Fisheries-induced evolution in morphology:

A selection experiment on the Trinidadian guppy (Poecilia reticulata).

Eihab Ahmed Idris

Thesis in partial fulfillment of the degree

Master of Science in Biology

Department of Biology

University of Bergen (Norway)

June 2016

To Mahdia

Acknowledgements

Abstract

For the past century, the human population has been growing, and the human technology has been advancing, both at historically-unprecedented rates. As a result, the overall impact of human activities on the lives and deaths of the other species has increased to the extent that, in many situations, human-induced mortality has overwhelmed natural mortality. One such situation is commercial fisheries, some of which have been subjected to heavy exploitation for several decades. Because fish harvesting is inherently selective for certain traits (e.g. size and behavior), and because these traits have genetic basis, the evolutionary theory predicts that fishing is capable of driving evolutionary changes in the exploited stocks, and that these changes might take place within contemporary time scales (i.e. decades). Recently, evidence from multiple lines of inquiry (i.e. theoretical modelling, field observations, and selection experiments) are converging on the conclusion that fisheries-induced evolution in the life-history traits (e.g. age and size at maturation, and growth rate) is rapid and ubiquitous, therefore, it has potentially serious consequences on fisheries yield as well as on the viability of the exploited populations. Less is known about the indirect evolutionary effects of fishing on traits other than those directly under selection by fishing gear; these effects might occur either due to genetic/developmental correlations with the directly-selected traits, or as adaptive responses to the demographic/environmental consequences of fishing activities. The present study was conducted with the aim of exploring the potential for indirect fisheries-induced evolution in two aspects of fish morphology, namely coloration and shape; these aspects were chosen because they influence mate choice and swimming performance, respectively.

The present study is based on a long-term selection experiment on the effects of size-selective fishing in the Trinidadian guppy (*Poecilia reticulata*). The selection experiment consisted of two sub-experiments: the harvest and the assay experiments. In the harvest experiment, nine populations of guppies were maintained for multiple generations in the laboratory, and each was assigned to one of three harvesting regimes (three populations were assigned for each of the three treatments): for the positive harvest, large individuals were preferentially removed; for the negative harvest, small individuals were preferentially removed; for the random harvest, the removal of individuals was non-preferential. Harvesting was repeated every six weeks. The assay experiment was conducted once each year, during which, four pregnant females were randomly chosen from each population and were reared under common garden conditions, as well as their first- and second-generation descendants. The selection experiment lasted for three years, approximately equivalent to four guppy generations. The investigation of morphology was based on photographs for all the males from two harvests (the harvests 2 and 4) and for the second-generation males from three assays (the assays 1, 2 and 4). Investigation of the color changes was performed by measuring the numbers and the relative areas of orange and black spots. Investigation of the shape changes was performed by using geometric morphometrics for analyzing landmark coordinates.

The investigation of the morphology of male guppies revealed a number of changes through time as well as differences between the treatments; some of the differences observed in the harvest experiment did not persist in the assay experiment, indicating that they represent phenotypically-plastic responses to the rearing conditions. Other differences were probably genetic in origin, as they occurred in both experiments. Regarding guppy's color, the black coloration decreased during the early stage of the experiment, then increased during the late stage. Moreover, and only for the positive line, the orange coloration increased during the early stage of the experiment, then decreased during the late stage. The early increase of the orange coloration in the positive line was explained as an adaptive response to selection against large size; because the orange color is an expensive yet sexually attractive trait in guppies, evolving more orange reflects increasing investment by the males on reproductive effort, consistent with the theoretical predictions regarding the effect of selection against large size on the life history. The late decrease of the orange coloration in the positive line was explained by the relaxation of selection for increased reproductive effort following the reduction in population density and, consequently, the greater availability of resources per capita. Regarding guppy's shape, the most robust differences were observed between the generations, rather than between the treatments. Specifically, late-generation guppies showed a narrowing in the caudal peduncle, in the dorsal and the anal fins, and also in the dorsal side of the body. The observed changes, especially the narrowing of the caudal peduncle, were explained as adaptive responses to improve steady swimming performance (related to competitive ability) at the expense of unsteady swimming performance (related to predator-escape ability), driven by the release from natural predation. The differences between the treatments, though mostly statistically significant, were neither consistent in direction nor robust across experiments. Interestingly, the treatments differed more clearly during the middle of the experiment than at the end; this pattern was explained as a consequence of the unique demographic histories (i.e. population dynamics) of the treatment lines, that led to variation in the intensity of selection for competitive ability and, consequently, to variable rates of shape change. Overall, the results of the study indicate that indirect fisheries-induced evolution in morphology is possible and could be taking place in the wild; the results also highlight the possibility that the fluctuating selection induced by the demographic consequences of fishing could have important ecological effects on the harvested populations and their biological communities.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1   Human-induced evolution

Due to the unprecedented increase in human population size and technological power, the ecological impact of humans now exceeds that of any other species on Earth. Inevitably, an ecological impact of such scale will cause substantial evolutionary consequences. Human activities affect the survival and reproduction of animals, plants and microbes around them; these anthropogenic pressures are often strong in magnitude, consistent through time and non-random with respect to the traits that renders organisms most susceptible to them. Consequently, the evolutionary theory predicts that, if some of these targeted traits show heritable variation, susceptible organisms will experience adaptive genetic changes driven by the anthropogenic pressures, that is, they will undergo human-induced evolution. Supporting this prediction, there is a large and growing number of documented cases from diverse taxa for organisms evolving in response to anthropogenic pressures (Palumbi, 2001). Moreover, there are evidence that the rate of phenotypic change in human-altered environments is substantially higher than that in unaltered environments (Hendry *et al*., 2008; Darimont *et al*., 2009; Devine *et al*., 2012) although some of these changes are due to phenotypic plasticity, it is safe to conclude that human-driven evolutionary change is widespread, and that it underlies some of the most dramatic episodes of contemporary evolution (e.g. Jachmann *et al*., 1995; Baquero and Blazquez, 1997; Majerus, 1998).

It is important to note that 'human-induced evolution' is a broad term, and that the mechanisms by which humans induce evolution are quite diverse. We must first distinguish between the 'indirect' and the 'direct' modes of human-induced evolution; by interfering with climatic patterns, biogeographical barriers and nutrient cycles, and by degrading and fragmenting natural habitats, humans impose novel selective pressures on many organisms, even though the humans themselves are not the selective agents in these cases. These are examples of indirect human-induced evolution, in which humans do not select for certain phenotypes, but, rather, alter the environment in a way that selects for certain phenotypes (Stockwell *et al*., 2003). The outcomes of indirect human-induced evolution are unpredictable from the human value perspective, that is, there is no general reason to expect these outcomes to be 'desirable' or 'undesirable' in a systematic manner, as that will depend on the details of each individual case.

In other cases, however, the differential reproduction of individuals of the targeted species is directly caused by human activities without the mediation of environmental change. This is the case when humans systematically favor the breeding of certain individuals, or when they systematically remove certain individuals from the breeding population. The first scenario represents what Darwin called 'artificial selection', in which the breeders of domesticated species intentionally choose individuals with the desirable phenotypes for reproduction, with the aim of increasing the frequency of these particular phenotypes in the next generation (Darwin, 1958). The second scenario occurs when the individuals of the targeted species are massively killed by

humans, either to harvest them (i.e. through hunting or fishing) or to get rid of them (i.e. through the application of drugs, insecticides or herbicides). In these cases, humans are selecting, unintentionally, for the phenotypic traits that facilitate resistance to the anthropogenic mortality (e.g. antibiotic, insecticide and herbicide resistance) (Palumbi, 2001), or the traits that minimize the impact of anthropogenic mortality on reproductive success (e.g. early maturation, higher reproductive effort) (Heino and Godø, 2002). Note that the outcomes of the direct form of human-induced evolution are often predictable from a human value perspective, as intentional selection is expected to lead to desirable changes (i.e. enhancing the selected breed), while unintentional selection is expected to lead to undesirable changes (i.e. the evolution of resistance, and harvest-induced evolution) (Allendorf and Hart, 2009).

A wide variety of wild animals are regularly exploited by humans, including both vertebrates (e.g. fish, ungulates) and invertebrates (e.g. gastropods, arthropods) (Fenberg and Roy, 2008). The harvesting of wild populations is often intensive, sometimes resulting in mortality rates that exceed natural mortality (e.g. fish: Heino and Godø, 2002, Hutchings and Fraser, 2008). In addition, harvesting is almost always non-random; traits such as age, size, maturity, color, ornaments, spatial location and behavior renders the individuals differentially susceptible to being caught by hunters and fishermen. For many exploited species, harvesting has been going on for prolonged periods of time, sometimes for centuries (Allendorf and Hart, 2009). The intensity, selectivity and temporal consistency of harvesting pressure indicate that harvest-induced selection is capable of generating strong directional selection if the traits it acts upon have genetic basis. Consistent with this, harvest-induced evolution was documented in diverse cases, from the lack of tusks in elephants (Jachmann *et al.*, 1995) to the earlier maturation in many commercially-exploited fish species (Heino *et al.*, 2015). There are important gaps in our knowledge of harvest-induced evolution, especially regarding its relative ubiquity compared with the phenotypically-plastic responses to harvesting, and its relative importance compared with the demographic and environmental effects of harvesting (Fenberg and Roy, 2008). However, there are three reason suggesting that the likely consequences of harvest-induced evolution are potentially serious, therefore, they merit a precautionary consideration in management and conservation plans: first, harvest-induced evolution leads to the paradoxical situation in which humans select against the traits they prefer most, with the intensity of their selection growing with the extent of their preference, consequently, undesirability is inherent in this evolutionary process (Allendorf and Hart, 2009). Secondly, there are indications that selective harvest can act against natural selection (Edeline *et al.*, 2007) as well as sexual selection (Hutchings and Rowe, 2008), indicating that this process does not only dilute the traits we consider desirable, but also the traits that contribute to population growth and viability and, ultimately, to fisheries yield (Conover and Munch, 2002). Thirdly, there are theoretical (Dunlop *et al.*, 2009) and empirical (Conover *et al.*, 2009) grounds for the belief that the reversal of the undesirable changes caused by fisheries-induced evolution might take longer time than the time it took these changes to evolve in the first place; this is because fishing can impose strong selection differentials, while the relaxation of fishing can only impose weak selection in the reverse direction; this effect is sometimes termed 'Darwin's debt' (Pandolfi, 2009).

1.1.1   Fisheries-induced evolution

Fishing pressure has been steadily increasing all over the world during the past century, driven by the increasing demand from the growing human population, and also by the increasing efficiency of fishing technology (Heino and Godø, 2002). As a result, fishing became a major source of mortality for commercially exploited fish species (Jørgensen *et al*., 2007); a study by the FAO (1999) has shown that most marine fish stocks are either exploited to their maximum capacity, overexploited, collapsed or recovered. Exploitation rates as high as 60-80% were reported for some of the stocks of the Atlantic cod (Ratner and Lande, 2001). Fishing, whether recreational or commercial, is highly selective with respect to size. Commercial fishermen prefer big fish, because they give greater yield per effort, and are therefore more profitable. The situation is the same with recreational fishers, as they value a big fish much more than a small fish. This bias is further strengthened by fishing regulations that protect fish below a minimum size without also protecting fish above a maximum size (Heino and Godø, 2002; Allendorf and Hard, 2009). The scientific literature contains reports of size-selective harvesting from more than 76 fish species (Fenberg and Roy, 2008). Because of its intensity and selectivity, fishing exerts a strong directional selection against large size in fish; to the extent of which size is genetically determined, and to the extent of which size is correlated with other genetically-determined traits, size-selective fishing is expected to act as a selective force driving contemporary evolution (Heino and Godø, 2002).

Size is a key aspect of fish biology, as it correlates with growth, maturation and fecundity. These life history traits show considerable heritabilities in fish, and are therefore capable of undergoing adaptive changes responding to selection (Law, 2000; Fenberg and Roy, 2008). By taking time to grow to a large size before maturing, the fish invests on higher reproductive success in the future (because larger fish typically have higher fecundity) at the expense of no reproduction in the present. The theory of life history evolution predicts that selection will push toward an optimal (in terms of lifetime reproductive success) tradeoff between current and future reproduction (i.e. between the allocation of resources to reproduction versus growth and maintenance), and that this optimal tradeoff depends critically on mortality regimes; the value of reproduction at a certain future age decreases if the probability of the fish surviving to that age decreases. Consequently, species subjected to high adult mortality rates are expected to invest more on current reproduction than on future reproduction, resulting in fast life histories, characterized by early maturation, high early-life reproductive effort and slow post-maturation growth. To the contrary, species subjected to low adult mortality rates are expected to evolve slower life histories, characterized by prolonged growth, delayed maturation and low early-life reproductive effort. The primary effect of size-selective fishing is that it disrupts the natural regime of size- and age-specific mortality, by preferentially removing larger, older individuals from the breeding population. For species with slow life histories, such as many of the commercially exploited fish, this preferential removal will decrease the value of future reproduction (and, consequently, of growing to a large size), therefore imposing a selection for early maturation, slow post-maturation growth and high reproductive effort. Note that the same pattern of selective pressures can occur even if fishing mortality is not

selective; a uniformly elevated mortality across all size and age classes will also reduce the probability of survival at the long term, therefore reducing the value of future reproduction and selecting for a fast life history (Law, 2000; Heino and Godø, 2002; Heino and Dieckmann, 2009; Heino *et al.*, 2015). A number of theoretical models confirmed that these general predictions from life history theory apply under the conditions of commercial fisheries (e.g. Dunlop *et al.*, 2009).

Observational studies of phenotypic changes in exploited fish stocks have reported a decreasing age and size at maturity (i.e. earlier maturation) for a number of species, including the North Sea plaice (Rijnsdorp, 1993), the Atlantic cod (Olsen *et al.*, 2004), and some species of the Atlantic (Quinn *et al.*, 2006) and pacific salmon (Ricker, 1981), confirming the predictions from life history theory. The rate of phenotypic change was the highest for heavily exploited stocks, and it declined when exploitation stopped, consistent with the hypothesis that fishing is the selective agent (Devine *et al.*, 2012). However, an inherent limitation of observational studies is their inability to distinguish genetic from environmentally-induced changes (Law *et al.*, 2000). This problem is further complicated by the fact that, in this case, those environmentally-induced changes are expected to show a similar trend toward earlier maturation, thereby constituting confounding factors rather than mere noise. One such factor is phenotypic plasticity; fishing pressure is expected to reduce intra-specific competition by removing a large proportion of the competitors, resulting in a greater availability of food per individual. Food availability can lead to enhanced growth and, consequently, to earlier maturation which occurs independently of any genetic change (Law, 2000, Heino and Godø, 2002; Kuparinen and Merila, 2007; Hutchings and Frasier, 2008; Heino *et al.*, 2015), therefore, disentangling such 'compensatory responses' to fishing (Trippel, 1995) from the evolutionary responses to fishing is necessary in order to interpret the observed trend as evidence for the latter, rather than the former. In addition, water temperature also correlates positively with fish growth, and it showed an increasing trend during the past decades, as a result of global warming. Finally, a direct demographic effect of fishing is that larger, late maturing fish are repeatedly removed from the population, consequently, smaller, early maturing fish will increase in relative frequency even if no genetic change has occurred (Kuparinen and Merila, 2007; Heino *et al.*, 2015).

Disentangling fisheries-induced evolution from the confounding factors, whether plastic, demographic or environmental is an important task for more than the academic reasons (Jørgensen *et al.*, 2007). If a phenotypic change is not accompanied by a genetic change (e.g. a change due to phenotypic plasticity), then no matter how large the change is, it will not accumulate through generations, therefore, it is readily reversible once the environmental conditions were restored to their original state. On the other side, an evolutionary change, even if it was slight between any two successive generations, can accumulate through generations, potentially leading to radical shifts over prolonged periods of time (Kuparinen and Merila, 2007). Reversing an evolutionary change is more difficult than reversing a plastic response, especially if the relevant genetic variation was lost (see Smith *et al.*, 1991). For fisheries-induced evolution, the reversal is rendered even more difficult by the asymmetry between the strong selection for early maturation imposed

by fishing, and the weak selection for late maturation when fishing is stopped (de Roos *et al*., 2006; Conover *et al*., 2009).

The standard approach for disentangling the genetic and plastic components of maturation schedules is the probabilistic maturation reaction norms (PMRNs) (Dieckmann and Heino, 2007). The key idea behind it is to separate maturation changes that reflect differences in growth rates (interpreted as reflecting plastic responses to environmental variation), from maturation changes that are independent of growth (interpreted as reflecting genetic changes in the timing of maturation). The method describes the probabilities of maturation for a range of sizes-at-ages; the probability of maturation is calculated specifically for each size-at-age class (i.e. for each growth rate), on the condition that the individual has survived to that age and grew to that size. As a result, the probabilities of maturation are rendered insensitive to the environmentally-induced variation in survival and growth (Heino and Godø, 2002; Heino and Dieckmann, 2009; Heino *et al*., 2015). Note that the method accounts only for the phenotypic plasticity related to growth, leaving out other potential factors such as body condition and social structure. Note, also, that the method does not account for the possibility of evolution in growth strategies, which is a serious possibility under the scenario of fisheries-induced evolution (Kraak, 2007; Kuparinen and Merila, 2007; Marshal and McAdam 2007). However, the value of this method does not depend on its ability to remove all non-genetic variation, but, rather, on its ability to remove a particularly problematic source of non-genetic variation; a source that can mimic the effects of fisheries-induced evolution, that is, growth- related phenotypic plasticity (Dieckmann and Heino, 2007; Heino and Dieckmann, 2009). Since their introduction, probabilistic maturation reaction norms (PMRNs) became widely applied, and confirmed the occurrence of fisheries-induced evolution in a number of exploited fish species that showed a trend toward earlier maturation (see Heino *et al*., 2015), most notably the Atlantic cod (e.g. Barot *et al*., 2004; Olsen *et al*., 2004; Pardoe *et al*., 2009).

Size-selective fishing can also lead to the evolution of lower growth rates; if the fish become vulnerable to harvesting only after reaching a minimum size, then the fish that delay growing to the vulnerable size class will survive longer, and reproduce more, given that the reproductive benefit of large size is sufficiently delayed and the mortality risk is sufficiently high (Heino and Godø, 2002). Moreover, fishing can select directly against fast growth (i.e. in a size-independent way); fast-growing individuals tend to forage more actively than slow-growing individuals, consequently, they are more susceptible to capture by fishing gears, regardless to their size (Biro and Post, 2008). Although many exploited fish species show a trend toward decreasing size (Fenberg and Roy, 2008), explaining this in terms of the evolution of slower growth rate has been problematic, for two reasons: first, there is extensive phenotypic plasticity in growth, due to fisheries-related factors (e.g. density) as well as fisheries unrelated factors (e.g. water temperature). These factors are expected to act jointly in the opposite direction (i.e. reduced density and increased temperature both enhance growth) thereby diluting the effects of the potential evolutionary change toward slow growth. Secondly, growth rate is strongly correlated with maturation and fecundity, as a result, an apparent trend toward slower growth could be a

consequence of selection for earlier maturation and/or higher reproductive effort (Heino and Godø, 2002; Enberg *et al*., 2012; Heino *et al*., 2015). Nevertheless, fisheries-induced evolution in growth was documented in the laboratory (Conover and Munch, 2002), and also in the field (Swain *et al*., 2007; Pardoe *et al*., 2009).

Because elevated mortality (whether size-selective or size-neutral) decreases the value of future reproduction relative to current reproduction (Heino and Kaitala, 1999), fishing is expected to select for increased reproductive effort at age. Measuring reproductive effort is challenging, however, a number of proxies can be estimated from field data, such as relative gonad weight (gonadosomatic index) (Heino and Godø, 2002). Few studies have been conducted so far (e.g. North Sea plaice: Rijnsdorp *et al*., 2005; van Walraven *et al*., 2010), and they reported a mixture of positive and negative results (Heino *et al*., 2015). A shared difficulty that faces the investigations of fisheries-induced evolution in growth rate and reproductive effort is the absence of a statistical method analogous to probabilistic maturation reaction norms (PMRNs), as a result, demonstrating that a phenotypic trend is due to evolutionary change is even more difficult for growth and fecundity than it is for maturation (Heino and Dieckmann, 2009). Perhaps the only way to overcome this difficulty is by directly comparing the relevant genes of growth and fecundity between stocks that are subjected to variable levels of exploitation; unfortunately, a full molecular genetic-based methodology does not appear to be a practical option for the short term (Heino and Dieckmann, 2009; Heino *et al*., 2015).

Fishing is not only selective for size, but also for behavior. There are two ways by which fish behavior can render them differentially susceptible to fishing mortality: first, for many stocks, exploitation pressures vary systematically over space (i.e. geographic position and depth) and time. Therefore, if there are heritable differences between individuals in their distributional and migrational characteristics, these behavioral traits will evolve so as to maximize fishing avoidance (Heino and Godø, 2002). Studies on the Atlantic salmon (Quinn *et al*., 2006) and the sockeye salmon (Quinn *et al*., 2006) have documented behavioral changes in the run timing that appear to be adaptive responses to fishing. Secondly, depending on their mechanism of action, fishing gears are selective for different personality traits. Active gears (i.e. trawls) capture more of the individuals with delayed predator-avoidance reaction, thereby selecting for increased vigilance. Passive gears (i.e. traps, gill nets) capture more of the bold, explorative and active individuals, thereby selecting for increased shyness and decreased activity (Heino and Godø, 2002; Biro and Post, 2008; Diaz Pauli *et al*., 2015). Importantly, these behavioral changes reduce vulnerability to capture by diverting some of the time that is normally invested on other activities, such as searching for food and mates, potentially rendering the individuals less ecologically efficient and less reproductively successful (Diaz Pauli *et al*., 2015). The behavioral selectivity of fishing gears was demonstrated experimentally (Cooke *et al*., 2007; Biro and Post, 2008, Diaz Pauli *et al*., 2015), but detecting them in the field has only recently became technologically feasible, therefore, sufficiently-long time series data on the frequencies of different behavioral types are currently lacking (Heino *et al*., 2015). However, by considering the above-mentioned evidence on the

selectivity of fishing gears together with the evidence on the heritability of behavioral traits (Philipp *et al*., 2009; Chervet *et al*., 2011; Ariyomo *et al*. 2013), it seems likely that considerable fisheries-induced behavioral evolution is taking place in the wild. In addition to the direct selection by fishing gears, behavioral traits can evolve responding to fisheries-induced selection on a correlated trait; for example, selection against fast growth can affect traits such as consumption rate and willingness to forage (Walsh *et al*., 2006). Selection against vulnerability to angling can affect the correlated traits of aggression and dominance (Cooke *et al*., 2007; Sutter *et al*., 2012), as well as the traits that are, in turn, correlated with them, such as parental care and mating success (Sutter *et al*., 2012). Further evolutionary changes in behavior can result from fisheries-induced sexual selection, triggered by the reduced mate-encounter rates and the biased population sex ratios; both are possible consequences of heavy, selective fishing (Rowe and Hutchings, 2003).

1.1.1.1 Fisheries-induced evolution in morphology

In addition to life history and behavior, fisheries selection can also affect morphology. Morphological traits show greater heritabilities than life history traits (Mousseau and Roff, 1987; Roff, 1997), and their ability to undergo evolutionary change is demonstrated by their response to selective breeding in aquaculture (Heino and Godø, 2002). Fishing gears can select directly for morphological changes; for example, fast-swimming individuals are more likely to escape the approaching volume of a trawl, therefore, they suffer less mortality from active fishing gears (Izyumov *et al*., 2002). In fish, swimming performance is correlated with morphology, especially with the size of the caudal region (Langerhans *et al*., 2004). Consequently, fishing by trawls can increase the frequency of individuals with larger caudal region. A study on the walleye Pollock has shown that the mean number of the vertebrae (a proxy for length) of the caudal region has increased in the wild following a 30 years of intensive fishing. Individuals with more vertebrae (and therefore longer caudal region) were found to be more resistant to trawl simulators in the laboratory (Izyumov *et al*., 2002). However, these results are merely suggestive of an evolutionary change, due to the limited sample size of the study, and the absence of a systematic treatment of the alternative explanations.

Passive gears like gill-nets can also select directly on morphology. In order for the fish to get caught by a gill-net, the ratio between the girth of the fish at the position of entanglement (usually the maximum girth) and the mesh perimeter of the net should be close to one. A fish that is too slim for the meshes can slip through the net, while a fish that is too deep for the meshes will not get entangled in the net (Reis and Pawson, 1999). Given that the mesh size is usually standardized by fisheries managers; gill-nets can exert consistent selection on body depth under sufficient fishing pressure. Selection on depth can be disruptive, that is, favoring slim and deep shapes and acting against the intermediates that fall within the range of the gill-nets (Heino and Godø, 2002). However, given that fish depth and length are positively correlated (Regier, 1969), and that gill-nets target longer (and hence, on overage, deeper) individuals, it is expected that only slim shapes

will be selected for. The potential costs of an evolutionary change toward slimmer shapes involve reduced space for gonads, as well as reduced burst swimming speed (Heino and Dieckmann, 2009). Evidence for the morphological selectivity of gill-nets are substantial (Hamley, 1975; Reis and Pawson, 1999), but the evidence that the selected-for morphological change does occur in the wild is currently lacking (Heino and Godø, 2002). The most likely reason for this lack of evidence is the lack of effort to find them, since the possibility of fisheries-induced morphological evolution has received much less attention than that focused on fisheries-induced life history evolution. Further research is needed in order to investigate whether the theoretical predictions converge with actual changes in the harvested stocks.

