroneously identified tags being used in further analysis. Although the number of tags identified by the current project was similar to that seen in a previous RAD sequencing project of the guppy (Willing *et al.* 2011) and other fish species (Palaiokostas *et al.* 2013), direct comparisons with other studies are not possible due to variation in the enzymes used and genome size.

### 2.5.1 Experimental considerations of the design

#### 2.5.1.1 Identifying SNPs under selection

We used both “Arlequin” and “BayeScan” to detect outlier SNPs between large and small selection lines, and compared individuals before and after selection. As “Arlequin” has been criticised for identifying large numbers of false positives (Pérez-Figueroa *et al.* 2010a, Vilas *et al.* 2012, Lotterhos and Whitlock 2014), a SNP was considered to be showing signs of selection only if it was identified as an outlier in at least two of the selection lines (6 out of the 8 pairwise comparisons). Applying such constraints reduced the number of outlying SNPs identified with the “Arlequin” approach from 11,724 to 28, thereby removing a large number of false positives.

Although “BayeScan” utilised here has also been found to identify false positives, it has been shown to perform more robustly when detecting true outliers than other techniques (Narum and Hess 2011, Vilas *et al.* 2012, Lotterhos and Whitlock 2014). We therefore combined the results from “BayeScan” with those from “Arlequin” to set a less stringent cut-off for identifying outliers. Employing the above combination, any SNP identified as an outlier in one of the pairwise comparisons using the “BayeScan”, as well as at least four of the pairwise comparisons using “Arlequin”, was proposed as being under selection. Using the above approach, a total of 34 SNPS were identified as showing signs of selection.

None of the 53 SNPs considered to be under selection were identified as outliers in all 8 of the pairwise comparisons. However, 19 of these SNPs were identified as outliers in all of the pairwise comparisons that included a large line, while 2 were identified as an outlier in all of the comparisons which included a small line. The latter suggests that the lack of identified outliers in all comparisons is due to variation in the response to selection between the small and large lines, rather than the identification of false positives.

2.5.1.2 Selection or drift?

Although designation of outlier SNPs in multiple comparisons significantly reduces the chance of false positives, it does not exclude the possibility of elevated divergence due to drift acting on one replicate within each selection regime (e.g. a change in allele frequency in L1 and S1 but not in L2 and S2). We therefore bootstrapped the allele frequencies of each SNP within each line (Uusi-Heikkilä *et al.* 2015) to examine the consistency of the changes in allele frequencies between the replicates. Using such an approach, 16 out of 53 SNPs exhibited overlapping 95% allele frequency confidence intervals between large and small lines, and are therefore likely to be under the influence of drift. These SNPs were therefore no longer considered to be showing signs of selection.

The above approach for distinguishing between the effects of selection and drift on a locus may potentially be conservative for two reasons. Firstly, it is possible that the intensity of selection acting on a particular gene differs between two replicates, which may be due to interactions between loci or due to the frequency of the beneficial allele in the founder population. Secondly, it is expected that many of the SNPs identified as under selection are not the causal variant but are linked to it. A SNP in tight linkage would appear to show a very marked change in allele frequency while the allele frequency change in a SNP in loose linkage would be much lower. Therefore variation in the degree of linkage in the different selection lines could result in variation in the change in allele frequency between the replicates.

Under both of these scenarios it would be expected that the direction of allele frequency change for both replicates within a selection regime would be the same. For four of the sixteen SNPs that we removed due to overlapping confidence intervals, allele frequency differences were seen to be moving in the same direction between the two replicates. Two of these SNPs were also located on the sex chromosome within 18.7MB and 49.2MB (24234_71 and 58137_23 respectively) of another SNP under selection. It is therefore more likely the allele frequency changes in these SNPs are due to selection rather than drift.

By examining the changes in allele frequencies at SNPs identified as outliers we were able to exclude 12 SNPs showing signs of genetic drift rather than selection. However, caution should be taken when comparing results between replicates.

2.5.1.3 Inconsistency between SNPs on the same tag

Of the 37 SNPs under selection, 45% (17) were located on a tag which also contained at least one other SNP which did not show any signs of selection. As the mean length of a tag was 93bp, it is extremely unlikely that recombination has occurred among SNPs on the same tag this many times following only three generations of selection. To test the likelihood of more than one SNP occurring within 93bp, a simulation was run based on the number of bases covered by the tags sequenced and the number of polymorphisms found in the guppy genome (figure 2.9). This suggested that 4%

of the SNPs sequenced would be expected to be located within 100bp of another SNP, with a further 53% of the SNPs expected to fall within 1000bp. It is therefore unlikely that the SNPs located on the same tag are result of sequencing or genotyping error.

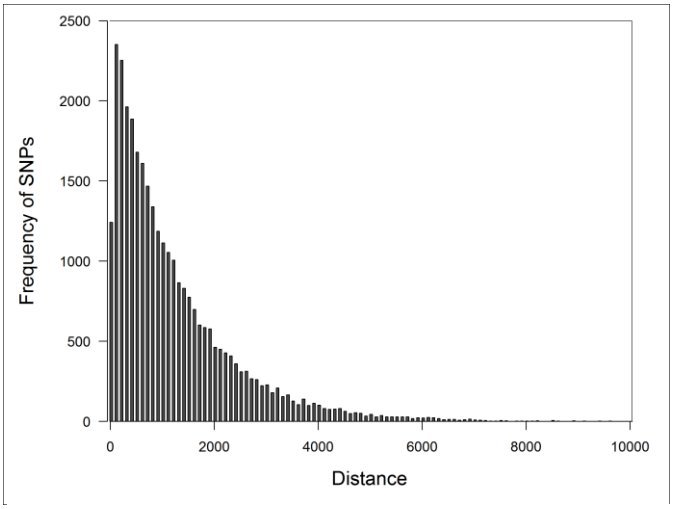



**Figure 2.9:** Expected distribution of the distance (in base pairs) between two adjacent SNPs assuming an equal mutation rate across the genome.

The presence of SNPs on the same tag showing different responses to selection is therefore thought to be a result of ancestral SNPs and variation in the degree of linkage between the causal variant and the SNP being considered, which could also be classed as a form of incomplete hitchhiking. Apart from recombination, it is also possible for variation in the degree of linkage to occur as a result of the age of the SNPs. If two SNPs (one causal and one neutral) are present in the population and occur at high frequency, the degree of linkage between them will be high. If a mutation then creates a second neutral SNP on one haplotype containing the causal SNP, although a new mutation will be in linkage with the causal SNP, linkage will not be as tight as between the first two SNPs. Therefore if selection then occurs on the causal SNP, the neutral SNP in strong linkage will show a high level of divergence between the selected lines, while the second neutral SNP not in strong linkage will not. A schematic explaining this further can be seen in appendix IV.

The potential for two SNPs to be in such close physical proximity and yet show a very different amount of linkage to the causal variant highlights the importance of examining as much of the genome as feasibly possible. Furthermore, it shows that the SNPs identified as under selection here are likely to be linked to the casual variant even though other SNPs on the same tag are showing no signs of selection.

2.5.1.4 Difference between selection regimes

Of the 37 SNPs under selection, only 9 showed a significant change in allele frequency between the $F_2$ and $F_6$ in the small selected lines, while 29 showed a significant change in the large lines (also see figure 2.4). Despite the fact that the SL changed significantly across all selection lines, the underlying genetic change was relatively small in the small lines compared to the large. Such variation could be the result of selection for large body size having had a strong effect on a small number of loci, while selection for small body size has had a much smaller effect on a large number of loci. If selection for large size has affected fewer loci than selection for small size there are two possible factors which could explain this variation.

The first possibility considers the origin of the founder fish. As discussed in section 1.5 (see also section 3.2.2), guppies are found in freshwater streams where waterfalls separate upper river populations from populations downstream. Migration primarily occurs in one direction (upstream to downstream) (Shaw *et al.* 1991, Crispo *et al.* 2006, Barson *et al.* 2009, Willing *et al.* 2010) and as a result downstream populations are genetically more diverse (Shaw *et al.* 1991, Crispo *et al.* 2006, van Oosterhout *et al.* 2006, Willing *et al.* 2010). Downstream populations are also smaller in body size, in part due to the higher levels of predation they experience (Reznick and Bryga 1987b, Reznick, Rodd, *et al.* 1996).

The founder fish used in van Wijk *et al.* (2013), were collected from a downstream population and therefore genetic diversity was high, and the average SL was low (van Wijk *et al.* 2013). Due to unidirectional flow of migration, it is likely that any alleles coding for large size in a downstream population would exist at low frequency and wouldn't persist for long. Therefore many of the alleles present in the founder fish may have been associated with small size with only a small number associated with large size. The increase in SL during the first three generations of selection would have increased the frequency of these 'large' alleles in the founder population, but it would not have increased the number of alleles responsible for large size. Selection for large body size increased the frequency of such alleles further, creating large variation at a small number of SNPs. At the same time, selection for small body size would have reduced the frequency of the alleles coding for large size and resulting in the genetic structure of the $F_6$ small lines being very similar to that of the $F_1$. A review of evolve and re-sequence studies (E&R) showed that the history of the founder population can influence the power the experiment to identify all of the variants under selection (Schlötterer *et al.* 2015).

The second explanation which could lead to selection for large size affecting fewer loci than selection for small size is variation in the genetic mechanisms underpinning small and large body size. The observed pattern of variation could result from selection for large body size acting on a small number of large effect loci while selection for small body size is acting on a large number of small effect loci. The effect of inbreeding on body size in a wide range of species highlights the fact that increased homozygosity at a large number of loci can result in smaller body size (Goldish 1996, Beekman *et al.* 1999, Rzewuska *et al.* 2005, Honan 2008, Lacy and Horner 2012). Although inbreeding was not identified in this study (van Wijk *et al.* 2013), it is possible that the small body size observed is an indirect result of the accumulation of a large number of small (and potentially deleterious) changes, while a large body size is

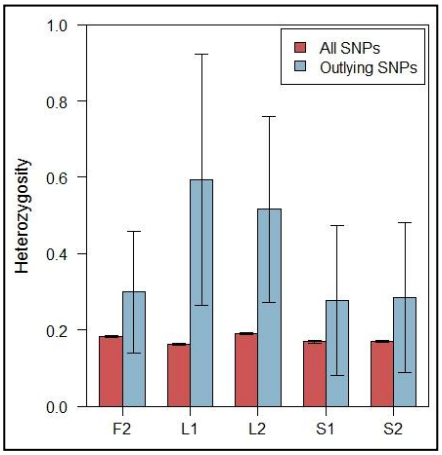



**Figure 2.10:** Heterozygosity of all SNPS and those showing signs of selection in each of the five populations.

the result of direct selection on a small number of alleles involved in pathways linked to body size. If such was the case it would be expected that selection on a large number of genes linked would result in reduced heterozygosity in the small lines. Although there was no variation in heterozygosity between the lines across all genotyped SNPs, observed heterozygosity in SNPs identified as being under selection was significantly higher in the large lines than it was in the small lines (figure 2.10).

However, although the observed variation in genetic response could be the result of selection for large size affecting fewer loci than selection for small size, it could also be explained by variation in the level of linkage disequilibrium. For example, it could be that the selection on large body size acted on a small number of alleles which were in linkage disequilibrium with a large number of other loci across the chromosome. If selection acted on only one or two loci which were part of a much larger linkage block then our analysis would have identified signs of selection across this entire block. On an autosomal chromosome recombination makes a linkage block this large unlikely, however the level of recombination across the guppy sex chromosome is unknown. It is therefore not possible to determine how many loci are actually under selection on chromosome 12 and how many are simply in linkage with the loci under selection. Concurrently, the selection for small body size might have acted on the same loci as selection for large body size, but the size of the linkage block these loci are present on in the small selection lines might have been smaller. A lower level of linkage between the causal loci and the rest of the chromosome in the small selected lines would have resulted in our analysis identifying only a small number of loci showing signs of selection in the small lines.

### 2.5.2 Hard or soft selective sweep?

The results of the genome-wide analysis of the experimental selection lines highlighted only a small number of SNPs showing signs of selection. The majority of these SNPs (86%) were located on the sex chromosome which may indicate a large amount of linkage across this chromosome. All of the five SNPs showing signs of selection on the autosome were located on independent chromosomes. Despite the apparent linkage between the SNPs on the sex chromosome it is more likely that the observed genetic variation results from a soft selective sweep.

Traditionally, both hard and soft selective sweeps can create a region of reduced diversity around a mutation under selection (Burke 2012), although the size of the region affected and the degree of the reduction in diversity would differ. Following a hard selective sweep, in which a novel mutation rapidly becomes fixed, the reduction in diversity around the allele in question is severe, and the area affected large due to the strong linkage between it and the variation around it. Following a soft selective sweep, in which selection acts on standing variation, the drop in diversity does not cover such a large area and is not as significant, making it more difficult to detect. It has been considered traditionally that hard selective sweeps act on few large effect alleles, while soft selective sweeps act on a larger number of small effect alleles. Moreover, it has also been pointed out that there no

reason why a single standing genetic variant could not be rapidly swept to fixation as with new mutations (Orr and Betancourt 2001, Jensen 2014). When combined with the potential for the signatures of an old hard sweep to be very similar to that of a soft sweep this has led to questions surrounding our ability to distinguish between hard and soft sweeps (Jensen 2014, Schrider *et al.* 2015, Stephan 2015).

High levels of genetic variation (and selection for body size in the founding population) in our study, taken together with the short time over which selection was imposed (making new mutations extremely unlikely), and the lack of a significant drop in diversity surrounding any of the outlying SNPs, suggest that the observed changes in allele frequencies likely derived from a soft selective sweep. The observed soft selective sweep is likely to represent an example of an incomplete sweep because very few alleles under selection reached fixation. Pavlidis *et al.* (2012) showed that as the optimum phenotype for polygenic traits can be obtained by a range of allele combinations, the more loci involved in a trait the lower the chance the alleles involved will reach fixation. It is therefore possible that for such a polygenic trait as body size, fixation for the alleles underpinning it may never be reached.

### 2.5.3 Sex linkage of SNPs under selection

Of the 37 SNPs under selection, 32 (86%) were located on the sex chromosome, which highlights the sex-linked nature of the phenotypic traits under selection here. Although the guppy has an X/Y sex determining system (Tripathi *et al.* 2009; Lisachov *et al.* 2015), the majority of the sex chromosome is pseudoautosomal (Nanda *et al.* 2014, Lisachov *et al.* 2015). It is thought that there are three distinct regions of the guppy Y chromosome, the male specific non-recombining region 1 (MSNR1), the male specific non-recombining region 2 (MSNR1), the freely recombining region 1 (FR1) and the freely recombining region 2 (FR2) (see figure 6 in Lisachov *et al.* 2015). The Y chromosome is homologous with the X at all but the MSNR2 but exhibits reduced recombination in both the MSNR1 and MSNR2 (Lisachov *et al.* 2015).

As the reference genome used in this study originated from a female guppy (Fraser *et al.* 2014), it is expected that any region of the Y chromosome not homologous to the X will be excluded in the assembly, the adjacent region on X chromosome. Indeed, although the length of the Y chromosome has been found to differ between different strains (Nanda *et al.* 2014), even at its shortest the sex chromosome is thought to be the longest chromosome in the guppy genome, yet the sequence for this linkage group is only the sixth longest in the assembly. It is possible, however, that as the homology between the X and Y of the MSNR1 and the MSNR2 are unknown that sequences from these regions in the Y would have successfully been mapped to the adjacent X sequence. If this was the case it would be expected that a drop in the quality of the mapping would be visible in these regions, however no such drop was observed.

The analyses used in the current chapter focus on the relative measure of divergence, $F_{st}$ which compares within-population diversity to between-population diversity. Regions of reduced recombination have reduced within-population diversity (Charlesworth *et al.* 1997) which will lead to inflated measures of relative divergence such as $F_{st}$ (Cruickshank and Hahn 2014). Reduced recombination is expected on a sex chromosome, although recombination is thought to occur freely in the pseudo-autosomal sections of the guppy sex chromosomes. It is possible that the inflated $F_{st}$ values seen across the sex chromosome in the current project are simply the result of reduced diversity across this chromosome. In order to assess this we compared the observed heterozygosity of a subset of autosomal SNPs to SNPs located on the sex chromosome. Comparisons of 5 random subsets for each population (25 in total) found a significant difference in diversity between the sex chromosome and the autosomes only twice (out of 25) (Appendix II). However, when considering the Y chromosome, in particular the non-recombining fragments, heterozygosity could be considered simply as a measure of the differences between the X and Y chromosome and not a measure of diversity on the Y chromosome. Using the data available here, there is no appropriate way to determine whether the inflated $F_{st}$ values seen across the sex chromosome are a result of selection on Y-linked loci or an artefact of reduced recombination.

It is not possible to know whether Y-specific reads have been included in our analysis. However, the strong signs of selection across the sex chromosome, the phenotypic response which was observed only in males and a previous observation of sex specific segregation of alleles linked to body size (Tripathi *et al.* 2009b) indicate Y linkage of the genes under selection in this study. Further to this, the reduced levels of recombination identified (Nanda *et al.* 2014) are likely to have led to increased linkage which would explain the elevated $F_{st}$ across the whole chromosome. It would also be expected that such increased linkage would result in linkage between SNPs present in the Y-specific region and SNPs present in the freely recombining regions.

### 2.5.4 Putative function of genes under selection

#### 2.5.4.1 Genes located on the sex chromosome

Only one of the 37 SNPs showing signs of selection was located at the second codon position within an exon and therefore synonymous. This SNP was located on chromosome 12 in a gene coding for cysteine conjugate-beta lyase (*ccbl1*). One other SNP likely to be a synonymous mutation (first coding position in the exon) is also located on chromosome 12, in the netrin receptor *unc5d*-like gene (*unc5d*). For any SNP in our study, unless the causal allele is Y-linked, the gene must be differentially expressed in order to explain the lack of phenotypic change in females. A recent study of the transcriptome found transcripts within 100bp of both the *ccbl1* gene and the UNC5D gene to be differentially expressed in males and females. Both of these genes have the potential to be contributing to growth and maturation, however neither has been previously linked to either.

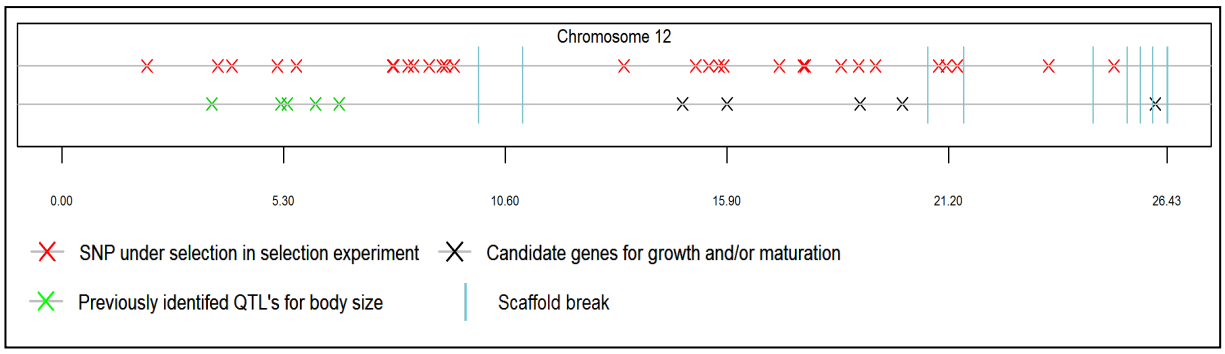



**Figure 2.11:** Location of outlying SNPs, candidate genes and previously identified QTLs for body size on chromosome 12. Vertical blue lines show where one scaffold ends and another begins.

The remaining SNPs were either located in intragenic or intergenic regions on both the sex chromosome and the autosomes. On the sex chromosome, 30 of the outlying SNPs appear to fall into four broad groups according to the scaffold on which they are located.

The first group is located on scaffold 10 (between 0 and 9.96MB) and in close proximity to previously identified QTL for body size in the guppy (figure 2.11) (Tripathi *et al.* 2009b). However, it does not contain candidate genes for growth or body size. Our attention therefore turns to the genes in which the SNPs under selection are located. Of the fourteen SNPs located on scaffold 10, thirteen are located within ten genes. The SNPs located in two of these genes are likely to be synonymous mutations and have already been discussed (*ccbl1* and *unc5d*). Of the remaining eight genes, three are noteworthy. Two (*pcsk5* and *cdk5*) have both been identified as being involved in organ development while the kinesin-like protein *kif2a* (*kif2a*) may contribute to growth by controlling the amount of energy available to a cell. Both the *kif2a* and the *cdk5* genes have been found to be differentially expressed in male and female guppies (Sharma *et al.* 2014). In addition to the genes above it is possible that sequence data is missing from either side of this scaffold and the SNPs here are linked to genes located on the missing sequences.

The majority of the second group of SNPs are located on scaffold 13 (11.01-20.70MB), however given the observed similarities in allele frequencies between these SNPs and those on scaffold 188 (20.70-21.55MB) and the SNP at the distal end of scaffold 13, the three SNPs located on scaffold 188 are also included. Of the 11 genes (14 SNPs) in this group, four are noteworthy. Two of these genes (*oc2* and *shrm3*-like) have been linked to organ development in zebrafish (liver and lateral line respectively) (Matthews *et al.* 2008, Ernst *et al.* 2012). The two other genes (*kcnn3* and *tmem38b*) are involved in the maintenance of intracellular calcium levels (Adelman *et al.* 2012, Volodarsky *et al.* 2013). In medaka (*Oryzias latipes*), an increase in intracellular calcium ions has been found to take place in response to an increase in the gonadotropin-releasing hormone (Strandabø *et al.* 2013) and therefore changes in these genes may have been related to changes in the level of gonadotropin.

It is possible that some of the genes discussed above are under selection. However, it is unlikely that these SNPs were synonymous mutations, and the level of linkage across this chromosome is thought to be high. Therefore the causal variant may also be located in one of the four candidate genes found on this chromosome. The first candidate gene on this scaffold is an NADH dehydrogenase gene has been associated with body size in rainbow trout (*Oncorhynchus mykiss*), and found to be differentially expressed in male and female guppies (Sharma *et al.* 2014). Also located on this chromosome and differentially expressed in male and female guppies (Sharma *et al.* 2014) is the transforming growth factor beta receptor type 3 (*tgfß r-III*). *tgfß r-iii* has been shown to increase growth by negatively regulating the TGF/SMAD cascade (Eickelberg *et al.* 2002) as well as acting as a receptor for the insulin growth factor binding protein 3 (Wu *et al.* 2000). While this gene contained no SNP showing signs of selection, SNP 59969_19 was located only 34.89KB from the start of the gene.

The remaining two candidate genes on this scaffold are both related to the hormone prolactin. A wide range of functions have been attributed to the hormone prolactin such as metabolism, lactation, immune response and osmoregulation in fish (Manzon 2002, Boutet *et al.* 2007). As well as being linked to growth rate and body size in agricultural species (Bhattacharya *et al.* 2011, Lü *et al.* 2011), *prl* has also been linked to both growth and reproduction in fish (Shepherd *et al.* 1997, Whittington and Wilson 2013, Velan *et al.* 2015). One of these genes, (*prl*) has been found to be significantly upregulated in female guppies (Sharma *et al.* 2014). As guppy females grow throughout their lifetime while growth in males is significantly reduced after maturation the up-regulation of *prl* genes in female guppies supports the hypothesis that this gene plays a role in growth and body size.

2.5.4.2 Genes located on the autosomes

Only five of the SNPs showing signs of selection are located on autosomes, four of which are located within a gene (table 2.4). The SNP on chromosome 14 is not located within a gene, which suggests that it is either a false positive, or is in linkage with the causal variant. The closest gene to this SNP is the voltage-dependent T-type calcium channel subunit alpha-1H gene. As discussed previously, cellular calcium levels have been found to increase as a result of an increase in gonadotropin which may be linked to the changes in the timing of maturation observed.

The four remaining SNPs are located on chromosomes 3, 11, 17 and 23 in the genes *atic*, *macf1*, *myo10* and *ddx51* respectively. The *myo10* and *ddx51* genes are not thought to be linked to body size or maturation timing. The protein produced by the *atic* gene (located on chromosome 3) catalyses steps in the de-novo synthesis of purine, which has been shown to act as a growth limiter in cultured fibroblast cells (Kondo *et al.* 2000). It could be expected therefore that changes in the de-novo synthesis of purine may be associated with changes in body size.

It is also possible that the SNP identified as being under selection on chromosome 3 is linked to the causal variant. Within 0.8MB of this SNP is a candidate genes for maturation (cytochrome P450 which encodes the aromatase protein). Inhibition of aromatase, which is the closest of these genes to the SNP, has been shown to be induce early maturation in fish (Antonopoulou *et al.* 1995, Charan *et al.* 2013). The SNP on chromosome 11 showing signs of selection is located in the *macf1* gene on chromosome 11 which is not thought to be linked to body size. The candidate gene closest to this SNP (0.8MB) is another NADH dehydrogenase gene which as discussed in section 2.2.6 has been previously linked to body size (Salem *et al.* 2012).

As it is unlikely that a gene located on an autosomal chromosome would be Y linked, any gene under selection and located on an autosome in our study would be expected to be differentially expressed in males and females. While only one of the SNPs showing signs of selection is within 100bp of a transcript differentially expressed between males and females (SNP 255284_22 on Chromosome 3), all four candidate genes discussed above are located in a differentially expressed transcript (Sharma *et al.* 2014). In particular the aromatase candidate gene located on chromosome 3 was found to be significantly up-regulated in females.

2.5.4.3 Genes located on unassembled scaffolds

Two of the SNPs showing signs of selection were located on scaffolds which had not successfully been assigned to a linkage group. It is thereby not possible to determine whether these loci belong on the sex chromosome or an autosome, though since these scaffolds have been annotated, it is possible to consider putative function. Neither of the two SNPs is located within a gene. The *dcc* gene which is located 0.027MB from a SNP identified as under selection, is the involved in axon guidance in early development and has been linked to morphogenesis of the pectoral fin in Zebrafish (Fricke and Chien 2005).

### 2.5.5 Consequences for fisheries-induced evolution

As discussed in section 1.4, the importance of findings presented here for FIE specifically, and more generally for size-selective changes in wild populations, lies in the number of loci involved and the degree of change observed in these loci. Data indicate that selection for small body size might affect a large number of loci while selection for large body size might have a stronger effect on fewer loci. Due to the size limits imposed on fish caught by fisheries, and the greater value of fish of large size (Gwinn *et al.* 2015), FIE selects for small size. Studies of harvested populations show that fish in these populations are smaller than their ancestors and mature earlier, which is consistent with the pattern seen in the small selected lines in this study (Ricker 1981, Olsen *et al.* 2004, Swain *et al.* 2007, Fenberg and Roy 2008, Neuheimer and Taggart 2010, Hutchings and Rangeley 2011).

Previous results from the selection experiment utilised here have shown that size selective harvesting can cause genetic change as well as phenotypic change (van Wijk *et al.* 2013). The

results from the RAD sequencing performed here show that size selective harvesting has had a genome wide impact. It is unlikely that we have been able to identify the causal genomic variants underpinning the shifts in life history traits. However, the regions undergoing genetic change here are consistent with regions previously linked to body size in the guppy. Therefore our results offer direct support for the genome wide effect of FIE. As well as providing evidence of genomic change by size selective harvesting our results also offer an insight into the potential for recovery in exploited populations.

If selection for small body size is underpinned by a large number of loci a reduction in the level of diversity would be expected. Therefore, if the pattern of genetic change seen in this study is similar to that in the wild, it would be expected that harvested wild populations would exhibit reduced levels of diversity. Several studies have shown reduced genetic diversity (Hutchinson *et al.* 2003; Hoarau *et al.* 2005; Hauser *et al.* 2002; Kenchington 2003; Pinsky & Palumbi 2014 although see also Ruzzante *et al.* 2001; Poulsen *et al.* 2006). Although it could be argued that selection on a large number of small effect loci could be easier to reverse due to the small change in allele frequencies, the short time for which our study was run compared to the time over which fisheries-induced evolution has taken place, makes it very likely that selection on wild populations has had a much larger effect on a large number of loci. Therefore, to return the observed traits to their pre-harvesting values, selection would have to be equally as strong as fishing, and in the opposing direction for a long time. Following fishing moratorium selection for large body size (as would be expected under natural selection) would be the only selection pressure following fishing moratorium. Under such a scenario, the genetic changes observed in the large lines here suggest that selection for large body size would only act on a smaller number of loci and would not result in a genome-wide recovery of genetic diversity. Conover *et al.* (2009) showed that phenotypic changes as a result of selection for body size over five generations did begin to return to their pre-selection values following five generations during which harvesting was halted. Following the logic above, the recovery Conover *et al.* (2009) observed could be explained by change at numerous loci only being negligible due to the limited number of generations over which selection took place. In agreement with such a theory, populations of Atlantic cod which have been harvested for decades have showed very little recovery following a dramatic reduction in harvesting (Hutchings and Rangeley 2011).

# Chapter 3: Examining candidate loci in wild populations

## 3.1 Abstract

In the rivers in Trinidad guppies exist in what has been described as a 'natural laboratory'. Waterfalls form barriers to larger fish migration, creating significant differences in the level of predation experienced by fish above and below these waterfalls. Above the waterfalls, where the level of predation is low, fish have evolved to grow larger and mature later. In the pools below the waterfalls, where the level of predation is high, fish are smaller and mature earlier. The changes in life history traits in response to predation in these wild populations are similar to the changes seen in response to harvesting and in the van Wijk (2011) selection experiment. Consequently, the wild populations offered a unique opportunity to provide support for the selection identified in the previous chapter. To examine this, a selection of the SNPs showing signs of selection in Chapter Two were genotyped in fish from one upstream and one downstream site from each of nine rivers. Signs of selection were identified using $F_{st}$ outlier analysis and analysis of allele frequency changes. None of the SNPs showed consistent signs of selection across the nine rivers studied. The apparent lack of selection at these loci is probably the result of a lack of convergent evolution, although the experimental design may have also contributed.

## 3.2 Introduction

### 3.2.1 Using wild populations to validate inferences from experimental selection

Laboratory experiments have been invaluable in determining the genetic basis of evolutionary processes such as the agents driving selection (Harshman and Hoffmann 2000), the loci or genomic regions underpinning quantitative traits (Ollivier *et al.* 1997, Jacobsson *et al.* 2005, Yang *et al.* 2007), the contribution of phenotypic plasticity (Scheiner 2002, Garland and Kelly 2006) and the heritability of the trait in question (Visscher *et al.* 2008). Despite the accepted importance of laboratory experiments in increasing our understanding of evolution (Fuller *et al.* 2005), such studies have their limitations (Huey and Rosenzweig 2009), and findings should be validated in wild scenarios where appropriate. The consequences of not performing such complementary empirical tests are illustrated in the case of the evolution of insecticide resistance, where laboratory selection experiments have yielded different results to those found in a natural setting (McKenzie and Batterham 1994). Early laboratory selection experiments demonstrated that pesticide resistance was under polygenic control in *Drosophila* (Crow 1957, McKenzie *et al.* 1992, Morton 1993). However, resistance in natural populations is more often underpinned by a limited number of large effect genes (Carrière and Roff 1995, ffrench-Constant 1996, McKenzie 2000, Raymond *et al.* 2001, Rinkevich *et al.* 2007). These differences appear to have been primarily caused by the reduced size of the gene pool in a laboratory setting, though the length of time the strains had been bred in the laboratory are also likely to have influenced outcomes (Ffrench-Constant 2013). Although many studies highlight the need to confirm the results of laboratory selection experiments in wild populations, these tests are still not widely performed.

One early example of a selection experiment where the results were tested in the wild is that of the LDH gene in the estuarine killifish *Fundulus heteroclitus*. Using a laboratory selection experiment, Powers *et al.* (1991) showed that two allozyme alleles were linked to a range of temperature dependant traits including swimming performance, metabolism, developmental rates, and overall survivorship at high temperatures (see also DiMichele & Powers (1991)). Populations of *F. heteroclitus* found on the Atlantic coasts of the United States are distributed along a latitudinal thermal gradient. Studies of these wild populations have shown that fish at the extremes of the range are fixed for alternate allozyme alleles in the LDH gene (Powers and Schulte 1998). Other studies where genetic variation associated with phenotypic traits were identified in the laboratory before being verified in wild populations, include genes mediating flowering time in barley (Turner *et al.* 2005, Jones *et al.* 2008) and *Arabidopsis* (Aranzana, Kim, Zhao, Bakker, *et al.* 2005), the mutation encoding antimalarial drug resistance in *Plasmodium falciparum* (Ariey *et al.* 2014), nesting building in the house mouse (Lynch 1980; 1992) and the genomic response to predation and parasitism seen in *Daphnia magna* (Orsini *et al.* 2012). The existence of clear patterns of genetic and phenotypic diversity in populations of Trinidadian guppies, *Poecilia reticulata*, provides a potentially insightful opportunity to compare inferences from the laboratory (Chapter 2), with patterns in the wild.

### 3.2.2 The guppy system

The geology and the river system found on the northern range of Trinidad have created what has previously been described as a 'natural laboratory' (Haskins *et al.* 1961). Several parallel rivers drain the northern slopes of the range whilst the southern slopes are drained by several rivers which flow into one of two main drainages, the Caroni (to the west) and the Oropuche (to the east). All main rivers contain guppies, *Poecilia reticulata*. Early genetic studies suggested a 'two arcs hypothesis' of phylogeography with guppies from the Caroni drainage and rivers along the northern coast originating in the Orinoco, a river which runs through Venezuela, while those in the Oropuche drainage originating in a separate drainage in South America. The hypothesis stemmed from the high levels of genetic divergence observed between populations in the Oropuche and Caroni drainages which are thought to have been separated for 600,000 to 1.2 million years (Carvalho *et al.* 1991, Fajen and Breden 1992, Alexander *et al.* 2006). Under the two arcs hypothesis it would be expected that differentiation between populations in the Caroni and Oropuche would be high, and on the other hand, low between the Caroni and the rivers on the northern slopes (Suk and Neff 2009). However, more recent genetic analyses have found that both of the populations in two drainages on the south side of the range are highly differentiated from rivers on the northern slopes, and that populations in these three drainages make up three distinct lineages (Alexander *et al.* 2006, Schories *et al.* 2009, Suk and Neff 2009, Willing *et al.* 2010).

### 3.2.3 Adaptation to predation regimes

Many of the rivers located on both the northern and southern slopes are segregated by waterfalls large enough to act as a barrier to upward migration and colonisation of many species of fish (Magurran 2005). Although guppies have been able to colonise pools both above and below these waterfalls, many predators have not. For individuals living in the headwaters above the waterfalls, the risk of predation from predatory fish as adults is low. Typically these sections of river contain one species of predatory fish, the killifish (*Rivulus hartii*), which is an occasional predator and only a significant threat to smaller and juvenile individuals (Mattingly & Butler 1994). In contrast, those living downstream of the waterfalls experience much higher predation where they are targeted by a range of predators such as the pike cichlid (*Crenicichla alta*), the wolf fish (*Hoplias malabaricus*) and the characin (*Astyanax bimaculatus*), many of which will selectively predate on large, mature individuals (Magurran and Phillip 2001). In addition to the predation by fish, aerial and invertebrate predators also feed on guppies (Magurran 2005). Several avian predators such as green kingfishers (*Chloroceryle americana*), American pygmy kingfishers (*C. aenea*), belted kingfishers (*Ceryle alcyon*) and great kiskadees (*Pitangus sulphuratus*) are thought to feed on guppies. The fishing bat (*Noctilio leporinus*) is also a potential aerial predator (Templeton and Shriner 2004, Magurran 2005). The main invertebrates to prey on guppies are freshwater prawns (*Macrobrachium*), although the level of predation by prawns has been debated (Liley and Luyten 1985, Endler 1991). Both aerial and invertebrate predators are thought to affect guppy anti-predator behaviour however neither impose size selective predation and are therefore thought to have little effect on the variation in life history traits (Rodd and Reznick 1991, Templeton and Shriner 2004).

As well as the notable difference in predation regimes, populations from above and below waterfalls also differ in a range of phenotypic, life history and behavioural traits. The first of these was noted by Endler (1980; 1984) who identified that the colouration of male guppies varied as a result of a combination of natural and sexual selection, with natural selection for dull colour patterns dominating in the highly predated populations, and sexual selection for bright coloured males dominating in populations with low predation rates (see also Houde (1997) and Martin & Johnsen (2007)). Guppies have also been found to differ in behavioural traits (Seghers 1974, Breden *et al.* 1987, Magurran and Seghers 1990, Magurran *et al.* 1992, 1995, Houde 1997, O'Steen *et al.* 2002, Kelley and Magurran 2003, Ghalambor *et al.* 2004), life history traits (Reznick *et al.* 1996; Reznick 1987; Reznick & Endler 1982), sex ratio (Haskins *et al.* 1961, Seghers 1974, Peterson and Small 2005), parasite resistance (Oosterhout 2003, Oosterhout and Smith 2007), morphology (Langerhans and DeWitt 2004), swimming performance (Ghalambor *et al.* 2004) and diet (Zandonà *et al.* 2011, Sullam *et al.* 2014) between the high and low predation sites. The genetic basis of some traits has been examined using laboratory studies and common garden experiments (Endler 1980, Reznick and Endler 1982, Reznick and Bryga 1996, O'Steen *et al.* 2002). The traits of interest here included variation in life history strategies in relation to predation regime. Fish from low predation, upstream populations mature later, grow larger, allocate less energy to reproduction

and produce smaller broods of larger offspring than those in higher predation, downstream populations (Reznick *et al.* 1996; Reznick 1987; Reznick & Endler 1982).

While the level of predation by fish appears to have a direct effect on life history traits (Reznick and Bryga 1996), predation alone cannot explain all observed variation and other pressures such as parasitism (Cable and van Oosterhout 2007) and population density are likely to be factors. To be the sole driver, the mortality risk from predation would need to be age-specific, that is, the risk of mortality for adults would be higher in high predation sites than in low predation sites, while the predation risk for juveniles would not differ (Travis, Reznick, and Bassar 2014). However, this appears not to be the case with the mortality risk the same across the age classes (Mattingly & Butler 1994; Reznick *et al.* 1996). It has therefore been suggested that feedback between ecological and evolutionary processes are playing a role in the evolution of the life history patterns observed (Reznick *et al.* 2002, Bassar 2010, Travis, Reznick, and Bassar 2014). As guppy populations colonise low predation sites, the population grows, and competition for food resources drives a shift from eating low numbers of high quality invertebrates to large amounts of low quality algae (de Villemereuil and López-Sepulcre 2011, Palkovacs *et al.* 2011, Zandonà *et al.* 2011, Sullam *et al.* 2014). In turn, the increase in time and energy allocated to resource acquisition drives a shift in life history patterns towards later maturity and smaller broods of higher quality offspring (Bassar *et al.* 2013, 2015).

Perhaps the most important aspect of guppy populations present in Trinidad is the replication of high / low predation sites. Upstream and downstream populations within a river are more genetically related to each other than to populations in other rivers (Willing *et al.* 2010), and gene flow between sites is limited, and occurs primarily in the downstream direction. Despite the genetic divergence, high predation sites are phenotypically more similar to each high predation sites in other rivers than they are to their corresponding low predation site. The guppy system in Trinidad therefore represents adaptive phenotypic variation which has evolved independently across multiple sites. Systems such as this, which have also been so well studied, do not occur often in nature and provide an opportunity to examine patterns from experimental studies where certain factors can be controlled and manipulated.

By utilising the ‘natural laboratory’ that is the river system in Northern Trinidad and the guppies found within, here we will assess the importance of candidate loci identified in Chapter 2 of the thesis, by examining selection in wild populations experiencing corresponding selection pressures. If similar signs of selection can be detected at the previously identified candidate loci in wild populations it would offer support for their role in influencing observed trends in life history traits.

**Table 3.1:** Number and location of samples utilised in chapter 3.

| Site number | Drainage | River | Course | Years sampled | UTP (20P) | | Nº of fish sampled | Symbol on map |
|---|---|---|---|---|---|---|---|---|
| | | | | | GPS E | GPS N | | |
| 1 | Caroni | Aripo | Upper | 2003 | 693277 | 1181843 | 33 | |
| 2 | Caroni | Aripo | Lower | 2003 | 694410 | 1177783 | 52 | |
| 3 | Caroni | Aripo | Upper | 2006 | 694030 | 1182128 | 26 | |
| 4 | Caroni | Aripo | Lower | 2006 | 694410 | 1177783 | 29 | |
| 5 | Caroni | Caura | Upper | 2003 | 679757 | 1182475 | 16 | |
| 6 | Caroni | Caura | Lower | 2003 | 678436 | 1177236 | 59 | |
| 7 | Caroni | Guanapo | Lower | 2003 | 691385 | 1174569 | 24 | |
| 8 | Caroni | Guanapo | Upper | 2003 | 690247 | 1182015 | 20 | |
| 9 | Caroni | Lopinot | Upper | 2003 | 683520 | 1182443 | 28 | |
| 10 | Caroni | Lopinot | Lower | 2003 | 683553 | 1175663 | 63 | |
| 11 | Marianne | Marianne | Lower | 2003 | 685890 | 1193642 | 26 | |
| 12 | Marianne | Marianne | Upper | 2003 | 685891 | 1192747 | 19 | |
| 13 | Oropuche | Oropuche | Lower | 2006 | 704394 | 1178967 | 12 | |
| 14 | Oropuche | Oropuche | Upper | 2006 | 702534 | 1185310 | 21 | |
| 15 | Oropuche | Turure | Lower | 2003 | 700541 | 1181127 | 27 | |
| 16 | Oropuche | Turure | Upper | 2003 | 700643 | 1181210 | 22 | |
| 17 | Yara | Yarra | Lower | 2006 | 680414 | 1194065 | 24 | |
| 18 | Yara | Yarra | Upper | 2006 | 683427 | 1189518 | 11 | |

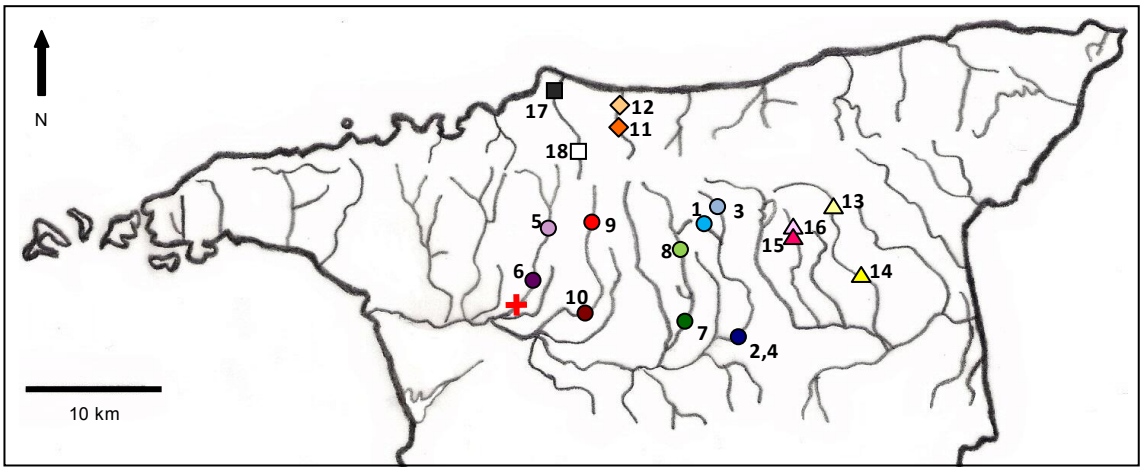



**Figure 3.1:** Map of the sites sampled across Northern Trinidad. ✚ Marks the site the original samples used in the selection experiment were collected from.

## 3.3 Methods

### 3.3.1 Fish sampled

A total of 512 adult male fish were sampled from seventeen sites in Northern Trinidad between 2003 and 2006, encompassing four drainages and eight rivers (Table 3.1, Figure 3.1- see also figure 1.2). Tissue samples and length data for each fish were provided by Cock Van Oosterhaut (UEA) and Jo Cable (Cardiff University). Rivers were chosen based on availability of samples from corresponding upstream and downstream sites in a river. In seven rivers, upstream and downstream sites were separated by large waterfalls which act as a barrier to upstream predator

movement. Although no large waterfalls are present between the upstream and downstream sites in the Oropuche River these sites were separated by approximately 4.14 miles of river containing several smaller waterfalls. Six of the eight rivers sampled exhibited a corresponding variation in predation regime, with high levels of predation found in the downstream sites and low levels in upstream sites. In order to assess any temporal variance in the allele frequencies, the same downstream site in the Aripo River was sampled in both 2003 and 2006. Two different sites in the upstream section of the Aripo were also sampled, one in 2003 and the other in 2006. The two upstream sites were located approximately 0.5 miles away from each other on separate forks of the Aripo River. In addition to the wild fish sampled, 109 fish from the four experimental selection lines were also genotyped. Where possible the $F_6$ fish utilised in the RAD sequencing were measured at two time points approximately 18 months apart. However as two measurements could not be obtained for all fish, these fish ($F_6$ fish from the selection line) were included in the SNP assay in order to try and identify time point one (T1) measurements. The Inclusion of these fish also allowed for levels of genotyping error to be assessed.

The standard length (SL) of each of the wild fish was measured before they were euthanised with an overdose of 0.02% tricaine methanesulfonate (MS222; Pharmaq, Fordingbridge, UK) and preserved in 90-100% ethanol. For all fish included in the assay DNA was extracted in Bangor using the salting out protocol as described by Domingues *et al.* (2010) (see also Appendix V).

### 3.3.2 Selection of SNPs to be genotyped

The SNPs chosen for genotyping wild fish were selected according to their genomic location and links to body size, growth or maturation (table 3.2). SNPs were identified from RAD sequencing on the laboratory selection lines (Chapter 2) and from candidate SNP identified by van Wijk (2011). They fall into one of three groups: (i) SNPs showing signs of selection between the large and small body size experimental selection lines; (ii) SNPs located in candidate genes or linked to body size which had previously been analysed in the experimental selection lines and (iii) SNPs which have not previously shown signs of selection and are therefore putatively neutral. Full details of the criteria used identify SNPs showing signs of selection can be found in sections 2.3 and 2.4.

| SNP name | Chr | SNP position | Type |
|---|---|---|---|
| 36113 | 1 | 2,105,912 | Putatively neutral |
| 313767 | 1 | 32,997,388 | Putatively neutral |
| 70445 | 2 | 5,083,588 | Putatively neutral |
| 91693 | 2 | 37,221,334 | Putatively neutral |
| 207392 | 3 | 17,146,271 | Putatively neutral |
| 294904 | 3 | 33,486,211 | Putatively neutral |
| 220371 | 4 | 23,977,279 | Putatively neutral |
| 173592 * | 5 | 19,813,546 | Putatively neutral |
| 124729 | 7 | 25,808,206 | Putatively neutral |
| 39628 | 8 | 10,002,874 | Putatively neutral |
| 97831 | 8 | 24,802,806 | Putatively neutral |
| 10802 | 9 | 18,715,175 | Putatively neutral |
| 183579 * | 10 | 23,378,889 | Putatively neutral |
| 22946 | 12 | 6,729,263 | Putatively neutral |
| 58198 * | 12 | 11,448,576 | Putatively neutral |
| 59508 | 12 | 17,841,297 | Putatively neutral |
| 60276 | 12 | 19,704,251 | Putatively neutral |
| 337518 * | 12 | 23,777,596 | Putatively neutral |
| 148158 | 12 | 26,336,920 | Putatively neutral |
| 353804 | 13 | 15,015,299 | Putatively neutral |
| 214079 | 14 | 4,331,016 | Putatively neutral |
| 334713 | 15 | 15,934,589 | Putatively neutral |
| 291080 | 16 | 9,689,711 | Putatively neutral |
| 283548 | 16 | 17,645,471 | Putatively neutral |
| 111347 | 17 | 5,616,917 | Putatively neutral |
| 343160 | 17 | 17,202,094 | Putatively neutral |
| 330994 | 18 | 9,293,050 | Putatively neutral |
| 369214 * | 19 | 10,098,638 | Putatively neutral |
| 150841 | 19 | 20,381,279 | Putatively neutral |
| 363926 | 20 | 6,141,039 | Putatively neutral |
| 120249 | 21 | 6,529,099 | Putatively neutral |
| 261690 | 23 | 10,906,454 | Putatively neutral |
| 20521 | 12 | 2,038,395 | RAD selected |
| 21765 | 12 | 4,068,940 | RAD selected |
| 22486 | 12 | 5,603,862 | RAD selected |
| 23539 | 12 | 7,907,396 | RAD selected |
| 57318 | 12 | 13,442,128 | RAD selected |
| 58232 | 12 | 15,466,340 | RAD selected |
| 58352 | 12 | 15,714,984 | RAD selected |
| 58413 | 12 | 15,824,981 | RAD selected |
| 59179 | 12 | 17,157,630 | RAD selected |
| 59448 | 12 | 17,727,953 | RAD selected |
| 108025 | 12 | 20,962,700 | RAD selected |
| 108125 | 12 | 21,151,202 | RAD selected |
| 108291 | 12 | 21,412,323 | RAD selected |
| 97286 | 17 | 30,057,250 | RAD selected |
| Myostatin | 2 | 45,914,011 | Candidate |
| TBC1 | 3 | 12,465,171 | Candidate |
| Prolactin_1 | 8 | 4,723,874 | Candidate |
| GH1 | 8 | 19,282,905 | Candidate |
| GH2_165 | 8 | 19,283,220 | Candidate |
| M009_403 | 12 | 3,584,373 | Candidate |
| MH30_Dreyer | 12 | 6,071,349 | Candidate |
| M1046_2 | 20 | 18,281,277 | Candidate |
| SBF1 | NA | NA | Candidate |

**Table 3.2:** SNPs which were successfully genotyped in all rivers. * denotes a SNP which was monomorphic in all rivers sampled.

Initial analysis identified 34 SNP outliers which were included in the assay. However, following further analysis of the experimental selection lines, 15 of these SNPs were not consistently identified as outliers and therefore were not included in analysis of the wild fish. Therefore, in all fish assayed from the selection lines, a total of 19 SNPs consistently showing signs of selection in the experimental selection lines, as identified with RAD sequencing, were selected. For the remainder of this chapter these SNPs will be referred to as RAD-associated SNPs. The initial study by van Wijk (2011) identified 14 SNPs which were either located within candidate genes for growth and maturation or had previously been identified as being linked to body size in guppies. Following optimisation of the SNP panel, 10 of these 14 SNPs were selected. Hereafter these SNPs will be referred to as candidate SNPs.

All SNPs identified as showing signs of selection from the RAD analysis which were included in the assay (RAD-associated SNPs) are located on the sex chromosome (Chr. 12). Of the SNPs identified by van Wijk (2011) (candidate SNPs), two were located on the sex chromosome, three on chromosome 8 and one each on chromosome 2, 3 and 20. The final candidate SNP which is located in the nuclear receptor 5A1 steroidogenic factor gene (SBF1) could not be assigned a position on the guppy genome.

In order to detect statistical outliers between the up and downstream populations a neutral genetic baseline is required. Therefore, as well as the SNPs described above, 36 putatively neutral SNPs were genotyped in all populations. These SNPs were identified as polymorphic in fish from the selection experiment (Section 2.4), but

showed no significant divergence between the large and small selected lines. While the lack of divergence means these SNPs are not under selection in experimental lines we could not exclude the possibility that they are involved in the other traits which have been shown to vary between high and low predation sites. As most of the SNPs included on the assay are located on the sex chromosome, six of the putatively neutral SNPs we chose were also located on this chromosome. The remaining 30 SNPs were spread across the genome.

### 3.3.3 SNP genotyping and evaluation

The 512 fish were genotyped with the selected SNPs using Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) (Sequenom, San Diego, CA, USA). This technology combines the use of a single base extension followed by mass spectrometry detection to determine the genotype of a SNP. After using PCR to amplify the region of interest a primer is annealed immediately adjacent to the SNP. An allele-specific single base extension (SBE) was then run with mass-modified ddNTPs and MALDI-TOF mass spectrometry to determine the mass of each molecule. The allele-specific SBE reaction means that each DNA sample will have a different mass, according to the allele present, thereby allowing the genotype of a sample to be determined (Griffin and Smith 2000).

To assess the accuracy of the genotyping, datasets from the experimental selection lines obtained both from RAD sequencing (Chapter 2) and the Sequenom assay were compared and any SNPs with different alleles were identified as errors. The error rate for 55 SNPs for all 512 individuals assayed was 16.47%. However, following the removal of four SNPs with an error rate greater than 10%, this dropped to only 3.4%. Any individual or SNP with a missing call rate of greater than 15% was removed from further analysis, leaving a data set with 51 SNPs for the analysis of wild fish populations.

### 3.3.4 Data analysis

Initial examinations of the data were performed by eye in GenAlEx 6.5 (Peakall and Smouse 2006, 2012) and GenePop 4.2 (Raymond and Rousset 1995, Rousset 2008). Mean heterozygosity across all loci was calculated for all sites. Global $F_{st}$ values between upstream and downstream sites within a river, rivers and drainages were calculated in GenePop 4.2. A Principal Components Analysis (PCA) (in full) was undertaken to visualise levels of structuring between the sampled populations. Genotype notations were converted into numbers (0, 1, 2) based on the presence of one allele chosen at random. For example a C/G SNP coded using the C allele would be converted into: G/G= 0, C/G= 1 and C/C= 2. Eigen values and principal component scores were then calculated in MultiVariate Statistical Package 3.1 (MVSP) (Kovach 2007).

### 3.3.5 Detecting SNPs under selection

Under very strong selection a SNP may deviate from Hardy Weinberg equilibrium (HWE) due to a bias in the number of individuals who survive to adulthood. We therefore used GenAlEx 6.5 was used to calculate whether the observed allele frequencies of each SNP were in HWE for each site independently. As the populations sampled could be split hierarchically, an AMOVA, implemented in Arlequin 3.5 (Excoffier and Lischer 2010), was used to assess the amount of variation which could be attributed to each level. Assessment was made at three levels: between rivers; between sites within rivers and within sites. If a SNP is under selection for body size, it would be expected that most of the observed variation would be attributed to differences between the sites within rivers. For SNPs under selection a significant divergence would be expected between upstream and downstream sites within a river. Therefore, $F_{st}$ values were calculated (between upstream and downstream sites within each river) across: all loci; all putatively neutral SNPs; all selected (candidate and RAD) SNPs and each SNP individually.

#### 3.3.5.1 Outlier analysis

In addition to the AMOVA analysis, three tests to identify $F_{st}$ outliers were used because the tests had slight variations in their implementation of this methodology. The first was the method implemented in fdist ('fdist method') which as discussed in Section 2.3 examines the relationship between the heterozygosity and $F_{st}$. It is also able to detect outliers by simulating the boundaries of a neutral distribution (Beaumont and Nichols 1996). The second method used Lositan (Antao *et al.* 2008) and Arlequin (Excoffier *et al.* 2009). One potential problem with the 'fdist method' is the potential for outliers to inflate the simulated neutral boundaries. To address this problem, Lositan runs the simulation twice: the first time the entire data set is used to simulate the neutral distribution and detect outliers while the second time outliers detected in the first run are not used in the simulations of the neutral distribution. As it is possible that the putatively neutral SNPs selected from the analysis of the experimental selection lines are not in fact neutral, using Lositan may help to prevent these SNPs over inflating the neutral distribution. Lositan was run with 50,000 simulations and a forced neutral mean $F_{st}$. Lositan uses a simple infinite island model and where more complex structure exists, its use could lead to false positives being identified. Consequently, we used Arlequin to allow the implementation of a hierarchical structure to detect outliers using 100,000 simulations under a hierarchical island model with 10 groups of 100 demes. For each pairwise comparison (upstream vs. downstream site within a river) SNPs that were monomorphic in both the upstream and downstream site were removed from the outlier analyses.

As both Lositan and Arlequin have been shown be exhibit high type I and type II error rates, the third method we used was a Bayesian approach. As discussed in Section 2.3, this technique examines the likelihood of two models of differentiation, one with and one without selection. Such methodology was implemented in BayeScan using the default parameters (20 pilot runs of 5,000 iterations and an additional burn in of 50,000 iterations followed by 100,000 iterations). BayeScan

also calculates a q-value for each locus which is defined as the minimum FDR at which a locus may become significant. For this analysis any locus with a q-value ≤ 0.05 was considered an outlier.

3.3.5.2 Examination of allele frequencies

In addition to the outlier analyses, allele frequencies of each selected SNP (candidate and RAD-associated) were examined within each population, bootstrapped and the resulting values used to calculate 95% confidence intervals (Uusi-Heikkilä *et al.* 2015). Although it would not be possible to distinguish between patterns resulting from genetic drift or selection, it is unlikely that similar patterns of allelic divergence in both wild populations and the experimental selection lines would result from genetic drift.

3.3.5.3 Allelic association

Tests for significant associations between SL and genotype were performed using the R package SNPassoc [R, version 3.1.1, (R development core team 2014)]. SNPs were tested in each pairwise comparison independently. Each association analysis considered 5 inheritance models (codominant, dominant, recessive, overdominant and log-additive) and the most appropriate model was decided based on the lowest p-value. Where p-values could not distinguish between models, the lowest Aikaike Information Criterion (AIC) was used.

## 3.4 Results

### 3.4.1 Phenotypic variation

In six rivers sampled (including samples from both 2003 and 2006 for the Aripo River: table 3.1) fish from sites with a higher level of predation (downstream sites) had a significantly smaller standard length (SL) than fish sampled from low predation sites (upstream sites) within the same river (figure

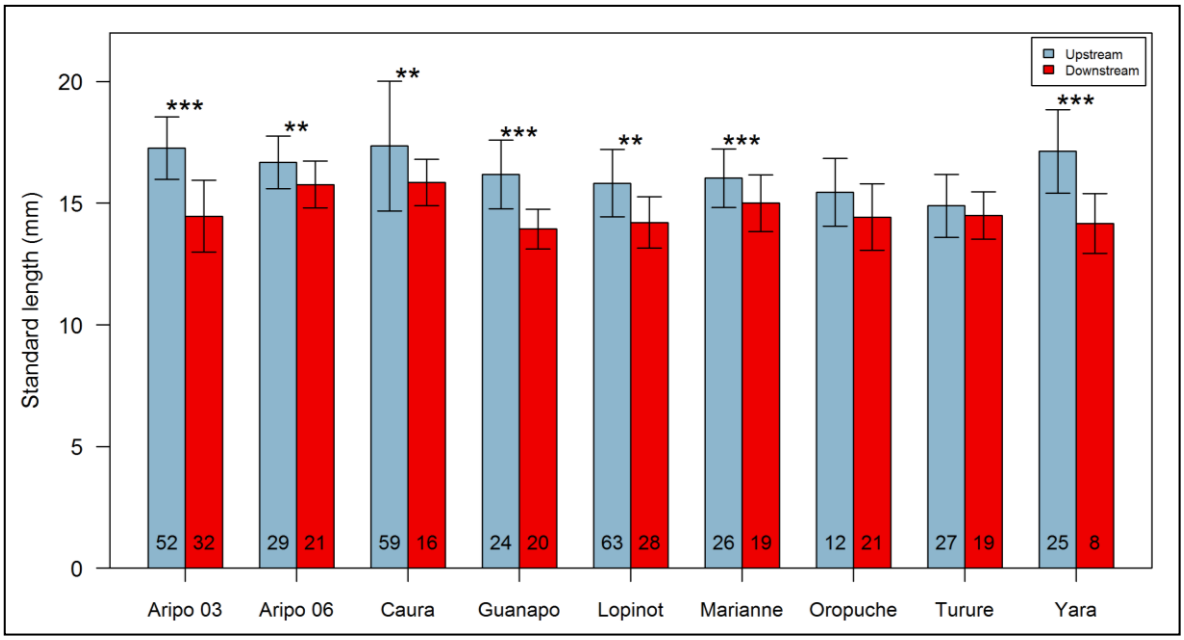



**Figure 3.2:** Mean body size of fish from upstream and downstream sites within each river. Values within each bar show the number of fish successfully genotyped at each site. *** p ≤ 0.001, ** p ≤ 0.01, * p ≤ 0.05.

3.2). No significant difference in SL was observed between upstream and downstream sites for samples collected from the Oropuche and the Turure rivers (Mann Whitney tests: U=80 p=0.089 and U=216 p=0.362 respectively).

### 3.4.2 Evaluation of SNP genotyping

Four fish from the wild populations and three from the selected lines could not be genotyped at any of the SNPs included on the assay, while one fish from the wild populations had 34% of genotypes which could not be scored and was therefore removed from further analysis. Contamination was identified in a further six fish which were consequently removed from further analysis. Three putatively neutral, two RAD associated SNPs and one of the candidate SNPs were not genotyped in any individuals. Following removal of these 14 fish and 6 SNPs the rate of missing genotypes was only 0.67%.

The error rate in the fish from the $F_6$ selection lines, calculated as the percentage of genotypes which were not the same in both the RAD analysis and the Sequenom data set, was 3.4%. However as some error will result from incorrect genotypes in the RAD data set it is expected that the error rate for the fish from the wild populations will be lower. Finally five SNPs which were monomorphic in all of the wild populations sampled were removed from further analysis. After initial examination and cleaning, the data set comprised of genotypes for 28 putatively neutral (Pneu), 14 RAD associated and 9 candidate SNPs (Psel will be used to refer to the combination of the 14 RAD associated and 9 candidate SNPs, i.e. those putatively under selection), in 501 individuals sampled from 8 rivers. The SNPs which were successfully genotyped can be seen in table 3.2 while their genomic locations can be seen in appendix VI.

**Table 3.3:** Percentage of monomorphic SNPs per river.

| River | Total | RAD selected | Candidate selected | Putatively neutral |
|---|---|---|---|---|
| Aripo 03 | 22 | 6 | 0 | 16 |
| Aripo 06 | 30 | 8 | 0 | 22 |
| Caura | 4 | 0 | 0 | 4 |
| Guanapo | 28 | 6 | 0 | 22 |
| Lopinot | 12 | 6 | 0 | 6 |
| Marianne | 64 | 18 | 10 | 36 |
| Oropuhce | 50 | 16 | 2 | 32 |
| Turure | 32 | 10 | 2 | 20 |
| Yara | 42 | 6 | 2 | 34 |

For analysis of upstream and downstream sites within a river only polymorphic loci were included. The average number of monomorphic SNPs per river was 31.6% (range 4-64%) (table 3.3). The largest number of monomorphic SNPs was those thought to be neutral (mean 20.6% per river), while RAD associated and candidate SNPs showed a much higher level of polymorphism (mean number of monomorphic SNPs per river 8.4% and 1.5% respectively). Ascertainment bias is expected to have led to the number of SNPs in the Caura River population being only 4%.

### 3.4.3 Population structure

For analysis of population structure we excluded SNPs at the loci GH1 and Myostatin as both markers deviated from the expectations under Hardy Weinberg Equilibrium.

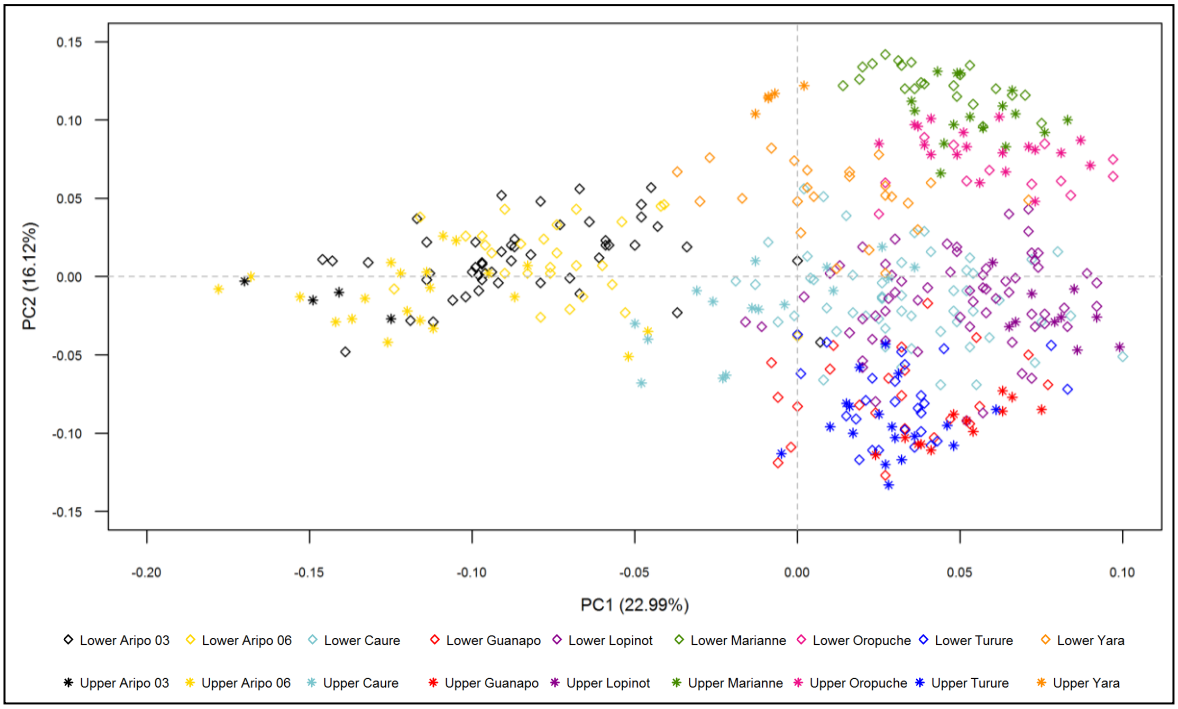



**Figure 3.3:** Principal components analysis of all sites using all SNPs.

PCA and $F_{st}$ values were used to investigate the structuring among and within rivers. PCA broadly grouped rivers according to the geographical separation observed (figure 3.3). One exception are the samples from the Turure which despite being located in the Oropuche drainage clustered with samples from the Caroni drainage, which has previously been identified (Willing *et al.* 2010). Within the Caroni drainage, samples from the Aripo River from both 2003 and 2006 clustered together but were separate from the other rivers in the drainage where clear separation was not observed.

Global $F_{st}$ values showed a high level of differentiation between all rivers (table 3.4). When SNPs were separated according to whether they were putatively neutral or putatively selected, $F_{st}$ values from Psel SNPs were significantly higher than $F_{st}$ values from Pneu SNPs (Mann Whitney test: U=470, p= 0.045) (table 3.4, section B and C). However this may be an artefact of the low levels of polymorphism observed in the Pneu SNPs (mean observed heterozygosity (Ho) in Pneu SNPs= 0.101 ± 0.057, Ho in Psel SNPs= 0.198 ± 0.174). Across all SNPs $F_{st}$ values between rivers at upland sites were significantly higher than those between lowland sites (Mann Whitney test: U= 162, $p < 0.001$) (appendix VII). For all rivers samples, upstream and downstream within a river were more closely related to each other than to other rivers (appendix VII).

Between upland and lowland sites global $F_{st}$ values showed high and significant differentiation ($F_{st} \geq 0.1$) in the Aripo, Caura, Guanapo, Lopinot and Yara sites while differentiation was comparatively low ($F_{st} \leq 0.05$), although still significant, in Marianne (table 3.5, A). No allelic differentiation was detected between the upland and lowland sites in the Oropuche and Turure. PCA utilising all SNPs

**Table 3.4:** Fst values between rivers with upstream and downstream sites combined. (A) All SNPs, (B) putatively neutral SNPs and (C) RAD selected and candidate SNPs. All $F_{st}$ values are highly significant ($p \leq 0.001$). Colours of the river names show the drainage to which they belong (blue=Caroni, red= Marianne, green= Oropuche and purple= Yara). Colours of the values indicate the size of the Fst value (red= maximum and green= minimum).

| (A) | Aripo 03 | Aripo 06 | Caure | Guanapo | Lopinot | Marianne | Oropuche | Turure |
|---|---|---|---|---|---|---|---|---|
| Aripo 06 | 0.02 | | | | | | | |
| Caure | 0.30 | 0.25 | | | | | | |
| Guanapo | 0.44 | 0.42 | 0.25 | | | | | |
| Lopinot | 0.39 | 0.35 | 0.13 | 0.24 | | | | |
| Marianne | 0.49 | 0.48 | 0.35 | 0.59 | 0.38 | | | |
| Oropuche | 0.45 | 0.41 | 0.25 | 0.51 | 0.32 | 0.43 | | |
| Turure | 0.41 | 0.39 | 0.21 | 0.05 | 0.21 | 0.54 | 0.46 | |
| Yara | 0.32 | 0.28 | 0.15 | 0.40 | 0.29 | 0.32 | 0.28 | 0.35 |

| (B) | Aripo 03 | Aripo 06 | Caure | Guanapo | Lopinot | Marianne | Oropuche | Turure |
|---|---|---|---|---|---|---|---|---|
| Aripo 06 | 0.04 | | | | | | | |
| Caure | 0.24 | 0.18 | | | | | | |
| Guanapo | 0.40 | 0.35 | 0.12 | | | | | |
| Lopinot | 0.33 | 0.29 | 0.09 | 0.17 | | | | |
| Marianne | 0.41 | 0.35 | 0.33 | 0.53 | 0.40 | | | |
| Oropuche | 0.38 | 0.32 | 0.27 | 0.46 | 0.32 | 0.45 | | |
| Turure | 0.33 | 0.29 | 0.09 | 0.04 | 0.16 | 0.45 | 0.41 | |
| Yara | 0.35 | 0.27 | 0.15 | 0.39 | 0.25 | 0.18 | 0.34 | 0.33 |

| (C) | Aripo 03 | Aripo 06 | Caure | Guanapo | Lopinot | Marianne | Oropuche | Turure |
|---|---|---|---|---|---|---|---|---|
| Aripo 06 | 0.10 | | | | | | | |
| Caure | 0.32 | 0.28 | | | | | | |
| Guanapo | 0.47 | 0.47 | 0.30 | | | | | |
| Lopinot | 0.41 | 0.38 | 0.15 | 0.27 | | | | |
| Marianne | 0.53 | 0.55 | 0.37 | 0.63 | 0.37 | | | |
| Oropuche | 0.48 | 0.45 | 0.23 | 0.54 | 0.32 | 0.42 | | |
| Turure | 0.45 | 0.44 | 0.27 | 0.05 | 0.24 | 0.59 | 0.50 | |
| Yara | 0.31 | 0.28 | 0.15 | 0.40 | 0.31 | 0.38 | 0.25 | 0.37 |

**Table 3.5:** Fst values between upstream and downstream sites across (a) all SNPs, (B) putatively neutral SNPs and (C) RAD selected and candidate SNPs. *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$. Colours of the indicate the size of Fst value (red= maximum and green= minimum).

| | All (A) | Neutral (B) | Selected (B) |
|---|---|---|---|
| Aripo 03 | 0.30*** | 0.36*** | 0.25*** |
| Aripo 06 | 0.12*** | 0.10*** | 0.14*** |
| Caura | 0.25*** | 0.18*** | 0.29*** |
| Guanapo | 0.18*** | 0.23*** | 0.12*** |
| Lopinot | 0.24*** | 0.17*** | 0.29*** |
| Marianne | 0.06** | 0.06** | 0.04** |
| Oropuhce | 0.01 | 0.01 | 0.00 |
| Turure | 0.01 | 0.00 | 0.02* |
| Yara | 0.22*** | 0.53*** | 0.30*** |

showed distinct clustering between upland and lowland sites in the Caura, Guanapo, Lopinot and Yara (figure 3.4).

To assess the degree of divergence attributable to the Pneu and Psel SNPs, $F_{st}$ values were calculated and PCA run on each set of SNPs independently. If the Psel SNPs were under selection in these wild river populations, the divergent selection acting on the Psel SNPs would lead to the structure identified when utilising only the Psel would be more cl ear-cut than that seen when utilising only the Pneu SNPs. $F_{st}$ values calculated using only Psel were significantly higher than those using only Pneu SNPs in the in the Caura and Lopinot only (U=0.0, p=0.034) (table 3.5, section B and C). For the Guanapo, Yara and Aripo 2003 samples $F_{st}$ values were higher when calculated using only Pneu SNPs, although not significantly (U=2, p=0.275), while values were similar for both sets in the Aripo 2006, Marianne, Oropuche, and Turure (U=7.5, p=0.885). PCA was able to separate between the upland and lowland sites when only utilising the Psel SNPs in two of the rivers (Caura and the Lopinot) (figure 3.4).

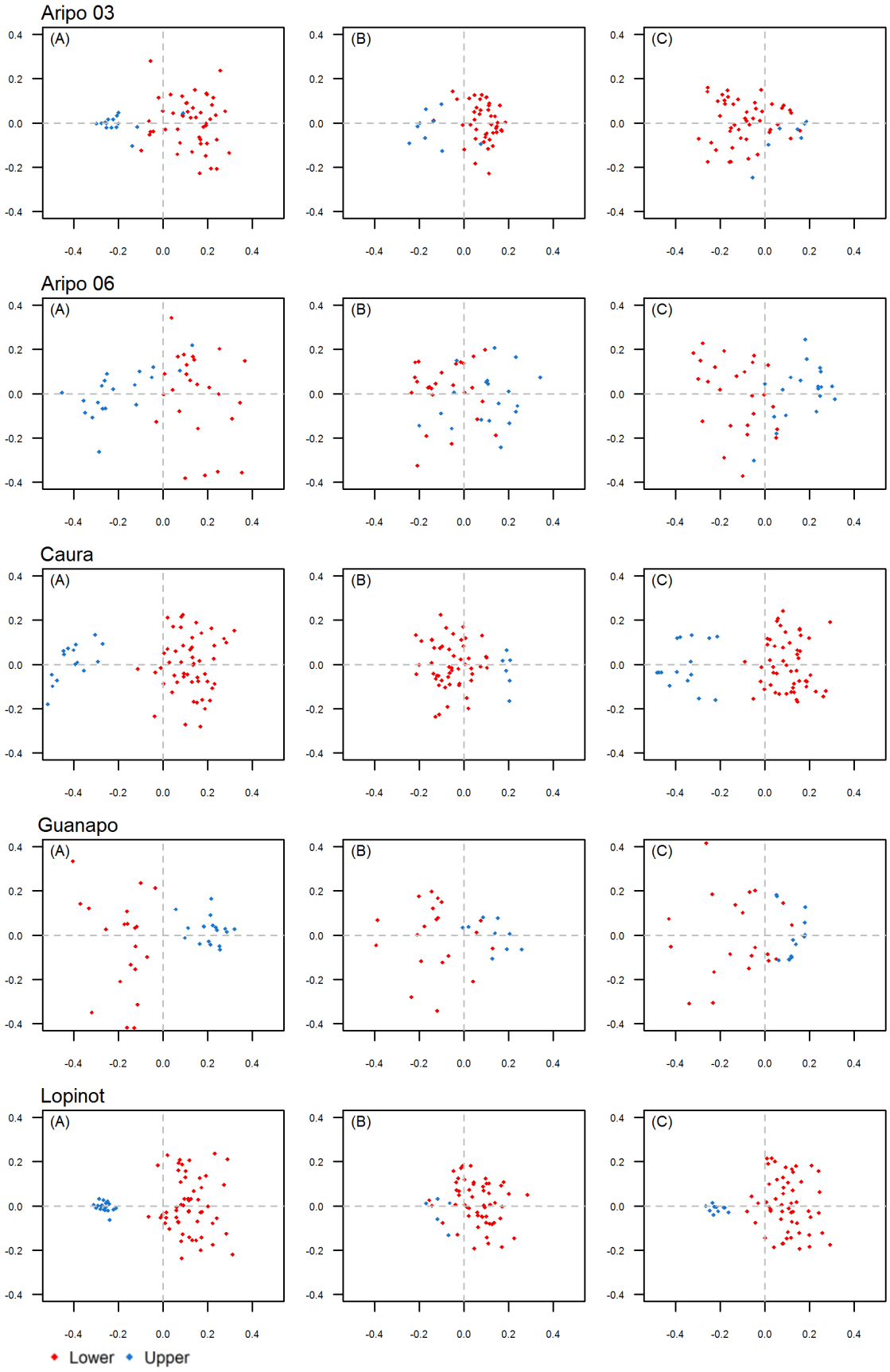



**Figure 3.4:** Principal components analysis of upstream and downstream site within each river (X= PC1 and Y= PC2). (A) all SNPs genotyped, (B) putatively neutral SNPs, (C) RAD selected and candidate SNPs. Percentage of variation accounted by each PC for each analysis can be seen in appendix XII.

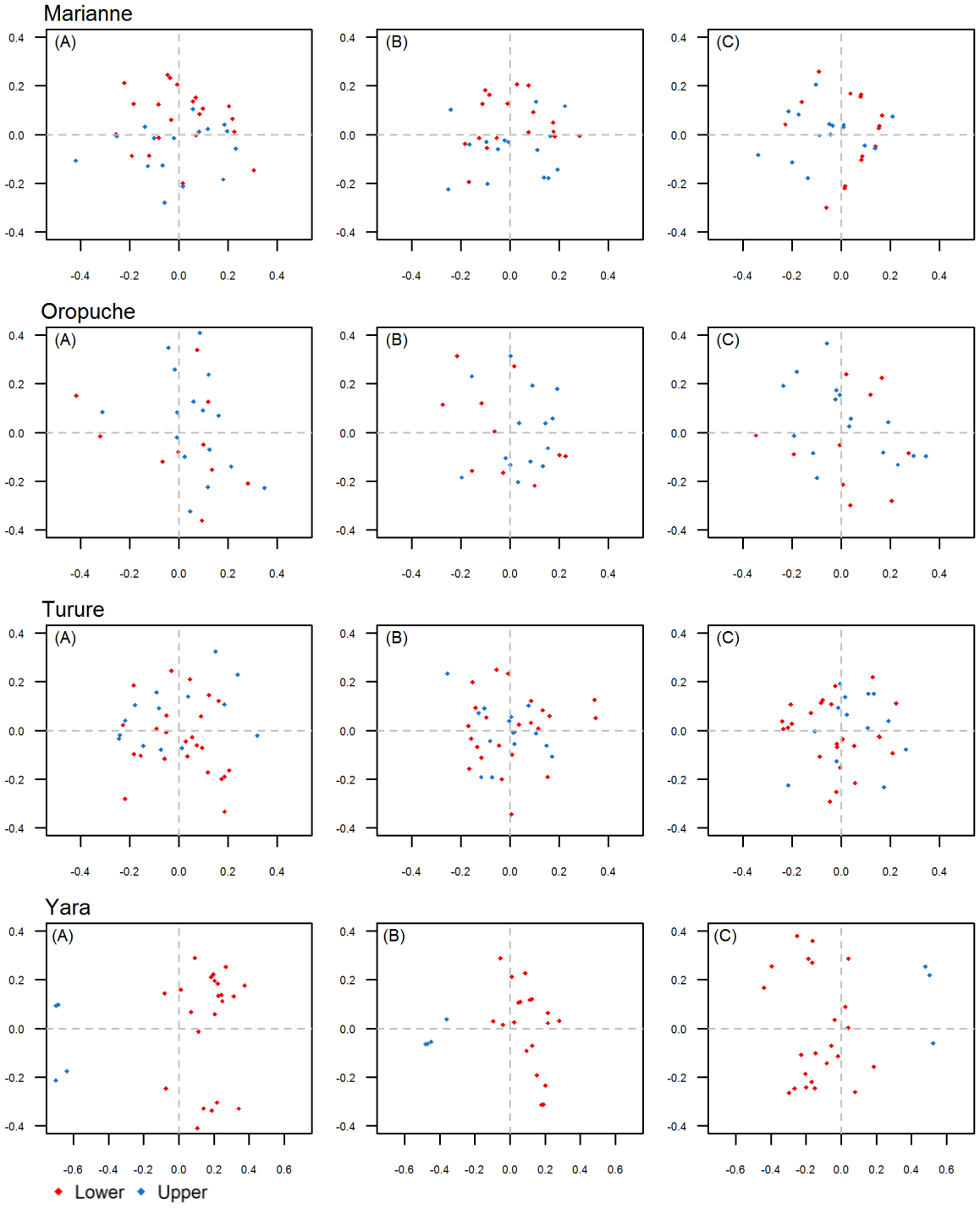



**Figure 3.4:** Continued

**Table 3.6:** Deviations from Hardy Weinberg equilibrium. Grey boxes denote an significant p value (p ≤ 0.01).

| Loci | 36113 | 313767 | 70445 | 91693 | Myostatin | TBC1 | 207392 | 294904 | 220371 | 124729 | Prolactin_1 | 39628 | GH1 | GH2_165 | 97831 | 10802 | 20521 | M009_403 | 21765 | 22486 | MH30_Dreyer | 22946 | 23539 | 57318 | 58232 | 58352 | 58413 | 59179 | 59448 | 59508 | 60276 | 108025 | 108125 | 108291 | 148158 | 353804 | 214079 | 334713 | 291080 | 283548 | 111347 | 343160 | 97286 | 330994 | 150841 | 363926 | M1046_2 | 120249 | 261690 | SBF1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Type | N | N | N | N | SR | SR | N | N | N | N | SR | N | SR | SR | N | N | SL | SR | SL | SL | SR | N | SL | SL | SL | SL | SL | SL | SL | N | N | SL | SL | SL | N | N | N | N | N | N | N | SL | N | N | N | SR | N | N | |
| Chromosome | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 7 | 8 | 8 | 8 | 8 | 8 | 9 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 13 | 14 | 15 | 16 | 16 | 17 | 17 | 17 | 18 | 19 | 20 | 20 | 21 | 23 | |
| BP | 2,105,912 | 32,997,388 | 5,083,588 | 37,221,334 | 45,914,011 | 12,465,171 | 17,146,271 | 33,486,211 | 23,977,279 | 25,808,206 | 4,723,874 | 10,002,874 | 19,282,905 | 19,283,220 | 24,802,806 | 18,715,175 | 2,038,395 | 3,584,373 | 4,068,940 | 5,603,862 | 6,071,349 | 6,729,263 | 7,907,396 | 13,442,128 | 15,466,340 | 15,714,984 | 15,824,981 | 17,157,630 | 17,727,953 | 17,841,297 | 19,704,251 | 20,962,700 | 21,151,202 | 21,412,323 | 26,336,920 | 15,015,299 | 4,331,016 | 15,934,589 | 9,689,711 | 17,645,471 | 5,616,917 | 17,202,094 | 30,057,250 | 9,293,050 | 20,381,279 | 6,141,039 | 18,281,277 | 6,529,099 | 10,906,454 | |
| Lower Aripo 03 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Aripo 06 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Caura | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Guanapo | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Lopinot | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Marianne | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Oropuche | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Turure | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lower Yara | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Aripo 03 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Aripo 06 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Caura | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Guanapo | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Lopinot | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Marianne | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Oropuche | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Turure | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Upper Yara | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

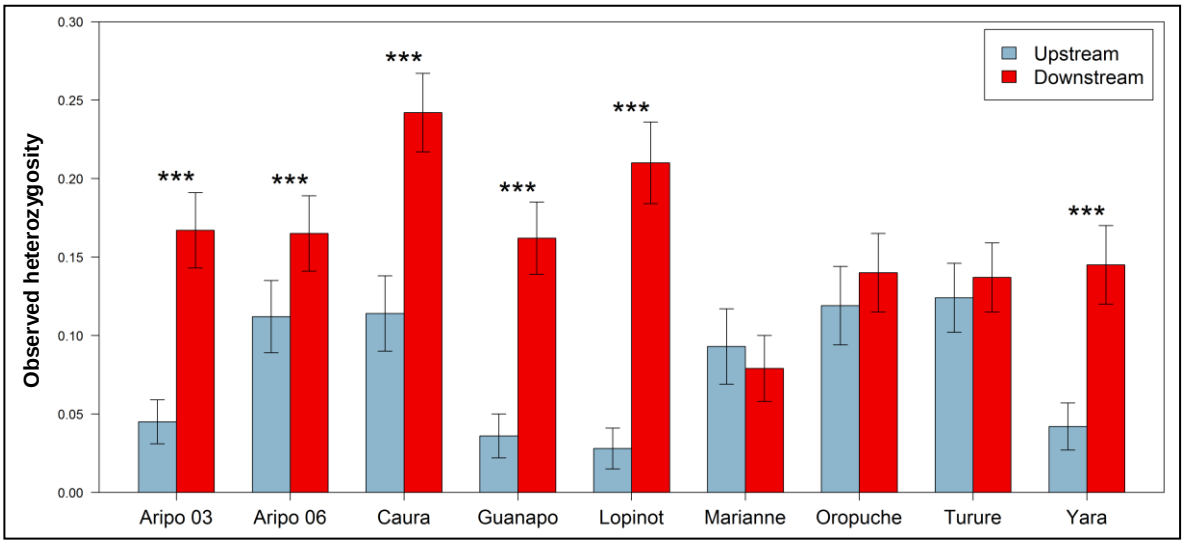



Figure 3.5: Mean observed heterozygosity across all SNPs in the upstream and downstream sites within each river. *** p ≤ 0.001, ** p ≤ 0.01, * p ≤ 0.05.

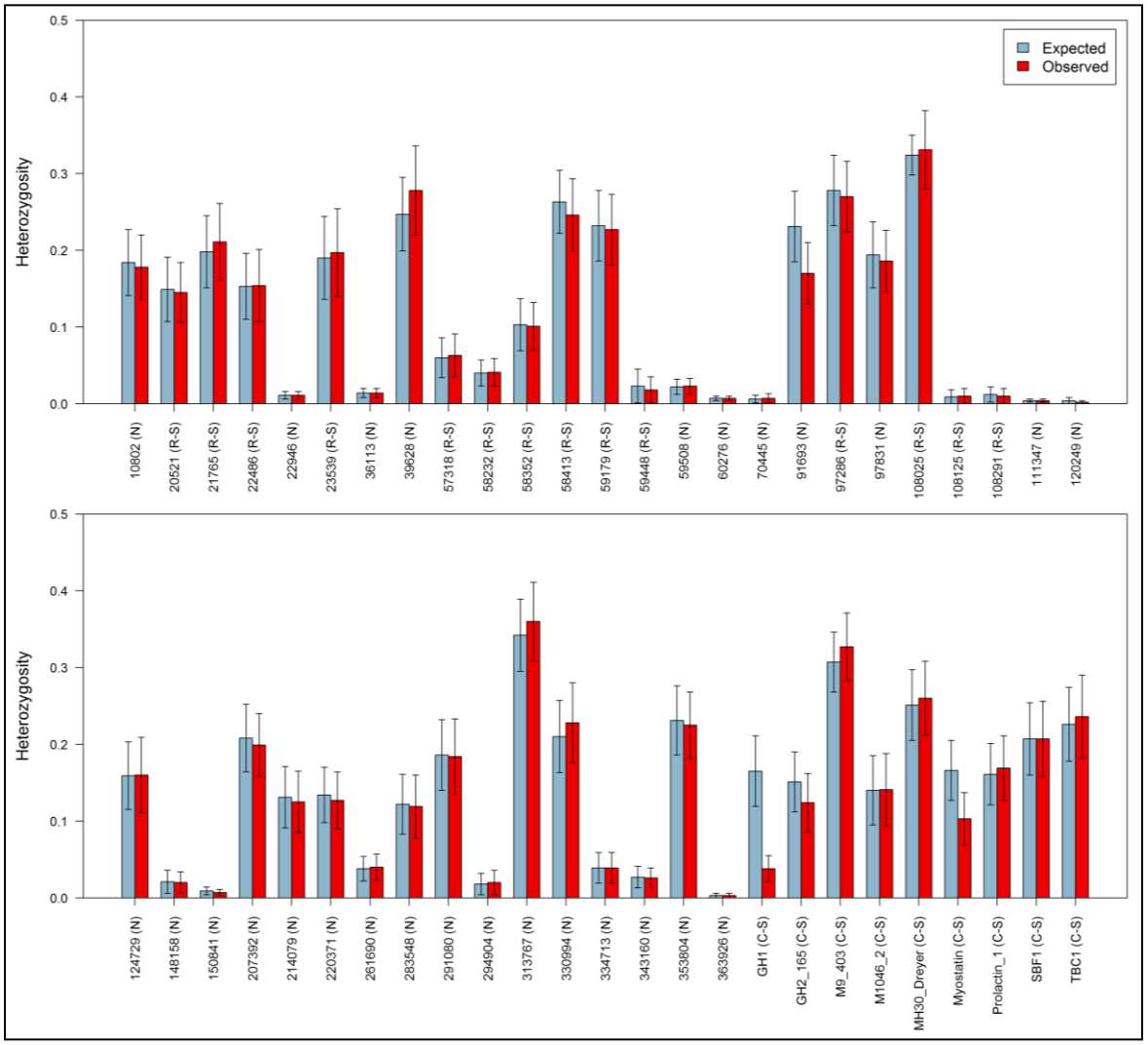



**Figure 3.6:** Expected and observed levels of heterozygosity for the 50 polymorphic SNPs successfully genotyped. R-S = RAD selected SNPs, C-S= Candidate SNPs.

### 3.4.4 Genetic diversity

Genetic diversity was low overall, with a mean expected heterozygosity (He) across all sites of only 0.120 (range 0.028-0.242). Although there was an overall trend for higher He in the lowland sites, these differences were only significant in 6 of the 9 pairs of sites sampled (figure 3.5). The one exception was observed in the Marianne where He in the upland site was higher: however, this difference was not significant and the number of polymorphic SNPs successfully genotyped in the Marianne was low.

Analysis of HWE did not identify any deviations from HWE at the population level (table 3.6). At the locus level, two loci consistently deviated from HWE (GH1 and Myostatin). For both SNPs the deviation from HWE was caused by a deficit of heterozygotes. Deviation from HWE could be an indicator of strong selection so both markers were included in any analysis aimed at identifying selection.

### 3.4.5 Identifying SNPs under selection

To detect SNPs showing signs of selection in the rivers sampled a range of different techniques were used.

#### 3.4.5.1 Hardy Weinberg Equilibrium

As previously mentioned, a SNP under strong selection may deviate from HWE. For example, if selection was sufficiently strong that only fish with allele X at SNP 1 could survive to adulthood, then by the time of sampling the allele frequencies of SNP 1 would not be in HWE. Only two of the SNPs examined exhibited significant and consistent deviations from HWE (Myostatin and GH1). Although the difference between expected and observed heterozygosity were not significant for either, a deficit of heterozygotes was observed for both SNPs (figure 3.6). Deviations from HWE were observed in other SNPs however these were not consistent across sites (table 3.6).

#### 3.4.5.2 $F_{st}$ values

Individual locus $F_{st}$ values were high between upland and lowland sites with an average of 40% (± 23%) of polymorphic SNPs exhibiting significant levels of divergence ($P \leq 0.05$) (table 3.7) and 18% (± 17%) exhibiting highly significant differentiation ($P \leq 0.001$). However, as can be seen by the global $F_{st}$ values (table 3.5), high levels of genetic differentiation were present in both putatively neutral and selected SNPs.

SNPs associated with body size should exhibit an elevated $F_{st}$ in rivers only with a corresponding difference in SL. Considering only the rivers with significant differences in SL, ten SNPs consistently had a significant $F_{st}$ ($p \leq 0.05$ in 5 out of 7 comparisons), however, five of these were putatively neutral SNPs. High and significant levels of differentiation were consistently observed in some of the putatively neutral SNPs, for example 313767 and 39628 (table 3.7). While these SNPs

**Table 3.7:** Individual loci $F_{st}$ values between upstream and downstream sites within a river. Blank cells denote a SNP which is monomorphic in the respective river pair. Red cells denote very highly significant $F_{st}$ ($P \leq 0.001$), amber cells denote highly significant $F_{st}$ ($P \leq 0.01$), green cells denote significant $F_{st}$ ($P \leq 0.05$) and grey cells denote an insignificant $F_{st}$.

| Locus | Aripo 03 | | Aripo 06 | | Caura | | Guanapo | | Lopinot | | Marianne | | Oropuche | | Turure | | Yara | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value | $F_{st}$ | P value |
| 10802 | | | 0.03 | 0.253 | 0.19 | 0.001 | 0.00 | 1.000 | 0.10 | 0.006 | 0.00 | 0.420 | 0.00 | 1.000 | 0.00 | 0.744 | 0.61 | 0.000 |
| 20521 | 0.04 | 0.076 | 0.07 | 0.034 | 0.38 | 0.000 | 0.02 | 0.495 | 0.07 | 0.019 | | | 0.00 | 0.613 | 0.00 | 1.000 | 0.35 | 0.002 |
| 21765 | 0.51 | 0.000 | 0.42 | 0.000 | 0.43 | 0.000 | 0.16 | 0.004 | 0.04 | 0.052 | | | | | 0.00 | 0.829 | 0.03 | 1.000 |
| 22486 | 0.03 | 0.152 | 0.00 | 1.000 | 0.26 | 0.000 | 0.00 | 1.000 | 0.56 | 0.000 | 0.04 | 0.178 | 0.00 | 0.432 | 0.00 | 1.000 | 0.06 | 1.000 |
| 22946 | 0.00 | 0.571 | | | 0.00 | 0.575 | | | 0.00 | 1.000 | 0.00 | 1.000 | | | | | | |
| 23539 | 0.33 | 0.000 | 0.11 | 0.013 | 0.35 | 0.000 | 0.15 | 0.004 | 0.07 | 0.012 | | | | | | | 0.40 | 0.001 |
| 36113 | | | 0.00 | 1.000 | 0.00 | 0.567 | 0.01 | 0.493 | 0.00 | 0.571 | | | | | 0.00 | 1.000 | | |
| 39628 | 0.18 | 0.000 | 0.11 | 0.031 | 0.30 | 0.000 | 0.23 | 0.001 | 0.10 | 0.001 | 0.06 | 0.034 | 0.08 | 0.124 | 0.00 | 0.153 | 0.04 | 0.293 |
| 57318 | | | 0.00 | 1.000 | 0.00 | 0.797 | 0.00 | 1.000 | 0.02 | 0.174 | 0.01 | 0.417 | | | | | 0.04 | 0.414 |
| 58232 | 0.00 | 1.000 | | | 0.06 | 0.072 | 0.00 | 1.000 | 0.06 | 0.026 | | | | | 0.00 | 1.000 | 0.03 | 1.000 |
| 58352 | 0.00 | 1.000 | 0.01 | 0.501 | 0.43 | 0.000 | 0.02 | 0.490 | 0.09 | 0.002 | | | | | 0.00 | 1.000 | 0.10 | 0.130 |
| 58413 | 0.00 | 0.835 | 0.00 | 0.658 | 0.11 | 0.015 | 0.00 | 1.000 | 0.43 | 0.000 | 0.05 | 0.113 | 0.04 | 0.224 | 0.00 | 0.423 | 0.13 | 0.056 |
| 59179 | 0.08 | 0.007 | 0.03 | 0.213 | 0.22 | 0.001 | 0.04 | 0.169 | 0.30 | 0.000 | | | 0.00 | 0.412 | 0.12 | 0.006 | 0.09 | 0.152 |
| 59448 | | | | | 0.42 | 0.000 | | | | | | | | | | | | |
| 59508 | 0.00 | 1.000 | | | | | | | 0.00 | 1.000 | 0.00 | 1.000 | | | 0.00 | 1.000 | 0.08 | 0.256 |
| 60276 | 0.00 | 0.523 | 0.00 | 1.000 | 0.00 | 1.000 | | | 0.00 | 1.000 | | | | | | | | |
| 70445 | | | | | 0.01 | 0.328 | | | 0.00 | 1.000 | | | | | | | | |
| 91693 | 0.11 | 0.007 | 0.00 | 0.572 | 0.63 | 0.000 | 0.26 | 0.000 | 0.05 | 0.050 | | | 0.08 | 0.128 | 0.02 | 0.832 | 0.05 | 0.376 |
| 97286 | 0.33 | 0.000 | 0.15 | 0.006 | 0.49 | 0.000 | 0.05 | 0.120 | 0.21 | 0.000 | 0.00 | 1.000 | 0.03 | 0.363 | 0.09 | 0.250 | 0.44 | 0.000 |
| 97831 | 0.11 | 0.003 | 0.15 | 0.003 | 0.03 | 0.190 | 0.21 | 0.003 | 0.00 | 0.545 | 0.03 | 0.153 | 0.00 | 0.456 | 0.00 | 0.493 | 0.29 | 0.011 |
| 108025 | 0.04 | 0.074 | 0.12 | 0.011 | 0.32 | 0.000 | 0.00 | 0.760 | 0.33 | 0.000 | 0.03 | 0.224 | 0.00 | 1.000 | 0.00 | 1.000 | 0.37 | 0.001 |
| 108125 | | | | | 0.18 | 0.008 | | | | | | | | | | | | |
| 108291 | 0.00 | 1.000 | | | 0.18 | 0.008 | | | | | | | | | | | | |
| 111347 | 0.00 | 1.000 | | | 0.00 | 1.000 | | | 0.00 | 1.000 | | | | | | | | |
| 120249 | | | | | 0.00 | 0.731 | | | | | | | | | | | | |
| 124729 | 0.60 | 0.000 | 0.34 | 0.000 | 0.06 | 0.056 | 0.21 | 0.002 | 0.02 | 0.180 | | | | | 0.00 | 0.675 | 0.00 | 0.705 |
| 148158 | | | | | 0.06 | 0.060 | | | 0.06 | 0.043 | | | | | | | | |
| 150841 | | | | | 0.00 | 0.724 | 0.00 | 1.000 | 0.00 | 0.548 | | | | | | | | |
| 207392 | 0.16 | 0.000 | 0.12 | 0.015 | 0.10 | 0.019 | 0.26 | 0.001 | 0.12 | 0.001 | | | 0.00 | 0.382 | 0.00 | 0.716 | | |
| 214079 | 0.05 | 0.021 | 0.00 | 0.505 | 0.00 | 0.696 | 0.07 | 0.114 | 0.28 | 0.000 | | | | | 0.00 | 0.335 | | |
| 220371 | | | 0.00 | 1.000 | 0.10 | 0.018 | 0.00 | 1.000 | 0.06 | 0.014 | 0.02 | 0.156 | 0.00 | 1.000 | 0.02 | 1.000 | 0.53 | 0.000 |
| 261690 | 0.04 | 0.060 | 0.00 | 0.437 | 0.01 | 0.351 | | | 0.07 | 0.013 | | | | | | | | |
| 283548 | 0.00 | 0.524 | | | 0.08 | 0.016 | 0.19 | 0.004 | 0.14 | 0.000 | | | 0.13 | 0.030 | 0.00 | 0.785 | | |
| 291080 | 0.58 | 0.000 | 0.13 | 0.008 | 0.07 | 0.041 | 0.25 | 0.001 | 0.09 | 0.004 | 0.01 | 0.425 | 0.00 | 1.000 | 0.00 | 0.061 | 0.36 | 0.000 |
| 294904 | 0.00 | 0.514 | 0.00 | 1.000 | | | 0.11 | 0.012 | | | | | | | | | | |
| 313767 | 0.56 | 0.000 | 0.09 | 0.028 | 0.24 | 0.000 | 0.44 | 0.000 | 0.39 | 0.000 | 0.17 | 0.005 | 0.01 | 0.292 | 0.00 | 1.000 | 0.08 | 0.147 |
| 330994 | 0.12 | 0.001 | 0.06 | 0.045 | 0.00 | 0.384 | 0.11 | 0.025 | 0.20 | 0.000 | | | 0.08 | 0.126 | 0.00 | 0.123 | | |
| 334713 | 0.00 | 1.000 | | | 0.03 | 0.253 | | | 0.11 | 0.002 | | | | | 0.06 | 0.114 | | |
| 343160 | 0.00 | 1.000 | | | 0.05 | 0.085 | | | 0.02 | 0.262 | | | | | 0.01 | 0.297 | | |
| 353804 | 0.08 | 0.007 | 0.00 | 0.423 | 0.18 | 0.001 | 0.10 | 0.024 | 0.24 | 0.000 | 0.09 | 0.042 | 0.00 | 0.831 | 0.02 | 1.000 | 0.91 | 0.000 |
| 363926 | | | | | 0.00 | 0.596 | | | | | | | | | | | | |
| GH1 | 0.27 | 0.000 | 0.16 | 0.029 | 0.19 | 0.003 | 0.03 | 0.500 | 0.09 | 0.018 | | | 0.00 | 1.000 | | | 0.08 | 0.906 |
| GH2_165 | 0.05 | 0.047 | 0.05 | 0.132 | 0.36 | 0.000 | 0.17 | 0.005 | 0.28 | 0.000 | 0.00 | 0.712 | 0.18 | 0.043 | 0.06 | 0.748 | 0.09 | 0.118 |
| M1046_2 | 0.23 | 0.000 | 0.11 | 0.027 | 0.03 | 0.134 | 0.06 | 0.116 | 0.02 | 0.167 | | | | | | | 0.23 | 0.008 |
| M9_403 | 0.39 | 0.000 | 0.10 | 0.019 | 0.24 | 0.000 | 0.14 | 0.013 | 0.14 | 0.000 | 0.00 | 0.608 | 0.00 | 0.785 | 0.03 | 0.308 | 0.04 | 1.000 |
| MH30 | 0.15 | 0.000 | 0.05 | 0.122 | 0.29 | 0.000 | 0.00 | 1.000 | 0.34 | 0.000 | 0.01 | 0.421 | 0.00 | 0.803 | 0.00 | 0.700 | 0.53 | 0.000 |
| Myostatin | 0.05 | 0.076 | 0.00 | 0.494 | 0.10 | 0.007 | 0.08 | 0.114 | 0.27 | 0.000 | | | 0.00 | 0.802 | 0.00 | 1.000 | | |
| Prolactin_1 | 0.10 | 0.007 | 0.10 | 0.030 | 0.38 | 0.000 | 0.30 | 0.000 | 0.26 | 0.000 | | | 0.08 | 0.123 | 0.00 | 0.511 | 0.04 | 0.297 |
| SBF1 | 0.16 | 0.000 | 0.00 | 0.381 | 0.01 | 0.450 | 0.12 | 0.010 | 0.00 | 0.545 | 0.18 | 0.007 | 0.04 | 0.122 | 0.06 | 1.000 | 0.39 | 0.002 |
| TBC1 | 0.25 | 0.000 | 0.23 | 0.000 | 0.09 | 0.017 | 0.21 | 0.002 | 0.36 | 0.000 | | | 0.00 | 0.729 | 0.08 | 0.066 | 0.31 | 0.006 |

**Table 3.8:** Outlying SNPs as identified by Lositan and Arlequin and genetic variation attributable to variation between rivers within sites from a hierarchical AMOVA analysis. Blue boxes mark an significant outlier ($p \leq 0.01$) in Lositan only. Green boxes mark an significant outlier ($p \leq 0.01$) in Arelquin only. Red boxes mark an significant outlier ($p \leq 0.01$) in both Lositan and Arelquin.

| Locus | Type | Aripo 03 | Aripo 06 | Caura | Guanapo | Lopinot | Marianne | Oropuche | Turure | Yara | AMOVA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 36113 | N | | | | | | | | | | 0 |
| 313767 | N | | | | | | | | | | 33 |
| 70445 | N | | | | | | | | | | 0 |
| 91693 | N | | | | | | | | | | 16 |
| Myostatin | C-S | | | | | | | | | | 5 |
| TBC1 | C-S | | | | | | | | | | 19 |
| 207392 | N | | | | | | | | | | 11 |
| 294904 | N | | | | | | | | | | 9 |
| 220371 | N | | | | | | | | | | 11 |
| 124729 | N | | | | | | | | | | 37 |
| Prolactin_1 | C-S | | | | | | | | | | 12 |
| 39628 | N | | | | | | | | | | 13 |
| GH1 | C-S | | | | | | | | | | 9 |
| GH2_165 | C-S | | | | | | | | | | 11 |
| 97831 | N | | | | | | | | | | 7 |
| 10802 | N | | | | | | | | | | 15 |
| 20521 | R-S | | | | | | | | | | 17 |
| M9_403 | C-S | | | | | | | | | | 12 |
| 21765 | R-S | | | | | | | | | | 31 |
| 22486 | R-S | | | | | | | | | | 20 |
| MH30_Dreyer | C-S | | | | | | | | | | 16 |
| 22946 | N | | | | | | | | | | -1 |
| 23539 | R-S | | | | | | | | | | 15 |
| 57318 | R-S | | | | | | | | | | 0 |
| 58232 | R-S | | | | | | | | | | 3 |
| 58352 | R-S | | | | | | | | | | 21 |
| 58413 | R-S | | | | | | | | | | 15 |
| 59179 | R-S | | | | | | | | | | 14 |
| 59448 | R-S | | | | | | | | | | 41 |
| 59508 | N | | | | | | | | | | 0 |
| 60276 | N | | | | | | | | | | -1 |
| 108025 | R-S | | | | | | | | | | 18 |
| 108125 | R-S | | | | | | | | | | 15 |
| 108291 | R-S | | | | | | | | | | 9 |
| 148158 | N | | | | | | | | | | 5 |
| 353804 | N | | | | | | | | | | 15 |
| 214079 | N | | | | | | | | | | 5 |
| 334713 | N | | | | | | | | | | 7 |
| 291080 | N | | | | | | | | | | 22 |
| 283548 | N | | | | | | | | | | 11 |
| 111347 | N | | | | | | | | | | -1 |
| 343160 | N | | | | | | | | | | 3 |
| 97286 | R-S | | | | | | | | | | 24 |
| 330994 | N | | | | | | | | | | 4 |
| 150841 | N | | | | | | | | | | 0 |
| 363926 | N | | | | | | | | | | -1 |
| M1046_2 | C-S | | | | | | | | | | 6 |
| 120249 | N | | | | | | | | | | 0 |
| 261690 | N | | | | | | | | | | 3 |
| SBF1 | C-S | | | | | | | | | | 7 |

were not under selection in the experimental lines studied, it is possible that they are influenced by selection in the wild populations.

### 3.4.5.3 $F_{st}$ outlier analysis: Lositan

Lositan identified 32 SNPs outlier SNPs ($p \leq 0.05$) in at least one comparison of upland and lowland sites within a river (Table 3.8). The largest number of outliers was found in the Lopinot and Caura, (13 and 12 SNPs respectively). More than half (8 in each river) of the outliers in these two rivers were putatively neutral. Seven SNPs were outliers in the Aripo 2003 samples, while the number of outlying SNPs was much lower in the following rivers; Aripo 2006, Guanapo, Oropuche, Turure and Yara (mean = 2.4). Lositan did not detect any outliers in the Marianne (table 3.8). Across all the rivers sampled, no SNP was consistently identified as an outlier (one SNP was identified as an

outlier in 4 comparisons and 9 were identified as an outlier in 2 comparisons). A full table showing observed He, $F_{st}$ and p values obtained from Lositan can be found in appendix VIII.

### 3.4.5.4 $F_{st}$ outlier analysis: Arlequin

The Fdist approach, as implemented in Arlequin, identified 11 SNPs which were outliers, however none of these was an outlier in more than one pairwise comparison (upstream vs. downstream site within a river). The largest number of outliers identified by Arlequin was in the Oropuche (6) while no outliers were found in the Caura, Lopinot, Marianne and Yara comparisons. The mean number of outliers found in the remaining rivers (Aripo 2003, Aripo 2006, Guanapo, and Turure) was only 1.3. A full table showing observed He, $F_{st}$ and p values obtained from

Arlequin can be found in appendix IX.

3.4.5.5 $F_{st}$ outlier analysis: BayeScan

The Bayesian analysis, implemented in BayeScan, did not identify any outliers ($Q \leq 0.05$) in any of the rivers studied, with the lowest Q value observed being 0.8. A full table showing the posterior probability, logarithm of posterior odds alpha, Fst and Q values obtained from BayeScan can be found in appendix X.

3.4.5.6 AMOVA

Based on a hierarchical AMOVA, the majority of the genetic variability observed could be attributed to '*within site*' variation for 42 of the 50 SNPs and variation '*between rivers*' for the remaining 8 SNPs (table 3.8). Although the majority of variability observed was thus not attributable to variation between sites within a river in any SNP, such variation accounted for at least 20% of the total variation in 4 selected and 4 putatively neutral SNPs.

Negative variance components were observed in 6 SNPs at the '*between rivers*' level and 8 SNPs at the '*between sites within rivers*' level (table 3.8 and appendix XI). For SNPs where the negative value was attributable to variation '*between sites within rivers*', the accompanying p values were not significant and therefore indicate a lack of structure at this level for these SNPs. Values which are significantly negative imply that an allele is more related between populations than within them (Weir and Cockerham 1984, Weir 1996). The negative values which were attributed to variation '*between rivers*' were significant and, therefore, these values indicate that these SNPs were more similar between different rivers than within sites from the same river.

3.4.5.7 Allele frequency changes

Figure 3.7 shows the patterns of allele frequencies in the upland and lowland sites for each river as well as those SNPs which showed a significant allelic association with SL. Analysis of the allele frequency confidence intervals did not show consistent variation between upland and lowland sites in the rivers sampled for any of the selected SNPs studied. Such a lack of consistency indicates that none of the SNPs examined are under selection for body size in the rivers sampled here.

Allele frequency confidence intervals in the SNP Prolactin showed a similar pattern across 5 of the 7 rivers where it was polymorphic, and very low levels of polymorphism in the remaining two (figure 3.7). At the Prolactin SNP the pattern of allelic variation observed between rivers was also very similar to the pattern of variation observed in the selection lines. While some of the remaining SNPs showed similar patterns of allelic variation between the selection lines and upland and lowland sites within rivers, this pattern was found in no more than 3 of the 9 pairs of sites for any given SNP (for example 23539 and 21765).

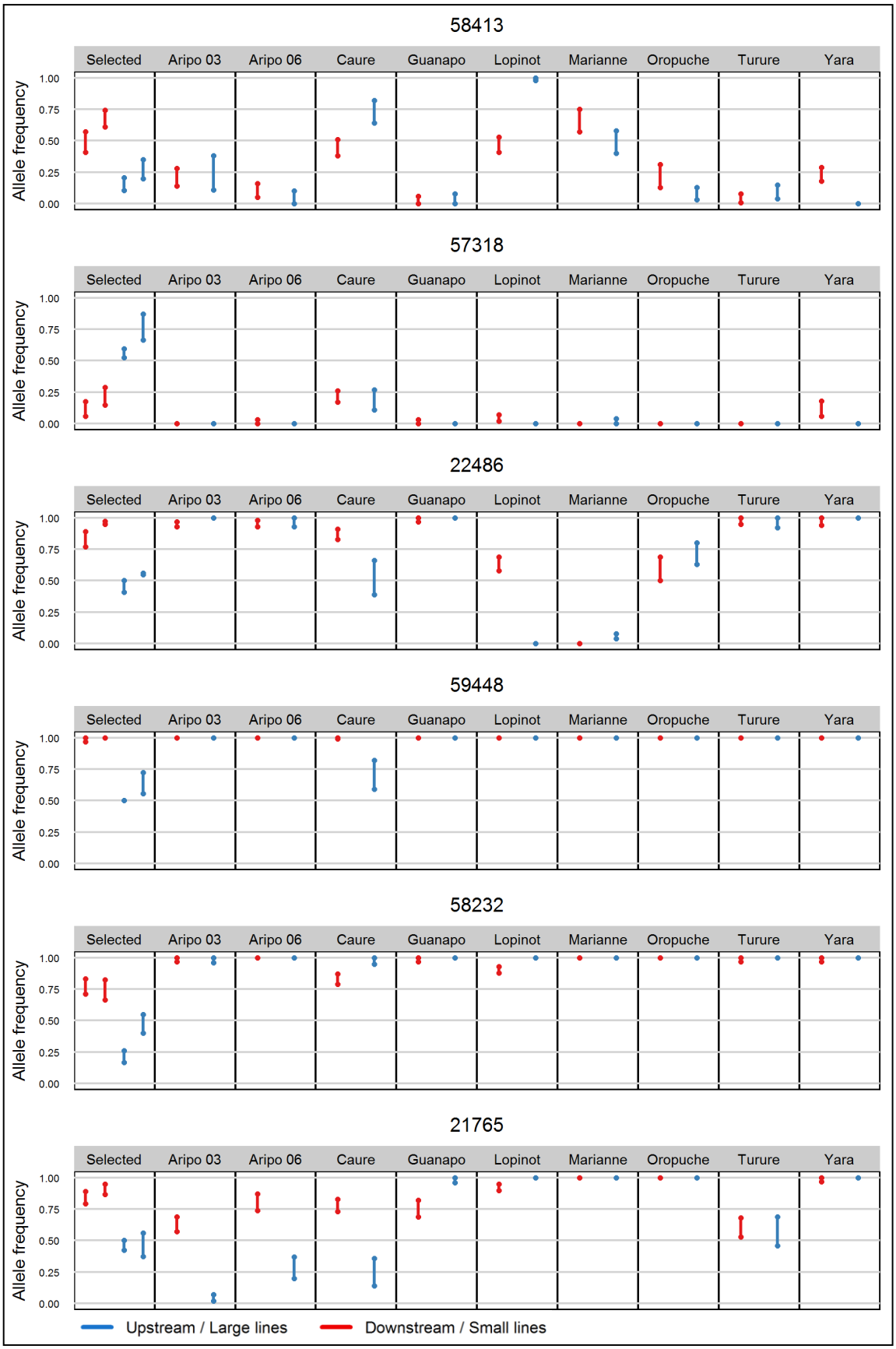



**Figure 3.7:** Allele frequency confidence intervals from upstream and downstream sites within each river and the experimental selection lines. In the river sites blue lines show the values for the upstream sites where the fish are typically larger while red lines show the downstream sites where the fish are typically smaller. In the experimental selection lines blue show the large selected lines (L1 and L2) and red show the small selected lines (S1 and S2).

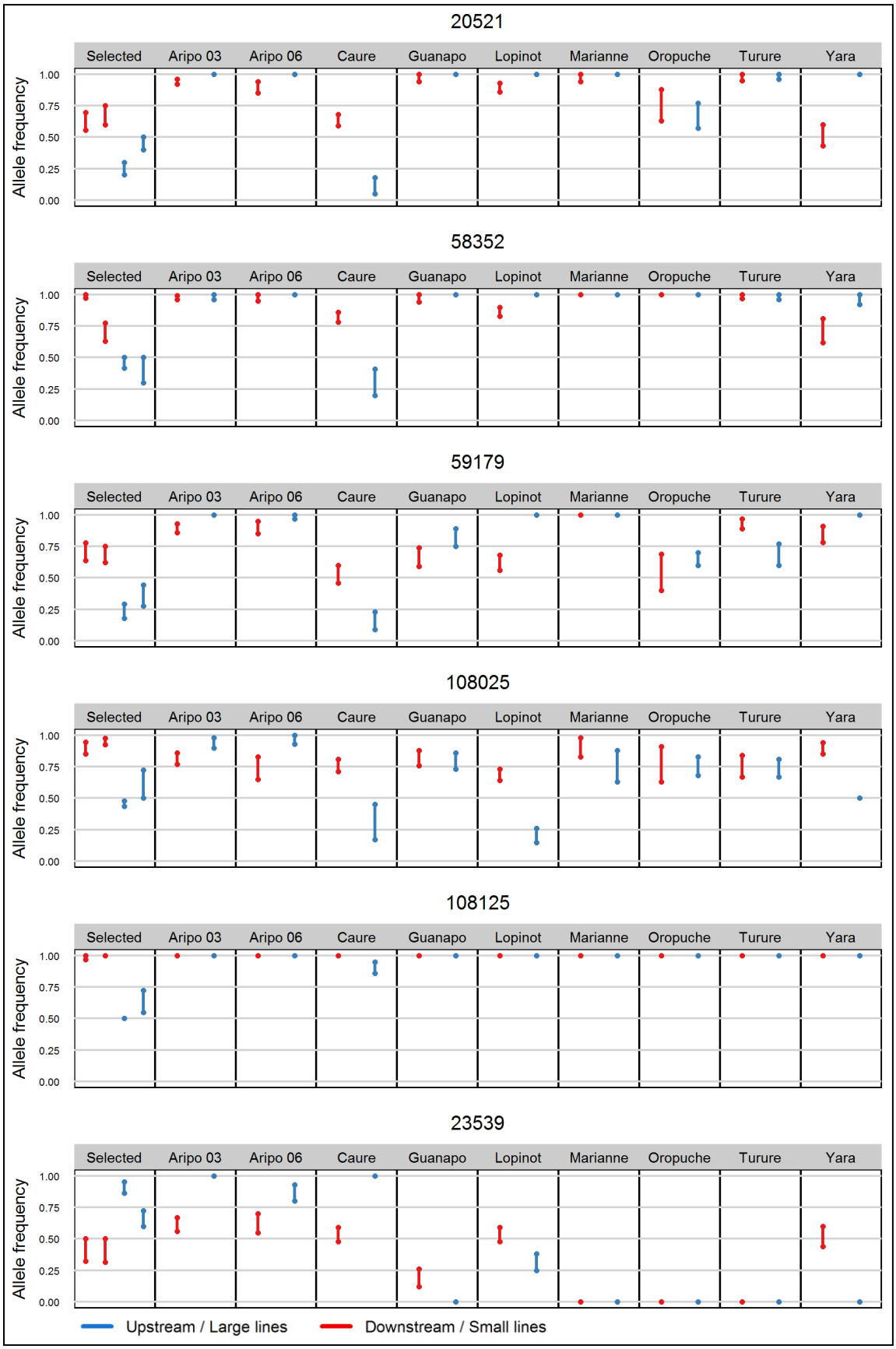



**Figure 3.7:** Continued

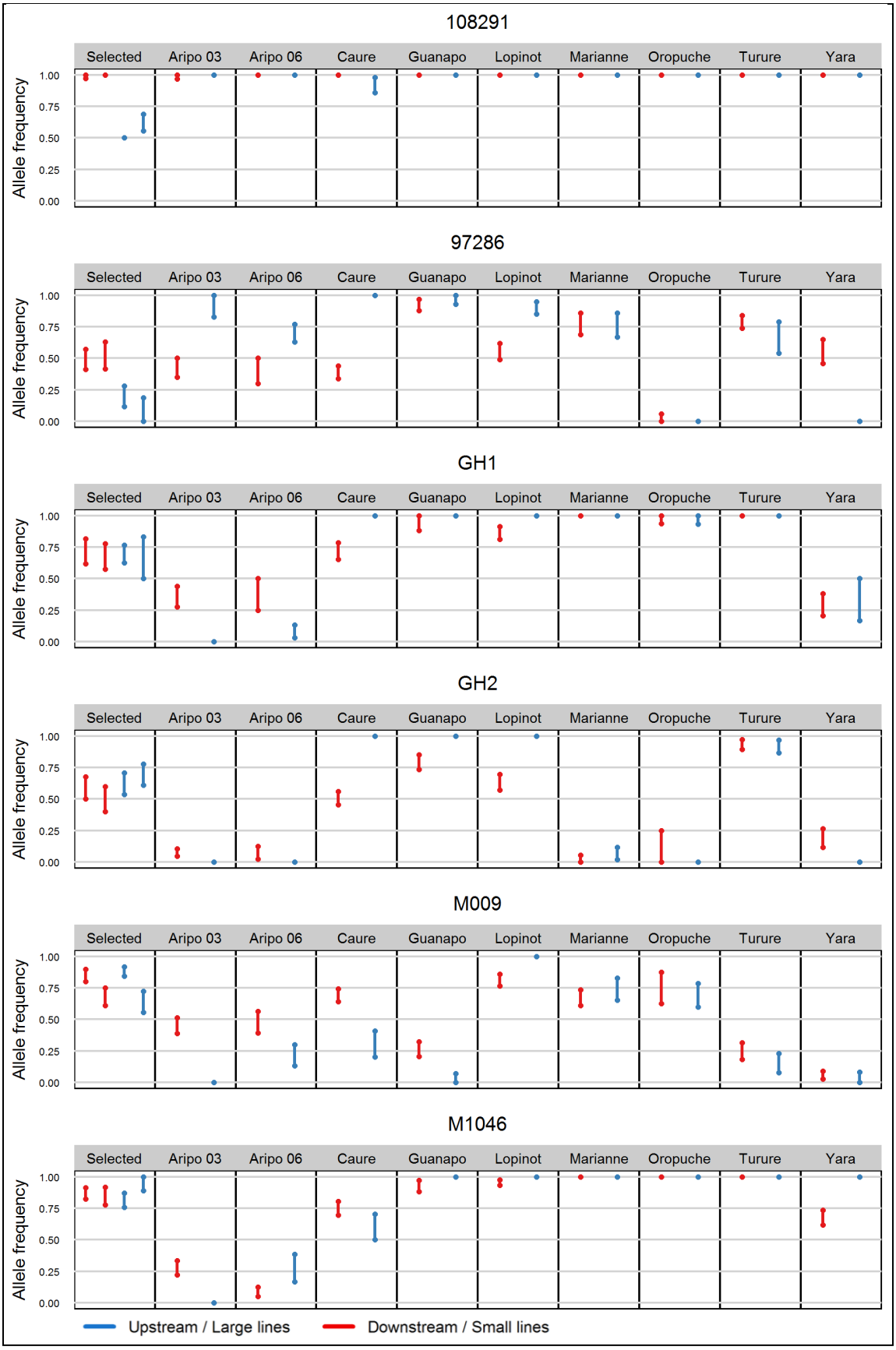



**Figure 3.7:** Continued

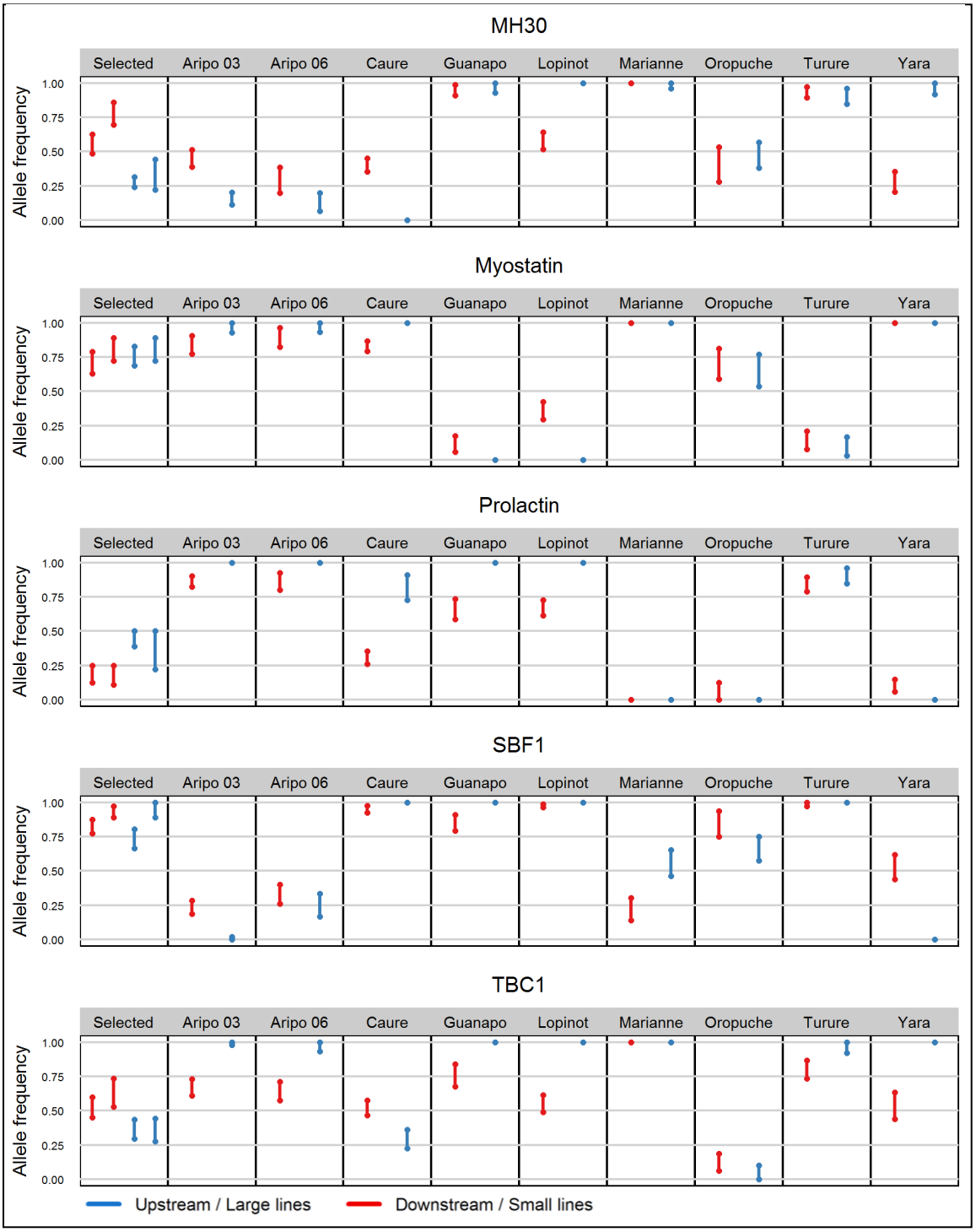



**Figure 3.7:** Continued

**Table 3.9:** SNPs showing an association with SL. Only values where the association is significant are shown. Inheritance models are : CD= codominant; D= Dominant; OD= overdominant; R= recessive LA= log additive. No significant associations were found in the Oropuche.

| | Aripo 03 | | | Aripo 06 | | | Caura | | | Guanapo | | | Lopinot | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model | P value | AIC | Model | P value | AIC | Model | P value | AIC | Model | P value | AIC | Model | P value | AIC |
| 20521 | CD | 0.044 | 355.3 | CD | 0.002 | 146.5 | R | 0.012 | 280.8 | - | - | - | CD | 0.002 | 311.3 |
| 21765 | CD | <0.001 | 321.3 | - | - | - | OD | 0.030 | 282.4 | LA | 0.005 | 157.2 | - | - | - |
| 22486 | - | - | - | - | - | - | LA | <0.001 | 270.9 | - | - | - | D | <0.001 | 300.0 |
| 23539 | D | 0.001 | 347.8 | - | - | - | D | 0.036 | 282.7 | LA | 0.007 | 162.4 | D | 0.031 | 316.6 |
| 57318 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 58232 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 58352 | - | - | - | - | - | - | R | 0.001 | 262.7 | - | - | - | CD | 0.049 | 314.6 |
| 58413 | - | - | - | - | - | - | D | 0.032 | 282.5 | - | - | - | D | <0.001 | 294.4 |
| 59179 | CD | 0.025 | 354.3 | CD | 0.024 | 151.2 | - | - | - | - | - | - | OD | 0.015 | 315.3 |
| 59448 | - | - | - | - | - | - | OD | <0.001 | 267.5 | - | - | - | - | - | - |
| 97286 | LA | 0.005 | 322.1 | - | - | - | LA | 0.016 | 278.3 | - | - | - | - | - | - |
| 108025 | - | - | - | - | - | - | R | 0.001 | 271.1 | - | - | - | LA | 0.007 | 303.7 |
| 108125 | - | - | - | - | - | - | CD | <0.001 | 259.1 | - | - | - | - | - | - |
| 108291 | - | - | - | - | - | - | CD | <0.001 | 259.1 | - | - | - | - | - | - |
| GH1 | LA | 0.026 | 331.8 | OD | 0.034 | 151.9 | LA | 0.029 | 267.1 | - | - | - | - | - | - |
| GH2_165 | CD | 0.049 | 342.6 | - | - | - | - | - | - | - | - | - | LA | <0.001 | 300.4 |
| M9_403 | CD | <0.001 | 324.5 | - | - | - | - | - | - | LA | 0.028 | 164.6 | OD | 0.026 | 316.3 |
| M1046_2 | D | <0.001 | 344.2 | - | - | - | - | - | - | - | - | - | - | - | - |
| MH30_Dreyer | D | <0.001 | 337.6 | R | 0.019 | 146.4 | D | 0.018 | 281.4 | - | - | - | D | <0.001 | 289.4 |
| Myostatin | - | - | - | - | - | - | - | - | - | D | 0.022 | 164.6 | D | 0.002 | 305.2 |
| Prolactin_1 | - | - | - | - | - | - | LA | 0.009 | 280.2 | D | 0.005 | 161.7 | D | 0.017 | 315.5 |
| SBF1 | D | <0.001 | 341.6 | - | - | - | - | - | - | - | - | - | - | - | - |
| TBC1 | D | 0.001 | 348.0 | - | - | - | LA | 0.027 | 279.2 | D | 0.007 | 162.4 | D | 0.000 | 307.7 |

| | Marianne | | | Turure | | | Yara | | | All rivers | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model | P value | AIC | Model | P value | AIC | Model | P value | AIC | Model | P value | AIC |
| 20521 | - | - | - | - | - | - | LA | 0.001 | 123.4 | CD | 0.005 | 524.0 |
| 21765 | - | - | - | - | - | - | - | - | - | CD | <0.001 | 500.2 |
| 22486 | CD | 0.037 | 149.8 | - | - | - | - | - | - | - | - | - |
| 23539 | - | - | - | - | - | - | D | 0.034 | 130.3 | LA | 0.002 | 522.3 |
| 57318 | - | - | - | - | - | - | - | - | - | - | - | - |
| 58232 | - | - | - | - | - | - | - | - | - | - | - | - |
| 58352 | - | - | - | - | - | - | - | - | - | - | - | - |
| 58413 | D | 0.029 | 149.3 | - | - | - | OD | 0.017 | 129.0 | - | - | - |
| 59179 | - | - | - | - | - | - | - | - | - | CD | 0.004 | 523.7 |
| 59448 | - | - | - | - | - | - | - | - | - | - | - | - |
| 97286 | - | - | - | - | - | - | LA | 0.025 | 129.7 | LA | 0.002 | 491.5 |
| 108025 | - | - | - | - | - | - | - | - | - | D | 0.018 | 484.7 |
| 108125 | - | - | - | - | - | - | - | - | - | - | - | - |
| 108291 | - | - | - | - | - | - | - | - | - | - | - | - |
| GH1 | - | - | - | - | - | - | - | - | - | R | 0.017 | 503.1 |
| GH2_165 | - | - | - | - | - | - | - | - | - | OD | 0.026 | 514.9 |
| M9_403 | - | - | - | - | - | - | - | - | - | CD | <0.001 | 503.7 |
| M1046_2 | - | - | - | - | - | - | OD | 0.049 | 131.0 | LA | 0.001 | 521.6 |
| MH30_Dreyer | - | - | - | - | - | - | R | 0.004 | 126.1 | CD | <0.001 | 500.9 |
| Myostatin | - | - | - | - | - | - | - | - | - | D | 0.024 | 515.5 |
| Prolactin_1 | - | - | - | - | - | - | - | - | - | - | - | - |
| SBF1 | - | - | - | - | - | - | D | 0.044 | 130.7 | D | 0.002 | 522.4 |
| TBC1 | - | - | - | D | <0.001 | 139.8 | D | <0.001 | 120.6 | D | 0.002 | 522.5 |

3.4.5.8 Association analysis

A significant allelic association with male SL was found in 21 of the 23 Psel SNPs, however, no patterns observed were consistent across rivers. Significant associations between SL and SNP genotypes were observed in at least one pairwise comparison for 21 of the 23 Psel SNPs (table 3.9). In the Aripo 2003, Caura and Lopinot the number of SNPs exhibiting a significant association was high (mean= 13.33). No significant associations were found in the Oropuche and only one in the Turure. In the remaining comparisons (Aripo 2006, Guanapo, Marianne and Yara) an average of 5 significant associations were found.

When considering consistency between rivers four SNPs showed a significant association with SL in at least 5 of the pairwise comparisons and global analysis of all rivers. Dominance of the T allele was observed for the TBC1 SNP in 5 of the pairwise comparisons and the global analysis. In these 5 rivers (Aripo 2003, Guanapo, Lopinot, Turure and Yara) the T allele was associated with large body size however, in the Caura, where a significant log-additive association was observed, this allele was associated with small body size. Although significant associations were also observed for SNPs 20521, 23539 and MH30_Dreyer in 5 pairwise comparisons the allele associated with size differed between rivers making it unlikely that the these alleles are linked to body size (table 3.9).

### 3.4.6 Signs of selection across all analyses

Both Arlequin and Lositan have been shown to have high type II error rates (Narum and Hess 2011, Vilas *et al.* 2012) and therefore SNPs identified as outliers in only one of these analyses are not considered to be under selection in this study. Only four SNPs were identified as outliers by both Arlequin and Lositan (two each in the Oropuche and Turure) but only in one pairwise comparison each. Two of these SNPs were putatively neutral (291080 and 283548), one was a candidate SNP (GH2_165) and one was a RAD associated SNP (59179).

The Prolactin SNP, which was identified as under selection in the experimental selection lines, was also identified as an outlier in the Aripo 2003, Caura and Lopinot by Lositan and in the Oropuche by Arlequin. The pattern of allele frequencies in these 4 rivers was consistent and the same as the pattern of allele frequency difference observed between experimental lines. Despite this, a significant allelic association with SL was only observed in three of the rivers sampled for this SNP.

Across both the analyses and the rivers sampled in our study no SNP showed consistent signs of selection for body size. The possible reasons for this will be discussed in section 3.5.3.

## 3.5 Discussion

In this study, the patterns of selection on body size and the nature of genetic change observed in SNPs in the experimental lines from Chapter 2 were compared to those observed in wild populations. A total of 50 SNPs were successfully genotyped in fish from eighteen sites comprised

of at least one upstream and downstream population in each of eight rivers located in Northern Trinidad. In agreement with predictions (Reznick *et al.* 1996; Reznick 1987; Reznick & Endler 1982; Willing *et al.* 2010), our study found that adult males were significantly smaller in upstream sites compared to downstream sites in the Aripo (in both 2003 and 2006), Caura, Guanapo, Lopinot, Marianne and Yara rivers. However, males did not differ in body size between upstream and downstream sites of the Oropuche and Turure rivers. The patterns of genetic variation observed correlated with the geographic regions (Shaw *et al.* 1991): levels of genetic diversity were found to be higher in downstream than in upstream sites in all rivers with the exception of the Turure. However, despite the significant phenotypic differences observed, we identified no consistent signs of selection for body size in any of the SNPs studied here.

This section will begin with a discussion of the phenotypic differences observed in guppies from the rivers sampled. In order to provide an indication of the reliability of the data we will then discuss the similarities between our data and previous studies before finally offering reasons as to why no consistent signs of selection were detected in fish genotyped from the wild at RAD-associated SNPs.

### 3.5.1 Phenotypic variation

Guppies in populations located in upstream and downstream sites within the same river have been shown to exhibit variation in a wide range of phenotypic traits. In our study we identified significant variation in body size in six of the eight rivers sampled. All targeted rivers in which a significant difference in SL was identified also exhibited predicted differences in predation regime, with high levels of predation found in lowland sites and low levels of predation in upland sites. The difference in body size between the large and small selection lines (mean = 1.97mm) was comparable to the size differences in the rivers which exhibited a significant difference (mean = 1.86mm). The similarity between the phenotypes observed in the wild and in our selection experiment provided a potentially insightful opportunity to assess the SNPs showing signs of selection in the experimental lines with fish from wild populations.

In the Turure and Oropuche samples no significant difference in body size was observed. Although Croft & Krause (2004) did find variation in body size between upland and lowland sites in the Turure, they also reported a corresponding difference in predation regime. In the two rivers for which no difference in SL was found in our study, no such variation in predation regime existed. In fact, high levels of predation were found in all sites sampled in both of these rivers in our study (Ryan S. Mohammed personal communication). The presence of sites, both with and without a difference in SL, would have helped to distinguish between the effects of genetic drift and selection if any signs of selection had been detected.

### 3.5.2: Population structure

3.5.2.1 Genetic structure between rivers

As previously found (Carvalho *et al.* 1991, Shaw *et al.* 1991, Alexander *et al.* 2006, Barson *et al.* 2009, Suk and Neff 2009, Willing *et al.* 2010, Fraser *et al.* 2014), the majority of samples from different rivers clustered according to drainage basin of origin. The one exception to this was the Turure River which clustered with samples in the Caroni drainage (figure 3.4). While the Turure is located within the Oropuche drainage, guppies from a high predation site in the Guanapo River (located in the Caroni drainage) were translocated in 1957 to a guppy free site in upstream Turure. Guppies sampled from the upstream Turure site in this study are known to be the descendants of the fish translocated from the Guanapo (C. van Oosterhout personal communication). Since the initial introduction, the translocated fish have largely replaced the native population in the downstream sites as well (Shaw *et al.* 1991, Becher and Magurran 2000, Suk and Neff 2009). The initial introduction and subsequent migration of these fish explains why the Turure samples clustered in the Caroni drainage instead of Oropuche drainage.

The clustering seen with the PCA was not as clearly defined as previously found in an analysis based on ~1000 SNPs (Willing *et al.* 2010). However, concordance of the observed structure between our data and previous studies (Carvalho *et al.* 1991, Shaw *et al.* 1991, Alexander *et al.* 2006, Barson *et al.* 2009, Suk and Neff 2009, Willing *et al.* 2010, Fraser *et al.* 2014) suggest that a lack of clear differentiation is a result of the smaller numbers of polymorphic SNPs employed here. Although genetic structuring between rivers was not as clearly defined with PCA, $F_{st}$ values between the rivers were high and similar to those found in other studies (Carvalho *et al.* 1991, Shaw *et al.* 1991, Barson *et al.* 2009, Suk and Neff 2009, Willing *et al.* 2010, Fraser *et al.* 2014).

3.5.2.2 Genetic structure within rivers

Several previous studies have highlighted the substantial differentiation which exists between sites located along the same river (Carvalho *et al.* 1991, Shaw *et al.* 1991, Crispo *et al.* 2006, Barson *et al.* 2009, Suk and Neff 2009, Willing *et al.* 2010, Fraser *et al.* 2014). Waterfalls, which can be found in most of the rivers in Trinidad, limit migration and gene flow between upstream and downstream sites (Crispo *et al.* 2006, Barson *et al.* 2009, Fraser *et al.* 2014). These waterfalls also mean that any gene flow which does take place, primarily occurs from upstream to downstream sites with very little gene flow in the opposite direction (Crispo *et al.* 2006, Barson *et al.* 2009, Fraser *et al.* 2014). Each of the rivers in Trinidad contains multiple upstream sites located in tributaries feeding into the main river. Due to the mountainous geography of the island, these sites are often isolated from each other and genetically differentiated (Barson *et al.* 2009). As the tributaries feed into the main river, gene flow from these isolated populations feeds into fewer downstream populations. Such typically unidirectional gene flow from multiple upstream populations into fewer downstream populations creates downstream populations which exhibit higher levels of diversity. Furthermore, stochastic forces such as founder effects decrease the diversity in upstream populations. Seasonal flash flooding which has been shown to decrease guppy biomass significantly (Grether *et al.* 2001)

might exacerbate founder effects. These flooding events could increase migration from upstream to downstream sites but might also allow for active upstream migration (Barson *et al.* 2009).

In our study the expected pattern of reduced diversity in upland sites was observed in the majority of the rivers sampled. As an exception, in the Marianne, diversity was higher in the upstream site. However only 18 of the SNPs genotyped in this river were polymorphic (minor allele frequency (MAF) > 0), and four of these only had a MAF ≤ 0.01. Previously Crispo *et al.* (2006) undertook an in depth study of 20 sites along the Marianne and, found He to be significantly higher in downstream sites than in upstream sites. Similar pattern of diversity in the Marianne has also been found in other studies (Crispo *et al.* 2006, Willing *et al.* 2010). The findings of these studies suggest that the inconsistent pattern of He observed in our study is likely to be a result of the low number of polymorphic SNPs and is not a true representation of the diversity present in sites in the Marianne.

As discussed above, one of the reasons for the increased level of diversity in the downstream sites was the mixing of fish from several genetically diverse upstream populations which could potentially cause a Wahlund effect within the downstream populations. $F_{st}$ can ultimately be described as a measure of the drop in heterozygosity in a sub-population relative to the total population as a result of genetic drift. Therefore, if observed heterozygosity is less than expected heterozygosity in the downstream populations as a consequence of the Wahlund effect, the calculations of the $F_{st}$ between each upstream/downstream site could be inflated. As the outlier analysis utilised in the current project focuses on detecting SNPs with unexpectedly high $F_{st}$ values, the Wahlund effect might have led to SNPs being falsely identified as showing signs of selection. However, it would also be expected that if the Wahlund effect was present in the downstream sites the SNPs would be out of HWE in these populations.

As well as significant variation in the levels of genetic diversity between upstream and downstream sites we also observed high $F_{st}$ values between sites within a river in five of the eight rivers studied. In three of the eight rivers (Marianne, Oropuche and Turure), $F_{st}$ values showed low levels of divergence between upstream and downstream sites. Although the divergence in the Marianne was low in comparison with the other rivers sampled, the $F_{st}$ of 0.06 was significant. Once again, previous studies (Crispo *et al.* 2006, Willing *et al.* 2010) have found much higher levels of divergence between sites in the Marianne than we observed and it is therefore likely that the $F_{st}$ value found here is an underestimation of the true genetic differentiation. The $F_{st}$ values obtained for the Turure and Oropuche showed no differentiation between the sites within these rivers. Similar values were found in previous studies (Suk and Neff 2009, Willing *et al.* 2010, Fraser *et al.* 2014) and our values are therefore likely to be a good representation of the true differentiation present.

The large genetic differentiation between upland and lowland sites is primarily driven by the large waterfalls which limit gene flow though, in the Marianne, geographic distance has also been found

to reduce dispersal (Crispo *et al.* 2006). The sites sampled in the Oropuche in the study presented here were not separated by large waterfalls but the geographic distance between them was substantial (4.14 miles). The lack of genetic differentiation between the upland and lowland sites on the Oropuche therefore suggests that geographic distance does not limit gene flow between these sites.

The two Turure sites sampled in the study presented here are separated by a drop of approximately 1 m over a series of small waterfalls spread across a 300 m section of river (R. Mohammed personal communication). As these sites are the result of an introduction, the time span in which divergence has taken place in much shorter than any of the other sites within the rivers. The lack of genetic divergence observed between these sites is therefore likely to be the result of a combination of close proximity of the sites and the recent timing of the divergence.

Analyses of more than one upstream site within the Aripo highlight the importance of confirming the results obtained across multiple sites within a river. The level of genetic divergence observed between the samples collected from the downstream Aripo site in 2003 and 2006 were very low, which suggests that temporal variation between sites is likely to be low. However, in the 2003 sampling, the differentiation between the upstream and downstream sites was higher than in the 2006 sampling (a different upstream site and the same downstream site). These results suggest that variation in the pair of sites sampled in 2003 and the pair of sites sampled in 2006 is caused by variation in the upstream sites rather than the time span between samples. Indeed, large temporal genetic variation has been found in guppies of the Aripo River by previous workers studying immune genes of the MHC (Major Histocompatibility Complex). These genes are known to be under strong selection (van Oosterhout *et al.* 2006, Llaurens *et al.* 2012), and computer simulations indicate that temporal genetic divergence can significantly exceed spatial divergence under some scenarios of selection (McMullan and Van Oosterhout 2012).

Analysis of the population structure, both between and within rivers, although not the main aim of this study, offers the opportunity to assess the accuracy of the data set. Results from the Marianne River suggest that the number of SNPs genotyped in this river does not give a true representation of the genetic variation present. However, with the exception of the Marianne, the concordance between our study and previous work provides a high level of confidence and allows us to examine the potential reasons for an apparent lack of selection at target loci.

### 3.5.3: Detecting SNPs under selection

Although 17 of the SNPs examined here were found to show consistent signs of selection in experimental lines which had been selected for body size (van Wijk 2011a), there was no clear trend of such associations with body size in the wild populations studied. The potential reasons for this, which include both limitations of study design and underlying biological factors, are now considered.

3.5.3.1 Experimental design

All $F_{st}$ outlier analyses used to detect signs of selection rely on the principle that loci under diversifying selection should exhibit larger divergence ($F_{st}$) between populations than neutral loci (Lewontin and Krakauer 1973). Thus, a number of neutral markers are required to estimate the neutral levels of divergence and allow loci under selection to be detected. In our study of wild fish, we successfully genotyped 27 polymorphic SNPs exhibiting very low $F_{st}$ values in the experimental selection lines studied in Chapter 2. The low divergence exhibited by these SNPs in the experimental lines shows that they were not under selection for body size, and were subsequently classified as putatively neutral. However, on average, half of the putatively neutral SNPs in any river exhibited intermediate levels of divergence ($F_{st} \geq 0.05$), while 10% exhibited very high levels of divergence ($F_{st} \geq 0.25$). This high level of divergence in our putatively neutral SNPs resulted in the neutral simulated $F_{st}$ being high, making it difficult to detect SNPs with high levels of divergence as a result of selection. Essentially, if the mean dataset $F_{st}$ is high, the $F_{st}$ of the simulated neutral distribution will also be high. SNPs with a high $F_{st}$ as a result of divergent selection will therefore fall within the neutral distribution and will not be detected as outliers. In other words, random genetic drift has inflated the level of genetic differentiation of the putatively neutral SNPs, and hence, that signal of diversifying selection in the putative selective SNPs may be overshadowed by the effects of drift, making selection difficult to detect.

The high levels of divergence observed in our putatively neutral loci between upland and lowland populations in the wild suggests that some of them were not neutral. Guppy populations, which experience different levels of predation, are known to differ in a wide range of traits, not just body size, and it is possible that the putatively neutral SNPs are actually under selection for other traits. However, studies using a number of markers have found considerable divergence between upland and lowland populations in these rivers which is mainly driven by genetic drift and not selection (Crispo *et al.* 2006, Willing *et al.* 2010). Therefore the divergence observed in the putatively neutral SNPs is most likely to be a result of genetic drift rather than selection.

The inflated neutral $F_{st}$ as a result of divergence caused by genetic drift rather than selection will also be exacerbated by the polygenic nature of body size. The van Wijk *et al.* (2013) experiment exposed fish to very strong selection for only three generations. The contribution of an allele under strong selection over a short time frame will be influenced by its frequency in the founder population (Falconer and Mackay 1966, Illingworth and Bird 2009, Nosil 2012). For example, an allele with a large effect, which only occurs at low frequency in the founder population, will not contribute much to the selective response, whereas a modest effect SNP, which occurs at an intermediate frequency in the founder population, will result in a large selective response. Therefore SNPs identified as being under selection in the van Wijk *et al.* (2013) experiment may actually only be small effect loci which occurred at an intermediate frequency in the founder population. When a high level of

genetic drift and a large number of small effect loci are combined, it will be difficult to detect real signs of selection at a SNP.

3.5.3.2 Reduced recombination

The strength and timing of the selection imposed on the experimental lines compared with the wild populations will also have resulted in variation in the level of linkage between the causal variant and the SNP being studied i.e. not all causal variants and physically linked SNPs will necessarily be in high statistical linkage (Smit-Mcbride *et al.* 1988). In our selection experiment, a SNP in linkage with the variant under selection at the start of the study will still be in linkage with it three generations later. However, in the wild populations where selection has been underway for much longer, recombination between causal variants and SNPs under study may have reduced the association between the two (Falconer and Mackay 1966, Lynch and Walsh 1998). The SNPs studied here may have been in linkage with the causal variant in the experimental selection lines but not in the wild populations. Fraser *et al.* (2014) looked for signs of selection across the genomes of both wild and translocated guppy populations and found higher linkage disequilibrium in outlier regions than non-outlier regions for translocated populations but not for wild populations. The linkage disequilibrium decay in wild populations will add to the difficulties in detecting the causal variant for body size and may explain why SNPs that show signs of selection in our experimental lines show no signs of selection in the wild populations.

3.5.3.3 Convergent evolution

One further possibility for the lack of consistent signs of selection across the rivers sampled here may arise from variation in body size arising from river-specific gene-phenotype associations. Convergent evolution, which here refers to phenotypic traits which have evolved independently in multiple environments or lineages, can be seen both within (mouth morphology in cichlids: Muschick *et al.* 2012, limb morphology in Anolis lizards: Losos 2009 and temperature tolerance in *Drosophila melanogaster*: Schluter 2000) and across a wide range of species (for example echolocation in mammals: Parker *et al.* 2013 and wing pattern mimicry in butterflies: Reed *et al.* 2011). However, despite its prevalence, we know relatively little about whether the genetic mechanisms underpinning these traits are also similar. Changes in the same amino acid have been found to mediate a convergent phenotype (Stewart *et al.* 1987, Hoekstra *et al.* 2006, Wierer *et al.* 2012). These examples however, are rare and it has been suggested that the genetic mechanisms underpinning these traits are often not the same (Foote *et al.* 2015). Further, theoretical studies have confirmed that convergent evolution is less likely to occur in polygenic traits such as body size than those underpinned by few genetic mechanisms (Orr 2005).

Recently a study used RAD sequencing to examine phenotypic convergent evolution in wild and translocated populations of guppy (Fraser *et al.* 2014). Despite the large number of genome-wide loci considered, the regions of the genome identified as under selection were not consistent across wild populations. One suggestion proposed by the authors for the absence of genetic convergence

was that different loci are under selection in different populations. Life history traits such as body size are polygenic, and often determined by many loci; for example height in humans can influenced by nearly 300,000 SNPs (Yang *et al.* 2010). There are therefore a huge number of genetic pathways which could lead to the same phenotype. The absence of consistent signs of selection in our study offers further support to the hypothesis that genetic convergent evolution in guppy populations is not found because the genes which mediate traits such as body size likely differ across populations. Furthermore, the absence of signs of selection in the study by Fraser *et al.* 2014 suggests that it is unlikely that the low density of SNPs contributed to the absence of consistent signs of selection in our study.

3.5.3.4 Pleiotropy

Pleiotropy is the phenomenon where one locus controls multiple apparently unrelated phenotypic traits (Stearns 2010). Although examples of pleiotropy can be seen in a number of species (*Pisum sativum*: Mendel 1866, vestigial gene in *Drosophila melanogaster*: Miglani 2002, frizzle gene in *Gallus gallus domesticus*: Landauer and Upham 1936, phenylketonuria gene in humans: Paul 2000), it can be difficult to differentiate the true pleiotropy and the effect of two or more separate loci which are closely linked (Flint and and Mackay 2009). In guppies pleiotropy may be responsible for females preference for males with larger orange spots (Rodd *et al.* 2002) and a link between male brain size and a range of sexual traits (Kotrschal *et al.* 2015). Furthermore it has been suggested that antagonistic pleiotropy, where the pleiotropic allele is beneficial for one trait but deleterious for another, occurs in Y-linked loci responsible for both male ornamentation and survivability (Brooks 2000, Bolstad *et al.* 2012). It is therefore possible that the lack of SNPs showing consistent signs of selection in the current study is due to the loci responsible for body size being pleiotropic. Such pleiotropy at these loci could result in these loci being under varying selection pressure in the different rivers despite the consistent differences in predation regime.

3.5.3.5 Heterosis

When considering the potential genetic mechanisms which can determine body size and the effect these might have had on the results of our study, it is also important to consider heterosis. Sometimes known as hybrid vigour, heterosis describes the improved performance of hybrid, and therefore heterozygous, offspring relative to their homozygous parents (Birchler *et al.* 2003). The two main theories which have been proposed to explain heterosis are dominance and overdominance (Birchler *et al.* 2010). The dominance hypothesis states that the increased vigour of the hybrid is due to the masking of recessive alleles by superior dominant alleles (Davenport 1908, Bruce 1910, Jones 1917). Under the dominance scenario, homozygosity for the superior allele would result in the same phenotype as the hybrid. The overdominance hypothesis attributes heterosis to heterozygote advantage per se, that is when the heterozygote is superior to homozygotes of either allele (East 1908, Shull 1908). The classic example of heterosis due to overdominance can be seen in sickle cell anaemia whereby heterozygosity for the recessive sickle allele will give the individual immunity from malaria without them suffering from sickle cell anaemia.

Size is a one of the most common phenotypes to be effected by heterosis, with hybrid offspring being larger than their parent (Gama *et al.* n.d., Drabo *et al.* 1984, Gjerde and Refstie 1984, Hedgecock *et al.* 1995, Bryden *et al.* 2004, Meyer *et al.* 2010, Groszmann *et al.* 2014). If the large body size found in the upstream populations was underpinned by overdominance heterosis it is possible that the outlier analysis utilised would not detect it. For example, if large body size was solely the result of heterozygote advantage it would be possible that the allele frequencies in the upstream and downstream sites would be the same, but that the proportion of heterozygotes was much higher in the upstream site. Under this scenario the $F_{st}$ between the two would be low despite the marked difference in their genetic composition. Although heterosis has not been widely examined in wild populations of the guppy, studies which have considered its effect on body size have found it to be minimal (Nakadate *et al.* 2003, Shikano and Taniguchi 2003).

### 3.5.4: Comparing results with a previous study

Of the nine candidate SNPs analysed three (M30_Dreyer, Prolactin and M9_403) had previously shown consistent signs of selection in the experimental selection lines (van Wijk 2011a). van Wijk (2011) found consistent signs of selection and patterns of allelic variation between the large and small selection lines in the M30_Dreyer SNP: however, this pattern was not observed in our analysis of wild fish (figure 3.7).