Morphological changes in exploited stocks can occur as secondary responses to changes in other traits driven by fishing pressure, including life history and behavioral traits (Heino and Godø, 2002). The possibility of indirect, fisheries-induced selection on morphology has not been investigated so far, but it seems quite likely; the reason is that fishing preferentially removes larger, older, active and fast-growing individuals from the breeding populations; these are key fitness-determinant traits and are correlated with many other traits (Walsh *et al*., 2006, Conover and Baumann, 2009). Selection experiments have shown that size-selective fishing can have a broad range of indirect effects, including declines in egg volume, larval growth rate, food consumption rate and conversion efficiency (Walsh *et al*., 2006). Poor growth in the larval and the adult stages can have negative effects on the morphological traits that reflect body condition, such as muscle volume, weight and coloration. Further changes in morphology can arise from the interaction between fisheries selection and sexual selection (see Hutchings and Rowe, 2008). For example, one possible scenario is that the greater allocation of resources toward reproductive effort (as a response to fisheries induced selection), will result in a greater investment in secondary sexual characteristics, such as coloration. However, a contrary scenario is that the decrease in abundance caused by fishing, and the subsequent decrease in mate-encounter rate, will reduce the reproductive advantage of secondary sexual characteristics; if there are few available mates, the costs of performing mate choice might be too high for the choosing sex, and less-selective mating renders sexual advertisements like color less important. A similar effect can be expected from the fact that larger, older and fast growing individuals typically have the highest quality sexual advertisements (e.g. the brightest colors); when fishing preferentially removes these individuals, the variability in sexually-selected traits can be greatly reduced, resulting in a further reduction in the importance of mate choice (see Rowe and Hutchings, 2003). Clearly, it is difficult to make general predictions regarding the direction and the extent of the indirect effects of fisheries selection on morphology, as these will be influenced by the species-specific details of ecology and behavior (e.g. the mating system). The present study is an attempt to apply the experimental approach in order investigate the possibility of indirect morphological evolution under size-selective fishing.

1.1.1.2 The experimental approach to fisheries-induced evolution

There are four major approaches for the study of fisheries-induced evolution: field observations, selection experiments, mathematical models and molecular-genetic methods (Conover and Baumann, 2009). The challenge that faces all these approaches is two-fold: first, to disentangle evolutionary changes from environmental variation, and second, to identify fishing, in particular, as the cause of the evolutionary change (Dieckmann and Heino, 2007). A conclusive argument for fisheries-induced evolution must eliminate the possibility that the observed phenotypic change is due to non-genetic factors (e.g. phenotypic plasticity), and also the possibility that an environmental factor, other than fishing (e.g. raising water temperatures), was the selective agent responsible for the evolutionary change. No study based on a single methodology can satisfy both criteria, not even in principle. Statistical analysis of field observations does not account for all sources of environmental variation (Marshal and Browman, 2007; Marshal and McAdam, 2007). Selection experiments are conducted on systems far simpler than those of interest to fisheries science (Diaz Pauli and Heino, 2014); extrapolating from the former to the latter involves an inevitable element of uncertainty. The conclusions of the mathematical models depend on the assumption that all the relevant factors are accounted for and their magnitudes estimated correctly (Conover and Baumann, 2009); a highly uncertain assumption given the simplifications that are usually adopted to make the mathematics tractable. Molecular genetic analysis is hampered by the gaps in our knowledge of the genetic mechanisms underlying fish development and behavior (Heino *et al.*, 2015); even when this gap is filled, it will take a long time to accumulate genetic data on sufficient time scales for the study of contemporary evolution (i.e. decades); even if there is a way around this difficulty, genetic changes, by themselves, do not identify the selective agent that caused them. In the light of the inherent limitations of all methodologies, the standard argument for fisheries-induced evolution (e.g. Heino and Godø, 2002; Kuparinen and Merila, 2007; Heino and Dieckmann, 2009; Heino *et al.*, 2015) has been to present the core evidence obtained from careful statistical analysis of long-term field data (e.g. probabilistic maturation reaction norms of age and size at maturity), then supplementing them by evidence from experiments, molecular genetics and mathematical modelling; these are not attempted as independent arguments for fisheries-induced evolution, but, rather, as replies to various objections that can be raised against the core evidence. By combining multiple evidence from different and complementary methodologies, the case for fisheries-induced evolution in particular species (e.g. the Atlantic cod: Olsen *et al.*, 2004; Dunlop *et al.*, 2009; Hemmer-Hansen *et al.*, 2014) becomes compelling, and when the evidence from several commercially-exploited species are combined, the general case for the reality of fisheries-induced evolution becomes overwhelming.

Seen in this light, the true value of selection experiments lies in their ability to supplement observational studies by providing information that are unattainable by these studies. Through control, experiments can completely disentangle genetic changes from confounding environmental factors (e.g. phenotypic plasticity). Through replication, experiments can identify harvesting as the cause of the genetic change. By allowing the investigator to set the selection differential imposed

by harvesting, and measure the resulting evolutionary response, experiments enable the precise estimation of the heritabilities of candidate traits. As a result, selection experiments can demonstrate the evolvability of different traits under fishing pressures and their rates of evolution, in addition, they reveal the underlying genetic covariances that might accelerate or slow the rate of evolution as estimated from selection on a single trait. None of the above tasks can be achieved, unambiguously, from observational studies (Conover and Baumann, 2009; Diaz Pauli and Heino, 2014).

Typically, a selection experiment starts with a homogeneous population which is then randomly divided into two or more groups. The first stage in the experiment (i.e. the treatment stage) involves subjecting the groups to systematically-different treatments, in order to induce divergence between the groups. Depending on the nature of the experimental interference, selection experiments can be classified as 'artificial' or 'natural'. In artificial selection experiments (e.g. Conover and Munch, 2002; Cooke *et al.*, 2007), the investigator acts directly as the selective agent, by determining the number and type of breeders in each generation. This approach enables a precise control over the intensity of selection, at the expense of rendering the outcomes of the experiment less relevant to wild populations, where the interactions between the various selective and ecological forces plays a key role in determining the direction and the rate of the evolutionary change. In natural selection experiments (e.g. Edeline *et al.*, 2007; Drake *et al.*, 1997) the investigator imposes selection indirectly by controlling the environment under which individuals survive and reproduce, rather than by directly interfering with their reproduction. The advantage of this approach is that, while retaining some control over the experiment, it allows 'surprises' to take place because other factors (demographic, ecological and evolutionary) can interact with and alter the anticipated outcome of selection. The second stage in the experiment (i.e. the common garden stage) involves rearing offspring from the different groups under identical environmental conditions. The standard practice is to conduct measurements on the second generation of common garden-reared individuals, in order to ensure that all environmental variability was removed, including variability in the parental environment (i.e. maternal effects). If phenotypic differences among the groups were developed during the treatment stage and persisted during the common-garden stage, that constitutes strong evidence that the differences in question are genetic, and if the experiment is well designed, genetic differences between the groups can only be caused by the differential treatments to the which the groups were subjected (Conover and Baumann, 2009; Diaz Pauli and Heino, 2014).

Selection experiments have contributed valuable insights to the study of fisheries-induced evolution, confirming the assumptions and the findings of observational studies and also pointing to new, unexplored aspects of the phenomenon (Conover and Baumann, 2009; Diaz Pauli and Heino, 2014). Experiments confirmed that fishing pressure can cause phenotypic changes in individual size, maturation schedule (Van Wijk *et al.*, 2013), growth rate (Conover and Munch, 2002), and vulnerability to angling (Cooke *et al.*, 2007). In a number of cases, the phenotypic changes were found to persist under common garden conditions, indicating that they represent

genuine evolution induced by fishing (Conover and Munch, 2002; Philipp *et al*., 2009; Van Wijk *et al*., 2013). Experiments also confirmed that significant evolutionary changes can accumulate in a few generations of selective fishing (e.g. Conover and Munch, 2002), indicating that the rate of fisheries-induced evolution is fast. Size-selective and behavior-selective fishing was shown experimentally to result in lower harvested biomass and lower catch rate, respectively (Conover and Munch, 2002; Philipp *et al*., 2009), confirming the relevance of fisheries-induced evolution to fisheries management. Furthermore, experiments have shown that the rate of reversal of fisheries-induced evolution can be slower than the rate of fisheries-induced evolution (Conover *et al*., 2009), confirming theoretical predictions (e.g. Dunlop *et al*., 2009) and field observations (Swain *et al*., 2007). In addition to supporting observational studies, selection experiments highlighted key aspects of fisheries-induced evolution that did not receive much attention in observational studies (Conover and Baumann, 2009). Experiments showed that selection on a single trait (e.g. size, vulnerability to capture) could have a broad range of effects, anticipated and unanticipated, on other traits. Selection against large size in the Atlantic silverside was found to affect growth, fecundity, egg size, larval traits, swimming performance, food consumption rate, foraging and anti-predator behavior (Walsh *et al*., 2006). Selection against angling vulnerability in the largemouth bass was shown to affect growth rate, metabolic rate, gonadosomatic index, aggression and dominance, parental care, mating success and reproductive fitness (Cooke *et al*., 2007; Redpath *et al*., 2009; Nannini *et al*., 2011; Sutter *et al*., 2012). Selection against vulnerability to gillnets in the rainbow trout has led to a decrease in the frequency of fast-growing genotypes (Biro and Post, 2008). Through its secondary effects on correlated traits, selective fishing could have much stronger (and less reversible) effects on population persistence and productivity (Walsh *et al*., 2006, Hutchings and Fraser, 2008). Another novel insight from selection experiments is that fisheries-induced evolution in growth rates might represent a form of counter-gradient variation (CnGV), that is, a genetic change in the reverse direction to (and therefore masked by) that of phenotypic plasticity; the compensatory effects of fishing might cause faster growth rates, while selection by fishing may be increasing the frequency of slow-growing genotypes at the same type, the net result would be little or no phenotypic change in the growth rate (Conover and Baumann, 2009).

A famous example of selection experiments simulating fisheries-induced evolution is the work on the Atlantic silverside conducted by Conover and his colleagues (Conover *et al*., 2005). The Atlantic silverside is a marine fish, common along the east coast of North America, where it shows a clinal adaptive variation in growth rates (and a number of related traits) between different latitudes, despite extensive gene flow across the species range. This suggests the occurrence of genetic variation in growth rates that is tuned by continuous stabilizing selection, indicating that this trait can evolve in response to anthropogenic fishing mortality. The species has a short generation time of one year, and can easily be maintained in large captive populations, making it convenient as an experimental model. In this experiment (Conover *et al*., 2005), six captive populations were established from a common founding population, sampled from the middle part of the species range. The populations were then subjected to three different treatments (two

replicas per each treatment), each involving the removal of 90% of the individuals from the breeding population; in the large-harvested treatment, the largest 90% of the individuals were harvested, while in the small-harvested treatment, the smallest 90% of the individuals were harvested. A random-harvested treatment was also established, in which 90% of the individuals were harvested regardless to size. Harvesting regimes were repeated in each generation. The experiment was designed to answer three broad questions: first, can growth rates evolve under size-selective fishing? How fast can they evolve? can growth rate evolution affect population productivity? second, what other evolutionary changes can be induced by selection on size? How broad is the network of traits correlated with size? To what extent can these correlations enhance or constrain fisheries-induced evolution? Third, assuming that size-selective fishing causes evolutionary changes, are these changes reversible? How fast can they be reversed once fishing ceases? Regarding the first question, the harvested biomass (i.e. yield) declined rapidly in the large-harvested line, after four generations of selective harvesting; the decline was a result of the evolution of lower growth rate (Conover and Munch, 2002). Regarding the second question, size-selective fishing induced correlated responses in multiple traits, including larval, physiological, reproductive and behavioral traits (Walsh *et al.*, 2006). Regarding the third question, the large harvested line showed slow, but significant reversal toward the pre-harvesting levels of growth and yield (Conover *et al.*, 2009).

The experiment on the Atlantic silverside demonstrates both the strengths and weaknesses of the experimental approach; unfortunately, the two seems to be tightly intertwined. For example, it is the same aspect that makes the species a convenient experimental model (i.e. short, non-overlapping generations) is what makes it a poor representative of commercially exploited fish, as these typically have long, overlapping generations. Likewise, the severe and precise harvesting regimes that enabled a significant evolutionary response to occur within a few generations is, at the same time, the reason why the experimental harvesting regimes do not resemble actual fisheries, as these are neither that heavy nor that selective. The precise control of the investigators over factors such as physical conditions, density and recruitment is the reason why the results are so 'clean' and straightforward to interpret, however, they are also the reason why the results are difficult to extrapolate to uncontrolled, wild populations where the physical environment is changing rapidly (due to global warming and other anthropogenic effects), and where density-dependent feedbacks, intra-specific interactions as well as natural and sexual selection are continuously acting and interacting, altering the outcome of fishing selectivity in complex ways that can be synergistic, antagonistic, or both (Conover and Baumann, 2009; Diaz Pauli and Heino, 2014).

The design of any selection experiment is a tradeoff between multiple considerations; first, there is the logistical considerations; the experiment need to be conducted in a spatial scale (e.g. tanks, ponds) and temporal scale (i.e. years) that is practical from the perspective of individual scientists and funding agencies, since both have only limited resources to invest in any project (years of professional career and money, respectively); An experiment that lasts decades could be

interesting, but it is very impractical. The logistical considerations limit the range of life histories that can be studied (resulting in a taxonomic bias), and also the intensity of selective pressures that can be imposed (resulting in a design bias). Second, there is the design considerations; a well-designed experiment is one whose success (i.e. the occurrence of significant among-group differences in individuals reared under common garden conditions) can only be explained by the differential treatments. To achieve this ease of interpretation, the investigator need to maintain all the factors other than the experimental treatment strictly homogeneous across treatments, a demand that severely limits the complexity that is allowed in a well-designed experiment. An experiment that grows too complex will become as messy and open to multiple interpretations as reality is, and then it will offer no advantage over observational studies. Third, there is the generality considerations; the experiment needs to retain a reasonable similarity with the system it aims to describe, especially for the factors known to be relevant to the phenomenon under study. For fisheries-induced evolution, a similarity in life history between wild and experimental organisms and in density-dependence between wild and experimental environments is crucial if the experimental findings are to be extrapolated to wild populations. An experiment that fails to meet the generality considerations could be neat and remarkable, but it offers no rigorous support to the theory of fisheries-induced evolution in the wild, simply because the situation could be very different in the wild (for a detailed discussion, see Diaz Pauli and Heino, 2014).

To conclude, a perfect selection experiment is inherently impossible because of the tradeoff between practical and design considerations on one side, and generality considerations on the other side. The optimal design for any particular study depends on the relative importance of the different considerations, seen from the perspective of the resources of the investigator and the goals of the investigation. For example, a study aiming only at estimating the heritability of a trait and its potential for evolution under selective fishing will assign less importance to the generality consideration than a study aiming to predict the actual evolutionary response that will occur in commercial fisheries (Conover and Baumann, 2009).

1.2   The Trinidadian guppy (*Poecilia reticulata*: Poeciliidae)

The Trinidadian guppy *Poecilia reticulata* (family: poeciliidae) is a freshwater fish, native to the mountain streams of north-eastern South America and neighboring islands, including Trinidad, and was introduced to many other tropical areas (Reznick and Endler, 1982). Guppies have small body size (16 mm standard length for mature males and 18-35 mm for mature females) and short generation time (2-3 months). Guppies are sexually dimorphic with respect to growth, size and coloration; unlike females, male guppies stop growing after reaching maturity, and are therefore smaller than females. Unlike the colorful males, female guppies have drab, uniform color (i.e. sexual dichromism). Guppies are characterized by internal fertilization, facilitated by the modification of the male's anal fin into a gonopodium. Females can mate multiple times, and are able to store sperm for months. Once matured, they produce broods of live-born offspring every

three to four weeks throughout their lives (Reznick and Ghalambor, 2005; Diaz Pauli, 2012; Gordon *et al.*, 2015).

Natural populations of guppies show systematic variation in life history, morphology and behavior depending on their location. A substantial body of scientific research has focused on understanding the evolutionary origins of this geographic variation, especially within the Northern Range mountains of the island of Trinidad. It is now well established that most of this variation is explained by the corresponding variation in predation regimes. the streams where guppies occur are punctuated by waterfalls that acts as barriers, preventing the upstream migration of predators. Guppies living downstream the waterfalls are subjected to more predation than those living upstream, because dangerous predatory fish are present only downstream. The phenotypic differences between upstream and downstream guppy populations were explained as adaptive responses to local predation regimes. Supporting this explanation, the adaptive divergence between 'high-predation' and 'low-predation' ecotypes has evolved independently in many streams. Furthermore, a series of introduction experiments have demonstrated that the high-predation ecotype can evolve rapidly to the low-predation ecotype when transplanted in a low-predation, upstream location, thereby constituting one of the best-known cases of contemporary evolution (Reznick and Ghalambor, 2005; Gordon *et al.*, 2015; see below for more details).

There are several factors that make the guppy an ideal model species for experimental studies in general, including their small body size, short generation time and continuous reproduction, together with their amenability to experimental investigation, and their high survival rate in the laboratory (Diaz Pauli, 2015). For evolutionary studies in particular, guppies offer the valuable advantage of being one of a few fish species in which there is a substantial knowledge on the selective forces that shaped their life history, morphology and behavior, as well as the tradeoffs between these forces (Reznick and Endler, 1982). It should also be noted that the utility of a model species has an element of self-propagation; an important reason for the usefulness of guppies as model species is that many researchers have found them useful in the past, worked on them, and established standard protocols for measuring various traits associated with their life history (e.g. maturation: Reznick *et al.*, 1990), morphology (e.g. color pattern: Endler, 1980) and behavior (e.g. boldness: Burns, 2008).

1.2.1   The guppy as a model for the study of fisheries-induced evolution

Essentially, the prediction that selective fishing will induce fish to evolve is derived from a general prediction from life history theory, namely, that changes in the rates of age- and size-specific mortality will select for adaptive changes in life history traits. As a result, any source of extrinsic mortality can simulate the effect of fishing, be it nets or traps, parasites or predators; the only condition is that the intensity and the size-selectivity of the mortality source must be comparable to that which occurs in fisheries. This is the reason why the classical work on life history evolution in guppies by Reznick and collaborators represents a compelling argument for fisheries-induced

evolution, even though it was not originally intended to address this issue (Reznick and Ghalambor, 2005).

As noted above, guppies that occur in the Northern Range mountains in Trinidad show adaptive variation in multiple traits as a response to local predation pressures. Guppies from downstream the waterfalls co-occur with dangerous predatory fish that prey specifically on guppies, including the pike cichlid (*Crenicichla alta*), while those from upstream the waterfalls co-occur with weaker predators, including the killifish (*Rivulus hartii*) that occasionally prey on guppies (Reznick and Endler, 1982). Field investigations confirmed that guppies living downstream the waterfalls are subjected to substantially higher mortality rates than those living upstream (Reznick *et al.*, 1996a). Importantly, the pike cichlid preys preferentially on large, mature guppies, while the killifish preys predominately on small, immature guppies (Reznick and Bryga, 1987). Note that the difference in mortality regimes between low- and high-predation environments resembles the difference between unexploited and exploited stocks in two key aspects: first, the overall mortality rate is higher in both high-predation and exploited environments; second, larger size classes suffer disproportionately higher mortality rates in both environments. Therefore, the predictions from life history theory are qualitatively the same for high-predation guppies and heavily-exploited, commercially-important species. The difference however, is that the biology of guppies and their unusual geographical context allow scientists to conduct a range of thorough investigations (including mark-recapture studies, common garden experiments and introduction experiments) that are difficult, and sometimes impossible to perform in commercial fish species, especially for long-lived marine ones (Reznick and Ghalambor, 2005).

Assessment of the life history traits of wild guppies was performed in two ways: first, through direct measurements on guppies caught from the field, and second, through the investigation of common garden-reared, second generation offspring of wild-caught guppies. Regarding field measurements, guppies from high-predation localities were found to be smaller than those from low-predation localities, and they mature at younger ages and smaller sizes. They also had greater reproductive allocation, that is, they allocate more of their body weight to developing offspring (Reznick and Endler, 1982; Reznick *et al.*, 1996b). These observations confirm the predictions of the theory of life history evolution. The above differences were found to persist when investigated more thoroughly under common garden conditions; the second generation descendants of guppies from high-predation localities were smaller and matured earlier at smaller sizes than their counterparts from low-predation localities. Moreover, the reproductive effort of lab-reared, high-predation guppies (i.e. the proportion of available resources devoted to offspring production) was significantly higher than that of lab-reared, low-predation guppies; specifically, they produced more litters, and have more and smaller offspring per litter. Importantly, the same differences in life history between high- and low-predation guppies emerged repeatedly in multiple streams, even where the communities of guppy predators are different (i.e. such as between the north and south slopes of the Northern Range mountains) (Reznick and Bryga, 1996). Again, this agrees with

theory, because the predicted life history changes should depend only on the rate and the selectivity of mortality, not on the source of mortality.

In order to conclusively demonstrate the causal relationship between predation and the observed life history differences, a series of introduction experiments were performed. These experiments were facilitated by the fact that waterfalls differ in their flow rate and, consequently, in their efficiency as barriers for upstream migration; some exclude only dangerous predators while others exclude both guppies and dangerous predators (Endler, 1980). In the first experiment, guppies were translocated from a downstream location to an upstream location in the same stream that contains the killifish (a weak predator) but lacks guppies and dangerous predators. To provide replication, the introduction experiment was performed in two separate streams. The introduced guppies successfully colonized their new locations, and experienced significantly lower mortality rate than that in their original, downstream locations (Reznick *et al.*, 1996a). Years after, guppies were collected from the introduction locations and the original location, and their second generation, common garden-reared offspring were investigated for potential differences in life history traits. As predicted by theory and suggested from observations in non-manipulated streams, the decrease in mortality rate has led to a significant increase in age and size at maturity for males (after 4 years) (Reznick and Bryga, 1987), and for females (after 7.5 years) as well as a significant decrease in early life reproductive effort (after 11 years) (Reznick *et al.*, 1990). In the second experiment, a dangerous predator (the pike cichlid) was translocated from a downstream location to an upstream location that contains only guppies and small predators. The pike cichlids successfully colonized their new location, imposing higher mortality on the resident guppy population. As expected from theory and earlier observations, the increase in mortality rate caused a significant decrease in age at maturity for both males and females after five years (Reznick, 1997). By directly manipulating mortality rates to test whether guppies will evolve in the predicted direction, the investigators were able to show that the presence of dangerous predators is the cause of the differences in life history, not just a factor that is correlated with these differences. They also showed that life history evolution can be fast enough to be observed during the lifetime of one investigator (Reznick and Ghalambor, 2005).

The findings of these studies can reasonably be extended to commercial fisheries, with the introduction of pike cichlids simulating the establishment of fisheries, and the introduction of guppies simulating the cessation of fishing. Guppies have overlapping generations, just like most commercially exploited species. Moreover, the natural setting in which the experiment was conducted allows the ecological processes to interact with and alter the effects of changing the mortality rate. The predator-induced mortality in guppies are comparable to, and often lower than, fisheries-induced mortality in exploited species (Reznick and Ghalambor, 2005; Conover and Baumann, 2009). Based on the above reasons, it could be argued that life history evolution in guppies offer a stronger case for fisheries-induced evolution than the influential experiment by Conover and Munch (2002), where the model species had discrete generations, the ecological factors were neutralized, the selection was artificial and the imposed mortality was severe. The

evolutionary response of age and size at maturity for the introduced guppies was estimated to represent a 5-15% increase after 7-12 generations of selection (Reznick *et al*., 1997). If a similar change per generation occurred in exploited species, a similar evolutionary response will take place in a time scale of several decades. Since fishing mortality often exceeds predation mortality, the actual response of some of the exploited species is expected to be much faster. The introduction experiment also suggests that life history evolution is an inherently symmetrical process, because guppies responded adaptively to both the increase and the decrease in mortality rate, and there was no clear indication that the magnitudes of the opposite responses were different. This result is relevant to the ongoing debate on the reversibility of fisheries-induced evolution (Reznick and Ghalambor, 2005).

Guppies have also served as a model species for selection experiments that aimed specifically to simulate fisheries-induced evolution (Diaz Pauli and Heino, 2014). In these experiments, size selective mortality was imposed directly by the investigator, rather than indirectly by exploiting the size-preferences of natural predators. The advantage of this approach is that it demonstrates that the size-bias of mortality, in particular, is the cause of the phenotypic change. Experiments that rely on predation mortality cannot accomplish this, because the predators can be selective for traits other than size (e.g. guppy's predators exert selection on coloration and behavior as well as size 'Endler, 1980, Magurran *et al*., 1992'), and that additional selectivity can be the actual cause of the phenotypic change, thereby confounding interpretation.

Kasperski and Kozlowski (1993) performed a natural selection experiment in which the treatment group was subjected to size-selective harvesting, imposed by removing one large individual every four weeks (alternating between males and females); given their average population size of 47 individuals, this is equal to a selection intensity of 2%. Selective harvesting lasted for 15 months, then followed by a common garden experiment that lasted for one generation. The investigators reported a reduced size at maturity for the treatment group compared to the control group. However, these differences did not persist under the common garden conditions, leading the investigators to conclude that the observed differences reflect phenotypic plasticity rather than genetic change. The negative result obtained by this study is probably due to the short duration of the study (Diaz Pauli and Heino, 2014); the investigators estimated that the 15 months' period is equivalent to two to three generations, however, taking the overlap between generations into consideration, the actual number of generations is probably less than that.

Van Wijk *et al*. (2013) conducted an artificial selection experiment with two treatments and one control group; for the two treatments, harvesting was selective for small and large body sizes, while harvesting was random with respect to size in the control group. Harvesting consisted of selecting 20% (i.e. the 20% smallest, the 20% largest and 20% randomly chosen) of the males from each group for representation in the next generation, while females were not subjected to size-selective harvesting. Harvesting regimes were repeated for three generations under controlled conditions. Significant phenotypic changes in the predicted directions were reported for male standard length as well as age and size at maturity (i.e. a decrease in male standard length

accompanied by a decrease in age and size at maturity for the small-selected line and the opposite responses in the large-selected line). The unique contribution of this study is that the different lines were screened for differences in allelic frequencies at two types of loci; first, candidate loci, previously identified as related to standard length from quantitative trait loci (QTL) mapping (Tripathi *et al*., 2009), and second, neutral microsatellite loci, used to assess the potential effects of inbreeding. Significant differences between the lines were detected at four candidate loci, but no significant difference between the lines was found at the neutral loci. Overall, these results strongly indicate that size-selective harvesting caused genetic changes (i.e. evolutionary changes) in the selected lines, at least for male length. Note that the findings of this study are not directly applicable to commercially exploited fish (van Wijk *et al*., 2013: Diaz Pauli and Heino, 2014), not only because of the general limitations of artificial selection setups discussed earlier (i.e. non-overlapping generations, severe and precise selection, control of ecological and demographic processes), but also because male guppies have determinate growth, unlike most exploited species, moreover, the four candidate genes were all linked to the Y chromosome, suggesting that they might be related to the determinate growth mode of the males.