Although it is possible that this is a result of genotyping error in our data set, the high level of concordance between genotypes obtained from RAD sequencing and the Sequenom assay (96.6%) make it unlikely that this is an assay-wide problem. It could, however, be the result of a SNP-specific genotyping error. One further explanation is that the genotypes in both studies could be correct. In such case as the fish genotyped in here were not the same as the fish genotyped by van Wijk (2011), the signs of selection identified by van Wijk (2011) may not be as consistent as was first thought.

### 3.5.5: Concluding remarks

Using selection experiments to identify candidate genes for investigation of wild populations has been used successfully to identify the genetic mechanisms underpinning trait variance (Aranzana, Kim, Zhao, Bakker, *et al.* 2005, Turner *et al.* 2005, Jones *et al.* 2008, Orsini *et al.* 2012, Ariey *et al.* 2014). However, using this technique, we were unable to identify consistent signs of selection in any SNPs identified from RAD analysis of size-selected guppy lines (Van Wijk *et al.* 2013). Several possibilities for such observations are proposed, and it is likely that all, to varying degrees have contributed. The results from this chapter and other recent work (Fraser *et al.* 2014) suggest that detecting genetic variants associated with body size in wild populations of guppies is likely to be challenging. In order to detect such associations in wild populations, comparison recently diverged translocations, where linkage between the causal variant and nearby variation is tighter, may be

especially insightful, as well as techniques which are able to detect small effect loci (Bourret *et al.* 2014, Brieuc *et al.* 2015).

# Chapter 4: Epigenetic changes following a size selection experiment

## 4.1 Abstract

The previous two chapters of the thesis have examined the genetic change underpinning variation in life history traits observed in guppies. However, where evolution is rapid, epigenetic change has also been found to drive major shifts in life history. To assess the contribution of epigenetics to the shifts in life history traits observed by van Wijk (2011), the following chapter examines the patterns of DNA methylation in fish from the selection experiment. Using methylation sensitive AFLP analysis, DNA methylation was examined in four tissues from fish in the four selection lines (F6 generation) and two tissues in the fish from the pre-selection (F2 generation). Patterns of DNA methylation varied significantly between tissue types. When all of the 91 loci analysed were considered, no variation in the level of methylation between the selection lines was detected. The level of DNA methylation across all loci increased during the course of the selection experiment, however this increase was the same across all four selection lines. When the loci were examined individually, only one showed significant variation between the selection lines. Given the complexity of the traits being examined, it is unlikely that DNA methylation has not contributed to the variation observed. However, given the results described here, the shifts in life history traits observed by van Wijk (2011) are thought to primarily be the result of genetic change rather than epigenetic change.

## 4.2 Introduction

Evolutionary change is typically considered to occur slowly over many millions of years. However, it can also occur rapidly, with species evolving and adapting to their environment over tens of generations (and sometimes less: Reznick 1987, Khater *et al.* 2014). While contemporary evolution can, and does, take place in populations not affected by human actions, anthropogenic activity often results in rapid evolution. Anthropogenic pressures such as heavy exploitation or pollution, can induce phenotypic change by acting as a direct selection pressure and/or by modifying environmental conditions and driving natural selection. Harvesting creates selection pressure which usually acts to drive evolution in the opposite direction from that which would be favoured by natural selection and therefore has the potential to create maladapted populations. In fisheries, natural selection would usually favour slower growth, later maturation and ultimately larger body size. However, the high levels of harvesting experienced by most exploited fish populations have led to smaller body size, with fish maturing earlier and at a smaller size. Although still controversial, there is now a large body of evidence, including findings from the current study, which suggests that such changes are likely to be underpinned by genetic, change rather than phenotypic plasticity.

Rapid evolutionary change has been observed in life history traits if the alewife, *Alosa pseudoharengus*, in traits very similar to those seen in exploited fish populations (Ricker 1981, Olsen *et al.* 2004, Swain *et al.* 2007, Fenberg and Roy 2008, Neuheimer and Taggart 2010, Hutchings and Rangeley 2011). However these changes are not underpinned by changes in the

DNA sequence. In its native form, the alewife exhibits an anadromous life cycle, but following a rapid range expansion between 1860 and 1955 several populations became landlocked in the great lakes of North America, and a new freshwater morph evolved. The freshwater morph matures earlier, has slower adult growth, smaller size at maturation and reduced fecundity in comparison to the anadromous morph. Czesny *et al.* (2012) examined the contribution of regulatory modifications and nucleotide changes to the rapid evolution of the freshwater morph, and found the phenotypic differences were a result of large regulatory changes, rather than coding changes. Regulatory change may be underpinned by epigenetic changes (Armstrong 2013, Allis *et al.* 2015). However, despite the similarity between the phenotypic changes observed in the rapid evolution of the freshwater morph and overexploited fish populations, to date, no study has considered the contribution of epigenetic changes to the phenotypic change resulting from overexploitation in fisheries.

Regulatory change can also be underpinned by genetic change, that is, changes to the DNA sequence (Carroll 2005). For example, pelvic reduction in sticklebacks, which is thought to be linked to the evolution of hind limbs, is the result of sequence changes in the regulatory region of the *Pitx1* gene (Shapiro *et al.* 2004). As well as small base changes such as SNPs and indels, gene duplication and rearrangements can also influence regulation. Although regulatory change can be underpinned by both genetic and epigenetic mechanisms, here we focus on the potential contribution of epigenetic change to the observed shifts in life history traits.

Historically the term epigenetics referred to all processes which regulated the final phenotype (Felsenfeld 2014). Today a number of different definitions exist and different fields of study have different views of the limitations of the term debated (Bird 2007, Richards *et al.* 2010, Doğan *et al.* 2013, Felsenfeld 2014, Deans and Maggert 2015). Here we refer to epigenetics as a study of heritable changes which cannot be explained by variation or changes in the DNA sequence. Such a description could apply to a number of epigenetic mechanisms including chromatin remodelling, histone modification, non-coding RNA and DNA methylation. Here, we focus on DNA methylation which is not only the most widely studied epigenetic mark in humans (Heyn *et al.* 2013, Yamada and Yonezawa 2013, Jones *et al.* 2015) but is also the subject of an increasing number of studies which consider its role in wild populations (Herrera and Bazaga 2010, Lira-Medeiros, Parisod, Fernandes, *et al.* 2010, Massicotte *et al.* 2011, Kilvitis *et al.* 2014, Wenzel and Piertney 2014, Liu, Sun, *et al.* 2015).

### 4.2.1 The effect of DNA methylation

The addition of a methyl group to the cytosine base, usually to a CpG site, is known to influence gene expression. In vertebrates, the only unmethylated regions of the genome are CpG islands (CGI) which are typically found in the promotor regions of genes (Jones 2012). Methylation of these sites can prevent transcription, acting as a 'lock' to silence genes and is thought to be involved in sex chromosome inactivation (Sharp *et al.* 2011, Cotton *et al.* 2015), imprinting (Li *et al.* 1993, Reik

and Walter 2001, Lucifero *et al.* 2002) and silencing transposable elements (Ikeda and Nishimura 2015). The effect that methylation can have on life history traits has been widely studied in humans, (Heijmans and Tobi 2008, Fryer *et al.* 2011, Slomko *et al.* 2012, Burgess 2013, Gentilini *et al.* 2013, Lomniczi *et al.* 2013) but less so in plants and animals (Shindo *et al.* 2006, Cortijo *et al.* 2014, Zhong *et al.* 2014, Alvarado *et al.* 2015, Cao *et al.* 2015, Hu *et al.* 2015, Yang *et al.* 2015). Studies have, however, shown that DNA methylation can influence development (Navarro-Martín *et al.* 2011, Ding *et al.* 2013) and growth (Wolff *et al.* 1998, Zhong *et al.* 2014, Alvarado *et al.* 2015, Cao *et al.* 2015) in a range of animals.

### 4.2.2 DNA methylation in fish

Before the potential contribution of DNA methylation to the observed shift in life history traits can be assessed, it is important to consider what is already known about DNA methylation in the target group. Initially methylation research in fish focused on determining the pattern of DNA methylation (Feng *et al.* 2010) and identifying the enzymatic pathway responsible for methylation (Mhanni and McGowan 2004, Mackay *et al.* 2009, Fang *et al.* 2013). Over the last two decades, interest in DNA methylation has grown rapidly, and to date, DNA methylation in fish has been studied in relation to cancer (Mirbahai *et al.* 2011, Wu *et al.* 2011), embryonic development (Andersen, Reiner, *et al.* 2012), gonad growth (Pierron *et al.* 2014), sex determination (Shao *et al.* 2014), aquaculture (Moghadam *et al.* 2015), body temperature (Varriale and Bernardi 2006a), morphology (Smith, Smith, *et al.* 2014) and body size (Zhong *et al.* 2014).

All species of fish examined so far exhibit a global pattern of methylation similar to that found in

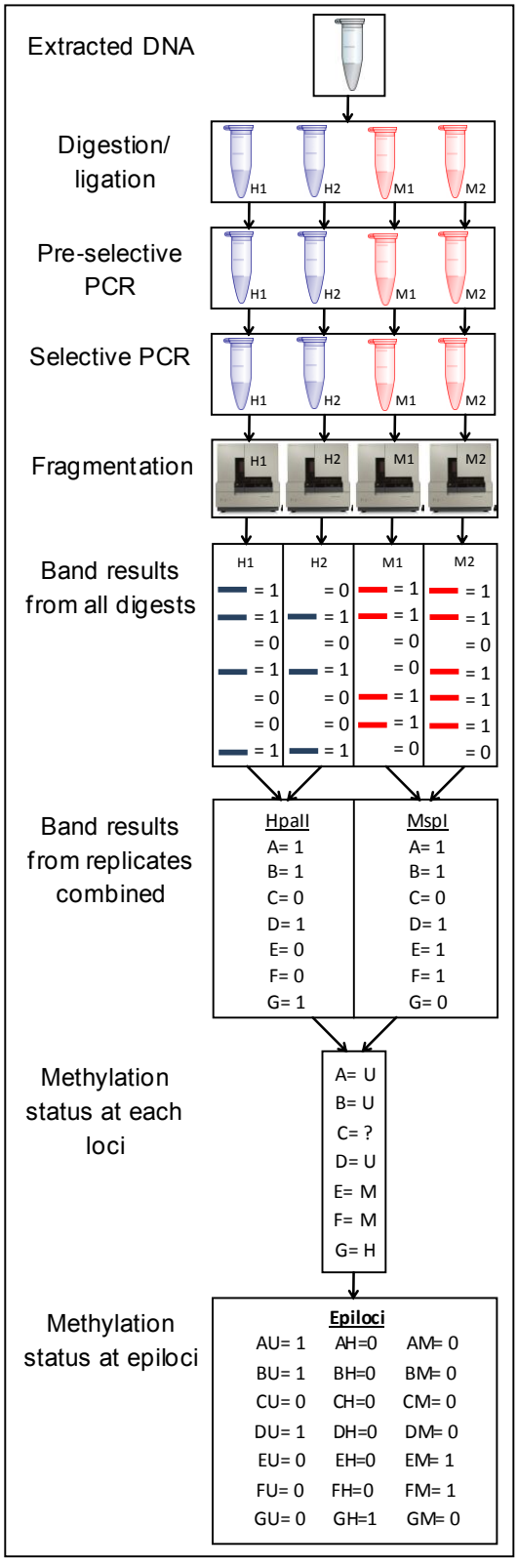



**Figure 4.1:** Schematic showing the MSAP process from DNA extraction to epiloci values. H1= HpaII digest replicate 1; H2= HpaII digest replicate 1; M1= MspI digest replicate 1; M2= MspI digest replicate 2

mammals, although the total level of methylation in fish has been shown to be higher with ~80% of CpG sites methylated (~74% in mammals) (Jabbari *et al.* 1997, Feng *et al.* 2010). The level of DNA methylation is also higher in tropical/temperate fish species than it is in polar species (Varriale and Bernardi 2006a, Varriale 2014). Work with both fish and reptiles suggests that the higher level of methylation found in fish (and also reptiles) is a result of a lower rate of deamination in cold blooded vertebrates (Varriale & Bernardi 2006a; 2006b). The spontaneous deamination of 5-methylcytosine yields thymine and therefore the high level of deamination is thought to explain the deficit of CpG sites found in most vertebrate genomes (Shen *et al.* 1994) and offers another potential way in which DNA methylation could influence a population's evolution.

Of particular interest here are studies in which DNA methylation has been shown to have an effect on variation in body size and growth rate in fish. For example, in the Tilapia (*Oreochromis niloticus*), DNA methylation in the promotor region of the growth hormone (GH) gene was negatively correlated with GH mRNA expression and growth rate (Zhong *et al.* (2014). Sexual size dimorphism in Tilapia may also be the result of higher levels of methylation in the GH promotor region in females (adult males weigh approximately twice as much as females and are ~23mm longer [Bhatta *et al.* 2012]). In freshwater and marine ecotypes of the three spined stickleback (*Gasterosteus aculeatus*), differential methylation regions associated with several growth related genes including an insulin-like growth factor receptor, has been identified (Smith *et al.* (2014)). It should be noted, however, that phenotypic variation between freshwater and marine ecotypes in the stickleback have been strongly associated with genetic changes (Baird *et al.* 2008, Hohenlohe *et al.* 2010, Catchen *et al.* 2013, Guo *et al.* 2015), and it is therefore unlikely that DNA methylation is independent of genetic change in this scenario.

### 4.2.3 Methylation sensitive amplified polymorphism

Over the last 30 years a wide range of techniques for studying DNA methylation have been developed (see Shen & Waterland 2007 and Moghadam *et al.* 2015 for reviews of relevant techniques). Bisulphite sequencing is considered to be the gold-standard for methylation analysis, and when combined with MPS is a very powerful approach for studying the entire methylome (Tost and Gut 2007, Ku *et al.* 2011). However this approach is still costly and not suited to the large sample sizes usually required in population-based studies.

The most commonly used technique for studying methylation in larger studies, particularly when using wild populations, is methylation sensitive amplified polymorphism analysis (MSAP) (Schrey *et al.* 2013). MSAP is a modification of the standard amplified fragment length polymorphism analysis which allows numerous loci to be analysed in many individuals for a fraction of the cost of most other techniques. By using the rare cutter *Eco*RI in combination with two common cutters which recognise the same site, but differ in their sensitivity to methylation, MSAP can use variation in the banding pattern to identify methylation (Figure 4.1). The two isoschizomers, *Hpa*II and *Msp*I recognise the restriction site: 5'-CCGG-3'. Both enzymes will cut when non-methylated sites are

tested, while *Hpa*II will not cut if the internal C is fully methylated (methylation present on both strands) or hemi-methylated (methylation only present on one strand) and *Msp*I will not cut if the external C is fully methylated or if both Cs are methylated (see Table 4.1). The main limitations of MSAP are that methylation can only be detected at CCGG sites, and that the sites analysed are anonymous and may not be linked to the trait in question. Although it is not possible to target specific loci using MSAP, studies which have found variation in methylation patterns have enabled subsequent sequencing of loci of interest in the quest to assess functional importance (Wenzel and Piertney 2014).

**Table 4.1:** Sites which will and won't be cut by enzymes HpaII and MspI. **C** indicates a methylated cytosine.

| | | MspII | | | | | |
|---|---|---|---|---|---|---|---|
| | | Cuts | | Doesn't cut | | | |
| HpaII | Cuts | CCGG<br>GGCC | | **C**CGG<br>GGCC | | | |
| | Doesn't cut | C**C**GG<br>GG**C**C | C**C**GG<br>GGCC | **C**CGG<br>GGC**C** | C**C**GG<br>GG**CC** | C**C**GG<br>GGCC | M<br>M |

By enabling the detection of a large number of epiloci across the genome in numerous individuals, MSAP is a quick, cheap and powerful technique for studying methylation. In particular the utility of MSAP in detecting variation in DNA methylation between populations has been seen in a range of studies involving both plants (Xiong *et al.* 1999, Portis *et al.* 2004, Takata *et al.* 2005, Li *et al.* 2008, Salmon *et al.* 2009, Gao *et al.* 2010, Herrera and Bazaga 2010, Long *et al.* 2011) and animals (Taylor and Blouin 2010, Massicotte *et al.* 2011, Morán and Pérez-Figueroa 2011, Rodríguez López *et al.* 2012, Schrey *et al.* 2013, Wenzel and Piertney 2014). Thus, the MSAP approach was used in our study to assess the contribution of DNA methylation to the observed phenotypic variation.

In the present study we investigate the methylation patterns in the Trinidadian guppy, *Poecilia reticulate*, from the four selection lines generated by the experimental selection on body size (van Wijk *et al.* 2013). By analysing four tissue types we will be able to assess the level of DNA methylation found in guppies and assess the contribution of DNA methylation to the shifts in life history patterns observed. In addition, the level of DNA methylation in fish from the generation before selection will also examined to determine whether DNA methylation patterns changed significantly during the selection experiment.

## 4.3 Methods

### 4.3.1 Samples

*P. reticulata* were originally collected from the Tacarigua river, Northern Trinidad (10°38'49.5"N, 61°22'47.2"W) in March 2008. Once transferred to Bangor, individuals were subjected to three generations of random breeding and three generations of experimental selection for body size. Full

details of the collection of fish and the selection experiment can be found in Chapter Two and van Wijk (2011).

Tissue samples for DNA methylation studies were collected from adult males in the $F_2$ generation (before the experimental selection began: n=10) and the $F_6$ generation (the generation after selection: n=60). For fish from the $F_2$ generation, samples were obtained from the tail tissue shortly after the fish had reached maturation. In order to obtain samples from the $F_6$ generations the majority of fish were euthanised with an overdose of 0.02% tricaine methanesulfonate (MS222; Pharmaq, Fordingbridge, UK). A few $F_6$ fish (n=6) died of natural causes and were preserved in 90-100% ethanol within 12 hours of death. The age of all sampled $F_6$ fish was between 1-2 years. Following collection, all $F_6$ fish were dissected and tissue from the liver, brain, gonads and tail was collected. DNA was extracted from all samples using the salting out technique as described in Domingues *et al.* (2010). These four tissues were chosen for practicality and because they have previously been shown to have differing patterns of methylation (Rodríguez López *et al.* 2012).

**Table 4.2:** Samples used in MSAP methylation analysis.

| | Brain | Liver | Gonads | Tail |
|---|---|---|---|---|
| **F2** | - | - | - | 10 |
| **L1** | 20 | 20 | 20 | 20 |
| **L2** | 9 | 10 | 7 | 10 |
| **S1** | 14 | 15 | 15 | 13 |
| **S2** | 13 | 11 | 13 | 12 |

Environmental stress is known to lead to changes in DNA methylation (Labra *et al.* 2002, Steward *et al.* 2002, Sollars *et al.* 2003, Anway *et al.* 2005, Meaney and Szyf 2005, Brown *et al.* 2009, Boyko *et al.* 2010, Herrera and Bazaga 2010, Verhoeven *et al.* 2010, Vandegehuchte and Janssen 2011). Although the selection experiments in this study were designed to reduce the stress experienced by the fish, the selection for body size necessitates handling the fish and may therefore have led to increased stress. The collection and transport of wild fish from Trinidad to Bangor also has the potential to have caused stress. Care was taken to ensure the fish from all five selection lines were treated in the same way throughout the experiment, however, it is possible that small variations in stress may have led to variation in the epigenome between the different lines. The use of wild fish also necessitated a change of diet for the first generation of fish. As nutrition has been shown to cause changes in epigenetic marks, particularly in the embryo (Wolff *et al.* 1998, Ashworth *et al.* 2009, Chmurzynska 2010), it is possible that the change in diet may have had an effect, although it is unlikely that this will have differed between the selection lines.

Variability in quality of DNA extracts meant that all four tissue types could not be analysed for all individuals. The number of samples of each tissue type and from each selection line, including the $F_2$, can be seen in Table 4.2.

### 4.3.2 Methylation sensitive AFLP

For each sample 100 ng of genomic DNA was digested with *Eco*RI and one of the two isoschizomers, *Hpa*II or *Msp*I. Double stranded adaptors were prepared by combining single stranded oligonucleotides (table 4.3) and heating to 90°C for 5 minutes before cooling by 1.5°C per minute for 45 minutes. Twenty µl digestion/ligation reactions were incubated at 37°C for 3 hours and each contained: 5 U of *Eco*RI, 5 U of *Hpa*II/*Msp*I, 2 µl of restriction enzyme buffer (NEB buffer 1 for *Hpa*II and NEB buffer 4 for *Msp*I), 5pmol of *Eco*RI adaptor, 50 pmol of *Hpa*II/*Msp*I adaptor, 0.6 U of T4 DNA ligase and 2 µl of T4 DNA ligase buffer. For each sample, both the *Hpa*II and the *Msp*I digestion was replicated twice, meaning that all samples were digested in a total of four reactions.

For each digest a pre-selective and selective PCR were then performed, each in a total reaction volume of 10 µl comprising: 0.625 U Thermostart™ *taq* DNA polymerase, 1x Thermostart™ reaction buffer, 0.8 mM dNTPs, 1.5 mM $MgCl_2$, 0.2 µl of *Eco*RI primer (10 µM) and 0.2 µl of either the *Hpa*II or *Msp*I primer (10 µM). The pre-selective PCR contained 1 µl of product of the digestion/ligation reaction, while the selective PCR contained 1 µl of the pre-selective PCR product. Sequences of the primers used in the pre-selective and selective PCR can be found in table 4.3. All PCR's were carried out in a BioRad DNA engine Tetrad 2 Peltier Thermal Cycler with the following thermal profile: 5 minutes at 94°C; 35 cycles of 30 seconds at 92°C, 1 minute 30 seconds at 52°C and 30 seconds at 72°C; 10 minutes at 72°C.

**Table 4.3:** Samples used in MSAP methylation analysis.

| Name | Function | Sequence |
|---|---|---|
| Ad.HpaII/MspI + FW | Adaptor | GGTTCTAGACTCATC |
| Ad.HpaII/MspI + RV | Adaptor | GACGATGAGTCTAGAA |
| Ad.EcoRI + FW | Adaptor | CTCGTAGACTGCGTACC |
| Ad.EcoRI + RV | Adaptor | AATTGGTACGCAGTCTAC |
| Pre.EcoRI | Pre-selective primer | GACTGCGTACCAATTCA |
| Pre.HpaII/MspI | Pre-selective primer | GATGAGTCTAGAACGGT |
| EcoRI + ACT | Selective Primer | GACTGCGTACCAATTCACT |
| HpaII/MspI + TAC | Selective Primer | GATGAGTCTAGAACGGTAC |

The products of the digestion/ligation, pre-selective PCR and selective PCR were initially visualised with a 5% agarose gel to check for amplification success before the selective PCR products were resolved on 36 cm capillaries using POP7 polymer in a Applied Biosystems 3130 Bioanalyser (Applied Biosystems, Foster City, CA) using the 500 LIZ size standard.

### 4.3.3 Statistical analysis

The resulting MSAP banding patterns were analysed in GeneMapper v4 (Applied Biosystems, Foster City, CA). To reduce the potential impact of size homoplasy (Caballero *et al.* 2008), only bands larger than 100 bp or smaller than 500 bp were considered. The threshold for band detection

was set at 50 relative fluorescent units (RFU) however all sites which were thought to be methylated (i.e. *Hpa*II+/MpsI- or *Hpa*II-/*Msp*I+ fragments) were also visually assessed.

To give an estimate of the number of potential EcoRI-HpaII/MspI sites present in the guppy genome an in silco digest of the reference genome was performed using the R package SimRAD (Lepais and Weir 2014). This analysis identified a total of 26,812 fragments containing both an EcoRI and a MspII/HpaII cut site within the selected size range of 100-500bp. However, as the reference genome is not complete, this is likely to be a underestimate of the number of potential fragments.

When examining the individual banding pattern at each site, any band which was present in at least one of the two replicates was called as present (1) while a band had to be absent from both replicates to be called as absent (0). Only bands with repeatability (1 - frequency discordant states between replicates) of ≥ 0.9 were considered for further analysis. Once the presence/absence of each band was determined at each site, the methylation state of each locus in each individual was determined. As discussed in section 4.2.3, the use of the *Msp*I and *Hpa*II enzymes can produce four different banding patterns, three of which are informative (Table 4.1) in mammals. For the purposes of this study the banding patterns will be referred to as: unmethylated, in which a site is successfully cut by both enzymes (H+ / M+); hemi-methylated, in which a site is only cut by the *Hpa*II enzyme (H+ / M-); fully methylated, in which a site is cut only by the *Msp*I enzyme (H- / M+) and uninformative, in which a site is not cut by either enzyme (H- / M-).

To identify those sites susceptible to methylation, an error threshold ($e_T$) was calculated where $e_T = e_m + e_h - 2\, e_m e_h$ when $e_m$ is the number of discordant *Msp*I scores and $e_h$ is the number of discordant *Hpa*II scores (Herrera and Bazaga 2010). When calculated across all bands with a repeatability of ≥0.9, the $e_T$ was 0.07. Sites with a methylation frequency above this threshold were classed as methylation susceptible loci (MSL) while those with a methylation frequency below were classed as non-methylated loci (NML). The methylation state of each MSL in each individual was then coded into three binary epi-loci representing fully methylated, hemi-methylated or unmethylated states. Uninformative sites are coded as 0 for all three epiloci. The banding pattern, methylation state and resulting epiloci coding can be seen in table 4.4. To code the MSL into epiloci the 'Mix2' algorithm in the R script *msap_calc* was used (Schulz *et al.* 2013).

**Table 4.4:** MSAP banding patterns and the resulting epiloci codes.

| Banding pattern | Epiloci | | |
|---|---|---|---|
| | H | M | U |
| HpaII +/ MspI + | 0 | 0 | 1 |
| HpaII +/ MspI - | 0 | 1 | 0 |
| HpaII -/ MspI + | 1 | 0 | 0 |
| HpaII -/ MspI - | 0 | 0 | 0 |

Once the methylation state of each individual had been converted into epiloci, variability in the pattern of methylation between tissue types and selection lines could be assessed. Variability among tissues was considered by grouping all $F_6$ samples according to tissue type. Variability

among the selection lines was analysed by grouping according to selection line within each tissue type.

#### 4.3.4 Genome-wide methylation and Epigenetic structure

Variability in genome-wide levels of methylation was assessed by estimating the mean frequency of the fully methylated, hemi-methylated and unmethylated sites for each group. The frequency of 'hemi-methylated' sites described in vertebrates using the MSAP technique has led to discussion about other ways this banding pattern could be formed (Fulneček and Kovařík 2014). It is possible that these bands result from a fragment containing both a fully methylated and unmethylated site (see section 4.5.1). To ensure that separating these sites did not bias our analysis when considering the genome wide level of methylation, we also analysed the frequency of fully methylated/hemi-methylated sites combined and unmethylated/hemi-methylated sites combined. The Shannon diversity index, calculated as $i = \sum pi \log_2 pi$ where $pi$ is the frequency of a present ('1') band the population, was used to assess the epigenetic diversity in each group. Statistical significance of any observed variation between methylation frequency and epigenetic diversity was assessed using a Kruskal-Wallis test in SPSS (Version 22.0).

The epigenetic structure, both between tissue types and between selection lines within tissues, was assessed using pairwise Φpt values and Principal Coordinates Analysis (PCoA). Φpt is a measure analogous to $F_{st}$ which can be used for binary or haploid data (Maguire *et al.* 2002). PCoA and Φpt values were both calculated in GenAlEx 6.501 (Peakall and Smouse 2012).

#### 4.3.5 Level of methylation at individual epiloci

As well as examining genome-wide levels of methylation, the patterns of methylation between selection lines (within each tissue type), at individual epiloci were also investigated. GenAlEx 6.501 (Peakall and Smouse 2012) was used to calculate individual epilocus Φpt values for each pairwise comparison (L1/S1, L1/S2, L2/S1 and L2/S2 for all four tissue types, as well as $F_2$/L1, $F_2$/L2, $F_2$/S1 and $F_2$/S2 for the samples from the tail tissue).

In addition to pairwise Φpt values, $F_{st}$ outlier analyses (Fdist and Bayesian approaches) were also used to identify epiloci showing elevated levels of differentiation between the selection lines. The Fdist approach was implemented in Mcheza (Antao and Beaumont 2011). Mcheza utilises DFdist, a modified version of Fdist, which enables dominant markers to be analysed. Allele frequencies are calculated according to Zhivotovsky (1999) before the distribution of $F_{st}$ values expected under neutral conditions is determined, allowing the identification of loci falling outside the neutral distribution. Mcheza was run with 50,000 simulations and a forced neutral mean $F_{st}$. Epiloci with a p value ≤0.05 were identified as outliers. A Bayesian approach, as implemented in BayeScan, examines the likelihood of two models of differentiation, one with and one without selection. BayeScan was run using the default parameters (20 pilot runs of 5,000 iterations and an additional

burn in of 50,000 iterations followed by 100,000 iterations) and any locus with a q-value ≤ 0.05 was considered an outlier.

## 4.4 Results

A total of 91 bands had a repeatability of ≥ 0.9. Across all 91 bands the $e_T$ was 0.07 and resulted in 80 MSL and 11 NML. Because the number of NML was low and the genetic diversity of the selection lines has already been analysed, the 11 NML were not considered further. Of the 80 MSL, 15 were uninformative in more than 30% of the samples processed and were therefore not included in further analysis. To assess the patterns of methylation in guppies, 65 MSL were transformed by MSAP_calculation into 162 polymorphic binary epiloci.

### 4.4.1 Analysis of tissue types

#### 4.4.1.1 Frequency of methylation

The frequency of both hemi-methylation and full methylation varied significantly among tissue types ($X^2$=19.56, p<0.001 and $X^2$=16.87, p=0.001 respectively). Similarly, the frequency of full/ hemi-methylated sites combined and unmethylated/hemi-methylated sites combined varied significantly among tissues ($X^2$=11.37, p=0.010 and $X^2$=13.44, p=0.004 respectively). The frequency of genome wide hemi-methylation was lowest in the brain and liver tissues while the frequency of full methylation was lowest in the gonads and the tail (see Figure 4.2). Combining the hemi-methylated sites with the unmehtylated and fully methylated sites reduced the variation between the liver and brain for the fully methylated sites but increased it for the unmehtylated sites. The genome wide absence of methylation also varied significantly with tissue type ($X^2$= 11.67, p=0.009).

#### 4.4.1.2 Epigenetic diversity

Across all selection lines and epiloci the Shannon diversity index varied significantly ($X^2$=9.752, p=0.021). Diversity was higher in the gonad and liver tissue (mean=0.59) than in the brain and tail tissue (mean=0.53) (Figure 4.3). However, when analysing variation in the level of diversity between all four tissue types within a selection line, the observed diversity was only significantly different in the L1 and S2 lines ($X^2$=9.38, p=0.025 and $X^2$=11.34, p=0.01 respectively).

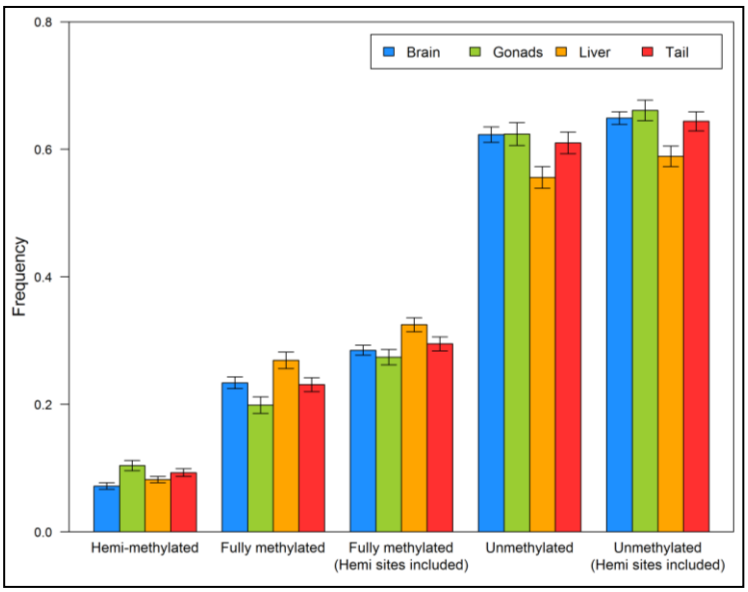
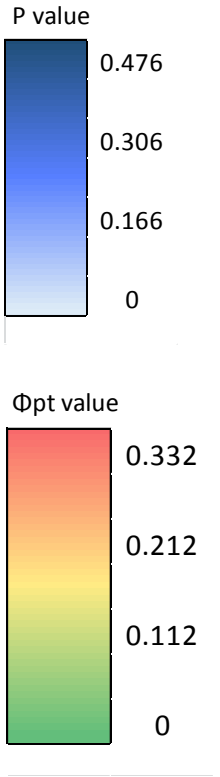



**Figure 4.2:** genome wide methylation frequency across all selection lines. SE shown in error bars

**Table 4.5:** Φpt values and their significance between the selection lines and tissues. Φpt values are shown below the diagonal, p values above the diagonal

| | | Brain | | | | Gonads | | | | Liver | | | | Tail | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | L1 | L2 | S1 | S2 | L1 | L2 | S1 | S2 | L1 | L2 | S1 | S2 | L1 | L2 | S1 | S2 |
| Brain | L1 | | 0.006 | 0.011 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | L2 | 0.083 | | 0.010 | 0.001 | 0.001 | 0.012 | 0.001 | 0.001 | 0.001 | 0.009 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | S1 | 0.044 | 0.061 | | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | S2 | 0.119 | 0.124 | 0.144 | | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| Gonads | L1 | 0.150 | 0.121 | 0.207 | 0.218 | | 0.003 | 0.003 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.033 | 0.001 | 0.001 |
| | L2 | 0.177 | 0.053 | 0.190 | 0.169 | 0.062 | | 0.001 | 0.001 | 0.001 | 0.021 | 0.001 | 0.001 | 0.001 | 0.476 | 0.001 | 0.002 |
| | S1 | 0.145 | 0.156 | 0.171 | 0.193 | 0.059 | 0.142 | | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.046 | 0.004 | 0.016 | 0.001 |
| | S2 | 0.164 | 0.164 | 0.221 | 0.099 | 0.113 | 0.092 | 0.099 | | 0.001 | 0.001 | 0.001 | 0.016 | 0.001 | 0.007 | 0.067 | 0.441 |
| Liver | L1 | 0.198 | 0.193 | 0.258 | 0.232 | 0.094 | 0.132 | 0.178 | 0.163 | | 0.009 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | L2 | 0.247 | 0.109 | 0.262 | 0.223 | 0.139 | 0.052 | 0.222 | 0.171 | 0.094 | | 0.002 | 0.010 | 0.001 | 0.021 | 0.002 | 0.001 |
| | S1 | 0.190 | 0.183 | 0.226 | 0.202 | 0.151 | 0.155 | 0.151 | 0.143 | 0.057 | 0.114 | | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | S2 | 0.202 | 0.183 | 0.241 | 0.112 | 0.176 | 0.140 | 0.200 | 0.063 | 0.106 | 0.095 | 0.079 | | 0.001 | 0.001 | 0.001 | 0.001 |
| Tail | L1 | 0.173 | 0.176 | 0.220 | 0.229 | 0.039 | 0.137 | 0.025 | 0.098 | 0.147 | 0.240 | 0.164 | 0.209 | | 0.001 | 0.008 | 0.001 |
| | L2 | 0.177 | 0.092 | 0.218 | 0.191 | 0.037 | 0.000 | 0.101 | 0.062 | 0.154 | 0.109 | 0.174 | 0.167 | 0.084 | | 0.002 | 0.001 |
| | S1 | 0.201 | 0.181 | 0.221 | 0.189 | 0.096 | 0.123 | 0.056 | 0.032 | 0.171 | 0.205 | 0.129 | 0.139 | 0.062 | 0.083 | | 0.008 |
| | S2 | 0.274 | 0.267 | 0.332 | 0.219 | 0.191 | 0.190 | 0.184 | 0.000 | 0.242 | 0.266 | 0.223 | 0.116 | 0.155 | 0.135 | 0.070 | |



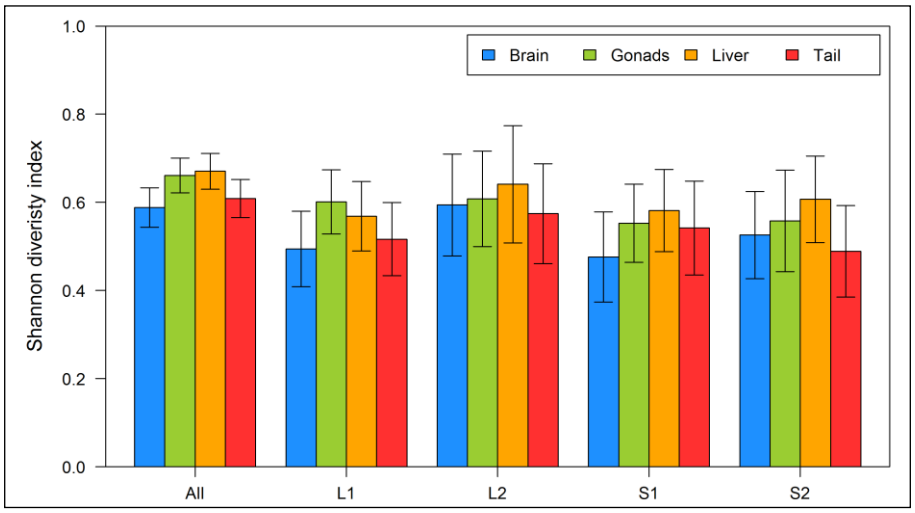


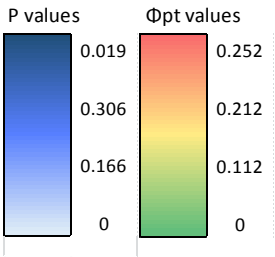

**Figure 4.3:** Shannon diversity index across tissue and selection lines. All shows the data from all selection lines combined.

4.4.1.3 Epigenetic structure

When calculated using all selection lines combined, Φpt between tissues were all significant and ranged from 0.017 to 0.149. As Table 4.5 shows, epigenetic differentiation was high between the brain and the three other tissues sampled and also between the liver and tail tissues. Epigenetic differentiation between the gonads and the liver and the gonads and the tail was considerably lower. The pattern of genetic differentiation observed when the selection lines were combined (L1 and L2 combined and S1 and S2 combined), was the same as when the samples were grouped by selection line and tissue type (Table 4.6). In all four lines the brain showed the highest level of epigenetic differentiation. The liver and the tail showed no epigenetic differentiation in two of the lines (L2 and S2) and low epigenetic differentiation in the remaining two lines (L1 and S1). Values of Φpt observed between the tissue types were significantly higher than those observed between the selection lines within tissues (Mann Whitney U= 381, p ≤ 0.001).

**Table 4.6:** Φpt values and their significance between the selection regimes and tissues (L= L1 and L2 combined, S= S1 and S2 combined). Φpt values are shown below the diagonal, p values above the diagonal

| | | Brain | | Gonads | | Liver | | Tail | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | L | S | L | S | L | S | L | S | F2 |
| Brain | L | | 0.007 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | S | 0.028 | | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| Gonads | L | 0.110 | 0.162 | | 0.001 | 0.001 | 0.001 | 0.011 | 0.001 | 0.001 |
| | S | 0.115 | 0.119 | 0.060 | | 0.001 | 0.001 | 0.019 | 0.006 | 0.001 |
| Liver | L | 0.164 | 0.201 | 0.078 | 0.142 | | 0.001 | 0.001 | 0.001 | 0.001 |
| | S | 0.157 | 0.152 | 0.130 | 0.106 | 0.050 | | 0.001 | 0.001 | 0.001 |
| Tail | L | 0.131 | 0.170 | 0.022 | 0.022 | 0.128 | 0.146 | | 0.001 | 0.001 |
| | S | 0.203 | 0.198 | 0.122 | 0.036 | 0.185 | 0.120 | 0.072 | | 0.001 |
| | F2 | 0.252 | 0.225 | 0.201 | 0.153 | 0.201 | 0.167 | 0.211 | 0.210 | |



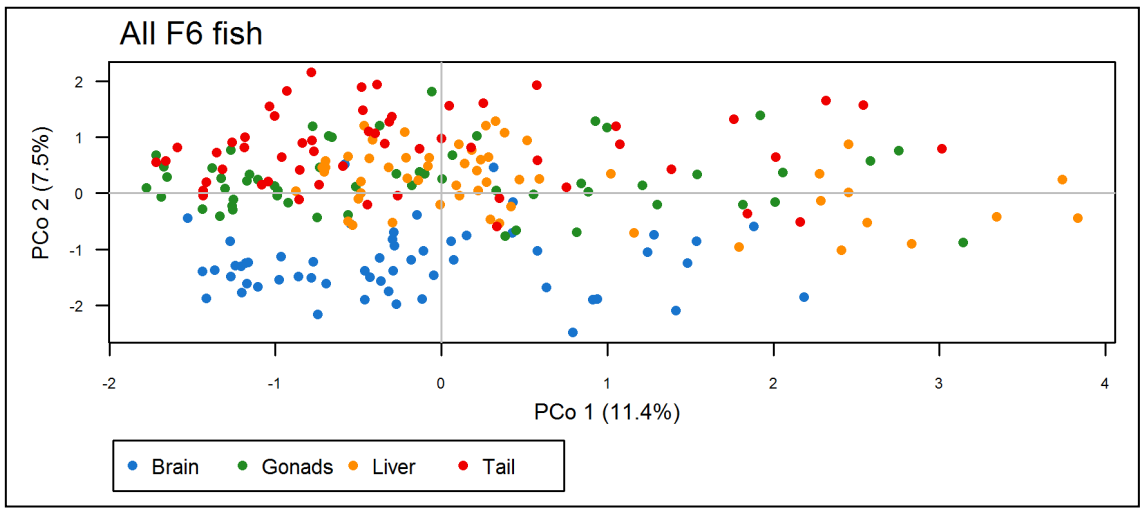



**Figure 4.4:** PCoA showing structure of tissues.

4.4.1.4 Principal coordinates analysis

PCoA of epigenetic variation between tissue types broadly agreed with the structure indicated by the Φpt values. When plotted across all selection lines, the brain tissue clustered separately from the other three tissues (Figure 4.4). Despite the high Φpt values between the liver and the tail tissues, the PCoA did not cluster these tissues separately.

## 4.4.2 Analysis of selection lines

As the pattern of methylation varied between the different tissue types sampled, analysis of methylation in the selection lines and in the $F_2$ fish was performed on each tissue type separately.

4.4.2.1 Frequency of methylation

The frequency of methylation for each selection line in each tissue type are illustrated in Figure 4.5. Table 4.7 shows the results of Kruskall Wallis tests of the variation between the selection lines and the $F_2$. In the brain tissues methylation types (hemi-methylated, fully methylated, unmethylated, hemi-methylated/fully methylated combined, and unmethylated/hemi-methylated combined) did not differ significantly across the $F_6$ selection lines (L1, L2, S1 and S2). In the gonads, all five types of methylation varied significantly when considering all four selection lines. However, when the two large lines were combined and compared against the combination of the two small lines, variation in only the frequency of hemi-methylation remained significant. In the liver the frequency of hemi-

**Table 4.7:** Significant differences in the frequency of methylation between the selection lines.

| Comparison | Selection lines | | | | | L vs S | | | | | F2 vs L | | | | | F2 vs S | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Methylation type | H | F | U | F + H | U + H | H | F | U | F + H | U + H | H | F | U | F + H | U + H | H | F | U | F + H | U + H |
| Brain | NS | NS | NS | NS | NS | NS | NS | NS | NS | NS | - | - | - | - | - | - | - | - | - | - |
| Gonads | ** | * | * | * | * | *** | NS | NS | NS | NS | - | - | - | - | - | - | - | - | - | - |
| Liver | * | NS | NS | NS | NS | NS | NS | NS | NS | NS | - | - | - | - | - | - | - | - | - | - |
| Tail | * | * | * | NS | * | NS | ** | * | * | ** | NS | NS | ** | NS | ** | * | ** | NS | * | NS |

methylation varied significantly between all the selection lines. This variation, however, was not significant when the large and small lines were combined, indicating a treatment-specific effect.

In tail tissues, the frequency of hemi-methylation, full methylation, no methylation and no methylation/hemi-methylated combined all varied significantly between the selection lines (Table 4.7). When the large and small selection lines were combined the frequency of full methylation varied significantly between the $F_2$ and the large lines (L1 and L2 combined) and the $F_2$ and the small lines (S1 and S2 combined) (Table 4.7). The frequency of no methylation only varied significantly between the $F_2$ and the large lines, while the frequency of both hemi-methylation only varied significantly between the $F_2$ and the small lines (S1 and S2 combined). For both F2 comparisons combining the hemi-methylated sites with the fully methylated and unmethylated sites did not change the significance of the result.

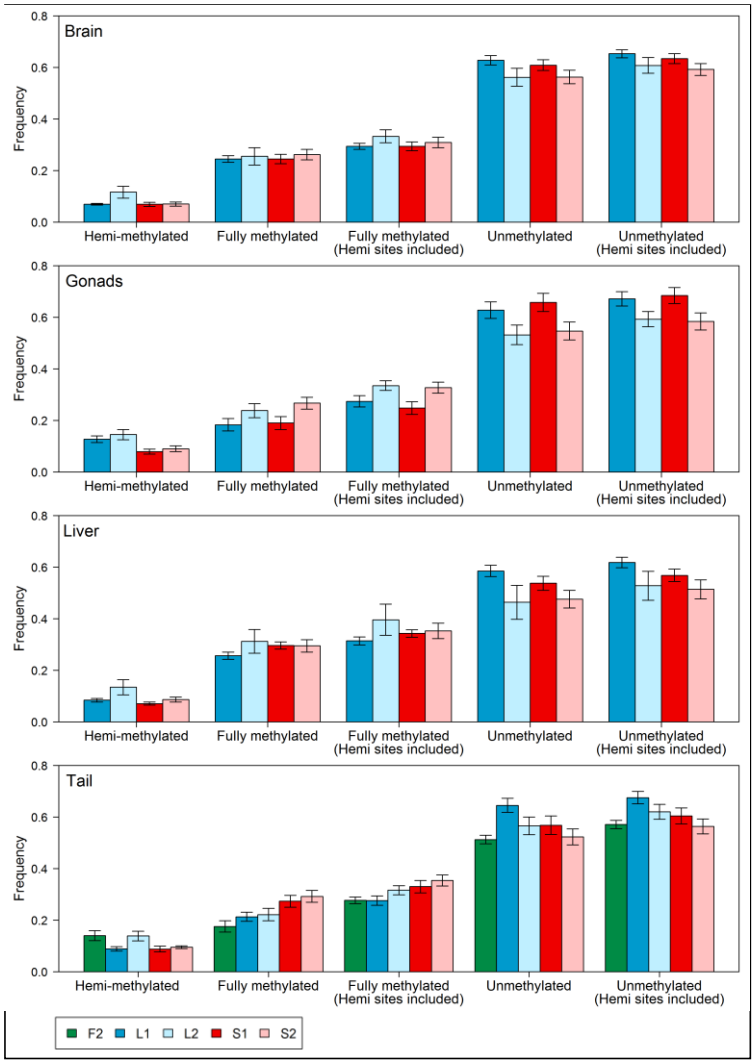



**Figure 4.5:** genome wide methylation frequency in each selection line for each tissue. SE shown in error bars

4.4.2.2 Epigenetic diversity

Epigenetic diversity for the 4 selection lines, as measured by the Shannon diversity index (*i*), is illustrated in Figure 4.3, and the results of the Kruskall Wallis tests in Table 4.8. For the $F_2$ tail tissue (not included in figure 4.3) the Shannon diversity index is 0.573 (SE ± 0.121). The observed variation in the epigenetic diversity between the selection lines was only significant in the brain. However, the diversity was highest in the L2 and S2 and lowest in the L1 and S1. Therefore when the large and small lines were combined, there was no significant difference in epigenetic diversity. Conversely, the epigenetic diversity in the gonads was not significantly different when all four selection lines were considered separately, but did differ significantly when the large (L1 and L2) and small lines (S1 and S2) were combined. Epigenetic diversity in the gonads was higher in the

large selected lines than in the small selected lines. The $F_2$ did not differ significantly in either the large (L1 and L2 combined) or small lines (S1 and S2 combined).

4.4.2.3 Epigenetic structure

Although the level of epigenetic differentiation between the selection lines (from the same tissue type) was significantly lower than that observed between the different tissue types (Mann Whitney U=735, p≤0.001), the Φpt values observed between the selection lines were still high and significant (Table 4.5). The Φpt values between the selection line replicates (e.g. between L1 and L2) were similar to those between lines from opposing selection regimes (e.g. between L1 and S1). When the replicates were combined, the Φpt values were low (Table 4.6). The high level of epigenetic differentiation between the four selection lines (L1, L2 S1 and S2), as well as the low level of epigenetic differentiation between the selection regimes (L and S), suggest that any epigenetic structuring between the selection lines in our study was driven by selection on body size. Φpt values were high and significant between the $F_2$ and the selection lines, both when the replicates were analysed separately (Table 4.5) and when the replicates were combined, ($F_2$ vs L: Φpt = 0.211 p= 0.001, $F_2$ vs S: Φpt = 0.210 p= 0.001).

**Table 4.8:** Significant differences in the diversity of methylation between the selection lines.

| Comparison | Selection lines | L vs S | F2 vs L | F2 vs S |
|---|---|---|---|---|
| **Brain** | ** | NS | - | - |
| **Gonads** | NS | * | - | - |
| **Liver** | NS | NS | - | - |
| **Tail** | NS | NS | NS | NS |

As well as global Φpt values, Φpt values between the selection lines were calculated for each epilocus. Across all of the pairwise comparisons for each epilocus (20 comparisons for 162 epiloci), only 12% exhibited a significantly elevated Φpt ($p \leq 0.05$). In order to be potentially linked to body size, we expected a methylated site to show divergence in more than one of the selected replicates and, therefore, have a significantly elevated Φpt in at least 3 of the pairwise comparisons which were considered for each tissue type. When considering the samples taken from the brain, the examined epiloci did not exhibit a significantly elevated Φpt in more than two of the pairwise comparisons (Table 4.9). In tissues from the gonads, four epiloci (h107, u107, m104, u104), representing 2 MSL, were significantly elevated in 3 of the 4 pairwise comparisons while one further epilocus (u337) had a significantly elevated Φpt in all 4 pairwise comparisons (Table 4.10). Pairwise comparisons of samples from the liver showed one (m312) epilocus which had a significantly elevated Φpt in 3 comparisons and one which had a significantly elevated Φpt in all 4 pairwise comparisons (Table 4.11). In the tail tissue only 3 epiloci (u176, m176, h119), which represented 2 MSL, had a significantly elevated Φpt in 3 of the pairwise comparisons (Table 4.12).

**Table 4.9:** Epilocus by epilocus Φpt values from brain tissue samples. Lowest values are shown in green, highest in red. Only those loci which are significantly elevated in at least one comparison are shown.

| | L1/S1 | | L1/S2 | | L2/S1 | | L2/S2 | |
|---|---|---|---|---|---|---|---|---|
| **u102** | 0.20 | | 0.57 | ** | 0.00 | | 0.26 | |
| **m102** | 0.20 | | 0.57 | *** | 0.00 | | 0.10 | |
| **u104** | 0.19 | * | 0.00 | | 0.00 | | 0.01 | |
| **h119** | 0.00 | | 0.46 | ** | 0.06 | | 0.65 | ** |
| **u119** | 0.00 | | 0.39 | ** | 0.25 | | 0.71 | *** |
| **m136** | 0.29 | * | 0.00 | | 0.00 | | 0.25 | |
| **u136** | 0.29 | * | 0.00 | | 0.00 | | 0.39 | * |
| **h165** | 0.09 | | 0.08 | | 0.00 | | 0.52 | ** |
| **u165** | 0.36 | ** | 0.00 | | 0.00 | | 0.35 | * |
| **m180** | 0.00 | | 0.00 | | 0.41 | * | 0.38 | * |
| **u222** | 0.00 | | 0.35 | * | 0.00 | | 0.10 | |
| **m222** | 0.00 | | 0.35 | * | 0.00 | | 0.26 | |
| **u224** | 0.07 | | 0.18 | | 0.43 | * | 0.29 | |
| **m224** | 0.00 | | 0.18 | | 0.41 | * | 0.15 | |
| **u226** | - | | 0.32 | * | - | | 0.20 | |
| **m226** | - | | 0.32 | * | 0.19 | | 0.00 | |
| **u244** | 0.00 | | 0.08 | | 0.14 | | 0.40 | * |
| **m244** | 0.00 | | 0.36 | ** | 0.06 | | 0.56 | ** |
| **m259** | 0.14 | | 0.00 | | 0.41 | * | 0.03 | |
| **u259** | 0.09 | | 0.29 | * | 0.41 | * | 0.00 | |
| **u263** | 0.00 | | 0.27 | * | 0.00 | | 0.19 | |
| **u275** | 0.00 | | 0.23 | * | 0.09 | | 0.00 | |
| **h275** | 0.15 | | 0.54 | *** | 0.17 | | 0.53 | ** |
| **m278** | - | | 0.23 | * | 0.05 | | 0.00 | |
| **u312** | 0.05 | | 0.46 | ** | 0.00 | | 0.65 | ** |
| **m312** | 0.00 | | 0.42 | ** | 0.02 | | 0.68 | *** |
| **m318** | 0.00 | | 0.03 | | 0.48 | * | 0.67 | *** |
| **u318** | 0.01 | | 0.08 | | 0.68 | ** | 0.78 | *** |
| **u326** | 0.00 | | 0.18 | | 0.33 | * | 0.00 | |
| **u336** | 0.00 | | 0.36 | ** | 0.01 | | 0.15 | |
| **m336** | 0.00 | | 0.36 | ** | 0.01 | | 0.15 | |
| **u341** | 0.01 | | 0.12 | | 0.00 | | 0.38 | * |
| **m341** | 0.01 | | 0.12 | | 0.00 | | 0.38 | * |
| **u351** | 0.08 | | 0.47 | *** | 0.00 | | 0.18 | |
| **m356** | 0.37 | ** | 0.00 | | 0.05 | | 0.26 | |
| **u356** | 0.50 | *** | 0.00 | | 0.13 | | 0.12 | |
| **u359** | 0.00 | | 0.58 | ** | 0.15 | | 0.44 | * |
| **u363** | 0.00 | | 0.36 | ** | 0.00 | | 0.03 | |
| **h384** | 0.00 | | 0.00 | | 0.33 | * | 0.32 | |
| **m384** | 0.05 | | 0.01 | | 0.55 | ** | 0.03 | |

**Table 4.10:** Epilocus by epilocus Φpt values from gonad tissue samples. Lowest values are shown in green, highest in red. Only those loci which are significantly elevated in at least one comparison are shown.

| | L1/S1 | | L1/S2 | | L2/S1 | | L2/S2 | |
|---|---|---|---|---|---|---|---|---|
| **h107** | 0.65 | *** | 0.36 | ** | 0.62 | ** | 0.30 | |
| **h119** | 0.00 | | 0.37 | ** | 0.00 | | 0.27 | |
| **h149** | 0.23 | * | 0.00 | | 0.62 | *** | 0.11 | |
| **h165** | 0.13 | | 0.31 | * | 0.00 | | 0.00 | |
| **h166** | 0.06 | | 0.33 | * | 0.59 | ** | 0.00 | |
| **h275** | 0.10 | | 0.25 | * | 0.00 | | 0.00 | |
| **h286** | 0.13 | | 0.01 | | 0.30 | * | 0.14 | |
| **h372** | 0.03 | | 0.00 | | 0.49 | ** | 0.04 | |
| **m104** | 0.18 | | 0.33 | * | 0.36 | * | 0.53 | * |
| **m161** | 0.00 | | 0.40 | ** | 0.00 | | 0.38 | * |
| **m176** | 0.02 | | 0.60 | *** | 0.00 | | 0.48 | * |
| **m213** | 0.00 | | 0.21 | | 0.00 | | 0.41 | * |
| **m224** | 0.11 | | 0.04 | | 0.47 | ** | 0.37 | * |
| **m278** | 0.21 | * | 0.03 | | 0.10 | | 0.00 | |
| **m312** | 0.28 | * | 0.10 | | 0.52 | ** | 0.27 | |
| **m318** | 0.18 | | 0.69 | *** | 0.10 | | 0.59 | ** |
| **m337** | 0.26 | * | 0.20 | | 0.35 | * | 0.27 | |
| **m341** | 0.00 | | 0.15 | | 0.00 | | 0.46 | ** |
| **m345** | 0.00 | | 0.00 | | 0.30 | * | 0.25 | |
| **m351** | 0.00 | | 0.56 | *** | 0.03 | | 0.27 | |
| **m384** | 0.00 | | 0.37 | ** | 0.00 | | 0.00 | |
| **m386** | 0.10 | | 0.30 | * | 0.00 | | 0.00 | |
| **u104** | 0.18 | | 0.33 | * | 0.36 | * | 0.53 | ** |
| **u107** | 0.73 | *** | 0.37 | * | 0.62 | *** | 0.19 | |
| **u119** | 0.00 | | 0.19 | | 0.01 | | 0.38 | * |
| **u140** | 0.03 | | 0.00 | | 0.29 | * | 0.04 | |
| **u166** | 0.00 | | 0.24 | | 0.39 | ** | 0.00 | |
| **u176** | 0.02 | | 0.60 | ** | 0.00 | | 0.48 | ** |
| **u199** | 0.00 | | 0.50 | ** | 0.14 | | 0.12 | |
| **u213** | 0.30 | * | 0.00 | | 0.01 | | 0.11 | |
| **u224** | 0.12 | | 0.13 | | 0.49 | ** | 0.51 | * |
| **u244** | 0.10 | | 0.00 | | 0.26 | * | 0.12 | |
| **u245** | 0.25 | * | 0.18 | | 0.00 | | 0.00 | |
| **u278** | 0.21 | * | 0.03 | | 0.00 | | 0.00 | |
| **u286** | 0.13 | | 0.01 | | 0.30 | * | 0.14 | |
| **u304** | 0.00 | | 0.05 | | 0.23 | | 0.35 | * |
| **u312** | 0.01 | | 0.00 | | 0.53 | ** | 0.53 | ** |
| **u318** | 0.04 | | 0.40 | ** | 0.00 | | 0.24 | |
| **u336** | 0.02 | | 0.29 | * | 0.36 | * | 0.01 | |
| **u337** | 0.26 | * | 0.34 | * | 0.35 | * | 0.46 | * |
| **u341** | 0.00 | | 0.15 | | 0.01 | | 0.35 | * |
| **u345** | 0.00 | | 0.05 | | 0.44 | ** | 0.19 | |
| **u351** | 0.00 | | 0.43 | ** | 0.00 | | 0.15 | |
| **u356** | 0.37 | ** | 0.00 | | 0.35 | * | 0.00 | |
| **u359** | 0.00 | | 0.52 | *** | 0.34 | * | 0.02 | |
| **u372** | 0.08 | | 0.08 | | 0.49 | ** | 0.00 | |
| **u384** | 0.01 | | 0.19 | | 0.53 | ** | 0.00 | |
| **u386** | 0.18 | | 0.51 | ** | 0.00 | | 0.19 | |

D

**Table 4.11:** Epilocus by epilocus Φpt values from liver tissue samples. Lowest values are shown in green, highest in red. Only those loci which are significantly elevated in at least one comparison are shown.

| | L1/S1 | | L1/S2 | | L2/S1 | | L2/S2 | |
|---|---|---|---|---|---|---|---|---|
| **u104** | 0.19 | * | 0.13 | | 0.35 | | 0.29 | |
| **m104** | 0.19 | * | 0.13 | | 0.35 | | 0.29 | |
| **u107** | 0.41 | ** | 0.00 | | 0.35 | | 0.00 | |
| **h107** | 0.49 | ** | 0.01 | | 0.45 | * | 0.00 | |
| **m122** | 0.21 | * | 0.00 | | 0.00 | | 0.09 | |
| **u154** | 0.28 | * | 0.27 | * | 0.04 | | 0.02 | |
| **m154** | 0.28 | * | 0.27 | * | 0.04 | | 0.02 | |
| **m161** | 0.00 | | 0.46 | ** | 0.00 | | 0.31 | |
| **u166** | 0.02 | | 0.31 | * | 0.08 | | 0.29 | |
| **h166** | 0.06 | | 0.35 | * | 0.00 | | 0.44 | * |
| **u168** | - | | 0.32 | * | 0.32 | | 0.00 | |
| **u176** | - | | 0.32 | * | - | | 0.17 | |
| **m176** | - | | 0.32 | * | - | | 0.17 | |
| **h180** | - | | - | | 0.49 | * | 0.46 | * |
| **u199** | 0.00 | | 0.46 | *** | 0.26 | | 0.00 | |
| **m224** | 0.27 | * | 0.53 | *** | 0.26 | | 0.59 | ** |
| **m240** | 0.00 | | 0.00 | | 0.37 | | 0.44 | * |
| **u243** | 0.00 | | 0.29 | * | 0.01 | | 0.00 | |
| **m243** | 0.00 | | 0.29 | * | 0.01 | | 0.00 | |
| **u244** | 0.04 | | 0.00 | | 0.27 | | 0.50 | * |
| **h244** | - | | 0.03 | | 0.49 | * | 0.24 | |
| **u259** | 0.00 | | 0.53 | ** | 0.19 | | 0.00 | |
| **m263** | 0.06 | | 0.13 | | 0.31 | * | 0.42 | * |
| **m278** | 0.33 | * | 0.31 | ** | - | | - | |
| **u278** | 0.33 | ** | 0.31 | ** | - | | - | |
| **u280** | 0.10 | | 0.25 | * | 0.00 | | 0.00 | |
| **m280** | 0.10 | | 0.25 | * | 0.00 | | 0.00 | |
| **m304** | - | | 0.23 | * | 0.12 | | 0.00 | |
| **u312** | 0.18 | | 0.53 | *** | 0.25 | | 0.59 | ** |
| **m312** | 0.50 | ** | 0.56 | ** | 0.72 | *** | 0.79 | *** |
| **m318** | 0.00 | | 0.05 | | 0.25 | | 0.59 | ** |
| **u326** | 0.02 | | 0.05 | | 0.49 | * | 0.00 | |
| **u338** | 0.00 | | 0.08 | | 0.00 | | 0.50 | * |
| **m338** | 0.08 | | 0.08 | | 0.00 | | 0.44 | * |
| **u345** | 0.00 | | 0.00 | | 0.45 | * | 0.19 | |
| **m353** | 0.21 | * | 0.18 | * | 0.00 | | 0.00 | |
| **m356** | 0.03 | | 0.08 | | 0.49 | * | 0.00 | |
| **u356** | 0.21 | * | 0.01 | | 0.66 | *** | 0.00 | |
| **u359** | 0.03 | | 0.35 | ** | 0.00 | | 0.24 | |
| **u363** | 0.00 | | 0.29 | * | 0.00 | | 0.20 | |
| **u368** | 0.40 | ** | 0.13 | | 0.25 | | 0.02 | |
| **m368** | 0.40 | ** | 0.13 | | 0.25 | | 0.02 | |
| **h372** | 0.00 | | 0.00 | | 0.66 | ** | 0.75 | ** |
| **u379** | 0.00 | | 0.22 | * | 0.00 | | 0.00 | |
| **m384** | 0.36 | ** | 0.00 | | 0.37 | | 0.00 | |
| **m386** | 0.00 | | 0.27 | * | 0.00 | | 0.02 | |
| **u386** | 0.05 | | 0.45 | ** | 0.00 | | 0.19 | |
| **m395** | 0.10 | | 0.25 | * | 0.00 | | 0.00 | |
| **u395** | 0.04 | | 0.36 | ** | 0.08 | | 0.00 | |

**Table 4.12:** Epilocus by epilocus Φpt values from tail tissue samples for the selection lines. Lowest values are shown in green, highest in red. Only those loci which are significantly elevated in at least one comparison are shown.

| | L1/S1 | | L1/S2 | | L2/S1 | | L2/S2 | |
|---|---|---|---|---|---|---|---|---|
| **u104** | 0.00 | | 0.41 | ** | 0.10 | | 0.69 | ** |
| **m104** | 0.00 | | 0.41 | ** | 0.10 | | 0.69 | *** |
| **u107** | 0.00 | | 0.10 | | 0.38 | * | 0.07 | |
| **h107** | 0.00 | | 0.10 | | 0.38 | * | 0.07 | |
| **u109** | 0.00 | | 0.27 | * | 0.00 | | 0.42 | * |
| **m109** | 0.00 | | 0.27 | * | 0.00 | | 0.42 | * |
| **u119** | 0.19 | | 0.36 | ** | 0.23 | | 0.41 | * |
| **h119** | 0.32 | * | 0.51 | *** | 0.18 | | 0.42 | * |
| **h123** | 0.18 | | 0.00 | | 0.32 | * | 0.00 | |
| **u161** | 0.00 | | 0.11 | | 0.49 | ** | 0.04 | |
| **m161** | 0.08 | | 0.51 | ** | 0.07 | | 0.00 | |
| **h165** | 0.00 | | 0.27 | * | 0.14 | | 0.00 | |
| **h166** | 0.00 | | 0.47 | ** | 0.00 | | 0.40 | * |
| **u176** | 0.03 | | 0.39 | * | 0.38 | * | 0.78 | *** |
| **m176** | 0.03 | | 0.39 | ** | 0.38 | * | 0.78 | ** |
| **m199** | 0.25 | * | 0.12 | | 0.00 | | 0.00 | |
| **u199** | 0.25 | * | 0.57 | *** | 0.00 | | 0.14 | |
| **u213** | 0.32 | ** | 0.40 | ** | 0.00 | | 0.00 | |
| **m213** | 0.54 | *** | 0.61 | *** | 0.07 | | 0.15 | |
| **m226** | 0.32 | * | 0.04 | | 0.00 | | 0.00 | |
| **u243** | 0.35 | * | 0.00 | | 0.29 | | 0.00 | |
| **m243** | 0.35 | * | 0.00 | | 0.29 | | 0.00 | |
| **u244** | 0.00 | | 0.00 | | 0.44 | * | 0.41 | * |
| **m245** | 0.35 | * | 0.02 | | 0.14 | | 0.00 | |
| **u245** | 0.54 | *** | 0.03 | | 0.30 | * | 0.00 | |
| **m259** | 0.00 | | 0.21 | * | 0.19 | | 0.00 | |
| **u259** | 0.00 | | 0.51 | *** | 0.18 | | 0.00 | |
| **u280** | 0.44 | ** | 0.10 | | 0.10 | | 0.00 | |
| **m280** | 0.44 | ** | 0.10 | | 0.10 | | 0.00 | |
| **u312** | 0.00 | | 0.06 | | 0.20 | | 0.40 | * |
| **m318** | 0.00 | | 0.07 | | 0.21 | | 0.61 | *** |
| **u341** | 0.02 | | 0.00 | | 0.38 | * | 0.17 | |
| **m341** | 0.02 | | 0.00 | | 0.38 | * | 0.17 | |
| **u342** | 0.27 | * | 0.00 | | 0.11 | | 0.00 | |
| **u345** | 0.03 | | 0.15 | | 0.46 | ** | 0.33 | * |
| **u351** | 0.08 | | 0.41 | ** | 0.00 | | 0.25 | |
| **m356** | 0.03 | | 0.36 | * | 0.00 | | 0.06 | |
| **u359** | 0.00 | | 0.51 | *** | 0.00 | | 0.42 | * |
| **u368** | 0.32 | * | 0.26 | * | 0.03 | | 0.00 | |
| **m368** | 0.32 | * | 0.26 | * | 0.03 | | 0.00 | |
| **u372** | 0.25 | * | 0.40 | * | 0.00 | | 0.00 | |
| **u384** | 0.00 | | 0.00 | | 0.30 | * | 0.06 | |
| **h384** | 0.08 | | 0.00 | | 0.32 | * | 0.15 | |
| **u386** | 0.00 | | 0.78 | *** | 0.00 | | 0.80 | *** |
| **m386** | 0.00 | | 0.69 | *** | - | | 0.71 | *** |

**Table 4.13:** Epilocus by epilocus Φpt values from tail tissue samples for the F2. Lowest values are shown in green, highest in red. Only those loci which are significantly elevated in at least one comparison are shown.

| | F2/L1 | | F2/L2 | | F2/S1 | | F2/S2 | |
|---|---|---|---|---|---|---|---|---|
| **u102** | 0.21 | | 0.44 | * | 0.32 | | 0.15 | |
| **m102** | 0.21 | | 0.44 | * | 0.32 | * | 0.15 | |
| **m104** | 0.00 | | 0.00 | | 0.10 | | 0.69 | ** |
| **u104** | 0.00 | | 0.00 | | 0.10 | | 0.69 | *** |
| **u107** | 0.00 | | 0.56 | ** | 0.00 | | 0.27 | |
| **h107** | 0.29 | * | 0.89 | *** | 0.29 | | 0.67 | ** |
| **h119** | 0.61 | *** | 0.56 | * | 0.13 | | 0.00 | |
| **u119** | 0.67 | *** | 0.78 | *** | 0.29 | * | 0.16 | |
| **u126** | 0.30 | * | 0.33 | | 0.19 | | 0.37 | * |
| **h126** | 0.30 | * | 0.33 | | 0.38 | * | 0.37 | * |
| **u150** | 0.48 | ** | 0.76 | ** | 0.68 | *** | 0.90 | *** |
| **h150** | 0.76 | *** | 0.89 | *** | 0.79 | *** | 0.90 | *** |
| **u154** | 0.15 | | 0.49 | * | 0.10 | | 0.06 | |
| **m154** | 0.28 | | 0.48 | * | 0.23 | | 0.19 | |
| **m161** | 0.30 | * | 0.00 | | 0.00 | | 0.00 | |
| **u165** | 0.35 | * | 0.00 | | 0.33 | * | 0.00 | |
| **h165** | 0.49 | *** | 0.11 | | 0.46 | ** | 0.07 | |
| **u176** | 0.13 | | 0.00 | | 0.38 | * | 0.78 | ** |
| **m176** | 0.13 | | 0.00 | | 0.38 | * | 0.78 | ** |
| **m199** | 0.08 | | 0.62 | ** | 0.58 | ** | 0.45 | * |
| **u213** | 0.00 | | 0.34 | | 0.44 | * | 0.53 | ** |
| **m213** | 0.00 | | 0.24 | | 0.58 | ** | 0.67 | *** |
| **m224** | 0.18 | | 0.44 | * | 0.06 | | 0.07 | |
| **m226** | 0.57 | ** | 0.09 | | 0.00 | | 0.29 | |
| **u226** | 0.57 | ** | 0.44 | * | 0.07 | | 0.48 | ** |
| **u240** | 0.33 | * | 0.21 | | 0.00 | | 0.05 | |
| **u243** | 0.00 | | 0.00 | | 0.46 | ** | 0.07 | |
| **m243** | 0.00 | | 0.00 | | 0.46 | ** | 0.07 | |
| **m244** | 0.28 | * | 0.22 | | 0.06 | | 0.00 | |
| **h244** | 0.46 | ** | 0.13 | | 0.38 | * | 0.17 | |
| **u245** | 0.46 | ** | 0.20 | | 0.00 | | 0.16 | |
| **h245** | 0.57 | ** | 0.24 | | 0.49 | ** | 0.48 | ** |
| **u259** | 0.57 | ** | 0.00 | | 0.44 | * | 0.00 | |
| **h260** | 0.30 | * | 0.00 | | 0.19 | | 0.04 | |
| **h263** | 0.34 | * | 0.22 | | 0.27 | | 0.25 | |

| | F2/L1 | | F2/L2 | | F2/S1 | | F2/S2 | |
|---|---|---|---|---|---|---|---|---|
| **u263** | 0.37 | * | 0.49 | * | 0.56 | ** | 0.06 | |
| **u266** | 0.13 | | 0.44 | * | 0.07 | | 0.00 | |
| **u275** | 0.13 | | 0.00 | | 0.38 | * | 0.15 | |
| **h275** | 0.28 | | 0.21 | | 0.55 | ** | 0.30 | * |
| **u278** | 0.25 | * | 0.37 | | 0.44 | * | 0.58 | ** |
| **m278** | 0.49 | ** | 0.48 | * | 0.68 | ** | 0.80 | *** |
| **h286** | 0.00 | | 0.07 | | 0.10 | | 0.40 | * |
| **u312** | 0.22 | * | 0.00 | | 0.33 | * | 0.53 | * |
| **u318** | 0.29 | * | 0.37 | | 0.29 | | 0.00 | |
| **m318** | 0.35 | * | 0.89 | *** | 0.48 | * | 0.06 | |
| **u326** | 0.42 | ** | 0.21 | | 0.44 | * | 0.30 | |
| **h326** | 0.57 | ** | 0.44 | * | 0.49 | * | 0.48 | * |
| **m336** | 0.29 | * | 0.20 | | 0.70 | ** | 0.69 | ** |
| **m337** | 0.21 | | 0.00 | | 0.49 | * | 0.48 | ** |
| **h342** | 0.34 | * | 0.22 | | 0.08 | | 0.25 | |
| **m345** | 0.14 | | 0.89 | *** | 0.38 | * | 0.43 | * |
| **u345** | 0.29 | * | 0.22 | | 0.02 | | 0.00 | |
| **u351** | 0.38 | * | 0.22 | | 0.13 | | - | |
| **m353** | 0.22 | | 0.49 | * | 0.20 | | 0.27 | |
| **u353** | 0.22 | | 0.62 | ** | 0.33 | * | 0.41 | * |
| **u356** | 0.48 | ** | 0.13 | | 0.48 | ** | 0.06 | |
| **u359** | 0.35 | * | 0.21 | | 0.23 | | 0.00 | |
| **h363** | 0.43 | * | 0.09 | | 0.17 | | 0.29 | |
| **u363** | 0.75 | *** | 0.48 | * | 0.55 | ** | 0.53 | * |
| **m368** | 0.57 | ** | 0.24 | | 0.00 | | 0.04 | |
| **u368** | 0.67 | *** | 0.37 | | 0.08 | | 0.15 | |
| **h372** | 0.00 | | 0.44 | * | 0.06 | | 0.07 | |
| **u372** | 0.86 | *** | 0.44 | * | 0.46 | ** | 0.33 | * |
| **u384** | 0.10 | | 0.00 | | 0.30 | * | 0.06 | |
| **h384** | 0.18 | | 0.44 | * | 0.00 | | 0.07 | |
| **m386** | 0.30 | * | 0.33 | | 0.38 | * | 0.15 | |
| **u386** | 0.54 | ** | 0.56 | * | 0.44 | * | 0.04 | |
| **u395** | 0.34 | * | 0.00 | | 0.27 | | 0.00 | |
| **h395** | 0.34 | * | 0.00 | | 0.27 | | 0.25 | |
| **u410** | 0.30 | * | 0.13 | | 0.38 | * | 0.17 | |

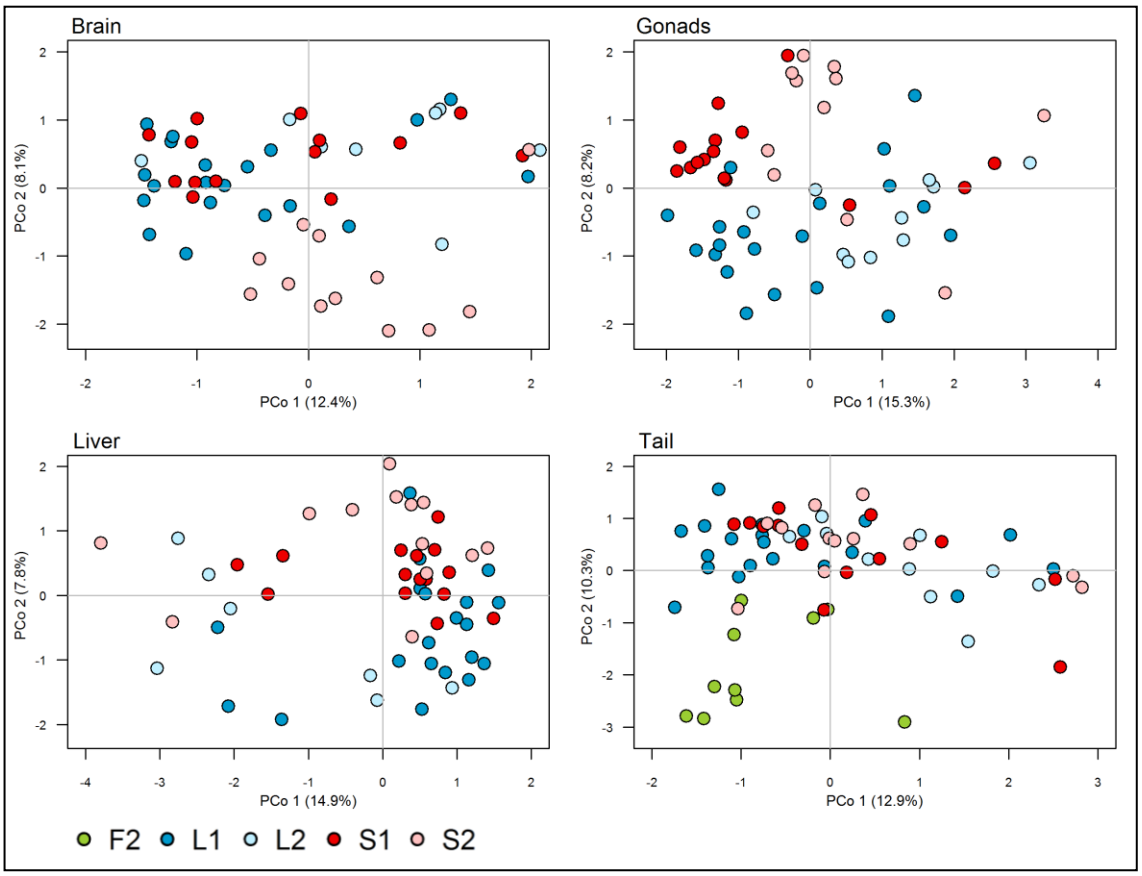



**Figure 4.6:** PCoA showing structure of selection lines.

Differentiation between the $F_6$ and the $F_2$ tail tissue samples showed a larger number of epiloci which were significantly differentiated (Table 4.13). Overall, 14 epiloci had a significantly elevated Φpt in 3 of the 4 pairwise comparisons, while an additional 6 epiloci had a significantly elevated Φpt in all 4 comparisons.

4.4.2.4 Principal coordinates analysis

PCoA was used to further examine the extent of epigenetic differentiation between selection lines and, for the tail tissue samples, the $F_2$ (Figure 4.6). In analysis of the brain tissue, the S2 samples clustered separately from the other three lines. However, there was no separation between the L1, L2 and S1. Although there is some overlap, the PCoA of samples from the gonads shows separation between the large selected lines and the small selected lines. A similar pattern of structuring can be seen in the analysis of samples from the liver, though the overlap is greater in the latter tissue. In tail tissue, none of the $F_6$ selection lines clustered separately, though $F_2$ samples did cluster separately from $F_6$ fish.

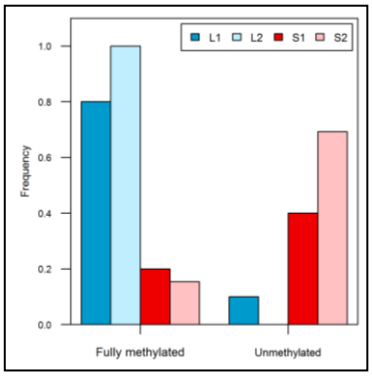



**Figure 4.7:** Frequency of methylation at MSL 312 in samples from the liver tissue.

**Table 4.14**: Outliers identified by Mcheza. Only those epiloci which were identified as an outlier in at least one pairwise comparison are shown. Grey boxes mark outliers with a $p \leq 0.05$

| | Brain | | | | Gonads | | | | Liver | | | | Tail | | | | Tail | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | F2 vs. L1 | F2 vs. L2 | F2 vs. S1 | F2 vs S2 |
| m-102 | ■ | ■ | | | | | | | | | | | | | | | | | | |
| u-102 | | ■ | | | | | | | | | | | | | | | | | | |
| m-104 | | | | | | | | | | | ■ | | | | | ■ | | | | ■ |
| u-104 | | | | | | | | | | | ■ | | | | | ■ | | | | |
| h-107 | | | | | ■ | | ■ | | | | | | | | | | | ■ | | |
| u-107 | | | | | ■ | | | | ■ | | ■ | | | | | | | | | |
| h-119 | | | | ■ | | | | | | | | | | | | | | | | |
| u-119 | | | | ■ | | | | | | | | | | | | | | | | |
| m-136 | ■ | | | | | | | | | | | | | | | | | | | |
| u-140 | | | ■ | | | | | | | | | | | | | | | | | |
| h-149 | | | | | | | ■ | | | | | | | | | | | | | |
| h-150 | | | | | | | | | | | | | | | | | ■ | | ■ | |
| u-150 | | | | | | | | | | | | | | | | | | | | ■ |
| u-154 | | | | | | | | | ■ | | | | | | | | | | | |
| m-161 | | | | | | | | | | ■ | | | | | | | | | | |
| u-165 | ■ | | | | | | | | | | | | | | | | | | | |
| h-166 | | | | | | | ■ | | | | | ■ | | | | | | | | |
| u-168 | | | | | | | | | | ■ | | | | | | | | | | |
| m-176 | | | | | | ■ | | | | ■ | | | | | | ■ | | | | |
| u-176 | | | | | | | | | | ■ | | | | | | ■ | | | | |
| h-180 | | | | | | | | | | | ■ | ■ | | | | | | | | |
| u-199 | | | | | | | | | | ■ | | | | | | | | | | |
| m-213 | | | | | | | | | | | | ■ | | | | | | | | |
| m-224 | | | | | | | | | | | ■ | | | | | | | | | |
| u-224 | | | ■ | | | | | ■ | | | | | ■ | | | | | | | |
| m-226 | ■ | | ■ | | | | | | | | | | | | | | | | | |
| u-226 | ■ | | | | | | | | | | | | | | | | ■ | | | |
| h-244 | | | | | | | | | | | ■ | | | | | | ■ | | | |
| m-244 | | | | ■ | | | | | | | | | | | | | | | | |
| h-245 | | | | | | | | | | | | | | | | | ■ | | | |

| | Brain | | | | Gonads | | | | Liver | | | | Tail | | | | Tail | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | L1 vs. S1 | L1 vs. S2 | L2 vs. S1 | L2 vs S2 | F2 vs. L1 | F2 vs. L2 | F2 vs. S1 | F2 vs S2 |
| u-245 | | | | | | | | | | | | | ■ | | | | | | | |
| h-263 | | | | | | | | | | | | | | | | | ■ | | | |
| h-275 | | ■ | | ■ | | | | | | | | | | | | | | | | |
| m-278 | | | | | | | | | ■ | | | | | | | | | | | ■ |
| u-278 | | | | | | | | | ■ | | | | | | | | | | | |
| m-280 | | | | | | | | | | | | | ■ | | | | | | | |
| m-312 | | ■ | | ■ | | | ■ | | ■ | ■ | ■ | ■ | | | | | | | | |
| u-312 | | ■ | | ■ | | | ■ | | | | | ■ | | | | | | | | |
| m-318 | | | ■ | | | ■ | | | | | | ■ | | | | ■ | | | | |
| u-318 | | | ■ | ■ | | | | | | | | | | | | | | | | |
| h-326 | | | | | | | | | | | | | | | | | ■ | | | |
| u-326 | | | ■ | | | | | | | | | | | | | | | | | |
| m-336 | | | | | | | | | | | | | | | | | | | ■ | ■ |
| m-337 | | | | | ■ | | | | | | | | | | | | | | ■ | |
| m-341 | | | | | | | | ■ | | | | | | | | | | | | |
| h-342 | | | | | | | | | | | | | | | | | ■ | | | |
| m-351 | | | | | | ■ | | | | | | | | | | | | | | |
| m-356 | ■ | | | | | | | | | | ■ | | | ■ | | | | | | |
| u-356 | ■ | | | | ■ | | | | | | ■ | | | | | | | | | |
| u-359 | | ■ | | | | | | | | | | | | | | | | | | |
| m-368 | | | | | | | | | ■ | | | | ■ | | | | ■ | | | |
| u-368 | | | | | | | | | ■ | | | | ■ | | | | ■ | | | |
| h-372 | | | | | | | | | | | ■ | ■ | | | | | | | | |
| u-372 | | | | | | | | | | | | | | | | | ■ | | | |
| m-384 | | | ■ | | | | | | ■ | | | | | | | | | | | |
| u-384 | | | | | | | ■ | | | | | | | | | | | | | |
| m-386 | | | | | | ■ | | | | | | | | ■ | | ■ | | | | |
| u-386 | | | | | ■ | ■ | | | | | | | | ■ | | ■ | | | | |
| h-395 | | | | | | | | | | | | | | | | | ■ | | | |

4.4.2.5 Epigenetic outlier analysis

For each of tissues analysed, the two large replicates were compared against each of the small replicates in both BayeScan and the Mcheza (4 comparisons per tissue type). In addition, for the samples from the tail tissue, each of the four selection lines was compared to the $F_2$. BayeScan did not identify any outlying epi-loci across any of the comparisons. Mcheza identified 59 epiloci that were outliers in at least one of the comparisons that represented 40 MSL (Table 4.14). However, only one epilocus (m312) was identified as an outlier all of the 4 comparisons for each tissue. As can be seen in Figure 4.7, the frequency of full methylation in liver at locus 312 was significantly higher in the large line ($X^2$= 24.932, $p \leq 0.001$), while the frequency of no methylation was significantly higher in the small line ($X^2$= 13.717, $p \leq 0.001$).

## 4.5 Discussion

This chapter examined the contribution of DNA methylation to the observed phenotypic variation between the large and small guppy selection lines and assessed the effect of imposing selection on the level of DNA methylation. MSAP was used to assess DNA methylation in four tissue types from fish in small and large-selected lines of *P. reticulata* (van Wijk *et al.* 2013), and in fish from before and after selection was imposed. The results indicate that DNA methylation in the guppy is tissue-specific with the pattern of methylation observed in the brain being significantly different from the pattern seen in samples taken from liver, tail and gonads. No significant differences in the level of methylation were observed between the large and small selection lines, suggesting that methylation at these loci did not contribute to the phenotypic shifts in the life history traits observed. The fragments examined here represent less than 0.34% of the potential EcoRI-HpaII/MspI fragments, and a considerably smaller proportion of the CG sites in the guppy genome. It is therefore not possible to be certain that the overall genome wide level of methylation is not associated with the selection regimes, using the available data. Only one of the epiloci analysed (m312) showed variation in the level of methylation which was consistent between the selection lines.

### 4.5.1 Scoring methylation sensitive amplified polymorphisms

While MSAP has been widely used to examine variance in methylation there is still little consistency in how the resulting peaks are scored and analysed (Schulz *et al.* 2013). To date, five different scoring techniques have been described, which provide the user with different levels of information. The need to determine the appropriate scoring technique arises from the necessity of transforming the four state matrix produced by the fragment analysis into a binary data format which can be statistically analysed. The four states which can be obtained from the fragment analysis are: *Msp*I+/*Hpa*II+, *Msp*I+/*Hpa*II-, *Msp*I-/*Hpa*II+ and *Msp*I-/*Hpa*II- with each state resulting from a different methylation pattern at the locus in question (see Table 4.1). When both enzymes cannot cut, (*Msp*I-/*Hpa*II-), it is not possible to distinguish between sequence variation and some types of methylation and therefore all of the scoring techniques treat individuals with this banding pattern as uninformative.

Of the five scoring techniques previously used only two have been widely used for analysis of methylation (Salmon *et al.* 2008, Herrera and Bazaga 2010, Paun *et al.* 2010, Vergeer *et al.* 2012, Yan *et al.* 2014, Medrano *et al.* 2014, Schulz *et al.* 2014, Sun *et al.* 2014, Wenzel and Piertney 2014, Avramidou *et al.* 2015, Nicotra *et al.* 2015). The first only distinguishes between methylated and non-methylated fragments by scoring *Msp*I+/*Hpa*II- and *Msp*I-/*Hpa*II+ fragments as present ('1'), *Msp*I+/*Hpa*II+ fragments as absent ('0') and *Msp*I-/*Hpa*II- fragments as missing data ('N/A'). The limitation of this technique, however, is that the methylation pattern which produces *Msp*I+/*Hpa*II- fragments is expected to be different from the pattern which produces *Msp*I-/*Hpa*II+ fragments. By grouping *Msp*I+/*Hpa*II- and *Msp*I-/*Hpa*II+ together, it is possible that variation in the pattern of methylation between populations could be missed (Lira-Medeiros *et al.* 2010; Schulz *et al.* 2013). The second scoring technique, therefore, creates up to three epiloci each of which is then scored according to the presence or absence of the methylation type for that epilocus (Schulz *et al.* 2013, 2014, Medrano *et al.* 2014, Sun *et al.* 2014, Avramidou *et al.* 2015). When using such an approach, an unmethylated epilocus would be scored according to the presence or absence of an *Msp*I+/*Hpa*II+ fragment (the presence of which would indicate an unmethylated site). A methylated epilocus would be scored according to the presence or absence of an *Msp*I+/*Hpa*II-

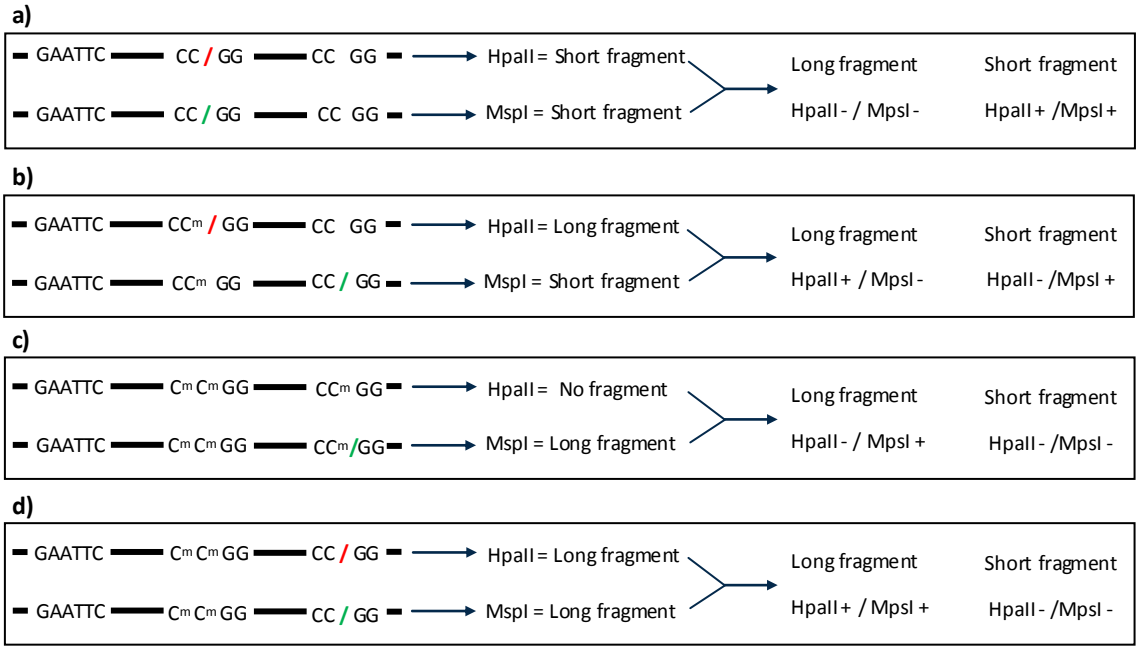



**Figure 4.8:** Figure showing how two CCGG cut sites can produce all four possible banding patterns when cut with HpaII and MspI. $C^m$ denotes a methylated cytosine, / denotes where the MspI enzyme would cut, / denotes where the HpaII enzyme would cut and GAATTC is the EcoRI cut site. a) If there is no methylation at either cut site digestion with the two enzymes will cut at the first site producing two small fragments. If studying the long fragment only, a banding pattern of HpaII - / MspI - will be produced b) If the internal C in the first cut site is methylated, the HpaII digest will cut at this site producing a short fragment while the MspI digest will not cut at the first site but will cut the second site. The banding pattern at the long fragment will therefore be HpaII + / MspI -. c) If the first site is fully methylated while the second has methylation at the internal C the resulting digests will produce a banding pattern at the long fragment of HpaII - / MspI +. d) Finally if the first site is fully methylated and the second site has no methylation a long fragment will be produced by both digests giving a long fragment banding pattern of HpaII + / MspI +.

fragment (the presence of which would indicate a methylated site). A hemi-methylated locus would be scored according to the presence or absence of an *Msp*I-/*Hpa*II+ fragment (the presence of which is thought to indicate a hemi-methylation of the external C) (figure 4.8). As the remaining three scoring techniques either only consider genetic variation (Lira-Medeiros *et al.* 2010) or have not been widely used they will not be discussed here, although a detailed review can be found in Schulz *et al.* (2013).

In our examination of methylation patterns in the guppy, we used the second technique to score the bands produced by the MSAP analysis. The third epilocus category produced with this technique has been purported to describe the presence or absence of hemi-methylation at the external C. As the cut site of the two isoschizomers is 5'-CCGG-3' methylation of the external C would represent a case of CpHpG methylation. According to the results of our analyses if the hemi-methylated banding pattern only represents methylation at the external C, then CpCpG methylation occurs at a frequency of between 7.2% and 10.4% in the guppy genome (Figure 4.2). Such an estimate is significantly higher than previous estimates of non-CpG methylation in vertebrates, which yields a frequency of CpHpG methylation of 1.22% in fish and 0.17-0.30% in other vertebrates (Feng *et al.* 2010). Although the frequency of apparently hemi-methylated fragments in our analysis is similar to that found in other studies which utilise the MSAP (Yang, Zhang, *et al.* 2011, Sun *et al.* 2014, Wenzel and Piertney 2014), to our knowledge no study using a different technique to study methylation has found such high levels of non-CpG methylation in vertebrates. The ambiguity in the level of non-CpG being reported with MSAP brings into question how else the *Msp*I-/*Hpa*II+ banding pattern could be produced.

While it is possible that the lack of an *Msp*I band at these sites is caused by ineffective cutting, the high level of repeatability observed here and in other studies indicates that cutting error is unlikely. The most likely explanation is the presence of a second 5'-CCGG-3' cut site closer to the *Eco*RI cut site which has varying levels of DNA methylation (Fulneček and Kovařík 2014). If two CCGG sites were present, figure 4.9, b illustrates how such an *Hpa*II+/*Msp*I- banding pattern could be generated from the methylation of an internal C.

It should be noted that, if methylation of an internal C could generate a *Hpa*II+/*Msp*I- banding pattern, where internal cytosine methylation did result in a *Hpa*II+/*Msp*I- banding pattern (as in Figure 4.9, b), a second shorter fragment with the *Hpa*II-/*Msp*I+ would also be produced. Analysis

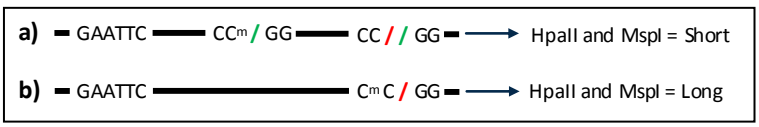



**Figure 4.9:** Fragments produced from a combined HpaII and MspI digest. A fragment with two cut sites close together would produce only a short fragment, while a fragment with only one hemi-methylated site would only produce a long fragment.

of the short fragment would appear to falsely indicate that it resulted from methylation of the internal cytosine at the second cut site. While only considering either the short or long fragment resulting from the sequence, it would be appropriate to simply accept all four possible banding patterns as four different epiloci, as they would result from varying levels of methylation at the two cut sites. However, if all CCGG sites in the genome were to be considered, it is possible that both the long and short fragment would be considered as separate epiloci, leading to incorrect interpretation of the methylation pattern, or potentially one $CC^{m}GG$ being analysed twice.

Fulneček & Kovařík (2014) suggested that running an additional digestion with both enzymes could be one way of determining whether a *Hpa*II+/*Msp*I- banding pattern is the result of $C^{m}CGG$ methylation or $CC^{m}GG$ methylation. Using this technique, a sequence with two CCGG sites where the first one is internally methylated would produce only a short fragment, while a site with only one CCGG site where the external C is methylated would produce only a long fragment (Figure 4.9).

Most MSAP analyses utilise multiple selective PCR primers in order to reduce the number of fragments analysed and in some cases the results from these primers are combined for analysis. Our results are based only on data from one pair of selective primers which considerably reduces the likelihood of having analysed both the short and long fragment from a pair of cut sites. It is, however, more likely that the high number of *Hpa*II+/*Msp*I- fragments seen in our analysis result from $CC^{m}GG$ methylation which has a second cut site nearby than from a large number of sites with a $C^{m}CGG$ pattern of methylation.

When considering the biological function of the methylated sites it is important to question whether it is appropriate to separate the *Hpa*II+/*Msp*I- and *Hpa*II-/*Msp*I+ banding patterns if both result from $CC^{m}GG$ methylation. However, it should be noted that although the *Hpa*II+/*Msp*I- does result from $CC^{m}GG$ methylation, these bands must also contain one unmethylated CCGG site. Therefore, when examining the global pattern of methylation, it would not be appropriate to separate out the two banding patterns but instead to add one unmethylated and one methylated site for each *Hpa*II+/*Msp*I- banding pattern. However, when analysing loci individually (e.g. for outlier analysis), it would not be appropriate to combine *Hpa*II+/*Msp*I- and *Hpa*II-/*Msp*I+ banding patterns as, although both result from $CC^{m}GG$ methylation, the cut site producing the two different patterns must be different. In our study we analysed the global methylation frequency by considering the *Hpa*II+/*Msp*I- banding pattern both separate from the *Hpa*II-/*Msp*I+ and combined with the methylated and unmethylated sites. By doing this we were able to show that there was no bias in considering the *Hpa*II-/*Msp*I+ separately in our data.

Despite the fact that it is more likely that *Hpa*II+/*Msp*I- represent internally $CC^{m}GG$ than hemi-methylated $C^{m}CGG$, it is not possible to confirm this from the available data. Therefore, for the remainder of the discussion, sites with a *Hpa*II+/*Msp*I- will still be referred to as hemi-methylated, while sites with a *Hpa*II-/*Msp*I+ will be referred to as fully methylated.

### 4.5.2 Tissue-specific DNA methylation

The results of our study identified tissue-specific DNA methylation in three of the four selection lines. When the data from all four lines was analysed together, significant variation in the level of DNA methylation was observed. Analysis of the epigenetic structure between the tissues revealed significant variation between the brain and the other three tissues studied.

Variation in the pattern of DNA methylation has been identified in many other species (Yang, Zhang, *et al.* 2011, Rodríguez López *et al.* 2012, Sun *et al.* 2014, Covelo-Soto *et al.* 2015), though not in many fish. Where variation has been investigated, eye, brain, liver and muscle tissues exhibited distinct patterns of DNA methylation in salmon (Rodríguez López *et al.* 2012), whereas in the closely related brown trout, liver, heart and kidneys showed very similar methylation patterns, while brain and muscle tissues yielded distinct patterns (Covelo-Soto *et al.* 2015). The sensitivity of our analysis to detect tissue-specific methylation patterns in *P. reticulata* provides evidence of the utility of the MSAP technique as a method of studying patterns of methylation in selection lines. The results of our study and those elsewhere also highlight the need to choose the tissue used to study DNA methylation with care and ideally examine more than one tissue type wherever possible.

### 4.5.2 Patterns of DNA methylation between the selection lines

When considering the impact of genome wide methylation on body size in our selection lines, based on the frequency of methylation discussed above, there are only two cases where methylation could be associated with body size. The frequency of hemi-methylation was significantly higher in the gonad tissue of the large lines, while the frequency of full methylation in the tail tissue was significantly higher in the small lines. The frequency of no methylation also varied significantly between the large and small lines, values for the L2 and S1 were the same, therefore it is unlikely that the observed variation is linked to body size. Despite the observed frequency differences in the frequency of hemi-methylation and full methylation, analysis of the Φpt values and the PCoA suggest that there is no global epigenetic structure present between the selection lines. The results of the analysis of DNA methylation in our selection experiment therefore support the conclusion that variation in the levels of genome wide DNA methylation does not play a significant role in the variation in life history traits observed between the selection lines studied here.

The majority of genetic change observed between the selection lines has been mapped to the sex chromosome (Fraser *et al.* 2014). However, as discussed in Section 2.5.3, a large portion of the Y chromosome in guppies is pseudoautosomal and recombines with the X chromosome. In addition, both our study and previous work (van Wijk 2011a) identified candidate SNPs located on the autosomal chromosomes. Therefore, the lack of observed variation in life history traits in the female fish would suggest that regulatory changes are playing a role in the variation observed. Indeed, the high level of plasticity observed in body size and other life history traits (Nylin and Gotthard 1998,

Wiens *et al.* 2014, Baker *et al.* 2015) makes it extremely likely that regulatory mechanisms do play a role in determining traits such as body size, at least within the context of selection and genomic backgrounds examined here. These regulatory mechanisms will be underpinned by both genetic changes to regulatory regions of genes and epigenetic changes, such as DNA methylation.

Although recent studies in which the level of DNA methylation has been experimentally manipulated have indicated that the global DNA methylation can influence quantitative traits like body size (Alvarado *et al.* 2015), it is more likely that locus-specific and not genome wide methylation is responsible for the regulatory role DNA methylation has on body size. The results of our study suggest that the level of methylation at locus 312 is playing a role in the variation observed between the selection lines. Further analysis of this locus would determine whether it was linked to a candidate gene which in turn is linked to a life history trait such as body size. It is also notable that the variation in methylation observed at this locus was only present in the liver, which again highlights the necessity of analysing a range of tissues when studying DNA methylation.

Genetic change in both coding and regulatory regions can and does control body size in most species, however given the complexity of the trait and the number of factors which can influence it, epigenetic change should also be expected to play a part. Several studies have now shown that DNA methylation can affect body size (Wolff *et al.* 1998, Zhong *et al.* 2014, Alvarado *et al.* 2015, Cao *et al.* 2015). It is therefore extremely likely that DNA methylation can and does control body size and other life history traits in vertebrates, and that DNA methylation is contributing to the observed shifts in life history traits observed in our study. However, the genetic variation observed between the selection lines appears to be much greater than the epigenetic variation and therefore, the observed shifts in life history traits over the course of our selection experiment are thought to be primarily underpinned by genetic change.

### 4.5.3 Changes in DNA methylation patterns after selection

We examined changes in methylation which occurred over the course of the experimental selection by comparing the pattern of DNA methylation in tail tissue from the $F_2$ (before selection generation) and $F_6$ generations, that is, the four selection lines following selection. Over the course of the experimental selection, the level of fully methylated sites increased in all four selected lines (Figure 4.5). In addition, a PCoA of all tail tissue samples shows separation between the $F_2$ and the $F_6$ fish. As shown by earlier analysis have shown (see Section 2.4), significant genetic differences exist between the $F_2$ fish and the $F_6$ fish in all four selection lines. It could, therefore, be possible that the results of the segregation in the PCoA are caused by variation between the $F_2$ and the $F_6$ as a direct result of these genetic differences. However, the frequency of uninformative sites (sites which produce no band and could be the result of a fully methylated site or a sequence change) does not differ significantly between the $F_2$ and the $F_6$ fish and genetic variation would not explain the increase in the frequency of fully methylated sites observed.

There are two possible explanations for the increase in frequency of methylation between the $F_2$ and the $F_6$ fish. The first considers the stress the experimental selection may have caused during the experiment and the potential for DNA methylation to be underpinning adaptation to captivity. Environmental stress has been shown to induce changes in DNA methylation (Labra *et al.* 2002, Steward *et al.* 2002, Sollars *et al.* 2003, Anway *et al.* 2005, Meaney and Szyf 2005, Brown *et al.* 2009, Boyko *et al.* 2010, Herrera and Bazaga 2010, Verhoeven *et al.* 2010, Vandegehuchte and Janssen 2011). The stress of handling to fish, and perhaps changes in the water chemistry (which would have had an equal effect on all selection lines), may have triggered an increase in the level of methylation. Adaptation to captivity is a widely recognised phenomenon which will have taken place in our experiment. The three generations of random breeding and controlled environment utilised throughout the selection experiment mean that adaptation to captivity should not have played a role in the variation in phenotypic traits observed between the $F_6$ selection lines. However the variation in methylation observed between the $F_2$ and the $F_6$ fish may be the result of ongoing adaptation to captivity which is underpinned by DNA methylation rather than genetic change.

It is equally plausible that the observed variation arises from sampling variance. Due to the timing of the experiment, the samples taken from the $F_2$ fish were taken very shortly after the fish had reached maturation, while samples were taken from the $F_6$ fish when they were between 1 and 2 years old. The evidence that DNA methylation patterns change with age is undeniable (reviewed in Jung & Pfeifer 2015). Indeed, in humans, methylation patterns at specific loci can be used as predictors of age (Bekaert *et al.* 2015, Zbieć-Piekarska *et al.* 2015). Although age-related methylation studies in fish have primarily focused on methylation patterns during early development, (Mhanni and McGowan 2004, MacKay *et al.* 2007, Rai *et al.* 2007, Wu *et al.* 2011), it would not be surprising to find variation in the patterns of methylation that younger and older fish exhibit. Further studies of the changes in the pattern of methylation with age in guppies would help to clarify whether the observed changes between the $F_2$ and the $F_6$ generations are a result of stress or variation in the age of the fish.

The selection experiment also included a control line which was not utilised in the analysis of this thesis. The control line was subjected to selection at the same time as the large and small lines, but the selection was random. The analysis undertaken by van Wijk (2011) showed that, while the body size of fish in the large selected line increased and in the small selected line decreased, the body size in the control line did not change. Therefore by analysing DNA methylation in the fish from the control selection line it would have been possible to further explore the potential effect adaptation to captivity has had on the methylation patterns. However, as the fish from the control line were sampled at the same age as the fish from the selection lines, it is unlikely that it would have been possible to distinguish between the effects of sampling variance and adaptation to captivity.

### 4.5.4 DNA methylation and fisheries-induced evolution

The results of both our study and those of van Wijk (2011) show that size selective harvesting can induce genetic change in genes underpinning life history traits such as body size and maturation. The analysis of DNA methylation patterns following experimental size selection show that the effects of genetic change in our experiment are greater and more consistent than the epigenetic change examined here. However, there is potential for selection to influence epigenetic modifications in wild populations (Angers *et al.* 2010, Verhoeven *et al.* 2010, Massicotte *et al.* 2011), and in turn, for these epigenetic modifications to be playing a role in the shifts in life history traits we are seeing in exploited populations of fish.

The term 'fisheries-induced evolution' implies that size selective harvesting imposed by fisheries acts as a strong selection pressure and drives the evolution of smaller body size and an earlier maturation age (amongst other traits). At present this definition focuses on the evolutionary effect harvesting can have on the genetics of exploited populations. The rise in awareness of the potential role of epigenetics has led to consideration of the possibility of selective pressures driving evolution on epigenetic variation as well as genetic variation (Angers *et al.* 2010, Verhoeven *et al.* 2010, Massicotte *et al.* 2011). For selection to act on variation in patterns of DNA methylation, there are three key conditions which must be fulfilled: (1) methylation marks must exhibit transgenerational inheritance, (2) variation in levels of DNA methylation which cause or contribute to phenotypic variation must be observed, and (3) there must be natural variation in the patterns of DNA methylation in wild populations. Whilst there are several studies that provide evidence of changes in the level of DNA methylation being passed from parent to offspring or even into the $F_3$ generation (Molinier *et al.* 2006, Vaughn *et al.* 2007, Johannes *et al.* 2009, Whittle *et al.* 2009, Franklin and Mansuy 2010, Skinner 2010, Verhoeven *et al.* 2010), very few have assessed the inheritance of methylation changes over more generations (Grossniklaus *et al.* 2013). As already discussed, there is a significant body of evidence that DNA methylation can influence phenotypic traits and the rapidly expanding field of ecological epigenetics has provided several examples of naturally occurring epialleles (Bossdorf *et al.* 2007, Massicotte *et al.* 2011, Kilvitis *et al.* 2014). One example that appears to fulfil all of the criteria, is the epimutation in the colourless non-ripening (*Cnr*) tomato. Manning *et al.* (2006) showed that the *Cnr* epimutation inhibits normal ripening and causes colourless tomatoes. It also exhibits transgenerational inheritance and substantial natural variation. If epigenetic variation can form the base for natural selection, it follows that the selection pressure arising from harvesting could similarly select for epialleles, as well as allelic variation in coding genes. .

Another way that DNA methylation could influence shifts in life history traits in exploited populations would be through the increased stress experienced by these populations. Fishing is expected to cause stress directly through the process of catching the fish and the stress caused to escapees from fishing gear, but also indirectly through the detrimental effect certain types of fishing gear can have on the environment (Hiddink *et al.* 2006, Clark *et al.* 2015). Given the large body of evidence

that environmental stress can modify DNA methylation, there is potential for these stressors to be changing DNA methylation patterns.

Finally, even if natural selection does not act directly on epigenetic mutations to drive evolutionary change, changes in DNA methylation have been shown to lead to genetic change (Shen *et al.* 1994). Therefore, anything which results in a change in the level of methylation, for example stress caused by fishing, may also lead to increased genetic variation for natural selection to act on.

### 4.5.5 Concluding remarks

To summarise here we show the shifts in life history traits observed following the selection experiment on guppies are primarily underpinned by genome wide genetic change rather than epigenetic change. It is extremely unlikely that epigenetic modifications such as DNA methylation do not play a role in body size and timing at maturation, however these are likely to be locus or epi-locus specific. Finally the potential for DNA methylation to be contributing to rapid evolutionary change highlights the importance of considering the role of epigenetic changes in FIE.

# Chapter 5. General Discussion

## 5.1 Abstract

Having shown that size selective harvesting can be underpinned by genome wide changes, the following chapter examines the limitations and wider applications of the study before considering possible future research. The main limitations of the study are the design of the selection experiment, the ability of $F_{st}$ outlier analyses to identify the loci underpinning complex polygenic traits and the lack of transcriptomic analysis. Discussion of the wider implications of the study examines the contribution of the results to fisheries induced evolution, the wider impacts of size-selective predation and the contribution of epigenetics to traits such as body size. Future research in the field of fisheries induced evolution should include a focus on the recovery potential of harvested populations and potential management strategies. To be able to identify the loci underpinning body size and maturation in the guppy, it will be necessary to obtain a reference sequence for the Y-chromosome.

## 5.2 Overview of key findings

The results described herein focused on genome-wide effect of size-selective harvesting by using a reduced representation sequencing approach that identified 37 SNPs associated with shifts in life history traits. Furthermore, the recently sequenced guppy reference genome (Fraser *et al.* 2014) enabled SNP identification and functional analysis of those genomic regions under selection.

The majority of SNPs showing signs of selection (86%) were located on the sex chromosome (chromosome 12), though signs of selection were also identified on chromosomes 3, 11, 14, 17 and 23. The large number of SNPs showing signs of selection that were located on the sex chromosome confirms the sex-linked control of body size, however it is not possible to rule out the effect of reduced recombination falsely inflating the $F_{st}$. The genomic response to selection for large size appeared to be much stronger than the genomic response to small size. Although we were able to identify SNPs associated with shifts in life history traits, functional analysis of the genes these SNPs were located in revealed no specific metabolic pathway associated with body size or timing of maturation. However given the increased level of linkage that is expected on the sex chromosome it is likely that some of the candidate genes on this chromosome are under selection, and that the SNPs identified are in linkage with the causal variant.

Once SNP-associated genomic regions under selection had been identified, they were then screened in wild populations of guppies exhibiting similar patterns of life history variation associated with size-selective predation. Analysis of these SNPs was unable to identify consistent signs of selection in any of the wild populations sampled. There are thought to be two reasons why SNPs showing selection in the selection experiment did not show consistent signs of selection in the wild populations. Firstly, the linkage between SNPs in the wild populations is likely to be a lot lower than

in the fish sampled following the selection experiment. Secondly, given the polygenic nature of life history traits there will be a various combinations of alleles which could produce the same phenotype. Previous studies looking for signs of selection in wild populations of guppies found little evidence of parallel genetic evolution, despite obvious parallel phenotypic evolution (Fraser *et al.* 2014). Based on the hypothesis of little parallel genetic evolution between the wild populations, it follows that the genes underpinning the shifts in life history traits seen in our selection lines are not the same as the genes underpinning the variation in life history traits seen in the wild populations.

The aim of the final data chapter presented was to examine the role that regulatory changes may have played in the phenotypic shifts observed. Analysis of DNA methylation patterns in the $F_2$ (before selection) and $F_6$ (after selection) generations found no significant variation in the overall level of methylation at the 91 loci examined that was associated with the observed phenotypic shifts. At one locus the pattern of DNA methylation was significantly higher in the fish from the large selection regime, suggesting that while genome wide patterns of methylation have had little effect, locus-specific methylation is playing a role. The genome wides pattern of DNA methylation did change significantly between the $F_2$ and the $F_6$, although the magnitude of the changes was the same across all selection lines. Given the history of the experimental fish used in the van Wijk *et al.* (2013) study, the changes in methylation between the $F_2$ and the $F_6$ are likely to be the result of ongoing adaptation to captivity.

## 5.3 Limitations of the study

### 5.3.1 The selection experiment

The selection experiment discussed here was the first to provide evidence of genetic change associated with phenotypic change as a result of harvesting. There are, however, several factors which limit the wider applicability of the study. When considering the selection experiment itself there are two points which must be considered. The first of these is the population structure used in the study. The experimental set up used in both the current study and by Conover and Munch (2002) involved selection on discrete generations and species with short generation times, however such a setup is not representative of the true scenario seen in commercially harvested populations. However, although the experimental setup used is a simplified version which limits ecological feedback, it is expected that this would reduce the efficiency of selection and increase the timeframe required to see such a significant phenotypic and genetic response (Conover and Munch 2002, Conover and Baumann 2009b). Wild populations have been subjected to size selective harvesting for considerably longer. Therefore, although the magnitude of change will be reduced, it is unlikely that it would be completely absent. The phenotypic and genetic changes seen in wild guppy populations (with overlapping generations and ecological feedback), in response to changes in predation support such a conclusion (Reznick and Ghalambor 2005).

The second issue which leads to questions about the applicability of the study to commercially harvested fish populations, is the harvesting of only males and the potential Y-linkage of the life history traits being examined. The results from the initial experiment (van Wijk *et al.* 2013) and those discussed in the current thesis (although see section 2.5.3) suggest that a number of the loci under selection are located on the Y-chromosome. Harvesting in the wild is not limited to males and it is unlikely that such a level of sex linkage is present. As with the experimental setup, it is likely that the Y-linkage will have accelerated the effect of selection in the current study in comparison to wild populations.

### 5.3.2 Identifying signs of selection

Throughout the current project $F_{st}$ outlier analyses have been used to identify signs of selection however these techniques have been criticised. The main criticisms focus on the large numbers of false positives detected by the various methods primarily as a result of models which don't fit the true demographic history (Narum and Hess 2011, Vilas *et al.* 2012, Bierne *et al.* 2013, De Mita *et al.* 2013, Fourcade *et al.* 2013, Lotterhos and Whitlock 2014). The large discrepancy between the numbers of outliers detected by different techniques in the present study suggests that many of these were false positives. Although detection of false positives can be reduced by comparing the results of multiple methods (Pérez-Figueroa *et al.* 2010b) and studying several pairs of populations (Vilas *et al.* 2012) the results from these analyses should be treated with caution and investigated further wherever possible.

However of greater significance to the current study, and perhaps the future of the $F_{st}$ outlier approach, is that this approach is only likely to identify large effect loci (Pritchard and Di Rienzo 2010, Lotterhos and Whitlock 2014). These methodologies are unable to selection on small effect loci and particularly covariation between such loci. Quantitative traits (including body size and age at maturation) are expected to be polygenic with adaptation of these traits being underpinned by subtle differences in a large number of covarying alleles (McKay and Latta 2002, Hancock *et al.* 2010, Nadeau and Jiggins 2010, Pritchard and Di Rienzo 2010, Yang *et al.* 2010, Le Corre and Kremer 2012, Messer and Petrov 2013, Bourret *et al.* 2014).

Their importance and prevalence in nature has led to the development of several methods to try and identify the genetic underpinning of such complex polygenic traits. One approach, known as genomic partitioning, groups loci by proximity and examines how much of the phenotypic variance observed can be explained by each group of loci (Yang, Manolio, *et al.* 2011, Robinson *et al.* 2013, Santure *et al.* 2013). Although genomic partitioning can be used to show that a trait is underpinned by a large number of small effect loci and identify regions of the genome which require further investigation (Robinson *et al.* 2013, Santure *et al.* 2013), this approach can't identify covariation in loci which are not in close proximity or determine the causal variants. Bourret *et al.* (2014) used a similar approach to genomic partitioning. However, instead of using proximity, loci were partitioned using a principle components analysis, allowing the identification of covariation. Another method

used to identify covarying small effect loci is a random forest algorithm which uses a tree-based ensemble machine learning tool to examine the link between phenotype and multiple SNPs (Boulesteix *et al.* 2012, Stephan *et al.* 2015, Wellenreuther and Hansson 2016). Random forest has been used for a while to identify the loci underpinning polygenic traits used in medicine (Shi *et al.* 2005, Cordell 2009, Tang *et al.* 2009, Xu *et al.* 2011) but has only recently started being used in wild populations (Holliday *et al.* 2012, Brieuc *et al.* 2015, Hornoy *et al.* 2015, Laporte *et al.* 2015, 2016, Pavey *et al.* 2015).

Due to the polygenic nature, complex traits such as body size and age at maturation $F_{st}$ should no longer be considered the best approach to identifying the loci under selection. However, it is unlikely that there will be one method which can be applied to all scenarios and therefore, in most cases, a combination of approaches will be optimal (Sork *et al.* 2013, Bourret *et al.* 2014, Wenzel and Piertney 2014, Laporte *et al.* 2015).

### 5.3.3 Transcriptomic analysis

Even with the development of new techniques for detecting signs of selection, whether those loci are underpinning polygenic or mendelian traits, it will not be possible to say for certain what effect these loci are having on the trait in question using DNA analysis alone. In human disease genetics identification of pathogenic variants is carried out by using functional analysis which can include: analysis of gene expression; examination of sequence homology; predictions of protein features; biochemical studies and protein interactions (Rugarli 2006). Such a comprehensive level of analysis for all of the potential loci identified here would have been beyond the time frame and budget of this project. However, analysis of the transcriptome or the specific transcripts of the loci identified would have provided further evidence that these loci were contributing to the observed shifts in life history traits.

Comparative transcriptomics has been used to identify loci under selection in a number of different species (Güimil *et al.* 2005, Andersen *et al.* 2008, Chen *et al.* 2008, Whitehead *et al.* 2010, Baldo *et al.* 2011, Koenig *et al.* 2013). If the candidate loci have already been identified, quantitative reverse-transcription PCR (RT-PCR) can be used to compare the level of gene expression at these loci between populations, species or before and after treatment (Güimil *et al.* 2005, Lelandais *et al.* 2008). If a reference genome is available, it can be used to design a microarray to examine a large number of loci across the genome (Oleksiak *et al.* 2002, Gilad *et al.* 2006, Hoffmann and Willi 2008). As with DNA analysis, the advent of next-generation sequencing has, combined with RT-PCR, enabled the analysis of large numbers of loci in significant numbers of individuals where no *a priori* knowledge of the genome is available (Chen *et al.* 2008, Baldo *et al.* 2011). In our study, transcriptome analysis could have provided information about the variation in expression levels between the selection lines and about the changes in expression levels over the course of the experiment. Furthermore, as gene expression varies between tissues, transcriptome analysis of multiple tissues could have provided additional information on the tissue variation identified in the

DNA methylation analysis. Unfortunately, due to the timing and practical constraints of the experiment, the samples collected during the experiment were not appropriate for RNA extraction. The selection experiment was brought to an end by the outbreak of a disease (*Mycobacterium marinum*) which led to the death of a number of fish. Although the bodies of all fish were preserved, variation in the timing of the deaths and the preservation method used during this period (100% ETOH) rendered transcriptome analysis impossible.

## 5.4 Importance of the guppy sex chromosome

As mentioned previously, the majority of the SNPs showing signs of selection in the study presented here were located on the sex chromosome (chromosome 12). Previous work has identified QTLs linked to body size on the sex chromosome, though there remains much that is unknown about the evolution of this chromosome, particularly in the degree of linkage.

Early studies utilised linkage mapping to identify 3 distinct regions on the sex chromosome, two male-specific non-recombining sections (MSNR1 and MSNR2) separated by a freely recombining section (FR) (Tripathi *et al.* 2009a, b). These studies showed that while recombination across the whole chromosome (between the X and Y) is reduced compared to the autosomes, recombination takes place freely in the FR region and although very rare, is not completely absent in the MSNR1. MSNR2, thought to be located at the distal end of the chromosome, was identified as the diverged section of the chromosome where the sex loci and previously identified Y-linked colour loci (Winge and Ditlevsen 1947, Haskins *et al.* 1970) could be found Tripathi *et al.* (2009a).

Contrary to this, cytological studies had found regular pairing and association between the distal end of the X and Y, suggesting regular recombination in this region (Traut and Winking 2001, Nanda *et al.* 2014). This discrepancy was explained by Lisachov *et al.* (2015) who found a fourth freely recombining region located below the MSNR2 at the distal end of the Y chromosome which had not been identified previously due to a lack of markers. They further hypothesised that the reduced level of recombination found in the MSNR1 was a result of the suppressive effect of the centromere, rather than a lack of sequence homology. It is likely, however, that such a lack of recombination has led to some divergence between the X and Y chromosomes in this region.

The lack of conclusive information about the degree of recombination between the X and the Y chromosome makes it difficult to determine how far the linkage between the causal variant and the SNPs identified here extends. For example if, as suggested by previous studies (Nanda *et al.* 2014, Lisachov *et al.* 2015), the majority of the chromosome is pseudoautosomal and freely recombines between the X and the Y, then it is likely that the SNPs identified here are linked to several different causal variants across the chromosome. However, if the level of recombination across this chromosome is reduced, then the sections of the chromosome in linkage will be longer and the SNPs identified here linked to only a small number of causal variants.

The issue of how many causal variants are present on the sex chromosome is also complicated by the lack of knowledge about the reference sequence for this chromosome. As the reference genome was created from only one female fish, it is unlikely that it contains any Y-specific sequences. Furthermore the reference sequence from the sex chromosome is incomplete (Axel Künster, personal communication), and it is unknown which section/s of the chromosome (FR1/FR2 etc) sequences represent. In order to fully clarify how many variants are under selection, both in the experiment described here and in wild populations, it will be necessary obtain a full sequence of the Y chromosome and to obtain a full understanding of the level of recombination between it and the X chromosome.

## 5.5 Wider ecological impacts of size-selective predation

The evidence presented here and in previous studies (Conover and Munch 2002, van Wijk *et al.* 2013) has shown the direct impacts size selective harvesting can have on life history traits. It is important, however, to consider the effects that size selective predation can also have on the wider ecosystem. The natural variation in predation regime across wild populations of guppies provides an ideal model for examination of such ecological effects.

Initial studies of ecological change showed that the release from predation that occurs between lowland and upland sites in wild guppy populations (higher predation in lowland sites, lower predation in upland sites) leads to shifts in diet type and consumption rate (Palkovacs *et al.* 2009, 2011) and, in turn, some associated shifts in ecosystem structure (algal and invertebrate biomass) and function (primary productivity, decomposition rates and nutrient cycling) (Bassar *et al.* 2010). In sites with high levels of predation, guppies were found to primarily consume invertebrates rather than algae, which led to lower levels of leaf decomposition (Palkovacs and Post 2009, Bassar *et al.* 2010, 2013, 2015). Guppies in high predation sites also excreted higher levels of nutrients which when combined with the low level of algal consumption leads to higher levels of primary productivity (Bassar *et al.* 2010).

The results of an ongoing project (The Guppy Project: http://cnas.ucr.edu/guppy/) confirmed the influence of varying life history and diet on ecosystem processes (Travis, Reznick, and Bassar 2014, El-Sabaawi, Bassar, *et al.* 2015, El-Sabaawi, Marshall, *et al.* 2015) and showed that the effect of phenotype can as strong as the effect of environmental variables such as light (El-Sabaawi, Bassar, *et al.* 2015). Furthermore by introducing guppies to sites with varying biodiversity The Guppy Project has also been able to examine the impact the different life history traits have on other species (Walsh and Reznick 2011, Furness *et al.* 2012, Furness and Reznick 2014).

The impacts that trait change can have on the wider ecosystem can also be seen in the alewife (*Alosa pseudoharengus*) where populations which have become landlocked have evolved life

history traits similar to those seen in harvested fish populations and guppies from sites with high levels of predation (Palkovacs *et al.* 2008). Studies of the alewife have found that the landlocked populations have changed the zooplankton biomass and community structure in introduced lakes (Post *et al.* 2008, Palkovacs and Post 2009), as well as having a significant impact on the evolution of their prey, *Daphnia ambigua* (Walsh and Post 2011). Importantly these studies have also shown that the effect of different phenotypes on the ecosystem is greater than the effect of the presence or absence of the alewife (Palkovacs and Post 2009).

The effects of predation regime on the wider ecosystem discussed above could be transferred to harvested fish populations in two ways. Firstly, treating the harvested fish as the prey and fisheries as the predator, which is how FIE is traditionally viewed, requires consideration of contributions the harvested fish have on ecosystem processes such as those discussed above. Secondly, would be consideration of ecosystem wide impacts if the harvested fish are treated as the predator. In such a scenario, the harvesting of the predator or changes in feeding behaviour may lead to a release from predation for any species it preys on. In turn this release from predation might lead to changes such as those seen in the low predation guppy populations described above.

## 5.6 Application to fisheries induced evolution

The results from the van Wijk (2011) study provided evidence that shifts in life history traits resulting from size selective harvesting are associated with change at specific candidate loci. The results from the present study show that size selective harvesting can have a genome wide impact. The structure of the selection experiment and the Y-linked nature of the loci under selection (see section 2.5.3) mean that caution is required when making direct comparisons between results described here and commercially harvested populations. However, providing an unequivocal answer to the question of whether genetic change is underpinning the phenotypic shifts seen in commercially harvested fish populations would require a study using species with long generation times, populations with overlapping generations, and a large number of generations. Such a study would require an impractical budget and time frame. Furthermore, phenotypic change has already occurred in commercially harvested populations, therefore, waiting for unequivocal evidence that genetic change is underpinning the observed shifts before taking action will only increase the severity of the situation.

The current study provides evidence of genomic change in a simplified experimental design. In order to provide evidence of genetic change which is more directly applicable to commercially harvested populations, it will be necessary to determine the effects of size selective harvesting in a more complex experimental design. Such an experiment, in which populations of guppies have been selected for large or small body size but also with overlapping generations and subject to density feedback and natural selection, is ongoing (M. Henio, personal communication). However, these fish have yet to show evidence of significant phenotypic change. It will also be necessary to

provide evidence based on analysis of species with life history characteristics similar to those seen in commercially harvested populations. The guppy however, which was an ideal species for the experiments described herein, exhibits significant sexual dimorphism in the traits being examined (both body size and maturation) which are underpinned by Y-linked loci. As the majority of commercially harvested fish do not exhibit such dimorphism, a similar experiment on a more representative fish species would be more directly comparable.

Although experiments such as those discussed above would provide more conclusive evidence that phenotypic shifts in commercially harvested populations are underpinned by genomic change, the large body of evidence which supports the existence of FIE (Jorgensen *et al.* 2007, Kuparinen and Merilä 2007, Conover and Baumann 2009a, Dunlop *et al.* 2009, Sharpe and Hendry 2009) highlights the need for action to be taken now. Despite the evidence of change, management tools in use for some species still rely on data which is at least 30 years old (Hilborn *et al.* 2010, van Walraven *et al.* 2010, Heino *et al.* 2013). The primary strategies which have been suggested focus on modifying reference points used in fisheries managements (Hutchings 2009, Heino *et al.* 2013). Suggested modifications to the reference points include increasing the minimum size limit, introducing a maximum size limit (either in combination with, or instead of, a minimum size limit) and reducing the total catch (Conover and Munch 2002, Law 2007, Hutchings and Fraser 2008, Eikeset *et al.* 2013). It has also been suggested that marine protected areas, or no-take zones, might reduce the effect of fisheries induced evolution (Dunlop *et al.* 2009). However, despite the number of times these suggestions have been made, there is currently only limited evidence these would prevent fisheries induced evolution (Jorgensen *et al.* 2007, Dunlop *et al.* 2009, Hutchings 2009).

## 5.7 Recovery potential of harvested populations

Whether generated by natural predation or anthropogenic change, the ecosystem wide impacts of trait change are likely to be significant even if the trait changes are caused by phenotypic plasticity or regulatory change. It was not possible to examine the recovery potential of the fish in the selection experiment presented here due to the outbreak of disease (*Mycobacterium marinum*). Although not all of the fish were lost during the outbreak, the numbers where significantly reduced, leading to a population bottleneck and therefore making it unfeasible to examine the genetic impact of a halt in harvesting. As the infection did not occur until the $F_6$ fish were at least 1 year old and the fish from all of the selection lines were all equally affected, it is very unlikely to have had impact on the results discussed here. While we were unable to directly examine the recovery potential of our selected lines, the importance of fully understanding the genome wide changes which are underpinning these trait changes lies in our ability to understand their reversibility.

The potential for traits affected by FIE to return to their pre-harvest values has been examined several times by utilising eco-genetic models (Dunlop *et al.* 2009, Enberg *et al.* 2009, Kuparinen

and Hutchings 2012, Marty *et al.* 2015, Uusi-Heikkilä *et al.* 2015). These models suggested that following a fishing moratorium the recovery would be slow, though as many of these studies were based on probabilistic maturation norms (PMRN), it is likely they have underestimated the complexity found in real populations. The only study to have experimentally examined the potential for recovery (Conover *et al.* 2009) showed that phenotypic changes did begin to recover once size selective harvesting was stopped. As discussed in 2.4.5, the recovery observed by Conover *et al.* (2009) may be the result of the small number of generations selection was imposed over.

The suggested reasons for the slow recovery rate have primarily focused on the strength of selection (Dunlop *et al.* 2009, Enberg *et al.* 2009, Kuparinen and Hutchings 2012). The selection pressure imposed by fishing is has been shown to be significantly higher than the pressure imposed by natural selection (Mertz and Myers 1998). Therefore fishing has the ability to drive very rapid shifts in phenotypic traits, but once a moratorium on fishing is imposed, because the strength of natural selection is presumably smaller, it will take a correspondingly long time for phenotypic traits to recover to pre-harvest levels.

In addition to putatively reduced selection pressures following cessation of harvesting, it has also been suggested that the loss of genetic variability caused by fishing will reduce the adaptability of exploited populations (Marty *et al.* 2015). Put simply, the rarity of new beneficial mutations will mean that an exploited population's ability to evolve to its pre-harvest phenotype will be determined by the level of standing genetic variation. If the standing variation has been eroded or changed in relation to key ecologically significant traits by fishing, then the population's ability to recover will be limited. Our results indicate that the life history shifts observed in FIE are underpinned by changes at a large number of small effect loci. The effect of rapid evolution of on a large number of small effect loci is likely to result in a much greater loss of standing variation than if selection were acting on a small number of large effect loci.

Contrary to predictions (Enberg *et al.* 2009, Kuparinen and Hutchings 2012) a recent study of Atlantic cod (*Gadus morhua*) populations has identified significant signs of population growth and biomass rebuilding (Rose and Rowe 2015). While these findings suggest the biomass of overexploited populations may show signs of recovery, the changes in size distributions observed by Rose & Rowe (2015) are the result of changes in age structure rather than age at size. It is therefore possible that the biomass recovery of these populations is the result of fish which would previously have been harvested surviving for longer. The rapid recovery of biomass is similar to that predicted by Enberg *et al.* (2009), who modelled the recovery of Atlantic cod. Under their model, although an initial rapid recovery (0-50 years post moratorium) in population biomass was seen, it was closer to 700 years before populations reached their pre-harvest levels of biomass and >2000 years before life history traits fully recovered.

## 5.7 Importance of epigenetic change

Our results show that the phenotypic changes observed following the size selective harvesting undertaken here are the result of primarily genetic change rather than changes in DNA methylation. It should, however, be noted that only a small number of methylated loci were examined here, and these loci were anonymous. Given the large body of evidence that phenotypic changes can be underpinned by epigenetic change it is extremely likely that epigenetic modifications are playing a role in the rapid shifts in life history traits observed as a result of size selective harvesting.

As only 91 epigenetic loci were analysed in the current project, it was not possible to say whether the overall level of DNA methylation across the genome had changed. The level of DNA methylation is regulated by a group of enzymes known as DNA methyltransferases (Dnmts) (Jones 2012; Briones & Muegge 2012). In mammals, the DNMTs consist of DNMT1, DNMT2, DNMT3a, DNMT3b and DNMT3L and can be categorised according to the primary function. The group of DNMT3 proteins, also known as the *de novo* methyltransferases, are vital in setting up the pattern of DNA methylation which takes places in early development (Jones 2012), while the DNMT1 enzyme is important for maintaining the levels of DNA methylation during DNA replication (Smith *et al.* 2011). However, although these two groups of DNMTs can be categorised by their primary function, reports have suggested that the pathways of methylation may overlap (Robertson and Jones 2000, Chen *et al.* 2003, Estève *et al.* 2005) and that neither can act without the participation of the other (Jones 2012). Unlike the DNMT1 and the DNMT3 gene family, DNMT2 is not strictly a DNA methyltransferase as it is thought to act on RNA rather than DNA (Rai *et al.* 2007). Research in zebrafish has identified highly conserved orthologues of the DNMT1 and DNMT2 which both do similar jobs to the DNMT group of enzymes (Smith *et al.* 2011). Unlike mammals, the zebrafish has been found to have at least six of the *de novo* methyltransferase (DNMT3) genes (Mhanni *et al.* 2001, Smith *et al.* 2005). As the pattern of DNA methylation is determined and maintained by the DNMT enzymes, several studies have used real-time reverse transcription PCR (qPCR) to quantify the level of their mRNA transcripts (Attwood and Richardson 2004, Shimoda *et al.* 2005, Smith *et al.* 2011, Campos *et al.* 2012, Fang *et al.* 2013, Sarabi and Naghibalhossaini 2015, Xia *et al.* 2015). Had appropriate samples been collected, such a technique would have provided a useful overview of any changes in the level of global DNA methylation over the course of our selection experiment, and between the small and large selection lines upon completion.

Although epigenetic change has, to our knowledge, not been considered in relation to FIE there are many studies indicating that rapid shifts in selection pressure can drive epigenetic change (Bonduriansky *et al.* 2012). One example of epigenetic change following a rapid shift in selection pressure can be seen in studies of invasive species. Studies of the house sparrow (*Passer domesticus*) showed that recently introduced populations in Kenya had higher levels of DNA methylation than established populations (Schrey *et al.* 2012). Furthermore, the introduced populations with the highest epigenetic diversity had the lowest genetic diversity and vice versa

(Liebl *et al.* 2013, 2015). Similar patterns of reduced genetic diversity and high levels of epigenetic diversity have also been found in species of Japanese knotweed (*Fallopia japonica*) (Richards *et al.* 2012) and Alligator weed (*Alternanthera philoxeroides*) (Gao *et al.* 2010).

The potential for epigenetic change to be underpinning rapid phenotypic shifts also raises the question of whether epigenetic changes are playing a role in the rapid phenotypic shifts seen in translocated populations of guppies. Numerous experiments have shown that guppies translocated from a site with high levels of predation to a site with low levels of predation will rapidly evolve phenotypic traits similar to those of other low predation sites (Endler 1980, Reznick and Bryga 1987a, Reznick 1990, Karim *et al.* 2007). Studies have also shown significant genetic divergence between these sites 57 years after the introduction (Fitzpatrick *et al.* 2014). Despite the years of studies, the genomic underpinning of many of the traits, which evolve following a change in predation, remains unclear. It is therefore possible that at least in the period immediately following the introduction, epigenetic changes may be contributing.

## 5.8 Future work

The results of the work presented here provide evidence of the genome wide effect size selective harvesting can have. However many questions remain unanswered in relation to the genomic effects of fisheries-induced evolution, as well as the genomic architecture underpinning life history shifts in the guppy.

Due to the complex nature of polygenic traits such as the life history traits examined here, the task of determining the genes underpinning these traits is not easy. In the case of *P.reticulata*, findings show that a significant amount of the genetic variation in these traits is located on the sex chromosome (chromosome 12). Therefore elucidation of genes underpinning these traits requires a complete reference sequence for both the X and the Y-chromosomes. Furthermore a thorough understanding of the level of recombination between these chromosomes will help to determine the degree of linkage and number of genes underpinning the changes observed.

During the current study, we used RAD sequencing to reduce the amount of sequencing required to examine the entire genome. Such an approach results in sequencing being spread across both the non-coding and coding regions of the genome. As there is now a draft reference genome for the guppy, it would be possible to use design probes to capture only the coding regions of the genome and sequence these (Grover *et al.* 2012). The exome sequencing approach has been widely used to examine the mutations responsible for human diseases (Ng *et al.* 2010, DaRe *et al.* 2013, Warr *et al.* 2015) but has also been used in non-model species for identification of candidate genes (Cosart *et al.* 2011, Ahonen *et al.* 2013, McClure *et al.* 2014, Pankin *et al.* 2014). As only a small proportion of genomes are thought to be coding, exome sequencing dramatically reduces the amount of sequence data required for each individual, enabling more individuals to be studied.

When the current project began, the guppy reference genome was not available and therefore exome sequencing was not an option. It would be interesting to see if exome sequencing a number of fish from both the selection experiment and the wild populations could identify any additional candidate genes for the life history traits examined. However, as the current guppy reference genome is incomplete, it is unlikely that probes designed from this draft would capture the entire exome. Therefore, while exome sequencing might identify additional candidate loci, it is unlikely to provide a complete picture.

In order to determine the role epigenetics is playing it would also be interesting to examine epigenetic variation in guppies from high predation and low predation sites. Initially this could simply be done by analysing the levels of DNA methylation between the sites, including recent translocations. As discussed in section 5.7 global methylation could also be assessed by examining level of the DNMT methylation enzymes.

Being able to determine the genes underpinning shifts in life history traits in overexploited fish populations would allow the changes in these genes to be determined. However given the complexity of the traits and the large number of species involved, such an approach remains impractical. Therefore, given the evidence provided by this and van Wijk *et al.* (2013) which shows that size selective harvesting can induce genetic change, it is important that the focus shifts to the ecosystem-wide effects that these changes can exert, as well as management implications. As previously discussed, ecosystem wide effects are already beginning to be examined both in guppies (Travis, Reznick, and Bassar 2014, El-Sabaawi, Bassar, *et al.* 2015, El-Sabaawi, Marshall, *et al.* 2015) and other species (Post *et al.* 2008, Palkovacs and Post 2009, Walsh and Post 2011). Although potential management implications are also being discussed (Kuparinen and Merilä 2007, Heino *et al.* 2013, Laugen *et al.* 2014) they have yet to be implicated.

Although it is important to prevent exploitation of fish populations driving selection in the way it has been shown to do, for many populations it is likely that detrimental effects have already been exerted. Therefore a better understanding of the potential for recovery of the phenotypic traits is vital. Several studies have used eco-genetic models, though the results of these studies do not completely agree with the only experimental study to examine the timeframe for recovery (Conover *et al.* 2009). There is, therefore a need for studies in which both the selection and recovery period are examined. Experimental selection utilising a more realistic design in which the effect of selection over a larger number of generations followed by a prolonged period of recovery are needed to give a more realistic understanding of the recovery of harvested populations. Additionally it would be interesting to use selection experiments to examine the effect of the various management regimes which have been selected, particularly the maximum size limit.

In summary, there is now a strong body of evidence that size selective harvesting can cause genome-wide changes, some of which at least underpin shifts in life history traits. There is now an

escalating need to take action to reduce the impact of future harvesting, both on the target species and the ecosystem as a whole, as well as identifying management strategies to allow traits to recover.

# References


Adelman, J. P., Maylie, J., and Sah, P., 2012. Small-Conductance Ca 2+ -Activated K + Channels: Form and Function. *Annual Review of Physiology*, 74 (1), 245–269.

Agrama, H. A., Eizenga, G. C., and Yan, W., 2007. Association mapping of yield and its components in rice cultivars. *Molecular Breeding*, 19 (4), 341–356.

Agrillo, C., Piffer, L., Bisazza, A., and Butterworth, B., 2012. Evidence for two numerical systems that are similar in humans and guppies. *PLoS One*.

Ahonen, S. J., Arumilli, M., and Lohi, H., 2013. A CNGB1 frameshift mutation in Papillon and Phalène dogs with progressive retinal atrophy. *PloS one*, 8 (8), e72122.

Akey, J. M., 2012. Parallel selection: evolution's surprising predictability. *Current biology*, 22 (10), R407–9.

Alexander, H., Taylor, J., Wu, S., and Breden, F., 2006. Parallel evolution and vicariance in the guppy (Poecilia reticulata) over multiple spatial and temporal scales. *Evolution*.

Allendorf, F. W. and Hard, J. J., 2009. Human-induced evolution caused by unnatural selection through harvest of wild animals. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl (Supplement_1), 9987–94.

Allfrey, V. G., Faulknet, R., and Mirsky, A. E., 1964. Acetylation and methylation of histones and their possible role in the regulation of rna synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 51 (5), 786–94.

Allis, C. D., Caparros, M., Jenuwein, T., and Reinberg, D., 2015. *Epigenetics*. Second. Cold Spring Harbor Laboratory Press.

Almasy, L., Hixson, J. E., Rainwater, D. L., Cole, S., Williams, J. T., Mahaney, M. C., VandeBerg, J. L., Stern, M. P., MacCluer, J. W., and Blangero, J., 1999. Human pedigree-based quantitative-trait-locus mapping: localization of two genes influencing HDL-cholesterol metabolism. *American journal of human genetics*, 64 (6), 1686–93.

Almazan-Rueda, P., van Helmond, A. T. M., Verreth, J. A. J., and Schrama, J. W., 2005. Photoperiod affects growth, behaviour and stress variables in Clarias gariepinus. *Journal of Fish Biology*, 67 (4), 1029–1039.

Altshuler, D., Daly, M., and Lander, E., 2008. Genetic mapping in human disease. *science*.

Alvarado, S., Rajakumar, R., Abouheif, E., and Szyf, M., 2015. Epigenetic variation in the Egfr gene generates quantitative variation in a complex trait in ants. *Nature Communications*, 6 (October), 1–9.

Amish, S. J., Hohenlohe, P. A., Painter, S., Leary, R. F., Muhlfeld, C., Allendorf, F. W., and Luikart, G., 2012. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. *Molecular ecology resources*, 12 (4), 653–60.

Ammerpohl, O., Martín-Subero, J. I., Richter, J., Vater, I., and Siebert, R., 2009. Hunting for the 5th base: Techniques for analyzing DNA methylation. *Biochimica et biophysica acta*, 1790 (9), 847–62.

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q., and Postlethwait, J. H., 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 188 (4), 799–808.

Amthor, H., Nicholas, G., McKinnell, I., Kemp, C. F., Sharma, M., Kambadur, R., and Patel, K., 2004. Follistatin complexes Myostatin and antagonises Myostatin-mediated inhibition of myogenesis. *Developmental biology*, 270 (1), 19–30.

Amundsen, T., 2003. Fishes as models in studies of sexual selection and parental care. *Journal of Fish Biology*, 63 (s1), 17–52.

Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., Félix, M.-A., and Kruglyak, L., 2012. Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nature Genetics*, 44 (3), 285–290.

Andersen, I. S., Reiner, A. H., Aanes, H., Alestrom, P., and Collas, P., 2012. Developmental features of DNA methylation during activation of the embryonic zebrafish genome. *Genome Biology*, 13 (7), R65.

Andersen, M. R., Vongsangnak, W., Panagiotou, G., Salazar, M. P., Lehmann, L., and Nielsen, J., 2008. A trispecies Aspergillus microarray: comparative transcriptomics of three Aspergillus species. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (11), 4387–92.

Anderson, J. T., Wagner, M. R., Rushworth, C. A., Prasad, K. V. S. K., and Mitchell-Olds, T., 2014. The evolution of quantitative traits in complex environments. *Heredity*, 112 (1), 4–12.

Andersson, M. and Simmons, L. W., 2006. Sexual selection and mate choice. *Trends in ecology & evolution*, 21 (6), 296–302.

Angers, B., Castonguay, E., and Massicotte, R., 2010. Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. *Molecular Ecology*, 19 (7), 1283–1295.

Ansorge, W. J., 2009. Next-generation DNA sequencing techniques. *New biotechnology*, 25 (4), 195–203.

Antao, T. and Beaumont, M. a., 2011. Mcheza: A workbench to detect selection using dominant markers. *Bioinformatics*, 27 (12), 1717–1718.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., and Luikart, G., 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC bioinformatics*, 9 (1), 323.

Antonopoulou, E., Mayer, I., Berglund, I., and Borg, B., 1995. Effects of aromatase inhibitors on sexual maturation in Atlantic salmon, Salmo salar, male parr. *Fish physiology and biochemistry*, 14 (1), 15–24.

Anway, M., Cupp, A., Uzumcu, M., and Skinner, M., 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*, 308 (5727), 1466–1469.

Aranzana, M., Kim, S., Zhao, K., and Bakker, E., 2005. Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genet*.

Aranzana, M. M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., and Nordborg, M., 2005. Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS genetics*, 1 (5), e60.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Ménard, S., Rogers, W. O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.-C., Fairhurst, R. M., Benoit-Vical, F., Mercereau-Puijalon, O., and Ménard, D., 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, 505 (7481), 50–5.

Armstrong, L., 2013. *Epigenetics*. Garland Science.

Arnold, B., Corbett-Detig, R. B., Hartl, D., and Bomblies, K., 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular ecology*, 22 (11), 3179–90.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25 (1), 25–9.

Ashworth, C. J., Toma, L. M., and Hunter, M. G., 2009. Nutritional effects on oocyte and embryo development in mammals: implications for reproductive efficiency and environmental sustainability. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364 (1534), 3351–61.

Atchley, W. R., Fitch, W. M., and Bronner-Fraser, M., 1994. Molecular evolution of the MyoD family

of transcription factors. *Proceedings of the National Academy of Sciences*, 91 (24), 11522–11526.

Attwood, J. and Richardson, B., 2004. Relative quantitation of DNA methyltransferase mRNA by real-time RT-PCR assay. *Methods in molecular biology (Clifton, N.J.)*, 287, 273–83.

Audzijonyte, A., Kuparinen, A., Gorton, R., and Fulton, E. A., 2013. Ecological consequences of body size decline in harvested fish species: positive feedback loops in trophic interactions amplify human impact. *Biology letters*, 9 (2), 20121103.

Avramidou, E. V., Ganopoulos, I. V., Doulis, A. G., Tsaftaris, A. S., and Aravanopoulos, F. a., 2015. Beyond population genetics: natural epigenetic variation in wild cherry (Prunus avium). *Tree Genetics & Genomes*, 11 (5), 95.

Ayala, F. J., Powell, J. R., and Dobzhansky, T., 1971. Polymorphisms in Continental and Island Populations of Drosophila willistoni. *Proceedings of the National Academy of Sciences*, 68 (10), 2480–2483.

Azzi, A., Dallmann, R., Casserly, A., Rehrauer, H., Patrignani, A., Maier, B., Kramer, A., and Brown, S. A., 2014. Circadian behavior is light-reprogrammed by plastic DNA methylation. *Nature neuroscience*, 17 (3), 377–82.

Backström, N., Brandström, M., Gustafsson, L., Qvarnström, A., Cheng, H., and Ellegren, H., 2006. Genetic mapping in a natural population of collared flycatchers (Ficedula albicollis): conserved synteny but gene order rearrangements on the avian Z chromosome. *Genetics*, 174 (1), 377–86.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, 3 (10), e3376.

Baker, J. A., Wund, M. A., Heins, D. C., King, R. W., Reyes, M. L., and Foster, S. A., 2015. Life-history plasticity in female threespine stickleback. *Heredity*, 115 (4), 322–34.

Bakulski, K. M., Dolinoy, D. C., Sartor, M. A., Paulson, H. L., Konen, J. R., Lieberman, A. P., Albin, R. L., Hu, H., and Rozek, L. S., 2012. Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. *Journal of Alzheimer's disease : JAD*, 29 (3), 571–88.

Balasubramanian, S., Schwartz, C., Singh, A., Warthmann, N., Kim, M. C., Maloof, J. N., Loudet, O., Trainer, G. T., Dabi, T., Borevitz, J. O., Chory, J., and Weigel, D., 2009. QTL mapping in new Arabidopsis thaliana advanced intercross-recombinant inbred lines. *PloS one*, 4 (2), e4318.

Baldo, L., Santos, M. E., and Salzburger, W., 2011. Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome biology and evolution*, 3, 443–55.

Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Valè, G., Toppino, L., and Rotino, G. L., 2011. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC genomics*, 12 (1), 304.

Barot, S., Heino, M., Morgan, M., and Dieckmann, U., 2005. Maturation of Newfoundland American plaice: long-term trends in maturation reaction norms despite low fishing mortality? *ICES Journal of Marine Science*, 62 (1), 56–64.

Barson, N. J., Cable, J., and Van Oosterhout, C., 2009. Population genetic analysis of microsatellite variation of guppies (Poecilia reticulata) in Trinidad and Tobago: evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. *Journal of evolutionary biology*, 22 (3), 485–97.

Bassar, R., 2010. Bridging the gap between ecology and evolution: integrating density regulation and life-history evolution. *Annals of the New York Academy of Sciences*, 1206, 17–34.

Bassar, R. D., Heatherly, T., Marshall, M. C., Thomas, S. a., Flecker, A. S., and Reznick, D. N., 2015. Population size structure dependent fitness and ecosystem consequences in trinidadian guppies. *Journal of Animal Ecology*, 84 (4), 955–968.

Bassar, R. D., Lopez-Sepulcre, A., Reznick, D. N., and Travis, J., 2013. Experimental Evidence for

Density-Dependent Regulation and Selection on Trinidadian Guppy Life Histories. *The American Naturalist*, 181 (1), 25–38.

Bassar, R. D., Marshall, M. C., Lopez-Sepulcre, a., Zandona, E., Auer, S. K., Travis, J., Pringle, C. M., Flecker, a. S., Thomas, S. a., Fraser, D. F., and Reznick, D. N., 2010. Local adaptation in Trinidadian guppies alters ecosystem processes. *Proceedings of the National Academy of Sciences*, 107 (8), 3616–3621.

Baulier, L., Heino, M., Lilly, G., and Dieckmann, U., 2006. Body condition and evolution of maturation of Atlantic cod in Newfoundland. *ICES report*.

Baxter, S. W., Davey, J. W., Johnston, J. S., Shelton, A. M., Heckel, D. G., Jiggins, C. D., and Blaxter, M. L., 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PloS one*, 6 (4), e19315.

Baylin, S., 2005. DNA methylation and gene silencing in cancer. *Nature clinical practice Oncology*.

Beaumont, M. A. and Nichols, R. A., 1996. Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society B: Biological Sciences*, 263 (1377), 1619–1626.

Becher, S. A., Russell, S. T., and Magurran, A. E., 2002. Isolation and characterization of polymorphic microsatellites in the Trinidadian guppy (Poecilia reticulata). *Molecular Ecology Notes*, 2 (4), 456–458.

Becher, S. and Magurran, A., 2000. Gene flow in Trinidadian guppies. *Journal of Fish Biology*, 56 (2), 241–249.

Becker, P. B. and Workman, J. L., 2013. Nucleosome remodeling and epigenetics. *Cold Spring Harbor perspectives in biology*, 5 (9), a017905–.

Beebee, T. and Rowe, G., 2008. *An introduction to molecular ecology*. Oxford University Press.

Beekman, S. P. A., Kemp, B., Louwman, H. C. M., and Colenbrander, B., 1999. Analyses of factors influencing the birth weight and neonatal growth rate of Cheetah (Acinonyx jubatus) cubs. *Zoo Biology*, 18 (2), 129–139.

Bekaert, B., Kamalandua, A., Zapico, S. C., Van de Voorde, W., and Decorte, R., 2015. Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*, 10 (10), 1–9.

Bergman, Y. and Cedar, H., 2013. DNA methylation dynamics in health and disease. *Nature structural & molecular biology*, 20 (3), 274–81.

Bernstein, E. and Allis, C. D., 2005. RNA meets chromatin. *Genes & development*, 19 (14), 1635–55.

Bhatta, S., Iwai, T., Miura, T., Higuchi, M., and Maugars, G., 2012. Differences Between Male and Female Growth and Sexual Maturation in Tilapia ( Oreochromis Mossambicus ). *Kathmandu University Journal of Science, Engineering and Technology*, 8 (II), 57–65.