1.2.2   Color evolution in guppies

Theories of sexual selection assume that the elaboration of secondary sexual traits (i.e. colors, ornaments and displays) in prey species enhances mating success at the expense of reducing survival rate, by rendering the organisms more susceptible to visually-hunting predators. Sexual selection by mates is expected to favor conspicuousness, while natural selection by predators is expected to favor crypsis. The outcome of these opposing selective forces is a compromise that serve the goals of mate attraction and predator avoidance but does not maximize either goal. A key prediction of this classical view is that, if predation intensity varies spatially or temporally, the conspicuousness of sexual signals should vary accordingly, specifically, at the times and places of low predation rate, sexual selection should predominate, resulting in more conspicuous signals, while at the times and places of high predation rate, natural selection should predominate, resulting in more cryptic signals (Haskins *et al*., 1961; Endler, 1978).

The Trinidadian guppy *Poecilia reticulata* represents an ideal model to test this prediction. Male guppies are extremely polymorphic in nature, to the extent that no two males are alike. The color pattern of male guppies is a mosaic of patches that vary in color, size, shape and brightness; all these aspects are quantifiable, and all affect overall conspicuousness. Guppy colors are classified into multiple categories: carotenoid colors (orange, red and yellow), melanic colors (various shades of black) and iridescent/structural colors (blue, violet, green and silver); the color categories differ in their physiology, visibility to conspecifics and predators as well as in the information they reveal about the bearer's condition (Endler, 1978; 1983). Guppies color patterns are under the control of multiple genes, some of which are linked to the Y chromosome (Brooks and Endler, 2001), in addition to an important plastic component (e.g. Kodric-Brown, 1989; Ruell *et al*., 2013).

The above points indicate that the color pattern of the guppy is capable of undergoing complex responses to selection, because the different elements of the color pattern can evolve independently of each other. As a result, high-precision analysis of the interaction between natural and sexual selection is possible, because detailed quantitative predictions can be formulated and tested.

In contrast with the complexity of guppies' color patterns, the predation regimes to which guppies are subjected are rather simple. Field studies have shown that fish are the most important predators of guppies, moreover, the intensity of fish predation varies consistently across the upstream-downstream axis; guppies that live upstream the waterfalls co-occur with weak predators and experience lower mortality rates than guppies living downstream the waterfalls. Consequently, if the color pattern is, in fact, a compromise between natural selection and sexual selection, then the relaxation of the former should lead to an increase in the relative importance of the latter. The predicted outcome is more conspicuous coloration in low-predation locations. The magnitude of the difference between low- and high-predation locations may vary for the different elements of the color pattern (e.g. number of spots, size of spots), as well as for the different colour categories (i.e. carotenoid, melanic and iridescent) (Endler, 1978; 1983).

The prediction that guppies respond to low predation by becoming more colorful was investigated in a remarkable series of studies conducted by Endler (1978; 1980; 1982; 1983). A detailed investigation of the color patterns in natural populations confirmed that guppies from low-predation locations are more colorful (and therefore more conspicuous) than those from high-predation locations. For structural colors, this is achieved through a reduction in the numbers of spots. For pigment colors (carotenoid and melanic), this is achieved through a reduction in the numbers and sizes of spots. Observations from several streams revealed the same differences between high- and low-predation guppies; importantly, these include streams from both the north and the south slopes of the Northern Range mountains, where the predator communities are substantially different from each other. This indicates that the differences in coloration are correlated with differential predation rather than with local factors related to a specific stream or a specific predator. However, observational studies cannot go beyond correlation to establish that the differential predation is, in fact, the cause of the differential coloration.

In order to rigorously establish the causal connection, Endler (1980) went on to conduct two parallel experiments; a greenhouse experiment and an introduction experiment. In the greenhouse experiment, ten ponds were established as mimics of the natural streams in the Northern Range mountains. The starting population consisted of wild guppies sampled from several locations in multiple streams, in order to ensure maximum genetic variation. The founders were then assigned to different ponds in such a way that each pond started with guppies originating from all the sampling locations. Later, a dangerous predator, the pike cichlid, was added to four ponds, to impose a high predation rate similar to that in downstream locations. A weak predator, the killifish, was added to four ponds, to impose a low predation rate similar to that in upstream locations. Finally, two ponds were treated as control and left without predators. Therefore, the experiment tested simultaneously for the effects of high predation, low predation and no predation. After 14

months from the introduction of predators, a survey was conducted for the color patterns in the experimental population. The results of the greenhouse experiment converged with field observations, firmly confirming the theoretical predictions. As predicted, high-predation ponds had fewer and smaller color spots than low-predation ponds, moreover, no-predation ponds had more and larger spots than predation ponds, whether experiencing low or high predation. As observed in the field, structural colors responded to the changes in predation intensity by changing the numbers of spots, while pigment colours responded primarily by changing the sizes of spots.

To demonstrate that the greenhouse results were not caused by the artificial nature of the greenhouse environment, Endler (1980) performed an introduction experiment, in which guppies were transferred from high-predation, downstream location to an upstream location that contains the weak predator, the killifish, but lacks guppies. Two years later, a survey was conducted for the color patterns in the introduction site, together with two controls: the original site and a nearby low-predation site which is similar to the introduction site in all aspects except that it naturally contains guppies. Remarkably, the results of the introduction experiment replicated the results of both the greenhouse experiment and the field study; the color patterns in the introduction site diverged from the high-predation control (the original site), and converged with the low-predation control; the increase in coloration was significant for all the color categories. Overall, this series of studies represent a strong case (indeed, the closest one can get to a "proof'' in biology) for the hypothesis that predatory fishes select for cryptic coloration in guppies. The evidence for an opposing role of sexual selection, though compelling, are less direct. For example, the more conspicuous coloration of the no-predation ponds compared to the low-predation ponds in the greenhouse experiment indicates that conspicuousness increases through active selection by female choice, rather than as a consequence of the relaxation of selection by predators.

Later research provided further support to Endler's conclusions, by confirming some of his key assumptions; these assumptions were reasonable to make, but they were not specifically tested in his work. Firstly, the assumption that more colorful, conspicuous males suffer higher mortality rate because they are preferentially attacked by predators was confirmed from field studies as well as laboratory investigations; Reznick *et al.* (1996a) showed that adult males in the wild experience higher mortality than do females and juveniles that lack conspicuous coloration, and that the differences persist even when the effects of the size are removed. Godin and McDonough (2003) were able to show, in an experiment, that brightly-colored males are preferentially approached, attacked and captured by a dangerous predatory fish (the blue acara cichlid) compared to similar-sized, drab-colored males, and therefore they incur higher mortality risk. Secondly, the assumption that brightly-colored males are preferentially chosen by females was also confirmed in both the field and the laboratory. Gordon *et al.* (2015) showed, through a detailed, individual-based monitoring of an introduced population of guppies, that males with more orange and black coloration have higher reproductive success. Houde (1987) showed experimentally that females discriminate among the males based on the relative area of orange in the color pattern, preferring the males with larger orange area. Similarly, Kodric-Brown (1989) conducted an experiment in

which males with variable levels of orange coloration were directly observed during courtship and their mating success was measured. He was able to show that males with more orange coloration were preferred by females and had higher mating success. It is likely that females prefer orange-colored males because carotenoid pigments cannot be synthesized by guppies and therefore must be obtained from diet (Rothschild, 1975), consequently, the extent and brightness of orange coloration indicates the male feeding ability and, indirectly, the quality of his genes (Endler, 1980; 1983). Preferring the males with more and brighter orange coloration is adaptive from the perspective of the ''good genes'' hypothesis, since these males are better foragers (Kodric-Brown, 1989; Grether, 2000) and more resistant to parasites (Houde and Torio, 1992), among other traits. The findings of Endler's work were replicated in new introductions (Gordon *et al*., 2015) as well as in later field studies (Winemiller *et al*., 1990).

It is important to keep in mind that the simple scenario of guppy coloration as a dynamic balance between the opposing demands of predator avoidance and female attraction is only a part of the actual story. A number of other factors interact with, and complicate the outcome of selection by predators and females; some of these factors were investigated by Endler (e.g. 1980), and others by later researchers. Firstly (and perhaps most importantly) there is the effects of the visual background (Endler, 1978; 1983); the notions of a 'conspicuous' or a 'cryptic' color pattern acquire a specific meaning only by reference to a certain background against which the color pattern is naturally seen; conspicuous patterns are those that contrast with the background, while cryptic patterns are those that match the background. Predators are selecting for patterns that match the background, while mates are selecting for patterns that contrast with the background. The mountain streams where guppies naturally occur are characterized by clear waters and gravel bottoms; the colors and sizes of the gravel varies within and between streams. Endler (1980) tested the potential effects of the visual background on guppy coloration in his greenhouse experiment; within each predation level (high, low, none), some ponds had fine gravel while others had course gravel. Endler (1980) predicted that guppies living with predators will evolve to match the background (i.e. smaller spots in fine-gravel ponds and larger spots in course-gravel ponds), and that the match will be better for the guppies living with the dangerous predator, the pike cichlid. He also predicted that guppies living without predators will show the reverse trend, evolving a contrast with the background (i.e. larger spots in fine-gravel ponds and smaller spots in course-gravel ponds). The results of the greenhouse experiment confirmed all these predictions.

Secondly, despite the convenience of defining conspicuousness on the basis of human perception, the explanatory value of this notion depends on its being defined in terms of the perception of the actual agents of selection, that is, conspecifics and predators. Since predators and their prey can be distantly related, it is possible for the perception of conspecifics and predators to work differently, and these differences can be exploited by selection. Specifically, if a certain signal (e.g. a range of wavelengths) is conspicuous to conspecifics but cryptic to predators, then selection will favor the modes of intra-specific communication that are mediated by this ''private signal''. The advantage of private signals is that they increase conspicuousness to conspecifics without also

increasing conspicuousness to predators, thereby avoiding the compromise between mate attraction and predator avoidance (Endler, 1978; 1983). A possible example of private signals is provided by the invertebrate predator of guppies, the freshwater prawn. This species is most abundant in locations where fish predation is low (i.e. upstream locations). Importantly, the visual system of prawns is sensitive to short wavelengths such as blue, but is relatively insensitive to long wavelengths such as orange and red (compared to fish vision). It is therefore expected that guppies that live where the prawns are abundant should have less blue and more orange and red that guppies that live where the prawns are rare. This prediction received variable degrees of support from a number of studies (Endler, 1983, Millar *et al*., 2006; Kemp *et al*., 2008).

Thirdly, although female guppies generally prefer more colorful males, female choice (and, consequently, the intensity of selection for conspicuousness) is not a fixed factor, but varies across populations. Endler and Houde (1995) conducted an assessment of female choice from 11 locations in Trinidad; they found that females from different populations varied not only in the traits they prefer, but also in the degree of their choosiness. The mean female preference in each population was found to be similar to the mean value of the male trait in the same populations, as a result, females typically prefer males from their own population. The correlation between the female preference and the male trait is most evident for orange colouration (Houde and Endler, 1990). This behavioural variation may reflect adaptive evolution to local factors, such as predation intensity, that affect the benefits and costs of female choice for different male traits. Because female preference varies across populations, a reduction in predation intensity will cause an increase in male coloration only if the females in the population do prefer more colorful males. The lack of such preference may explain the puzzling results of Karim *et al*. (2007) introduction experiment, in which guppies were transferred from high- to low-predation location but did not show the expected increase in male coloration.

Variation in physical habitat features such as stream width and depth as well as canopy openness was found to correlate with variation in color patterns, suggesting a potential influence of the former on the latter (Millar *et al*., 2006). Finally, there are evidence that conspicuous coloration imposes a fitness cost that is independent of predation risk (Brooks, 2000; Gordon *et al*., 2015). Despite all the complicating factors discussed above, the basic view of guppy colors as a compromise between natural and sexual selection remains unchallenged by the current evidence.

1.2.3   Shape evolution in guppies

The body shape of fishes reflects various aspects of their ecology and behavior, affecting their feeding habits, antipredator responses and sexual displays (Burns *et al*., 2009). The most direct effect of fish shape is on swimming performance, as evidence suggest that shape features are constrained by a tradeoff between sustained swimming and burst swimming (Webb, 1984). Importantly, burst swimming performance enables fast-start escapes, that is, the sudden, unsteady swimming movements that are generated by the muscles of the caudal peduncle and performed to

escape from predators (Walker, 1997; Ghalambor *et al*., 2004). The speed of fast-start escapes can be enhanced by adopting an elongated/streamlined body shape with a deeper caudal peduncle and a smaller head; the deeper caudal peduncle increases thrust, while the smaller head reduces drag (Langerhans *et al*., 2004). It is therefore expected that predation should select for these shape features.

Guppies provide an opportunity to study the adaptive divergence of shapes under spatially-variable selection. As discussed earlier, natural populations of guppies are distributed across a predation gradient, with downstream populations experiencing higher predation mortality than upstream populations (Endler, 1978). As a result, the intensity of selection for fast-start escapes should vary along the upstream-downstream axis, specifically, the fitness value of efficient burst swimming (relative to efficient sustained swimming) should be higher for guppies that frequently encounter dangerous predators, since their reproductive success will depend critically on their ability to survive those encounters. It is possible, therefore, that the high-predation and low-predation ecotypes differ in their shape features as they differ in their life history, color patterns and behavior (Langerhans and DeWitt, 2004).

Langerhans and DeWitt (2004) conducted a morphological investigation of the shapes of wild-collected males from high- and-low predation populations (three populations from each ecotype). Consistent with predictions, they found that high-predation males have more elongate bodies and deeper caudal peduncles than low-predation males. They interpreted these differences as adaptive responses to the variable predation pressure in the two environments. This interpretation is supported by two parallel lines of evidence: first, the observations from the same study that similar shape differences between the high- and low-predation localities were observed for two other poeciliid species, *Gambusia affinis* and *Brachyrhaphis rhabdophora*, that inhabit different geographic regions and co-exist with different predator communities. The shared responses between the three species were found to outweigh the unique responses of each species to the high-predation/low-predation gradient, as they were better predictors of the predation regime of origin. The parallel evolution of the intra-specific shape differences in the three species was taken as strong evidence that these differences were caused by natural selection acting differently along the predation gradients (Langerhans and Dewitt, 2004). Secondly, high-predation guppies were found to escape predators faster and to survive encounters with predators more often than low-predation guppies, suggesting that their characteristic shape features do indeed enhance fast-start escapes (Ghalambor *et al*., 2003). Faster escape speeds were also reported for high-predation individuals of *Gambusia affinis* (Langerhans *et al*., 2004). Langerhans and DeWitt (2004) also found that high-predation individuals have a lower position of the eyes than low-predation individuals. This trait was interpreted as an adaptation for predator detection; because predators typically attack poeciliids from below, a lower eye position enables a better detection of approaching predators.

Hendry *et al*. (2006) studied the shape differences between high-predation and low-predation ecotypes as manifested by wild-collected, male and female guppies. The aim of the study was to investigate whether the sexes have evolved parallel or unique responses to the predation gradient.

Overall, the shared component of the phenotypic divergence was weaker than the unique component, indicating that the sexual differences in morphology and behavior have led to different adaptive responses to predation for males and females. Predation had a stronger effect on the male shape than on the female shape; this may be because the female shape is heavily constrained by the requirements of live-bearing (Ghalambor *et al.*, 2004), or because females suffer less predation than males (Reznick *et al.*, 1996a). High-predation males had smaller heads and deeper caudal peduncles, consistent with predictions and earlier observations, however, their overall body shape was deeper rather than more elongated.

Alexander *et al.* (2006) investigated divergence among guppy populations across the entire range of the species, by comparing the spatial patterns of the variation in neutral genetic markers (nuclear and mitochondrial) with the spatial variation in the morphological traits that are thought to evolve under predation pressure (color pattern, body length and body shape). The aim of the study was to test the standard view that the divergence between the ecotypes has evolved independently several times. Consistent with the standard view, high-predation and low-predation ecotypes in each stream were often more closely related to each other genetically than either of them is related to similar ecotypes in the other streams. Moreover, the genetic distances between the neutral loci were not significantly correlated with morphological distances, suggesting that selection by current ecological conditions (e.g. predation) is the driving force for morphological variation. The study found that high-predation individuals have elongate body shape compared to low-predation individuals; no other difference in shape was detected between the two ecotypes.

Clearly, the shape differences between high- and low-predation ecotypes were not consistently replicated in different studies, unlike the case with color pattern and life history. One reason for this is that the shape features may vary between streams, as a result, studies conducted in different streams can yield different results (Hendry *et al.*, 2006). A second reason is that shape features are strongly affected by physical habitat features that vary between high-predation and low-predation localities. For example, in Hendry *et al.* (2006) study, canopy openness was found to be the most important determinant of female shape, as females in open canopy sites had more distended abdomens; this is most likely because canopy openness is correlated positively with resource availability and, consequently, with reproductive effort. Supporting this link, females from open canopy sites had more growing embryos than those from closed canopy sites. Water flow was an important determinant of body shape for both sexes; at high flow sites, males and females had deeper caudal peduncles and smaller heads. These features may reflect the need for more thrust and less drag to meet the higher hydrodynamic requirements of swimming under high flow rate.

A third reason that could affect the consistency of the studies of guppy shape is phenotypic plasticity. Because the above-described studies were based on the assessment of wild-caught individuals, they cannot distinguish between genetic shape variation and environmentally-induced shape variation. Burns *et al.* (2009) performed multiple comparisons between the shapes of wild-collected and lab-reared males and females from high- and low predation populations (two populations from each predation regime). Their findings indicate a substantial role of phenotypic

plasticity in determining guppy shapes, moreover, the pattern of plasticity seemed to vary among populations and between sexes, further complicating the observed shape variation. Burns *et al.* (2009) observed differences between populations in the depth of the caudal peduncle, but these differences did not correlate with the predation regimes in the populations of origin. The comparison between wild-caught and lab-reared guppies showed that the shapes of the individuals from some populations were sensitive to rearing conditions (i.e. were phenotypically plastic), while the individuals from other populations did not show much plasticity; again, the differences in the pattern of plasticity did not correlate with predation regimes. The comparison between the responses of males and females indicated that males showed less plasticity than females (i.e. less difference in tail depth between wild and laboratory individuals); this is likely to a be a consequence of the intra-sexual differences in habitat use. In the wild, males exploit microhabitats with relatively still water, similar to the low-flow conditions in the laboratory; as a result, they experienced less change in the rearing conditions between the wild and the laboratory, unlike females. The authors concluded that the effect of predation in their study was probably overwhelmed by the effects of local habitat features and rearing conditions, moreover, they argued that the assessment of wild-caught individuals alone is not a reliable indicator of selection by predators, rather, common garden experiments should also be conducted to quantify the contribution of plasticity to the observed differences between predation regimes (Burns *et al.*, 2009).

Dowdall *et al.* (2012) investigated the role of predator-induced plasticity in a number of guppy traits, including head morphology. They reared family lines of guppies that were obtained from high- and low- predation localities (from two streams) under two sets of experimental conditions, one containing predator cues and the other lacking predator cues, in order to mimic high- and low-predation environments. When guppies were reared under the ancestral, high predation conditions, all the lines converged on the characteristic high-predation head morphology, developing a more fusiform head and a lower position of the eye. However, when reared without the predator cues, guppies from low-predation localities diverged in their head morphology, developing a stouter head shape and an upper position of the eye. Similar patterns of plasticity were observed for the other traits investigated in the study. These results indicate that low-predation guppies are more plastic in their phenotypic expression than high-predation guppies. The authors concluded that the evolution of the derived low-predation ecotype from the ancestral high-predation ecotype was accompanied by the evolution of greater phenotypic plasticity (Dowdall *et al.*, 2012).

1.3   The guppy selection experiment

Experimental studies on fisheries-induced evolution have been criticized for their lack of realism (Hilborn, 2006). This general criticism can be disentangled into a number of specific objections: first, that the selective pressures imposed by the experimenters are higher than those realized in recreational and commercial fisheries (Brown *et al.*, 2008; Hilborn and Minte-Vera, 2008).

Second, that the controlled, simplified experimental environments poorly resemble the natural environments where fisheries-induced evolution is hypothesized to occur, specifically, the experimenters typically impose artificial constraints on population size and the potential for natural and sexual selection (Hilborn, 2006; Conover and Baumann, 2009). Third, the model species used in the experiments are different in many respects from the species targeted by fishing; some of these differences, such as the model species having discrete generations while the exploited species having overlapping generations, might lead to substantially different rates of evolution (Hilborn, 2006; Diaz Pauli and Heino, 2014). Because the ultimate value of the experiments on fisheries-induced evolution rests on the applicability of their conclusions to actual fisheries, the unrealism objection is a serious one.

The selection experiment on guppies described in this thesis attempted to consider the above objections in its design, in order to represent a realistic simulation of the ecological and the evolutionary processes taking places in the fisheries. The harvesting intensities that were imposed during the experiment were similar to those occurring in commercial fisheries. Moreover, the model species, the guppy (*Poecilia reticulata*) have overlapping generations, similar to the exploited species. The experimental populations were age- and size-structured, in addition, the experimental environment allowed for density-dependent feedbacks to take place, and also for natural and sexual selection to occur. As a result, complex interactions between harvesting selection and other ecological and evolutionary processes were not prevented by the design of the experiment, thereby allowing for unanticipated responses to occur and alter the expected outcome of the harvesting selection. The main aim of the selection experiment was to investigate the effects of size-selective harvesting on the evolution of life history traits (maturation, growth and fecundity) under realistic experimental conditions (see Diaz Pauli, 2012).

1.4   Objectives of the study

The present study investigates the potential of size-selective harvesting to induce morphological evolution in exploited fish populations, by using the Trinidadian guppy as a model species. The study focuses on two aspects of guppy morphology, namely, the color and the shape of the males; both color and shape are well-studied traits in guppies, and there are well-established procedures for quantifying them from photographs (i.e. for color: measuring the number and the area of color spots 'Endler, 1980'; for shape: geometric morphometrics analysis 'Langerhans and DeWitt, 2004'). Moreover, the potential for guppy color and shape to evolve under contemporary selective pressures was established from previous studies on the adaptation of guppies to different predation regimes (e.g. Endler, 1980; Langerhans and DeWitt, 2004). Color and shape represent important aspects of fish biology, and both are relevant to the long-term sustainability of the exploited populations; male color is a key trait for female choice (Houde, 1987; Kodric-Brown, 1989; Brooks, 2000), as it indicates the condition and, indirectly, the genetic quality of males (Endler, 1980; 1983), consequently, maintaining the natural variation in male colors is important to sustain

the function of sexual selection and, therefore, to ensure healthy population recruitment in the long term (Rowe and Hutchings, 2003). Likewise, shape affects the swimming performance and therefore directly affects the survival of individuals (e.g. by rendering them more of less capable of escaping predators and fishing gear 'Ghalambor *et al*., 2004'). The question of whether color and shape can evolve under size-selective fishing is therefore relevant to the debate on the impact of fisheries-induced evolution on the productivity and the sustainability of the exploited populations.

Unlike previous studies, the present investigation is specifically focused on the indirect responses to selection, that is, on the morphological changes that arise as a byproduct of selection on other traits. It is well-established that both color and shape can evolve when they are subjected to direct selection; the color selectivity of predatory fish (Endler, 1980), as well as the shape selectivity of gillnets (Reis and Pawson, 1999), were shown to be capable of inducing adaptive changes in color and shape, respectively. A possibility that has not yet been investigated is that color and shape might evolve under fishing pressure even where they are not the traits under direct selection. In the present study, we investigate this possibility in the Trinidadian guppy by using an experimental setup. Because the indirect effects of fishing selection are often unanticipated (Conover and Baumann, 2009), the study does not begin from specific *apriori* predictions about the identity of the changes that will occur during the experiment, only the general hypothesis of there being a difference in morphology (i.e. in male color and/or shape) between the experimental populations that were subjected to different regimes of size-selective harvesting; this hypothesis was tested against the null hypothesis of no morphological difference between these populations.

2. Materials and Methods

2.1 The selection experiment

Below is an overview of the selection experiment on guppies, conducted by the EvoFish research group at the Department of Biology, University of Bergen; a detailed description is presented in Diaz Pauli (2012).

2.1.1   Founder populations

The guppies used in this experiment have descended from wild-caught individuals that were collected from a low-predation site at the Yarra river in Trinidad in April 2009. The original population (F0) was composed of seventy-six females; all assumed to be storing sperm from wild males. They were kept separately in 2-litre aquaria, and their offspring were transferred from those aquaria once they were born. Their offspring (F1) were kept brood-wise in two-liter aquaria until sexing was possible, then separated according to sex.

Between March and July 2010, nine replicated populations were established from the offspring of the original population, each with 140 individuals (70 males and 70 females). To ensure a uniform genetic composition among the founder populations, individuals were assigned to populations in such a way that each F0 female is represented by offspring in each of the nine populations. This is important to ensure that any subsequent genetic differences between the populations are not due to founder effects. Each of the nine populations was maintained in 400-litre aquarium in a flow-through system. All aquaria (before and after the establishment of the founder populations) were maintained at a standard temperature and lighting, and were fed a standard diet. The populations were given four months to grow before the first harvesting event.

The setup of the experiment allows for interaction between individuals from a broad range of ages and sizes, so that ecological feedbacks (caused by competition and cannibalism), as well as natural and sexual selection can take place and alter the outcome of the harvesting regimes. This makes the experiment a more realistic simulation of actual fisheries.

2.1.2   Harvesting regimes

Each of the nine populations (will be referred to as 'harvest populations' from now on) were assigned to one of three lines of harvesting (will be referred to as 'treatments'): 1) positive harvest, in which a proportion P of the fish larger than 16 mm in standard length was removed. 2) negative harvest, in which a proportion p of the fish smaller than 16 mm in standard length was removed. 3) random harvest, in which a proportion p of the fish was removed, regardless to their size. Each

treatment was performed in three replicate populations. The harvesting proportion P was set at 25%, 33% or 50%, depending on the population size. These harvesting rates were chosen such as to be within the range observed in commercial fisheries (e.g. Olsen and Moland, 2011).