Bhattacharya, T. K., Chatterjee, R. N., Sharma, R. P., Niranjan, M., Rajkumar, U., and Reddy, B. L. N., 2011. Polymorphism in the prolactin promoter and its association with growth traits in chickens. *Biochemical genetics*, 49 (5-6), 385–94.

Bierne, N., Roze, D., and Welch, J. J., 2013. Pervasive selection or is it…? why are F ST outliers sometimes so frequent? *Molecular Ecology*, 22 (8), 2061–2064.

Birchler, J. A., Auger, D. L., and Riddle, N. C., 2003. In search of the molecular basis of heterosis. *The Plant cell*, 15 (10), 2236–9.

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A., 2010. Heterosis. *The Plant cell*, 22 (7), 2105–12.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. *Genes & development*, 16 (1), 6–21.

Bird, A., 2007. Perceptions of epigenetics. *Nature*.

Bird, A. and Wolffe, A., 1999. Methylation-induced repression—belts, braces, and chromatin. *Cell*, 99 (5), 451–454.

Biro, P. A. and Post, J. R., 2008. Rapid depletion of genotypes with fast growth and bold personality traits from harvested fish populations. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (8), 2919–22.

Björklund, M., Ranta, E., Kaitala, V., Bach, L. A., Lundberg, P., and Stenseth, N. C., 2009. Quantitative trait evolution and environmental change. *PloS one*, 4 (2), e4521.

Björnsson, B. T., Taranger, G. L., Hansen, T., Stefansson, S. O., and Haux, C., 1994. The interrelation between photoperiod, growth hormone, and sexual maturation of adult Atlantic salmon (Salmo salar). *General and comparative endocrinology*, 93 (1), 70–81.

Blackstone, N. and Joslyn, A., 1984. Utilization and preference for the introduced gastropod Littorina littorea(L.) by the hermit crab Pagurus longicarpus(Say) at guilford,. *Journal of experimental marine biology and …*, 80 (1), 1–9.

Blouin, M., Thuillier, V., Cooper, B., Amarasinghe, V., Cluzel, L., Araki, H., and Grunau, C., 2010. No evidence for large differences in genomic methylation between wild and hatchery steelhead (Oncorhynchus mykiss). *Canadian Journal of Fisheries and Aquatic Sciences*, 67 (2), 217–224.

Boeuf, G. and Le Bail, P.-Y., 1999. Does light have an influence on fish growth? *Aquaculture*, 177 (1-4), 129–152.

Bœuf, G. and Payan, P., 2001. How should salinity influence fish growth? *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 130 (4), 411–423.

Bohlin, T., Sundstrom, L. F., Johnsson, J. I., Hojesjo, J., and Pettersson, J., 2002. Density-dependent growth in brown trout: effects of introducing wild and hatchery fish. *Journal of Animal Ecology*, 71 (4), 683–692.

Bolstad, G. H., Pélabon, C., Larsen, L.-K., Fleming, I. A., Viken, A., and Rosenqvist, G., 2012. The effect of purging on sexually selected traits through antagonistic pleiotropy with survival. *Ecology and evolution*, 2 (6), 1181–94.

Bonduriansky, R., Crean, A. J., and Day, T., 2012. The implications of nongenetic inheritance for evolution in changing environments. *Evolutionary applications*, 5 (2), 192–201.

Borash, D. J., Teotónio, H., Rose, M. R., and Mueller, L. D., 2000. Density-dependent natural selection in Drosophila: Correlations between feeding rate, development time and viability. *Journal of Evolutionary Biology*, 13 (FEBRUARY), 181–187.

Bossdorf, O., Richards, C. L., and Pigliucci, M., 2007. Epigenetics for ecologists. *Ecology Letters*, 11 (2), 106–115.

Boulesteix, A.-L., Janitza, S., Kruppa, J., and König, I., 2012. Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics, (129).

Bourret, V., Dionne, M., and Bernatchez, L., 2014. Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular ecology*, 23 (18), 4444–57.

Boutet, I., Lorin-Nebel, C., De Lorgeril, J., and Guinand, B., 2007. Molecular characterisation of prolactin and analysis of extrapituitary expression in the European sea bass Dicentrarchus labrax under various salinity conditions. *Comparative biochemistry and physiology. Part D, Genomics & proteomics*, 2 (1), 74–83.

Bower, N. I. and Johnston, I. a, 2010. Discovery and characterization of nutritionally regulated genes associated with muscle growth in Atlantic salmon, (79), 114–130.

Boyko, A., Blevins, T., Yao, Y., and Golubov, A., 2010. Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of Dicer-like proteins. *PloS one*, 5 (3), e9514.

Brachi, B., Morris, G. P., and Borevitz, J. O., 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology*, 12 (10), 232.

Brander, K., 1995. The effect of temperature on growth of Atlantic cod (Gadus morhua L.). *ICES Journal of Marine Science*, 52 (1), 1–10.

Breden, F., Scott, M., and Michel, E., 1987. Genetic differentiation for anti-predator behaviour in the Trinidad guppy, Poecilia reticulata. *Animal Behaviour*, 35 (2), 618–620.

Breden, F. and Stoner, G., 1987. Male predation risk determines female preference in the Trinidad guppy. *Nature*, 329 (6142), 831–833.

Breseghello, F. and Sorrells, M., 2006. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. *Genetics*, 172 (2), 1165–1177.

Brieuc, M. S. O., Ono, K., Drinan, D. P., and Naish, K. A., 2015. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (Oncorhynchus tshawytscha). *Molecular ecology*, 24 (11), 2729–46.

Brooks, R., 2000. Negative genetic correlation between male sexual attractiveness and survival. *Nature*, 406 (6791), 67–70.

Brooks, R. and Endler, J. A., 2007. Direct and indirect sexual selection and quantitative genetics of male traits in guppies (poecilia reticulata). *Evolution*, 55 (5), 1002–1015.

Brouwer, D. J. and St Clair, D. A., 2004. Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 108 (4), 628–38.

Browman, H. I., Law, R., and Marshall, C. T., 2008. The role of fisheries-induced evolution. *Science (New York, N.Y.)*, 320 (5872), 47–50; author reply 47–50.

Browman, H. I., Law, R., Marshall, C. T., Jorgensen, C., Enberg, K., Dunlop, E. S., Arlinghaus, R., Boukal, D. S., Brander, K., Ernande, B., Gardmark, a, Johnston, F., Matsumura, S., Pardoe, H., Raab, K., Silva, a, Vainikka, a, Dieckmann, U., Heino, M., and Rijnsdorp, a D., 2008. The role of fisheries-induced evolution. *Science*, 320 (5872), 48–50.

Brown, C., Hobday, A., and Ziegler, P., 2008. Darwinian fisheries science needs to consider realistic fishing pressures over evolutionary time scales. *Marine Ecology Progress Series*, 369, 257–266.

Brown, E. A., Pilkington, J. G., Nussey, D. H., Watt, K. A., Hayward, A. D., Tucker, R., Graham, A. L., Paterson, S., Beraldi, D., Pemberton, J. M., and Slate, J., 2013. Detecting genes for variation in parasite burden and immunological traits in a wild population: testing the candidate gene approach. *Molecular ecology*, 22 (3), 757–73.

Brown, K. H., Schultz, I. R., and Nagler, J. J., 2009. Lack of a heritable reproductive defect in the offspring of male rainbow trout exposed to the environmental estrogen 17alpha-ethynylestradiol. *Aquatic toxicology (Amsterdam, Netherlands)*, 91 (1), 71–4.

Bruce, A. B., 1910. THE MENDELIAN THEORY OF HEREDITY AND THE AUGMENTATION OF VIGOR. *Science (New York, N.Y.)*, 32 (827), 627–8.

Bryden, C. A., Heath, J. W., and Heath, D. D., 2004. Performance and heterosis in farmed and wild Chinook salmon (Oncorhynchus tshawytscha) hybrid and purebred crosses. *Aquaculture*, 235 (1-4), 249–261.

Burgess, D. J., 2013. Human epigenetics: Showing your age. *Nature reviews. Genetics*, 14 (1), 6.

Burke, M., Dunham, J., and Shahrestani, P., 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature*.

Burke, M. K., 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings. Biological sciences / The Royal Society*, 279 (1749), 5029–38.

Caballero, A., Quesada, H., and Rolán-Alvarez, E., 2008. Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, 179 (1), 539–554.

Cable, J. and van Oosterhout, C., 2007. The impact of parasites on the life history evolution of guppies (Poecilia reticulata): the effects of host size on parasite virulence. *International journal for parasitology*, 37 (13), 1449–58.

Cain, A. J. and Sheppard, P. M., 1950. Selection in the polymorphic land snail Cepaea nemoralis. *Heredity*, 4 (3), 275–94.

Campbell, N. R., LaPatra, S. E., Overturf, K., Towner, R., and Narum, S. R., 2014. Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. *G3 (Bethesda, Md.)*, 4 (12), 2473–81.

Campos, C., Fernandes, J. M. O., Conceição, L. E. C., Engrola, S., Sousa, V., and Valente, L. M. P., 2013. Thermal conditions during larval pelagic phase influence subsequent somatic growth of Senegalese sole by modulating gene expression and muscle growth dynamics. *Aquaculture*, 414-415, 46–55.

Campos, C., Valente, L. M. P., and Fernandes, J. M. O., 2012. Molecular evolution of zebrafish dnmt3 genes and thermal plasticity of their expression during embryonic development. *Gene*, 500 (1), 93–100.

Candaele, J., Demuynck, K., Mosoti, D. R., Mosoti, D. R., Beemster, G. T. S., Inze, D., Nelissen, H., Inzé, D., and Nelissen, H., 2014. Differential Methylation during Maize Leaf Growth Targets Developmentally Regulated Genes. *Plant physiology*, 164 (3), 1350–64.

Cao, J., Wei, C., Liu, D., Wang, H., Wu, M., Xie, Z., Capellini, T. D., Zhang, L., Zhao, F., Li, L., Zhong, T., Wang, L., Lu, J., Liu, R., Zhang, S., Du, Y., Zhang, H., and Du, L., 2015. DNA methylation Landscape of body size variation in sheep. *Scientific Reports*, 5, 13950.

Cao, X. and Jacobsen, S. E., 2002. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 4 (suppl 4), 16491–8.

Capuano, F., Mülleder, M., Kok, R., Blom, H. J., Ralser, M., Mülleder, M., and Kok, R., 2014. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Analytical chemistry*, 86 (8), 3697–702.

Cariou, M., Duret, L., and Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and evolution*, 3 (4), 846–52.

Carlson, S. and Seamons, T., 2008. A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. *Evolutionary Applications*, 1 (2), 222–38.

Carrière, Y. and Roff, D. A., 1995. Change in genetic architecture resulting from the evolution of insecticide resistance: a theoretical and empirical analysis. *Heredity*, 75 (6), 618–629.

Carroll, S. B., 2005. Evolution at two levels: on genes and form. *PLoS biology*, 3 (7), e245.

Carroll, S. P., Hendry, A. P., Reznick, D. N., and Fox, C. W., 2007. Evolution on ecological time-scales. *Functional Ecology*, 21 (3), 387–393.

Carvalho, G. R., Shaw, P. W., Hauser, L., Seghers, B. H., and Magurran, A. E., 1996. Artificial introductions, evolutionary change and population differentiation in Trinidadian guppies (Poecilia reticulata: Poeciliidae). *Biological Journal of the Linnean Society*, 57 (3), 219–234.

Carvalho, G., Shaw, P., Magurran, A., and Seghers, B., 1991. Marked genetic divergence revealed by allozymes among populations of the guppy Poecilia reticulata (Poeciliidae), in Trinidad. *Biological Journal of the Linnean Society*, 42 (3), 389–405.

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., Franks, C. D., and Kresovich, S., 2008. Community Resources and Strategies for Association Mapping in Sorghum. *Crop Science*, 48 (1), 30.

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q., and Cresko, W. A., 2013. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular ecology*, 22 (11), 2864–83.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H., 2011. Stacks: building and genotyping Loci de novo from short-read sequences. *G3*, 1 (3), 171–82.

Charan, R., Suresh Babu, P. P., Venugopal, G., Chadha, N. K., and Sreeramamurty, K. B., 2013. Effect of aromatase inhibitors on the ovarian development of stunted yearlings of rohu (Labeo rohita): a preliminary study. *Aquaculture International*, 22 (2), 689–697.

Charlesworth, B., Nordborg, M., and Charlesworth, D., 1997. The effects of local selection,

balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, 70 (02), 155–174.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.-C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B. R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R. J., and Georges, M., 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature genetics*, 40 (4), 449–54.

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., and Li, E., 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and cellular biology*, 23 (16), 5594–605.

Chen, Z., Cheng, C.-H. C., Zhang, J., Cao, L., Chen, L., Zhou, L., Jin, Y., Ye, H., Deng, C., Dai, Z., Xu, Q., Hu, P., Sun, S., Shen, Y., and Chen, L., 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (35), 12944–9.

Cheng, W. T., Lee, C. H., Hung, C. M., Chang, T. J., and Chen, C. M., 2000. Growth hormone gene polymorphisms and growth performance traits in Duroc, Landrace and Tao-Yuan pigs. *Theriogenology*, 54 (8), 1225–37.

Chitramuthu, B. P., Baranowski, D. C., Cadieux, B., Rousselet, E., Seidah, N. G., and Bennett, H. P. J., 2010. Molecular cloning and embryonic expression of zebrafish PCSK5 co-orthologues: functional assessment during lateral line development. *Developmental dynamics : an official publication of the American Association of Anatomists*, 239 (11), 2933–46.

Chiyo, P. I., Obanda, V., and Korir, D. K., 2015. Illegal tusk harvest and the decline of tusk size in the African elephant. *Ecology and Evolution*, 5 (22), 5216–5229.

Chmurzynska, A., 2010. Fetal programming: link between early nutrition, DNA methylation, and complex diseases. *Nutrition reviews*, 68 (2), 87–98.

Chutimanitsakun, Y., Nipper, R. W., Cuesta-Marcos, A., Cistué, L., Corey, A., Filichkina, T., Johnson, E. A., and Hayes, P. M., 2011. Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC …*, 12, 4.

Cingolani, O. H., Kirk, J. a., Seo, K., Koitabashi, N., Lee, D. -i., Ramirez-Correa, G., Bedja, D., Barth, a. S., Moens, a. L., and Kass, D. a., 2011. Thrombospondin-4 Is Required for Stretch-Mediated Contractility Augmentation in Cardiac Muscle. *Circulation Research*, 109 (12), 1410–1414.

Clark, M. R., Althaus, F., Schlacher, T. A., Williams, A., Bowden, D. A., and Rowden, A. A., 2015. The impacts of deep-sea fisheries on benthic communities: a review. *ICES Journal of Marine Science*, fsv123–.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E., 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, 452 (7184), 215–9.

Colicchio, J. M., Miura, F., Kelly, J. K., Ito, T., and Hileman, L. C., 2015. DNA methylation and gene expression in Mimulus guttatus. *BMC genomics*, 16 (1), 507.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., and Kingsley, D. M., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science (New York, N.Y.)*, 307 (5717), 1928–33.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., and Kingsley, D. M., 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS biology*, 2 (5), E109.

Coltman, D. D. W., O'Donoghue, P., Jorgenson, J. J. T. J., Hogg, J. T., Strobeck, C., and Festa-Bianchet, M., 2003. Undesirable evolutionary consequences of trophy hunting. *Nature*, 426 (6967), 655–8.

Coltman, D., O'Donoghue, P., and Jorgenson, J., 2003. Undesirable evolutionary consequences of

trophy hunting. *Nature*.

Conover, D. and Baumann, H., 2009a. PERSPECTIVE: The role of experiments in understanding fishery-induced evolution. *Evolutionary Applications*.

Conover, D. O., 2007. Fisheries: nets versus nature. *Nature*, 450 (7167), 179–80.

Conover, D. O. and Baumann, H., 2009b. The role of experiments in understanding fishery-induced evolution. *Evolutionary applications*, 2 (3), 276–90.

Conover, D. O. and Munch, S. B., 2002. Sustaining fisheries yields over evolutionary time scales. *Science (New York, N.Y.)*, 297 (5578), 94–6.

Conover, D. O., Munch, S. B., and Arnott, S. A., 2009. Reversal of evolutionary downsizing caused by selective harvest of large fish. *Proceedings. Biological sciences / The Royal Society*, 276 (1664), 2015–20.

Conte, G. L., Arnegard, M. E., Peichel, C. L., and Schluter, D., 2012. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences*, 279 (1749), 5039–5047.

Cooke, S., Suski, C., and Ostrand, K., 2007. Physiological and Behavioral Consequences of Long-Term Artificial Selection for Vulnerability to Recreational Angling in a Teleost Fish. *Physiological and Biochemical Zooology: Ecological and Evolutionary approaches*, 80 (5), 480–90.

Cordell, H. J., 2009. Detecting gene-gene interactions that underlie human diseases. *Nature reviews. Genetics*, 10 (6), 392–404.

Le Corre, V. and Kremer, A., 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular ecology*, 21 (7), 1548–66.

Cortijo, S., Wardenaar, R., Colome-Tatche, M., Gilly, a., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.-M., Wincker, P., Roudier, F., Jansen, R. C., Colot, V., and Johannes, F., 2014. Mapping the Epigenetic Basis of Complex Traits. *Science*, 343 (6175), 1145–1148.

Cosart, T., Beja-Pereira, A., Chen, S., Ng, S. B., Shendure, J., and Luikart, G., 2011. Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC genomics*, 12 (1), 347.

Costa, F. F., 2005. Non-coding RNAs: New players in eukaryotic biology. *Gene*, 357 (2), 83–94.

Cotton, A. M., Price, E. M., Jones, M. J., Balaton, B. P., Kobor, M. S., and Brown, C. J., 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Human molecular genetics*, 24 (6), 1528–39.

Courtenay, W. R. . J., Meffe, G. K., Meffe, G. K., and Snelson, F. F. . J., 1989. Small fishes in strange places: a review of introduced poeciliids., 319–331.

Cousyn, C., De Meester, L., Colbourne, J., Brendonck, L., Vershuren, D., and Volckaert, F., 2001. Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (11), 6256–6260.

Covelo-Soto, L., Leunda, P. M., Pérez-Figueroa, A., and Morán, P., 2015. Genome-wide methylation study of diploid and triploid brown trout (Salmo trutta L.). *Animal genetics*, 46 (3), 280–8.

Crews, D., 2008. Epigenetics and its implications for behavioral neuroendocrinology. *Frontiers in neuroendocrinology*, 29 (3), 344–357.

Crews, D., Gore, A. C., Hsu, T. S., Dangleben, N. L., Spinetta, M., Schallert, T., Anway, M. D., and Skinner, M. K., 2007. Transgenerational epigenetic imprints on mate preference. *Proceedings of the National Academy of Sciences*, 104 (14), 5942–5946.

Crispo, E., Bentzen, P., Reznick, D. N., Kinnison, M. T., and Hendry, A. P., 2006. The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology*, 15 (1), 49–62.

Croft, D. and Krause, M., 2004. Is sexual segregation in the guppy, Poecilia reticulata, consistentwith the predation risk hypothesis? *Environmental Biology of Fishes*.

Croft, D. P., Botham, M. S., and Krause, J., 2009. Is sexual segregation in the guppy, Poecilia reticulata, consistent with the predation risk hypothesis? *Environmental Biology of Fishes*, 71, 127–133.

Crow, J. F., 1957. Genetics of Insect Resistance to Chemicals. *Annual Review of Entomology*, 2 (1), 227–246.

Cruickshank, T. E. and Hahn, M. W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, 23 (13), 3133–57.

Cubas, P., Vincent, C., and Coen, E., 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, 401, 157–161.

Curi, R. A., Oliveira, H. N., Silveira, A. C., and Lopes, C. R., 2005. Effects of polymorphic microsatellites in the regulatory region of IGF1 and GHR on growth and carcass traits in beef cattle. *Animal Genetics*, 36 (1), 58–62.

Czesny, S., Epifanio, J., and Michalak, P., 2012. Genetic divergence between freshwater and marine morphs of alewife (Alosa pseudoharengus): a 'next-generation' sequencing analysis. *PloS one*, 7 (3), e31803.

Daborn, P., Boundy, S., Yen, J., Pittendrigh, B., and ffrench-Constant, R., 2001. DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Molecular genetics and genomics : MGG*, 266 (4), 556–63.

Dalla, E., Mignone, F., Verardo, R., Marchionni, L., Marzinotto, S., Lazarević, D., Reid, J. F., Marzio, R., Klarić, E., Licastro, D., Marcuzzi, G., Gambetta, R., Pierotti, M. A., Pesole, G., and Schneider, C., 2005. Discovery of 342 putative new genes from the analysis of 5'-end-sequenced full-length-enriched cDNA human transcripts. *Genomics*, 85 (6), 739–51.

DaRe, J. T., Vasta, V., Penn, J., Tran, N.-T. B., and Hahn, S. H., 2013. Targeted exome sequencing for mitochondrial disorders reveals high genetic heterogeneity. *BMC medical genetics*, 14 (1), 118.

Darimont, C. T., Carlson, S. M., Kinnison, M. T., Paquet, P. C., Reimchen, T. E., and Wilmers, C. C., 2009. Human predators outpace other agents of trait change in the wild. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (3), 952–4.

Darwin, C., 1859. *On the Origin of Species*. London, UK: John Murray.

Daufresne, M., Lengfellner, K., and Sommer, U., 2009. Global warming benefits the small in aquatic ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (3), 12788–12793.

Davenport, C. B., 1908. DEGENERATION, ALBINISM AND INBREEDING. *Science (New York, N.Y.)*, 28 (718), 454–5.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., and Blaxter, M. L., 2013. Special features of RAD Sequencing data: implications for genotyping. *Molecular ecology*, 22 (11), 3151–64.

Davey, J. W., Hohenlohe, P. a, Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews. Genetics*, 12 (7), 499–510.

Deans, C. and Maggert, K. a., 2015. What Do You Mean, 'Epigenetic'? *Genetics*, 199 (4), 887–896.

Deaton, A. M. and Bird, A., 2011. CpG islands and the regulation of transcription. *Genes & development*, 25 (10), 1010–22.

Deeb, N. and Lamont, S., 2003. Use of a novel outbred by inbred F1 cross to detect genetic markers for growth. *Animal genetics*.

Dendrinos, P. and Thorpe, J. P., 1985. Effects of reduced salinity on growth and body composition in the European bass Dicentrarchus labrax (L.). *Aquaculture*, 49 (3-4), 333–358.

Deng, H.-W., Deng, X.-T., Conway, T., Xu, F.-H., Heaney, R., and Recker, R. R., 2002. Determination of Bone Size of Hip, Spine, and Wrist in Human Pedigrees by Genetic and Lifestyle Factors. *Journal of Clinical Densitometry*, 5 (1), 45–56.

Denson, M. R., Stuart, K. R., Smith, T. I. J., Weirlch, C. R., and Segars, A., 2003. Effects of Salinity

on Growth, Survival, and Selected Hematological Parameters of Juvenile Cobia Rachycentron canadum. *Journal of the World Aquaculture Society*, 34 (4), 496–504.

Derrien, V., Couillault, C., Franco, M., Martineau, S., Montcourrier, P., Houlgatte, R., and Chavrier, P., 2002. A conserved C-terminal domain of EFA6-family ARF6-guanine nucleotide exchange factors induces lengthening of microvilli-like membrane protrusions. *Journal of cell science*, 115, 2867–2879.

Desrochers, A., 2010. Morphological response of songbirds to 100 years of landscape change in North America. *Ecology*, 91 (6), 1577–1582.

Dhavan, R. and Tsai, L.-H., 2001. A DECADE OF CDK5. *Nature Reviews Molecular Cell Biology*, 2 (10), 749–759.

Dieckmann, U. and Heino, M., 2007. Probabilistic maturation reaction norms: their history, strenghts and limitations. *Marine Ecology Progress Series*, 355, 253–269.

DiMichele, L. and Powers, D. a., 1991. Allozyme variation, developmental rate, and differential mortality in the teleost Fundulus heteroclitus. *Physiological zoology*, 64 (6), 1426–1443.

Ding, Y., He, F., Wen, H., Li, J., Ni, M., Chi, M., Qian, K., Bu, Y., Zhang, D., Si, Y., and Zhao, J., 2013. DNA methylation status of cyp17-II gene correlated with its expression pattern and reproductive endocrinology during ovarian development stages of Japanese flounder (Paralichthys olivaceus). *Gene*, 527 (1), 82–8.

Diopere, E., Hellemans, B., Volckaert, F. A. M., and Maes, G. E., 2013. Identification and validation of single nucleotide polymorphisms in growth- and maturation-related candidate genes in sole (Solea solea L.). *Marine genomics*, 9, 33–8.

Doerge, R. W., 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nature reviews. Genetics*, 3 (1), 43–52.

Doğan, N., Yazıcı, Z., Şişman, T., and Aşkin, H., 2013. Acute toxic effects of fenpyroximate acaricide on Guppy (Poecilia reticulata Peters, 1859). *Toxicology and industrial health*, 29 (8), 716–21.

Domingues, C. P. C., Creer, S., Taylor, M. I., Queiroga, H., and Carvalho, G. R., 2010. Genetic structure of Carcinus maenas within its native range: Larval dispersal and oceanographic variability. *Marine Ecology Progress Series*, 410, 111–123.

Drabo, I., Redden, R., Smithson, J. B., and Aggarwal, V. D., 1984. Inheritance of seed size in cowpea (Vigna unguiculata (L.) Walp.). *Euphytica*, 33 (3), 929–934.

Dreyer, C., Hoffmann, M., Lanz, C., Willing, E.-M., Riester, M., Warthmann, N., Sprecher, A., Tripathi, N., Henz, S. R., and Weigel, D., 2007. ESTs and EST-linked polymorphisms for genetic mapping and phylogenetic reconstruction in the guppy, Poecilia reticulata. *BMC genomics*, 8 (1), 269.

Dunlop, E. S., Baskett, M. L., Heino, M., and Dieckmann, U., 2009. Propensity of marine reserves to reduce the evolutionary effects of fishing in a migratory species. *Evolutionary applications*, 2 (3), 371–93.

Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., and Chatain, B., 2008. Heritabilities and GxE interactions for growth in the European sea bass (Dicentrarchus labrax L.) using a marker-based pedigree. *Aquaculture*, 275 (1-4), 81–87.

Dwyer, W. P. and Piper, R. G., 1987. Atlantic Salmon Growth Efficiency as Affected by Temperature. *The Progressive Fish-Culturist*, 49 (1), 57–59.

East, E. M., 1908. Inbreeding in corn. *Reports of the Connecticut Agricultural Experiment Station for Years 1907–1908.*, (419-428).

Edeline, E., Carlson, S. M., Stige, L. C., Winfield, I. J., Fletcher, J. M., James, J. Ben, Haugen, T. O., Vøllestad, L. A., and Stenseth, N. C., 2007. Trait changes in a harvested population are driven by a dynamic tug-of-war between natural and harvest selection. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (40), 15799–804.

Edeline, E., Le Rouzic, A., Winfield, I. J., Fletcher, J. M., James, J. Ben, Stenseth, N. C., and Vøllestad, L. A., 2009. Harvest-induced disruptive selection increases variance in fitness-related traits. *Proceedings. Biological sciences / The Royal Society*, 276 (1676), 4163–71.

Edley, M. T. and Law, R., 1988. Evolution of life histories and yields in experimental populations of Daphnia magna. *Biol J Linn Soc*, 34 (4), 309–327.

Eickelberg, O., Centrella, M., Reiss, M., Kashgarian, M., and Wells, R. G., 2002. Betaglycan inhibits TGF-beta signaling by preventing type I-type II receptor complex formation. Glycosaminoglycan modifications alter betaglycan function. *The Journal of biological chemistry*, 277 (1), 823–9.

Eikeset, A. M., Richter, A., Dunlop, E. S., Dieckmann, U., and Stenseth, N. C., 2013. Economic repercussions of fisheries-induced evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (30), 12259–64.

El-Sabaawi, R., Bassar, R. D., Rakowski, C., Marshall, M. C., Bryan, B. L., Thomas, S. N., Pringle, C. M., Reznick, D. N., and Flecker, A. S., 2015. Intraspecific phenotypic differences in fish affect ecosystem processes as much as bottom-up factors. *Oikos*, 124 (9), EV1–EV11.

El-Sabaawi, R., Marshall, M. C., Bassar, R. D., López-Sepulcre, A., Palkovacs, E. P., and Dalton, C., 2015. Assessing the effects of guppy life history evolution on nutrient recycling: from experiments to the field. *Freshwater Biology*, 60 (3), 590–601.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, 6 (5), e19379.

Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., and Holzapfel, C. M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, 107 (37), 16196–16200.

Enberg, K. and Heino, M., 2007. Fisheries-induced life history changes in herring (Clupea harengus).

Enberg, K., Jørgensen, C., Dunlop, E. S., Heino, M., and Dieckmann, U., 2009. Implications of fisheries-induced evolution for stock rebuilding and recovery. *Evolutionary applications*, 2 (3), 394–414.

Enberg, K., Jørgensen, C., Dunlop, E. S., Varpe, Ø., Boukal, D. S., Baulier, L., Eliassen, S., and Heino, M., 2012. Fishing-induced evolution of growth: concepts, mechanisms and the empirical evidence. *Marine Ecology*, 33 (1), 1–25.

Endler, J., 1980. Natural selection on color patterns in Poecilia reticulata. *Evolution*, 34 (1), 76–91.

Endler, J., 1984. Natural and sexual selection on color patterns in poeciliid fishes. *Evolutionary ecology of neotropical freshwater fishes*, 9 (2), 173–190.

Endler, J., 1995. Multiple-trait coevolution and environmental gradients in guppies. *Trends in Ecology & Evolution*, 10 (1), 22–29.

Endler, J. A., 1988. Sexual selection and predation risk in guppies. *Nature*, 332 (6165), 593–594.

Endler, J. A., 1991. Interactions between predators and prey. *In*: *Behavioural Ecology: An evolutionary approach*. Oxford: Blackwell Scientific publications, 169–196.

Engelhard, G. H. and Heino, M., 2004. Maturity changes in Norwegian spring-spawning herring Clupea harengus: compensatory or evolutionary responses?

Erickson, D. L., Fenster, C. B., Stenøien, H. K., and Price, D., 2004. Quantitative trait locus analyses and the study of evolutionary process. *Molecular ecology*, 13 (9), 2505–22.

Ernst, S., Liu, K., Agarwala, S., Moratscheck, N., Avci, M. E., Dalle Nogare, D., Chitnis, A. B., Ronneberger, O., and Lecaudey, V., 2012. Shroom3 is required downstream of FGF signalling to mediate proneuromast assembly in zebrafish. *Development*, 139 (24), 4571–81.

Estève, P.-O., Chin, H. G., and Pradhan, S., 2005. Human maintenance DNA (cytosine-5)-methyltransferase and p53 modulate expression of p53-repressed promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (4), 1000–5.

Excoffier, L., Hofer, T., and Foll, M., 2009. Detecting loci under selection in a hierarchically structured population. *Heredity*, 103 (4), 285–98.

Excoffier, L. and Lischer, H. E. L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, 10 (3),

564–7.

Fajen, A. and Breden, F., 1992. Mitochondrial DNA sequence variation among natural populations of the Trinidad guppy, Poecilia reticulata. *Evolution*, 46 (5), 1457–1465.

Falconer, D. S., 1992. Early selection experiments. *Annual review of genetics*, 26, 1–14.

Falconer, D. S. and Mackay, T. F. C., 1966. *Introduction to Quantitative Genetics*. Benjamin Cummings.

Fang, X., Corrales, J., and Thornton, C., 2013. Global and gene specific DNA methylation changes during zebrafish development. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 166 (1), 99–108.

Farr, J., 1975. The role of predation in the evolution of social behavior of natural populations of the guppy, Poecilia reticulata (Pisces: Poeciliidae). *Evolution*, 29 (1), 151–158.

Felsenfeld, G., 2014. A brief history of epigenetics. *Cold Spring Harbor perspectives in biology*, 6 (1), a018200–.

Fenberg, P. B. and Roy, K., 2008. Ecological and evolutionary consequences of size-selective harvesting: how much do we know? *Molecular Ecology*.

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E., Ukomadu, C., Sadler, K. C., Pradhan, S., Pellegrini, M., and Jacobsen, S. E., 2010. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (19), 8689–8694.

Ferguson, L., Lee, S. F., Chamberlain, N., Nadeau, N., Joron, M., Baxter, S., Wilkinson, P., Papanicolaou, A., Kumar, S., Kee, T.-J., Clark, R., Davidson, C., Glithero, R., Beasley, H., Vogel, H., Ffrench-Constant, R., and Jiggins, C., 2010. Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Molecular ecology*, 19 (1), 240–54.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R., 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular biology and evolution*, 31 (5), 1275–91.

Festa-Bianchet, M., Pelletier, F., Jorgenson, J. T., Feder, C., and Hubbs, A., 2014. Decrease in horn size and increase in age of trophy sheep in Alberta over 37 years. *The Journal of Wildlife Management*, 78 (1), 133–141.

ffrench-Constant, R., 1996. Ecological and evolutionary aspects of insecticide resistance. *Trends in Ecology & Evolution*, 11 (9), 391–392.

Ffrench-Constant, R. H., 2013. The molecular genetics of insecticide resistance. *Genetics*, 194 (4), 807–15.

Fielder, D. S. and Bardsley, W., 1999. A Preliminary Study on the Effects of Salinity on Growth and Survival of Mulloway Argyrosomus japonicus Larvae and Juveniles. *Journal of the World Aquaculture Society*, 30 (3), 380–387.

Fisher, R. A. and Ford, E. B., 1947. The spread of a gene in natural conditions in a colony of the moth Panaxia dominula L. *Heredity*, 1 (2), 143–174.

Fitzpatrick, S. W., Gerberich, J. C., Kronenberger, J. a., Angeloni, L. M., and Funk, W. C., 2014. Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, 18 (1), 37–43.

Flint, J. and and Mackay, T. F. C., 2009. Genetic architecture of quantitative traits in flies, mice and humans. *Genome Res.*, 19, 723–733.

Foll, M. and Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180 (2), 977–93.

Foote, A. D., Liu, Y., Thomas, G. W. C., Vinař, T., Alföldi, J., Deng, J., Dugan, S., van Elk, C. E., Hunter, M. E., Joshi, V., Khan, Z., Kovar, C., Lee, S. L., Lindblad-Toh, K., Mancia, A., Nielsen, R., Qin, X., Qu, J., Raney, B. J., Vijay, N., Wolf, J. B. W., Hahn, M. W., Muzny, D. M., Worley, K. C., Gilbert, M. T. P., and Gibbs, R. A., 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics*, 47 (3), 272–275.

Ford, E., 1975. *Ecological genetics*. Springer.

Ford, E. B., 2005. R. A. Fisher: An Appreciation. *Genetics*, 171 (2), 415–417.

Fourcade, Y., Chaput-Bardy, A., Secondi, J., Fleurant, C., and Lemaire, C., 2013. Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular ecology*, 22 (8), 2065–73.

Franklin, T. and Mansuy, I., 2010. Epigenetic inheritance in mammals: evidence for the impact of adverse environmental effects. *Neurobiology of disease*, 39 (1), 61–65.

Franssen, N., 2011. Anthropogenic habitat alteration induces rapid morphological divergence in a native stream fish. *Evolutionary Applications*, 4 (6), 791–804.

Fraser, B. A., Künstner, A., Reznick, D. N., Dreyer, C., and Weigel, D., 2014. Population genomics of natural and experimental populations of guppies ( Poecilia reticulata ). *Molecular Ecology*, 24 (2), n/a–n/a.

Fraser, B. A., Weadick, C. J., Janowitz, I., Rodd, F. H., and Hughes, K. A., 2011. Sequencing and characterization of the guppy (Poecilia reticulata) transcriptome. *BMC genomics*, 12 (1), 202.

Fricke, C. and Chien, C.-B., 2005. Cloning of full-length zebrafish dcc and expression analysis during embryonic and early larval development. *Developmental dynamics : an official publication of the American Association of Anatomists*, 234 (3), 732–9.

Fryer, A., Emes, R., Ismail, K., and Haworth, K., 2011. Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics*, 6 (1), 86–94.

Fuller, R. C., Baer, C. F., and Travis, J., 2005. How and When Selection Experiments Might Actually be Useful. *Integrative and comparative biology*, 45 (3), 391–404.

Fulneček, J. and Kovařík, A., 2014. How to interpret methylation sensitive amplified polymorphism (MSAP) profiles? *BMC genetics*, 15 (1), 2.

Furness, A. I. and Reznick, D. N., 2014. The comparative ecology of a killifish (Rivulus hartii) across aquatic communities differing in predation intensity. *Evolutionary Ecology Research*, 16, 249–265.

Furness, A. I., Walsh, M. R., and Reznick, D. N., 2012. Convergence of life-history phenotypes in a trinidadian killifish (rivulus hartii). *Evolution*, 66 (4), 1240–1254.

Gagnaire, P.-A., Normandeau, E., Pavey, S. A., and Bernatchez, L., 2013. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (Coregonus clupeaformis). *Molecular Ecology*, 22 (11), 3036–3048.

Gahr, S. A., Vallejo, R. L., Weber, G. M., Shepherd, B. S., Silverstein, J. T., and Rexroad, C. E., 2008. Effects of short-term growth hormone treatment on liver and muscle transcriptomes in rainbow trout (Oncorhynchus mykiss). *Physiological genomics*, 32 (3), 380–92.

Galindo, J., Grahame, J. W., and Butlin, R. K., 2010. An EST-based genome scan using 454 sequencing in the marine snail Littorina saxatilis. *Journal of evolutionary biology*, 23 (9), 2004–16.

Gama, L. T., Dickerson, G. E., Young, L. D., and Leymaster, K. A., n.d. Effects of breed, heterosis, age of dam, litter size, and birth weight on lamb mortality. *Journal of animal science*, 69 (7), 2727–2743.

Gao, H., Fang, M., Liu, J., and Zhang, Q., 2009. Bayesian shrinkage mapping for multiple QTL in half-sib families. *Heredity*, 103 (5), 368–76.

Gao, L., Geng, Y., Li, B., Chen, J., and Yang, J., 2010. Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. *Plant, cell & environment*, 33 (11), 1820–7.

Garcia-Rudaz, C., Luna, F., Tapia, V., Kerr, B., Colgin, L., Galimi, F., Dissen, G. a, Rawlings, N. D., and Ojeda, S. R., 2007. Fxna, a novel gene differentially expressed in the rat ovary at the time of folliculogenesis, is required for normal ovarian histogenesis. *Development*, 134, 945–957.

Garel, M., Cugnasse, J.-M., Maillard, D., Gaillard, J.-M., Hewison, A. J. M., and Dubray, D., 2007.

Selective harvesting and habitat loss produce long-term life history changes in a mouflon population. *Ecological Applications*, 17 (6), 1607–1618.

Garland, T. and Kelly, S. A., 2006. Phenotypic plasticity and experimental evolution. *The Journal of experimental biology*, 209 (Pt 12), 2344–61.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M., and Estoup, A., 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular ecology*, 22 (11), 3165–78.

Gavery, M. R. and Roberts, S. B., 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (Crassostrea gigas). *BMC genomics*, 11, 483.

Ge, W., Davis, M. E., Hines, H. C., Irvin, K. M., and Simmen, R. C. M., 2003. Association of single nucleotide polymorphisms in the growth hormone and growth hormone receptor genes with blood serum insulin-like growth factor I concentration and growth traits in Angus cattle. *Journal of animal science*, 81 (3), 641–8.

Gentilini, D., Mari, D., Castaldi, D., Remondini, D., Ogliari, G., Ostan, R., Bucci, L., Sirchia, S. M., Tabano, S., Cavagnini, F., Monti, D., Franceschi, C., Di Blasio, A. M., and Vitale, G., 2013. Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age*, 35 (5), 1961–73.

George, A., Visscher, P., and Haley, C., 2000. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics*, 156 (4), 2081–2092.

Ghalambor, C., Reznick, D., and Walker, J., 2004. Constraints on adaptive evolution: the functional trade-off between reproduction and fast-start swimming performance in the Trinidadian guppy (Poecilia reticulata). *The American Naturalist*, 161 (1), 38–50.

Gilad, Y., Oshlack, A., and Rifkin, S. A., 2006. Natural selection on gene expression. *Trends in genetics : TIG*, 22 (8), 456–61.

Gjedrem, T., 1983. Genetic variation in quantitative traits and selective breeding in fish and shellfish. *Aquaculture*, 33 (1-4), 51–72.

Gjerde, B. and Refstie, T., 1984. Complete diallel cross between five strains of Atlantic salmon. *Livestock Production Science*, 11 (2), 207–226.

Goetz, F., Rosauer, D., Sitar, S., Goetz, G., Simchick, C., Roberts, S., Johnson, R., Murphy, C., Bronte, C. R., and Mackenzie, S., 2010. A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (Salvelinus namaycush). *Molecular ecology*, 19 Suppl 1, 176–96.

Goldish, D., 1996. *Captive breeding causes small body size in Morpho peleides limpida ( Nymphalidae : Morphinae )*.

Gómez-Requeni, P., Kraemer, M. N., and Canosa, L. F., 2012. Regulation of somatic growth and gene expression of the GH-IGF system and PRP-PACAP by dietary lipid level in early juveniles of a teleost fish, the pejerrey (Odontesthes bonariensis). *Journal of comparative physiology. B, Biochemical, systemic, and environmental physiology*, 182 (4), 517–30.

Gonen, S., Lowe, N. R., Cezard, T., Gharbi, K., Bishop, S. C., and Houston, R. D., 2014. Linkage maps of the Atlantic salmon (Salmo salar) genome derived from RAD sequencing. *BMC Genomics*, 15 (1), 166.

Gonzalez, A., Valeiras, M., Sidransky, E., and Tayebi, N., 2014. Lysosomal integral membrane protein-2: a new player in lysosome-related pathology. *Molecular genetics and metabolism*, 111 (2), 84–91.

Gordon, S., López-Sepulcre, A., and Reznick, D., 2012. Predation-associated differences in sex linkage of wild guppy coloration. *Evolution*, 66 (3), 912–918.

Graeb, B. D. S., Dettmers, J. M., Wahl, D. H., and Cáceres, C. E., 2004. Fish Size and Prey Availability Affect Growth, Survival, Prey Selection, and Foraging Behavior of Larval Yellow Perch. *Transactions of the American Fisheries Society*, 133 (3), 504–514.

Gratten, J., Beraldi, D., Lowder, B. V, McRae, A. F., Visscher, P. M., Pemberton, J. M., and Slate, J., 2007. Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proceedings. Biological*

*sciences / The Royal Society*, 274 (1610), 619–26.

Greasley, S. E., Horton, P., Ramcharan, J., Beardsley, G. P., Benkovic, S. J., and Wilson, I. A., 2001. letters Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis, 8 (5), 402–406.

Grether, G., Millie, D., and Bryant, M., 2001. Rain forest canopy cover, resource availability, and life history evolution in guppies. *Ecology*, 82 (6), 1546–1559.

Griffin, T. J. and Smith, L. M., 2000. MALDI – TOF mass spectrometry. *Science*, 18, 77–84.

Grift, R., Heino, M., and Rijnsdorp, A., 2007. Three-dimensional maturation reaction norms for North Sea plaice. *Marine Ecology Progress Series*, 334, 213–224.

Groh, S., González-de-León, D., Khairallah, M. M., Jiang, C., Bergvinson, D., Bohn, M., Hoisington, D. A., and Melchinger, A. E., 1998. QTL Mapping in Tropical Maize: III. Genomic Regions for Resistance to Diatraea spp. and Associated Traits in Two RIL Populations. *Crop Science*, 38 (4), 1062.

Gross, R. and Nilsson, J., 1999. Restriction fragment length polymorphism at the growth hormone1 gene in Atlantic salmon (Salmo salar L.) and its association with weight among the offspring of a hatchery stock. *Aquaculture*, 173.

Gross, W. L., Roelofs, E. W., and Fromm, P. O., 1965. Influence of Photoperiod on Growth of Green Sunfish, Lepomis cyanellus. *Journal of the Fisheries Research Board of Canada*, 22 (6), 1379–1386.

Grossniklaus, U., Kelly, B., Ferguson-Smith, A. C., Pembrey, M., and Lindquist, S., 2013. Transgenerational epigenetic inheritance: how important is it? *Nature Reviews Genetics*, 14 (3), 228–235.

Groszmann, M., Gonzalez-Bayon, R., Greaves, I. K., Wang, L., Huen, A. K., Peacock, W. J., and Dennis, E. S., 2014. Intraspecific Arabidopsis hybrids show different patterns of heterosis despite the close relatedness of the parental genomes. *Plant physiology*, 166 (September), 265–280.

Grover, C. E., Salmon, A., and Wendel, J. F., 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *American journal of botany*, 99 (2), 312–9.

Güimil, S., Chang, H.-S., Zhu, T., Sesma, A., Osbourn, A., Roux, C., Ioannidis, V., Oakeley, E. J., Docquier, M., Descombes, P., Briggs, S. P., and Paszkowski, U., 2005. Comparative transcriptomics of rice reveals an ancient pattern of response to microbial colonization. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (22), 8066–70.

Guo, B., DeFaveri, J., Sotelo, G., Nair, A., and Merilä, J., 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC biology*, 13 (1), 19.

Guo, J., Su, Y., Shin, J., Shin, J., and Li, H., 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature neuroscience*, 17, 215–222.

Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., and Davidson, W. S., 2015a. Genome-Wide Association Study (GWAS) for Growth Rate and Age at Sexual Maturation in Atlantic Salmon (Salmo salar). *Plos One*, 10 (3), e0119730.

Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., and Davidson, W. S., 2015b. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (Salmo salar). *PloS one*, 10 (3), e0119730.

Gwinn, D. C., Allen, M. S., Johnston, F. D., Brown, P., Todd, C. R., and Arlinghaus, R., 2015. Rethinking length-based fisheries regulations: the value of protecting old and large fish with harvest slots. *Fish and Fisheries*, 16 (2), 259–281.

Hahn, M. A., Wu, X., Li, A. X., Hahn, T., and Pfeifer, G. P., 2011. Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks. *PLoS ONE*, 6 (4), e18844.

Haldane, J. B. S., 1942. The selective elimination of silver foxes in Eastern Canada. *Journal of Genetics*, 44 (2-3), 296–304.

Hancock, A. M., Alkorta-Aranburu, G., Witonsky, D. B., and Di Rienzo, A., 2010. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365 (1552), 2459–68.

Handford, P., Bell, G., and Reimchen, T., 1977. A gillnet fishery considered as an experiment in artificial selection. *Journal of the Fisheries Research Board of Canada*, 34 (7), 954–961.

Harbison, S. T., McCoy, L. J., and Mackay, T. F. C., 2013. Genome-wide association study of sleep in Drosophila melanogaster. *BMC genomics*, 14 (1), 281.

Hard, J., Gross, M., and Heino, M., 2008. Evolutionary consequences of fishing and their implications for salmon. *Evolutionary …*, 1 (2), 388–408.

Hard, J. J., Gross, M. R., Heino, M., Hilborn, R., Kope, R. G., Law, R., and Reynolds, J. D., 2008. Evolutionary consequences of fishing and their implications for salmon. *Evolutionary applications*, 1 (2), 388–408.

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., Stapleton, A. E., Vallabhaneni, R., Williams, M., Wurtzel, E. T., Yan, J., and Buckler, E. S., 2008. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, 319 (5861), 330–3.

Harshman, L. G. and Hoffmann, A. A., 2000. Laboratory selection experiments using Drosophila: what do they really tell us? *Trends in Ecology & Evolution*, 15 (1), 32–36.

Haskins, C. and Haskins, E., 1949. The role of sexual selection as an isolating mechanism in three species of poeciliid fishes. *Evolution*, 3 (2), 160–169.

Haskins, C. and Haskins, E., 1951. The inheritance of certain color patterns in wild populations of Lebistes reticulatus in Trinidad. *Evolution*, 5 (3), 216–225.

Haskins, C., Haskins, E., McLaughlin, J., and Hewitt, R., 1961. Polymorphism and population structure in Lebistes reticulatus, an ecological study. *In*: *Vertebrate Speciation*. University of Texas Press.

Haskins, C. P., Young, P., Hewitt, R. E., and Haskins, E. F., 1970. Stabilised heterozygosis of supergenes mediating certain Y-linked colour patterns in populations of Lebistes Reticulatus. *Heredity*, 25 (4), 575–589.

Haugen, T. O. and Vøllestad, L. A., 2001. A century of life-history evolution in grayling. *Genetica*, 112-113, 475–91.

Hauser, L., Adcock, G. J., Smith, P. J., Ramiréz, J. H. B., and Carvalho, G. ., 2002. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). *Proceedings of the National Academy of Sciences of the United States of America*, 99 (18), 11742–7.

Head, J. a., 2014. Patterns of DNA methylation in animals: An ecotoxicological perspective. *Integrative and Comparative Biology*, 54 (1), 77–86.

Hecht, B. C., Campbell, N. R., Holecek, D. E., and Narum, S. R., 2013. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular ecology*, 22 (11), 3061–76.

Hedgecock, D., McGoldrick, D. J., and Bayne, B. L., 1995. Hybrid vigor in Pacific oysters: an experimental approach using crosses among inbred lines. *Aquaculture*, 137 (1-4), 285–298.

Heijmans, B. and Tobi, E., 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences*, 105 (44), 17046–17049.

Heiman, M. L., Chen, Y., and Caro, J. F., 1998. Leptin participates in the regulation of glucocorticoid and growth hormone axes11This paper was delivered at the 23–25 October 1997 conference 'The Determination, Treatment, and Prevention of Obesity,' which was sponsored by the Institute of Nutrition, Un. *The Journal of Nutritional Biochemistry*, 9 (10), 553–559.

Heino, M., Baulier, L., Boukal, D. S., Ernande, B., Johnston, F. D., Mollet, F. M., Pardoe, H., Therkildsen, N. O., Uusi-Heikkila, S., Vainikka, A., Arlinghaus, R., Dankel, D. J., Dunlop, E. S., Eikeset, A. M., Enberg, K., Engelhard, G. H., Jorgensen, C., Laugen, A. T., Matsumura, S., Nussle, S., Urbach, D., Whitlock, R., Rijnsdorp, A. D., and Dieckmann, U., 2013. Can

fisheries-induced evolution shift reference points for fisheries management? *ICES Journal of Marine Science*, 70 (4), 707–721.

Heino, M., Díaz Pauli, B., and Dieckmann, U., 2015. Fisheries-Induced Evolution. *Annual Review of Ecology, Evolution, and Systematics*, 46 (1), annurev–ecolsys–112414–054339.

Heino, M. and Dieckmann, U., 2008. Detecting fisheries-induced life-history evolution: an overview of the reaction-norm approach. *Bulletin of Marine Science*, 83 (1), 69–93.

Heino, M., Dieckmann, U., and Godo, D., 2002. Estimating reaction norms for age and size at maturation with reconstructed immature size distributions: a new technique illustrated by application to Northeast Arctic cod. *ICES Journal of Marine Science*, 59 (3), 562–575.

Heino, M. and Godø, O., 2002. Fisheries-induced selection pressures in the context of sustainable fisheries. *Bulletin of Marine Science*, 70 (2), 639–656.

Hellman, A. and Chess, A., 2007. Gene body-specific methylation on the active X chromosome. *Science*, 315 (5815), 1141–3.

Hempel, G. and Blaxter, J. H. S., 1961. The Experimental Modification of Meristic Characters in Herring (Clupea harengus L.). *ICES Journal of Marine Science*, 26 (3), 336–346.

Hendry, A., Kinnison, M., and Heino, M., 2011. Evolutionary principles and their practical application. *Evolutionary principles and their practical application*, 4 (2), 159–183.

Hendry, A. P., Farrugia, T. J., and Kinnison, M. T., 2008. Human influences on rates of phenotypic change in wild animal populations. *Molecular ecology*, 17 (1), 20–9.

Hermisson, J. and Pennings, P. S., 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169 (4), 2335–52.

Herrera, C. and Bazaga, P., 2010. Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. *New Phytologist*, 187 (3), 867–876.

Hess, J. E., Caudill, C. C., Keefer, M. L., McIlraith, B. J., Moser, M. L., and Narum, S. R., 2014. Genes predict long distance migration and large body size in a migratory fish, Pacific lamprey. *Evolutionary Applications*, 7, 1192–1208.

Heyn, H., Moran, S., Hernando-Herraez, I., Sayols, S., Gomez, A., Sandoval, J., Monk, D., Hata, K., Marques-Bonet, T., Wang, L., and Esteller, M., 2013. DNA methylation contributes to natural human variation. *Genome research*, 23 (9), 1363–72.

Hiddink, J. G., Jennings, S., Kaiser, M. J., Queirós, A. M., Duplisea, D. E., and Piet, G. J., 2006. Cumulative impacts of seabed trawl disturbance on benthic biomass, production, and species richness in different habitats. *Canadian Journal of Fisheries and Aquatic Sciences*, 63 (4), 721–736.

Hiendleder, S., 2003. Mapping of QTL for Body Conformation and Behavior in Cattle. *Journal of Heredity*, 94 (6), 496–506.

Hilborn, R., 2006. Faith-based fisheries. *Fisheries*.

Hilborn, R., 2007. Moving to sustainability by learning from successful fisheries. *AMBIO: A Journal of the Human Environment*, 36 (4), 296–303.

Hilborn, R., Enberg, K., Jørgensen, C., and Mangel, M., 2010. Fishing-induced evolution and changing reproductive ecology of fish: the evolution of steepness. *Canadian Journal of Fisheries and Aquatic Sciences*, 67 (10), 1708–1719.

Hinds, D., Stokowski, R., and Patil, N., 2004. Matching strategies for genetic association studies in structured populations. *The American Journal of Human Genetics*, 74 (2), 317–325.

Hipp, A. L., Eaton, D. A. R., Cavender-Bares, J., Fitzek, E., Nipper, R., and Manos, P. S., 2014. A Framework Phylogeny of the American Oak Clade Based on Sequenced RAD Data. *PLoS ONE*, 9 (4), e93975.

Hirschhorn, J. and Daly, M., 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6 (95), 108.

Hoarau, G., Boon, E., Jongma, D. N., Ferber, S., Palsson, J., Van der Veer, H. W., Rijnsdorp, A. D.,

Stam, W. T., and Olsen, J. L., 2005. Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (Pleuronectes platessa L.). *Proceedings. Biological sciences / The Royal Society*, 272 (1562), 497–503.

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., and Crossland, J. P., 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313 (5783), 101–4.

Hoffmann, A. A. and Willi, Y., 2008. Detecting genetic responses to environmental change. *Nature reviews. Genetics*, 9 (6), 421–32.

Hoffmann, A. and Merilä, J., 1999. Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution*, 14 (3), 96–101.

Hohenlohe, P. A., Bassham, S., Currey, M., and Cresko, W. A., 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367 (1587), 395–408.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics*, 6 (2), e1000862.

Holliday, J. A., Wang, T., and Aitken, S., 2012. Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (Picea sitchensis) using random forest. *G3 (Bethesda, Md.)*, 2 (9), 1085–93.

Holloway, G., Povey, S., and Sibly, R., 1990. The effect of new environment on adapted genetic architecture. *Heredity*, 64, 323–330.

Honan, P., 2008. Notes on the biology, captive management and conservation status of the Lord Howe Island Stick Insect (Dryococelus australis) (Phasmatodea). *Journal of Insect Conservation*, 12 (3-4), 399–413.

Hornoy, B., Pavy, N., Gérardi, S., Beaulieu, J., and Bousquet, J., 2015. Genetic Adaptation to Climate in White Spruce Involves Small to Moderate Allele Frequency Shifts in Functionally Diverse Genes. *Genome biology and evolution*, 7 (12), 3269–85.

Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome biology*, 14 (10), R115.

Houde, A. E., 1988. Genetic difference in female choice between two guppy populations. *Animal Behaviour*, 36 (2), 510–516.

Houde, A. E., 1997. *Sex, Color, and Mate Choice in Guppies*. Princeton University Press.

Houston, R. D., Davey, J. W., Bishop, S. C., Lowe, N. R., Mota-Velasco, J. C., Hamilton, A., Guy, D. R., Tinch, A. E., Thomson, M. L., Blaxter, M. L., Gharbi, K., Bron, J. E., and Taggart, J. B., 2012. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics*, 13 (1), 244.

Hsia, H. C. and Schwarzbauer, J. E., 2005. Meet the tenascins: Multifunctional and mysterious. *Journal of Biological Chemistry*, 280 (4), 26641–26644.

Hu, Y., Morota, G., Rosa, G. J. M., and Gianola, D., 2015. Prediction of Plant Height in Arabidopsis thaliana Using DNA Methylation Data. *Genetics*, 201 (2), 779–93.

Huang, K. and Fan, G., 2010. DNA methylation in cell differentiation and reprogramming: an emerging systematic view. *Regenerative medicine*, 5 (4), 531–44.

Huang, N., Cogburn, L. A., Agarwal, S. K., Marks, H. L., and Burnside, J., 1993. Overexpression of a truncated growth hormone receptor in the sex-linked dwarf chicken: evidence for a splice mutation. *Molecular Endocrinology*, 7 (11), 1391–1398.

Huang, X., Paulo, M.-J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., and van Eeuwijk, F. A., 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (11), 4488–93.

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T.,

McMillan, W. O., and Joron, M., 2015. Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. *Heredity*, 114 (5), 515–524.

Huey, R. B. and Rosenzweig, F., 2009. Laboratory evolution meets catch-22. *In*: Garland, T. and Rose, M., eds. *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*. 671–702.

Hutchings, J., 2000. Collapse and recovery of marine fishes. *Nature*, 406, 882–885.

Hutchings, J. A., 2005. Life history consequences of overexploitation to population recovery in Northwest Atlantic cod ( Gadus morhua ). *Canadian Journal of Fisheries and Aquatic Sciences*, 62 (4), 824–832.

Hutchings, J. A., 2009. Avoidance of fisheries-induced evolution: management implications for catch selectivity and limit reference points. *Evolutionary applications*, 2 (3), 324–34.

Hutchings, J. A. and Fraser, D. J., 2008. The nature of fisheries- and farming-induced evolution. *Molecular ecology*, 17 (1), 294–313.

Hutchings, J. A. and Rangeley, R. W., 2011. Correlates of recovery for Canadian Atlantic cod ( Gadus morhua ) 1 1 This review is part of the virtual symposium 'Flagship Species – Flagship Problems' that deals with ecology, biodiversity and management issues, and climate impacts on species at risk a. *Canadian Journal of Zoology*, 89 (5), 386–400.

Hutchinson, W. F., van Oosterhout, C., Rogers, S. I., and Carvalho, G. R., 2003. Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (Gadus morhua). *Proceedings. Biological sciences / The Royal Society*, 270 (1529), 2125–32.

Hyma, K. E. and Fay, J. C., 2013. Mixing of vineyard and oak-tree ecotypes of Saccharomyces cerevisiae in North American vineyards. *Molecular Ecology*, 22 (11), 2917–2930.

Ikeda, Y. and Nishimura, T., 2015. The Role of DNA Methylation in Transposable Element Silencing and Genomic Imprinting. *In*: *Nuclear Functions in Plant Transcription, Signaling and development*. Springer, 13–29.

Illingworth, R. S. and Bird, A. P., 2009. CpG islands--'a rough guide'. *FEBS letters*, 583 (11), 1713–20.

Jabbari, K., Cacciò, S., Païs De Barros, J. P., Desgrès, J., and Bernardi, G., 1997. Evolutionary changes in CpG and methylation levels in the genome of vertebrates. *Gene*, 205 (1-2), 109–118.

Jachmann, H., Berry, P., and Imae, H., 1995. Tusklessness in African elephants: a future trend. *African Journal of Ecology*, 33 (3), 230–235.

Jacobsson, L., Park, H.-B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, Paul, B. ., and Anderson, L., 2005. Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens. *Genetical Research*, 86 (02), 115.

Jaenisch, R. and Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33, 245–254.

Jain, K. and Stephan, W., 2015. Response of Polygenic Traits Under Stabilizing Selection and Mutation when Loci Have Unequal Effects. *G3*, 5 (6), 1065–1074.

Jakobsdóttir, K. B., Pardoe, H., Magnússon, Á., Björnsson, H., Pampoulie, C., Ruzzante, D. E., and Marteinsdóttir, G., 2011. Historical changes in genotypic frequencies at the Pantophysin locus in Atlantic cod (Gadus morhua) in Icelandic waters: evidence of fisheries-induced selection? *Evolutionary applications*, 4 (4), 562–73.

Jensen, J. D., 2014. On the unfounded enthusiasm for soft selective sweeps. *Nature communications*, 5, 5281.

Johannes, F., Porcher, E., Teixeira, F. K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F., and Colot, V., 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS genetics*, 5 (6), e1000530.

Johansson, A., Pettersson, M., Siegel, P., and Carlborg, Ö., 2010. Genome-wide effects of long-

term divergent selection. *PLoS genetics*, 6 (11), e1001188.

Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., Pemberton, J. M., and Slate, J., 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular ecology*, 20 (12), 2555–66.

Jones, D. F., 1917. Dominance of Linked Factors as a Means of Accounting for Heterosis. *Proceedings of the National Academy of Sciences of the United States of America*, 3 (4), 310–2.

Jones, G. P., 1986. Food availability affects growth in a coral reef fish. *Oecologia*, 70 (1), 136–139.

Jones, H., Leigh, F. J., Mackay, I., Bower, M. a., Smith, L. M. J., Charles, M. P., Jones, G., Jones, M. K., Brown, T. a., and Powell, W., 2008. Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the fertile crescent. *Molecular Biology and Evolution*, 25 (10), 2211–2219.

Jones, J. C., Fan, S., Franchini, P., Schartl, M., and Meyer, A., 2013. The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular ecology*, 22 (11), 2986–3001.

Jones, M. J., Goodman, S. J., and Kobor, M. S., 2015. DNA methylation and healthy human aging. *Aging Cell*, 14 (6), 924–932.

Jones, P. a, 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews. Genetics*, 13 (7), 484–92.

Jones, P. a and Takai, D., 2001. The role of DNA methylation in mammalian epigenetics. *Science (New York, N.Y.)*, 293 (5532), 1068–1070.

Jorgensen, C., Enberg, K., and Dunlop, E., 2007. Managing evolving fish stocks. *Science*, 318 (5854), 1247–1248.

Jørgensen, C. and Holt, R. E., 2013. Natural mortality: Its ecology, how it shapes fish life histories, and why it may be increased by fishing. *Journal of Sea Research*, 75, 8–18.

Joulia-Ekaza, D. and Cabello, G., 2007. The myostatin gene: physiology and pharmacological relevance. *Current opinion in pharmacology*, 7 (3), 310–5.

Jung, M., Ching, A., Bhattramakki, D., Dolan, M., Tingey, S., Morgante, M., and Rafalski, A., 2004. Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 109 (4), 681–9.

Jung, M. and Pfeifer, G. P., 2015. Aging and DNA methylation. *BMC biology*, 13 (1), 7.

Kaikkonen, M. U., Lam, M. T. Y., and Glass, C. K., 2011. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular research*, 90 (3), 430–40.

Kaiya, H., Kojima, M., Hosoda, H., Riley, L. G., Hirano, T., Grau, E. G., and Kangawa, K., 2003. Identification of tilapia ghrelin and its effects on growth hormone and prolactin release in the tilapia, Oreochromis mossambicus. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 135 (3), 421–429.

Kakioka, R., Kokita, T., Kumada, H., Watanabe, K., and Okuda, N., 2013. A RAD-based linkage map and comparative genomics in the gudgeons (genus Gnathopogon, Cyprinidae). *BMC Genomics*, 14 (1), 32.

Kang, H., Zaitlen, N., and Wade, C., 2008. Efficient control of population structure in model organism association mapping. *Genetics*, 178 (3), 1709–1723.

Kang, J.-H., Lee, S.-J., Park, S.-R., and Ryu, H.-Y., 2002. DNA polymorphism in the growth hormone gene and its association with weight in olive flounder Paralichthys olivaceus. *Fisheries Science*, 68 (3), 494–498.

Kaplan, N. L., Hudson, R. R., and Langley, C. H., 1989. The 'hitchhiking effect' revisited. *Genetics*, 123 (4), 887–899.

Karasov, T., Messer, P. W., and Petrov, D. A., 2010. Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS genetics*, 6 (6), e1000924.

Karim, N., Gordon, S. P., Schwartz, A. K., and Hendry, A. P., 2007. This is not déjà vu all over again: male guppy colour in a new experimental introduction. *Journal of evolutionary biology*, 20 (4), 1339–50.

Kause, A., Paananen, T., Ritola, O., and Koskinen, H., 2007. Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *Journal of animal science*, 85 (12), 3218–27.

Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M. C., 2012. Experimental evolution. *Trends in ecology & evolution*, 27 (10), 547–60.

Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., Wittwer, S., and Seehausen, O., 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, 22 (11), 2848–2863.

Kelley, J. and Magurran, A., 2003. Effects of relaxed predation pressure on visual predator recognition in the guppy. *Behavioral Ecology and Sociobiology*, 54 (3), 225–232.

Kenchington, E., 2003. The Effects of Fishing on Species and Genetic Diversity. *In*: *Responsible fisheries in the marine ecosystem*. CABI publishing, 235–254.

Kendall, N. W., Dieckmann, U., Heino, M., Punt, A. E., and Quinn, T. P., 2014. Evolution of age and length at maturation of Alaskan salmon under size-selective harvest. *Evolutionary applications*, 7 (2), 313–22.

Kettlewell, H. B. D., 1958. A survey of the frequencies of Biston betularia (L.) (Lep.) and its melanic forms in Great Britain. *Heredity*, 12 (1), 51–72.

Keyte, A., Percifield, R., Liu, B., and Wendel, J., 2006. Infraspecific DNA methylation polymorphism in cotton (Gossypium hirsutum L.). *Journal of Heredity*, 97 (5), 444–450.

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., and Cooper, T. F., 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332 (6034), 1193–6.

Khater, M., Murariu, D., and Gras, R., 2014. Contemporary evolution and genetic change of prey as a response to predator removal. *Ecological Informatics*, 22, 13–22.

Khavari, D. A., Sen, G. L., and Rinn, J. L., 2010. DNA methylation and epigenetic control of cellular differentiation. *Cell cycle (Georgetown, Tex.)*, 9 (19), 3880–3.

Khoo, G., Lim, M. H., Suresh, H., Gan, D. K. Y., Lim, K. F., Chen, F., Chan, W.-K., Lim, T. M., and Phang, V. P. E., 2003. Genetic linkage maps of the guppy ( Poecilia reticulata): assignment of RAPD markers to multipoint linkage groups. *Marine biotechnology*, 5 (3), 279–93.

Kiialainen, A., Karlberg, O., Ahlford, A., Sigurdsson, S., Lindblad-Toh, K., and Syvänen, A.-C., 2011. Performance of microarray and liquid based capture methods for target enrichment for massively parallel sequencing and SNP discovery. *PloS one*, 6 (2), e16486.

Kilvitis, H. J., Alvarez, M., Foust, C. M., Schrey, A. W., Robertson, M., and Richards, C. L., 2014. Ecological Genomics. *In*: *Ecological genomics: Ecology and the evolution of genes and genomes*. Springer, 191–210.

Kim, Y. and Stephan, W., 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. *Genetics*, 160 (2), 765–777.

Kinnison, M. and Hendry, A., 2001. The pace of modern life II: from rates of contemporary microevolution to pattern and process. *In*: *Microevolution Rate, Pattern, Process*. Springer, 145–164.

Klose, R. and Bird, A., 2006. Genomic DNA methylation: the mark and its mediators. *Trends in biochemical sciences*, 31 (2), 89–97.

Koenig, D., Jiménez-Gómez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., Kumar, R., Covington, M. F., Devisetty, U. K., Tat, A. V, Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeples, M., Brady, S. M., Pauly, M., Weigel, D., Usadel, B., Fernie, A. R., Peng, J., Sinha, N. R., and Maloof, J. N., 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (28), E2655–62.

Kolss, M., Vijendravarma, R. K., Schwaller, G., and Kawecki, T. J., 2009. Life-history consequences of adaptation to larval nutritional stress in Drosophila. *Evolution; international journal of organic evolution*, 63 (9), 2389–401.

Kondo, M., Yamaoka, T., Honda, S., Miwa, Y., and Katashima, R., 2000. The Rate of Cell Growth Is Regulated by Purine Biosynthesis via ATP Production and G x to S Phase Transition 1, 128 (1), 57–64.

Kotrschal, A., Corral-Lopez, A., Zajitschek, S., Immler, S., Maklakov, A. A., and Kolm, N., 2015. Positive genetic correlation between brain size and sexual traits in male guppies artificially selected for brain size. *Journal of evolutionary biology*, 28 (4), 841–50.

Kovach, W. ., 2007. A MultiVariate Statistical Package for Windows, ver. 3.1. *Kovach Computing Services, Pentraeth, Wales, U.K.*

Kraak, S., 2007. Does the probabilistic maturation reaction norm approach disentangle phenotypic plasticity from genetic change? *Marine Ecology Progress Series*, 335, 295–300.

Ku, C. S., Naidoo, N., Wu, M., and Soong, R., 2011. Studying the epigenome using next generation sequencing. *Journal of Medical Genetics*, 48 (11), 721–730.

Kulis, M., Queirós, A. C., Beekman, R., and Martín-Subero, J. I., 2013. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochimica et biophysica acta*, 1829 (11), 1161–74.

Kuparinen, A. and Hutchings, J. A., 2012. Consequences of fisheries-induced evolution for population productivity and recovery potential. *Proceedings. Biological sciences / The Royal Society*, 279 (1738), 2571–9.

Kuparinen, A., Kuikka, S., and Merilä, J., 2009. Estimating fisheries-induced selection: traditional gear selectivity research meets fisheries-induced evolution. *Evolutionary Applications*, 2 (2), 234–243.

Kuparinen, A. and Merilä, J., 2007. Detecting and managing fisheries-induced evolution. *Trends in Ecology & Evolution*, 22 (12), 652–659.

Labra, M., Ghiani, A., Citterio, S., and Sgorbati, S., 2002. Analysis of cytosine methylation pattern in response to water deficit in pea root tips. *Plant Biology*, 4, 694–699.

Lacy, R. C. and Horner, B. E., 2012. Effects of inbreeding on skeletal development of Rattus villosissimus. *The Journal of heredity*, 87 (4), 277–87.

Landauer, W. and Upham, E., 1936. Weight and size of organs in frizzle fowl. *Storrs Agricultural Experiment Station Bulletin*, 210.

Lander, E. and Schork, N., 1994. Genetic dissection of complex traits. *Science*, 265, 2037–2048.

Langdon, Y. G. and Mullins, M. C., 2011. Maternal and zygotic control of zebrafish dorsoventral axial patterning. *Annual review of genetics*, 45, 357–77.

Langerhans, R. and DeWitt, T., 2004. Shared and unique features of evolutionary diversification. *The American Naturalist*, 164 (3), 335–349.

Laporte, M., Pavey, S. A., Rougeux, C., Pierron, F., Lauzent, M., Budzinski, H., Labadie, P., Geneste, E., Couture, P., Baudrimont, M., and Bernatchez, L., 2016. RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular ecology*, 25 (1), 219–37.

Laporte, M., Rogers, S. M., Dion-Côté, A.-M., Normandeau, E., Gagnaire, P.-A., Dalziel, A. C., Chebib, J., and Bernatchez, L., 2015. RAD-QTL Mapping Reveals Both Genome-Level Parallelism and Different Genetic Architecture Underlying the Evolution of Body Shape in Lake Whitefish (Coregonus clupeaformis) Species Pairs. *G3 (Bethesda, Md.)*, 5 (7), 1481–91.

Laugen, A. T., Engelhard, G. H., Whitlock, R., Arlinghaus, R., Dankel, D. J., Dunlop, E. S., Eikeset, A. M., Enberg, K., Jørgensen, C., Matsumura, S., Nusslé, S., Urbach, D., Baulier, L., Boukal, D. S., Ernande, B., Johnston, F. D., Mollet, F., Pardoe, H., Therkildsen, N. O., Uusi-Heikkilä, S., Vainikka, A., Heino, M., Rijnsdorp, A. D., and Dieckmann, U., 2014. Evolutionary impact assessment: accounting for evolutionary consequences of fishing in an ecosystem approach to fisheries management. *Fish and fisheries*, 15 (1), 65–96.

Laurent, L., Wong, E., Li, G., and Huynh, T., 2010. Dynamic changes in the human methylome during differentiation. *Genome Research*, 20, 320–331.

Law, R., 2007. Fisheries-induced evolution: present status and future directions. *Marine Ecology Progress Series*, 335, 295–300.

Leaché, A. D., Fujita, M. K., Minin, V. N., and Bouckaert, R. R., 2014. Species delimitation using genome-wide SNP data. *Systematic biology*, 63 (4), 534–42.

Lee, J., Izzah, N. K., Jayakodi, M., Perumal, S., Joh, H. J., Lee, H. J., Lee, S.-C., Park, J. Y., Yang, K.-W., Nou, I.-S., Seo, J., Yoo, J., Suh, Y., Ahn, K., Lee, J. H., Choi, G. J., Yu, Y., Kim, H., and Yang, T.-J., 2015. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. *BMC plant biology*, 15 (1), 32.

Lee, T. -f. T.-F., Zhai, J., and Meyers, B. C., 2010. Conservation and divergence in eukaryotic DNA methylation. *Proceedings of the National Academy of Sciences*, 107 (20), 9027–9028.

Lelandais, G., Tanty, V., Geneix, C., Etchebest, C., Jacq, C., and Devaux, F., 2008. Genome adaptation to chemical stress: clues from comparative transcriptomics in Saccharomyces cerevisiae and Candida glabrata. *Genome biology*, 9 (11), R164.

Lepais, O. and Weir, J. T., 2014. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources*, 14 (6), 1314–1321.

Lewontin, R. C. and Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, 74 (1), 175–95.

Li, E., Beard, C., and Jaenisch, R., 1993. Role for DNA methylation in genomic imprinting. *Nature*, 366 (6453), 362–5.

Li, Y., Shan, X., Liu, X., Hu, L., Guo, W., and Liu, B., 2008. Utility of the methylation-sensitive amplified polymorphism (MSAP) marker for detection of DNA methylation polymorphism and epigenetic population structure in a wild barley species (Hordeum brevisubulatum). *Ecological Research*, 23 (5), 927–930.

Liebl, A. L., Schrey, A. W., Andrew, S. C., Sheldon, E. L., and Griffith, S. C., 2015. Invasion genetics: lessons from a ubiquitous bird, the house sparrow (Passer domesticus). *Current Zoology*, 61 (3), 30.

Liebl, A. L., Schrey, A. W., Richards, C. L., and Martin, L. B., 2013. Patterns of DNA methylation throughout a range expansion of an introduced songbird. *Integrative and comparative biology*, 53 (2), 351–8.

Liley, N. R. and Luyten, P. H., 1985. Geographic Variation in the Sexual Behaviour of the Guppy, Poecilia Reticulata (Peters). *Behaviour*, 95 (1), 164–179.

Lira-Medeiros, C. F., Parisod, C., Fernandes, R. A. R., Mata, C. S., Cardoso, M. A., and Ferreira, P. C. G., 2010. Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS One*, 5 (4), e10326.

Lira-Medeiros, C., Parisod, C., and Fernandes, R., 2010. Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS One*.

Lisachov, A. P., Zadesenets, K. S., Rubtsov, N. B., and Borodin, P. M., 2015. Sex Chromosome Synapsis and Recombination in Male Guppies. *Zebrafish*, 00 (00), 1–7.

Lister, R. and Ecker, J. R., 2009. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome research*, 19 (6), 959–66.

Lister, R., Mukamel, E., Nery, J., and Urich, M., 2013. Global epigenomic reconfiguration during mammalian brain development. *Science*, 341 (6146), 1237905.

Lister, R., O'Malley, R., and Tonti-Filippini, J., 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, 133 (3), 523–536.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M., 2012. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology*, 2012, 251364.

Liu, S., Sun, K., Jiang, T., and Feng, J., 2015. Natural epigenetic variation in bats and its role in evolution. *The Journal of experimental biology*, 218 (Pt 1), 100–106.

Liu, S., Vallejo, R. L., Palti, Y., Gao, G., Marancik, D. P., Hernandez, A. G., and Wiens, G. D., 2015. Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Frontiers in genetics*, 6, 298.

Liu, Y.-Z., Xu, F.-H., Shen, H., Liu, Y.-J., Zhao, L.-J., Long, J.-R., Zhang, Y.-Y., Xiao, P., Xiong, D.-H., Dvornyk, V., Li, J.-L., Conway, T., Davies, K. M., Recker, R. R., and Deng, H.-W., 2004. Genetic dissection of human stature in a large sample of multiplex pedigrees. *Annals of human genetics*, 68 (Pt 5), 472–88.

Llaurens, V., McMullan, M., and van Oosterhout, C., 2012. Cryptic MHC polymorphism revealed but not explained by selection on the class IIb peptide-binding region. *Molecular biology and evolution*, 29 (6), 1631–44.

Lomniczi, A., Loche, A., Castellano, J. M., Ronnekleiv, O. K., Bosch, M., Kaidar, G., Knoll, J. G., Wright, H., Pfeifer, G. P., and Ojeda, S. R., 2013. Epigenetic control of female puberty. *Nature neuroscience*, 16 (3), 281–9.

Long, A. and Langley, C., 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research*, 9 (8), 720–731.

Long, Y., Xia, W., Li, R., Wang, J., and Shao, M., 2011. Epigenetic QTL mapping in Brassica napus. *Genetics*, 189 (3), 1093–1102.

Lorenzen, K. and Enberg, K., 2002. Density-dependent growth as a key mechanism in the regulation of fish populations: evidence from among-population comparisons. *Proceedings. Biological sciences / The Royal Society*, 269 (1486), 49–54.

Lory, P., Bidaud, I., and Chemin, J., 2006. T-type calcium channels in differentiation and proliferation. *Cell calcium*, 40 (2), 135–46.

Losos, J., 2009. *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles*. University of California Press.

Lotterhos, K. E. and Whitlock, M. C., 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular ecology*, 23 (9), 2178–92.

Lü, A., Hu, X., Chen, H., Dong, Y., and Pang, Y., 2011. Single nucleotide polymorphisms of the prolactin receptor (PRLR) gene and its association with growth traits in Chinese cattle. *Molecular biology reports*, 38 (1), 261–6.

Lu, X., Le Noble, F., Yuan, L., Jiang, Q., De Lafarge, B., Sugiyama, D., Bréant, C., Claes, F., De Smet, F., Thomas, J.-L., Autiero, M., Carmeliet, P., Tessier-Lavigne, M., and Eichmann, A., 2004. The netrin receptor UNC5B mediates guidance events controlling morphogenesis of the vascular system. *Nature*, 432 (7014), 179–86.

Lucifero, D., Mertineit, C., Clarke, H. J., Bestor, T. H., and Trasler, J. M., 2002. Methylation dynamics of imprinted genes in mouse germ cells. *Genomics*, 79 (4), 530–8.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R., 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS biology*, 8 (11), e1000506.

Lynch, C., 1992. Clinal variation in cold adaptation in Mus domesticus: verification of predictions from laboratory populations. *American Naturalist*, 139 (6), 1219–1236.

Lynch, C. B., 1980. Response to divergent selection for nesting behavior in Mus musculus. *Genetics*, 96 (3), 757–65.

Lynch, M. and Walsh, B., 1998. Genetics and analysis of quantitative traits.

Ma, Q., Liu, S. F., Zhuang, Z. M., Lin, L., Sun, Z. Z., Liu, C. L., Su, Y. Q., and Tang, Q. S., 2011. Genomic structure, polymorphism and expression analysis of growth hormone-releasing hormone and pituitary adenylate cyclase activating polypeptide genes in the half-smooth tongue sole (Cynoglossus semilaevis). *Genetics and molecular research : GMR*, 10 (4), 3828–46.

Machado, M. B. B., Alencar, M. M., Pereira, A. P., Oliveira, H. N., Casas, E., Coutinho, L. L., and Regitano, L. C. A., 2003. QTL affecting body weight in a candidate region of cattle chromosome 5. *Genetics and Molecular Biology*, 26 (3), 259–265.

MacKay, A. B., Mhanni, A. A., McGowan, R. A., and Krone, P. H., 2007. Immunological detection of changes in genomic DNA methylation during early zebrafish development. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada*, 50 (8), 778–85.

Mackay, T. F. C., 2009. Q&A: Genetic analysis of quantitative traits. *Journal of biology*, 8 (3), 23.

Mackay, T. F. C., Stone, E. a, and Ayroles, J. F., 2009. The genetics of quantitative traits: challenges and prospects. *Nature reviews. Genetics*, 10 (8), 565–77.

MacLean, D., Jones, J. D. G., and Studholme, D. J., 2009. In the news. *Nature Reviews Microbiology*, 7 (2), 96–97.

Maguire, T. L., Peakall, R., and Saenger, P., 2002. Comparative analysis of genetic diversity in the mangrove species Avicennia marina (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. *Theoretical and Applied Genetics*, 104 (2-3), 388–398.

Magurran, A., 2005. *Evolutionary ecology: the Trinidadian guppy*. Oxford University Press.

Magurran, A. E. A. E., Seghers, B. H., Carvalho, G. R., and Shaw, P. W., 1992. Behavioural Consequences of an Artificial Introduction of Guppies (Poecilia reticulata) in N. Trinidad: Evidence for the Evolution of Anti-Predator Behaviour in the Wild. *Proceedings of the Royal Society B: Biological Sciences*, 248 (1322), 117–122.

Magurran, A. E. and Seghers, B. H., 1991. Variation in Schooling and Aggression Amongst Guppy (Poecilia Reticulata) Populations in Trinidad. *Behaviour*, 118 (3), 214–234.

Magurran, A. E. and Seghers, B. H., 1994. Predator Inspection Behaviour Covaries With Schooling Tendency Amongst Wild Guppy, Poecilia Reticulata, Populations in Trinidad. *Behaviour*, 128 (1), 121–134.

Magurran, A. and Phillip, D., 2001. Evolutionary implications of large-scale patterns in the ecology of Trinidadian guppies, Poecilia reticulata. *Biological Journal of the Linnean Society*, 73 (1), 1–9.

Magurran, A. and Seghers, B., 1990. Population differences in predator recognition and attack cone avoidance in the guppy Poecilia reticulata. *Animal Behaviour*, 40 (3), 443–452.

Magurran, A., Seghers, B. H., Shaw, P. W., and Carvalho, G. R., 1995. The behavioral diversity and evolution of guppy, Poecilia reticulata, populations in Trinidad. *In*: *Advances in the Study of Behaviour*. Academic press, 155–202.

Malosetti, M., van der Linden, C. G., Vosman, B., and van Eeuwijk, F. A., 2007. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to Phytophthora infestans in potato. *Genetics*, 175 (2), 879–89.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., Giovannoni, J. J., and Seymour, G. B., 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature Genetics*, 38 (8), 948–952.

Manzon, L. A., 2002. The role of prolactin in fish osmoregulation: a review. *General and comparative endocrinology*, 125 (2), 291–310.

Mardis, E. R., 2008. The impact of next-generation sequencing technology on genetics. *Trends in genetics : TIG*, 24 (3), 133–41.

Mardis, E. R., 2013. Next-generation sequencing platforms. *Annual review of analytical chemistry*, 6, 287–303.

Marshall, C. and Browman, H., 2007. Disentangling the causes of maturation trends in exploited fish populations. *Marine Ecology Progress Series*, 335, 249.

Marshall, C. and McAdam, B., 2007. Integrated perspectives on genetic and environmental effects on maturation can reduce potential for errors of inference. *Marine Ecology Progress Series*, 335, 301–310.

Marshall, C. T., Pardoe, H., Vainikka, A., Thórdarson, G., Marteinsdóttir, G., and Heino, M., 2009. Temporal trends in probabilistic maturation reaction norms and growth of Atlantic cod (Gadus morhua) on the Icelandic shelf. *Canadian Journal of Fisheries and Aquatic Sciences*, 66 (10), 1719–1733.

Martin, C. and Johnsen, S., 2007. A field test of the Hamilton–Zuk hypothesis in the Trinidadian guppy (Poecilia reticulata). *Behavioral Ecology and Sociobiology*, 61 (12), 1897–1909.

Martinez, V. A., Hill, W. G., and Knott, S. A., 2002. On the use of double haploids for detecting QTL in outbred populations. *Heredity*, 88 (6), 423–431.

Marty, L., Dieckmann, U., and Ernande, B., 2015. Fisheries-induced neutral and adaptive evolution in exploited fish populations and consequences for their adaptive potential. *Evolutionary applications*, 8 (1), 47–63.

Martyniuk, C. J., Perry, G. M. L., Mogahadam, H. K., Ferguson, M. M., and Danzmann, R. G., 2003. The genetic architecture of correlations among growth-related traits and male age at maturation in rainbow trout. *Journal of Fish Biology*, 63 (3), 746–764.

Massicotte, R., Whitelaw, E., and Angers, B., 2011. DNA methylation: A source of random variation in natural populations. *Epigenetics*, 6 (4), 421–427.

Matthews, R. P., Lorent, K., and Pack, M., 2008. Transcription factor onecut3 regulates intrahepatic biliary development in zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists*, 237 (1), 124–31.

Mattingly, H. and Butler, M., 1994. Laboratory predation on the Trinidadian guppy: implications for the size-selective predation hypothesis and guppy life history evolution. *Oikos*, 69 (1), 54–64.

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S. J. M., Haussler, D., Marra, M. A., Hirst, M., Wang, T., and Costello, J. F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 466 (7303), 253–7.

Mayr, E., 1963. *Animal species and evolution.* Harvard University Press.

McClure, M. C., Bickhart, D., Null, D., Vanraden, P., Xu, L., Wiggans, G., Liu, G., Schroeder, S., Glasscock, J., Armstrong, J., Cole, J. B., Van Tassell, C. P., and Sonstegard, T. S., 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. *PloS one*, 9 (3), e92769.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular phylogenetics and evolution*, 66 (2), 526–38.

McKay, J. K. and Latta, R. G., 2002. Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, 17 (6), 285–291.

McKenzie, J. A., 2000. The character or the variation: the genetic analysis of the insecticide-resistance phenotype. *Bulletin of entomological research*, 90 (1), 3–7.

McKenzie, J. A., Parker, A. G., and Yen, J. L., 1992. Polygenic and single gene responses to selection for resistance to diazinon in Lucilia cuprina. *Genetics*, 130 (3), 613–20.

McKenzie, J. and Batterham, P., 1994. The genetic, molecular and phenotypic consequences of selection for insecticide resistance. *Trends in Ecology & Evolution*, 9 (5), 166–169.

McMullan, M. and Van Oosterhout, C., 2012. Inference of Selection Based on Temporal Genetic Differentiation in the Study of Highly Polymorphic Multigene Families. *PLoS ONE*, 7 (8), e42119.

McPherron, A. C. and Lee, S.-J., 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Sciences*, 94 (23), 12457–12461.

Meaney, M. J. and Szyf, M., 2005. Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues in clinical neuroscience*, 7 (2), 103–23.

Medrano, M., Herrera, C. M., and Bazaga, P., 2014. Epigenetic variation predicts regional and local intraspecific functional diversity in a perennial herb. *Molecular ecology*, 23 (20), 4926–38.

Mendel, G. J., 1866. Experiments in plant hybridization. *Verhandlungen des naturforschenden*

*Vereines in Brunn*, 4, 3–47.

Mertz, G. and Myers, R., 1998. A simplified formulation for fish production. *Canadian Journal of Fisheries and Aquatic Sciences*, 55 (2), 478–484.

Messer, P. W. and Petrov, D. A., 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*, 28 (11), 659–69.

Metzker, M. L., 2010. Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11 (1), 31–46.

Meyer, R. C., Kusterer, B., Lisec, J., Steinfath, M., Becher, M., Scharr, H., Melchinger, A. E., Selbig, J., Schurr, U., Willmitzer, L., and Altmann, T., 2010. QTL analysis of early stage heterosis for biomass in Arabidopsis. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 120 (2), 227–37.

Mhanni, A. A., Yoder, J. A., Dubesky, C., and McGowan, R. A., 2001. Cloning and sequence analysis of a zebrafish cDNA encoding DNA (cytosine-5)-methyltransferase-1. *Genesis (New York, N.Y. : 2000)*, 30 (4), 213–9.

Mhanni, A. and McGowan, R., 2004. Global changes in genomic methylation levels during early development of the zebrafish embryo. *Development genes and evolution*, 214 (8), 412–417.

Miglani, G. S., 2002. *Advanced Genetics*. CRC Press.

Miller, M., Atwood, T., and Eames, B., 2007. RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome biology*, 8, R105.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research*, 17 (2), 240–8.

Mirbahai, L., Yin, G., Bignell, J. P., Li, N., Williams, T. D., and Chipman, J. K., 2011. DNA methylation in liver tumorigenesis in fish from the environment. *Epigenetics*, 6 (11), 1319–33.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., and Vigouroux, Y., 2013. Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, 22, 1383–1399.

Mizoshita, K., Watanabe, T., Hayashi, H., Kubota, C., Yamakuchi, H., Todoroki, J., and Sugimoto, Y., 2004. Quantitative trait loci analysis for growth and carcass traits in a half-sib family of purebred Japanese Black (Wagyu) cattle. *Journal of Animal Science*, 82 (12), 3415–3420.

Moghadam, H., Mørkøre, T., and Robinson, N., 2015. Epigenetics—Potential for Programming Fish for Aquaculture? *Journal of Marine Science and Engineering*, 3 (2), 175–192.

Molinier, J., Ries, G., Zipfel, C., and Hohn, B., 2006. Transgeneration memory of stress in plants. *Nature*, 442, 1046–1049.

Mollet, F. M., Ernande, B., Brunel, T., and Rijnsdorp, A. D., 2010. Multiple growth-correlated life history traits estimated simultaneously in individuals. *Oikos*, 119 (1), 10–26.

Mollet, F. M., Kraak, S. B. M., and Rijnsdorp, A. D., 2007. Fisheries-induced evolutionary changes in maturation reaction norms in North Sea sole Solea solea. *Marine Ecology Progress Series*, 351, 189–199.

Moon, M. and Gomez, T. M., 2010. Balanced Vav2 GEF activity regulates neurite outgrowth and branching in vitro and in vivo. *Molecular and cellular neurosciences*, 44 (2), 118–28.

Moore, C. a C., Milano, S. K., and Benovic, J. L., 2007. Regulation of receptor trafficking by GRKs and arrestins. *Annual review of physiology*, 69, 451–482.

Morán, P. and Pérez-Figueroa, A., 2011. Methylation changes associated with early maturation stages in the Atlantic salmon. *BMC genetics*, 12 (86).

Morita, K. and Fukuwaka, M., 2006. Does size matter most? The effect of growth history on probabilistic reaction norm for salmon maturation. *Evolution*, 60 (7), 1516–1521.

Morita, K., Tsuboi, J., and Nagasawa, T., 2009. Plasticity in probabilistic reaction norms for maturation in a salmonid fish. *Biology letters*, 11 (12).

Morozova, O. and Marra, M. A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92 (5), 255–64.

Morton, R. A., 1993. Evolution of Drosophila insecticide resistance. *Genome*, 36 (1), 1–7.

Mosher, D. S., Quignon, P., Bustamante, C. D., Sutter, N. B., Mellersh, C. S., Parker, H. G., and Ostrander, E. A., 2007. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS genetics*, 3 (5), e79.

Mundy, N. I., 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings. Biological sciences / The Royal Society*, 272 (1573), 1633–40.

Muposhi, V. K., Gandiwa, E., and Makuza, S. M., 2015. Population and trophy quality trends of three gregarious herbivores in an insulated semi-arid savanna ecosystem (Cawston Ranch, Zimbabwe), 1997-2014. *Tropical Conversation Science*, 8 (2), 424–438.

Muschick, M., Indermaur, A., and Salzburger, W., 2012. Convergent evolution within an adaptive radiation of cichlid fishes. *Current biology*, 22 (24), 2362–2368.

Myles, S., Peiffer, J., Brown, P., and Ersoz, E., 2009. Association mapping: critical considerations shift from genotyping to experimental design. *American Society of Plant Biologists*, 21 (8), 2194–2202.

Nachman, M. W., Hoekstra, H. E., and D'Agostino, S. L., 2003. The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (9), 5268–73.

Nadeau, N. J. and Jiggins, C. D., 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in genetics : TIG*, 26 (11), 484–92.

Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., Baxter, S. W., Blaxter, M. L., Mallet, J., and Jiggins, C. D., 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular ecology*, 22 (3), 814–26.

Naegele, R. P., Ashrafi, H., Hill, T. A., Chin-Wo, S. R., Van Deynze, A. E., and Hausbeck, M. K., 2014. QTL mapping of fruit rot resistance to the plant pathogen Phytophthora capsici in a recombinant inbred line Capsicum annuum population. *Phytopathology*, 104 (5), 479–83.

Najarro, M. A., Hackett, J. L., Smith, B. R., Highfill, C. A., King, E. G., Long, A. D., and Macdonald, S. J., 2015. Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. *PLoS genetics*, 11 (11), e1005663.

Nakadate, M., Shikano, T., and Taniguchi, N., 2003. Inbreeding depression and heterosis in various quantitative traits of the guppy, Poecilia reticulata. *Aquaculture*, 220 (1-4), 219–226.

Nanda, I., Schories, S., Tripathi, N., Dreyer, C., Haaf, T., Schmid, M., and Schartl, M., 2014. Sex chromosome polymorphism in guppies. *Chromosoma*, 123 (4), 373–83.

Narum, S. R., Campbell, N. R., Kozfkay, C. C., and Meyer, K. A., 2010. Adaptation of redband trout in desert and montane environments. *Molecular ecology*, 19 (21), 4622–37.

Narum, S. R. and Hess, J. E., 2011. Comparison of F ST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11, 184–194.

Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L., and Piferrer, F., 2011. DNA methylation of the gonadal aromatase (cyp19a) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS genetics*, 7 (12), e1002447.

Neuheimer, A. B. and Taggart, C. T., 2010. Can changes in length-at-age and maturation timing in Scotian Shelf haddock (Melanogrammus aeglefinus) be explained by fishing? *Canadian Journal of Fisheries and Aquatic Sciences*, 67 (5), 854–865.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J., and Bamshad, M. J., 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*, 42 (1), 30–5.

Nicotra, A. B., Segal, D. L., Hoyle, G. L., Schrey, A. W., Verhoeven, K. J. F., and Richards, C. L., 2015. Adaptive plasticity and epigenetic variation in response to warming in an Alpine plant. *Ecology and Evolution*, 5 (3), 634–647.

Nie, Q., Sun, B., Zhang, D., Luo, C., Ishag, N. A., Lei, M., Yang, G., and Zhang, X., High diversity of the chicken growth hormone gene and effects on growth and carcass traits. *The Journal of heredity*, 96 (6), 698–703.

Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., Mittelholzer, C., Taranger, G.-L., Ogden, R., and Carvalho, G. R., 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (Gadus morhua). *BMC evolutionary biology*, 9 (1), 276.

Nishimura, S., Watanabe, T., Mizoshita, K., Tatsuda, K., Fujita, T., Watanabe, N., Sugimoto, Y., and Takasuga, A., 2012. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. *BMC genetics*, 13 (1), 40.

Nosil, P., 2012. *Ecological speciation*. London, UK: Oxford University Press.

Nuzzo, M. C. and Traill, L. W., 2014. What 50 years of trophy records illustrate for hunted African elephant and bovid populations. *African Journal of Ecology*, 52 (2), 250–253.

Nylin, S. and Gotthard, K., 1998. Plasticity in life-history traits. *Annual review of entomology*, 43, 63–83.

O'Steen, S., Cullum, A., and Bennett, A., 2002. Rapid evolution of escape ability in Trinidadian guppies (Poecilia reticulata). *Evolution*, 56 (4), 776–784.

Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J. W., Pourkazemi, M., McEwing, R., Eland, C., Vidotto, M., Sergeev, A., and Congiu, L., 2013. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular ecology*, 22 (11), 3112–23.

Oleksiak, M. F., Churchill, G. A., and Crawford, D. L., 2002. Variation in gene expression within and among natural populations. *Nature genetics*, 32 (2), 261–6.

Olendorf, R., Reudi, B., and Hughes, K., 2004. Primers for 12 polymorphic microsatellite DNA loci from the guppy (Poecilia reticulata). *Molecular Ecology Notes*, 4 (4), 668–371.

Ollivier, L., Messer, A. L., Rothschild, M. F., and Legault, C., 1997. The use of selection experiments for detecting quantitative trait loci. *Genetical Research*, 69 (03), 227–232.

Olsen, E. E. M., Heino, M., Lilly, G. R. G., Morgan, M. J., Brattey, J., Ernande, B., and Dieckmann, U., 2004. Maturation trends indicative of rapid evolution preceded the collapse of northern cod. *Nature*, 428 (6986), 932–935.

Olsen, H. G., Gomez-Raya, L., Våge, D. I., Olsaker, I., Klungland, H., Svendsen, M., Adnøy, T., Sabry, A., Klemetsdal, G., Schulman, N., Krämer, W., Thaller, G., Rønningen, K., and Lien, S., 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *Journal of dairy science*, 85 (11), 3124–30.

Oosterhout, C. Van, 2003. Marked variation in parasite resistance between two wild populations of the Trinidadian guppy, Poecilia reticulata (Pisces: Poeciliidae). *Biological Journal of the Linnean Society*, 79 (4), 645–651.

van Oosterhout, C., Joyce, D. a., Cummings, S. M., Blais, J., Barson, N. J., Ramnarine, I. W., Mohammed, R. S., Persad, N., and Cable, J., 2006. Balancing Selection, Random Genetic Drift, and Genetic Variation At the Major Histocompatibility Complex in Two Wild Populations of Guppies (Poecilia Reticulata). *Evolution*, 60 (12), 2562.

Oosterhout, C. Van and Smith, A., 2007. The guppy as a conservation model: implications of parasitism and inbreeding for reintroduction success. *Conservation …*.

Oosterhout, C. Van, Smith, A. A. M., Van Oosterhout, C., Smith, A. A. M., Hänfling, B., Ramnarine, I. W., Mohammed, R. S., and Cable, J., 2007. The guppy as a conservation model: Implications of parasitism and inbreeding for reintroduction success. *Conservation Biology*, 21 (6), 1573–1583.

Orpwood, J. E., Griffiths, S. W., and Armstrong, J. D., 2006. Effects of food availability on temporal activity patterns and growth of Atlantic salmon. *The Journal of animal ecology*, 75 (3), 677–85.

Orr, H. A., 2005. The probability of parallel evolution. *Evolution*, 59 (1), 216–220.

Orr, H. A. and Betancourt, A. J., 2001. Haldane's Sieve and Adaptation From the Standing Genetic Variation. *Genetics*, 157 (2), 875–884.

Orsini, L., Spanier, K. I., and DE Meester, L., 2012. Genomic signature of natural and anthropogenic stress in wild populations of the waterflea Daphnia magna: validation in space, time and experimental evolution. *Molecular ecology*, 21 (9), 2160–75.

Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N., and Weigel, D., 2008. Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Research*, 18, 2024–2033.

Paik, W. K. and Kim, S., 1971. Protein methylation. *Science*, 174 (4005), 114–9.

Palaiokostas, C., Bekaert, M., Khan, M. G. Q., Taggart, J. B., Gharbi, K., McAndrew, B. J., and Penman, D. J., 2013. Mapping and validation of the major sex-determining region in Nile tilapia (Oreochromis niloticus L.) Using RAD sequencing. *PloS one*, 8 (7), e68389.

Palkovacs, E. P., Dion, K. B., Post, D. M., and Caccone, A., 2008. Independent evolutionary origins of landlocked alewife populations and rapid parallel evolution of phenotypic traits. *Molecular ecology*, 17 (2), 582–97.

Palkovacs, E. P., Kinnison, M. T., Correa, C., Dalton, C. M., and Hendry, A. P., 2012. Fates beyond traits: ecological consequences of human-induced trait change. *Evolutionary Applications*, 5 (2), 183–191.

Palkovacs, E. P., Marshall, M. C., Lamphere, B. a., Lynch, B. R., Weese, D. J., Fraser, D. F., Reznick, D. N., Pringle, C. M., and Kinnison, M. T., 2009. Experimental evaluation of evolution and coevolution as agents of ecosystem change in Trinidadian streams. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364 (1523), 1617–1628.

Palkovacs, E. P. and Post, D. M., 2009. Experimental evidence that phenotypic divergence in predators drives community divergence in prey. *Ecology*, 90 (2), 300–305.

Palkovacs, E. P., Wasserman, B. A., and Kinnison, M. T., 2011. Eco-evolutionary trophic dynamics: loss of top predators drives trophic evolution and ecology of prey. *PloS one*, 6 (4), e18879.

Palumbi, S. R., 2001. Humans as the world's greatest evolutionary force. *Science (New York, N.Y.)*, 293 (5536), 1786–90.

Pankin, A., Campoli, C., Dong, X., Kilian, B., Sharma, R., Himmelbach, A., Saini, R., Davis, S. J., Stein, N., Schneeberger, K., and von Korff, M., 2014. Mapping-by-sequencing identifies HvPHYTOCHROME C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics*, 198 (1), 383–96.

PARCHMAN, T. L., GOMPERT, Z., MUDGE, J., SCHILKEY, F. D., BENKMAN, C. W., and BUERKLE, C. A., 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, 21 (12), 2991–3005.

Pardo-Diaz, C., Salazar, C., and Jiggins, C. D., 2015. Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, 6 (4), n/a–n/a.

Parker, J., Tsagkogeorga, G., Cotton, J. A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S. J., 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*, 502 (7470), 228–31.

Parks, J. S. and Brown, M. R., 1999. Transcription factors regulating pituitary development. *Growth Hormone & IGF Research*, 9, 2–11.

Parts, L., Cubillos, F. A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S. J., Molin, M., Zia, A., Simpson, J. T., Quail, M. A., Moses, A., Louis, E. J., Durbin, R., and Liti, G., 2011. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome research*, 21 (7), 1131–8.

Paterson, I. G., Crispo, E., Kinnison, M. T., Hendry, A. P., and Bentzen, P., 2005. Characterization of tetranucleotide microsatellite markers in guppy (Poecilia reticulata). *Molecular Ecology Notes*, 5 (2), 269–271.

Paul, D., 2000. A double-edged sword. *Nature*, 405 (6786), 515.

Pauli, B. D. and Heino, M., 2013. The importance of social dimension and maturation stage for the

probabilistic maturation reaction norm in Poecilia reticulata. *Journal of Evolutionary Biology*, 26 (10), 2184–2196.

Pauly, D., 1980. On the interrelationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. *ICES Journal of Marine Science*, 39 (2), 175–192.

Paun, O., Bateman, R. M., Fay, M. F., Hedrén, M., Civeyrel, L., and Chase, M. W., 2010. Stable epigenetic effects impact adaptation in allopolyploid orchids (Dactylorhiza: Orchidaceae). *Molecular biology and evolution*, 27 (11), 2465–73.

Pavey, S. A., Gaudin, J., Normandeau, E., Dionne, M., Castonguay, M., Audet, C., and Bernatchez, L., 2015. RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Current biology : CB*, 25 (12), 1666–71.

Pavlidis, P., Metzler, D., and Stephan, W., 2012. Selective sweeps in multilocus models of quantitative traits. *Genetics*, 192 (1), 225–39.

Peakall, R. and Smouse, P. E., 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6 (1), 288–295.

Peakall, R. and Smouse, P. E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics (Oxford, England)*, 28 (19), 2537–9.

Peckarsky, B. and McIntosh, A., 1998. Fitness and community consequences of avoiding multiple predators. *Oecologia*, 113 (4), 565–576.

Pegadaraju, V., Nipper, R., Hulke, B., Qi, L., and Schultz, Q., 2013. De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. *BMC genomics*, 14 (1), 556.

Pei, J. and Grishin, N. V, 2012. Unexpected diversity in Shisa-like proteins suggests the importance of their roles as transmembrane adaptors. *Cellular signalling*, 24 (3), 758–69.

Pennings, P. S. and Hermisson, J., 2006. Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. *Molecular biology and evolution*, 23 (5), 1076–84.

Pérez-Figueroa, a., García-Pereira, M. J., Saura, M., Rolán-Alvarez, E., and Caballero, a., 2010a. Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, 23 (1999), 2267–2276.

Pérez-Figueroa, a., García-Pereira, M. J., Saura, M., Rolán-Alvarez, E., and Caballero, a., 2010b. Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, 23 (10), 2267–2276.

Perry, G. M. L., Martyniuk, C. M., Ferguson, M. M., and Danzmann, R. G., 2005. Genetic parameters for upper thermal tolerance and growth-related traits in rainbow trout (Oncorhynchus mykiss). *Aquaculture*, 250 (1-2), 120–128.

Peterson, B. C. and Small, B. C., 2005. Effects of exogenous cortisol on the GH/IGF-I/IGFBP network in channel catfish. *Domestic animal endocrinology*, 28 (4), 391–404.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, 7 (5), e37135.

Pfender, W. F., Saha, M. C., Johnson, E. A., and Slabaugh, M. B., 2011. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in Lolium perenne. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 122 (8), 1467–80.

Pierron, F., Bureau du Colombier, S., Moffett, A., Caron, A., Peluhet, L., Daffe, G., Lambert, P., Elie, P., Labadie, P., Budzinski, H., Dufour, S., Couture, P., and Baudrimont, M., 2014. Abnormal ovarian DNA methylation programming during gonad maturation in wild contaminated fish. *Environmental science & technology*, 48 (19), 11688–95.

Piertney, S. B. and Webster, L. M. I., 2010. Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica*, 138 (4), 419–32.

Pinney, S. E., 2014. Mammalian Non-CpG Methylation: Stem Cells and Beyond. *Biology*, 3 (4), 739–51.

Pinsky, M. L. and Palumbi, S. R., 2014. Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular ecology*, 23 (1), 29–39.

Poissant, J., Réale, D., Martin, J., Festa-Bianchet, M., and Coltman, D., 2013. A quantitative trait locus analysis of personality in wild bighorn sheep. *Ecology and evolution*, 3 (3), 474–81.

Poland, J. a., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. L., 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, 7 (2), e32253.

Portela, A. and Esteller, M., 2010. Epigenetic modifications and human disease. *Nature biotechnology*, 28 (10), 1057–1068.

Portis, E., Acquadro, A., Comino, C., and Lanteri, S., 2004. Analysis of DNA methylation during germination of pepper (Capsicum annuum L.) seeds using methylation-sensitive amplification polymorphism (MSAP). *Plant Science*, 166 (1), 169–178.

Pörtner, H. ., Berdal, B., Blust, R., Brix, O., Colosimo, A., De Wachter, B., Giuliani, A., Johansen, T., Fischer, T., Knust, R., Lannig, G., Naevdal, G., Nedenes, A., Nyhammer, G., Sartoris, F. ., Serendero, I., Sirabella, P., Thorkildsen, S., and Zakhartsev, M., 2001. Climate induced temperature effects on growth performance, fecundity and recruitment in marine fish: developing a hypothesis for cause and effect relationships in Atlantic cod (Gadus morhua) and common eelpout (Zoarces viviparus). *Continental Shelf Research*, 21 (18-19), 1975–1997.

Post, D. M., Palkovacs, E. P., Schielke, E. G., and Dodson, S. I., 2008. Intraspecific variation in a predator affects community structure and cascading trophic interactions. *Ecology*, 89 (7), 2019–32.

Poulsen, N. A. A., Nielsen, E. E., Schierup, M. H., Loeschcke, V., and Grønkjaer, P., 2006. Long-term stability and effective population size in North Sea and Baltic Sea cod (Gadus morhua). *Molecular ecology*, 15 (2), 321–31.

Powers, D. a, Lauerman, T., Crawford, D., and DiMichele, L., 1991. Genetic mechanisms for adapting to a changing environment. *Annual review of genetics*, 25, 629–659.

Powers, D. A. and Schulte, P. M., 1998. Evolutionary adaptations of gene structure and expression in natural populations in relation to a changing environment: a multidisciplinary approach to address the million-year saga of a small fish. *The Journal of experimental zoology*, 282 (1-2), 71–94.

Price, A., Zaitlen, N., Reich, D., and Patterson, N., 2010. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11, 459–463.

Pritchard, J. K. and Di Rienzo, A., 2010. Adaptation - not by sweeps alone. *Nature reviews. Genetics*, 11 (10), 665–667.

Przeworski, M., Coop, G., and Wall, J. D., 2005. The signature of positive selection on standing genetic variation. *Evolution*, 59 (11), 2312.

Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics. *Current opinion in plant biology*, 5 (2), 94–100.

Rai, K., Chidester, S., Zavala, C. V, Manos, E. J., James, S. R., Karpf, A. R., Jones, D. A., and Cairns, B. R., 2007. Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes & development*, 21 (3), 261–6.

Rakyan, V. K., Chong, S., Champ, M. E., Cuthbert, P. C., Morgan, H. D., Luu, K. V. K., and Whitelaw, E., 2003. Transgenerational inheritance of epigenetic states at the murine AxinFu allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences*, 100 (5), 2538–2543.

Ramos, A. M., Crooijmans, R. P. M. A., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., Hansen, M. S., Hedegaard, J., Hu, Z.-L., Kerstens, H. H., Law, A. S., Megens, H.-J., Milan, D., Nonneman, D. J., Rohrer, G. A., Rothschild, M. F., Smith, T. P. L., Schnabel, R. D., Van Tassell, C. P., Taylor, J. F., Wiedmann, R. T., Schook, L. B., and Groenen, M. A. M., 2009. Design of a high density SNP

genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PloS one*, 4 (8), e6524.

Raymond, M., Berticat, C., and Weill, M., 2001. Insecticide resistance in the mosquito Culex pipiens: what have we learned about adaptation? *Contemporary Issues in Genetics and Evolution*, 8, 287–296.

Raymond, M. and Rousset, F., 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *J. Hered.*, 86 (3), 248–249.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., Jiggins, C. D., Chamberlain, N. L., Kronforst, M. R., Chen, R., Halder, G., Nijhout, H. F., and McMillan, W. O., 2011. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science (New York, N.Y.)*, 333 (6046), 1137–41.

Reik, W. and Walter, J., 2001. Genomic imprinting: parental influence on the genome. *Nature reviews. Genetics*, 2 (1), 21–32.

Reitzel, A. M., Herrera, S., Layden, M. J., Martindale, M. Q., and Shank, T. M., 2013. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular ecology*, 22 (11), 2953–70.

Remington, D., 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*, 98 (20), 11479–11484.

Renaut, S., Nolte, A. W., Rogers, S. M., Derome, N., and Bernatchez, L., 2011. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (Coregonus spp.). *Molecular ecology*, 20 (3), 545–59.

Reznick, D., 1982. The impact of predation on life history evolution in Trinidadian guppies: genetic basis of observed life history patterns. *Evolution*, 36 (6), 1236–1250.

Reznick, D., 1990. Plasticity in age and size at maturity in male guppies (Poecilia reticulata): an experimental evaluation of alternative models of development. *Journal of evolutionary Biology*, 3 (3-4), 185–203.

Reznick, D., 1997. Life history evolution in guppies (Poecilia reticulata): guppies as a model for studying the evolutionary biology of aging. *Experimental Gerontology*.

Reznick, D., Bassar, R., Travis, J., and Rodd, F. H., 2012. Life-history evolution in guppies VIII: the demographics of density regulation in guppies (Poecilia reticulata). *Evolution*, 66 (9), 2903–2915.

Reznick, D., Bryant, M. J., and Bashey, F., 2002. r - AND K -SELECTION REVISITED: THE ROLE OF POPULATION REGULATION IN LIFE-HISTORY EVOLUTION. *Ecology*, 83 (6), 1509–1520.

Reznick, D. and Bryga, H., 1987a. Life-history evolution in guppies (Poecilia reticulata): 1. Phenotypic and genetic changes in an introduction experiment. *Evolution*.

Reznick, D., Bryga, H., and Endler, J., 1990. Experimentally induced life-history evolution in a natural population. *Nature*, 346, 357–359.

Reznick, D., Butler, M., Rodd, F., and Ross, P., 1996. Life-history evolution in guppies (Poecilia reticulata) VI. Differential mortality as a mechanism for natural selection. *Evolution*, 50 (4), 1651–1660.

Reznick, D. and Endler, J., 1982. The impact of predation on life history evolution in Trinidadian guppies (Poecilia reticulata). *Evolution*, 36 (1), 160–177.

Reznick, D. and Ghalambor, C., 2005. Can commercial fishing cause evolution? Answers from guppies (Poecilia reticulata). *Canadian Journal of Fisheries and Aquatic Sciences*, 62 (4), 791–810.

Reznick, D. N. and Bryga, H., 1987b. Life-History Evolution in Guppies ( Poecilia reticulata ): I. Phenotypic and Genetic Changes in an Introduction. *Evolution*, 41 (6), 1370–1385.

Reznick, D. N. and Bryga, H. A., 1996. Life-history evolution in guppies (Poecilia reticulata: Poeciliidae). V. Genetic basis of parallelism in life histories. *American Naturalist*, 147 (3), 339–

359.

Reznick, D. N. and Ghalambor, C. K., 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*, 112-113 (1), 183–198.

Reznick, D. N., Ghalambor, C. K., and Crooks, K., 2008. Experimental studies of evolution in guppies: a model for understanding the evolutionary consequences of predator removal in natural communities. *Molecular ecology*, 17 (1), 97–107.

Reznick, D. N., Rodd, F. H., and Cardenas, M., 1996. Life-History Evolution in Guppies (Poecilia reticulata: Poeciliidae). IV. Parallelism in Life-History Phenotypes. *The American Naturalist*, 147 (3), 319.

Reznick, D., Shaw, F., Rodd, F., and Shaw, R., 1997. Evaluation of the rate of evolution in natural populations of guppies (Poecilia reticulata). *Science*, 275 (5308), 1934–1937.

Richards, C., Bossdorf, O., and Verhoeven, K., 2010. Understanding natural epigenetic variation. *New Phytologist*, 39 (237–257).

Richards, C. L., Schrey, A. W., and Pigliucci, M., 2012. Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. *Ecology Letters*, 15 (9), 1016–1025.

Richards, P. M., Liu, M. M., Lowe, N., Davey, J. W., Blaxter, M. L., and Davison, A., 2013. RAD-Seq derived markers flank the shell colour and banding loci of the Cepaea nemoralis supergene. *Molecular Ecology*, 22 (11), 3077–3089.

Ricker, W., 1981. Changes in the average size and average age of Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, 38 (12), 1636–1656.

Rijnsdorp, A. D., Grift, R. E., and Kraak, S. B., 2005. Fisheries-induced adaptive change in reproductive investment in North Sea plaice ( Pleuronectes platessa )? *Canadian Journal of Fisheries and Aquatic Sciences*, 62 (4), 833–843.

Rinkevich, F. D., Hamm, R. L., Geden, C. J., and Scott, J. G., 2007. Dynamics of insecticide resistance alleles in house fly populations from New York and Florida. *Insect biochemistry and molecular biology*, 37 (6), 550–8.

Rivrud, I. M., Sonkoly, K., Lehoczki, R., Csányi, S., Storvik, G. O., and Mysterud, A., 2013. Hunter selection and long-term trend (1881-2008) of red deer trophy sizes in Hungary. *Journal of Applied Ecology*, 50 (1), 168–180.

Robertson, K. D., 2005. DNA methylation and human disease. *Nature reviews. Genetics*, 6 (8), 597–610.

Robertson, K. D. and Jones, P. A., 2000. DNA methylation: past, present and future directions. *Carcinogenesis*, 21 (3), 461–7.

Robinson, M. R., Santure, A. W., Decauwer, I., Sheldon, B. C., and Slate, J., 2013. Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular ecology*, 22 (15), 3963–80.

Roda, F., Ambrose, L., Walter, G. M., Liu, H. L., Schaul, A., Lowe, A., Pelser, P. B., Prentis, P., Rieseberg, L. H., and Ortiz-Barrientos, D., 2013. Genomic evidence for the parallel evolution of coastal forms in the Senecio lautus complex. *Molecular ecology*, 22 (11), 2941–52.

Rodd, F. H., Hughes, K. a, Grether, G. F., and Baril, C. T., 2002. A possible non-sexual origin of mate preference: are male guppies mimicking fruit? *Proceedings. Biological sciences / The Royal Society*, 269 (1490), 475–481.

Rodd, F. H. and Reznick, D. N., 1991. Life-History Evolution in Guppies .3. the Impact of Prawn Predation on Guppy Life Histories. *Oikos*, 62 (1), 13–19.

Rodríguez López, C. M., Morán, P., Lago, F., Espiñeira, M., Beckmann, M., and Consuegra, S., 2012. Detection and quantification of tissue of origin in salmon and veal products using methylation sensitive AFLPs. *Food Chemistry*, 131 (4), 1493–1498.

Roesti, M., Hendry, A. P., Salzburger, W., and Berner, D., 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs.

*Molecular ecology*, 21 (12), 2852–62.

Roff, D., 1992. *The evolution of life histories: Theory and analysis*. Springer.

Roff, D., 2003. *Life history evolution*. Sinauer Associates Inc.

Rolshausen, G., Segelbacher, G., Hobson, K. A., and Schaefer, H. M., 2009. Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. *Current biology : CB*, 19 (24), 2097–101.

Roos, A. De, 2006. Evolutionary regime shifts in age and size at maturation of exploited fish stocks. *Proceedings of the Royal Society B: Biological Sciences*, 273 (1596).

Rose, G. a and Rowe, S., 2015. Northern cod comeback, 1798 (October), 1789–1798.

Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources*, 8 (1), 103–6.

Rowe, H., Renaut, S., and Guggisberg, A., 2011. RAD in the realm of next generation sequencing technologies. *Molecular Ecology*, 20 (17), 3499–3502.

Rubin, B. E. R., Ree, R. H., and Moreau, C. S., 2012. Inferring phylogenies from RAD sequence data. *PloS one*, 7 (4), e33394.

Rudge, D. W., 1999. Taking the Peppered Moth with a Grain of Salt. *Biology and Philosophy*, 14 (1), 9–37.

Ruzzante, D. E., Taggart, C. T., Doyle, R. W., and Cook, D., 2001. Stability in the historical pattern of genetic structure of Newfoundland cod (Gadus morhua) despite the catastrophic decline in population size from 1964 to 1994. *Conservation Genetics*, 2 (3), 257–269.

Rzewuska, K., Klewiec, J., and Martyniuk, E., 2005. Effect of inbred on reproduction and body weight of sheep in a closed Booroola flock. *Genetics*, 23 (4), 237–247.

Sahana, G., Guldbrandtsen, B., Bendixen, C., and Lund, M. S., 2010. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Animal genetics*, 41 (6), 579–88.

Salamin, N., Wüest, R. O., Lavergne, S., Thuiller, W., and Pearman, P. B., 2010. Assessing rapid evolution in a changing environment. *Trends in ecology & evolution*, 25 (12), 692–8.

Salem, M., Kenney, P. B., Rexroad, C. E., and Yao, J., 2006. Microarray gene expression analysis in atrophying rainbow trout muscle: a unique nonmammalian muscle degradation model. *Physiological genomics*, 28 (1), 33–45.

Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., Rexroad, C. E., and Yao, J., 2012. RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PloS one*, 7 (5), e36264.

Salinas, S., Perez, K. O., Duffy, T. A., Sabatino, S. J., Hice, L. A., Munch, S. B., and Conover, D. O., 2012. The response of correlated traits following cessation of fishery-induced selection. *Evolutionary applications*, 5 (7), 657–63.

Salmon, A., Bellis, H., Chable, V., and Manzanares-Dauleux, M., 2009. Identification of differentially expressed genes related to aberrant phenotypes in Brassica oleracea var. botrytis. *Plant Breeding*, 128 (6), 631–639.

Salmon, A., Clotault, J., Jenczewski, E., Chable, V., and Manzanares-Dauleux, M. J., 2008. Brassica oleracea displays a high level of DNA methylation polymorphism. *Plant Science*, 174 (1), 61–70.

Sánchez-Ramos, I., Cross, I., Mácha, J., Martínez-Rodríguez, G., Krylov, V., and Rebordinos, L., 2012. Assessment of Tools for Marker-Assisted Selection in a Marine Commercial Species: Significant Association between MSTN-1 Gene Polymorphism and Growth Traits. *The Scientific World Journal*, 2012, 1–10.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M., and Smith, M., 1977. Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, 265 (5596), 687–695.

Santure, A. W., De Cauwer, I., Robinson, M. R., Poissant, J., Sheldon, B. C., and Slate, J., 2013.

Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. *Molecular ecology*, 22 (15), 3949–62.

Sarabi, M. M. and Naghibalhossaini, F., 2015. Association of DNA methyltransferases expression with global and gene-specific DNA methylation in colorectal cancer cells. *Cell biochemistry and function*, 33 (7), 427–33.

Saze, H., Tsugane, K., Kanno, T., and Nishimura, T., 2012. DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation. *Plant & cell physiology*, 53 (5), 766–84.

Scaglione, D., Acquadro, A., Portis, E., Tirone, M., Knapp, S. J., and Lanteri, S., 2012. RAD tag sequencing as a source of SNP markers in Cynara cardunculus L. *BMC Genomics*, 13 (1), 3.

Scheiner, S. M., 2002. Selection experiments and the study of phenotypic plasticity1. *Journal of Evolutionary Biology*, 15 (6), 889–898.

Schielzeth, H., Kempenaers, B., Ellegren, H., and Forstmeier, W., 2012. QTL linkage mapping of zebra finch beak color shows an oligogenic control of a sexually selected trait. *Evolution; international journal of organic evolution*, 66 (1), 18–30.

Schlötterer, C., Kofler, R., Versace, E., Tobler, R., and Franssen, S. U., 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, 114 (5), 431–40.

Schluter, D., 2000. *The ecology of adaptive radiation*. Oxford University Press.

Schmidt, J., 1919. Racial studies in fishes. *Journal of Genetics*, 8 (3), 147–153.

Schories, S., Meyer, M., and Schartl, M., 2009. Description of Poecilia (Acanthophacelus) obscura n. sp.,(Teleostei: Poeciliidae), a new guppy species from western Trinidad, with remarks on P. wingei and. *Zootaxa*, 2266, 35–50.

Schrey, A., Coon, C., and Grispo, M., 2012. Epigenetic variation may compensate for decreased genetic variation with introductions: a case study using house sparrows (Passer domesticus) on two continents. *Genetics research International*, 979751.

Schrey, A. W., Alvarez, M., Foust, C. M., Kilvitis, H. J., Lee, J. D., Liebl, A. L., Martin, L. B., Richards, C. L., and Robertson, M., 2013. Ecological Epigenetics: Beyond MS-AFLP. *Integrative and Comparative Biology*, 53 (2), 340–350.

Schrider, D. R., Mendes, F. K., Hahn, M. W., and Kern, A. D., 2015. Soft Shoulders Ahead: Spurious Signatures of Soft and Partial Selective Sweeps Result from Linked Hard Sweeps. *Genetics*, 200 (1), 267–84.

Schulz, B., Eckstein, R. L., and Durka, W., 2013. Scoring and analysis of methylation-sensitive amplification polymorphisms for epigenetic population studies. *Molecular Ecology Resources*, 13 (4), 642–653.

Schulz, B., Eckstein, R. L., and Durka, W., 2014. Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. *Molecular ecology*, 23 (14), 3523–37.

Schulz, R. W., de França, L. R., Lareyre, J. J., LeGac, F., Chiarini-Garcia, H., Nobrega, R. H., and Miura, T., 2010. Spermatogenesis in fish. *General and Comparative Endocrinology*, 165 (3), 390–411.

Schuster, S. C., 2008. Next-generation sequencing transforms today's biology. *Nature methods*, 5 (1), 16–8.

Seghers, B., 1974. Schooling behavior in the guppy (Poecilia reticulata): an evolutionary response to predation. *Evolution*, 28 (3), 486–489.

Senn, H., Ogden, R., Cezard, T., Gharbi, K., Iqbal, Z., Johnson, E., Kamps-Hughes, N., Rosell, F., and McEwing, R., 2013. Reference-free SNP discovery for the Eurasian beaver from restriction site-associated DNA paired-end data. *Molecular Ecology*, 22 (11), 3141–3150.

Seymour, A., 1959. Effects of Temperature upon the Formation of Vertebrae and Fin Rays in Young Chinook Salmon. *Transactions of the American Fisheries Society*, 88 (1), 58–69.

Shackell, N., Frank, K., Fisher, J., Petri, B., and Leggett, W., 2009. Decline in top predator body size and changing climate alter trophic structure in an oceanic ecosystem. *Proceedings of the*

*Royal Society B: Biological Sciences*, 282 (1821).

Shao, C., Li, Q., Chen, S., Zhang, P., Lian, J., Hu, Q., Sun, B., Jin, L., Liu, S., Wang, Z., Zhao, H., Jin, Z., Liang, Z., Li, Y., Zheng, Q., Zhang, Y., Wang, J., and Zhang, G., 2014. Epigenetic modification and inheritance in sexual reversal of fish. *Genome Research*, 24 (4), 604–615.

Shao, C., Niu, Y., Rastas, P., Liu, Y., Xie, Z., Li, H., Wang, L., Jiang, Y., Tai, S., Tian, Y., Sakamoto, T., and Chen, S., 2015. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (Paralichthys olivaceus): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 22 (2), 161–70.

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D., and Kingsley, D. M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, 428 (6984), 717–723.

Sharma, E., Künstner, A., Fraser, B. A., Zipprich, G., Kottler, V. A., Henz, S. R., Weigel, D., and Dreyer, C., 2014. Transcriptome assemblies for studying sex-biased gene expression in the guppy, Poecilia reticulata. *BMC genomics*, 15 (1), 400.

Sharp, A. J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S. B., Dupre, Y., and Antonarakis, S. E., 2011. DNA methylation profiles of human active and inactive X chromosomes. *Genome research*, 21 (10), 1592–600.

Sharpe, D. and Hendry, A., 2009. Life history change in commercially exploited fish stocks: an analysis of trends across studies. *Evolutionary Applications*, 2 (3), 260–275.

Shaw, P. W., Carvalho, G. R., Magurran, A. E., and Seghers, B. H., 1991. Population differentiation in Trinidadian guppies ( Poecilia reticulata ): patterns and problems. *Journal of Fish Biology*, 39, 203–209.

Shaw, P. W., Carvalho, G. R., Magurran, A. E., and Seghersf, B. H., 1994. Factors affecting the distribution of genetic variability in the guppy, Poecilia reticulata. *Journal of Fish Biology*, 45 (5), 875–888.

Shaw, R. G. and Etterson, J. R., 2012. Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. *The New phytologist*, 195 (4), 752–65.

Shen, J. C., Rideout, W. M., and Jones, P. A., 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic acids research*, 22 (6), 972–976.

Shen, L. and Waterland, R. A., 2007. Methods of DNA methylation analysis. *Current opinion in clinical nutrition and metabolic care*, 10 (5), 576–81.

Shen, X., Yang, G., and Liao, M., 2007. Development of 51 genomic microsatellite DNA markers of guppy (Poecilia reticulata) and their application in closely related species. *Molecular Ecology Notes*, 7 (2), 302–306.

Shendure, J. and Ji, H., 2008. Next-generation DNA sequencing. *Nature biotechnology*, 26 (10), 1135–45.

Shepherd, B. S., Sakamoto, T., Nishioka, R. S., Richman, N. H., Mori, I., Madsen, S. S., Chen, T. T., Hirano, T., Bern, H. A., and Grau, E. G., 1997. Somatotropic actions of the homologous growth hormone and prolactins in the euryhaline teleost, the tilapia, Oreochromis mossambicus. *Proceedings of the National Academy of Sciences*, 94 (5), 2068–2072.

Shi, T., Seligson, D., Belldegrun, A. S., Palotie, A., and Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 18 (4), 547–57.

Shikano, T. and Taniguchi, N., 2003. DNA markers for estimation of inbreeding depression and heterosis in the guppy Poecilia reticulata. *Aquaculture Research*, 34 (11), 905–911.

Shimada, Y., Shikano, T., Kuparinen, A., Gonda, A., Leinonen, T., and Merilä, J., 2011. Quantitative genetics of body size and timing of maturation in two nine-spined stickleback (Pungitius pungitius) populations. *PloS one*, 6 (12), e28859.

Shimoda, N., Yamakoshi, K., Miyake, A., and Takeda, H., 2005. Identification of a gene required for

de novo DNA methylation of the zebrafish no tail gene. *Developmental dynamics : an official publication of the American Association of Anatomists*, 233 (4), 1509–16.

Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C., 2006. Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response. *Genes & development*, 20 (22), 3079–83.

Shokralla, S., Spall, J. L., Gibson, J. F., and Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular ecology*, 21 (8), 1794–805.

Shull, G. H., 1908. The Composition of a Field of Maize. *Journal of Heredity*, os-4 (1), 296–301.

Silliman, R. P., 1975. Selective and unselective exploitation of experimental populations of Tilapia mossambica. *Fishery Bulletin*, 73 (3), 495–507.

Sinclair, A., 2002. Measuring changes in the direction and magnitude of size-selective mortality in a commercial fish population. *Canadian Journal of …*.

Sinclair, A. F., Swain, D. P., and Hanson, J. M., 2001. Measuring changes in the direction and magnitude of size-selective mortality in a commercial fish population. *Canadian Journal of Fisheries and Aquatic Sciences*, 67 (5), 854–865.

Sittaramane, V. and Chandrasekhar, A., 2008. Expression of unconventional myosin genes during neuronal development in zebrafish. *Gene expression patterns : GEP*, 8 (3), 161–70.

Skelly, D., 2010. A climate for contemporary evolution. *BMC biology*, 8 (1), 136.

Skinner, M., 2010. Metabolic disorders: Fathers' nutritional legacy. *Nature*, 467, 922–923.

Slate, J., 2005. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular ecology*, 14 (2), 363–79.

Slate, J., Pemberton, J., and Visscher, P., 1999. Power to detect QTL in a free-living polygynous population. *Heredity*, 83, 327–336.

Slate, J., Visscher, P., MacGregor, S., Stevens, D., Tate, M., and Pemberton, J., 2002. A genome scan for quantitative trait loci in a wild population of red deer (Cervus elaphus). *Genetics*, 162 (4), 1863–1873.

Slomko, H., Heo, H., and Einstein, F., 2012. Minireview: epigenetics of obesity and diabetes in humans. *Endocrinology*, 153 (3).

Smith, A., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., Tylavsky, F. A., and Conneely, K. N., 2014. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC genomics*, 15 (1), 145.

Smith, G., Smith, C., Kenny, J. G., Chaudhuri, R. R., and Ritchie, M. G., 2014. Genome-Wide DNA Methylation Patterns in Wild Samples of Two Morphotypes of Threespine Stickleback (Gasterosteus aculeatus). *Molecular Biology and Evolution*, 32 (4), 888–895.

Smith, J. M. and Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetical research*, 23 (1), 23–35.

Smith, P., Bartley, D., Harvey, B., and Pullin, R., 2007. Issues, status and trends in deep-sea fishery genetic resources. *In*: *Workshop on Status and Trends in Aquatic Genetic Resources: A Basis for International Policy*. Food and Agriculture of the United Nations, 81–108.

Smith, T. H. L., Collins, T. M., and McGowan, R. A., 2011. Expression of the dnmt3 genes in zebrafish development: similarity to Dnmt3a and Dnmt3b. *Development genes and evolution*, 220 (11-12), 347–53.

Smith, T. H. L., Dueck, C. C., Mhanni, A. A., and McGowan, R. A., 2005. Novel splice variants associated with one of the zebrafish dnmt3 genes. *BMC developmental biology*, 5 (1), 23.

Smith, Z. D. and Meissner, A., 2013. DNA methylation: roles in mammalian development. *Nature reviews. Genetics*, 14 (3), 204–20.

Smit-Mcbride, Z., Moya, A., and Ayala, F. J., 1988. Linkage Disequilibrium in Natural and Experimental Populations of Drosophila-Melanogaster. *Genetics*, 120 (4), 1043–1051.

Sollars, V., Lu, X., Xiao, L., and Wang, X., 2003. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genetics*, 33, 70–74.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., Normandeau, É., Laroche, J., Larose, S., Jean, M., and Belzile, F., 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PloS one*, 8 (1), e54603.

Song, Z. J., Soller, M., and Genizi, A., 1999. The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. *Genetical Research*, 73 (01), 61–73.

Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., and Neale, D. B., 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, 9 (4), 901–911.

Spielman, R. and Ewens, W., 1996. The TDT and other family-based tests for linkage disequilibrium and association. *American journal of human genetics*, 59 (5), 983–989.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P., and Slate, J., 2010. Adaptation genomics: The next generation. *Trends in Ecology & Evolution*, 25 (12), 705–712.

Stearns, F. W., 2010. One hundred years of pleiotropy: a retrospective. *Genetics*, 186 (3), 767–73.

Stearns, S. and Koella, J., 1986. The evolution of phenotypic plasticity in life-history traits: predictions of reaction norms for age and size at maturity. *Evolution*, 40 (5), 893–913.

Stephan, J., Stegle, O., and Beyer, A., 2015. A random forest approach to capture genetic effects in the presence of population structure. *Nature communications*, 6, 7432.

Stephan, W., 2015. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, n/a–n/a.

Stephenson, J. F. J., van Oosterhout, C., and Cable, J., 2015. Pace of life, predators and parasites: predator-induced life-history evolution in Trinidadian guppies predicts decrease in parasite tolerance. *Biology Letters*, 11 (11), 20150806.

Stephenson, J. and Oosterhout, C. Van, 2015. Parasites of Trinidadian guppies: evidence for sex- and age-specific trait-mediated indirect effects of predators. *Ecology*, 96 (2), 489–498.

Steward, N., Ito, M., and Yamaguchi, Y., 2002. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *Journal of Biological Chemistry*, 277, 37741–37746.

Stewart, C. B., Schilling, J. W., and Wilson, A. C., 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature*, 330 (6146), 401–4.

Stiling, P., Moon, D., Hunter, M., Colson, J., Rossi, A., Hymus, G., and Drake, B., 2003. Elevated CO2 lowers relative and absolute herbivore density across all species of a scrub-oak forest. *Oecologia*, 134 (1), 82–87.

Stockwell, C. A., Hendry, A. P., and Kinnison, M. T., 2003. Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, 18 (2), 94–101.

Stokes, K. and Law, R., 2000. Fishing as an evolutionary force: 'Evolution'of fisheries science'. *Marine ecology. Progress series*, 208, 307–309.

Strandabø, R. a U., Hodne, K., Ager-Wick, E., Sand, O., Weltzien, F.-A., and Haug, T. M., 2013. Signal transduction involved in GnRH2-stimulation of identified LH-producing gonadotropes from lhb-GFP transgenic medaka (Oryzias latipes). *Molecular and cellular endocrinology*, 372 (1-2), 128–39.

Strauss, S. Y., Lau, J. a., and Carroll, S. P., 2006. Evolutionary responses of natives to introduced species: what do introductions tell us about natural communities? *Ecology Letters*, 9 (3), 357–374.

Suk, H. Y. and Neff, B. D., 2009. Microsatellite genetic differentiation among populations of the Trinidadian guppy. *Heredity*, 102 (5), 425–434.

Sullam, K. E., Dalton, C. M., Russell, J. a., Kilham, S. S., El-Sabaawi, R., German, D. P., and Flecker, A. S., 2014. Changes in digestive traits and body nutritional composition accommodate a trophic niche shift in Trinidadian guppies. *Oecologia*, 177 (1), 245–257.

Sun, Y., Hou, R., Fu, X., Sun, C., Wang, S., Wang, C., Li, N., Zhang, L., and Bao, Z., 2014.

Genome-Wide Analysis of DNA Methylation in Five Tissues of Zhikong Scallop, Chlamys farreri. *PLoS ONE*, 9 (1), e86232.

Suzuki, M. and Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, 9, 465–476.

Swain, D. P., Sinclair, A. F., and Mark Hanson, J., 2007. Evolutionary response to size-selective mortality in an exploited fish population. *Proceedings. Biological sciences / The Royal Society*, 274 (1613), 1015–1022.

Swaney, W., Kendal, J., Capon, H., Brown, C., and Laland, K., 2001. Familiarity facilitates social learning of foraging behaviour in the guppy. *Animal Behaviour*, 62 (3), 591–598.

Swinnen, S., Schaerlaekens, K., Pais, T., Claesen, J., Hubmann, G., Yang, Y., Demeke, M., Foulquié-Moreno, M. R., Goovaerts, A., Souvereyns, K., Clement, L., Dumortier, F., and Thevelein, J. M., 2012. Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome research*, 22 (5), 975–84.

Szalma, S. J., Hostert, B. M., Ledeaux, J. R., Stuber, C. W., and Holland, J. B., 2007. QTL mapping with near-isogenic lines in maize. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 114 (7), 1211–28.

Tabor, H. K., Risch, N. J., and Myers, R. M., 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*, 3 (5), 391–397.

Takahashi, T., Sota, T., and Hori, M., 2013. Genetic basis of male colour dimorphism in a Lake Tanganyika cichlid fish. *Molecular ecology*, 22 (11), 3049–60.

Takasuga, A., Watanabe, T., Mizoguchi, Y., Hirano, T., Ihara, N., Takano, A., Yokouchi, K., Fujikawa, A., Chiba, K., Kobayashi, N., Tatsuda, K., Oe, T., Furukawa-Kuroiwa, M., Nishimura-Abe, A., Fujita, T., Inoue, K., Mizoshita, K., Ogino, A., and Sugimoto, Y., 2007. Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. *Mammalian genome : official journal of the International Mammalian Genome Society*, 18 (2), 125–36.

Takata, M., Kishima, Y., and Sano, Y., 2005. DNA methylation polymorphisms in rice and wild rice strains: detection of epigenetic markers. *Breeding Science*, 55 (1), 57–63.

Tambasco, D. D., Paz, C. C. P., Tambasco-Studart, M., Pereira, A. P., Alencar, M. M., Freitas, A. R., Coutinho, L. L., Packer, I. U., and Regitano, L. C. A., 2003. Candidate genes for growth traits in beef cattle crosses Bos taurus x Bos indicus. *Journal of Animal Breeding and Genetics*, 120 (1), 51–56.

Tang, R., Sinnwell, J. P., Li, J., Rider, D. N., de Andrade, M., and Biernacka, J. M., 2009. Identification of genes and haplotypes that predict rheumatoid arthritis using random forests. *BMC Proceedings*, 3 (Suppl 7), S68.

Tanksley, S. D., Grandillo, S., Fulton, T. M., Zamir, D., Eshed, Y., Petiard, V., Lopez, J., and Beck-Bunn, T., 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 92 (2), 213–24.

Tao, W. J. and Boulding, E. G., 2003. Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (Salvelinus alpinus L.). *Heredity*, 91 (1), 60–69.

Teer, J. K., Bonnycastle, L. L., Chines, P. S., Hansen, N. F., Aoyama, N., Swift, A. J., Abaan, H. O., Albert, T. J., Margulies, E. H., Green, E. D., Collins, F. S., Mullikin, J. C., and Biesecker, L. G., 2010. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome research*, 20 (10), 1420–31.

Teer, J. K. and Mullikin, J. C., 2010. Exome sequencing: the sweet spot before whole genomes. *Human molecular genetics*, 19 (R2), R145–51.

Templeton, C. and Shriner, W., 2004. Multiple selection pressures influence Trinidadian guppy (Poecilia reticulata) antipredator behavior. *Behavioral Ecology*.

Tenaillon, M. and Sawkins, M., 2001. Patterns of DNA sequence polymorphism along chromosome

1 of maize (Zea mays ssp. mays L.). *Proceedings of the National Academy of Sciences*, 98 (16).

Thelen, T., 1991. Effects of harvest on antlers of simulated populations of elk. *The Journal of wildlife management*, 55 (2), 243–249.

Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M., and Kitts, P., 2013. Eukaryotic Genome Annotation Pipeline. *In*: *The NCBI Handbook*. National Center for Biotechnology Information (US).

Thorpe, J., 2007. Maturation responses of salmonids to changing developmental opportunities. *Marine Ecology Progress Series*, 335, 285–288.

Thresher, R. E., Koslow, J. A., Morison, A. K., and Smith, D. C., 2007. Depth-mediated reversal of the effects of climate change on long-term growth rates of exploited marine fish. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (18), 7461–5.

Thurber, C. S., Jia, M. H., Jia, Y., and Caicedo, A. L., 2013. Similar traits, different genes? Examining convergent evolution in related weedy rice populations. *Molecular ecology*, 22 (3), 685–98.

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., and Bird, C. E., 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1, e203.

Tost, J. and Gut, I. G., 2007. DNA methylation analysis by pyrosequencing. *Nature protocols*, 2 (9), 2265–75.

Traut, W. and Winking, H., 2001. Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Research*, 9, 659–672.

Travis, J., Reznick, D., and Bassar, R. D., 2014. *Eco-Evolutionary Dynamics*. Advances in Ecological Research. Elsevier.

Travis, J., Reznick, D., Bassar, R. D., López-Sepulcre, A., Ferriere, R., and Coulson, T., 2014. *Do Eco-Evo Feedbacks Help Us Understand Nature? Answers From Studies of the Trinidadian Guppy*. Elsevier.

Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C.-C., Febrer, M., and Uauy, C., 2012. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC plant biology*, 12, 14.

Tripathi, N., Hoffmann, M., Weigel, D., and Dreyer, C., 2009. Linkage analysis reveals the independent origin of Poeciliid sex chromosomes and a case of atypical sex inheritance in the guppy (Poecilia reticulata). *Genetics*, 182 (1), 365–74.

Tripathi, N., Hoffmann, M., Willing, E.-M., Lanz, C., Weigel, D., and Dreyer, C., 2009. Genetic linkage map of the guppy, Poecilia reticulata, and quantitative trait loci analysis of male size and colour variation. *Proceedings. Biological sciences / The Royal Society*, 276 (1665), 2195–208.

Trippel, E., 1995. Age at maturity as a stress indicator in fisheries. *Bioscience*, 45 (11), 759–771.

Turker, M. S., 2002. Gene silencing in mammalian cells and the spread of DNA methylation. *Oncogene*, 21 (35), 5388–93.

Turner, A., Beales, J., Faure, S., Dunford, R. P., and Laurie, D. A., 2005. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science (New York, N.Y.)*, 310 (5750), 1031–4.

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., and Nuzhdin, S. V, 2010. Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature genetics*, 42 (3), 260–3.

Turner, T. L., Stewart, A. D., Fields, A. T., Rice, W. R., and Tarone, A. M., 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. *PLoS genetics*, 7 (3), e1001336.

Unniappan, S. and Peter, R. E., 2005. Structure, distribution and physiological functions of ghrelin in fish. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*,

140 (4), 396–408.

Uusi-Heikkilä, S., Kuparinen, A., Wolter, C., Meinelt, T., O'Toole, A. C., and Arlinghaus, R., 2011. Experimental assessment of the probabilistic maturation reaction norm: condition matters. *Proceedings. Biological sciences / The Royal Society*, 278 (1706), 709–17.

Uusi-Heikkilä, S., Whiteley, A. R., Kuparinen, A., Matsumura, S., Venturelli, P. a., Wolter, C., Slate, J., Primmer, C. R., Meinelt, T., Killen, S. S., Bierbach, D., Polverino, G., Ludwig, A., and Arlinghaus, R., 2015. The evolutionary legacy of size-selective harvesting extends from genes to populations. *Evolutionary Applications*, n/a–n/a.

Uusi-Heikkilä, S., Wolter, C., Klefoth, T., and Arlinghaus, R., 2008. A behavioral perspective on fishing-induced evolution. *Trends in ecology & evolution*, 23 (8), 419–21.

Vainikka, A., Gardmark, A., Bland, B., and Hjelm, J., 2008. Two- and three-dimensional maturation reaction norms for the eastern Baltic cod, Gadus morhua. *ICES Journal of Marine Science*, 66 (2), 248–257.

Vandegehuchte, M. B. and Janssen, C. R., 2011. Epigenetics and its implications for ecotoxicology. *Ecotoxicology (London, England)*, 20 (3), 607–24.

Vandepitte, K., Honnay, O., Mergeay, J., Breyne, P., Roldán-Ruiz, I., and De Meyer, T., 2013. SNP discovery using Paired-End RAD-tag sequencing on pooled genomic DNA of Sisymbrium austriacum (Brassicaceae). *Molecular Ecology Resources*, 13 (2), 269–275.

Varriale, A., 2014. DNA Methylation, Epigenetics, and Evolution in Vertebrates: Facts and Challenges. *International journal of evolutionary biology*, 2014, 1–7.

Varriale, A. and Bernardi, G., 2006a. DNA methylation and body temperature in fishes. *Gene*, 385, 111–121.

Varriale, A. and Bernardi, G., 2006b. DNA methylation in reptiles. *Gene*, 385, 122–127.

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A., 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in biotechnology*, 27 (9), 522–30.

Vaughn, M. W., Tanurdzić, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P. D., Dedhia, N., McCombie, W. R., Agier, N., Bulski, A., Colot, V., Doerge, R. W., and Martienssen, R. A., 2007. Epigenetic natural variation in Arabidopsis thaliana. *PLoS biology*, 5 (7), e174.

Velan, A., Hulata, G., Ron, M., Slosman, T., Shirak, A., and Cnaani, A., 2015. Association between polymorphism in the Prolactin I promoter and growth of tilapia in saline-water. *Aquaculture Reports*, 1, 5–9.

Vergeer, P., Wagemaker, N. C. A. M., and Ouborg, N. J., 2012. Evidence for an epigenetic role in inbreeding depression. *Biology letters*, 8 (5), 798–801.

Verhoeven, K. J. F., Jansen, J. J., van Dijk, P. J., and Biere, A., 2010. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytologist*, 185 (4), 1108–1118.

Vilas, a., Pérez-Figueroa, a., and Caballero, a., 2012. A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology*, 25 (7), 1364–1376.

de Villemereuil, P. B. and López-Sepulcre, A., 2011. Consumer functional responses under intra- and inter-specific interference competition. *Ecological Modelling*, 222 (3), 419–426.

Viñas, J. and Piferrer, F., 2008. Stage-specific gene expression during fish spermatogenesis as determined by laser-capture microdissection and quantitative-PCR in sea bass (Dicentrarchus labrax) gonads. *Biology of reproduction*, 79 (July), 738–747.

Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J., 2012. Five years of GWAS discovery. *American journal of human genetics*, 90 (1), 7–24.

Visscher, P. M., Hill, W. G., and Wray, N. R., 2008. Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, 9 (4), 255–266.

Volodarsky, M., Markus, B., Cohen, I., Staretz-Chacham, O., Flusser, H., Landau, D., Shelef, I., Langer, Y., and Birk, O. S., 2013. A Deletion Mutation in TMEM38B Associated with

Autosomal Recessive Osteogenesis Imperfecta. *Human Mutation*, 34 (4), n/a–n/a.

Wallace, J. C., Kolbeinshavn, A. G., and Reinsnes, T. G., 1988. The effects of stocking density on early growth in Arctic charr, Salvelinus alpinus (L.). *Aquaculture*, 73 (1-4), 101–110.

van Walraven, L., Mollet, F. M., van Damme, C. J. G., and Rijnsdorp, A. D., 2010. Fisheries-induced evolution in growth, maturation and reproductive investment of the sexually dimorphic North Sea plaice (Pleuronectes platessa L.). *Journal of Sea Research*, 64 (1-2), 85–93.

Walsh, M. R., Munch, S. B., Chiba, S., and Conover, D. O., 2006. Maladaptive changes in multiple traits caused by fishing: impediments to population recovery. *Ecology letters*, 9 (2), 142–8.

Walsh, M. R. and Post, D. M., 2011. Interpopulation variation in a fish predator drives evolutionary divergence in prey in lakes. *Proceedings. Biological sciences / The Royal Society*, 278 (1718), 2628–37.

Walsh, M. R. and Reznick, D. N., 2011. Experimentally Induced Life-History Evolution in a Killifish in Response To the Introduction of Guppies. *Evolution*, 65 (4), 1021–1036.

Wang, B., Guo, W., Zhu, X., Wu, Y., Huang, N., and Zhang, T., 2006. QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton. *Euphytica*.

Wang, J., Marowsky, N. C., and Fan, C., 2014. Divergence of Gene Body DNA Methylation and Evolution of Plant Duplicate Genes. *PLoS ONE*, 9 (10), e110357.

Wang, S., Meyer, E., McKay, J. K., and Matz, M. V, 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature methods*, 9 (8), 808–10.

Wang, W. and Barratt, B., 2005. Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics*, 6, 109–118.

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M., 2015. Exome Sequencing: Current and Future Perspectives. *G3 (Bethesda, Md.)*, 5 (8), 1543–50.

Watanabe, T., Nakajima, M., Yoshida, M., and Taniguchi, N., 2004. Construction of six linkage groups in the guppy (Poecilia reticulata). *Animal genetics*, 35 (2), 147–8.

Watanabe, T., Yoshida, M., Nakajima, M., and Taniguchi, N., 2003. Isolation and characterization of 43 microsatellite DNA markers for guppy (Poecilia reticulata). *Molecular Ecology Notes*, 3 (4), 487–490.

Weir, B. and Cockerham, C., 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38 (6), 1358 – 1370.

Weir, B. S., 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates.

Wellenreuther, M. and Hansson, B., 2016. Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends in genetics : TIG*, 32 (3), 155–164.

Wenzel, M. a. and Piertney, S. B., 2014. Fine-scale population epigenetic structure in relation to gastrointestinal parasite load in red grouse (Lagopus lagopus scotica). *Molecular Ecology*, 4256–4273.

Werner, J., Borevitz, J., and Uhlenhaut, N., 2005. FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. *Genetics*, 170 (3), 1197–1207.

Whitehead, A., Triant, D. A., Champlin, D., and Nacci, D., 2010. Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population. *Molecular ecology*, 19 (23), 5186–203.

Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S., and Buckler, E. S., 2002. Genetic diversity and selection in the maize starch pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (20), 12959–62.

Whittington, C. M. and Wilson, A. B., 2013. The role of prolactin in fish reproduction. *General and Comparative Endocrinology*, 191, 123–136.

Whittle, C., Otto, S., Johnston, M., and Krochko, J., 2009. Adaptive epigenetic memory of ancestral temperature regime in Arabidopsis thaliana. *Botany*, 87 (6), 650–657.

Wielgoss, S., Barrick, J. E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C.,

Lenski, R. E., and Schneider, D., 2011. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With Escherichia coli. *G3 (Bethesda, Md.)*, 1 (3), 183–186.

Wiens, K. E., Crispo, E., and Chapman, L. J., 2014. Phenotypic plasticity is maintained despite geographical isolation in an African cichlid fish, Pseudocrenilabrus multicolor. *Integrative zoology*, 9 (1), 85–96.

Wierer, M., Schrey, A. K., Kühne, R., Ulbrich, S. E., and Meyer, H. H. D., 2012. A single glycine-alanine exchange directs ligand specificity of the elephant progestin receptor. *PloS one*, 7 (11), e50350.

van Wijk, S. J., 2011a. Fisheries-induced evolution : a genetic approach using selection experiments on Poecilia reticulata. Bangor University.

van Wijk, S. J., 2011b. Fisheries-induced evolution : a genetic approach using selection experiments on Poecilia reticulata.

van Wijk, S. J., Taylor, M. I., Creer, S., Dreyer, C., Rodrigues, F. M., Ramnarine, I. W., van Oosterhout, C., and Carvalho, G. R., 2013. Experimental harvesting of fish populations drives genetically based shifts in body size and maturation. *Frontiers in Ecology and the Environment*, 11 (4), 181–187.

Willing, E.-M., Bentzen, P., van Oosterhout, C., Hoffmann, M., Cable, J., Breden, F., Weigel, D., and Dreyer, C., 2010. Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular ecology*, 19 (5), 968–84.

Willing, E.-M., Hoffmann, M., Klein, J. D., Weigel, D., and Dreyer, C., 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics (Oxford, England)*, 27 (16), 2187–93.

Winge, O., 1937. Succession of broods in Lebistes. *Nature*, (140), 467.

Winge, Ö., 1922a. One-sided masculine and sex-linked inheritance inLebistes reticulatus. *Journal of Genetics*, 12 (2), 145–162.

Winge, Ö., 1922b. A peculiar mode of inheritance and its cytological explanation. *Journal of Genetics*, 12 (2), 137–144.

Winge, Ö., 1927. The location of eighteen genes inLebistes reticulatus. *Journal of Genetics*, 18 (1), 1–43.

Winge, Ø. and Ditlevsen, E., 1947. Colour inheritance and sex determination in Lebistes. *Heredity*, 1 (1), 65–83.