Harvesting events were separated by regular intervals of six weeks, during which the populations were allowed to grow and reproduce. The first harvest took place in October 2010. For each population, harvesting was performed in one day. On this day, all the fish were removed from the aquarium and sorted into large (above 16 mm standard length) or small (below 16 mm standard length). Fish from each size group were randomly divided into buckets. Then, the harvesting proportion was removed from the large fish (for the positive treatment), the small fish (for the negative treatment) or both (for the random treatment); this was performed by removing a specific number of buckets from the relevant size group/s. Harvested fish were killed by an overdose of MS-222 (Tricaine methanesulfonate), and their length, weight and maturity status were investigated. The experiment lasted for three years, approximately equivalent to four guppy generations.

2.1.3 Common garden experiments

In addition to the nine harvested populations, parallel experiments were performed with the specific aim of investigating the potential genetic basis of the potential phenotypic differences between harvest populations. In each experiment, four pregnant females were randomly chosen from each of the nine populations and were reared, separately, under common garden conditions. Their offspring (F1 males and females) were mated with each other to produce F2 offspring. Mating was allowed only between individuals from the same population. All descendants (F1 and F2) were also reared under common garden conditions. Investigation of phenotypic traits was performed on F2 offspring. If a systematic phenotypic difference was observed between F2 offspring from different populations, it is reasoned that these differences have heritable basis, since all F2 offspring, as well as their parents, were reared under identical conditions, therefore, the variable environments of the harvest populations cannot be the cause of such a difference, neither directly nor indirectly, such as through maternal effects (Chappell and Odell, 2004). Common garden experiments (will be referred to as 'assays') were performed once each year, from 2010 to 2013.

Note that the selection experiment lacks a real control, because the populations subjected to the random treatment (i.e. the obvious candidate for a control group) are expected to evolve responding to the conditions in the laboratory and also to repeated harvesting. Note, however, that the design of the experiment allows for effective statistical comparisons at multiple levels; firstly, comparisons across time, such as that between successive harvests or assays (both will be referred to as 'generations'). Secondly, comparisons between treatments lines, such as that between the positive, the negative and the random lines of a particular harvest or an assay, and finally, comparisons between harvest and assay populations. A phenotypic response to laboratory

conditions (e.g. absence of predatory fish, food availability, high density) is expected to manifest in comparisons across time, but not between treatments. On the other side, a phenotypic response to the harvesting regimes is expected to show in comparisons across time (e.g. harvest 4-positive and harvest 28-positive) as well as between treatments (e.g. harvest 28-positive and harvest 28-negative). On a second level, a phenotypic response due to plasticity is expected to manifest in comparisons between harvest but not assay populations. In contrast, a phenotypic response due to selection is expected to manifest in comparisons between census as well as between assay populations. Therefore, the experimental design is sensitive enough to distinguish between changes induced by the laboratory conditions and those induced by the harvesting regimes, and also between phenotypic plasticity and evolutionary change.

2.2   Investigation of morphological changes

Photographs of guppies were analyzed to extract color and shape information, and the obtained data were compared, statistically, between different treatments and generations. Only mature males were investigated. Males were distinguished from females by size and color, since males are smaller and more colorful. Mature males were distinguished from immature males by the shape of the gonopodium, which is a modified anal fin used by the males to inseminate the females. Measurements were performed every six weeks, but the current study investigated a subset of this data; for the harvest populations, investigation was limited to the harvests 4 (February 2011) and 28 (November 2013), separated by 23 harvesting episodes. For the assay populations, investigation was limited to the assays 1 (2010), 2 (2011) and 4 (2013). The first assay was performed before harvesting cycles were established. The second and the fourth assays were performed after one and three years (respectively) of regular harvesting.

2.2.1   Photography

Males were anaesthetized with 0.03% aqueous solution of tricaine methanesulfonate (MS-222). Each male was placed on its left side in a white background and blotted with a paper towel to reduce reflection from water droplets. The fins were carefully spread with a wet, fine paint brush. Then the male was photographed by a digital camera (Canon EOS 500D) mounted on a camera stand at a standard height above the background. To avoid shadows, illumination was provided by halogen bulbs situated above the left and right sides of the background (two on each side). A color standard and a millimeter scale were visible in the background for calibration purposes, moreover, a label with basic information about each male (e.g. date, harvest/assay number, population number, individual number) was also visible for identification purposes (figure 2.1). All photographs were taken at the maximum resolution possible for the camera; they were stored in JPEG format and labeled with unique codes. Note that the number and the area of the black spots increase after exposition to the anesthetic solution, while the orange spots remain unchanged

(personal observation; see also Millar *et al*., 2006). In order to minimize the handling time (and hence the mortality risk), only one, left-side photograph was taken for each male, however, males from harvest 28 were photographed four times; two on their left side and two on their right side; this is because these males were subjected to an additional test for measurement error (see below). After photography, all the males were placed in aerated water to recover from anesthesia, then returned to their designated aquaria.

Figure 2.1: A photograph of the left side of a male guppy. Also, showing a color standard, a ruler and a label.

2.2.2   Color changes

2.2.2.1. Image analysis:

All the photographs (census 4 and 28, together with assay 1, 2 and 4) were imported to the image analysis software ImageJ (version 1.50b, http://www.imageJ.nih.gov, Rasband, 1997). For each individual, coloration was quantified by manually delineating the area of each color spot in one side of the body, excluding those in the fins. The total area of that side, excluding the fins, was also delineated (figure 2.2). Area was measured in squared millimeters. The reason why the fins (dorsal, anal and caudal) were excluded is that their fragility makes standard positioning difficult, thereby reducing the reliability of measuring coloration from photographs. In addition, the number of spots from each color category was counted and recorded. All the measurements were performed by one person (E. I.), and the order in which populations were subjected to analysis was random with respect to the treatment and the generation.

Figure 2.2: The measurement of the coloration of a male guppy; the areas of orange and black spots were delineated, together with the total area of the body. Fins were excluded from the measurements of the colored area and the total area.

Coloration was inferred from two complementary parameters, namely, the number and the area of spots from each color category. This is the standard approach for quantifying guppy coloration, thereby rendering our findings comparable to previous studies (e.g. Endler, 1980; Pilastro, 2004; Millar *et al*., 2006, Weese *et al*., 2010, Gotanda and Hendry, 2014). Note that color spots sometimes grade into one another, thereby making the decision of whether a patch of color is sufficiently distinct to be counted, a partly subjective decision (personal observation). The area of each color is obtained by summing the area of all the individual spots of that color. To adjust for body size, the area of each color is divided by the total area of the body side, to get the relative area of the color. This adjustment renders the color estimates independent of size (i.e. so that big males do not appear more colorful just because they are big). This adjustment was particularly important because of the design of the selection experiment, since it imposes systematic size differences between the harvested populations; such differences are likely to bias absolute measurements of the colored area.

Male guppies display a diverse range of colors; following Endler (1978, 1980, 1991), nine colors could be distinguished: orange (includes red), yellow, black, fuzzy black, blue (includes purple), violet-blue, green, bronze-green and silver. The above colors can be grouped into three categories with known biological relevance, since they differ both in their physiological bases and selective importance (Grether, 2000; Brooks and Endler, 2001; Grether *et al*., 2005; Griffith *et al*., 2006); these categories are: carotenoid colors (orange, yellow and red), melanic colors (black, fuzzy black) and structural/iridescent colors (blue, violet, silver and green). In order to reduce the number of variables for the statistical analysis, all the spots that belong to the same color category were summed together. Furthermore, because the structural colors are not reliably represented under the lighting conditions in which males were photographed (personal observation; see also Endler and Mielke, 2005; Kemp *et al*, 2009; Gordon, *et al*., 2015), they were excluded. Consequently, only two color categories were included in the image analysis: black (includes fuzzy black) and orange (includes red and yellow).

To conclude, male coloration was represented by four metrics, measured from the photograph of each individual: 'number of orange spots', 'number of black spots', 'area of orange spots', and 'area of black spots'.

2.2.2.2. Statistical analysis

Statistical analysis was conducted separately on two datasets: 'harvest experiment' and 'assay experiment'; each dataset contains measurements from nine populations that were subjected to different treatments, in addition, measurements from each population were repeated for different generations (i.e. harvests/assays). Consequently, for each dataset, two major comparisons were performed; one between the treatments and one between the generations. The comparisons were performed separately, then jointly, for the four metrics of male coloration. All analysis was performed by using the statistical computing software R (version 3.1.2, https://www.r-project.org, R development Core team, 2012)

Firstly, a preliminary analysis was performed to explore the relations between orange and black coloration, as well as between the multiple metrics of coloration. From the assay dataset, the means of 'spot number' and 'relative area' were calculated for orange and black colors (by using the function 'mean x'), in order to determine which color has more spots and which color covers more area. The median and the standard deviation for the values of each color metric were also calculated (from the 'summary' function), to obtain a general overview of the data. In addition, and also from the assay dataset, a simple correlation test was performed (by using the function 'cor.test (x, y)') between each 'spot number' metric and its corresponding 'relative area' metric (i.e. 'number of orange spots' with 'relative area of orange', and 'number of black spots' with 'relative area of black'). The goal was to determine to which extent the spot numbers and the relative areas are independent of each other; if it turned out that they are strongly correlated, then including both metrics in the analysis would have been redundant. Data from all the treatments and the generations of the assay experiment were combined during this preliminary stage.

Secondly, univariate analysis was conducted by using linear mixed-effects models (from the library 'nlme'). The analysis was performed separately for each dataset, and also separately for each of the four color metrics within each dataset. To satisfy the model assumptions, all the values of the color metrics were transformed to obtain normality; spot numbers were subjected to square-root transformation (by using the function 'sqrt (x)'), while relative areas were subjected to arcsine root transformation (by using the function 'asin (sqrt(x/100))'). The transformed values were used as inputs for the mixed-effects models, and also for the illustrative graphs. The variation among the groups was investigated from five types of models: first, a unified model that simultaneously considers the effects of the generation, the treatment and the interaction between the generation and the treatment. The outputs of this model were obtained from the 'anova' function. Second, a model with the generation as the only main effect, which gives an estimation for the extent and the direction the differences between successive harvests/assays. The outputs of this model were

obtained from the 'summary' function. Third, a model with the generation as the only main effect, applied to a subset of the data that contains only one treatment; this model indicates the changes between generations that occurred in each treatment line, and therefore enables comparing the rate and the direction of change between the treatment lines. The outputs of this model were obtained from the 'summary function. Fourth, a model with the treatment as the only main effect, applied to a subset of the data that contains only the latest generation (harvest 28/assay 4); this models assesses the divergence between treatments by the end of the experiment, and therefore indicates the overall effect of the differential treatments. The outputs of this model were obtained from the 'anova' function. Finally, a model with the treatment as the only main effect, applied to a subset of the data that contains only two treatments from the latest generation (i.e. positive-negative, positive-random, and negative-random); this model enables the assessment of the extent and the direction of the differences between each pair of treatments by the end of the experiment. The outputs of this model were obtained from the 'summary' function.

All the above models were constructed with the population specified as a random effect. This is because, in the harvest dataset, same-population individuals were kept in the same tank, therefore they share the same environment. Moreover, in the assay dataset, F2 individuals share the family (i.e. were descended from the same parents) with some F2 individuals from the same population, but not with F2 individuals from other populations, because inter-population mating was not allowed in the experiment. Therefore, in both datasets, there were reasons to expect measurements from the same population to cluster with each other, violating the statistical assumption of independence, and demanding a mixed-effects treatment. Differences between the groups in the mean values for their color metrics were reported as percentage changes in the mean values between the relevant groups (e.g. 'The relative area of orange spots decreased by 41.2% between the harvests 4 and 28'); to obtain these percentages, the estimated means from the summary function were back-transformed to the original units (by using the functions: $(x)^2$ and $(sin(x))^2$ for the spot numbers and the relative areas, respectively). Groups differences were also illustrated graphically by simple plots (by using the function 'plot (y~x)'). Percentages and plots were cited in the text only where the differences between the groups are significant (here defined as '$p < 0.05$'), marginally significant (here defined as '$p \leq 0.10$'), or in the context of a comparison with significant/marginally significant differences.

Thirdly, Multivariate analysis of variance (MANOVA) was performed by using Wilks' lambda test (calculated from the 'MANOVA' and the 'summary' functions). MANOVA analyses were performed separately for the census and the assay datasets. All the color metrics (i.e. number of orange spots, relative area of orange spots, number of black spots and relative area of black spots) were combined as the dependent variables for the MANOVA models. The values of the color metrics were transformed as described above, to impose a normal distribution on the data. The structure of variation among the groups was investigated from four models, similar to the univariate models described above: first, a model with both the generation and the treatment as independent variables. Second, a model with the generation as an independent variable, applied to

a subset of the data that includes only one treatment line. Third, a model with the treatment as an independent variable, applied to a subset of the data that includes only the latest generation. Finally, a model with the treatment as an independent variable, applied to a subset of the data that includes only two treatments from the latest generation. In order to account for population clustering, the effect of the population, as well as the interaction between the population and the generation, were included as independent variables in all the MANOVA models. The output values for the effects of the population and the population-generation interaction, though mostly significant, were not reported among the results, because they are irrelevant to the hypotheses being tested. The univariate analysis investigates the identity and the amount of group differences in each color trait, while the multivariate analysis investigates the extent to which the combined effects of these differences contribute to overall group separation; therefore, the two approaches complement, rather than replicate, each other.

The null hypothesis Ho can be expressed in two claims: 1) male coloration does not vary with the generation or the treatment and 2) even if there is such a variation, it reflects phenotypic plasticity, not evolutionary change. The alternative hypothesis H1 can be expressed in the two contrary claims: 1) male coloration varies between the generations, the treatments, or both, and 2) this color variation is due to genetic differences between the generations, the treatments, or both, and therefore it indicates the occurrence of evolutionary change. The harvest dataset is relevant to the assessment of claim 1 (of both hypotheses), while the common garden dataset is relevant to the assessment of claim 2 (of both hypotheses). Statistical analysis was performed in order to distinguish between the two hypotheses with respect to both of their claims.

2.2.3　　Shape changes

2.2.3.1 Selection of landmarks

The shape features of male guppies were digitized as configurations of landmarks, and the shape differences between male groups were analyzed and visualized as the relative displacements in landmark positions after superimposition (Klingenberg, 2013). A landmark is a recognizable point in the study subjects that can be precisely located in each individual, and shows a one-to-one correspondence from individual to individual. Configurations of landmarks are fully describable by their (x, y) coordinates, therefore, they are liable to quantitative analysis, unlike actual images. Landmark data serve as the input for geometric morphometrics (Adams *et al*., 2004; Webster and Sheets, 2010). The choice of the landmarks was guided by two criteria: firstly, that the landmarks must be 'good', in the sense that they can be located with minimal measurement error, and secondly, that the landmarks must be 'informative' in the sense that their configuration yields shape information that are relevant to the specific questions addressed by the study. In addition, the choice of the landmarks was constrained by two factors: firstly, the morphology of male guppies, which determines the number of the recognizable points that can serve as landmarks, thereby

setting the upper threshold on how many landmarks can be chosen and, secondly, practical limitations; the more landmarks are chosen, the more the time it will take to place them in each individual and, consequently, the less the individuals that can be covered during the study period.

By taking both the criteria and the constraints into consideration, ten landmarks were chosen as the targets of the geometric morphometrics analysis (figure 2.3). The anatomical definitions of the chosen landmarks are listed in table 2.1. The numbering of the landmarks indicates the fixed order by which they were placed in the photographs. The shape information contained in the landmark data can be visualized in the configurations of these landmarks, as pairs of points or as groups of points; the pair 1-2 gives an indirect indication of the length of the head. The pairs 1-6 and 6-8 indicate the standard length and the caudal fin length, respectively. The pairs 3-4, 5-7 and 9-10 indicate the width of the dorsal, caudal and anal fins, respectively. The points 4, 5, 6, 7, 9 jointly demarcate the caudal peduncle, a muscular region whose shape affects swimming performance (Langerhans and DeWitt, 2004), and was therefore the center of focus for the shape analysis. Shape changes are most easily visualized with wireframe graphs (Figure 2.4), because they represent the changes in the context of the anatomical structures in which they occur (Klingenberg, 2013). The choice of the landmarks followed that of Jayawickrama (2013); a preliminary study that was conducted during an earlier stage of the guppy selection experiment. The methodological similarity between the two studies was intentional, in order to insure that their findings are fully comparable.
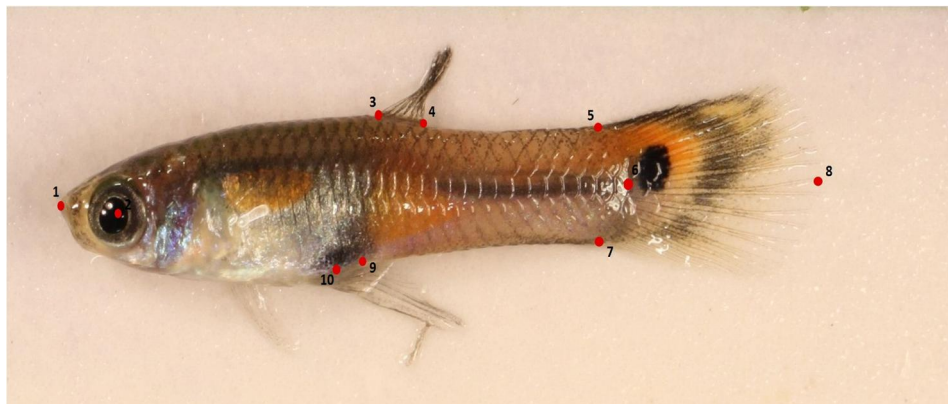
Figure 2.3: The locations of the ten landmarks used to digitize the shape features of male guppies.

Table 2.1: The anatomical definitions of the landmarks used for the geometric morphometrics analysis of male guppies (adopted from Jayawickrama, 2013).

Landmark number	Anatomical definition
1	Tip of the snout
2	Center of the eye orbit
3	Anterior extent of the dorsal fin insertion
4	Posterior extent of the dorsal fin insertion
5	Dorsal extent of the caudal fin insertion
6	Posterior tip of the caudal peduncle
7	Ventral extent of the caudal fin insertion
8	End of the caudal fin
9	Posterior extent of the anal fin insertion
10	Anterior extent of the anal fin insertion
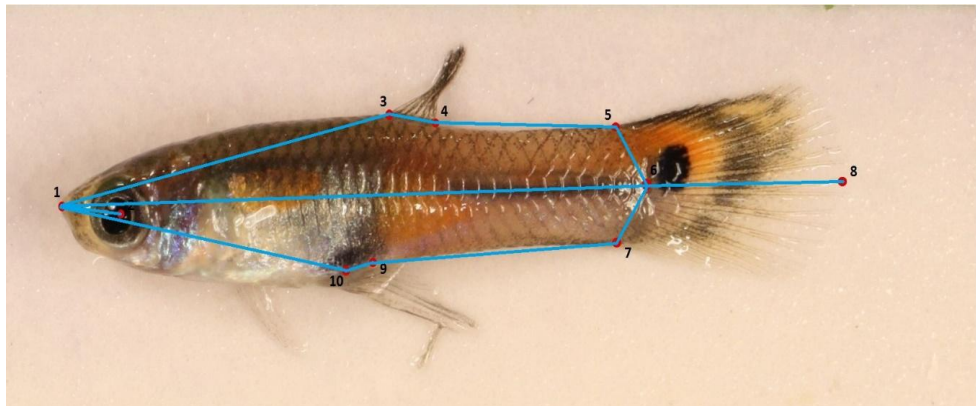
Figure 2.4: The configuration of the landmarks, illustrated by wireframes. The red dots indicate the positions of the landmarks. The blue lines connecting the red dots indicate the distances between the landmarks.

The chosen landmarks differ in their type as well as in their liability to measurement error. The landmarks 3, 4, 5, 7, 9 and 10 belong to type 1, because they occur at the intersections of different tissues. The landmarks 1 and 8 belong to type 2, because they represent points of maximum curvature. The landmark 6 belongs to type 3 because it occurs at the extreme point of a two-dimensional surface (Bookstein, 1991). The landmark 2 is a constructed landmark (Lele and Richtsmeier, 2001), because it is derived by geometric construction from the shape of the neighboring parts. Some landmarks can be located precisely, based only on their anatomical definitions; these are the landmarks 1, 2, 3, 4 and 8. Other landmarks, however, were fuzzy (Valeri *et al*., 1998); their locations in each individual were only broadly specified by their anatomical

definitions. Placing these landmarks was performed by taking the average of multiple trials. The landmarks 5, 6 and 7 often occurred within color spots where it was difficult to distinguish them from neighboring points. The fuzziness of the landmarks 9 and 10 was caused by the fragility of the anal fin, as it was difficult to be positioned in a standard fashion, and some of its rays were often absent.

2.2.3.2. Placement of landmarks

The photographs of the male guppies (from the harvests 4 and 28, together with the assays 2 and 4) were imported to the program tpsUtil (version 1.61, http://life.bio.sunysb.edu/morph/, Rohlf, 2009) in order to build '.tps' files that can be modified by tpsDig and analyzed by morphoJ. A separate '.tps' file was created for the harvest and the assay datasets; all subsequent analyses were conducted separately on the two datasets. Photographs from assay 1 were excluded from the shape analysis because Jayawickrama (2013) has already compared them with assay 2; the purpose of the present study is to track the shape changes that occurred between the middle and the late stages of the selection experiment (i.e. between the harvests 4 and 28, and between the assays 2 and 4), if there is any.

After conversion to the '.tps' format, the ten landmarks were identified from each photograph, and their coordinates were recorded in the relevant '.tps' file by using the program tpsDig2 (version 2.19, http://life.bio.sunysb.edu/morph/, Rohlf, 2004). The placement of the landmarks was performed by one person (E.I.). To ensure that the landmarks correspond reliably, the methodology for landmark placement (and the rules for settling problematic cases) were followed strictly and consistently across individuals and populations, but not across datasets; changes in the placement methodology (for the landmarks 2, 9 and 10) were introduced in the analysis of the assay dataset, guided by the experience that was gained from the earlier analysis of the harvest dataset. Because all the statistical comparisons were performed within, but not between, the two datasets, no bias could have resulted from these methodological improvements. If it was not possible to identify a certain landmark reliably from a certain photograph, the problematic landmark was labeled as missing.

2.2.3.3. Geometric morphometrics analysis

Following the digitization of the landmarks, the '.tps' file for each dataset was imported to morphoJ software for geometric morphometrics (version 1.06d, http://www.flywings.org.uk/morphoj_page.htm, Klingenberg, 2011) where all subsequent analysis was performed. Initially, procrustes superimposition (in which specimens were aligned by principal axis) was applied to the landmark coordinates, in order to purify the shape variation from non-shape variation, such as that caused by size, position and orientation. The resulting procrustes

coordinates were screened for outliers; if a certain individual deviates substantially from the mean shape, a treatment is attempted by swapping pairs of its landmarks under the assumption that these landmarks were mixed up during digitization; if swapping the landmarks didn't reduce the deviation, the outlier is excluded from the analysis. Individuals were grouped based on two criteria or 'classifiers': the generation (i.e. the harvests 4 and 28 for the harvest dataset and the assays 2 and 4 for the assay dataset) and the treatment (i.e. the positive, the negative and the random treatments). Since each generation in each dataset contains individuals from all the treatments, six groups were identified in each dataset (e.g. for the harvest dataset: harvest 4-positive, harvest 4-negative, harvest 4-random, harvest 28-positive, harvest 28-negative, and harvest 28-random). The wireframes were defined by drawing straight lines between specific landmarks, as illustrated in figure 2.4. In order to account for allometry (i.e. the dependence of shape on size), a regression analysis was performed between centroid size and procrustes coordinates. Size correction was particularly important because different groups were expected, given the design of the selection experiment, to vary in their mean size. The regression residuals, which represent the allometry-free shape variation, were used as the inputs for subsequent multivariate analysis.

In order to assess the reliability of the shape digitization, measurement error was quantified from a set of photographs of the males from harvest 28. Each male was photographed on the left and the right sides, and each side was photographed twice. Landmarks were placed separately on each of the four photographs. Following the procrustes superimposition, the screening for outliers, and the definition of the classifiers, measurement error was quantified by applying procrustes anova test to the dataset. In this test, the combined effects of imaging error and landmark placing error (as shown by the differences between the photographs of the same side of the same individual) was compared to the effect of a subtle biological factor, namely, fluctuating asymmetry (as shown by the differences between the photographs of the different sides of the same individual). This comparison gives an indication of the magnitude of error variation relative to biological variation, and therefore evaluates the reliability of the method of inferring shape changes by analyzing photographs

The purpose of the statistical analysis was to assess the extent of the shape variation among the groups, and also to identify the shape features that vary between these groups. The null hypothesis (Ho) asserts that: 1) the male shape does not vary with the treatment or the generation, and 2) even if there is such a variation, it is due to phenotypic plasticity, not evolutionary change. To the contrary, the alternative hypothesis asserts that: 1) male shape varies with the generations, the treatments, or both, and 2) the shape variation is due to genetic differences between the generations, the treatments, or both, therefore it indicates the occurrence of evolutionary change. The harvest dataset is relevant to the assessment of the claims regarding phenotypic variation, while the assay dataset is relevant to the assessment of the claims regarding genetic variation. In order to distinguish between the two hypotheses with respect to both of their claims, the regression residuals from all the groups were subjected to a series of comparisons by using multivariate ordination analysis.

Three types of statistical comparisons were performed; first, pairwise comparisons between generations (i.e. between harvest 4 and harvest 28, as well as between assay 2 and assay 4), in order to assess the shape changes that occurred through time as the guppies from all populations were adapting to their (shared) experimental environment. Second, pairwise comparisons between each group in the early generation and the corresponding group in the late generation (e.g. harvest 4-positive and harvest 28-positive), in order to investigate the shape changes that occurred through time within each treatment line. Pairwise comparisons were performed by using the discriminant function analysis test (DFA), as it is the best method for investigating the shape variation between two groups (Klingenberg, 2011). Third, comparisons were performed between the treatments within each generation (e.g. harvest 4-positive, harvest 4-negative, and harvest 4-random). The statistical method used for comparing the treatments was the canonical variate analysis test (CVA), as it is the best method for investigating the structure of the shape variation among more than two groups (Klingenberg, 2011). In order to obtain P values for the null hypotheses, permutation tests with 10,000 iterations were applied for both the DFA and CVA tests. All the above analyses were performed on the datasets that contain only left-side photographs; in addition, the CVA test on the treatments of harvest 28 was independently performed on the larger dataset that contain two photographs from each side per each individual, in order to test the robustness of the results. Because the two parallel CVA tests yielded similar outputs, only the outputs of the more comprehensive test were reported, to avoid unnecessary repetition. For all the above analysis, both the numerical and the graphical outputs were reported, including the Mahalanobis and the Procrustes distances and their respective probability estimates, the wireframe graphs of the shape changes, the plots of the discriminant and the cross validation scores (for the DFA test), and the ordination plots of the canonical scores (for the CVA test). To obtain an overview of the separation between all the six groups simultaneously, a generalized CVA test was performed that includes both the generation and the treatment as classifiers; only the plot of the canonical scores was reported for this test, because the rest of the outputs were equivalent to the outputs of the earlier tests. All the above analyses were performed separately on the harvest and the assay datasets.