Wolff, G. L., Kodell, R. L., Moore, S. R., and Cooney, C. A., 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J*, 12 (11), 949–957.

Wootton, R., 1999. *Ecology of teleost fishes*. Springer.

Wright, K. J., Baye, L. M., Olivier-mason, A., Mukhopadhyay, S., Sang, L., Kwong, M., Wang, W., Pretorius, P. R., Sheffield, V. C., Sengupta, P., Slusarski, D. C., and Jackson, P. K., 2011. An ARL3 – UNC119 – RP2 GTPase cycle targets myristoylated NPHP3 to the primary cilium, 2347–2360.

Wright, P., Gibb, F., Gibb, I., and Millar, C., 2011. Reproductive investment in the North Sea haddock: temporal and spatial variation. *Marine Ecology Progress Series*, 432, 149–160.

Wu, H. B., Kumar, A., Tsai, W. C., Mascarenhas, D., Healey, J., and Rechler, M. M., 2000. Characterization of the inhibition of DNA synthesis in proliferating mink lung epithelial cells by insulin-like growth factor binding protein-3. *Journal of cellular biochemistry*, 77 (2), 288–97.

Wu, J., Li, L.-T., Li, M., Khan, M. A., Li, X.-G., Chen, H., Yin, H., and Zhang, S.-L., 2014. High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *Journal of experimental botany*, 65 (20), 5771–81.

Wu, S.-F., Zhang, H., Hammoud, S. S., Potok, M., Nix, D. A., Jones, D. A., and Cairns, B. R., 2011. *The Zebrafish: Genetics, Genomics and Informatics*. Methods in Cell Biology. Elsevier.

Xia, L., Ma, S., Zhang, Y., Wang, T., Zhou, M., Wang, Z., and Zhang, J., 2015. Daily variation in global and local DNA methylation in mouse livers. *PloS one*, 10 (2), e0118101.

Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L., and Ren, B., 2012. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*, 148 (4), 816–31.

Xiong, L. Z., Xu, C. G., Maroof, M. A. S., and Zhang, Q., 1999. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Molecular and General Genetics MGG*, 261 (3), 439–446.

Xu, M., Tantisira, K. G., Wu, A., Litonjua, A. A., Chu, J., Himes, B. E., Damask, A., and Weiss, S. T., 2011. Genome Wide Association Study to predict severe asthma exacerbations in children using random forests classifiers. *BMC medical genetics*, 12 (1), 90.

Xu, Q., Feng, C. Y., Hori, T. S., Plouffe, D. a, Buchanan, J. T., and Rise, M. L., 2013. Family-specific differences in growth rate and hepatic gene expression in juvenile triploid growth hormone (GH) transgenic Atlantic salmon (Salmo salar). *Comparative biochemistry and physiology. Part D, Genomics & proteomics*, 8 (4), 317–33.

Xu, Y. X., Zhu, Z. Y., Lo, L. C., Wang, C. M., Lin, G., Feng, F., and Yue, G. H., 2006. Characterization of two parvalbumin genes and their association with growth traits in Asian seabass (Lates calcarifer). *Animal Genetics*, 37 (3), 266–268.

Yamada, N. and Yonezawa, S., 2013. Understanding Epigenetic Status: DNA Methylation and Cancer. *Journal of Cancer Biology & Research*.

Yamakoshi, K. and Shimoda, N., 2003. De novo DNA methylation at the CpG island of the zebrafish no tail gene. *Genesis (New York, N.Y. : 2000)*, 37 (4), 195–202.

Yan, X., Liu, X., Li, J., Zhao, H., Li, Q., Wang, Y., and Yuan, C., 2014. Genetic and epigenetic diversity of wild and cultivated soybean in local populations in Northern Huang Huai region of China, 7 (6), 415–423.

Yang, C., Zhang, M., Niu, W., Yang, R., Zhang, Y., Qiu, Z., Sun, B., and Zhao, Z., 2011. Analysis of DNA methylation in various swine tissues. *PloS one*, 6 (1), e16229.

Yang, H., Chang, F., You, C., Cui, J., Zhu, G., Wang, L., Zheng, Y., Qi, J., and Ma, H., 2015. Whole-genome DNA methylation patterns and complex associations with gene structure and expression during flower development in Arabidopsis. *The Plant Journal*, 81 (2), 268–281.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., and Visscher, P. M., 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, 42 (7), 565–9.

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M. G., Hill, W. G., Landi, M. T., Alonso, A., Lettre, G., Lin, P., Ling, H., Lowe, W., Mathias, R. A., Melbye, M., Pugh, E., Cornelis, M. C., Weir, B. S., Goddard, M. E., and Visscher, P. M., 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics*, 43 (6), 519–25.

Yang, J., Zhu, J., and Williams, R. W., 2007. Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics (Oxford, England)*, 23 (12), 1527–36.

Young, W. P., Wheeler, P. A., Coryell, V. H., Keim, P., and Thorgaard, G. H., 1998. A Detailed Linkage Map of Rainbow Trout Produced Using Doubled Haploids. *Genetics*, 148 (2), 839–850.

Yu, J. and Buckler, E. S., 2006. Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, 17 (0958-1669 (Print)), 155–160.

Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Bürkle, A., and Caiafa, P., 2015. Reconfiguration of DNA methylation in aging. *Mechanisms of Ageing and Development*, 151, 60–70.

Zandonà, E., Auer, S. K., Kilham, S. S., Howard, J. L., López-Sepulcre, A., O'Connor, M. P., Bassar, R. D., Osorio, A., Pringle, C. M., and Reznick, D. N., 2011. Diet quality and prey selectivity correlate with life histories and predation regime in Trinidadian guppies. *Functional Ecology*, 25 (5), 964–973.

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A.,

Kucharczyk, K., Płoski, R., and Branicki, W., 2015. Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic science international. Genetics*, 17, 173–9.

Zbinden, M., Haag, C. R., and Ebert, D., 2008. Experimental evolution of field populations of Daphnia magna in response to parasite treatment. *Journal of evolutionary biology*, 21 (4), 1068–78.

Zhang, K., Tian, J., Zhao, L., and Wang, S., 2008. Mapping QTLs with epistatic effects and QTL x environment interactions for plant height using a doubled haploid population in cultivated wheat. *Journal of genetics and genomics = Yi chuan xue bao*, 35 (2), 119–27.

Zheng, P., Allen, W., Roesler, K., and Williams, M., 2008. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nature Genetics*.

Zhivotovsky, L. A., 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular ecology*, 8 (6), 907–13.

Zhong, H., Xiao, J., Chen, W., Zhou, Y., Tang, Z., Guo, Z., Luo, Y., Lin, Z., Gan, X., and Zhang, M., 2014. DNA methylation of pituitary growth hormone is involved in male growth superiority of Nile tilapia (Oreochromis niloticus). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 171, 42–48.

Zhou, X. and Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature genetics*.

Zhu, J.-K., 2009. Active DNA demethylation mediated by DNA glycosylases. *Annual review of genetics*, 43, 143–66.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., and Henikoff, S., 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature genetics*, 39 (1), 61–9.

Zimmermann, F., Jørgensen, C., and Wilberg, M., 2015. Bioeconomic consequences of fishing-induced evolution: a model predicts limited impact on net present value. *Canadian Journal of Fisheries and Aquatic Sciences*, 72 (4), 612–624.

# Appendices

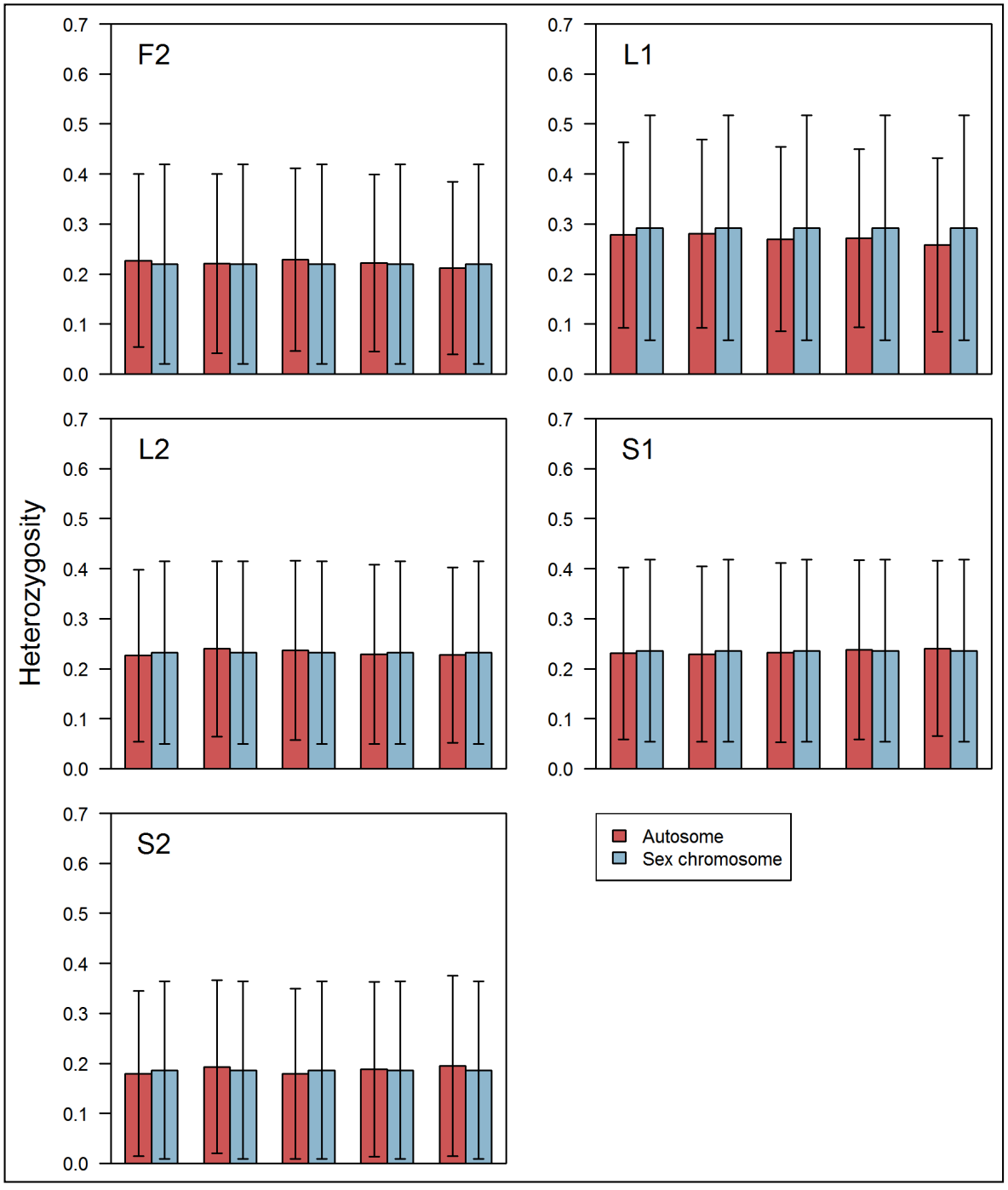
## Appendix I: Barcodes and adaptor sequences for RAD sequencing

| Sample | Barcode | Barcode length | Adaptor and barcode |
|---|---|---|---|
| L1 51 | CAGATA | 6 | TATCTGagatcggaagagcgtcgtg |
| L1 52 | GAAGTG | 6 | CACTTCagatcggaagagcgtcgtg |
| L1 53 | TAGCGGAT | 8 | ATCCGCTAagatcggaagagcgtcgtg |
| L1 54 | TATTCGCAT | 9 | ATGCGAATAagatcggaagagcgtcgtg |
| L1 55 | ATAGAT | 6 | ATCTATagatcggaagagcgtcgtg |
| L1 56 | CCGAACA | 7 | TGTTCGGagatcggaagagcgtcgtg |
| L1 57 | GGAAGACAT | 9 | ATGTCTTCCagatcggaagagcgtcgtg |
| L1 58 | AACGCACATT | 10 | AATGTGCGTTagatcggaagagcgtcgtg |
| L1 59 | GAGCGACAT | 9 | ATGTCGCTCagatcggaagagcgtcgtg |
| L1 60 | CCTTGCCATT | 10 | AATGGCAAGGagatcggaagagcgtcgtg |
| L1 61 | GGTATA | 6 | TATACCagatcggaagagcgtcgtg |
| L1 62 | TCTTGG | 6 | CCAAGAagatcggaagagcgtcgtg |
| L1 63 | GGTGT | 5 | ACACCagatcggaagagcgtcgtg |
| L1 64 | GGATA | 5 | TATCCagatcggaagagcgtcgtg |
| L1 65 | CTAAGCA | 7 | TGCTTAGagatcggaagagcgtcgtg |
| L1 66 | ATTAT | 5 | ATAATagatcggaagagcgtcgtg |
| L1 67 | GCGCTCA | 7 | TGAGCGCagatcggaagagcgtcgtg |
| L1 68 | ACTGCGAT | 8 | ATCGCAGTagatcggaagagcgtcgtg |
| L1 69 | TTCGTT | 6 | AACGAAagatcggaagagcgtcgtg |
| L1 70 | ATATAA | 6 | TTATATagatcggaagagcgtcgtg |
| L1 71 | TGGCAACAGA | 10 | TCTGTTGCCAagatcggaagagcgtcgtg |
| L1 72 | CTCGTCG | 7 | CGACGAGagatcggaagagcgtcgtg |
| L1 73 | GCCTACCT | 8 | AGGTAGGCagatcggaagagcgtcgtg |
| L1 74 | CACCA | 5 | TGGTGagatcggaagagcgtcgtg |
| L1 75 | AATTAG | 6 | CTAATTagatcggaagagcgtcgtg |
| L1 76 | GGAACGA | 7 | TCGTTCCagatcggaagagcgtcgtg |
| L1 77 | ACTGCT | 6 | AGCAGTagatcggaagagcgtcgtg |
| L1 BS1 | TGCTT | 5 | AAGCAagatcggaagagcgtcgtg |
| L1 BS2 | GCAAGCCAT | 9 | ATGGCTTGCagatcggaagagcgtcgtg |
| L1 BS3 | CGCACCAATT | 10 | AATTGGTGCGagatcggaagagcgtcgtg |
| L1 BS4 | CTCGCGG | 7 | CCGCGAGagatcggaagagcgtcgtg |
| L1 BS5 | AACTGG | 6 | CCAGTTagatcggaagagcgtcgtg |
| L1 BS6 | ATGAGCAA | 8 | TTGCTCATagatcggaagagcgtcgtg |
| L1 BS7 | CTTGA | 5 | TCAAGagatcggaagagcgtcgtg |
| L1 BS8 | GCGTCCT | 7 | AGGACGCagatcggaagagcgtcgtg |
| L1 BS9 | ACCAGGA | 7 | TCCTGGTagatcggaagagcgtcgtg |
| L1 BS10 | CCACTCA | 7 | TGAGTGGagatcggaagagcgtcgtg |
| L1 BS11 | TCACGGAAG | 9 | CTTCCGTGAagatcggaagagcgtcgtg |
| L1 BS12 | TATCA | 5 | TGATAagatcggaagagcgtcgtg |
| L1 BS13 | TAGCCAA | 7 | TTGGCTAagatcggaagagcgtcgtg |
| L2 51 | AATAAGAGT | 9 | ACTCTTATTagatcggaagagcgtcgtg |
| L2 52 | TACAAG_bot | 6 | CTTGTAagatcggaagagcgtcgtg |
| L2 53 | GGTGCACATT | 10 | AATGTGCACCagatcggaagagcgtcgtg |
| L2 54 | CTCTCGCAT | 9 | ATGCGAGAGagatcggaagagcgtcgtg |
| L2 BS1 | CAGAGGT | 7 | ACCTCTGagatcggaagagcgtcgtg |
| L2 BS2 | GCGTACAAT | 9 | ATTGTACGCagatcggaagagcgtcgtg |
| L2 BS3 | ACGCGCG | 7 | CGCGCGTagatcggaagagcgtcgtg |
| L2 BS4 | GTCGCCT | 7 | AGGCGACagatcggaagagcgtcgtg |
| L2 BS5 | AATAACCAA | 9 | TTGGTTATTagatcggaagagcgtcgtg |
| L2 BS6 | AATGAACGA | 9 | TCGTTCATTagatcggaagagcgtcgtg |
| L2 BS7 | ATGGCAA | 7 | TTGCCATagatcggaagagcgtcgtg |

| Sample | Barcode | Barcode length | Adaptor and barcode |
|---|---|---|---|
| L2 BS8 | GAAGCA | 6 | TGCTTCagatcggaagagcgtcgtg |
| L2 BS9 | AACGTGCCT | 9 | AGGCACGTTagatcggaagagcgtcgtg |
| L2 BS10 | CCTCG | 5 | CGAGGagatcggaagagcgtcgtg |
| S1 51 | CTCAT | 5 | ATGAGagatcggaagagcgtcgtg |
| S1 52 | ACGGTACT | 8 | AGTACCGTagatcggaagagcgtcgtg |
| S1 53 | GCGCCG | 6 | CGGCGCagatcggaagagcgtcgtg |
| S1 54 | CAAGT | 5 | ACTTGagatcggaagagcgtcgtg |
| S1 55 | GGAGTCAAG | 9 | CTTGACTCCagatcggaagagcgtcgtg |
| S1 56 | TGAAT | 5 | ATTCAagatcggaagagcgtcgtg |
| S1 57 | CATAT | 5 | ATATGagatcggaagagcgtcgtg |
| S1 58 | GTGACACAT | 9 | ATGTGTCACagatcggaagagcgtcgtg |
| S1 59 | TATGT | 5 | ACATAagatcggaagagcgtcgtg |
| S1 60 | TGCAGA | 6 | TCTGCAagatcggaagagcgtcgtg |
| S1 61 | CATCTGCCG | 9 | CGGCAGATGagatcggaagagcgtcgtg |
| S1 62 | GGACAG | 6 | CTGTCCagatcggaagagcgtcgtg |
| S1 63 | ATCTGT | 6 | ACAGATagatcggaagagcgtcgtg |
| S1 64 | AAGACGCT | 8 | AGCGTCTTagatcggaagagcgtcgtg |
| S1 65 | GAATGCAATA | 10 | TATTGCATTCagatcggaagagcgtcgtg |
| S1 66 | TAGCAG | 6 | CTGCTAagatcggaagagcgtcgtg |
| S1 67 | CTTAG | 5 | CTAAGagatcggaagagcgtcgtg |
| S1 68 | TTATTACAT | 9 | ATGTAATAAagatcggaagagcgtcgtg |
| S1 69 | GCCAACAAGA | 10 | TCTTGTTGGCagatcggaagagcgtcgtg |
| S1 70 | TGCCGCAT | 8 | ATGCGGCAagatcggaagagcgtcgtg |
| S1 71 | CGTGTCA | 7 | TGACACGagatcggaagagcgtcgtg |
| S1 72 | CAACCACACA | 10 | TGTGTGGTTGagatcggaagagcgtcgtg |
| S1 73 | GCTCCGA | 7 | TCGGAGCagatcggaagagcgtcgtg |
| S1 BS1 | CGTTCA | 6 | TGAACGagatcggaagagcgtcgtg |
| S1 BS2 | CATCACAAG | 9 | CTTGTGATGagatcggaagagcgtcgtg |
| S1 BS4 | TCCAG | 5 | CTGGAagatcggaagagcgtcgtg |
| S1 BS5 | AACTGAAG | 8 | CTTCAGTTagatcggaagagcgtcgtg |
| S1 BS6 | GATTCA | 6 | TGAATCagatcggaagagcgtcgtg |
| S2 51 | CAAGCCAATT | 10 | AATTGGCTTGagatcggaagagcgtcgtg |
| S2 52 | TTGCGCT | 7 | AGCGCAAagatcggaagagcgtcgtg |
| S2 53 | CGCAGACACT | 10 | AGTGTCTGCGagatcggaagagcgtcgtg |
| S2 54 | TGTGGA | 6 | TCCACAagatcggaagagcgtcgtg |
| S2 55 | TGGATA | 6 | TATCCAagatcggaagagcgtcgtg |
| S2 56 | ATAGCGT | 7 | ACGCTATagatcggaagagcgtcgtg |
| S2 57 | CCATAGA | 7 | TCTATGGagatcggaagagcgtcgtg |
| S2 58 | GGCACGCAT | 9 | ATGCGTGCCagatcggaagagcgtcgtg |
| S2 59 | ATTAACAATT | 10 | AATTGTTAATagatcggaagagcgtcgtg |
| S2 60 | CAATA | 5 | TATTGagatcggaagagcgtcgtg |
| S2 61 | TAGTCCAT | 8 | ATGGACTAagatcggaagagcgtcgtg |
| S2 62 | CGTGACCT | 8 | AGGTCACGagatcggaagagcgtcgtg |
| S2 63 | CTTCAGA | 7 | TCTGAAGagatcggaagagcgtcgtg |
| S2 64 | ATCTGCAACA | 10 | TGTTGCAGATagatcggaagagcgtcgtg |
| S2 65 | AAGGA | 5 | TCCTTagatcggaagagcgtcgtg |
| S2 66 | TTACT | 5 | AGTAAagatcggaagagcgtcgtg |
| S2 67 | TTATCCAT | 8 | ATGGATAAagatcggaagagcgtcgtg |
| S2 68 | GGATTG | 6 | CAATCCagatcggaagagcgtcgtg |
| S2 69 | GACGTGA | 7 | TCACGTCagatcggaagagcgtcgtg |

| Sample | Barcode | Barcode length | Adaptor and barcode |
|---|---|---|---|
| S2 70 | GACGGCA | 7 | TGCCGTCAgatcggaagagcgtcgtg |
| S2 BS1 | CGTCTG | 6 | CAGACGAgatcggaagagcgtcgtg |
| S2 BS2 | TCTGA | 5 | TCAGAAgatcggaagagcgtcgtg |
| S2 BS3 | AACTT | 5 | AAGTTAgatcggaagagcgtcgtg |
| S2 BS4 | GAGTCACAAT | 10 | ATTGTGACTCAgatcggaagagcgtcgtg |
| S2 BS5 | CGGTTGCAT | 9 | ATGCAACCGAgatcggaagagcgtcgtg |
| S2 BS6 | GTCCTGCCA | 9 | TGGCAGGACAgatcggaagagcgtcgtg |
| S2 BS7 | GTTACA | 6 | TGTAACAgatcggaagagcgtcgtg |
| S2 BS8 | GCGGA | 5 | TCCGCAgatcggaagagcgtcgtg |
| S2 BS9 | ATGATACG | 8 | CGTATCATAgatcggaagagcgtcgtg |
| F2 A8 | CTGTTG | 6 | CAACAGAgatcggaagagcgtcgtg |
| F2 A15 | TTCAGCCAGT | 10 | ACTGGCTGAAAgatcggaagagcgtcgtg |
| F2 A16 | TCACA | 5 | TGTGAAgatcggaagagcgtcgtg |
| F2 A20 | GTCGT | 5 | ACGACAgatcggaagagcgtcgtg |
| F2 A21 | ACGCTAA | 7 | TTAGCGTAgatcggaagagcgtcgtg |
| F2 A23 | ATAGG | 5 | CCTATAgatcggaagagcgtcgtg |
| F2 A28 | CCTGCCA | 7 | TGGCAGGAgatcggaagagcgtcgtg |
| F2 A31 | TAAGACA | 7 | TGTCTTAAgatcggaagagcgtcgtg |
| F2 A34 | TGAGA | 5 | TCTCAAgatcggaagagcgtcgtg |
| F2 B6 | AATGCAG | 7 | CTGCATTAgatcggaagagcgtcgtg |
| F2 B15 | CCGTGA | 6 | TCACGGAgatcggaagagcgtcgtg |
| F2 B20 | GCCAGACATT | 10 | AATGTCTGGCAgatcggaagagcgtcgtg |
| F2 B26 | GTGCG | 5 | CGCACAgatcggaagagcgtcgtg |
| F2 B27 | TTACACA | 7 | TGTGTAAAgatcggaagagcgtcgtg |
| F2 B28 | CCGTCACAGT | 10 | ACTGTGACGGAgatcggaagagcgtcgtg |
| F2 B32 | CTGTGT | 6 | ACACAGAgatcggaagagcgtcgtg |
| F2 B35 | CGCGCCG | 7 | CGGCGCGAgatcggaagagcgtcgtg |
| F2 C7 | CTAACA | 6 | TGTTAGAgatcggaagagcgtcgtg |
| F2 C10 | TGAAGCAACT | 10 | AGTTGCTTCAAgatcggaagagcgtcgtg |
| F2 C11 | TGACGT | 6 | ACGTCAAgatcggaagagcgtcgtg |
| F2 C12 | ACTGAG | 6 | CTCAGTAgatcggaagagcgtcgtg |
| F2 C18 | GCGCACT | 7 | AGTGCGCAgatcggaagagcgtcgtg |
| F2 C30 | GGTAAGCA | 8 | TGCTTACCAgatcggaagagcgtcgtg |
| F2 C34 | AATCGGAGG | 9 | CCTCCGATTAgatcggaagagcgtcgtg |
| F2 D4 | TGGAGCCT | 8 | AGGCTCCAAgatcggaagagcgtcgtg |
| F2 D9 | GATGGCCAT | 9 | ATGGCCATCAgatcggaagagcgtcgtg |
| F2 D12 | ACAACGCAT | 9 | ATGCGTTGTAgatcggaagagcgtcgtg |
| F2 D14 | GGCGGACGA | 9 | TCGTCCGCCAgatcggaagagcgtcgtg |
| F2 D18 | CCGTACCACT | 10 | AGTGGTACGGAgatcggaagagcgtcgtg |
| F2 D21 | GTAACG | 6 | CGTTACAgatcggaagagcgtcgtg |
| F2 D24 | TCCTCACAT | 9 | ATGTGAGGAAgatcggaagagcgtcgtg |
| F2 D25 | TCGTA | 5 | TACGAAgatcggaagagcgtcgtg |
| F2 E5 | GTATTGACT | 9 | AGTCAATACAgatcggaagagcgtcgtg |
| F2 E6 | GCTCA | 5 | TGAGCAgatcggaagagcgtcgtg |
| F2 E9 | AATGTA | 6 | TACATTAgatcggaagagcgtcgtg |
| F2 E10 | GGAGAGCAT | 9 | ATGCTCTCCAgatcggaagagcgtcgtg |
| F2 E15 | CCATG | 5 | CATGGAgatcggaagagcgtcgtg |
| F2 E23 | CGCTCACACA | 10 | TGTGTGAGCGAgatcggaagagcgtcgtg |
| F2 E33 | TGTTACG | 7 | CGTAACAAgatcggaagagcgtcgtg |
| F2 E37 | GATTGGAAGA | 10 | TCTTCCAATCAgatcggaagagcgtcgtg |

**Appendix II: Plots showing the diversity of the sex chromosome (Chr 12) compared to a subset of autosomal SNPs. Only 2 of the 5 subsets in the F2 population showed a significant difference.**



| Linkage group | Assecion number | Start | Gene ID | Protein | Length | Protein description |
|---|---|---|---|---|---|---|
| LG1 | NC_024331.1 | 5166547 | 103479325 | XP_008432025.1 | 486 | PREDICTED: somatostatin receptor type 3 |
| LG1 | NC_024331.1 | 5550654 | 103477614 | XP_008429170.1 | 295 | PREDICTED: somatostatin receptor type 2-like |
| LG1 | NC_024331.1 | 9522896 | 103459798 | XP_008400257.1 | 124 | PREDICTED: NADH dehydrogenase |
| LG1 | NC_024331.1 | 17395195 | 103463290 | XP_008404769.1 | 259 | PREDICTED: insulin-like growth factor-binding protein 4 isoform X1 |
| LG1 | NC_024331.1 | 18433705 | 103466424 | XP_008410195.1 | 127 | PREDICTED: NADH dehydrogenase |
| LG1 | NC_024331.1 | 27812960 | 103470526 | XP_008417272.1 | 208 | PREDICTED: NADH dehydrogenase |
| LG1 | NC_024331.1 | 33622543 | 103472096 | XP_008419703.1 | 352 | PREDICTED: NADH dehydrogenase |
| LG2 | NC_024332.1 | 9181506 | 103473346 | XP_008421731.1 | 248 | PREDICTED: insulin-like growth factor-binding protein 5 |
| LG2 | NC_024332.1 | 9190987 | 103473326 | XP_008421702.1 | 289 | PREDICTED: insulin-like growth factor-binding protein 2 isoform X1 |
| LG2 | NC_024332.1 | 18747206 | 103480080 | XP_008433065.1 | 349 | PREDICTED: growth hormone-regulated TBC protein 1-A-like isoform X1 |
| LG2 | NC_024332.1 | 21294971 | 103482419 | XP_008436788.1 | 104 | PREDICTED: NADH dehydrogenase |
| LG2 | NC_024332.1 | 21820147 | 103482342 | XP_008436673.1 | 355 | PREDICTED: NADH dehydrogenase |
| LG2 | NC_024332.1 | 33336088 | 103461664 | XP_008402174.1 | 613 | PREDICTED: insulin-like growth factor 2 mRNA-binding protein 2 isoform X1 |
| LG2 | NC_024332.1 | 33653430 | 103461653 | XP_008402164.1 | 780 | PREDICTED: 6-phosphofructokinase, liver type |
| LG3 | NC_024333.1 | 4911962 | 103462200 | XP_008403070.1 | 424 | PREDICTED: putative gonadotropin-releasing hormone II receptor isoform X1 |
| LG3 | NC_024333.1 | 6120547 | 103462176 | XP_008403020.1 | 422 | PREDICTED: gonadotropin-releasing hormone II receptor-like |
| LG3 | NC_024333.1 | 9832512 | 103462065 | XP_008402794.1 | 517 | PREDICTED: aromatase |
| LG3 | NC_024333.1 | 12431897 | 103461976 | XP_008402620.1 | 876 | PREDICTED: transforming growth factor-beta receptor-associated protein 1 |
| LG3 | NC_024333.1 | 12459077 | 103461973 | XP_008402616.1 | 339 | PREDICTED: growth hormone-regulated TBC protein 1 |
| LG3 | NC_024333.1 | 22638619 | 103462710 | XP_008403875.1 | 113 | PREDICTED: follitropin subunit beta |
| LG3 | NC_024333.1 | 26714273 | 103462605 | XP_008403720.1 | 1403 | PREDICTED: insulin-like growth factor 1 receptor |
| LG3 | NC_024333.1 | 33408921 | 103462939 | XP_008404233.1 | 288 | PREDICTED: troponin T, fast skeletal muscle isoforms-like |
| LG4 | NC_024334.1 | 8584192 | 103463426 | XP_008404980.1 | 108 | PREDICTED: NADH dehydrogenase |
| LG4 | NC_024334.1 | 8969045 | 103463415 | XP_008404963.1 | 507 | PREDICTED: pituitary adenylate cyclase-activating polypeptide type I receptor-like isoform X1 |
| LG4 | NC_024334.1 | 9210790 | 103463405 | XP_008404945.1 | 283 | PREDICTED: insulin-like growth factor-binding protein 3 |
| LG4 | NC_024334.1 | 9235470 | 103463404 | XP_008404944.1 | 255 | PREDICTED: insulin-like growth factor-binding protein 1 |
| LG4 | NC_024334.1 | 11340762 | 103463349 | XP_008404865.1 | 262 | PREDICTED: NADH dehydrogenase |
| LG4 | NC_024334.1 | 14956612 | 103463995 | XP_008406009.1 | 373 | PREDICTED: glutamine synthetase-like |
| LG4 | NC_024334.1 | 15295789 | 103463976 | XP_008405978.1 | 818 | PREDICTED: transforming growth factor beta receptor type 3 |
| LG4 | NC_024334.1 | 27538810 | 103464135 | XP_008406234.1 | 185 | PREDICTED: NADH dehydrogenase |
| LG5 | NC_024335.1 | 3474862 | 103464626 | XP_008407069.1 | 1210 | PREDICTED: myosin heavy chain, fast skeletal muscle-like |

| Linkage group | Assecion number | Start | Gene ID | Protein | Length | Protein description |
|---|---|---|---|---|---|---|
| LG5 | NC_024335.1 | 3491926 | 103464625 | XP_008407068.1 | 1392 | PREDICTED: myosin heavy chain, fast skeletal muscle-like |
| LG5 | NC_024335.1 | 3511683 | 103464607 | XP_008407049.1 | 1937 | PREDICTED: myosin heavy chain, fast skeletal muscle isoform X1 |
| LG5 | NC_024335.1 | 8606151 | 103464393 | XP_008406672.1 | 99 | PREDICTED: NADH dehydrogenase |
| LG5 | NC_024335.1 | 11325210 | 103464708 | XP_008407198.1 | 241 | PREDICTED: troponin T, cardiac muscle isoforms-like |
| LG5 | NC_024335.1 | 18875389 | 103464970 | XP_008407637.1 | 169 | PREDICTED: myosin regulatory light chain 2, ventricular/cardiac muscle isoform |
| LG5 | NC_024335.1 | 20968575 | 103464885 | XP_008407496.1 | 250 | PREDICTED: myogenin isoform X1 |
| LG5 | NC_024335.1 | 22557454 | 103465143 | XP_008407970.1 | 188 | PREDICTED: NADH dehydrogenase |
| LG5 | NC_024335.1 | 25487041 | 103465332 | XP_008408247.1 | 192 | PREDICTED: insulin-like growth factor-binding protein 3 |
| LG5 | NC_024335.1 | 26315741 | 103465370 | XP_008408349.1 | 778 | PREDICTED: 6-phosphofructokinase, muscle type-like |
| LG5 | NC_024335.1 | 27388251 | 103465401 | XP_008408401.1 | 142 | PREDICTED: somatoliberin isoform X1 |
| LG6 | NC_024336.1 | 542839 | 103465649 | XP_008408889.1 | 232 | PREDICTED: troponin T, fast skeletal muscle isoforms-like isoform X1 |
| LG6 | NC_024336.1 | 6275538 | 103465824 | XP_008409218.1 | 240 | PREDICTED: troponin T, cardiac muscle isoforms-like |
| LG6 | NC_024336.1 | 13327600 | 103466006 | XP_008409487.1 | 156 | PREDICTED: leptin-A-like |
| LG6 | NC_024336.1 | 17567964 | 103466159 | XP_008409776.1 | 223 | PREDICTED: insulin-like growth factor II |
| LG6 | NC_024336.1 | 22654936 | 103466654 | XP_008410609.1 | 1415 | PREDICTED: insulin-like growth factor 1 receptor |
| LG6 | NC_024336.1 | 23620987 | 103466634 | XP_008410576.1 | 394 | PREDICTED: gonadotropin-releasing hormone II receptor-like |
| LG6 | NC_024336.1 | 26763968 | 103466459 | XP_008410249.1 | 499 | PREDICTED: aromatase-like |
| LG6 | NC_024336.1 | 31345364 | 103466813 | XP_008410897.1 | 198 | PREDICTED: fibroblast growth factor 6-like |
| LG7 | NC_024337.1 | 2829468 | 103467061 | XP_008411337.1 | 382 | PREDICTED: growth hormone secretagogue receptor type 1-like |
| LG7 | NC_024337.1 | 2866184 | 103466898 | XP_008411023.1 | 161 | PREDICTED: troponin C, skeletal muscle |
| LG7 | NC_024337.1 | 2874663 | 103466900 | XP_008411024.1 | 160 | PREDICTED: troponin C, skeletal muscle-like |
| LG7 | NC_024337.1 | 4466392 | 103466958 | XP_008411135.1 | 196 | PREDICTED: insulin-like growth factor I, juvenile form isoform X2 |
| LG7 | NC_024337.1 | 5389550 | 103466994 | XP_008411222.1 | 144 | PREDICTED: NADH dehydrogenase |
| LG7 | NC_024337.1 | 11792192 | 103467230 | XP_008411649.1 | 110 | PREDICTED: somatostatin-1B-like |
| LG7 | NC_024337.1 | 13336485 | 103467898 | XP_008412848.1 | 201 | PREDICTED: insulin-like growth factor-binding protein 6 |
| LG7 | NC_024337.1 | 21232538 | 103467588 | XP_008412336.1 | 199 | PREDICTED: troponin C, slow skeletal and cardiac muscles isoform X1 |
| LG7 | NC_024337.1 | 25503303 | 103467433 | XP_008412032.1 | 156 | PREDICTED: follitropin subunit beta-like |
| LG7 | NC_024337.1 | 25834694 | 103467415 | XP_008412005.1 | 281 | PREDICTED: troponin T, cardiac muscle isoform X1 |
| LG7 | NC_024337.1 | 28258280 | 103467300 | XP_008411776.1 | 780 | PREDICTED: 6-phosphofructokinase, muscle type-like |
| LG7 | NC_024337.1 | 29986107 | 103468041 | XP_008413100.1 | 81 | PREDICTED: NADH dehydrogenase |
| LG7 | NC_024337.1 | 30391748 | 103468062 | XP_008413125.1 | 161 | PREDICTED: troponin C, slow skeletal and cardiac muscles-like |

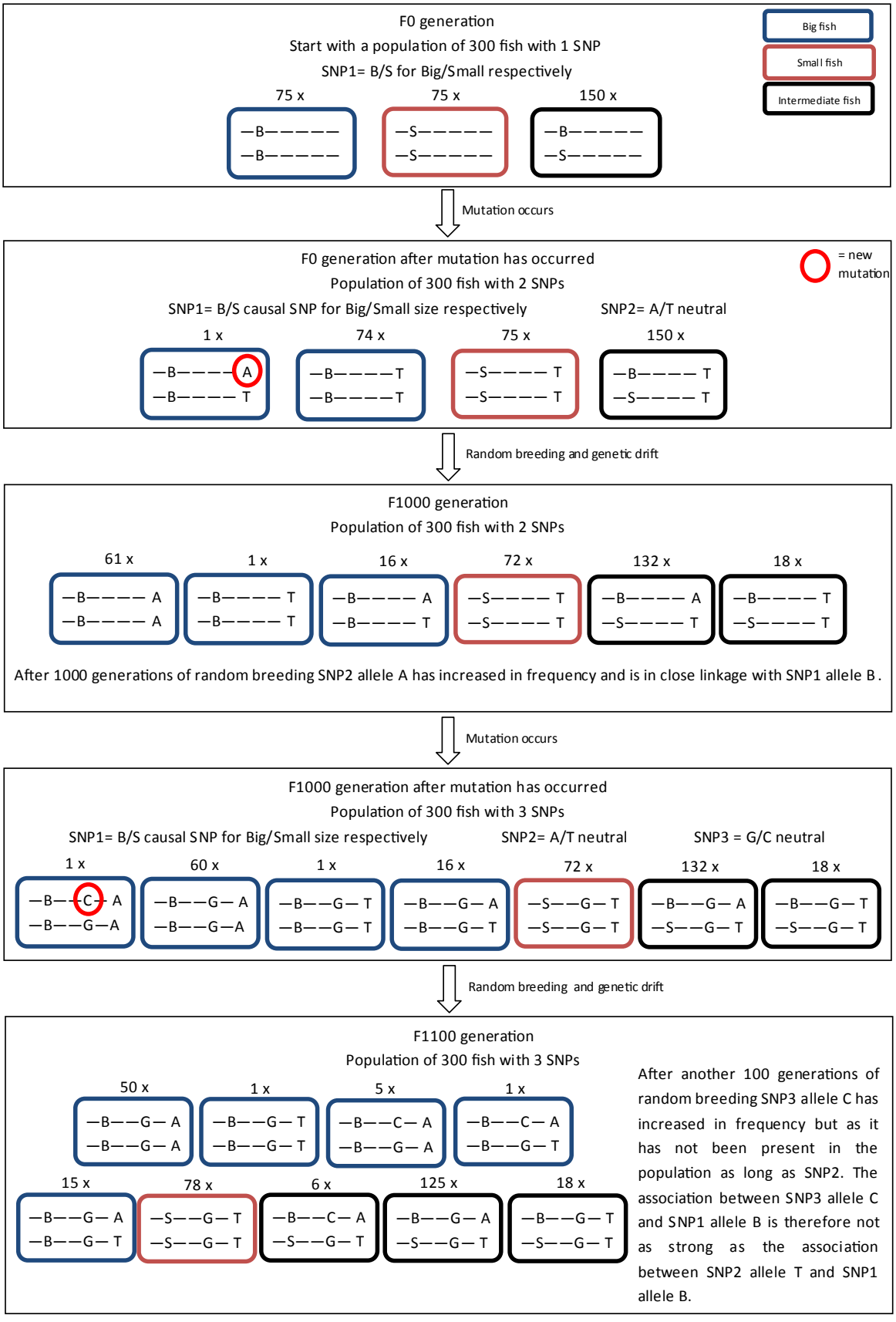
| Linkage group | Assecion number | Start | Gene ID | Protein | Length | Protein description |
|---|---|---|---|---|---|---|
| LG8 | NC_024338.1 | 2030874 | 103468198 | XP_008413316.1 | 601 | PREDICTED: insulin-like growth factor 2 mRNA-binding protein 1 |
| LG8 | NC_024338.1 | 3923604 | 103468503 | XP_008413845.1 | 1174 | PREDICTED: leptin receptor |
| LG8 | NC_024338.1 | 4722896 | 103468295 | XP_008413482.1 | 212 | PREDICTED: prolactin-1-like |
| LG8 | NC_024338.1 | 4972510 | 103468510 | XP_008413850.1 | 214 | PREDICTED: troponin T, slow skeletal muscle-like |
| LG8 | NC_024338.1 | 6383808 | 103468515 | XP_008413856.1 | 515 | PREDICTED: transforming growth factor beta receptor type 3-like |
| LG8 | NC_024338.1 | 7925447 | 103468432 | XP_008413740.1 | 128 | PREDICTED: NADH dehydrogenase |
| LG8 | NC_024338.1 | 11903773 | 103468864 | XP_008414460.1 | 370 | PREDICTED: somatostatin receptor type 2-like |
| LG8 | NC_024338.1 | 14197489 | 103468881 | XP_008414483.1 | 149 | PREDICTED: myosin heavy chain, fast skeletal muscle-like |
| LG8 | NC_024338.1 | 14209668 | 103468880 | XP_008414482.1 | 1138 | PREDICTED: myosin heavy chain, fast skeletal muscle-like |
| LG8 | NC_024338.1 | 14253373 | 103468749 | XP_008414253.1 | 179 | PREDICTED: myosin heavy chain, fast skeletal muscle-like |
| LG8 | NC_024338.1 | 15074657 | 103468697 | XP_008414162.1 | 577 | PREDICTED: insulin-like growth factor-binding protein complex acid labile subunit |
| LG8 | NC_024338.1 | 17389327 | 103468910 | XP_008414515.1 | 82 | PREDICTED: NADH dehydrogenase |
| LG8 | NC_024338.1 | 17688680 | 103468934 | XP_008414551.1 | 244 | PREDICTED: myosin regulatory light chain 2, skeletal muscle isoform isoform X1 |
| LG8 | NC_024338.1 | 19282483 | 103469032 | XP_008414724.1 | 204 | PREDICTED: somatotropin |
| LG8 | NC_024338.1 | 22003474 | 103469309 | XP_008415131.1 | 364 | PREDICTED: somatostatin receptor type 5-like isoform X1 |
| LG8 | NC_024338.1 | 23274501 | 103469245 | XP_008415033.1 | 217 | PREDICTED: NADH dehydrogenase |
| LG8 | NC_024338.1 | 27665377 | 103469447 | XP_008415353.1 | 468 | PREDICTED: somatostatin receptor type 5-like |
| LG9 | NC_024339.1 | 12928490 | 103469937 | XP_008416240.1 | 520 | PREDICTED: prolactin receptor-like |
| LG9 | NC_024339.1 | 14592583 | 103469994 | XP_008416318.1 | 170 | PREDICTED: NADH dehydrogenase |
| LG9 | NC_024339.1 | 14614689 | 103469996 | XP_008416319.1 | 351 | PREDICTED: follistatin isoform X1 |
| LG9 | NC_024339.1 | 14728925 | 103470002 | XP_008416329.1 | 172 | PREDICTED: myosin regulatory light chain 2, atrial isoform |
| LG9 | NC_024339.1 | 16900445 | 103470120 | XP_008416523.1 | 489 | PREDICTED: histone-lysine N-methyltransferase SMYD1-like isoform X1 |
| LG9 | NC_024339.1 | 22968878 | 103470360 | XP_008416972.1 | 272 | PREDICTED: insulin-like growth factor-binding protein 3 receptor |
| LG9 | NC_024339.1 | 27353509 | 103470524 | XP_008417266.1 | 365 | PREDICTED: prolactin-releasing peptide receptor-like |
| LG10 | NC_024340.1 | 12202417 | 103471155 | XP_008418206.1 | 102 | PREDICTED: NADH dehydrogenase |
| LG10 | NC_024340.1 | 16727971 | 103471547 | XP_008418843.1 | 70 | PREDICTED: NADH dehydrogenase |
| LG10 | NC_024340.1 | 21209985 | 103471446 | XP_008418677.1 | 414 | PREDICTED: transforming growth factor beta-2 |
| LG10 | NC_024340.1 | 27124922 | 103471210 | XP_008418279.1 | 158 | PREDICTED: putative fibroblast growth factor 1 |
| LG10 | NC_024340.1 | 32243330 | 103471790 | XP_008419204.1 | 288 | PREDICTED: insulin-like growth factor-binding protein 7 |
| LG10 | NC_024340.1 | 32364321 | 103471784 | XP_008419195.1 | 671 | PREDICTED: transforming growth factor-beta-induced protein ig-h3 isoform X1 |
| LG11 | NC_024341.1 | 5681372 | 103472060 | XP_008419646.1 | 170 | PREDICTED: NADH dehydrogenase |

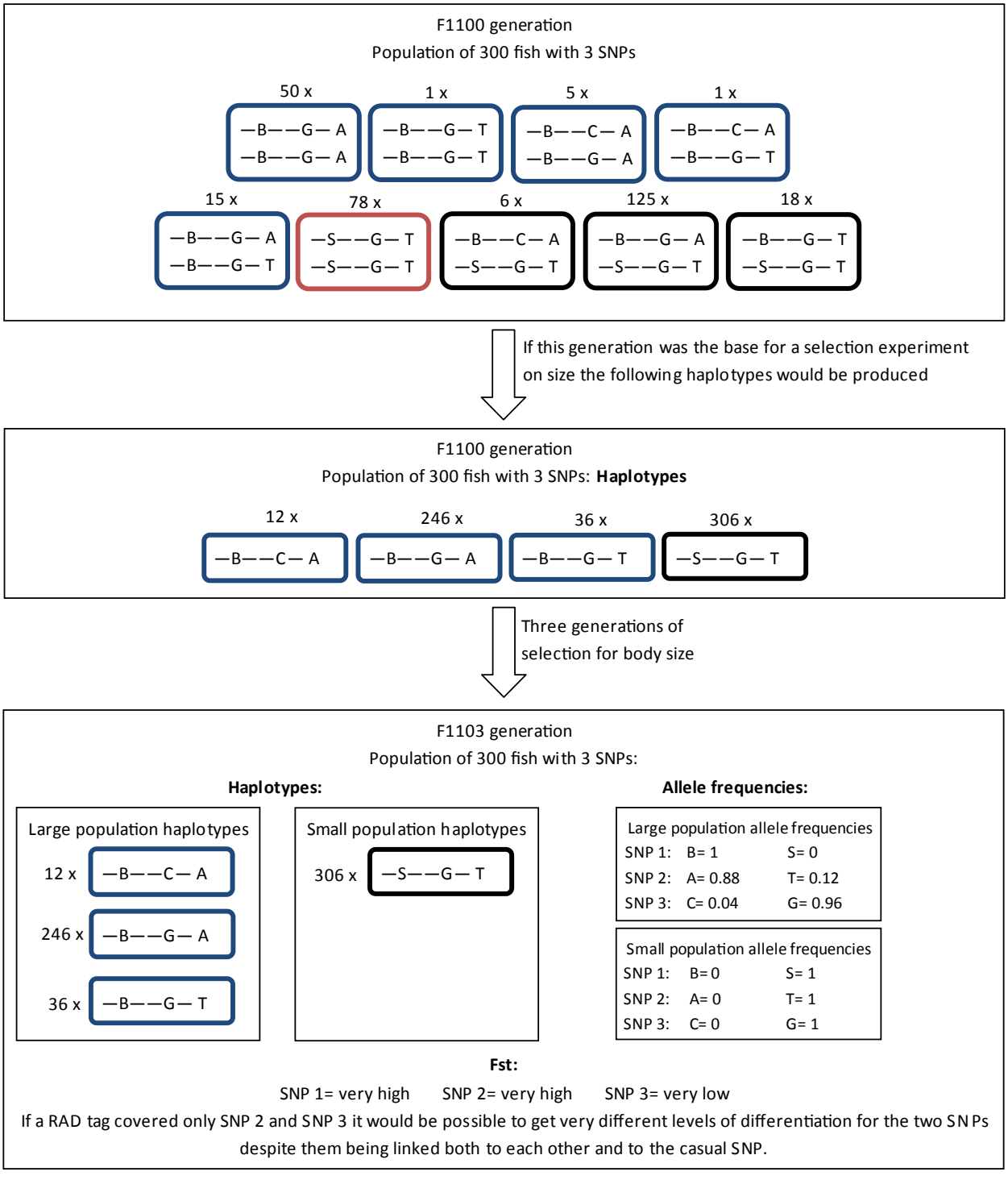
| Linkage group | Assecion number | Start | Gene ID | Protein | Length | Protein description |
|---|---|---|---|---|---|---|
| LG11 | NC_024341.1 | 10760794 | 103472224 | XP_008419904.1 | 608 | PREDICTED: insulin-like growth factor 2 mRNA-binding protein 3 isoform X1 |
| LG11 | NC_024341.1 | 17513645 | 103472506 | XP_008420426.1 | 472 | PREDICTED: NADH dehydrogenase |
| LG11 | NC_024341.1 | 22368099 | 103472807 | XP_008420871.1 | 175 | PREDICTED: NADH dehydrogenase |
| LG11 | NC_024341.1 | 26016782 | 103472969 | XP_008421071.1 | 106 | PREDICTED: NADH dehydrogenase |
| LG11 | NC_024341.1 | 27748514 | 103473019 | XP_008421153.1 | 254 | PREDICTED: apolipoprotein A-IV-like |
| LG12 | NC_024342.1 | 14841552 | 103473651 | XP_008422265.1 | 172 | PREDICTED: NADH dehydrogenase |
| LG12 | NC_024342.1 | 15907889 | 103473713 | XP_008422373.1 | 360 | PREDICTED: prolactin-releasing peptide receptor-like |
| LG12 | NC_024342.1 | 19087903 | 103473840 | XP_008422612.1 | 582 | PREDICTED: transforming growth factor beta receptor type 3-like |
| LG12 | NC_024342.1 | 20099139 | 103473871 | XP_008422714.1 | 638 | PREDICTED: prolactin receptor |
| LG12 | NC_024342.1 | 26143513 | 103474100 | XP_008423053.1 | 473 | PREDICTED: histone-lysine N-methyltransferase SMYD1 |
| LG13 | NC_024343.1 | 1491367 | 103474182 | XP_008423207.1 | 309 | PREDICTED: growth hormone secretagogue receptor type 1-like |
| LG13 | NC_024343.1 | 1495058 | 103474185 | XP_008423208.1 | 308 | PREDICTED: growth hormone secretagogue receptor type 1-like |
| LG13 | NC_024343.1 | 16970194 | 103474611 | XP_008423941.1 | 384 | PREDICTED: prolactin-releasing peptide receptor-like |
| LG13 | NC_024343.1 | 20887924 | 103474759 | XP_008424184.1 | 230 | PREDICTED: somatolactin-like isoform X2 |
| LG13 | NC_024343.1 | 21474000 | 103474775 | XP_008424214.1 | 546 | PREDICTED: transforming growth factor beta regulator 1 |
| LG13 | NC_024343.1 | 23174396 | 103474828 | XP_008424299.1 | 263 | PREDICTED: apolipoprotein A-I |
| LG13 | NC_024343.1 | 28009732 | 103475017 | XP_008424604.1 | 110 | PREDICTED: NADH dehydrogenase |
| LG13 | NC_024343.1 | 30883059 | 103475119 | XP_008424748.1 | 380 | PREDICTED: transforming growth factor beta-1-like |
| LG13 | NC_024343.1 | 31781732 | 103475163 | XP_008424814.1 | 668 | PREDICTED: insulin-like growth factor-binding protein complex acid labile subunit isoform X1 |
| LG13 | NC_024343.1 | 33346383 | 103475296 | XP_008425038.1 | 119 | PREDICTED: somatostatin-1-like |
| LG13 | NC_024343.1 | 33365659 | 103475319 | XP_008425065.1 | 121 | PREDICTED: somatostatin-2-like, partial |
| LG14 | NC_024344.1 | 10840083 | 103475823 | XP_008425945.1 | 1240 | PREDICTED: latent-transforming growth factor beta-binding protein 3 isoform X1 |
| LG15 | NC_024345.1 | 200878 | 103476545 | XP_008427205.1 | 1452 | PREDICTED: latent-transforming growth factor beta-binding protein 1 isoform X1 |
| LG15 | NC_024345.1 | 7550752 | 103476626 | XP_008427325.1 | 147 | PREDICTED: gonadotropin subunit beta-2-like isoform X1 |
| LG15 | NC_024345.1 | 19974001 | 103477044 | XP_008428134.1 | 693 | PREDICTED: lutropin-choriogonadotropic hormone receptor-like |
| LG15 | NC_024345.1 | 20339660 | 103477059 | XP_008428157.1 | 712 | PREDICTED: lutropin-choriogonadotropic hormone receptor-like |
| LG15 | NC_024345.1 | 21143109 | 103477164 | XP_008428321.1 | 292 | PREDICTED: somatostatin receptor type 5-like |
| LG15 | NC_024345.1 | 29500075 | 103477379 | XP_008428682.1 | 366 | PREDICTED: early growth response protein 2b-like |
| LG16 | NC_024346.1 | 831048 | 103477459 | XP_008428807.1 | 185 | PREDICTED: gonadotropin-releasing hormone II receptor-like |
| LG16 | NC_024346.1 | 2802913 | 103477594 | XP_008429014.1 | 331 | PREDICTED: NADH dehydrogenase (ubiquinone) complex I, assembly factor 6 isoform X1 |
| LG16 | NC_024346.1 | 7691682 | 103477850 | XP_008429393.1 | 255 | PREDICTED: apolipoprotein A-I-like |

| Linkage group | Assecion number | Start | Gene ID | Protein | Length | Protein description |
|---|---|---|---|---|---|---|
| LG16 | NC_024346.1 | 7695411 | 103477849 | XP_008429392.1 | 255 | PREDICTED: apolipoprotein A-I-like |
| LG16 | NC_024346.1 | 10893463 | 103478105 | XP_008429843.1 | 500 | PREDICTED: NADH dehydrogenase |
| LG16 | NC_024346.1 | 31404151 | 103478768 | XP_008431023.1 | 428 | PREDICTED: prolactin regulatory element-binding protein |
| LG17 | NC_024347.1 | 544166 | 103478849 | XP_008431148.1 | 129 | PREDICTED: NADH dehydrogenase |
| LG17 | NC_024347.1 | 18560113 | 103479587 | XP_008432331.1 | 371 | PREDICTED: glutamine synthetase |
| LG17 | NC_024347.1 | 19933656 | 103479557 | XP_008432277.1 | 285 | PREDICTED: insulin-like growth factor-binding protein 3 |
| LG17 | NC_024347.1 | 19941683 | 103479556 | XP_008432276.1 | 247 | PREDICTED: insulin-like growth factor-binding protein 1 |
| LG17 | NC_024347.1 | 22267930 | 103479727 | XP_008432545.1 | 244 | PREDICTED: NADH dehydrogenase |
| LG17 | NC_024347.1 | 23402902 | 103479754 | XP_008432607.1 | 237 | PREDICTED: myosin heavy chain, clone 203-like |
| LG17 | NC_024347.1 | 25799058 | 103479841 | XP_008432735.1 | 478 | PREDICTED: pituitary adenylate cyclase-activating polypeptide type I receptor-like isoform X1 |
| LG17 | NC_024347.1 | 25830368 | 103479894 | XP_008432799.1 | 306 | PREDICTED: pituitary adenylate cyclase-activating polypeptide type I receptor-like |
| LG17 | NC_024347.1 | 28273848 | 103479939 | XP_008432878.1 | 803 | PREDICTED: 6-phosphofructokinase type C isoform X1 |
| LG18 | NC_024348.1 | 3065585 | 103480169 | XP_008433256.1 | 467 | PREDICTED: NADH dehydrogenase |
| LG18 | NC_024348.1 | 10451989 | 103480475 | XP_008433651.1 | 109 | PREDICTED: somatostatin-1A-like |
| LG18 | NC_024348.1 | 21739197 | 103480965 | XP_008434404.1 | 551 | PREDICTED: myosin heavy chain, cardiac muscle isoform-like |
| LG19 | NC_024349.1 | 3435802 | 103481176 | XP_008434704.1 | 382 | PREDICTED: NADH dehydrogenase |
| LG19 | NC_024349.1 | 5257306 | 103481239 | XP_008434813.1 | 247 | PREDICTED: troponin T, slow skeletal muscle isoform X1 |
| LG19 | NC_024349.1 | 8247676 | 103481803 | XP_008435786.1 | 66 | PREDICTED: NADH dehydrogenase |
| LG19 | NC_024349.1 | 8458283 | 103481365 | XP_008434992.1 | 283 | PREDICTED: transforming growth factor beta-1-induced transcript 1 protein-like isoform X1 |
| LG19 | NC_024349.1 | 9818913 | 103481443 | XP_008435100.1 | 348 | PREDICTED: growth hormone-inducible transmembrane protein |
| LG19 | NC_024349.1 | 10058771 | 103481449 | XP_008435115.1 | 202 | PREDICTED: NADH dehydrogenase |
| LG19 | NC_024349.1 | 12435168 | 103481549 | XP_008435305.1 | 644 | PREDICTED: lutropin-choriogonadotropic hormone receptor-like, partial |
| LG19 | NC_024349.1 | 21314507 | 103481922 | XP_008435968.1 | 1133 | PREDICTED: myosin heavy chain, cardiac muscle isoform-like isoform X2 |
| LG19 | NC_024349.1 | 24934206 | 103482062 | XP_008436195.1 | 381 | PREDICTED: somatostatin receptor type 2-like |
| LG19 | NC_024349.1 | 24972850 | 103482063 | XP_008436196.1 | 339 | PREDICTED: somatostatin receptor type 2-like |
| LG20 | NC_024350.1 | 2579506 | 103482255 | XP_008436502.1 | 841 | PREDICTED: 6-phosphofructokinase type C-like isoform X1 |
| LG20 | NC_024350.1 | 10267068 | 103482442 | XP_008436824.1 | 127 | PREDICTED: NADH dehydrogenase |
| LG20 | NC_024350.1 | 10419025 | 103482439 | XP_008436817.1 | 122 | PREDICTED: prolactin-releasing peptide |
| LG20 | NC_024350.1 | 22478531 | 103456858 | XP_008394895.1 | 172 | PREDICTED: myosin regulatory light chain 2, smooth muscle minor isoform |
| LG20 | NC_024350.1 | 22692540 | 103456866 | XP_008394910.1 | 232 | PREDICTED: pituitary adenylate cyclase-activating polypeptide isoform X1 |
| LG21 | NC_024351.1 | 360616 | 103457053 | XP_008395183.1 | 452 | PREDICTED: NADH dehydrogenase |

| Linkage group | Assecion number | Start | Gene ID | Protein | Length | Protein description |
|---|---|---|---|---|---|---|
| LG21 | NC_024351.1 | 2353077 | 103457292 | XP_008395551.1 | 222 | PREDICTED: myosin-2 heavy chain, non muscle-like |
| LG21 | NC_024351.1 | 6761968 | 103457593 | XP_008396090.1 | 407 | PREDICTED: transforming growth factor beta-3-like |
| LG21 | NC_024351.1 | 11820826 | 103457382 | XP_008395682.1 | 57 | PREDICTED: NADH dehydrogenase |
| LG21 | NC_024351.1 | 16010425 | 103457915 | XP_008396615.1 | 733 | PREDICTED: heat shock protein HSP 90-alpha |
| LG21 | NC_024351.1 | 16015409 | 103457917 | XP_008396616.1 | 724 | PREDICTED: heat shock protein HSP 90-alpha 1 |
| LG21 | NC_024351.1 | 25553568 | 103458043 | XP_008396809.1 | 175 | PREDICTED: NADH dehydrogenase |
| LG22 | NC_024352.1 | 3200974 | 103458132 | XP_008396944.1 | 329 | PREDICTED: myosin-2 heavy chain-like isoform X1 |
| LG22 | NC_024352.1 | 7244496 | 103458418 | XP_008397434.1 | 384 | PREDICTED: growth hormone secretagogue receptor type 1-like |
| LG22 | NC_024352.1 | 10519383 | 103458496 | XP_008397564.1 | 1978 | PREDICTED: latent-transforming growth factor beta-binding protein 2 isoform X1 |
| LG22 | NC_024352.1 | 18585431 | 103459051 | XP_008398493.1 | 408 | PREDICTED: transforming growth factor beta-3 |
| LG22 | NC_024352.1 | 21719083 | 103458913 | XP_008398267.1 | 695 | PREDICTED: reticulon-1 isoform X1 |
| LG22 | NC_024352.1 | 22338652 | 103458878 | XP_008398218.1 | 407 | PREDICTED: prolactin-releasing peptide receptor-like |
| LG23 | NC_024353.1 | 535857 | 103459109 | XP_008398577.1 | 101 | PREDICTED: NADH dehydrogenase |
| LG23 | NC_024353.1 | 1556150 | 103459130 | XP_008398655.1 | 240 | PREDICTED: myogenic factor 5 |
| LG23 | NC_024353.1 | 1561875 | 103459131 | XP_008398656.1 | 240 | PREDICTED: myogenic factor 6 isoform X1 |
| LG23 | NC_024353.1 | 2010563 | 103459141 | XP_008398690.1 | 284 | PREDICTED: troponin T, fast skeletal muscle-like |
| LG23 | NC_024353.1 | 2270134 | 103459155 | XP_008398704.1 | 147 | PREDICTED: NADH dehydrogenase |
| LG23 | NC_024353.1 | 2615734 | 103459170 | XP_008398725.1 | 207 | PREDICTED: fibroblast growth factor 6-like |
| LG23 | NC_024353.1 | 3426591 | 103459218 | XP_008398820.1 | 184 | PREDICTED: leptin-B-like |
| LG23 | NC_024353.1 | 9962367 | 103459481 | XP_008399289.1 | 289 | PREDICTED: troponin T, cardiac muscle isoforms-like isoform X1 |
| LG23 | NC_024353.1 | 12911013 | 103459568 | XP_008399437.1 | 116 | PREDICTED: NADH dehydrogenase |
| LG23 | NC_024353.1 | 17691806 | 103459740 | XP_008399773.1 | 230 | PREDICTED: prolactin-like |
| Un | NW_007615957.1 | 6 | 103461378 | XP_008401905.1 | 783 | PREDICTED: myosin heavy chain, fast skeletal muscle-like, partial |
| Un | NW_007616313.1 | 6 | 103461448 | XP_008401968.1 | 292 | PREDICTED: myosin heavy chain, fast skeletal muscle-like, partial |
| Un | NW_007617004.1 | 6 | 103461559 | XP_008402053.1 | 281 | PREDICTED: myosin heavy chain, fast skeletal muscle-like, partial |
| Un | NW_007615029.1 | 59373 | 103460411 | XP_008400788.1 | 94 | PREDICTED: myosin heavy chain, embryonic smooth muscle isoform-like |
| Un | NW_007615023.1 | 272748 | 103460258 | XP_008400586.1 | 322 | PREDICTED: transcription factor jun-D-like |
| Un | NW_007615014.1 | 834575 | 103459895 | XP_008400049.1 | 100 | PREDICTED: NADH dehydrogenase |
| Un | NW_007615017.1 | 866581 | 103460057 | XP_008400261.1 | 1887 | PREDICTED: latent-transforming growth factor beta-binding protein 4 |

## Appendix IV: Schematic explaining how it is possible to have multiple SNPs on the same TAG showing a different response to selection.



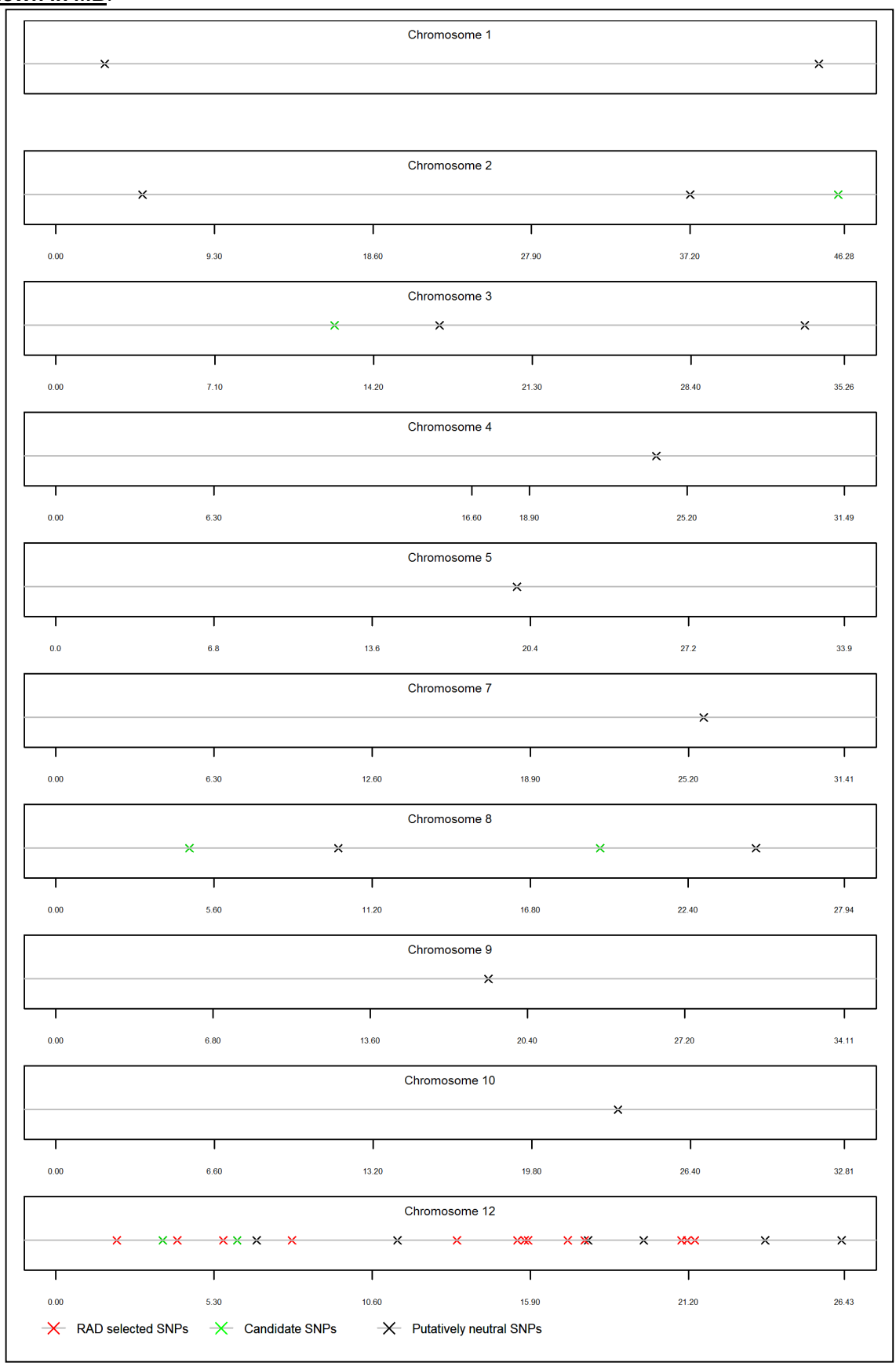

F1100 generation
Population of 300 fish with 3 SNPs
50 x
—B——G— A
—B——G— A
1 x
—B——G— T
—B——G— T
5 x
—B——C— A
—B——G— A
1 x
—B——C— A
—B——G— T
15 x
—B——G— A
—B——G— T
78 x
—S——G— T
—S——G— T
6 x
—B——C— A
—S——G— T
125 x
—B——G— A
—S——G— T
18 x
—B——G— T
—S——G— T
If this generation was the base for a selection experiment on size the following haplotypes would be produced
F1100 generation
Population of 300 fish with 3 SNPs: Haplotypes
12 x
—B——C— A
246 x
—B——G— A
36 x
—B——G— T
306 x
—S——G— T
Three generations of selection for body size
F1103 generation
Population of 300 fish with 3 SNPs:
Haplotypes:
Large population haplotypes
12 x —B——C— A
246 x —B——G— A
36 x —B——G— T
Small population haplotypes
306 x —S——G— T
Allele frequencies:
Large population allele frequencies
SNP 1: B= 1 S= 0
SNP 2: A= 0.88 T= 0.12
SNP 3: C= 0.04 G= 0.96
Small population allele frequencies
SNP 1: B= 0 S= 1
SNP 2: A= 0 T= 1
SNP 3: C= 0 G= 1
Fst:
SNP 1= very high SNP 2= very high SNP 3= very low
If a RAD tag covered only SNP 2 and SNP 3 it would be possible to get very different levels of differentiation for the two SNPs despite them being linked both to each other and to the casual SNP.

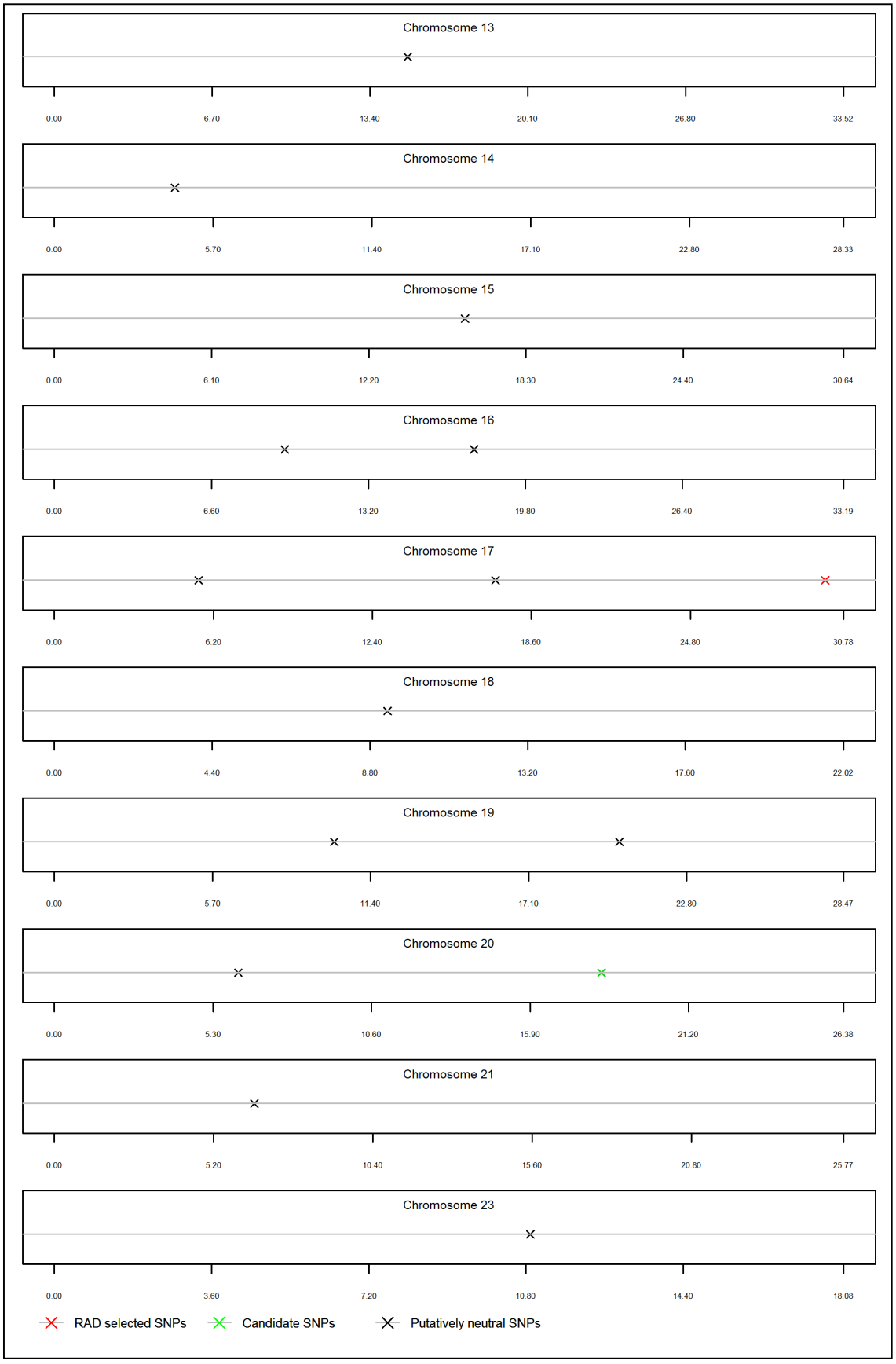
## Appendix V: Extraction protocol for tissue samples taken from wild fish.

1. Place the tissue sample in 500µl of TEN buffer*, add 5µl of proteinase K and incubate overnight at 55ºC.
2. Vortex the sample and remove 300µl of buffer/proteinase K mix. Place in a new 1.5ml eppendorf tube.
3. Add 100µl of 5M NaCl and centrifuge for 5 minutes at 14,000rpm.
4. Remove the supernatant and place in a new eppendorf tube. Be careful to avoid the precipitate at the base of the tube.
5. Add 800µl of 100% ice cold ethanol.
6. Leave overnight at –20ºC.
7. Centrifuge at 4ºC for 30 minutes at 14,000rpm.
8. Pour off the ethanol.
9. Add 1ml of 70% ethanol.
10. Centrifuge for 5 minutes at 13,000rpm.
11. Pour off the ethanol and allow to drain upside down.
12. Dry DNA pellet at 50ºC for 10 minutes.
13. Re-suspend DNA pellet in 50µl of $H_20$.
14. Dissolve DNA by incubating the samples 37ºC for 30 minutes.

* TEN: 0.4M NaCl, 10mM Tris-HCl pH 8.0 and 2mM EDTA pH 8.0 with 2% SDS in a 9:1 ratio

**Appendix VI: Location of genotyped SNPs on the guppy genome. Chromosome size is shown in MB**.




Chromosome 13
0.00
6.70
13.40
20.10
26.80
33.52
Chromosome 14
0.00
5.70
11.40
17.10
22.80
28.33
Chromosome 15
0.00
6.10
12.20
18.30
24.40
30.64
Chromosome 16
0.00
6.60
13.20
19.80
26.40
33.19
Chromosome 17
0.00
6.20
12.40
18.60
24.80
30.78
Chromosome 18
0.00
4.40
8.80
13.20
17.60
22.02
Chromosome 19
0.00
5.70
11.40
17.10
22.80
28.47
Chromosome 20
0.00
5.30
10.60
15.90
21.20
26.38
Chromosome 21
0.00
5.20
10.40
15.60
20.80
25.77
Chromosome 23
0.00
3.60
7.20
10.80
14.40
18.08
RAD selected SNPs
Candidate SNPs
Putatively neutral SNPs

**Appendix VII: $F_{st}$ values between all wild sites sampled.**

| | Downstream Aripo 03 | Downstream Aripo 06 | Downstream Caura | Downstream Guanapo | Downstream Lopinot | Downstream Marianne | Downstream Oropuche | Downstream Turure | Downstream Yara | Upstream Aripo 03 | Upstream Aripo 06 | Upstream Caura | Upstream Guanapo | Upstream Lopinot | Upstream Marianne | Upstream Oropuche | Upstream Turure |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Downstream Aripo 06 | 0.01 | | | | | | | | | | | | | | | | |
| Downstream Caura | 0.21 | 0.20 | | | | | | | | | | | | | | | |
| Downstream Guanapo | 0.32 | 0.33 | 0.16 | | | | | | | | | | | | | | |
| Downstream Lopinot | 0.25 | 0.24 | 0.05 | 0.16 | | | | | | | | | | | | | |
| Downstream Marianne | 0.42 | 0.46 | 0.35 | 0.53 | 0.36 | | | | | | | | | | | | |
| Downstream Oropuche | 0.37 | 0.35 | 0.18 | 0.38 | 0.21 | 0.51 | | | | | | | | | | | |
| Downstream Turure | 0.35 | 0.38 | 0.18 | 0.03 | 0.17 | 0.56 | 0.44 | | | | | | | | | | |
| Downstream Yara | 0.26 | 0.25 | 0.15 | 0.36 | 0.22 | 0.41 | 0.32 | 0.40 | | | | | | | | | |
| Upstream Aripo 03 | 0.30 | 0.37 | 0.44 | 0.60 | 0.48 | 0.77 | 0.72 | 0.63 | 0.61 | | | | | | | | |
| Upstream Aripo 06 | 0.10 | 0.12 | 0.28 | 0.39 | 0.32 | 0.60 | 0.50 | 0.44 | 0.39 | 0.18 | | | | | | | |
| Upstream Caura | 0.44 | 0.44 | 0.25 | 0.45 | 0.28 | 0.68 | 0.53 | 0.49 | 0.40 | 0.70 | 0.53 | | | | | | |
| Upstream Guanapo | 0.47 | 0.51 | 0.27 | 0.18 | 0.24 | 0.73 | 0.66 | 0.15 | 0.52 | 0.83 | 0.64 | 0.65 | | | | | |
| Upstream Lopinot | 0.57 | 0.62 | 0.36 | 0.51 | 0.24 | 0.71 | 0.70 | 0.50 | 0.65 | 0.86 | 0.73 | 0.69 | 0.74 | | | | |
| Upstream Marianne | 0.40 | 0.42 | 0.30 | 0.45 | 0.30 | 0.06 | 0.44 | 0.49 | 0.37 | 0.76 | 0.57 | 0.65 | 0.70 | 0.67 | | | |
| Upstream Oropuche | 0.38 | 0.37 | 0.23 | 0.44 | 0.26 | 0.51 | 0.01 | 0.49 | 0.32 | 0.72 | 0.52 | 0.57 | 0.67 | 0.71 | 0.47 | | |
| Upstream Turure | 0.31 | 0.33 | 0.16 | 0.04 | 0.16 | 0.51 | 0.39 | 0.01 | 0.35 | 0.61 | 0.40 | 0.46 | 0.18 | 0.52 | 0.44 | 0.44 | |
| Upstream Yara | 0.27 | 0.29 | 0.23 | 0.34 | 0.26 | 0.31 | 0.35 | 0.37 | 0.24 | 0.69 | 0.44 | 0.58 | 0.59 | 0.73 | 0.29 | 0.35 | 0.27 |

**Appendix VIII: Values from Lositan outlier analysis. N/A values denote a SNP which was monomorphic in the respective river.**

| Locus | Aripo 03 | | | Aripo 06 | | | Caura | | | Guanapo | | | Lopinot | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Fst | Het | P value | Fst | Het | P value | Fst | Het | P value | Fst | Het | P value | Fst | Het | P value |
| 10802 | N/A | N/A | N/A | 0.035 | 0.052 | 0.319 | 0.279 | 0.286 | 0.104 | 0.000 | 0.021 | 0.500 | 0.144 | 0.151 | 0.062 |
| 108025 | 0.050 | 0.223 | 0.787 | 0.142 | 0.282 | 0.272 | 0.304 | 0.589 | 0.454 | -0.020 | 0.300 | 0.998 | 0.345 | 0.600 | 0.466 |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | 0.065 | 0.094 | 0.051 | N/A | N/A | N/A | N/A | N/A | N/A |
| 108291 | 0.010 | 0.019 | 0.500 | N/A | N/A | N/A | 0.065 | 0.094 | 0.051 | N/A | N/A | N/A | N/A | N/A | N/A |
| 111347 | 0.000 | 0.010 | 0.500 | N/A | N/A | N/A | 0.009 | 0.017 | 0.500 | N/A | N/A | N/A | 0.000 | 0.008 | 0.500 |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | 0.026 | 0.034 | 0.029 | N/A | N/A | N/A | N/A | N/A | N/A |
| 124729 | 0.581 | 0.691 | 0.246 | 0.310 | 0.436 | 0.081 | 0.103 | 0.110 | 0.059 | 0.234 | 0.250 | 0.186 | 0.032 | 0.040 | 0.005 |
| 148158 | N/A | N/A | N/A | N/A | N/A | N/A | 0.104 | 0.112 | 0.000 | N/A | N/A | N/A | 0.096 | 0.103 | 0.222 |
| 150841 | N/A | N/A | N/A | N/A | N/A | N/A | 0.026 | 0.034 | 0.026 | 0.000 | 0.021 | 0.500 | 0.016 | 0.024 | 0.109 |
| 20521 | 0.049 | 0.058 | 0.016 | 0.088 | 0.103 | 0.000 | 0.424 | 0.602 | 0.348 | 0.021 | 0.042 | 0.674 | 0.096 | 0.103 | 0.222 |
| 207392 | 0.150 | 0.435 | 0.653 | 0.123 | 0.521 | 0.370 | 0.162 | 0.169 | 0.308 | 0.275 | 0.445 | 0.252 | 0.168 | 0.175 | 0.000 |
| 214079 | 0.068 | 0.077 | 0.047 | 0.004 | 0.089 | 0.733 | -0.013 | 0.428 | 0.982 | 0.085 | 0.104 | 0.331 | 0.345 | 0.393 | 0.248 |
| 21765 | 0.543 | 0.614 | 0.160 | 0.413 | 0.630 | 0.091 | 0.427 | 0.640 | 0.346 | 0.172 | 0.263 | 0.386 | 0.064 | 0.071 | 0.066 |
| 220371 | N/A | N/A | N/A | 0.000 | 0.017 | 0.500 | 0.162 | 0.169 | 0.308 | 0.000 | 0.021 | 0.500 | 0.088 | 0.095 | 0.000 |
| 22486 | 0.039 | 0.048 | 0.187 | -0.021 | 0.094 | 1.000 | 0.214 | 0.477 | 0.497 | 0.000 | 0.021 | 0.500 | 0.640 | 0.643 | 0.021 |
| 22946 | -0.003 | 0.039 | 0.692 | N/A | N/A | N/A | 0.026 | 0.034 | 0.029 | N/A | N/A | N/A | 0.000 | 0.008 | 0.500 |
| 23539 | 0.379 | 0.385 | 0.001 | 0.118 | 0.414 | 0.411 | 0.462 | 0.466 | 0.081 | 0.170 | 0.188 | 0.063 | 0.075 | 0.511 | 0.736 |
| 261690 | 0.058 | 0.067 | 0.334 | 0.003 | 0.141 | 0.854 | 0.035 | 0.043 | 0.158 | N/A | N/A | N/A | 0.104 | 0.111 | 0.000 |
| 283548 | 0.010 | 0.019 | 0.500 | N/A | N/A | N/A | 0.137 | 0.144 | 0.210 | 0.213 | 0.229 | 0.088 | 0.192 | 0.198 | 0.207 |
| 291080 | 0.626 | 0.630 | 0.024 | 0.131 | 0.531 | 0.349 | 0.122 | 0.129 | 0.396 | 0.277 | 0.292 | 0.059 | 0.128 | 0.135 | 0.322 |
| 294904 | 0.010 | 0.019 | 0.500 | 0.000 | 0.017 | 0.500 | N/A | N/A | N/A | 0.128 | 0.146 | 0.279 | N/A | N/A | N/A |
| 313767 | 0.595 | 0.657 | 0.160 | 0.090 | 0.518 | 0.473 | 0.342 | 0.347 | 0.060 | 0.462 | 0.520 | 0.055 | 0.460 | 0.529 | 0.199 |
| 330994 | 0.146 | 0.154 | 0.282 | 0.056 | 0.329 | 0.543 | -0.005 | 0.503 | 1.000 | 0.103 | 0.471 | 0.516 | 0.264 | 0.270 | 0.000 |
| 334713 | 0.000 | 0.010 | 0.500 | N/A | N/A | N/A | 0.068 | 0.076 | 0.000 | N/A | N/A | N/A | 0.152 | 0.159 | 0.000 |
| 343160 | 0.000 | 0.010 | 0.500 | N/A | N/A | N/A | 0.103 | 0.110 | 0.059 | N/A | N/A | N/A | 0.040 | 0.048 | 0.049 |
| 353804 | 0.098 | 0.155 | 0.633 | -0.003 | 0.328 | 0.879 | 0.265 | 0.271 | 0.172 | 0.106 | 0.125 | 0.000 | 0.312 | 0.317 | 0.000 |
| 36113 | N/A | N/A | N/A | 0.000 | 0.017 | 0.500 | 0.026 | 0.034 | 0.029 | 0.021 | 0.042 | 0.674 | 0.008 | 0.016 | 0.500 |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | 0.017 | 0.025 | 0.127 | N/A | N/A | N/A | N/A | N/A | N/A |
| 39628 | 0.206 | 0.258 | 0.418 | 0.140 | 0.155 | 0.120 | 0.409 | 0.414 | 0.111 | 0.233 | 0.563 | 0.362 | 0.136 | 0.143 | 0.223 |
| 57318 | N/A | N/A | N/A | 0.000 | 0.017 | 0.500 | -0.017 | 0.325 | 1.000 | 0.000 | 0.021 | 0.500 | 0.040 | 0.048 | 0.049 |
| 58232 | -0.011 | 0.048 | 0.985 | N/A | N/A | N/A | 0.089 | 0.190 | 0.623 | 0.000 | 0.021 | 0.500 | 0.088 | 0.095 | 0.000 |
| 58352 | -0.013 | 0.058 | 0.990 | 0.018 | 0.036 | 0.616 | 0.400 | 0.619 | 0.382 | 0.021 | 0.042 | 0.674 | 0.130 | 0.137 | 0.300 |
| 58413 | -0.024 | 0.357 | 1.000 | -0.007 | 0.146 | 0.868 | 0.123 | 0.522 | 0.649 | -0.023 | 0.090 | 1.000 | 0.503 | 0.540 | 0.116 |
| 59179 | 0.097 | 0.106 | 0.313 | 0.034 | 0.122 | 0.640 | 0.251 | 0.517 | 0.477 | 0.042 | 0.392 | 0.726 | 0.376 | 0.381 | 0.159 |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | 0.238 | 0.285 | 0.281 | N/A | N/A | N/A | N/A | N/A | N/A |
| 59508 | 0.000 | 0.010 | 0.500 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.000 | 0.008 | 0.500 |
| 60276 | 0.010 | 0.019 | 0.500 | 0.000 | 0.017 | 0.500 | 0.009 | 0.017 | 0.500 | N/A | N/A | N/A | 0.000 | 0.008 | 0.500 |
| 70445 | N/A | N/A | N/A | N/A | N/A | N/A | 0.043 | 0.051 | 0.044 | N/A | N/A | N/A | 0.000 | 0.008 | 0.500 |
| 91693 | 0.134 | 0.229 | 0.528 | -0.008 | 0.420 | 0.883 | 0.735 | 0.737 | 0.031 | 0.277 | 0.292 | 0.059 | 0.052 | 0.500 | 0.787 |
| 97286 | 0.343 | 0.555 | 0.424 | 0.154 | 0.538 | 0.360 | 0.609 | 0.612 | 0.048 | 0.064 | 0.083 | 0.418 | 0.241 | 0.456 | 0.450 |
| 97831 | 0.136 | 0.144 | 0.393 | 0.175 | 0.190 | 0.129 | 0.060 | 0.068 | 0.204 | 0.234 | 0.250 | 0.186 | 0.016 | 0.024 | 0.109 |
| GH1 | 0.333 | 0.340 | 0.200 | 0.185 | 0.402 | 0.277 | 0.280 | 0.287 | 0.098 | 0.064 | 0.083 | 0.418 | 0.137 | 0.144 | 0.211 |
| GH2_165 | 0.068 | 0.077 | 0.047 | 0.070 | 0.086 | 0.385 | 0.479 | 0.483 | 0.027 | 0.191 | 0.208 | 0.264 | 0.358 | 0.363 | 0.051 |
| M009_403 | 0.437 | 0.442 | 0.136 | 0.103 | 0.491 | 0.381 | 0.234 | 0.571 | 0.559 | 0.148 | 0.294 | 0.354 | 0.192 | 0.198 | 0.207 |
| M1046_2 | 0.272 | 0.279 | 0.000 | 0.104 | 0.323 | 0.400 | 0.030 | 0.454 | 0.847 | 0.064 | 0.083 | 0.418 | 0.040 | 0.048 | 0.049 |
| MH30_Dreyer | 0.157 | 0.468 | 0.607 | 0.053 | 0.360 | 0.568 | 0.402 | 0.407 | 0.010 | -0.022 | 0.106 | 1.000 | 0.419 | 0.424 | 0.007 |
| Myostatin | 0.068 | 0.208 | 0.730 | 0.015 | 0.157 | 0.708 | 0.162 | 0.169 | 0.308 | 0.106 | 0.125 | 0.000 | 0.347 | 0.352 | 0.116 |
| Prolactin_1 | 0.126 | 0.135 | 0.000 | 0.123 | 0.138 | 0.049 | 0.403 | 0.622 | 0.377 | 0.319 | 0.333 | 0.080 | 0.328 | 0.333 | 0.209 |
| SBF1 | 0.186 | 0.239 | 0.481 | -0.004 | 0.426 | 0.857 | 0.043 | 0.051 | 0.044 | 0.128 | 0.146 | 0.279 | 0.016 | 0.024 | 0.109 |
| TBC1 | 0.286 | 0.332 | 0.251 | 0.250 | 0.375 | 0.126 | 0.093 | 0.508 | 0.704 | 0.234 | 0.250 | 0.186 | 0.440 | 0.444 | 0.087 |

| Locus | Marianne | | | Oropuche | | | Turure | | | Yara | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Fst | Het | P value | Fst | Het | P value | Fst | Het | P value | Fst | Het | P value |
| 10802 | 0.001 | 0.500 | 0.544 | -0.034 | 0.369 | 1.000 | -0.010 | 0.135 | 0.492 | 0.629 | 0.734 | 0.362 |
| 108025 | 0.034 | 0.303 | 0.452 | -0.034 | 0.369 | 1.000 | -0.025 | 0.387 | 1.000 | 0.264 | 0.500 | 0.551 |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 108291 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 111347 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 124729 | N/A | N/A | N/A | N/A | N/A | N/A | -0.017 | 0.451 | 0.681 | 0.064 | 0.083 | 0.438 |
| 148158 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 150841 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 20521 | N/A | N/A | N/A | -0.016 | 0.417 | 0.563 | -0.020 | 0.061 | 1.000 | 0.468 | 0.479 | 0.091 |
| 207392 | N/A | N/A | N/A | 0.012 | 0.304 | 0.278 | -0.020 | 0.167 | 0.763 | N/A | N/A | N/A |
| 214079 | N/A | N/A | N/A | N/A | N/A | N/A | 0.004 | 0.251 | 0.303 | N/A | N/A | N/A |
| 21765 | N/A | N/A | N/A | N/A | N/A | N/A | -0.021 | 0.482 | 0.793 | 0.000 | 0.021 | 0.500 |
| 220371 | 0.021 | 0.429 | 0.506 | -0.033 | 0.085 | 0.500 | -0.022 | 0.178 | 0.838 | 0.599 | 0.682 | 0.288 |
| 22486 | 0.027 | 0.053 | 0.556 | 0.003 | 0.464 | 0.343 | -0.021 | 0.086 | 1.000 | 0.021 | 0.042 | 0.554 |
| 22946 | 0.000 | 0.019 | 0.500 | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 23539 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | 0.511 | 0.521 | 0.182 |
| 261690 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 283548 | N/A | N/A | N/A | 0.121 | 0.474 | 0.009 | -0.014 | 0.222 | 0.563 | N/A | N/A | N/A |
| 291080 | 0.000 | 0.026 | 0.448 | -0.032 | 0.163 | 0.977 | 0.113 | 0.130 | 0.004 | 0.468 | 0.479 | 0.091 |
| 294904 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 313767 | 0.172 | 0.551 | 0.143 | 0.004 | 0.508 | 0.343 | -0.022 | 0.500 | 0.847 | 0.022 | 0.141 | 0.918 |
| 330994 | N/A | N/A | N/A | 0.043 | 0.083 | 0.500 | 0.042 | 0.295 | 0.091 | N/A | N/A | N/A |
| 334713 | N/A | N/A | N/A | N/A | N/A | N/A | 0.053 | 0.149 | 0.074 | N/A | N/A | N/A |
| 343160 | N/A | N/A | N/A | N/A | N/A | N/A | 0.014 | 0.095 | 0.231 | N/A | N/A | N/A |
| 353804 | 0.088 | 0.471 | 0.230 | -0.030 | 0.500 | 0.913 | -0.018 | 0.119 | 0.688 | 0.936 | 0.938 | 0.168 |
| 36113 | N/A | N/A | N/A | N/A | N/A | N/A | 0.000 | 0.019 | 0.500 | N/A | N/A | N/A |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 39628 | 0.051 | 0.518 | 0.374 | 0.043 | 0.083 | 0.500 | 0.041 | 0.484 | 0.073 | 0.085 | 0.104 | 0.185 |
| 57318 | 0.000 | 0.026 | 0.448 | N/A | N/A | N/A | N/A | N/A | 1.000 | 0.106 | 0.125 | 0.009 |
| 58232 | N/A | N/A | N/A | N/A | N/A | N/A | 0.000 | 0.019 | 0.500 | 0.000 | 0.021 | 0.500 |
| 58352 | N/A | N/A | N/A | N/A | N/A | N/A | -0.022 | 0.044 | 0.500 | 0.152 | 0.329 | 0.660 |
| 58413 | 0.042 | 0.508 | 0.416 | 0.025 | 0.279 | 0.173 | -0.007 | 0.149 | 0.447 | 0.213 | 0.229 | 0.304 |
| 59179 | N/A | N/A | N/A | -0.001 | 0.486 | 0.312 | 0.150 | 0.343 | 0.002 | 0.149 | 0.167 | 0.191 |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 59508 | -0.023 | 0.106 | 0.915 | N/A | N/A | N/A | 0.000 | 0.019 | 0.500 | 0.000 | 0.063 | 0.749 |
| 60276 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 70445 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | N/A | N/A | N/A |
| 91693 | N/A | N/A | N/A | 0.043 | 0.083 | 0.500 | -0.020 | 0.385 | 0.806 | 0.149 | 0.167 | 0.191 |
| 97286 | -0.023 | 0.365 | 0.907 | 0.000 | 0.042 | 101.000 | 0.018 | 0.401 | 0.178 | 0.553 | 0.563 | 0.000 |
| 97831 | 0.033 | 0.190 | 0.463 | 0.003 | 0.464 | 0.343 | -0.007 | 0.424 | 0.450 | 0.133 | 0.188 | 0.338 |
| GH1 | N/A | N/A | N/A | -0.033 | 0.085 | 0.500 | N/A | N/A | 1.000 | -0.036 | 0.453 | 1.000 |
| GH2_165 | -0.009 | 0.111 | 0.750 | 0.130 | 0.167 | 0.014 | -0.018 | 0.164 | 0.695 | 0.170 | 0.188 | 0.019 |
| M009_403 | -0.016 | 0.415 | 0.827 | -0.024 | 0.405 | 0.730 | 0.009 | 0.335 | 0.279 | -0.044 | 0.117 | 1.000 |
| M1046_2 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 1.000 | 0.319 | 0.333 | 0.052 |
| MH30_Dreyer | 0.000 | 0.026 | 0.448 | -0.026 | 0.496 | 0.710 | -0.018 | 0.164 | 0.695 | 0.599 | 0.682 | 0.288 |
| Myostatin | N/A | N/A | N/A | -0.029 | 0.431 | 0.829 | -0.022 | 0.241 | 0.850 | N/A | N/A | N/A |
| Prolactin_1 | N/A | N/A | N/A | 0.043 | 0.083 | 0.500 | -0.006 | 0.237 | 0.423 | 0.085 | 0.104 | 0.185 |
| SBF1 | 0.177 | 0.528 | 0.118 | 0.042 | 0.389 | 0.105 | 0.000 | 0.019 | 0.500 | 0.511 | 0.521 | 0.182 |
| TBC1 | N/A | N/A | N/A | -0.020 | 0.179 | 0.572 | 0.078 | 0.235 | 0.035 | 0.426 | 0.438 | 0.242 |

**Appendix IX: Values from Arelquin outlier analysis. N/A values denote a SNP which was monomorphic in the respective river.**

| Locus | Aripo 03 | | | Aripo 06 | | | Caura | | | Guanapo | | | Lopinot | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | He | FST | P value | He | FST | P value | He | FST | P value | He | FST | P value | He | FST | P value |
| 10802 | N/A | N/A | N/A | 0.059 | 0.025 | 0.288 | 0.400 | 0.189 | 0.273 | 0.023 | -0.004 | 0.374 | 0.198 | 0.099 | 0.407 |
| 108025 | 0.243 | 0.045 | 0.177 | 0.304 | 0.128 | 0.285 | 0.562 | 0.322 | 0.162 | 0.298 | -0.020 | 0.065 | 0.618 | 0.334 | 0.290 |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | 0.046 | 0.180 | 0.105 | N/A | N/A | N/A | N/A | N/A | N/A |
| 108291 | 0.023 | 0.003 | 0.096 | N/A | N/A | N/A | 0.046 | 0.180 | 0.105 | N/A | N/A | N/A | N/A | N/A | N/A |
| 111347 | 0.012 | -0.005 | 0.350 | N/A | N/A | N/A | 0.026 | -0.009 | 0.246 | N/A | N/A | N/A | 0.011 | -0.007 | 0.280 |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | 0.052 | 0.002 | 0.031 | N/A | N/A | N/A | N/A | N/A | N/A |
| 124729 | 0.670 | 0.601 | 0.061 | 0.405 | 0.338 | 0.033 | 0.165 | 0.055 | 0.355 | 0.266 | 0.216 | 0.136 | 0.054 | 0.016 | 0.219 |
| 148158 | N/A | N/A | N/A | N/A | N/A | N/A | 0.167 | 0.056 | 0.356 | N/A | N/A | N/A | 0.138 | 0.065 | 0.412 |
| 150841 | N/A | N/A | N/A | N/A | N/A | N/A | 0.053 | 0.003 | 0.037 | 0.023 | -0.004 | 0.374 | 0.033 | 0.005 | 0.115 |
| 20521 | 0.070 | 0.036 | 0.370 | 0.118 | 0.071 | 0.309 | 0.668 | 0.377 | 0.249 | 0.045 | 0.016 | 0.257 | 0.138 | 0.065 | 0.412 |
| 207392 | 0.409 | 0.161 | 0.446 | 0.514 | 0.126 | 0.272 | 0.248 | 0.098 | 0.413 | 0.462 | 0.266 | 0.154 | 0.229 | 0.121 | 0.362 |
| 214079 | 0.093 | 0.052 | 0.415 | 0.096 | 0.001 | 0.006 | 0.437 | -0.013 | 0.076 | 0.112 | 0.075 | 0.334 | 0.476 | 0.279 | 0.166 |
| 21765 | 0.655 | 0.506 | 0.151 | 0.622 | 0.418 | 0.046 | 0.629 | 0.435 | 0.122 | 0.276 | 0.158 | 0.294 | 0.096 | 0.040 | 0.350 |
| 220371 | N/A | N/A | N/A | 0.020 | -0.006 | 0.324 | 0.248 | 0.098 | 0.413 | 0.023 | -0.004 | 0.374 | 0.128 | 0.059 | 0.390 |
| 22486 | 0.058 | 0.028 | 0.339 | 0.095 | -0.021 | 0.070 | 0.401 | 0.263 | 0.191 | 0.023 | -0.004 | 0.374 | 0.728 | 0.560 | 0.103 |
| 22946 | 0.035 | 0.000 | 0.401 | N/A | N/A | N/A | 0.052 | 0.002 | 0.031 | N/A | N/A | N/A | 0.011 | -0.007 | 0.280 |
| 23539 | 0.436 | 0.330 | 0.157 | 0.431 | 0.111 | 0.315 | 0.603 | 0.345 | 0.175 | 0.201 | 0.155 | 0.208 | 0.522 | 0.073 | 0.279 |
| 261690 | 0.081 | 0.044 | 0.393 | 0.149 | 0.000 | 0.003 | 0.066 | 0.008 | 0.109 | N/A | N/A | N/A | 0.148 | 0.071 | 0.431 |
| 283548 | 0.023 | 0.003 | 0.096 | N/A | N/A | N/A | 0.212 | 0.079 | 0.413 | 0.245 | 0.196 | 0.160 | 0.258 | 0.141 | 0.331 |
| 291080 | 0.690 | 0.583 | 0.103 | 0.525 | 0.133 | 0.287 | 0.192 | 0.068 | 0.391 | 0.308 | 0.252 | 0.126 | 0.179 | 0.090 | 0.410 |
| 294904 | 0.023 | 0.003 | 0.096 | 0.020 | -0.006 | 0.324 | N/A | N/A | N/A | 0.156 | 0.112 | 0.279 | 0.000 | 0.000 | -1.000 |
| 313767 | 0.696 | 0.560 | 0.146 | 0.513 | 0.091 | 0.369 | 0.473 | 0.239 | 0.185 | 0.539 | 0.442 | 0.043 | 0.610 | 0.395 | 0.166 |
| 330994 | 0.182 | 0.119 | 0.366 | 0.344 | 0.052 | 0.273 | 0.502 | -0.005 | 0.124 | 0.463 | 0.106 | 0.382 | 0.344 | 0.201 | 0.264 |
| 334713 | 0.012 | -0.005 | 0.350 | N/A | N/A | N/A | 0.115 | 0.031 | 0.283 | N/A | N/A | N/A | 0.209 | 0.109 | 0.385 |
| 343160 | 0.012 | -0.005 | 0.350 | N/A | N/A | N/A | 0.165 | 0.055 | 0.355 | N/A | N/A | N/A | 0.065 | 0.022 | 0.271 |
| 353804 | 0.179 | 0.082 | 0.366 | 0.334 | -0.004 | 0.170 | 0.381 | 0.176 | 0.290 | 0.135 | 0.095 | 0.296 | 0.399 | 0.242 | 0.219 |
| 36113 | N/A | N/A | N/A | 0.020 | -0.006 | 0.324 | 0.052 | 0.002 | 0.031 | 0.045 | 0.014 | 0.234 | 0.022 | -0.001 | 0.428 |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | 0.039 | -0.004 | 0.337 | N/A | N/A | N/A | N/A | N/A | N/A |
| 39628 | 0.295 | 0.176 | 0.356 | 0.176 | 0.119 | 0.234 | 0.548 | 0.298 | 0.170 | 0.567 | 0.231 | 0.302 | 0.189 | 0.096 | 0.401 |
| 57318 | N/A | N/A | N/A | 0.020 | -0.006 | 0.324 | 0.333 | -0.017 | 0.051 | 0.023 | -0.004 | 0.374 | 0.065 | 0.022 | 0.271 |
| 58232 | 0.046 | -0.010 | 0.103 | N/A | N/A | N/A | 0.251 | 0.057 | 0.252 | 0.023 | -0.004 | 0.374 | 0.128 | 0.059 | 0.390 |
| 58352 | 0.058 | -0.013 | 0.061 | 0.041 | 0.011 | 0.189 | 0.579 | 0.429 | 0.103 | 0.045 | 0.016 | 0.257 | 0.182 | 0.092 | 0.406 |
| 58413 | 0.352 | -0.023 | 0.011 | 0.153 | -0.015 | 0.100 | 0.545 | 0.116 | 0.371 | 0.089 | -0.023 | 0.059 | 0.628 | 0.431 | 0.146 |
| 59179 | 0.127 | 0.077 | 0.414 | 0.133 | 0.028 | 0.188 | 0.581 | 0.219 | 0.401 | 0.398 | 0.041 | 0.177 | 0.470 | 0.299 | 0.142 |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | 0.176 | 0.418 | 0.108 | N/A | N/A | N/A | N/A | N/A | N/A |
| 59508 | 0.012 | -0.005 | 0.350 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.011 | -0.007 | 0.280 |
| 60276 | 0.023 | 0.003 | 0.096 | 0.020 | -0.006 | 0.324 | 0.026 | -0.009 | 0.246 | N/A | N/A | N/A | 0.011 | -0.007 | 0.280 |
| 70445 | N/A | N/A | N/A | N/A | N/A | N/A | 0.078 | 0.014 | 0.164 | N/A | N/A | N/A | 0.011 | -0.007 | 0.280 |
| 91693 | 0.259 | 0.116 | 0.408 | 0.415 | -0.007 | 0.154 | 0.842 | 0.635 | 0.115 | 0.310 | 0.257 | 0.117 | 0.511 | 0.050 | 0.227 |
| 97286 | 0.583 | 0.325 | 0.304 | 0.544 | 0.152 | 0.312 | 0.743 | 0.493 | 0.171 | 0.090 | 0.053 | 0.386 | 0.513 | 0.211 | 0.271 |
| 97831 | 0.171 | 0.110 | 0.373 | 0.214 | 0.151 | 0.180 | 0.103 | 0.025 | 0.254 | 0.266 | 0.216 | 0.136 | 0.033 | 0.005 | 0.115 |
| GH1 | 0.387 | 0.282 | 0.200 | 0.425 | 0.172 | 0.193 | 0.402 | 0.192 | 0.269 | 0.090 | 0.055 | 0.367 | 0.191 | 0.099 | 0.392 |
| GH2_165 | 0.092 | 0.049 | 0.393 | 0.098 | 0.056 | 0.348 | 0.621 | 0.361 | 0.174 | 0.223 | 0.175 | 0.185 | 0.450 | 0.284 | 0.162 |
| M9_403 | 0.496 | 0.386 | 0.105 | 0.501 | 0.100 | 0.315 | 0.568 | 0.235 | 0.318 | 0.307 | 0.139 | 0.342 | 0.258 | 0.141 | 0.331 |
| M1046_2 | 0.323 | 0.231 | 0.247 | 0.303 | 0.113 | 0.320 | 0.422 | 0.036 | 0.151 | 0.090 | 0.055 | 0.367 | 0.065 | 0.022 | 0.271 |
| MH30_Dreyer | 0.491 | 0.148 | 0.417 | 0.374 | 0.050 | 0.270 | 0.540 | 0.291 | 0.163 | 0.107 | -0.022 | 0.060 | 0.519 | 0.343 | 0.114 |
| Myostatin | 0.229 | 0.059 | 0.233 | 0.166 | 0.012 | 0.081 | 0.248 | 0.098 | 0.413 | 0.135 | 0.095 | 0.296 | 0.440 | 0.275 | 0.174 |
| Prolactin_1 | 0.160 | 0.102 | 0.377 | 0.156 | 0.103 | 0.262 | 0.653 | 0.382 | 0.209 | 0.353 | 0.298 | 0.091 | 0.417 | 0.256 | 0.201 |
| SBF1 | 0.275 | 0.159 | 0.382 | 0.431 | -0.005 | 0.175 | 0.078 | 0.014 | 0.164 | 0.157 | 0.115 | 0.265 | 0.033 | 0.005 | 0.115 |
| TBC1 | 0.377 | 0.248 | 0.273 | 0.404 | 0.228 | 0.126 | 0.531 | 0.087 | 0.309 | 0.266 | 0.216 | 0.136 | 0.537 | 0.358 | 0.120 |

| Locus | Marianne | | | Oropuche | | | Turure | | | Yara | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | He | FST | P value | He | FST | P value | He | FST | P value | He | FST | P value |
| 10802 | 0.502 | 0.001 | 0.012 | 0.367 | -0.033 | 0.134 | 0.141 | -0.011 | 0.344 | 0.751 | 0.612 | 0.180 |
| 108025 | 0.289 | 0.042 | 0.291 | 0.367 | -0.033 | 0.134 | 0.386 | -0.024 | 0.139 | 0.434 | 0.357 | 0.186 |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 108291 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 111347 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 124729 | N/A | N/A | N/A | N/A | N/A | N/A | 0.448 | -0.016 | 0.266 | 0.119 | 0.017 | 0.180 |
| 148158 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 150841 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 20521 | N/A | N/A | N/A | 0.425 | -0.017 | 0.382 | 0.063 | -0.021 | 0.180 | 0.594 | 0.356 | 0.271 |
| 207392 | N/A | N/A | N/A | 0.322 | 0.007 | 0.068 | 0.170 | -0.020 | 0.199 | N/A | N/A | N/A |
| 214079 | N/A | N/A | N/A | N/A | N/A | N/A | 0.261 | 0.002 | 0.031 | N/A | N/A | N/A |
| 21765 | N/A | N/A | N/A | N/A | N/A | N/A | 0.481 | -0.021 | 0.204 | 0.030 | -0.029 | 0.048 |
| 220371 | 0.421 | 0.023 | 0.185 | 0.087 | -0.033 | 0.142 | 0.177 | -0.022 | 0.172 | 0.747 | 0.535 | 0.289 |
| 22486 | 0.045 | 0.040 | 0.135 | 0.453 | 0.005 | 0.049 | 0.083 | -0.020 | 0.198 | 0.060 | -0.014 | 0.322 |
| 22946 | 0.022 | -0.006 | 0.311 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 23539 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.635 | 0.398 | 0.246 |
| 261690 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 283548 | N/A | N/A | N/A | 0.449 | 0.132 | 0.017 | 0.228 | -0.015 | 0.280 | N/A | N/A | N/A |
| 291080 | 0.022 | 0.008 | 0.233 | 0.165 | -0.033 | 0.141 | 0.149 | 0.092 | 0.030 | 0.594 | 0.356 | 0.271 |
| 294904 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 313767 | 0.547 | 0.173 | 0.111 | 0.505 | 0.005 | 0.042 | 0.500 | -0.022 | 0.171 | 0.097 | 0.075 | 0.312 |
| 330994 | N/A | N/A | N/A | 0.063 | 0.081 | 0.041 | 0.310 | 0.038 | 0.147 | N/A | N/A | N/A |
| 334713 | N/A | N/A | N/A | N/A | N/A | N/A | 0.164 | 0.044 | 0.120 | N/A | N/A | N/A |
| 343160 | N/A | N/A | N/A | N/A | N/A | N/A | 0.085 | 0.021 | 0.213 | N/A | N/A | N/A |
| 353804 | 0.459 | 0.091 | 0.191 | 0.499 | -0.030 | 0.183 | 0.122 | -0.019 | 0.209 | 0.956 | 0.906 | 0.126 |
| 36113 | N/A | N/A | N/A | N/A | N/A | N/A | 0.022 | -0.007 | 0.311 | N/A | N/A | N/A |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 39628 | 0.517 | 0.051 | 0.353 | 0.063 | 0.081 | 0.041 | 0.491 | 0.040 | 0.122 | 0.148 | 0.032 | 0.244 |
| 57318 | 0.022 | 0.008 | 0.233 | N/A | N/A | N/A | N/A | N/A | N/A | 0.176 | 0.048 | 0.276 |
| 58232 | N/A | N/A | N/A | N/A | N/A | N/A | 0.022 | -0.007 | 0.311 | 0.030 | -0.029 | 0.048 |
| 58352 | N/A | N/A | N/A | N/A | N/A | N/A | 0.043 | -0.021 | 0.128 | 0.399 | 0.109 | 0.268 |
| 58413 | 0.504 | 0.043 | 0.311 | 0.255 | 0.036 | 0.167 | 0.142 | -0.005 | 0.455 | 0.311 | 0.131 | 0.414 |
| 59179 | N/A | N/A | N/A | 0.478 | 0.000 | 0.356 | 0.318 | 0.165 | 0.004 | 0.231 | 0.080 | 0.326 |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 59508 | 0.107 | -0.023 | 0.085 | N/A | N/A | N/A | 0.022 | -0.007 | 0.311 | 0.034 | 0.080 | 0.068 |
| 60276 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 70445 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 91693 | N/A | N/A | N/A | 0.063 | 0.081 | 0.041 | 0.381 | -0.020 | 0.205 | 0.231 | 0.080 | 0.326 |
| 97286 | 0.364 | -0.023 | 0.087 | 0.031 | 0.025 | 0.060 | 0.391 | 0.021 | 0.215 | 0.674 | 0.441 | 0.291 |
| 97831 | 0.202 | 0.028 | 0.221 | 0.453 | 0.005 | 0.049 | 0.418 | -0.006 | 0.447 | 0.114 | 0.288 | 0.107 |
| GH1 | N/A | N/A | N/A | 0.087 | -0.033 | 0.142 | N/A | N/A | N/A | 0.439 | -0.034 | 0.061 |
| GH2_165 | 0.106 | -0.007 | 0.239 | 0.130 | 0.190 | 0.005 | 0.159 | -0.017 | 0.256 | 0.258 | 0.097 | 0.383 |
| M9_403 | 0.418 | -0.016 | 0.157 | 0.411 | -0.025 | 0.252 | 0.345 | 0.008 | 0.101 | 0.116 | -0.043 | 0.041 |
| M1046_2 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.436 | 0.220 | 0.311 |
| MH30_Dreyer | 0.022 | 0.008 | 0.233 | 0.497 | -0.026 | 0.242 | 0.159 | -0.017 | 0.256 | 0.747 | 0.535 | 0.289 |
| Myostatin | N/A | N/A | N/A | 0.434 | -0.030 | 0.185 | 0.243 | -0.022 | 0.165 | N/A | N/A | N/A |
| Prolactin_1 | N/A | N/A | N/A | 0.063 | 0.081 | 0.041 | 0.244 | -0.007 | 0.410 | 0.148 | 0.032 | 0.244 |
| SBF1 | 0.518 | 0.182 | 0.063 | 0.410 | 0.036 | 0.164 | 0.022 | -0.007 | 0.311 | 0.635 | 0.398 | 0.246 |
| TBC1 | N/A | N/A | N/A | 0.167 | -0.016 | 0.393 | 0.254 | 0.069 | 0.062 | 0.551 | 0.315 | 0.239 |

**Appendix X: Values from BayeScan outlier analysis.**

PP shows the posterior probability of the model including selection. Log10(PO) shows the logarithm of Posterior Odds to the base 10 for the model including selection. N/A values denote a SNP which was monomorphic in the respective river.

| Locus | Aripo 03 | | | | | Aripo 06 | | | | | Caura | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PP | Log10 (PO) | Q value | Alpha | Fst | PP | Log10 (PO) | Q value | Alpha | Fst | PP | Log10 (PO) | Q value | Alpha | Fst |
| 10802 | N/A | N/A | N/A | N/A | N/A | 0.088 | -1.018 | 0.905 | 0.015 | 0.156 | 0.085 | -1.032 | 0.906 | 0.021 | 0.352 |
| 108025 | 0.099 | -0.961 | 0.895 | -0.064 | 0.308 | 0.072 | -1.108 | 0.912 | -0.008 | 0.152 | 0.090 | -1.004 | 0.900 | -0.028 | 0.345 |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.100 | -0.954 | 0.892 | 0.015 | 0.352 |
| 108291 | 0.097 | -0.970 | 0.898 | 0.017 | 0.320 | N/A | N/A | N/A | N/A | N/A | 0.103 | -0.939 | 0.891 | 0.026 | 0.353 |
| 111347 | 0.092 | -0.995 | 0.902 | 0.027 | 0.322 | N/A | N/A | N/A | N/A | N/A | 0.107 | -0.920 | 0.888 | 0.017 | 0.352 |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.087 | -1.020 | 0.904 | 0.014 | 0.351 |
| 124729 | 0.068 | -1.134 | 0.911 | 0.004 | 0.318 | 0.074 | -1.095 | 0.910 | 0.019 | 0.155 | 0.087 | -1.020 | 0.904 | 0.011 | 0.351 |
| 148158 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.089 | -1.009 | 0.901 | 0.002 | 0.350 |
| 150841 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.091 | -1.002 | 0.900 | 0.006 | 0.350 |
| 20521 | 0.091 | -1.000 | 0.903 | 0.016 | 0.320 | 0.088 | -1.017 | 0.905 | 0.032 | 0.158 | 0.076 | -1.082 | 0.911 | -0.017 | 0.346 |
| 207392 | 0.091 | -1.001 | 0.903 | -0.037 | 0.312 | 0.078 | -1.073 | 0.909 | -0.012 | 0.152 | 0.082 | -1.047 | 0.906 | 0.017 | 0.352 |
| 214079 | 0.082 | -1.050 | 0.906 | 0.006 | 0.318 | 0.092 | -0.996 | 0.902 | -0.039 | 0.150 | 0.103 | -0.939 | 0.891 | -0.070 | 0.339 |
| 21765 | 0.067 | -1.141 | 0.912 | 0.014 | 0.319 | 0.079 | -1.067 | 0.908 | 0.039 | 0.158 | 0.082 | -1.050 | 0.907 | -0.019 | 0.346 |
| 220371 | N/A | N/A | N/A | N/A | N/A | 0.096 | -0.975 | 0.901 | 0.017 | 0.156 | 0.081 | -1.052 | 0.907 | 0.003 | 0.350 |
| 22486 | 0.087 | -1.021 | 0.904 | 0.013 | 0.319 | 0.097 | -0.968 | 0.900 | -0.041 | 0.150 | 0.092 | -0.992 | 0.897 | -0.040 | 0.343 |
| 22946 | 0.097 | -0.968 | 0.898 | -0.031 | 0.313 | N/A | N/A | N/A | N/A | N/A | 0.092 | -0.993 | 0.897 | 0.004 | 0.350 |
| 23539 | 0.092 | -0.992 | 0.900 | 0.052 | 0.325 | 0.069 | -1.130 | 0.913 | -0.008 | 0.152 | 0.088 | -1.013 | 0.903 | 0.039 | 0.355 |
| 261690 | 0.087 | -1.020 | 0.904 | 0.013 | 0.319 | 0.098 | -0.966 | 0.900 | -0.031 | 0.151 | 0.088 | -1.013 | 0.903 | 0.007 | 0.350 |
| 283548 | 0.092 | -0.992 | 0.900 | 0.018 | 0.320 | N/A | N/A | N/A | N/A | N/A | 0.081 | -1.055 | 0.908 | 0.014 | 0.351 |
| 291080 | 0.119 | -0.869 | 0.881 | 0.105 | 0.334 | 0.077 | -1.079 | 0.909 | -0.016 | 0.152 | 0.080 | -1.059 | 0.908 | 0.008 | 0.350 |
| 294904 | 0.097 | -0.968 | 0.898 | 0.019 | 0.320 | 0.095 | -0.980 | 0.901 | 0.020 | 0.157 | N/A | N/A | N/A | N/A | N/A |
| 313767 | 0.071 | -1.117 | 0.910 | 0.011 | 0.318 | 0.079 | -1.064 | 0.908 | -0.022 | 0.151 | 0.082 | -1.048 | 0.907 | 0.027 | 0.353 |
| 330994 | 0.091 | -0.997 | 0.902 | 0.020 | 0.320 | 0.082 | -1.051 | 0.907 | -0.024 | 0.151 | 0.112 | -0.898 | 0.886 | -0.091 | 0.337 |
| 334713 | 0.102 | -0.946 | 0.891 | 0.030 | 0.322 | N/A | N/A | N/A | N/A | N/A | 0.085 | -1.033 | 0.906 | 0.009 | 0.350 |
| 343160 | 0.099 | -0.958 | 0.893 | 0.026 | 0.322 | N/A | N/A | N/A | N/A | N/A | 0.084 | -1.040 | 0.906 | 0.007 | 0.350 |
| 353804 | 0.088 | -1.018 | 0.903 | -0.034 | 0.312 | 0.091 | -1.002 | 0.903 | -0.040 | 0.150 | 0.079 | -1.065 | 0.909 | 0.020 | 0.352 |
| 36113 | N/A | N/A | N/A | N/A | N/A | 0.099 | -0.960 | 0.899 | 0.027 | 0.158 | 0.099 | -0.958 | 0.893 | 0.005 | 0.350 |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.092 | -0.996 | 0.899 | 0.011 | 0.351 |
| 39628 | 0.076 | -1.085 | 0.909 | -0.014 | 0.315 | 0.097 | -0.970 | 0.900 | 0.044 | 0.159 | 0.078 | -1.074 | 0.910 | 0.031 | 0.354 |
| 57318 | N/A | N/A | N/A | N/A | N/A | 0.100 | -0.954 | 0.899 | 0.018 | 0.157 | 0.118 | -0.872 | 0.882 | -0.095 | 0.336 |
| 58232 | 0.098 | -0.962 | 0.896 | -0.034 | 0.313 | N/A | N/A | N/A | N/A | N/A | 0.093 | -0.989 | 0.896 | -0.044 | 0.343 |
| 58352 | 0.103 | -0.941 | 0.889 | -0.052 | 0.311 | 0.095 | -0.977 | 0.901 | 0.012 | 0.156 | 0.078 | -1.073 | 0.910 | -0.023 | 0.346 |
| 58413 | 0.111 | -0.903 | 0.885 | -0.082 | 0.306 | 0.083 | -1.044 | 0.906 | -0.025 | 0.151 | 0.090 | -1.005 | 0.901 | -0.051 | 0.342 |
| 59179 | 0.082 | -1.051 | 0.907 | 0.013 | 0.319 | 0.089 | -1.011 | 0.903 | -0.022 | 0.152 | 0.086 | -1.024 | 0.905 | -0.031 | 0.345 |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.079 | -1.068 | 0.909 | 0.002 | 0.349 |
| 59508 | 0.098 | -0.966 | 0.897 | 0.031 | 0.322 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 60276 | 0.084 | -1.035 | 0.905 | 0.012 | 0.319 | 0.099 | -0.960 | 0.899 | 0.029 | 0.158 | 0.089 | -1.012 | 0.902 | 0.021 | 0.352 |
| 70445 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.089 | -1.010 | 0.902 | 0.004 | 0.350 |
| 91693 | 0.082 | -1.048 | 0.906 | -0.031 | 0.313 | 0.101 | -0.948 | 0.897 | -0.046 | 0.149 | 0.112 | -0.897 | 0.885 | 0.093 | 0.363 |
| 97286 | 0.069 | -1.129 | 0.911 | -0.016 | 0.315 | 0.079 | -1.068 | 0.909 | -0.007 | 0.152 | 0.099 | -0.961 | 0.894 | 0.065 | 0.359 |
| 97831 | 0.077 | -1.079 | 0.908 | 0.021 | 0.320 | 0.101 | -0.950 | 0.898 | 0.052 | 0.160 | 0.091 | -0.997 | 0.899 | 0.005 | 0.350 |
| GH1 | 0.082 | -1.051 | 0.907 | 0.038 | 0.323 | 0.071 | -1.115 | 0.913 | 0.007 | 0.154 | 0.075 | -1.090 | 0.911 | 0.016 | 0.351 |
| GH2_165 | 0.085 | -1.032 | 0.905 | 0.017 | 0.320 | 0.086 | -1.028 | 0.906 | 0.026 | 0.157 | 0.086 | -1.025 | 0.905 | 0.038 | 0.355 |
| M9_403 | 0.095 | -0.979 | 0.899 | 0.062 | 0.327 | 0.074 | -1.096 | 0.911 | -0.011 | 0.152 | 0.092 | -0.994 | 0.898 | -0.045 | 0.343 |
| M1046_2 | 0.072 | -1.109 | 0.909 | 0.026 | 0.321 | 0.089 | -1.012 | 0.904 | -0.028 | 0.151 | 0.108 | -0.919 | 0.887 | -0.077 | 0.338 |
| MH30_Dreyer | 0.086 | -1.024 | 0.904 | -0.037 | 0.312 | 0.086 | -1.025 | 0.905 | -0.024 | 0.151 | 0.077 | -1.079 | 0.910 | 0.025 | 0.353 |
| Myostatin | 0.098 | -0.962 | 0.896 | -0.043 | 0.311 | 0.091 | -1.002 | 0.903 | -0.030 | 0.151 | 0.085 | -1.031 | 0.905 | 0.022 | 0.353 |
| Prolactin_1 | 0.080 | -1.058 | 0.907 | 0.019 | 0.320 | 0.089 | -1.012 | 0.904 | 0.034 | 0.158 | 0.080 | -1.059 | 0.908 | -0.024 | 0.345 |
| SBF1 | 0.078 | -1.073 | 0.908 | -0.018 | 0.314 | 0.101 | -0.951 | 0.898 | -0.049 | 0.149 | 0.088 | -1.018 | 0.904 | 0.000 | 0.349 |
| TBC1 | 0.071 | -1.115 | 0.910 | -0.007 | 0.316 | 0.073 | -1.106 | 0.912 | 0.015 | 0.155 | 0.096 | -0.975 | 0.895 | -0.057 | 0.341 |

| Locus | Guanapo | | | | | Lopinot | | | | | Marianne | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PP | Log10 (PO) | Q value | Alpha | Fst | PP | Log10 (PO) | Q value | Alpha | Fst | PP | Log10 (PO) | Q value | Alpha | Fst |
| 10802 | 0.089 | -1.010 | 0.903 | 0.018 | 0.326 | 0.091 | -0.999 | 0.902 | 0.009 | 0.432 | 0.090 | -1.003 | 0.905 | -0.032 | 0.103 |
| 108025 | 0.108 | -0.919 | 0.890 | -0.076 | 0.315 | 0.096 | -0.973 | 0.897 | -0.062 | 0.425 | 0.085 | -1.030 | 0.907 | -0.019 | 0.104 |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 108291 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 111347 | N/A | N/A | N/A | N/A | N/A | 0.098 | -0.964 | 0.895 | 0.026 | 0.434 | N/A | N/A | N/A | N/A | N/A |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 124729 | 0.082 | -1.047 | 0.909 | 0.038 | 0.329 | 0.093 | -0.988 | 0.899 | 0.010 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 148158 | N/A | N/A | N/A | N/A | N/A | 0.084 | -1.039 | 0.908 | 0.009 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 150841 | 0.092 | -0.996 | 0.898 | 0.016 | 0.326 | 0.092 | -0.994 | 0.900 | 0.012 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 20521 | 0.094 | -0.984 | 0.895 | 0.011 | 0.326 | 0.088 | -1.015 | 0.904 | 0.002 | 0.431 | N/A | N/A | N/A | N/A | N/A |
| 207392 | 0.083 | -1.046 | 0.909 | -0.024 | 0.321 | 0.087 | -1.023 | 0.906 | 0.011 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 214079 | 0.082 | -1.048 | 0.910 | 0.015 | 0.326 | 0.095 | -0.977 | 0.898 | -0.038 | 0.427 | N/A | N/A | N/A | N/A | N/A |
| 21765 | 0.084 | -1.040 | 0.909 | -0.022 | 0.321 | 0.085 | -1.030 | 0.907 | 0.005 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 220371 | 0.092 | -0.993 | 0.897 | 0.010 | 0.326 | 0.088 | -1.018 | 0.904 | 0.008 | 0.432 | 0.089 | -1.009 | 0.906 | -0.024 | 0.104 |
| 22486 | 0.090 | -1.007 | 0.902 | 0.013 | 0.326 | 0.079 | -1.065 | 0.910 | 0.034 | 0.435 | 0.095 | -0.980 | 0.903 | 0.023 | 0.109 |
| 22946 | N/A | N/A | N/A | N/A | N/A | 0.103 | -0.940 | 0.888 | 0.020 | 0.433 | 0.095 | -0.980 | 0.903 | 0.010 | 0.108 |
| 23539 | 0.086 | -1.028 | 0.906 | 0.020 | 0.327 | 0.113 | -0.893 | 0.884 | -0.088 | 0.422 | N/A | N/A | N/A | N/A | N/A |
| 261690 | N/A | N/A | N/A | N/A | N/A | 0.087 | -1.020 | 0.905 | 0.010 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 283548 | 0.082 | -1.048 | 0.910 | 0.027 | 0.328 | 0.083 | -1.046 | 0.908 | 0.006 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 291080 | 0.087 | -1.021 | 0.905 | 0.033 | 0.328 | 0.087 | -1.020 | 0.905 | 0.007 | 0.432 | 0.099 | -0.961 | 0.899 | 0.025 | 0.110 |
| 294904 | 0.081 | -1.052 | 0.910 | 0.012 | 0.326 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 313767 | 0.073 | -1.105 | 0.912 | 0.006 | 0.325 | 0.088 | -1.013 | 0.904 | -0.041 | 0.427 | 0.079 | -1.068 | 0.910 | 0.014 | 0.107 |
| 330994 | 0.109 | -0.911 | 0.888 | -0.064 | 0.316 | 0.080 | -1.061 | 0.909 | 0.018 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 334713 | N/A | N/A | N/A | N/A | N/A | 0.089 | -1.012 | 0.903 | 0.002 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 343160 | N/A | N/A | N/A | N/A | N/A | 0.087 | -1.023 | 0.906 | 0.006 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 353804 | 0.073 | -1.105 | 0.912 | 0.015 | 0.326 | 0.081 | -1.057 | 0.909 | 0.024 | 0.434 | 0.084 | -1.040 | 0.908 | -0.012 | 0.105 |
| 36113 | 0.089 | -1.010 | 0.903 | 0.010 | 0.326 | 0.091 | -0.997 | 0.901 | 0.015 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 39628 | 0.086 | -1.025 | 0.905 | -0.041 | 0.319 | 0.083 | -1.046 | 0.908 | 0.002 | 0.432 | 0.092 | -0.993 | 0.904 | -0.022 | 0.104 |
| 57318 | 0.090 | -1.006 | 0.901 | 0.013 | 0.326 | 0.092 | -0.995 | 0.900 | -0.001 | 0.431 | 0.091 | -1.002 | 0.905 | 0.018 | 0.109 |
| 58232 | 0.089 | -1.012 | 0.904 | 0.013 | 0.326 | 0.090 | -1.005 | 0.903 | 0.002 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| 58352 | 0.089 | -1.008 | 0.902 | 0.008 | 0.325 | 0.086 | -1.029 | 0.907 | -0.003 | 0.431 | N/A | N/A | N/A | N/A | N/A |
| 58413 | 0.114 | -0.891 | 0.886 | -0.074 | 0.315 | 0.080 | -1.061 | 0.909 | -0.022 | 0.429 | 0.081 | -1.054 | 0.909 | -0.021 | 0.104 |
| 59179 | 0.094 | -0.982 | 0.894 | -0.052 | 0.318 | 0.085 | -1.031 | 0.907 | 0.021 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 59508 | N/A | N/A | N/A | N/A | N/A | 0.096 | -0.973 | 0.897 | 0.026 | 0.434 | 0.095 | -0.981 | 0.903 | -0.029 | 0.104 |
| 60276 | N/A | N/A | N/A | N/A | N/A | 0.102 | -0.946 | 0.892 | 0.016 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 70445 | N/A | N/A | N/A | N/A | N/A | 0.099 | -0.957 | 0.894 | 0.020 | 0.433 | N/A | N/A | N/A | N/A | N/A |
| 91693 | 0.085 | -1.033 | 0.908 | 0.034 | 0.328 | 0.118 | -0.873 | 0.882 | -0.109 | 0.420 | N/A | N/A | N/A | N/A | N/A |
| 97286 | 0.085 | -1.030 | 0.907 | 0.010 | 0.325 | 0.102 | -0.944 | 0.891 | -0.071 | 0.424 | 0.094 | -0.986 | 0.904 | -0.041 | 0.103 |
| 97831 | 0.080 | -1.059 | 0.910 | 0.032 | 0.328 | 0.096 | -0.974 | 0.898 | 0.006 | 0.432 | 0.087 | -1.020 | 0.907 | -0.025 | 0.104 |
| GH1 | 0.084 | -1.035 | 0.908 | 0.011 | 0.325 | 0.078 | -1.071 | 0.911 | 0.003 | 0.432 | N/A | N/A | N/A | N/A | N/A |
| GH2_165 | 0.084 | -1.040 | 0.909 | 0.025 | 0.327 | 0.086 | -1.029 | 0.907 | 0.016 | 0.433 | 0.087 | -1.019 | 0.906 | -0.022 | 0.104 |
| M9_403 | 0.086 | -1.026 | 0.906 | -0.036 | 0.319 | 0.079 | -1.068 | 0.910 | 0.008 | 0.432 | 0.103 | -0.938 | 0.897 | -0.049 | 0.103 |
| M1046_2 | 0.086 | -1.026 | 0.906 | 0.014 | 0.326 | 0.089 | -1.010 | 0.903 | -0.002 | 0.431 | N/A | N/A | N/A | N/A | N/A |
| MH30_Dreyer | 0.106 | -0.928 | 0.891 | -0.069 | 0.316 | 0.080 | -1.063 | 0.910 | 0.019 | 0.433 | 0.095 | -0.978 | 0.901 | 0.024 | 0.110 |
| Myostatin | 0.086 | -1.029 | 0.907 | 0.011 | 0.326 | 0.084 | -1.038 | 0.908 | 0.023 | 0.434 | N/A | N/A | N/A | N/A | N/A |
| Prolactin_1 | 0.086 | -1.029 | 0.907 | 0.039 | 0.329 | 0.082 | -1.049 | 0.908 | 0.022 | 0.434 | N/A | N/A | N/A | N/A | N/A |
| SBF1 | 0.079 | -1.067 | 0.911 | 0.017 | 0.326 | 0.092 | -0.996 | 0.901 | 0.017 | 0.433 | 0.079 | -1.064 | 0.910 | 0.010 | 0.106 |
| TBC1 | 0.085 | -1.031 | 0.908 | 0.027 | 0.328 | 0.085 | -1.034 | 0.907 | 0.026 | 0.434 | N/A | N/A | N/A | N/A | N/A |

| Locus | Oropuche | | | | | Turure | | | | | Yara | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PP | Log10 (PO) | Q value | Alpha | Fst | PP | Log10 (PO) | Q value | Alpha | Fst | PP | Log10 (PO) | Q value | Alpha | Fst |
| 10802 | 0.083 | -1.042 | 0.909 | -0.031 | 0.073 | 0.086 | -1.029 | 0.908 | -0.018 | 0.045 | 0.071 | -1.115 | 0.913 | -0.004 | 0.231 |
| 108025 | 0.099 | -0.961 | 0.895 | -0.036 | 0.072 | 0.086 | -1.026 | 0.907 | -0.032 | 0.044 | N/A | N/A | N/A | N/A | N/A |
| 108125 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 108291 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 111347 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 120249 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 124729 | N/A | N/A | N/A | N/A | N/A | 0.095 | -0.978 | 0.902 | -0.030 | 0.044 | 0.094 | -0.985 | 0.903 | -0.041 | 0.227 |
| 148158 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 150841 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 20521 | 0.089 | -1.010 | 0.905 | -0.033 | 0.072 | 0.095 | -0.981 | 0.903 | -0.025 | 0.045 | 0.095 | -0.981 | 0.902 | 0.054 | 0.242 |
| 207392 | 0.085 | -1.032 | 0.908 | -0.027 | 0.073 | 0.089 | -1.012 | 0.906 | -0.029 | 0.044 | 0.094 | -0.986 | 0.903 | 0.027 | 0.238 |
| 214079 | N/A | N/A | N/A | N/A | N/A | 0.093 | -0.988 | 0.904 | -0.027 | 0.044 | 0.089 | -1.009 | 0.906 | 0.041 | 0.239 |
| 21765 | N/A | N/A | N/A | N/A | N/A | 0.095 | -0.981 | 0.903 | -0.030 | 0.044 | 0.086 | -1.024 | 0.907 | -0.016 | 0.230 |
| 220371 | 0.095 | -0.980 | 0.902 | -0.031 | 0.073 | 0.082 | -1.047 | 0.909 | -0.007 | 0.045 | 0.071 | -1.117 | 0.914 | -0.017 | 0.229 |
| 22486 | 0.085 | -1.034 | 0.908 | -0.022 | 0.073 | 0.091 | -0.999 | 0.905 | -0.019 | 0.045 | 0.092 | -0.995 | 0.905 | 0.003 | 0.233 |
| 22946 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 23539 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.096 | -0.975 | 0.901 | 0.059 | 0.243 |
| 261690 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 283548 | 0.079 | -1.068 | 0.910 | 0.006 | 0.075 | 0.094 | -0.984 | 0.904 | -0.025 | 0.045 | N/A | N/A | N/A | N/A | N/A |
| 291080 | 0.096 | -0.974 | 0.900 | -0.034 | 0.072 | 0.085 | -1.034 | 0.909 | -0.020 | 0.045 | 0.093 | -0.991 | 0.904 | -0.039 | 0.226 |
| 294904 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 313767 | 0.085 | -1.031 | 0.908 | -0.026 | 0.073 | 0.088 | -1.014 | 0.906 | -0.027 | 0.044 | 0.078 | -1.070 | 0.911 | -0.002 | 0.232 |
| 330994 | 0.091 | -0.999 | 0.904 | 0.035 | 0.078 | 0.086 | -1.029 | 0.908 | -0.032 | 0.044 | 0.086 | -1.026 | 0.908 | 0.039 | 0.239 |
| 334713 | N/A | N/A | N/A | N/A | N/A | 0.090 | -1.007 | 0.905 | 0.003 | 0.046 | N/A | N/A | N/A | N/A | N/A |
| 343160 | N/A | N/A | N/A | N/A | N/A | 0.096 | -0.974 | 0.899 | -0.015 | 0.045 | N/A | N/A | N/A | N/A | N/A |
| 353804 | 0.097 | -0.970 | 0.898 | -0.034 | 0.073 | 0.084 | -1.040 | 0.909 | -0.004 | 0.045 | 0.079 | -1.067 | 0.910 | 0.038 | 0.238 |
| 36113 | N/A | N/A | N/A | N/A | N/A | 0.095 | -0.977 | 0.901 | 0.000 | 0.046 | N/A | N/A | N/A | N/A | N/A |
| 363926 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 39628 | 0.091 | -1.002 | 0.904 | 0.037 | 0.078 | 0.086 | -1.029 | 0.908 | -0.030 | 0.044 | 0.100 | -0.955 | 0.899 | -0.050 | 0.225 |
| 57318 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.084 | -1.035 | 0.908 | 0.006 | 0.234 |
| 58232 | N/A | N/A | N/A | N/A | N/A | 0.099 | -0.959 | 0.897 | 0.011 | 0.047 | 0.092 | -0.995 | 0.905 | -0.004 | 0.232 |
| 58352 | N/A | N/A | N/A | N/A | N/A | 0.095 | -0.979 | 0.902 | -0.019 | 0.045 | 0.089 | -1.010 | 0.906 | -0.044 | 0.226 |
| 58413 | 0.080 | -1.059 | 0.910 | -0.013 | 0.074 | 0.088 | -1.014 | 0.906 | -0.021 | 0.045 | 0.084 | -1.040 | 0.909 | 0.016 | 0.235 |
| 59179 | 0.088 | -1.017 | 0.906 | -0.032 | 0.072 | 0.086 | -1.029 | 0.908 | 0.027 | 0.047 | 0.091 | -0.998 | 0.905 | -0.043 | 0.226 |
| 59448 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 59508 | N/A | N/A | N/A | N/A | N/A | 0.089 | -1.009 | 0.906 | -0.002 | 0.046 | 0.094 | -0.983 | 0.902 | 0.017 | 0.236 |
| 60276 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 70445 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 91693 | 0.095 | -0.977 | 0.901 | 0.030 | 0.077 | 0.080 | -1.059 | 0.911 | -0.016 | 0.045 | 0.091 | -0.998 | 0.905 | -0.054 | 0.225 |
| 97286 | 0.092 | -0.996 | 0.903 | 0.035 | 0.078 | 0.078 | -1.070 | 0.911 | 0.010 | 0.046 | 0.072 | -1.110 | 0.913 | -0.010 | 0.230 |
| 97831 | 0.081 | -1.056 | 0.910 | -0.017 | 0.073 | 0.093 | -0.991 | 0.905 | -0.026 | 0.044 | 0.087 | -1.019 | 0.907 | 0.035 | 0.239 |
| GH1 | 0.086 | -1.024 | 0.907 | -0.028 | 0.073 | 0.107 | -0.922 | 0.893 | 0.057 | 0.050 | 0.092 | -0.996 | 0.905 | -0.040 | 0.226 |
| GH2_165 | 0.110 | -0.906 | 0.890 | 0.087 | 0.083 | 0.083 | -1.046 | 0.909 | 0.005 | 0.046 | 0.099 | -0.961 | 0.900 | -0.057 | 0.224 |
| M9_403 | 0.094 | -0.985 | 0.902 | -0.032 | 0.073 | 0.081 | -1.055 | 0.910 | -0.011 | 0.045 | N/A | N/A | N/A | N/A | N/A |
| M1046_2 | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 0.079 | -1.069 | 0.910 | 0.029 | 0.237 |
| MH30_Dreyer | 0.089 | -1.010 | 0.905 | -0.041 | 0.072 | 0.093 | -0.988 | 0.904 | -0.024 | 0.044 | 0.075 | -1.094 | 0.912 | 0.036 | 0.237 |
| Myostatin | 0.093 | -0.989 | 0.903 | -0.032 | 0.073 | 0.081 | -1.057 | 0.911 | -0.016 | 0.045 | 0.085 | -1.030 | 0.908 | 0.038 | 0.239 |
| Prolactin_1 | 0.097 | -0.971 | 0.899 | 0.035 | 0.078 | 0.082 | -1.051 | 0.910 | -0.024 | 0.044 | 0.079 | -1.064 | 0.910 | -0.035 | 0.227 |
| SBF1 | 0.086 | -1.024 | 0.907 | -0.020 | 0.073 | 0.085 | -1.031 | 0.908 | 0.011 | 0.046 | 0.080 | -1.063 | 0.909 | -0.032 | 0.227 |
| TBC1 | 0.083 | -1.042 | 0.909 | -0.020 | 0.073 | 0.082 | -1.050 | 0.910 | 0.021 | 0.046 | 0.074 | -1.097 | 0.912 | 0.007 | 0.233 |

## Appendix XI: AMOVA results for each SNP from the analysis of the wild fish.

| SNP | Between rivers | | | | Between sites within rivers | | | | Within sites | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Varience component | % variation | FSC | P-value | Varience component | % variation | FST | P-value | Varience component | % variation | FCT | P-value |
| 10802 | 0.01 | 9.61 | 0.16 | 0.00 | 0.02 | 14.71 | 0.24 | 0.00 | 0.09 | 75.67 | 0.10 | 0.06 |
| 20521 | 0.02 | 15.95 | 0.20 | 0.00 | 0.02 | 16.81 | 0.33 | 0.00 | 0.09 | 67.23 | 0.16 | 0.05 |
| 21765 | 0.02 | 10.14 | 0.35 | 0.00 | 0.06 | 31.47 | 0.42 | 0.00 | 0.11 | 58.39 | 0.10 | 0.19 |
| 22486 | 0.08 | 38.30 | 0.33 | 0.00 | 0.04 | 20.36 | 0.59 | 0.00 | 0.08 | 41.34 | 0.38 | 0.04 |
| 22946 | 0.00 | 0.64 | -0.01 | 0.47 | 0.00 | -0.59 | 0.00 | 0.40 | 0.01 | 99.95 | 0.01 | 0.28 |
| 23539 | 0.08 | 30.85 | 0.22 | 0.00 | 0.04 | 15.23 | 0.46 | 0.00 | 0.14 | 53.91 | 0.31 | 0.01 |
| 36113 | 0.00 | 0.18 | 0.00 | 0.37 | 0.00 | 0.03 | 0.00 | 0.30 | 0.01 | 99.79 | 0.00 | 0.60 |
| 39628 | 0.02 | 12.08 | 0.15 | 0.00 | 0.03 | 13.48 | 0.26 | 0.00 | 0.14 | 74.44 | 0.12 | 0.06 |
| 57318 | 0.01 | 12.71 | 0.00 | 0.36 | 0.00 | -0.10 | 0.13 | 0.00 | 0.04 | 87.40 | 0.13 | 0.01 |
| 58232 | 0.00 | 4.11 | 0.03 | 0.02 | 0.00 | 3.13 | 0.07 | 0.00 | 0.04 | 92.76 | 0.04 | 0.19 |
| 58352 | 0.00 | 3.28 | 0.21 | 0.00 | 0.02 | 20.51 | 0.24 | 0.00 | 0.06 | 76.21 | 0.03 | 0.11 |
| 58413 | 0.04 | 16.81 | 0.18 | 0.00 | 0.03 | 14.61 | 0.31 | 0.00 | 0.16 | 68.58 | 0.17 | 0.05 |
| 59179 | 0.02 | 11.22 | 0.16 | 0.00 | 0.02 | 13.92 | 0.25 | 0.00 | 0.13 | 74.86 | 0.11 | 0.08 |
| 59448 | 0.00 | -16.87 | 0.35 | 0.00 | 0.00 | 40.79 | 0.24 | 0.00 | 0.01 | 76.09 | -0.17 | 0.83 |
| 59508 | 0.00 | 2.42 | 0.00 | 0.59 | 0.00 | -0.43 | 0.02 | 0.02 | 0.01 | 98.01 | 0.02 | 0.06 |
| 60276 | 0.00 | 0.12 | -0.01 | 0.63 | 0.00 | -0.84 | -0.01 | 0.91 | 0.01 | 100.72 | 0.00 | 0.88 |
| 70445 | 0.00 | 2.22 | 0.00 | 0.32 | 0.00 | 0.20 | 0.02 | 0.01 | 0.01 | 97.58 | 0.02 | 0.14 |
| 91693 | 0.06 | 24.72 | 0.21 | 0.00 | 0.04 | 16.15 | 0.41 | 0.00 | 0.14 | 59.13 | 0.25 | 0.03 |
| 97286 | 0.01 | 5.49 | 0.26 | 0.00 | 0.06 | 24.42 | 0.30 | 0.00 | 0.17 | 70.09 | 0.05 | 0.17 |
| 97831 | 0.02 | 20.18 | 0.08 | 0.00 | 0.01 | 6.63 | 0.27 | 0.00 | 0.09 | 73.19 | 0.20 | 0.02 |
| 108025 | -0.01 | -2.93 | 0.18 | 0.00 | 0.04 | 18.31 | 0.15 | 0.00 | 0.16 | 84.62 | -0.03 | 0.54 |
| 108125 | 0.00 | -7.10 | 0.14 | 0.01 | 0.00 | 14.76 | 0.08 | 0.00 | 0.00 | 92.34 | -0.07 | 0.87 |
| 108291 | 0.00 | -4.25 | 0.08 | 0.01 | 0.00 | 8.74 | 0.04 | 0.01 | 0.00 | 95.51 | -0.04 | 0.90 |
| 111347 | 0.00 | 0.43 | -0.01 | 1.00 | 0.00 | -1.29 | -0.01 | 0.97 | 0.00 | 100.86 | 0.00 | 0.56 |
| 120249 | 0.00 | 1.90 | 0.00 | 0.73 | 0.00 | -0.36 | 0.02 | 0.24 | 0.00 | 98.46 | 0.02 | 0.17 |
| 124729 | -0.01 | -5.13 | 0.35 | 0.00 | 0.05 | 37.07 | 0.32 | 0.00 | 0.09 | 68.06 | -0.05 | 0.58 |
| 148158 | 0.00 | 2.88 | 0.05 | 0.01 | 0.00 | 4.54 | 0.07 | 0.00 | 0.02 | 92.59 | 0.03 | 0.39 |
| 150841 | 0.00 | 0.76 | 0.00 | 0.48 | 0.00 | -0.28 | 0.00 | 0.42 | 0.01 | 99.53 | 0.01 | 0.41 |
| 207392 | 0.03 | 18.21 | 0.14 | 0.00 | 0.02 | 11.23 | 0.29 | 0.00 | 0.12 | 70.56 | 0.18 | 0.02 |
| 214079 | 0.14 | 57.53 | 0.11 | 0.00 | 0.01 | 4.78 | 0.62 | 0.00 | 0.09 | 37.69 | 0.58 | 0.00 |
| 220371 | 0.04 | 33.48 | 0.16 | 0.00 | 0.01 | 10.53 | 0.44 | 0.00 | 0.07 | 56.00 | 0.33 | 0.01 |
| 261690 | 0.00 | 0.76 | 0.03 | 0.01 | 0.00 | 3.41 | 0.04 | 0.00 | 0.03 | 95.83 | 0.01 | 0.54 |
| 283548 | 0.00 | 2.56 | 0.11 | 0.00 | 0.01 | 10.57 | 0.13 | 0.00 | 0.07 | 86.87 | 0.03 | 0.19 |
| 291080 | 0.07 | 29.56 | 0.31 | 0.00 | 0.05 | 22.04 | 0.52 | 0.00 | 0.11 | 48.39 | 0.30 | 0.04 |
| 294904 | 0.00 | 0.01 | 0.09 | 0.01 | 0.00 | 8.86 | 0.09 | 0.00 | 0.01 | 91.12 | 0.00 | 0.33 |
| 313767 | -0.02 | -7.41 | 0.30 | 0.00 | 0.08 | 32.74 | 0.25 | 0.00 | 0.19 | 74.67 | -0.07 | 0.67 |
| 330994 | 0.13 | 48.05 | 0.08 | 0.00 | 0.01 | 4.41 | 0.52 | 0.00 | 0.13 | 47.54 | 0.48 | 0.00 |
| 334713 | 0.00 | 1.76 | 0.07 | 0.00 | 0.00 | 6.50 | 0.08 | 0.00 | 0.03 | 91.74 | 0.02 | 0.51 |
| 343160 | 0.00 | 2.17 | 0.03 | 0.03 | 0.00 | 2.66 | 0.05 | 0.00 | 0.02 | 95.17 | 0.02 | 0.27 |
| 353804 | 0.09 | 34.10 | 0.23 | 0.00 | 0.04 | 15.18 | 0.49 | 0.00 | 0.13 | 50.73 | 0.34 | 0.02 |
| 363926 | 0.00 | 1.48 | -0.01 | 0.60 | 0.00 | -0.80 | 0.01 | 0.18 | 0.00 | 99.33 | 0.01 | 0.17 |
| GH1 | 0.11 | 46.77 | 0.16 | 0.00 | 0.02 | 8.52 | 0.55 | 0.00 | 0.10 | 44.71 | 0.47 | 0.01 |
| GH2_165 | 0.13 | 50.14 | 0.21 | 0.00 | 0.03 | 10.64 | 0.61 | 0.00 | 0.10 | 39.22 | 0.50 | 0.00 |
| M009_403 | 0.06 | 24.11 | 0.16 | 0.00 | 0.03 | 12.48 | 0.37 | 0.00 | 0.16 | 63.42 | 0.24 | 0.02 |
| M1046_2 | 0.13 | 56.25 | 0.13 | 0.00 | 0.01 | 5.53 | 0.62 | 0.00 | 0.09 | 38.23 | 0.56 | 0.00 |
| MH30_Dreyer | 0.06 | 23.42 | 0.21 | 0.00 | 0.04 | 16.40 | 0.40 | 0.00 | 0.15 | 60.18 | 0.23 | 0.03 |
| Myostatin | 0.13 | 53.52 | 0.11 | 0.00 | 0.01 | 5.17 | 0.59 | 0.00 | 0.10 | 41.31 | 0.54 | 0.00 |
| Prolactin_1 | 0.12 | 46.01 | 0.23 | 0.00 | 0.03 | 12.17 | 0.58 | 0.00 | 0.11 | 41.82 | 0.46 | 0.00 |
| SBF1 | 0.12 | 51.27 | 0.15 | 0.00 | 0.02 | 7.09 | 0.58 | 0.00 | 0.10 | 41.64 | 0.51 | 0.00 |
| TBC1 | 0.03 | 13.89 | 0.22 | 0.00 | 0.04 | 19.04 | 0.33 | 0.00 | 0.15 | 67.08 | 0.14 | 0.06 |

**Appendix XII: Percentage variation from the first two principal components for PCA of upstream and downstream sites within a river.**

| River | All SNPs (A) | | Neutral SNPs (B ) | | Selected SNPs (C ) | |
|---|---|---|---|---|---|---|
| | PC 1 | PC 2 | PC 1 | PC 2 | PC 1 | PC 2 |
| Airpo 03 | 34.2 | 9.2 | 42.3 | 13.0 | 35.0 | 10.0 |
| Aripo 06 | 21.4 | 12.1 | 24.1 | 18.3 | 22.3 | 13.8 |
| Caura | 23.8 | 7.8 | 23.7 | 13.8 | 24.1 | 6.1 |
| Guanapo | 26.2 | 13.9 | 36.6 | 13.5 | 25.2 | 17.4 |
| Lopinot | 25.4 | 10.3 | 29.8 | 16.0 | 28.6 | 10.2 |
| Marianne | 29.7 | 20.9 | 30.6 | 25.6 | 28.1 | 24.0 |
| Oropuche | 21.6 | 17.7 | 31.5 | 24.8 | 24.2 | 16.5 |
| Turure | 12.6 | 12.0 | 16.9 | 16.0 | 16.7 | 16.0 |
| Yara | 36.7 | 10.9 | 47.2 | 20.4 | 37.1 | 16.7 |