2.2.3.4 Interpretation of the shape graphs

The shape changes were inferred from the relative shifts in the positions of the landmarks between the starting and the target shapes in the wireframe graphs; the shifts were identified either visually or by the program ImageJ. Particular interest was given to the relative positions of the dorsal, the anal, and the caudal fin insertions, as they indicate changes in the relative size of the caudal peduncle, which is the region responsible for generating most swimming power (Langerhans and Dewitt, 2004). Other potential shape changes were also relevant to the purposes of the present study, including changes in the length of the body relative to its width, changes in the caudal fin length relative to the standard length, and changes in the width of the dorsal, the anal, and the caudal fins relative to the rest of the body; all these shape changes can be linked to aspects of the swimming performance and, consequently, to the individual's ability to search for food, escape

from predators, and display to females (Langerhans and Reznick, 2009). The interpretation of the shape changes was performed in accordance with the guidelines presented by Klingenberg (2013) in the context of discussing visualizations in geometric morphometrics.

Two points must be kept in mind while interpreting (and reading the interpretations of) the shape graphs: firstly, that the observed shifts in landmark positions are relative (Klingenberg, 2013); variation in absolute size, as well as its allometric effects on the shape, were statistically removed through procrustes superimposition and regression, respectively. The remaining variation is specifically about the relative positions of the body parts. Ideally, we would describe all the shifts in explicitly relative terms, however,

it was often easier to follow to use shorthand descriptions of the landmarks as lengthening, widening, moving anteriorly or posteriorly, dorsally or ventrally. The reader is reminded that, by that, it is meant that the landmarks in question are changing position relative to the other landmarks; changes in one body dimension (e.g. length) are relative to the other dimension (e.g. width), and movements in one body part are relative to the other parts. Secondly, the shape graphs are not reliable indicators of the magnitudes of the shape changes, because the changes are scaled differently in different graphs. Ideally, we would use the same scale factor for all the shape graphs, so that the reader can get an overview of which pairs of groups showed more change simply by looking at their graphs. However, shape graphs are constrained by a tradeoff between visibility and distortion (Klingenberg, 2013); the scale factor must be set such that it is neither too small so that the changes are invisible or too large so that the changes are distorted. Because the amount of the shape variation differed substantially between the comparisons, no single scale factor was adequate (i.e. make the changes visible without distorting them) for all the comparisons. The scale factor for each shape graph was stated in the figure legend.

3. Results

Overall, morphological analysis was performed on 692 male guppies; of these, 350 males belonged to harvest dataset and 342 males belonged to assay dataset. Regarding the census dataset, 234 males belonged to harvest 4, and 116 males belonged to harvest 28. Regarding the assay data set, 97 males belonged to assay 1, 83 males belonged to assay 2 and 162 males belonged to assay 4. The variable numbers reflect the variable demographic histories of the experimental populations, because the experiment did not impose a fixed population size. Each harvest/assay was composed of positive, negative and random treatments; each treatment was represented by three populations (Table 3.1). All the males were subjected to the color analysis, but the males of assay 1 were excluded from the shape analysis.

The results contain a large body of details, but it is not necessary that the reader goes through all of them. An overview of the results can be obtained by reading the synthesis sections at the end of each part (i.e. the sections 3.1.4 and 3.2.4), where the key findings are reported.

Table 3.1: The total numbers of individuals in each group of male guppies that was investigated during the study; the group is defined in terms of the harvest and the treatment.

Harvest/Assay	Positive	Negative	Random	Total
Harvest 4	100	85	49	234
Harvest 28	45	39	32	116
Assay 1	32	31	34	97
Assay 2	29	27	27	83
Assay 4	53	53	56	162

3.1 Color changes

3.1.1 Color metrics

The means of the color metrics (the spot numbers and relative areas of orange and black colors), as well as the correlations between them, were calculated from the assay dataset by combining all the assays and the treatments. For the number of spots, the mean of black spots was higher than the mean of orange spots. For the relative area of spots, the mean of orange spots was higher than the mean of black spots (table 3.2). This indicates that the orange spots were, on average, fewer but larger than the black spots. The metrics of black coloration (spot numbers and relative areas) correlated moderately with each other, while the metrics of orange coloration showed a weak correlation. The numbers of orange and black spots showed a very weak, but significant correlation. The relative areas of orange and black spots showed no significant correlation (table 3.3). This shows that orange and black coloration were generally poor indicators of each other.

Table 3.2: Descriptive statistics for the values of the color metrics, estimated from the assay dataset. The values from all the assays and the treatments were combined.

| Color metrics	Mean	Median	Standard deviation
Number of orange spots	3.091	3.000	0.951
Number of black spots	4.061	4.000	1.215
Area of orange spots	0.137	0.139	0.046
Area of black spots	0.064	0.062	0.020

Table 3.3: Correlations between the values of the color metrics, estimated from the assay dataset. The values from all the assays and the treatments were combined. Abbreviations: AO = area of orange spots, AB = area of black spots, NO = number of orange spots, NB = number of black spots. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Pair of color metrics	Df	t-value	r	P-value
AO ~ NO	340	6.539	0.337	< 0.0001 ***
AB ~ NB	340	9.291	0.450	< 0.0001 ***
NO ~ NB	340	2.749	0.147	0.0063 **
AO ~ AB	340	0.290	0.016	0.7717

3.1.2 Harvest populations

3.1.2.1. Univariate comparisons

Regarding orange coloration, the numbers of spots decreased by 35.7% between the harvests 4 and 28 (ANOVA: F 1, 338 = 3.943. P = 0.05) (figure 3.1). Statistical analysis indicates that the interaction between the harvest and the treatment was significant (ANOVA: F 2, 338 = 5.488. P = 0.01); consistent with this, figure 3.2 shows that all the treatments have experienced a decrease in the numbers of orange spots between harvests, but the randomly-selected line showed the least decrease. There was no significant difference between whole treatments (i.e. treatments that combine individuals from all generations; will be referred to as 'lines') (ANOVA: F 2, 6 = 2.351. P = 0.18). Pairwise comparisons between the harvests 4-positive and 28-positive, 4-negative and 28-negative and 4-random and 28-random showed variable rates of decrease in the numbers of orange spots, ranging from 54.6% for the positive line, and 24.4% for the negative line, to 14.2% for the random line. The decrease was highly significant for the positive line, marginally non-significant for the negative line, and non-significant for the random line. Overall, the treatments of harvest 28 did not differ significantly from each other (ANOVA: F 2, 6 = 2.820. P = 0.14). However, pairwise comparisons revealed a marginally non-significant difference between the positive and the random treatments; the mean number of orange spots for the random treatment was larger by 20.8% than that for the positive treatment (figure 3.3). Differences between the positive and the negative treatments, as well as between the negative and the random treatments, were non-significant (see table 3.4).
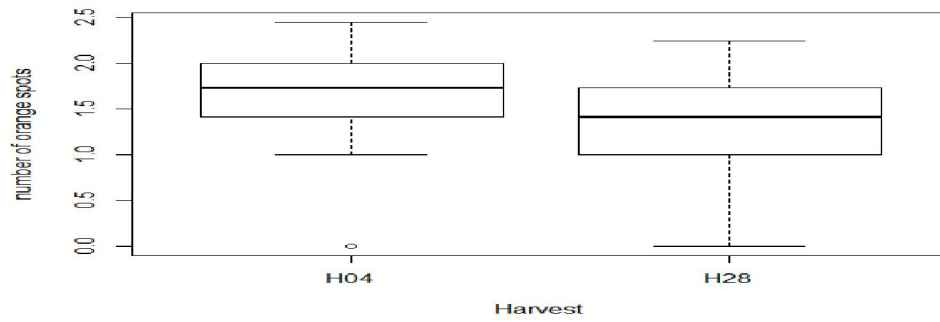
Figure 3.1: A comparison of the numbers of orange spots between the harvests 4 (H04) and 28 (H28). Groups belong to the harvest dataset. Spot numbers are square-root transformed.
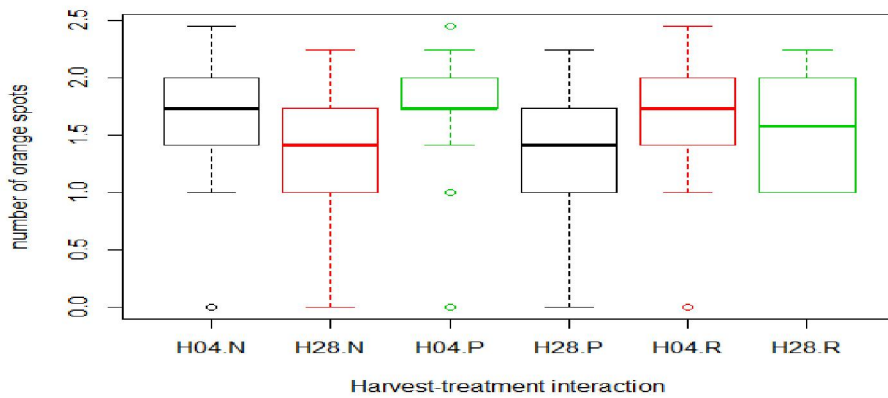
Figure 3.2: A comparison of the numbers of orange spots between the positive (P), the negative (N) and the random (R) treatments of harvest 4 (H04) and these of harvest 28 (H28). Groups belong to the harvest dataset. Spot numbers are square-root transformed.
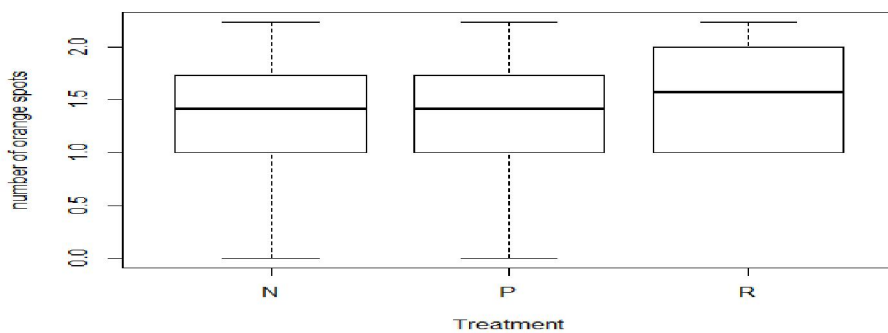
Figure 3.3: A comparison of the numbers of orange spots between the positive (P), the negative (N) and the random (R) treatments of harvest 28 (H28). Groups belong to the harvest dataset. Spot numbers are square-root transformed.

Table 3.4: Variation in the numbers of orange spots between the groups of the harvest dataset. The number of orange spots represents the dependent variable, while the harvest and the treatment represent the independent variables. Spot numbers are square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: H = harvest, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Difference	Std.error	Df	T-value	P-value
H 4, H 28	-0.338	0.063	340	-5.387	<0.0001 ***
H 4-P, H 28-P	-0.590	0.095	141	-6.207	<0.0001 ***
H 4-N, H 28-N	-0.210	0.112	120	-1.870	0.0639 -
H 4-R, H 28-R	-0.122	0.117	77	-1.045	0.2992
H 28-P, H 28-N	0.182	0.130	6	1.402	0.2103
H 28-P, H 28-R	0.321	0.137	6	2.338	0.0580 -
H 28-N, H 28-R	0.139	0.142	6	0.980	0.3648

The relative area of orange spots decreased by 47.3% between the harvests 4 and 28 (ANOVA: F 1, 338 = 6.509. P = 0.01) (figure 3.4). The effects of the harvest and the treatment interacted significantly (ANOVA: F 2, 338 = 4.512. P = 0.01) (figure 3.5). There was no significant difference between the treatment lines (ANOVA: F 2, 6 = 2.671. P = 0.15). Pairwise comparisons between the harvests 4-positive and 28-positive, 4-negative and 28-negative and 4-random and 28-random showed a consistent decrease in all the treatment lines; the positive line experienced the largest decrease in orange area (by 64.2%), the decreases observed for the negative and the random lines were 34.7% and 31.8%, respectively. Pairwise differences were highly significant for the positive line, and significant for the negative and the random lines. There was no significant difference between the treatments of harvest 28 (ANOVA: F 2, 6 = 1.068. P = 0.40). Pairwise differences, within harvest 28, between the positive and the negative, the positive and the random and the negative and the random treatments were all non-significant (see table 3.5).

Figure 3.4: A comparison of the relative areas of orange spots between the harvests 4 (H04) and 28 (H28). Groups belong to the harvest dataset. Relative areas are arcsine square-root transformed.
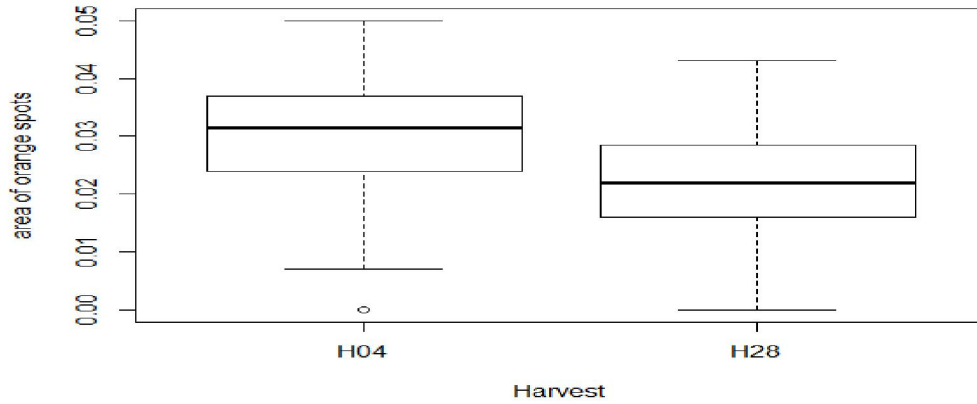

Figure 3.5: A comparison of the relative areas of orange spots between the positive (P), the negative (N) and the random (R) treatments of harvest 4 (H04) and these of harvest 28 (H28). Groups belong to the harvest dataset. Relative areas are arcsine square-root transformed.
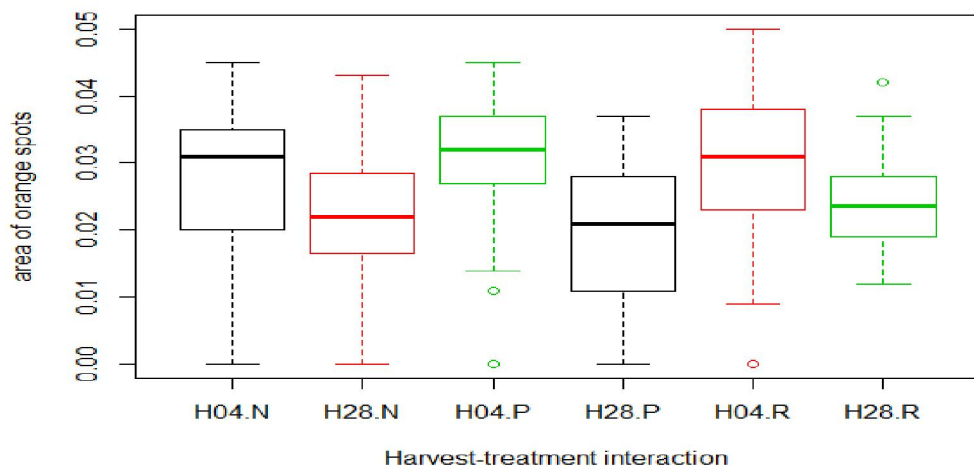
Table 3.5: Variation in the relative areas of orange spots between the groups of the harvest dataset. The relative area of orange spots represents the dependent variable, while the harvest and the treatment represent the independent variables. Relative areas are arcsine square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: H = harvest, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Difference	Std.error	Df	T-value	P-value
H 4, H 28	-0.008	0.001	340	-6.667	<0.0001 ***
H 4-P, H 28-P	-0.012	0.002	141	-6.862	<0.0001 ***
H 4-N, H 28-N	-0.005	0.002	120	-2.462	0.0152 *
H 4-R, H 28-R	-0.005	0.002	77	-2.197	0.0310 *
H 28-P, H 28-N	0.003	0.004	6	0.783	0.4636
H 28-P, H 28-R	0.006	0.004	6	1.458	0.1950
H 28-N, H 28-R	0.003	0.004	6	0.685	0.5191

Regarding black coloration, the numbers of spots showed no significant difference between the harvests (ANOVA: F 1, 338 = 1.457. P = 0.23), or between the treatment lines (ANOVA: F 2, 6 = 2.865. P = 0.13). In addition, there was no significant interaction between the harvest and the treatment (ANOVA: F 2, 338 = 0.195. P = 0.82). Pairwise comparisons, within each treatment line, between the harvests 4 and 28 did not reveal a significant change in any line. Within harvest 28, the treatments did not differ significantly in the numbers of black spots (ANOVA: F 2, 6 = 1.872, P-value = 0.23). Pairwise comparisons between the treatments of harvest 28 did not detect a significant difference in any pair (see table 3.6).

Table 3.6: Variation in the numbers of black spots between the groups of the harvest dataset. The number of black spots represents the dependent variable, while the harvest and the treatment represent the independent variables. Spot numbers are square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: H = harvest, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Difference	Std.error	Df	T-value	P-value
H 4, H 28	-0.042	0.036	340	-1.176	0.2404
H 4-P, H 28-P	-0.026	0.056	141	-0.476	0.6346
H 4-N, H 28-N	-0.076	0.061	120	-1.249	0.214
H 4-R, H 28-R	-0.036	0.072	77	-0.493	0.6231
H 28-P, H 28-N	-0.168	0.102	6	-1.649	0.1502
H 28-P, H 28-R	0.010	0.104	6	0.099	0.9240
H 28-N, H 28-R	0.179	0.105	6	1.696	0.1407

The relative area of black spots increased by 6.8% between the harvests 4 and 28 (figure 3.6). The effect of the harvest was non-significant when considered jointly with the effects of the treatment and the harvest-treatment interaction (ANOVA: F 1, 338 = 0.005. P = 0.94). However, a pairwise comparison between the two harvests (i.e. a comparison in which the harvest is the only main effect), showed the difference in the black area to be statistically significant. The black area did not differ significantly between the treatment lines (ANOVA: F = 2, 6 = 0.794. P = 0.49), and there was no significant interaction between the harvest and the treatment (ANOVA: F 2, 338 = 1.085. P = 0.34). Within the treatment lines, the black area increased (between the harvests) by 10.7% and 8.8% in the positive and the random lines, respectively; the difference was significant for the positive treatment and marginally non-significant for the random treatment. The negative line showed no significant change between the harvests (figure 3.7). Within harvest 28, no significant difference was detected between the treatments (ANOVA: F 2, 6 = 1.822. P = 0.24). Pairwise differences, within harvest 28, between the positive and the negative, the positive and the random, and the negative and the random treatments were non-significant (see table 3.7).
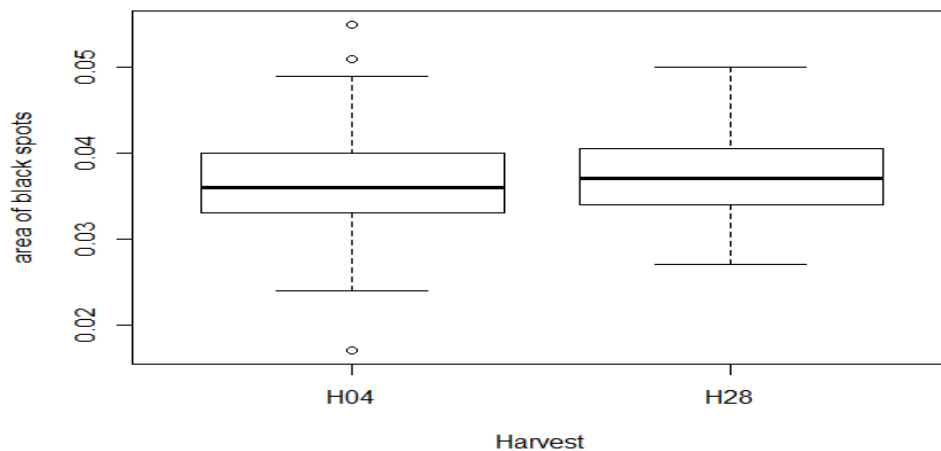
Figure 3.6: A comparison of the relative areas of black spots between the harvests 4 (H04) and 28 (H28). Groups belong to the harvest dataset. Relative areas are arcsine square-root transformed.

Figure 3.7: A comparison of the relative areas of black spots between the positive (P), the negative (N) and the random (R) treatments of harvest 4 (H04) and those of harvest 28 (H28). Groups belong to the harvest dataset. Relative areas are arcsine square-root transformed.
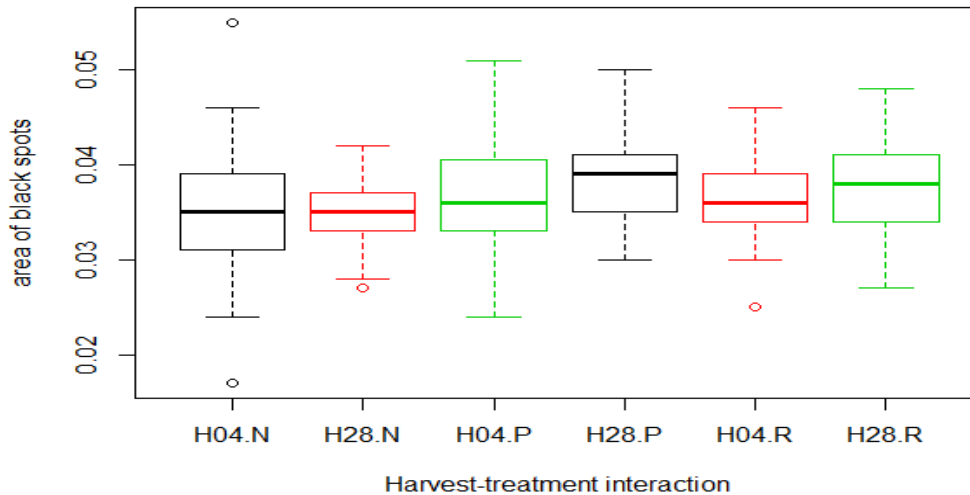
Table 3.7: Variation in the relative areas of black spots between the groups of the harvest dataset. The relative area of black spots represents the dependent variable, while the harvest and the treatment represent the independent variables. Relative areas are arcsine square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: H = harvest, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = $p < 0.001$, '**' = $p < 0.01$, '*' = $p < 0.05$, '-' = $p \leq 0.10$.

Comparison	Difference	Std.error	Df	T-value	P-value
H 4, H 28	0.001	0.001	340	2.140	0.0331 *
H 4-P, H 28-P	0.002	0.001	141	2.017	0.0456 *
H 4-N, H 28-N	0.0001	0.001	120	0.121	0.904
H 4-R, H 28-R	0.002	0.001	77	1.644	0.1042 -
H 28-P, H 28-N	-0.004	0.002	6	-1.748	0.1310
H 28-P, H 28-R	-0.0004	0.002	6	-0.188	0.8573
H 28-N, H 28-R	0.003	0.002	6	1.535	0.1758

3.1.2.2. Multivariate comparisons

Male color (estimated from the numbers and the relative areas of orange and black spots) varied across the harvests and between the treatments; general multivariate analysis through Wilks' lambda test showed the variation to be highly significant, indicating that the relevant groups (i.e. the harvests and the treatment lines) are well separated by color. The interaction between the harvest and the treatment was marginally significant. Pairwise comparisons between the harvests

4 and 28 within each treatment line yielded variable results. For the positive treatment, there was a highly significant change between the harvests. For the random treatment, the change was significant. For the negative treatment, no significant change was detected between the harvests. Overall, the positive, the negative and the random treatments of harvest 28 showed a highly significant difference in male color. However, comparisons between pairs of treatments (from harvest 28) did not detect a significant difference between the positive and the random or the negative and the random treatments; the only significant difference was between the positive and the negative treatments, and it was highly significant (see table 3.8).

Table 3.8: Variation in the male coloration between the groups of the harvest dataset. The numbers and the relative areas of orange and black spots represent the dependent variables, while the harvest, the treatment and the population represent the independent variables. The values for the population variable are not shown. Spot numbers are square-root transformed, and relative areas are arcsine square-root transformed. Statistical analysis was performed by using Wilks' lambda test, specified from the 'manova' and 'summary' functions in R. Abbreviations: H = harvest, T = treatment, H:T = harvest-treatment interaction, P = positive, N = negative, R = random, Wilks = Wilks' lambda values, approx F-value = approximate F-values, num Df = numerator's degrees of freedom, den Df = denominator's degrees of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Df	Wilks	num Df	den Df	approx F-value	P-value
H	1	0.855	4	329	13.936	< 0.0001 ***
T	2	0.925	8	658	3.282	0.0011 **
H:T	2	0.958	8	658	1.788	0.0764 -
H 4-P, H 28-P	1	0.696	4	136	14.822	< 0.0001 ***
H 4-N, H 28-N	1	0.941	4	115	1.801	0.1333
H 4-R, H 28-R	1	0.850	4	72	3.172	0.0186 *
H 28-P, H 28-N, H 28-R	2	0.759	8	208	3.845	0.0003 ***
H 28-P, H 28-N	1	0.667	4	32	3.985	0.0098 **
H 28-P, H 28-R	1	0.883	4	29	0.958	0.4455
H 28- N, H 28-R	1	0.842	4	27	1.264	0.3082

3.1.3   Assay populations

3.1.3.1. Univariate comparisons

Regarding orange coloration, the numbers of spots did not vary significantly between the assays (ANOVA: $F_{2, 327}$ = 0.117. P-value = 0.89) or between the treatment lines (ANOVA: $F_{2, 6}$ = 2.640. P-value = 0.15). In addition, there was no significant interaction between the assay and the treatment (ANOVA: $F_{4, 327}$ = 1.689. P = 0.15). However, pairwise comparisons between the assays showed that the number of orange spots has increased by 11.7% between the assays 1 and 2, then decreased by 13.8% between the assays 2 and 4 (figure 3.8). The changes between the assays 1 and 2, and between the assays 2 and 4 were significant and highly significant, respectively, while the overall change between the assays 1 and 4 was non-significant. Within the positive line, the numbers of orange spots fluctuated through time, increasing by 29.5% between the assays 1 and 2, then decreasing by 20.5% between the assays 2 and 4; both changes were statistically

significant. Within the random line, the numbers of orange spots also fluctuated, increasing by 4.8% between the assays 1 and 2, then decreasing by 16.7% between the assays 2 and 4; the changes were non- significant and highly significant, respectively. Within the negative line, no significant changes in the numbers of orange spots were detected between the assays (figure 3.9). Within assay 4, there was no significant difference between the treatments (ANOVA: F 2, 6 = 0.311. P-value = 0.74). Pairwise comparisons between the treatments of assays 4 showed that the differences between all the pairs (positive-negative, positive-random, and negative-random) are non-significant (see table 3.9).
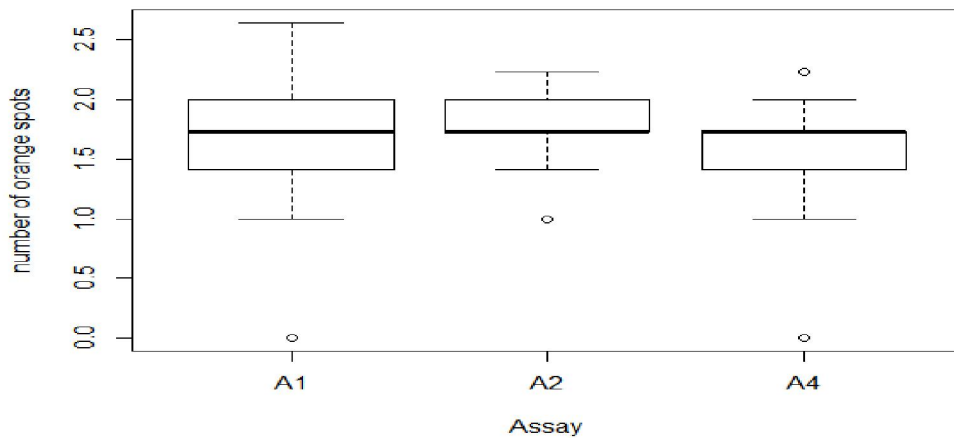
Figure 3.8: A comparison of the numbers of orange spots between the assays 1, 2 and 4 (A1, A2, and A4). Groups belong to the assay dataset. Spot numbers are square-root transformed.
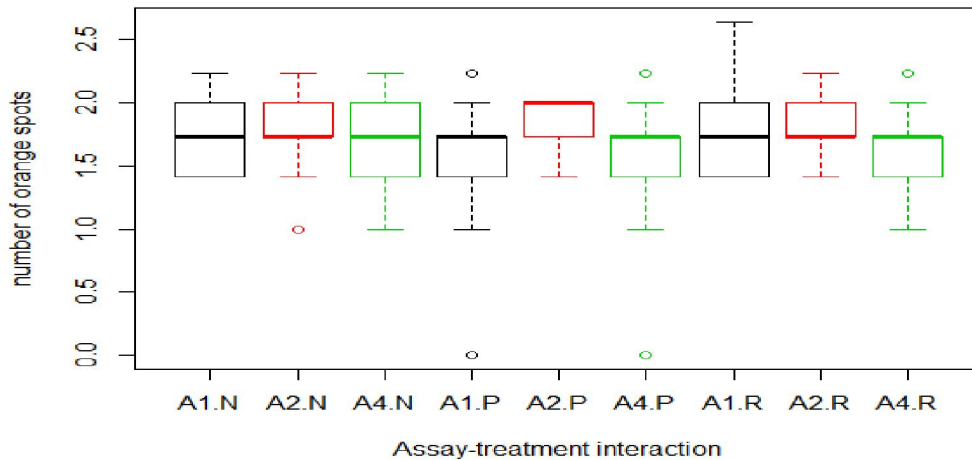
Figure 3.9: A comparison of the numbers of orange spots between the positive (P), the negative (N) and the random (R) treatments of the assays 1, 2 and 4 (A1, A2 and A4). Groups belong to the assay dataset. Spot numbers are square-root transformed.

Table 3.9: Variation in the numbers of orange spots between the groups of the assay dataset. The number of orange spots represents the dependent variable, while the assay and the treatment represent the independent variables. Spot numbers are square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: A = assay, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Difference	Std.error	Df	t-value	P-value
A1, A2	0.098	0.044	331	2.219	0.0272 *
A2, A4	-0.123	0.040	331	-3.096	0.0021 **
A1, A4	-0.025	0.038	331	-0.671	0.5026
A 1-P, A 2-P	0.225	0.087	109	2.579	0.0112 *
A 2-P, A 4-P	-0.177	0.079	109	-2.256	0.0260 *
A 1-P, A 4-P	0.048	0.076	109	0.626	0.5328
A 1-N, A 2-N	0.032	0.073	106	0.440	0.6611
A 2-N, A 4-N	-0.032	0.065	106	-0.487	0.6271
A 1-N, A 4-N	0.0001	0.062	106	0.002	0.9982
A 1-R, A 2-R	0.042	0.065	112	0.653	0.5151
A 2-R, A 4-R	-0.157	0.059	112	-2.665	0.0088 **
A 1-R, A 4-R	-0.115	0.055	112	-2.099	0.0380 *
A 4-P, A 4-N	0.040	0.056	6	0.723	0.4968
A 4-P, A 4-R	0.005	0.055	6	0.095	0.9276
A 4-N, A 4-R	-0.035	0.055	6	-0.638	0.5469

The relative area of orange spots showed apparent variation across the assays and between the treatment lines. The orange area increased by 11% between the assays 1 and 4 (figure 3.10). In addition, the random line had larger orange area than the negative and the positive lines (by 9.6% and 9.5%, respectively), respectively (figure 3.11). Differences between the assays, as well as between the treatment lines were marginally non-significant. (for the assay, ANOVA: F 2, 327 = 2.878. P = 0.06; for the treatment, ANOVA: F 2, 6 = 4.153, P = 0.07). There was a highly significant interaction between the assay and the treatment (ANOVA: F 4, 327 = 6.958. P < 0.01). Pairwise comparisons showed a marginally non-significant increase in the orange area by 11.1% between the assays 1 and 2, and a non-significant change between the assays 2 and 4, however, the overall increase across the assays was statistically significant (figure 3.10). For the positive treatment, the orange area fluctuated across the assays, substantially increasing by 61.4% between the assays 1 and 2, then decreasing by 26.7% between the assays 2 and 4; the changes were highly significant and significant, respectively. For the random treatment, the orange area also fluctuated but in the reverse direction, decreasing by 10.6% between the assays 1 and 2, then increasing by 7.1% between the assays 2 and 4; both changes, however, were non-significant. For the negative treatment, there was a non-significant decrease in the orange area between the assays 1 and 2 (by 5.5%), followed by a significant increase by 20.7% between the assays 2 and 4 (figure 3.12). Within assay 4, the differences between the treatments were non-significant (ANOVA: F 2, 6 = 254, P = 0.78); pairwise comparisons confirmed the absence of any significant difference between the treatments (see table 3.10).
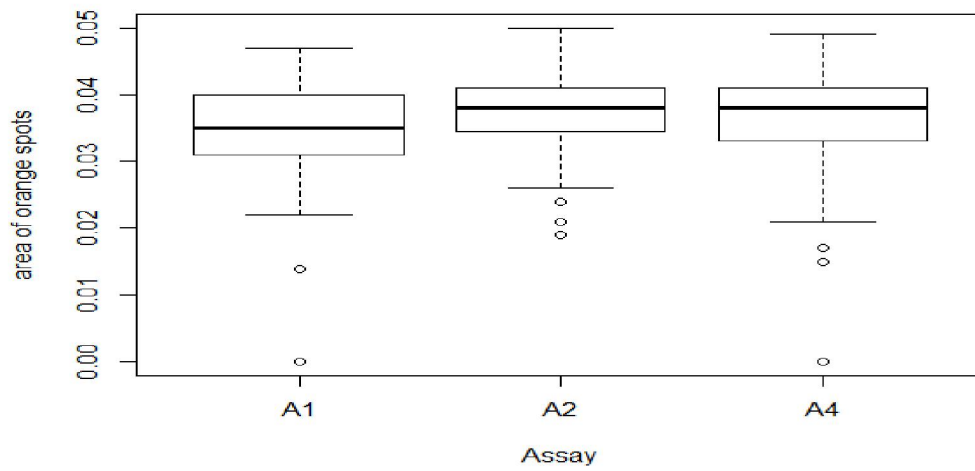
Figure 3.10: A comparison of the relative areas of orange spots between the assays 1, 2 and 4 (A1, A2, and A4). Groups belong to the assay dataset. Relative areas are arcsine square-root transformed.
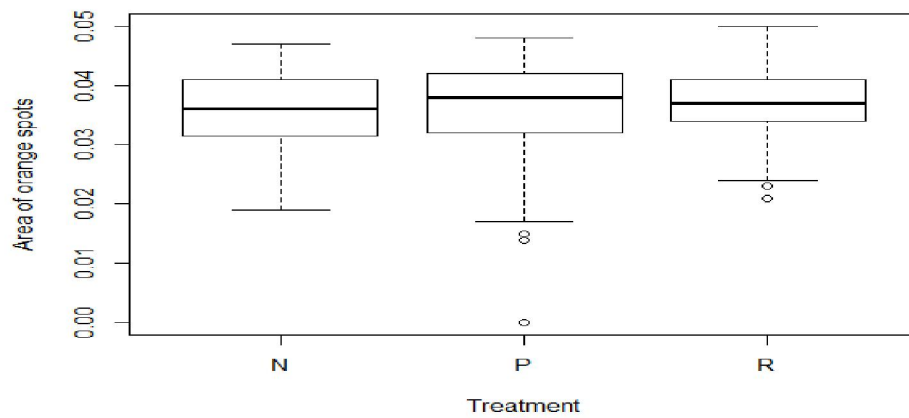
Figure 3.11: A comparison of the relative area of orange spots between the positive (P), the negative (N) and the random (R) lines of the assay dataset. Relative areas are arcsine square-root transformed.
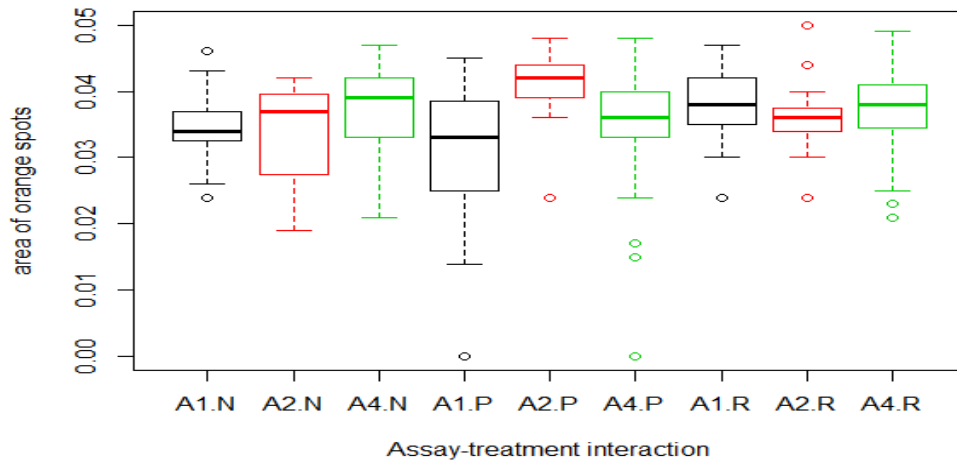
Figure 3.12: A comparison of the relative areas of orange spots between the positive (P), the negative (N) and the random (R) treatments of the assays 1, 2 and 4 (A1, A2 and A4). Groups belong to the assay dataset. Relative areas are arcsine square-root transformed.

Table 3.10: Variation in the relative areas of orange spots between the groups of the assay dataset. The relative area of orange spots represents the dependent variable, while the assay and the treatment represent the independent variables. Relative areas are arcsine square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: A = assay, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = $p < 0.001$, '**' = $p < 0.01$, '*' = $p < 0.05$, '-' = $p \leq 0.10$.

Comparison	Difference	Std.error	Df	t-value	P-value
A1, A2	0.002	0.001	331	1.846	0.0658 -
A2, A4	-0.00001	0.001	331	-0.012	0.9907
A1, A4	0.002	0.001	331	2.144	0.0327 *
A 1-P, A 2-P	0.009	0.002	109	4.373	< 0.0001 ***
A 2-P, A 4-P	-0.005	0.002	109	-2.568	0.0116 *
A 1-P, A 4-P	0.004	0.002	109	2.366	0.0197 *
A 1-N, A 2-N	-0.001	0.002	106	-0.597	0.5517
A 2-N, A 4-N	0.003	0.001	106	2.353	0.0205 *
A 1-N, A 4-N	0.002	0.001	106	1.765	0.0804 -
A 1-R, A 2-R	-0.002	0.002	112	-1.375	0.1719
A 2-R, A 4-R	0.001	0.001	112	0.966	0.3363
A 1-R, A 4-R	-0.001	0.001	112	-0.589	0.5571
A 4-P, A 4-N	0.001	0.003	6	0.535	0.6121
A 4-P, A 4-R	0.002	0.003	6	0.676	0.5243
A 4-N, A 4-R	0.0004	0.003	6	0.140	0.8932

Regarding black coloration, the numbers of spots did not differ significantly between the assays (ANOVA: $F_{2, 327} = 1.619$. $P = 0.20$) or between the treatments (ANOVA: $F_{2, 6} = 1.598$. $P = 0.28$). The effect of the treatment did not interact significantly with that of the assay (ANOVA: $F_{4, 327} = 1.396$. $P = 0.24$). Pairwise comparisons between the assays showed that the numbers of black spots decreased by 9.1% between the assays 1 and 2, then by 2.4% between the assays 2 and 4; the changes were significant and non-significant, respectively; the overall decrease across the assays was highly significant. For the random treatment, the numbers of black spots decreased by 11.8% between the assays 1 and 2, then by 7.5% between the assays 2 and 4; the changes were marginally non-significant and non-significant, respectively. However, the overall decrease across the assays was highly significant. For the negative treatment, there was a consistent but non-significant decrease in the numbers of black spots across the assays, leading to an overall, non-significant decrease by 11.1% between the assays 1 and 4. The positive treatment apparently fluctuated across assays, decreasing by 9.3% between the assays 1 and 2, then increasing by 6.8% between the assays 2 and 4; both changes, however, were statistically non-significant (figure 3.13). Within assay 4, there was no significant difference between the treatments (ANOVA: $F_{2, 6} = 0.142$, $P = 0.87$). Pairwise comparisons between the treatments of assay 4 confirmed that the differences between them are non-significant (see table 3.11).
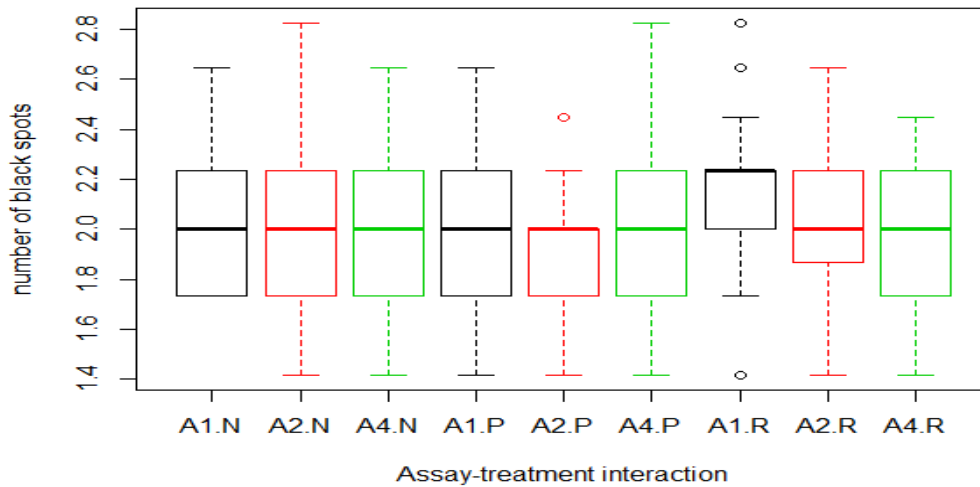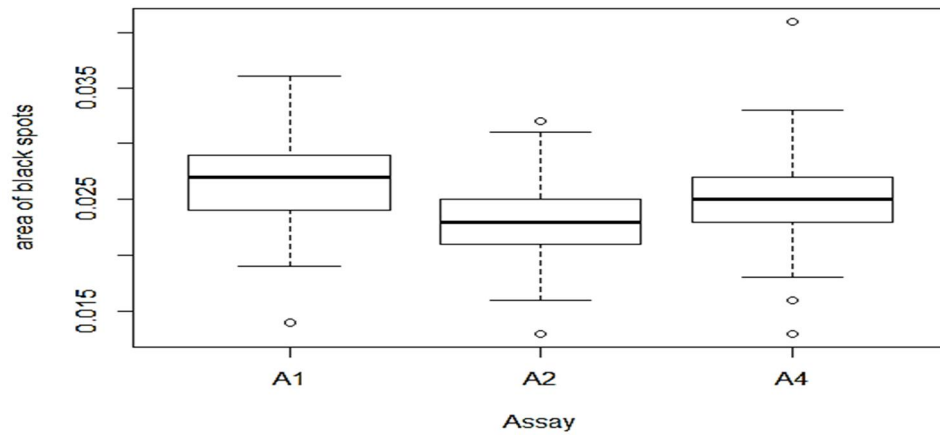

Figure 3.13: A comparison of the numbers of black spots between the positive (P), the negative (N) and the random (R) treatments of the assays 1, 2 and 4 (A1, A2 and A4). Groups belong to the assay dataset. Spot numbers are square-root transformed.
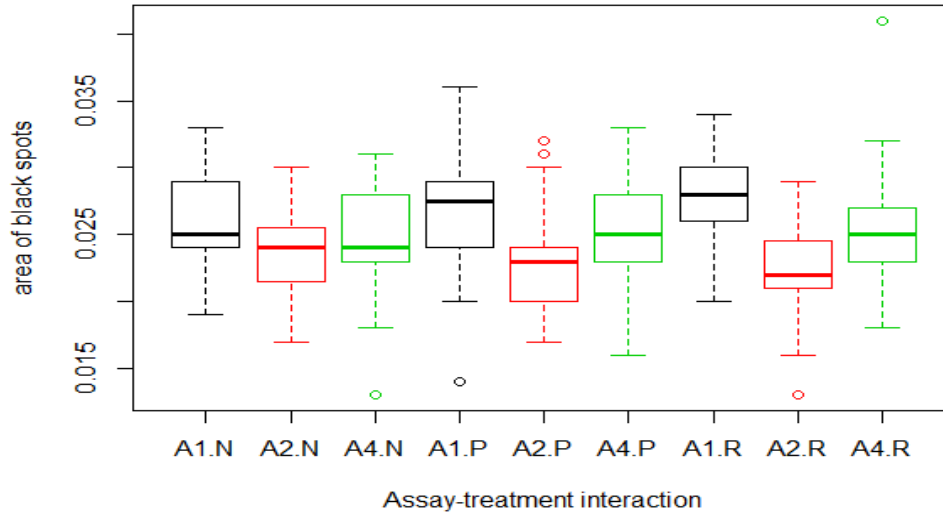
Table 3.11: Variation in the numbers of black spots between the groups of the assay dataset. The number of black spots represents the dependent variable, while the assay and the treatment represent the independent variables. Spot numbers are square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: A = assay, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Difference	Std.error	Df	t-value	P-value
A1, A2	-0.097	0.044	331	-2.207	0.0280 *
A2, A4	-0.025	0.040	331	-0.629	0.5300
A1, A4	-0.122	0.038	331	-3.239	0.0013 **
A 1-P, A 2-P	-0.095	0.069	109	-1.373	0.1726
A 2-P, A 4-P	0.067	0.063	109	1.072	0.2861
A 1-P, A 4-P	-0.028	0.060	109	-0.465	0.6430
A 1-N, A 2-N	-0.053	0.085	106	-0.623	0.5345
A 2-N, A 4-N	-0.064	0.076	106	-0.840	0.4030
A 1-N, A 4-N	-0.117	0.073	106	-1.604	0.1118
A 1-R, A 2-R	-0.132	0.073	112	-1.800	0.0745 -
A 2-R, A 4-R	-0.082	0.067	112	-1.237	0.2185
A 1-R, A 4-R	-0.214	0.062	112	-3.467	0.0007 ***
A 4-P, A 4-N	-0.051	0.101	6	-0.504	0.6324
A 4-P, A 4-R	-0.041	0.100	6	-0.404	0.6999
A 4-N, A 4-R	0.010	0.100	6	0.101	0.9230

The relative area of black spots fluctuated during the experiment, decreasing by 26.3% between the assays 1 and 2, then increasing by 13.4% between the assays 2 and 4 (figure 3.14). The anova test shows the changes across the assays to be statistically significant (ANOVA: F 2, 327 = 3.694. P = 0.03). Pairwise comparisons indicate that the differences between the assays 1 and 2, as well as between the assays 2 and 4, are highly significant. The treatment had no significant effect on the black area (ANOVA: F 2, 6 = 1.283. P = 0.34). In addition, there was no significant interaction between the assay and the treatment (ANOVA: F 4, 327 = 0.880. P = 0.48). All the treatment lines showed a similar direction of change, decreasing between the assays 1 and 2, then increasing between the assays 2 and 4, with the net change being a decrease in the black area over time. However, the treatment lines differed in the rate and the statistical significance of the change. For the random line, the black area decreased by 34.1% between the assays 1 and 2, then increased by 18.4% between the assays 2 and 4; both changes were highly significant. For the positive line, the black area decreased by 23.2%, then increased by 11.2%; the changes were highly significant and non-significant, respectively. For the negative line, the black area decreased by 19.4%, then increased by 9.1%; similar to the positive line, the difference between the assays 1 and 2 was highly significant, while the difference between the assays 2 and 4 was non-significant (figure 3.15). Within assay 4, the treatments showed no significant difference in the area of black spots (ANOVA: F 2, 6 = 0.020. P = 0.98). Moreover, pairwise comparisons between the positive and the negative, the positive and the random, and the negative and the random treatments did not detect any significant difference in the black area (see table 3.12).

Figure 3.14: A comparison of the relative areas of black spots between the assays 1, 2 and 4 (A1, A2, and A4). Groups belong to the assay dataset. Relative areas are arcsine square-root transformed.

Figure 3.15: A comparison of the relative areas of black spots between the positive (P), the negative (N) and the random (R) treatments of the assays 1, 2 and 4 (A1, A2 and A4). Groups belong to the assay dataset. Relative areas are arcsine square-root transformed.

Table 3.12: Variation in the relative areas of black spots between the groups of the assay dataset. The relative area of black spots represents the dependent variable, while the assay and the treatment represent the independent variables. Relative areas are arcsine square-root transformed. All statistical comparisons were performed as mixed-effects models, by using the library (nlme) in R, with 'population' specified as the random effect. The illustrated values represent the outputs of the summary function. The red borders separate the outputs of different models. Abbreviations: A = assay, P = positive, N = negative, R = random, Std.error = standard error, Df = degree of freedom. Significance levels: '***' = $p < 0.001$, '**' = $p < 0.01$, '*' = $p < 0.05$, '-' = $p \leq 0.10$.

Comparison	Difference	Std.error	Df	t-value	P-value
A1, A2	-0.004	0.0006	331	-6.749	< 0.0001 ***
A2, A4	0.002	0.0005	331	3.430	0.0007 ***
A1, A4	-0.002	0.0005	331	-4.255	< 0.0001 ***
A 1-P, A 2-P	-0.003	0.001	109	-3.184	0.0019 **
A 2-P, A 4-P	0.001	0.001	109	1.558	0.1221
A 1-P, A 4-P	-0.002	0.001	109	-2.045	0.0433 *
A 1-N, A 2-N	-0.003	0.001	106	-2.831	0.0055 **
A 2-N, A 4-N	0.001	0.001	106	1.368	0.1743
A 1-N, A 4-N	-0.002	0.001	106	-1.866	0.0648 -
A 1-R, A 2-R	-0.005	0.001	112	-5.494	< 0.0001 ***
A 2-R, A 4-R	0.002	0.001	112	2.824	0.0056 **
A 1-R, A 4-R	-0.003	0.001	112	-3.471	0.0007 ***
A 4-P, A 4-N	-0.00003	0.001	6	0.027	0.9791
A 4-P, A 4-R	0.0002	0.001	6	0.158	0.8795
A 4-N, A 4-R	0.0002	0.001	6	0.186	0.8588

3.1.3.2. Multivariate comparisons

Overall, male color (estimated from the numbers and the relative areas of orange and black spots) showed a highly significant change between the assays. Moreover, there was a highly significant interaction between the assay and the treatment. The treatment had no significant effect on coloration. To further investigate the color changes that occurred between the assays, pairwise comparisons were performed between the assays 1 and 2, as well as between the assays 2 and 4; both comparisons revealed highly significant differences. To investigate the potential variation among the treatment lines, the rate of color change was assessed independently from the positive, the negative and the random lines; all the treatment lines showed highly significant changes across the assays. Within assay 4, the treatments did not differ significantly in coloration. To further investigate the potential variation between the treatments, pairwise comparisons were conducted between the positive and the negative, the positive and the random, and the negative and the random treatments; no significant difference was found between any pair of treatments (see table 3.13).

Table 3.13: Variation in the male coloration between the groups of the assay dataset. The numbers and the relative areas of orange and black spots represent the dependent variables, while the assay, the treatment and the population represent the independent variables. The values for the population variable are not shown. Spot numbers are square-root transformed, and relative areas are arcsine square-root transformed. Statistical analysis was performed by using Wilks' lambda test, specified from the 'manova' and 'summary' functions in R. Abbreviations: H = harvest, T = treatment, H:T = harvest-treatment interaction, P = positive, N = negative, R = random, Wilks = Wilks' lambda values, approx F-value = approximate F-values, num Df = numerator's degrees of freedom, den Df = denominator's degrees of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Df	Wilks	num Df	den Df	approx F-value	P-value
A	2	0.792	8	624	9.664	< 0.0001 ***
T	2	0.980	8	624	0.775	0.6247
A:T	4	0.878	16	953.81	2.602	0.0005 ***
A1, A2	1	0.657	4	69	9.019	< 0.0001 ***
A2, A4	1	0.874	4	100	3.612	0.0086 **
A1, A4	1	0.891	4	107	3.289	0.0139 *
A 1-P, A 2-P, A 4-P	2	0.680	8	204	5.432	< 0.0001 ***
A 1-N, A 2-N, A 4-N	2	0.804	8	198	2.855	0.0050 **
A 1-R, A 2-R, A 4-R	2	0.625	8	210	6.951	< 0.0001 ***
A 4-P, A 4-N, A 4-R	2	0.777	8	300	0.960	0.623
A 4-P, A 4-N	1	0.848	4	44	1.970	0.1157
A 4-P, A 4-R	1	0.948	4	44	0.603	0.6628
A 4-N, A 4-R	1	0.960	4	44	0.458	0.7661

3.1.4 Trends in color data: a synthesis

In both the harvest and the assay datasets, a number of color differences were observed and corroborated statistically, across the generations as well as between the treatments. Some of the group differences that were observed in the harvest dataset were not mirrored in the assay dataset, indicating that phenotypic plasticity has contributed to the variation among the groups. For example, the general decrease in orange coloration between the harvests 4 and 28, which was evident from the spot numbers as well as the relative areas (figures 3.1 and 3.3), was not accompanied by a similar change between the assays 2 and 4.

Some of the observed changes were replicated in both the harvest and the assay datasets, indicating that their underlying causes are most likely genetic, and, consequently, that they represent genuine color evolution. The first case is the decrease in orange coloration within the positive line; a trend observed between the harvests 4 and 28 and also between the assays 2 and 4 (these are approximately simultaneous periods). The trend was evident from the numbers of orange spots and also from the relative areas of orange (figures 3.2, 3.5, 3.9 and 3.12). All the described changes were statistically significant (tables 3.4, 3.5, 3.9 and 3.10). Similar changes were observed for the negative and the random treatments in the harvest dataset, but not in the assay dataset (see the above-cited figures and tables); this suggests that the decrease in orange coloration was genetic only for the positive line. The second case is the increase in the relative area of black coloration in the positive and random treatments; a trend observed between the harvests 4 and 28 and also between the assays 2 and 4 (figures 3.7 and 3.15). The changes were evident from the percentage

differences as well as the graphs, however, they were significant only for the positive line in the harvest dataset and the random line in the assay dataset (tables 3.7 and 3.12). No similar changes were observed for the negative line in the harvest or the assay datasets (see the above-cited tables and figures). We conclude that the positive line has evolved less orange coloration during the period of the experiment; we also suggest that an increase in black coloration might have evolved in the positive and the random lines.

Because it contains data from three points in time (assay 1, assay 2 and assays 4), the assay dataset enables the investigation of the temporal consistency of color change through time. In 10 out of 12 cases, the color metrics fluctuated across the assays, rather than increasing or decreasing consistently. These include the positive, the negative and the random lines for the number of orange spots, the positive, the negative and the random lines for the relative area of orange spots, the positive, the negative and the random lines for the relative area of black spots, and the positive line for the number of black spots (tables 3.9, 3.10, 3.11 and 3.12). Note that this list is compiled by considering only the sign of the change between the assays; some of these differences were non-significant, however, that 10 of the observed changes showed fluctuation compared to only 2 that showed consistent change is a pattern that is unlikely to come by chance. Because the color changes mostly changed their direction through time, the changes across the whole experiment were mostly of less magnitude and less significance than the changes that took place during a stage in the experiment. For example, in the positive line, the numbers of orange spots changed significantly between the assays 1 and 2, and between the assays 2 and 4, but the overall change was non-significant (table 3.9). In the negative line, the area of orange spots changed significantly between the assays 2 and 4, but the overall change was marginally non-significant (table 3.10). In the random line, the numbers of orange spots changed highly significantly between the assays 2 and 4, while the overall change was significant (table 3.9).

Overall, the multivariate analyses showed that the males have experienced changes in coloration during the experiment; the trend was highly significant for both the harvest and the assay datasets. The differences in the male coloration between the treatment lines were highly significant for the harvest dataset but non-significant for the assay dataset (tables 3.8 and 3.13), indicating that the differences between the lines were less robust than the differences between the generations. The harvest-treatment interaction was marginally non-significant, while the assay-treatment interaction was highly significant, indicating that the treatment had an effect on the temporal changes in coloration, especially in the assay experiment. Consistent with this, pairwise comparisons between the generations within each treatment line showed that the lines varied in their rate of change (as estimated from the P values for the differences between the generations); the positive and the random lines showed greater changes than the negative line in both the harvest and the assay datasets. The treatments of the most recent generation (harvest 28, assay 4) showed no consistent differences, indicating that the color changes did not lead to stable phenotypic divergence (tables 3.8 and 3.13).

3.2   Shape changes

3.2.1   Measurement error

The photographs of the males from harvest 28 were subjected to Procrustes anova test, in order to quantify the measurement error. A statistical comparison was performed between biological variation (i.e. the effects of individual, side and individual-by-side interaction) and error variation (i.e. the combined effects of imaging error and landmark digitizing error). Comparisons were performed separately for the size and the shape components of variation. The effect of the individual was significant and highly significant for centroid size and shape, respectively. The side had no significant effect on size or shape, indicating that the left and the right sides were not systematically different from each other (i.e. no directional asymmetry). For both size and shape, the effect of the individual-by-treatment interaction was highly significant relative to the combined effects of imaging and digitizing errors, indicating that fluctuating asymmetry had a greater effect on size and shape variation than measurement error (table 3.14). Since fluctuating asymmetry is a subtle biological effect, the test shows that error variation had a negligible effect compared to biological variation.

Table 3.14: The outputs of the procrustes anova test for measurement error, estimated for the variation in the centroid size and the shape. The inputs were landmark coordinates from the males of harvest 28. Error represents the combined effects of imprecision in photography and landmark placement. The analysis was performed on a dataset containing four photographs for each individual, two per side, by using the program MorphoJ. Abbreviations: MS = mean of squares, Df = degree of freedom. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Type of variation	Effect	MS	Df	F-value	P-value
Centroid size	Individual	182955.42	119	1.49	0.0159 *
Centroid size	Side	31395.12	1	0.25	0.6146
Centroid size	Individual * side	123193.13	119	4698.44	< 0.0001 ***
Centroid size	Error	26.22	238	-	-
Shape	Individual	0.00051	1904	1.29	< 0.0001 ***
Shape	Side	0.00047	16	1.18	0.2761
Shape	Individual * side	0.00040	1904	94.40	< 0.0001 ***
Shape	Error	0.000004	3808	-	-

3.2.2   Harvest populations

3.2.2.1 Shape variation across harvests

By using Discriminant function analysis (DFA), pairwise comparisons were performed between the groups of harvest 4 and the equivalent groups of harvest 28. Four separate comparisons were performed in order to assess the shape variation between the harvests. In the first comparison, groups were defined in terms of harvest only, therefore, all the individuals in each harvest were grouped together, regardless to the treatment they were subjected to. The DFA test showed that

61

the harvests 4 and 28 differ highly significantly with respect to shape (table 3.15). The shape changes that correspond with this difference are illustrated in figure 3.16. Compared with harvest 4 males, the males of harvest 28 have considerably narrower anal fin, and slightly narrower dorsal fin. In addition, the dorsal and anal fins of harvest 28 males were shifted anteriorly relative to the anteroposterior axis of the body; this relative change was reflected in the anterior shift of the dorsal and the anal fins as well as the posterior shift along the anteroposterior axis. Because of this, the caudal peduncle became more elongated relative to the anterior part of the body. Moreover, the body shape became narrower, because the dorsal and the anal fins shifted closer to each other along the dorsoventral axis. The ventral shift of the dorsal fin was greater than the dorsal shift of the anal fin, as a result, the dorsal side of the body became narrower relative to the ventral side. The figures 3.17 and 3.18 (of discriminant scores and cross-validation scores, respectively) show that the two harvests are well separated by the above-described shape differences. Table 3.16 (of misclassification rates) shows that the individuals can be reliably assigned to each harvest based on their shape features.

Table 3.15: The outputs of discriminant function analysis (DFA) for the shape variation between the harvests in the harvest dataset, performed separately for the whole harvests and for each treatment line within each harvest. The inputs were the residuals of the regression of procrustes coordinates on centroid size, representing the allometry-free shape variation. The analysis was performed on a dataset containing one, left-side photograph for each individual, by using the program MorphoJ. The p values were calculated from permutation tests with 10,000 rounds. Abbreviations: H = harvest, P = positive, N = negative, R = random. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

| Comparison	Mahalanobis distance	p-value for Mahalanobis distance	Procrustes distance	P-value for Procrustes distance
H 4, H 28	5.2632	< 0.0001 ***	0.0545	< 0.0001 ***
H 4-P, H 28-P	5.3432	< 0.0001 ***	0.0516	< 0.0001 ***
H 4-N, H 28-N	5.3483	< 0.0001 ***	0.0638	< 0.0001 ***
H 4-R, H 28-R	6.6503	< 0.0001 ***	0.0515	< 0.0001 ***
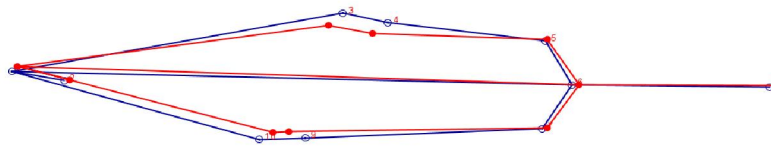
04 -- 28

Figure 3.16: Wireframe representation of the shape changes between the harvests, according to discriminant function analysis (DFA). The blue color represents the starting shape, which describes harvest 4. The red color represents the target shape, which describes harvest 28. All the treatments in each harvest were combined. Scale factor = 1.
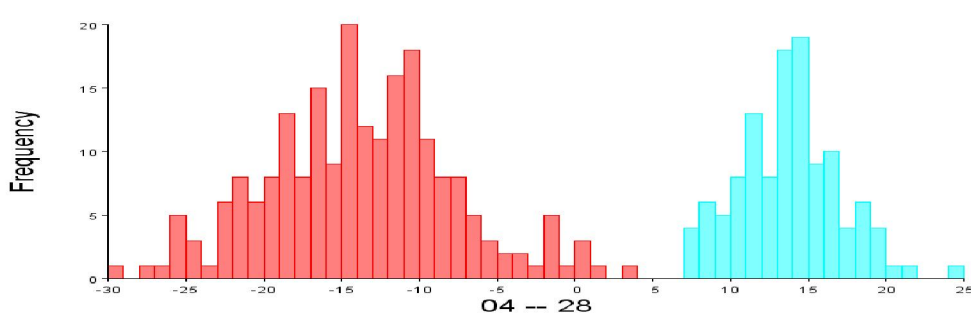
Figure 3.17: The discriminant scores from the Discriminant function analysis (DFA) for the comparison between the harvests. The low levels of the discriminant scores (on the left) represent harvest 4. The high levels of the discriminant scores (on the right) represent harvest 28. All the treatments in each harvest were combined. The color key is on the upper right corner.
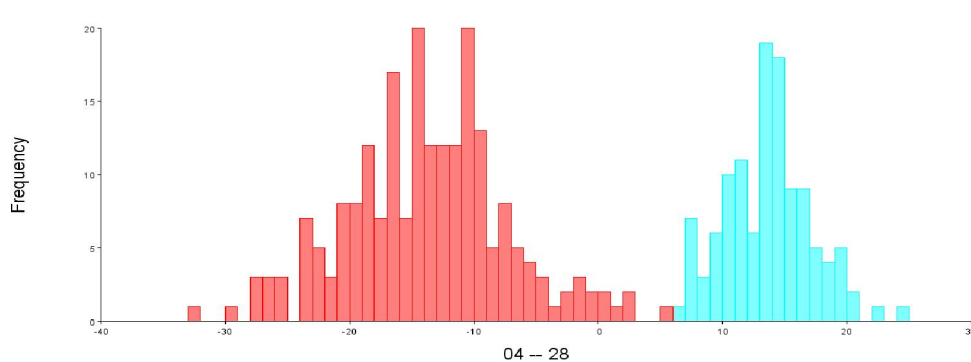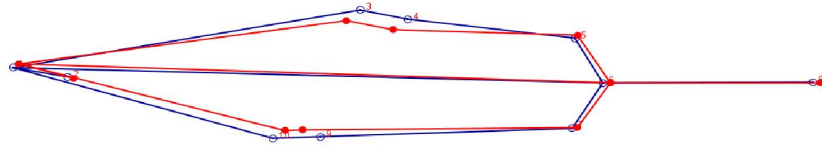

Figure 3.18: The cross-validation scores from the Discriminant function analysis (DFA) for the comparison between the harvests. The low levels of the discriminant scores (on the left) represent harvest 4. The high levels of the discriminant scores (on the right) represent harvest 28. All the treatments in each harvest were combined. The color key is on the upper right corner.

Table 3.16: The reliability of the discriminant function analysis for the harvest dataset, as indicated by the rates of misclassification, based on the discriminant and the cross-validation functions. Each value represents the number of misclassified individuals, divided by the total numbers of individuals in the group. Abbreviations: H = harvest, P = positive, N = negative, R = random.

Comparison	Discriminant function		Cross-validation function	
	H4 as H28	**H28 as H4**	**H4 as H28**	**H28 as H4**
H 4, H 28	4/213	0/117	6/213	0/117
H 4-P, H 28-P	0/98	0/48	4/98	0/48
H 4-N, H 28-N	2/81	0/38	3/81	0/38
H 4-R, H 28-R	0/34	0/31	1/34	0/31

For the rest of the pairwise comparisons, groups were defined in terms of both the harvest and the treatment. Therefore, comparisons were performed (by using DFA) between the harvests 4-positive and 28-positive, 4-negative and 28-negative, and between 4-random and 28-random. All the lines showed highly significant changes between the harvests (table 3.15). All the lines showed similar shape changes to each other, and to those observed in the comparison between total harvests. The males of harvest 28, from all the lines, had narrower body shapes, elongated caudal peduncles, considerably narrower anal fins and slightly narrower dorsal fins, than their counterparts from harvest 4 (figure 3.19). The harvests were well separated by their shape features in all the lines (figures 3.20 and 3.21). For all the lines, the misclassification rates were very low, under both the discriminant and the cross-validation functions (table 3.16).

a)

04.P -- 28.P

b)
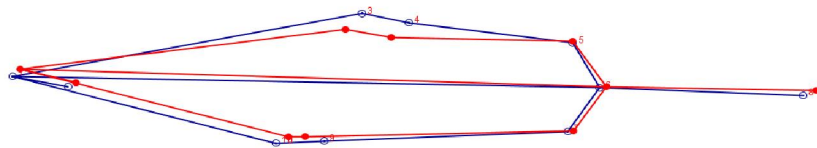
04.N -- 28.N

c)
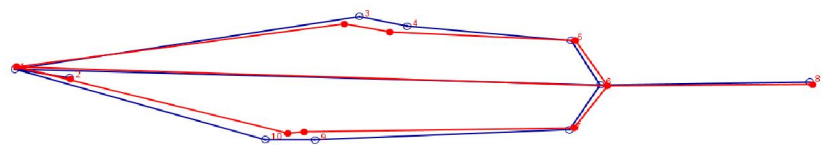
04.R -- 28.R

Figure 3.19: Wireframe representations of the shape changes between the harvests that occurred within: **a)** the positive line (P), **b)** the negative line (N), and **c)** the random line (R). Analysis was performed by using discriminant function analysis (DFA). The blue color represents the starting shape, which describes the treatment of harvest 4. The red color represents the target shape, which describes the treatment of harvest 28. Scale factor = 1.

a)
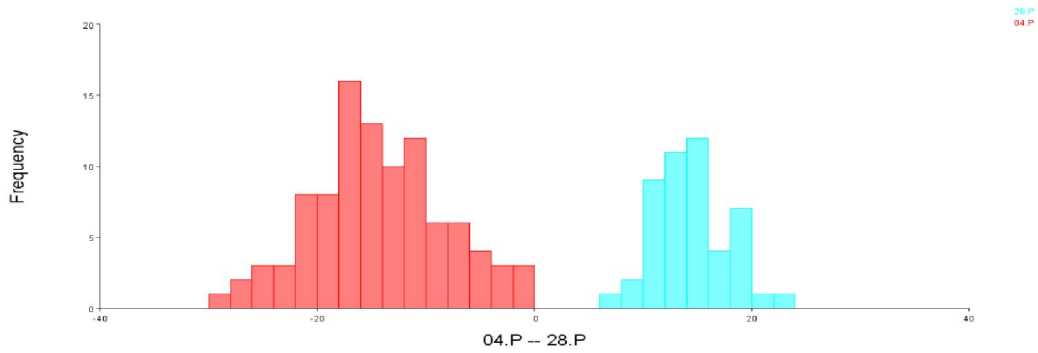
b)
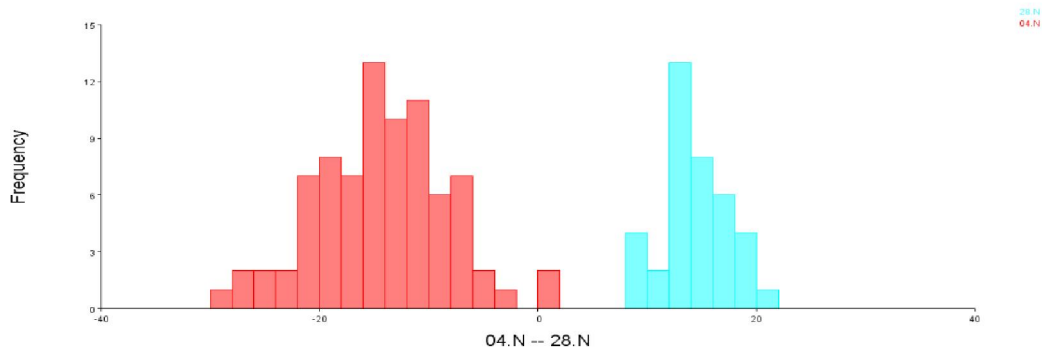
c)
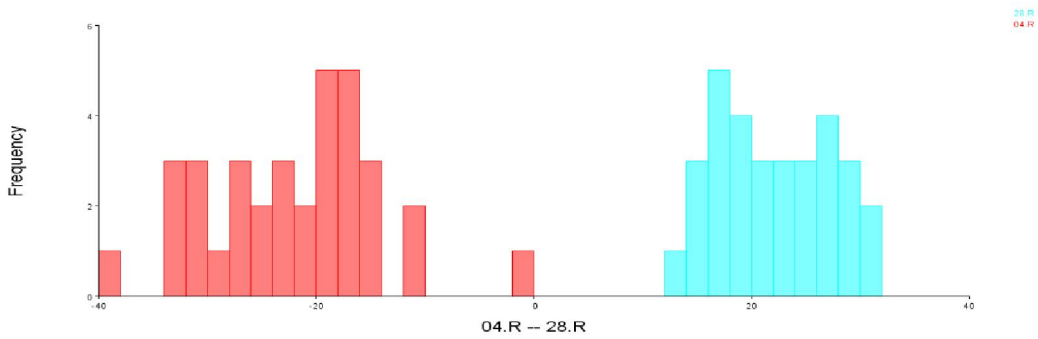
Figure 3.20: The discriminant scores from the Discriminant function analysis (DFA) for the comparison between the harvests of: **a)** the positive line (P), **b)** the negative line (N), and **c)** the random line (R). The low levels of the discriminant scores (on the left) represent the treatment of harvest 4. The high levels of the discriminant scores (on the right) represent the treatment of harvest 28. The color key is on the upper right corner.

a)
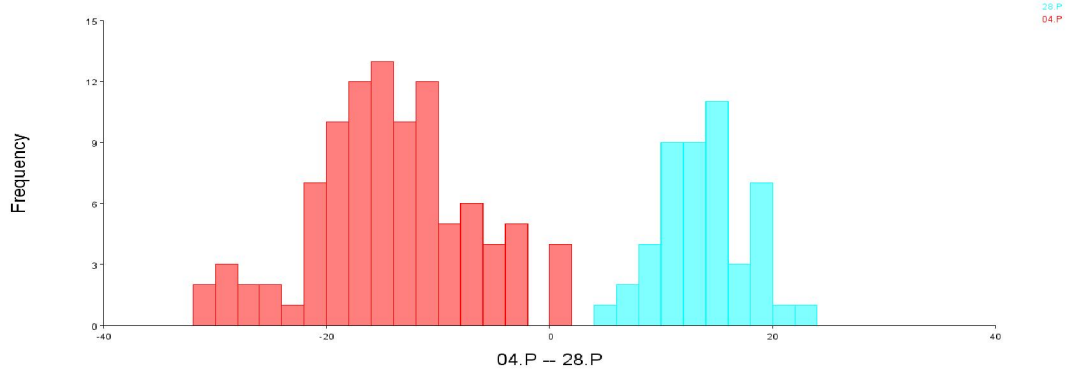
b)
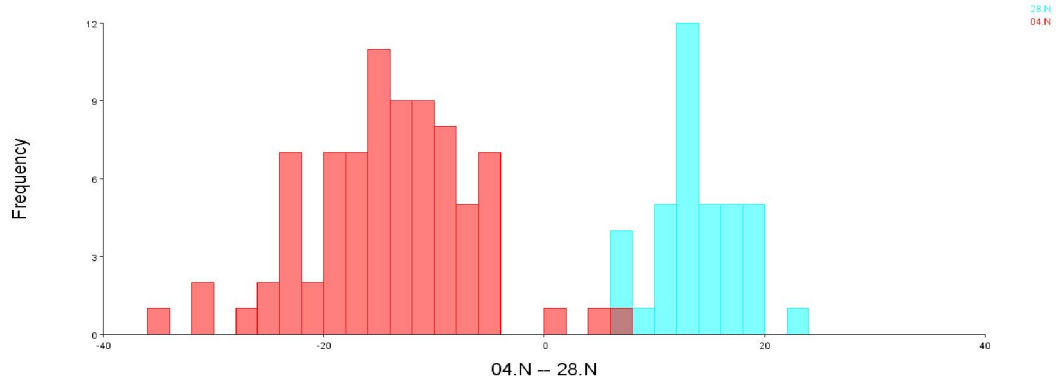
c)

Figure 3.21: The cross-validation scores from the Discriminant function analysis (DFA) for the comparison between the harvests of: **a)** the positive line (P), **b)** the negative line (N), and **c)** the random line (R). The low levels of the discriminant scores (on the left) represent the treatment of harvest 4. The high levels of the discriminant scores (on the right) represent the treatment of harvest 28. The color key is on the upper right corner.
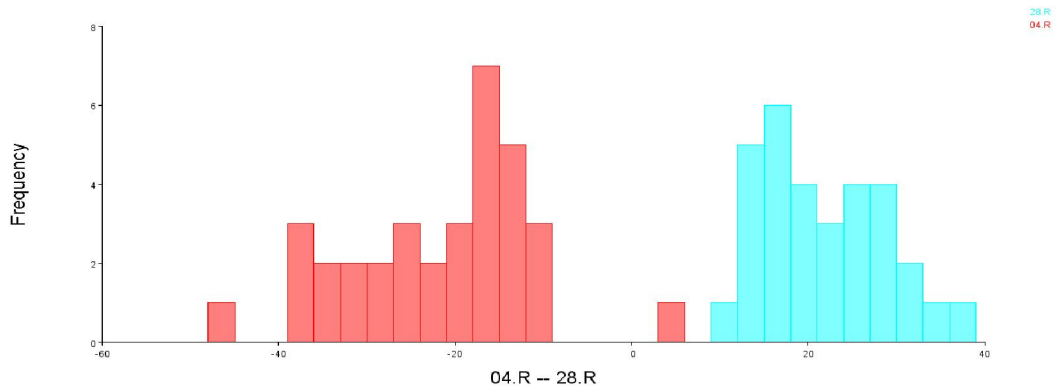
3.2.2.1 Shape variation between treatments

In order to investigate the shape variation between the positive, the negative and the random treatments of the harvest dataset, canonical variate analysis (CVA) was applied. Two separate CVA tests were performed, the first for harvest 4 and the second for harvest 28. Regarding harvest 4; the differences between the positive and the negative treatments, as well as that between the random and the negative treatments, were highly significant for both Mahalanobis and Procrustes distances. The difference between the positive and the random treatments was significant only for Mahalanobis distance (table 3.17). Figure 3.22 represents the ordination plot of the canonical scores; it shows that there is a considerable overlap between all the treatments. However, and consistently with the above p values; the negative treatment appears more separated from the positive and the random treatments than the two later treatments are separated from each other.

Table 3.17: The outputs of canonical variate analysis (CVA) for the shape variation among the treatments of the harvest dataset. The inputs were the residuals of the regression of procrustes coordinates on centroid size, representing the allometry-free shape variation. The analyses were performed separately for each harvest, but jointly for the pairwise comparisons within each harvest. The analysis of harvest 4 was performed on a dataset containing one, left-side photograph for each individual, while the analysis of harvest 28 was performed on an averaged dataset, combining data from four photographs for each individual, two per side. The p values were calculated from permutation tests with 10,000 rounds. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

| Comparison	Mahalanobis distance	P-value for Mahalanobis distance	Procrustes distance	P-value for Procrustes distance
H 4-P, H 4-N	1.3044	< 0.0001 ***	0.0207	0.0006 **
H 4-P, H 4-R	1.0736	0.0108 *	0.0138	0.2710
H 4-N, H 4-R	1.5262	< 0.0001 ***	0.0311	0.0006 **
H 28-P, H 28-N	1.1311	0.0238 *	0.0100	0.0625 -
H 28-P, H 28-R	1.6861	< 0.0001 ***	0.0147	0.0013 **
H 28-N, H 28-R	1.4297	0.0009 ***	0.0075	0.3735
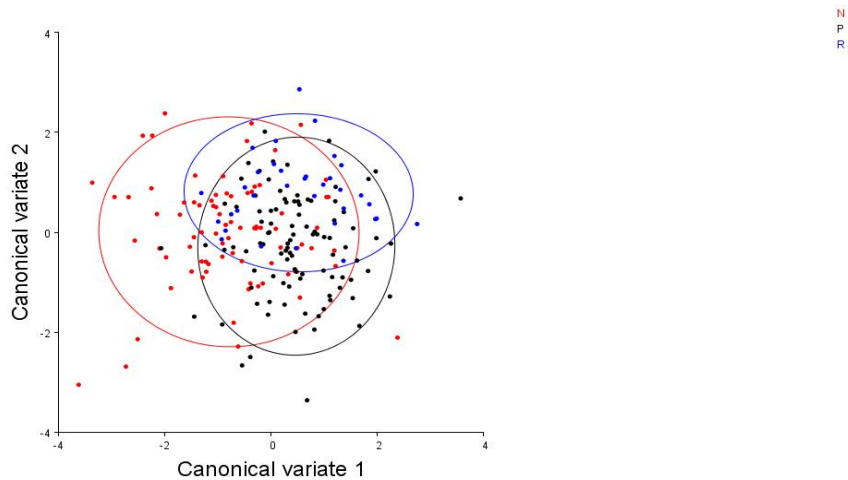
Figure 3.22: The ordination plot of canonical variate analysis (CVA) for the positive, the negative and the random treatments of harvest 4. Each dot represents one specimen. Each treatment is represented by a different color; the color key is at the upper right corner. Confidence ellipses were set at a probability of 0.9.

The difference between the negative treatment and the two other treatments is approximately described by the first canonical variate CV1 (which account for 74% of the total variance) (figure 3.23a); it shows an elongation in the caudal fin accompanied by a shortening in the standard length and the snout-eye distance (i.e. the distance between the tip of the snout and the center of the eye orbit). In addition, there is a slight narrowing in the base of the anal fin. There is a ventral shift in the position of the dorsal fin; Because of this shift, the overall body shape became narrower, and the dorsal side of the body became narrower relative to the ventral side. The base of the anal fin is shifted anteriorly, while the dorsal side of the caudal fin base is shifted posteriorly, as a result, the caudal peduncle became more elongated at the dorsal side. Note that the shape changes described above are interpreted as representing the difference between the negative treatment (the starting shape) and both the positive and the random treatments (the target shape). Note, also, that most of these differences are similar to those observed between the harvests.

There is extensive overlap between the positive and the random treatments along the vertical axis (CV2) of the ordination plot (figure 3.22), however, the individuals of the random treatment are mostly found at the higher values of CV2, while the individuals of the positive treatment are mostly found at the lower values of CV2, therefore, the shape changes described by CV2 can be interpreted as differences between the positive treatment (the starting shape) and the random treatment (the target shape). The second canonical variate (figure 3.23b) accounts for 26% of the total variance; it shows a widening at the bases of the dorsal, anal, and caudal fins. It also shows a ventral shift for both the dorsal and the anal fins; as a result, the dorsal side of the body became narrower, while the ventral side became deeper.

a)
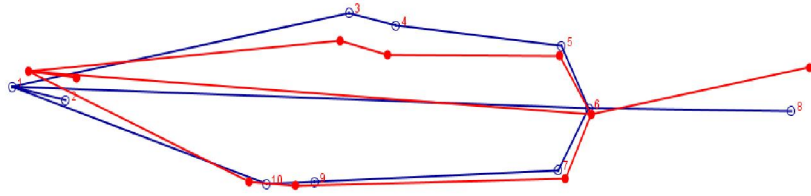
CV1

b)
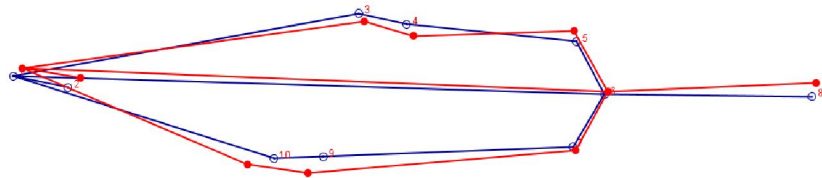
CV2

Figure 3.23: Wireframe representations of the shape changes that are associated with: **a)** the first canonical variate (CV1), and **b)** the second canonical variate (CV2), of canonical variate analysis (CVA) for the treatments of harvest 4. The blue color represents the starting shape, while the red color represents the target shape. The shape changes were exaggerated by a factor of 5.

Regarding harvest 28, there was a highly significant difference between the positive and the random treatments for both Mahalanobis and Procrustes distances. The difference between the positive and the negative treatments was significant for Mahalanobis distance and marginally non-significant for Procrustes distance. The difference between the negative and the random treatments was highly significant for Mahalanobis distance but non-significant for procrustes distance (table 3.17). The ordination plot of the canonical scores (figure 3.24) shows extensive overlap between the treatments along both the horizontal and the vertical axis, however, the individuals of the random treatment are mostly found at the higher levels of CV1, while the individuals of the two other treatments are mostly found at the lower levels of CV1, therefore, the shape changes described by CV1 can be interpreted as differences between the positive and the negative treatments on one side (the starting shape) and the random treatment on the other side (the target shape).
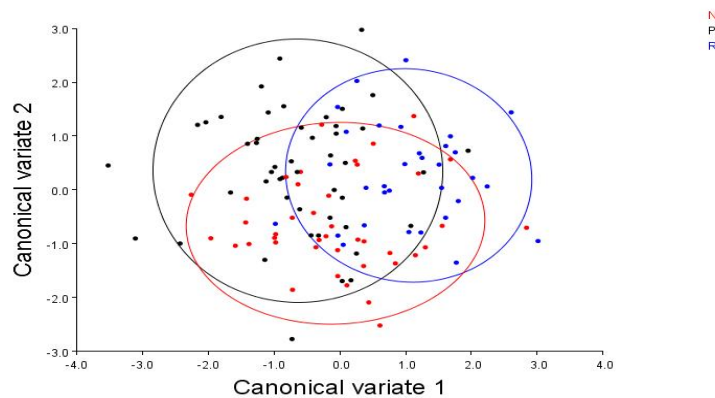
Figure 3.24: The ordination plot of canonical variate analysis (CVA) for the positive, the negative and the random treatments of harvest 28. Each dot represents one specimen. Each treatment is represented by a different color; the color key is at the upper right corner. Confidence ellipses were set at a probability of 0.9.

The first canonical variate (CV1) accounted for 69% of the total variance (figure 3.25a); it describes an upward shift along the anteroposterior axis, shown by the snout-eye distance, the standard length and the caudal fin length. This shift is accompanied by a downward shift in the position of the anal fin. In addition, the anal fin is shifted toward the ventral side, resulting in an increase in the area of the ventral side relative to that of the dorsal side. Finally, there is a shortening in the caudal fin length. The second canonical variate accounts for 31% of the total variance (figure 3.25b); it can be interpreted as representing the differences between the negative treatment (the starting shape) and the positive treatment (the target shape). It shows an upward shift in the position of the anal fin, in addition to a dorsal shift in the position of the dorsal fin.

a)

CV1

b)
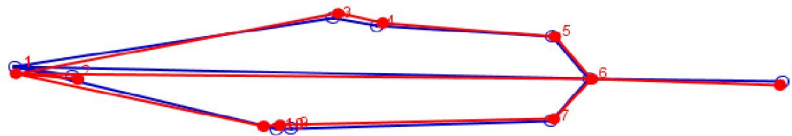
CV2

Figure 3.25: Wireframe representations of the shape changes that are associated with: **a)** the first canonical variate (CV1), and **b)** the second canonical variate (CV2), of canonical variate analysis (CVA) for the treatments of harvest 28. The blue color represents the starting shape, while the red color represents the target shape. The shape changes were exaggerated by a factor of 5.
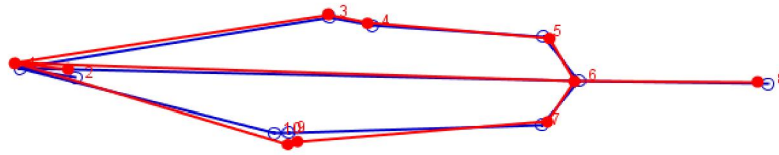
In order to compare the relative magnitudes of the differences between the treatments of harvest 4 to that between the treatments of harvest 28, an ordination plot was constructed for all the groups in the harvest dataset, including the positive, the negative and the random treatments of both harvests (figure 3.26). The unified plot shows that the separation between the harvests is substantially greater than the separation between the treatments within each harvest; the first canonical variate (CV1), which described the separation between harvests, accounted for 93% of the total variation among the groups. In addition, the individuals of harvest 4 are spread more widely along both the horizontal and the vertical axes compared to the individuals of harvest 28, indicating that a considerable reduction in the shape variation between the individuals has occurred during the period between the two harvests. The plot also shows that the treatments of harvest 28 overlap more extensively than the treatments of harvest 4, indicating that the shape variation between the treatments has decreased during the period between the two harvests.
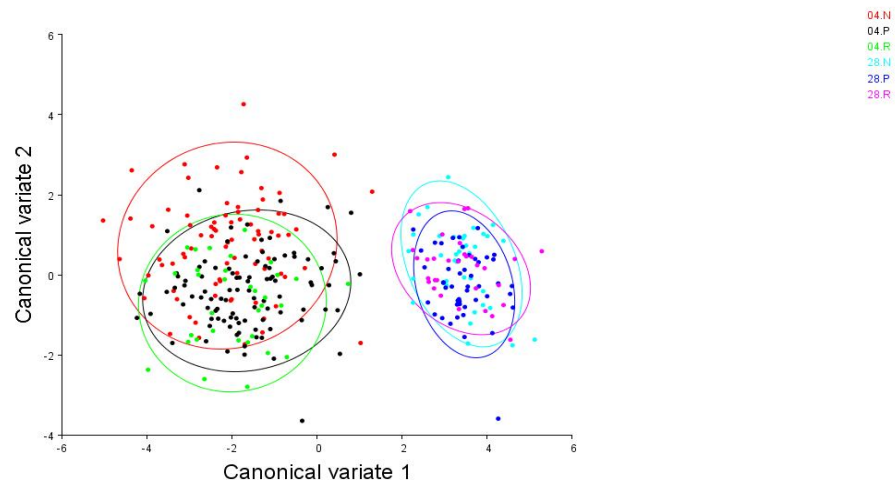
Figure 3.26: The ordination plot of canonical variate analysis (CVA) for the positive, the negative and the random treatments of the harvests 4 and 28. The groups are defined by the harvest and the treatment. Each dot represents one specimen. Each group is represented by a different color; the color key is at the upper right corner. Confidence ellipses were set at a probability of 0.9.

3.2.3 Assay populations

3.2.3.1 Shape variation across assays

By using discriminant function analysis (DFA), pairwise comparisons were performed between the assays 2 and 4 (regardless to the treatment), then between each treatment in assay 2 and its equivalent in assay 4. For the first comparison, the DFA test detected highly significant difference between the assays, estimated from both Mahalanobis and Procrustes distances (table 3.18). The shape changes that correspond with this statistical difference are illustrated in figure 3.27. Overall, the males of assay 4 have narrow, elongated body shapes compared to the males of assay 2. The

elongation of the body occurred along the snout-eye distance, the standard length and the caudal fin length. The narrowing of the body occurred because the dorsal and the anal fins moved closer to each other along the dorsoventral axis (i.e. the dorsal fin shifted ventrally, while the anal fin shifted dorsally). The dorsal fin showed greater shift along the dorsoventral axis than did the anal fin, leading to a decrease in the area of the dorsal side relative to that of the ventral side and, consequently, to an apparent dorsal shift for the anteroposterior axis. The basis of the dorsal and the anal fins became narrower, while the base of the caudal fin became wider. The dorsal and the anal fins show posterior shifts, accompanied by an anterior shift for the caudal fin, as a result, the caudal peduncle became shorter relative to the anterior part of the body. Note that the above-described differences represent the shape changes that occurred between assay 2 (the starting shape) and assay 4 (the target shape). The figures 3.28 and 3.29 show that, despite some overlap, the assays 2 and 4 are well separated by their shape features. The misclassification rates indicate that the shape was mostly a reliable indicator for assigning individuals to assays, however, few individuals from both groups were misclassified as members of the other group (table 3.19).

Table 3.18: The outputs of discriminant function analysis (DFA) for the shape variation across the assays in the assay dataset, performed separately for the whole assays and for each treatment line. The inputs were the residuals of the regression of procrustes coordinates on centroid size, representing the allometry-free shape variation. The analysis was performed on a dataset containing one, left-side photograph for each individual, by using the program MorphoJ. The p values were calculated from permutation tests with 10,000 rounds. Significance levels: '***' = p < 0.001, '**' = p < 0.01, '*' = p < 0.05, '-' = p ≤ 0.10.

Comparison	Mahalanobis distance	P-value for Mahalanobis distance	Procrustes distance	P-value for Procrustes distance
A 2, A 4	4.4376	< 0.0001 ***	0.0592	< 0.0001 ***
A 2-P, A 4-P	4.5991	< 0.0001 ***	0.0622	0.001 **
A 2-N, A 4-N	5.9047	< 0.0001 ***	0.0687	< 0.0001 ***
A 2-R, A 4-R	4.6392	< 0.0001 ***	0.0621	< 0.0001 ***
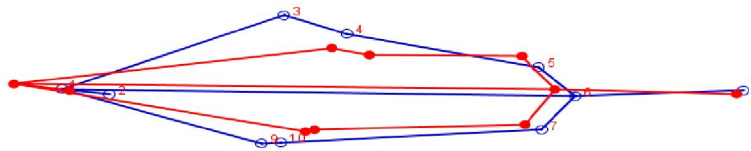
A2 -- A4

Figure 3.27: Wireframe representation of the shape changes between the assays, according to discriminant function analysis (DFA). The blue color represents the starting shape, which describes assays 2. The red color represents the target shape, which describes assay 4. All the treatments in each assay were combined. The shape changes were exaggerated by a factor of 3.
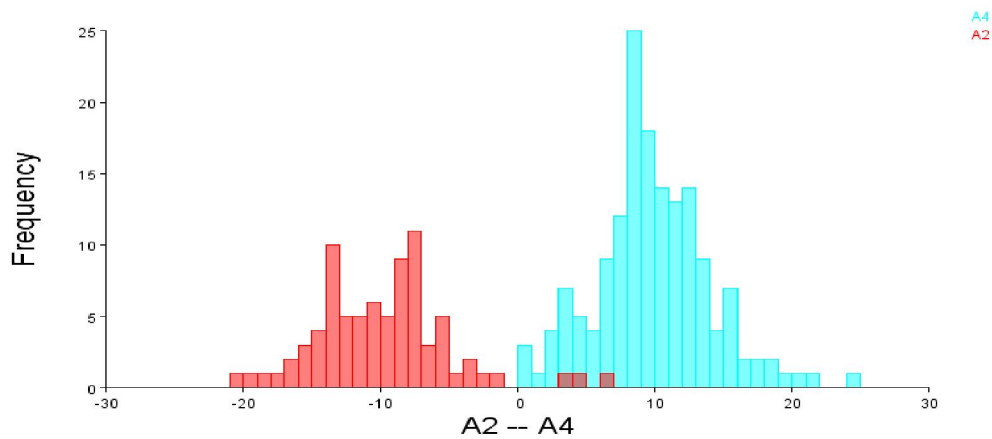
Figure 3.28: The discriminant scores from the Discriminant function analysis (DFA) for the comparison between the assays. The low levels of the discriminant scores (on the left) represent assay 2. The high levels of the discriminant scores (on the right) represent assay 4. All the treatments in each harvest were combined. The color key is on the upper right corner.

75
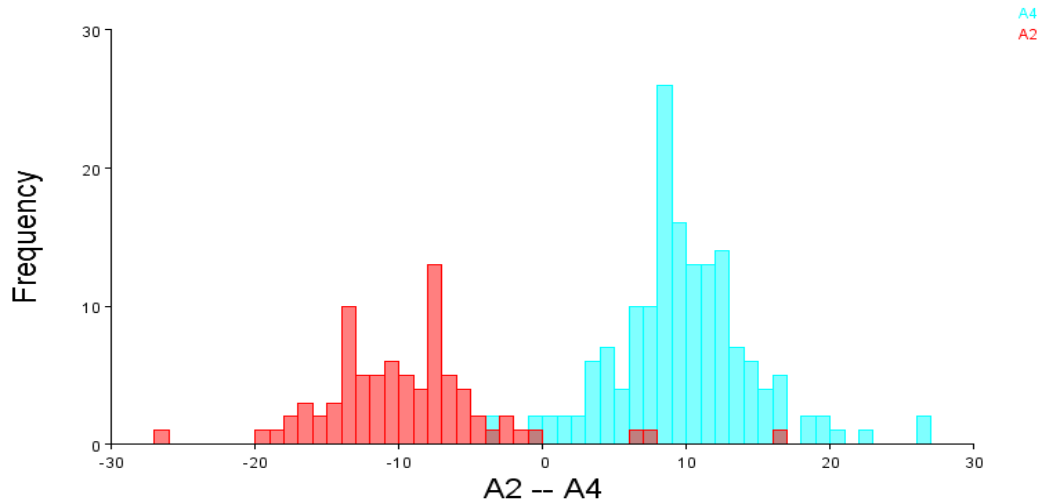
Figure 3.29: The cross-validation scores from the Discriminant function analysis (DFA) for the comparison between the assays. The low levels of the discriminant scores (on the left) represent assay 2. The high levels of the discriminant scores (on the right) represent assay 4. All the treatments in each harvest were combined. The color key is on the upper right corner.

Table 3.19: The reliability of the discriminant function analysis for the assay dataset, as indicated by the rates of misclassification, based on the discriminant and the cross-validation functions. Each value represents the number of misclassified individuals, divided by the total numbers of individuals in the group. Abbreviations: A = assay, P = positive, N = negative, R = random.

Comparison	Discriminant function		Cross-validation function	
	A2 as A4	**A4 as A2**	**A2 as A4**	**A4 as A2**
A2, A4	3/80	0/159	3/80	4/159
A 2-P, A 4-P	0/29	1/52	1/29	3/52
A 2-N, A 4-N	1/26	0/52	2/26	0/52
A 2-R, A 4-R	0/25	0/55	2/25	3/55

For the treatment-specific comparisons, all the treatment lines showed highly significant shape changes across the assays, as indicated by both Mahalanobis and Procrustes distances. All the lines showed similar shape changes to those observed between the whole assays, including the elongation of the body, the caudal fin and the snout-eye distance, the narrowing of the body and the anal fin, and the shortening of the caudal peduncle. However, the shape changes within the negative line (figure 3.30) were unique in three respects; first, the dorsal fin was widened rather than narrowed as in the two other lines. second, the dorsal shift of the anal fin was greater than the ventral shift of the dorsal fin, as a result, the area of the dorsal side has increased relative to that of the ventral side, contrary to what happened in the two other lines. Third, the base of the caudal fin did not become wider as in the two other lines, but became narrower. The positive and the random lines did not show notable deviation from the changes observed between the whole assays.

For all the treatment lines, the separation between the groups (figures 3.31 and 3.32), as well as the rates of misclassification (table 3.19), were similar to each other and to the whole assays.

a)

A2.P -- A4.P

b)

A2.N -- A4.N

c)

A2.R -- A4.R
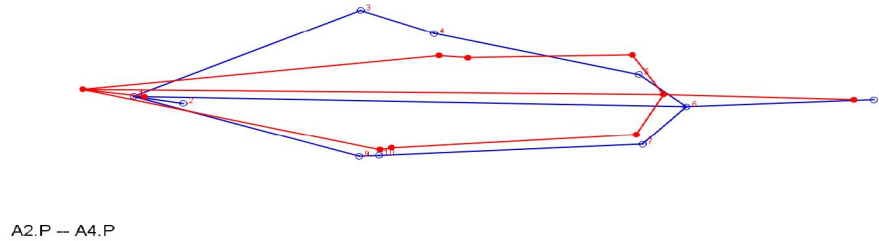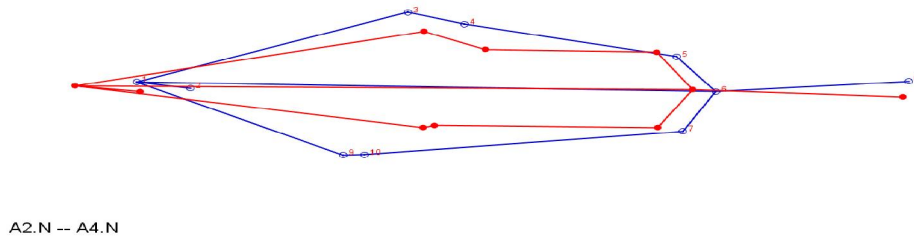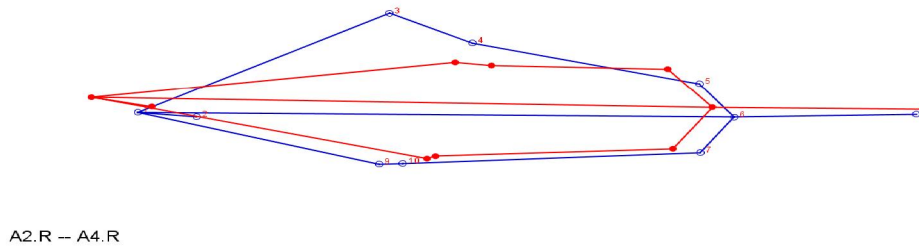
Figure 3.30: Wireframe representations of the shape changes between the assays that occurred within: **a)** the positive line (P), **b)** the negative line (N), and **c)** the random line (R). Analysis was performed by using discriminant function analysis (DFA). The blue color represents the starting shape, which describes the treatment of assay 2. The red color represents the target shape, which describes the treatment of assay 4. The shape changes were exaggerated by a factor of 3.

a)
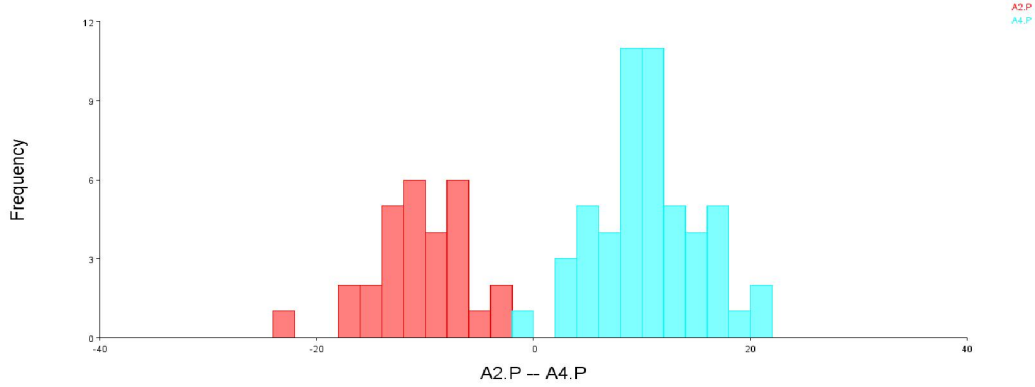
b)
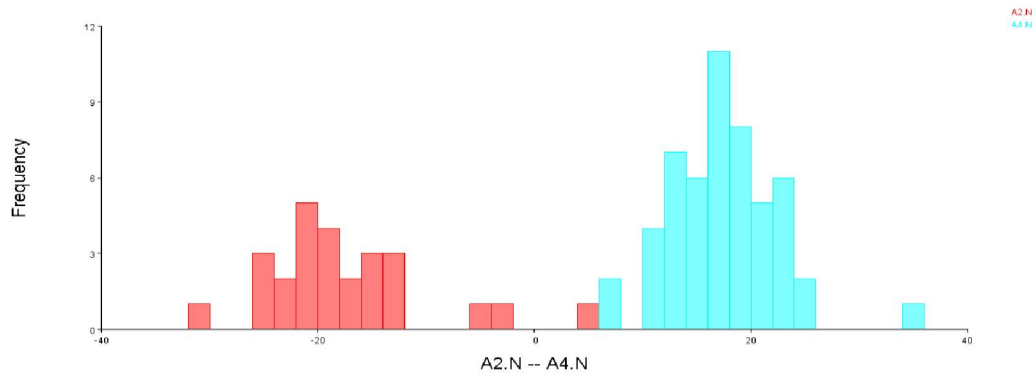
c)
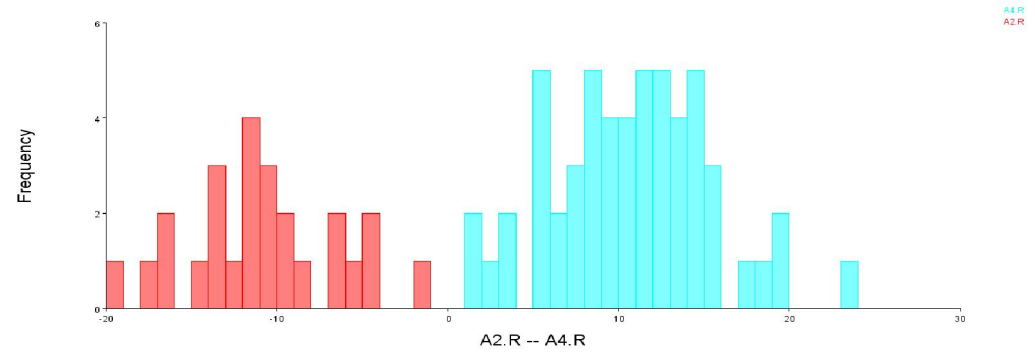
Figure 3.31: The discriminant scores from the Discriminant function analysis (DFA) for the comparison between the assays of: **a)** the positive line (P), **b)** the negative line (N), and **c)** the random line (R). The low levels of the discriminant scores (on the left) represent the treatment of assay 2. The high levels of the discriminant scores (on the right) represent the treatment of assay 4. The color key is on the upper right corner.

a)
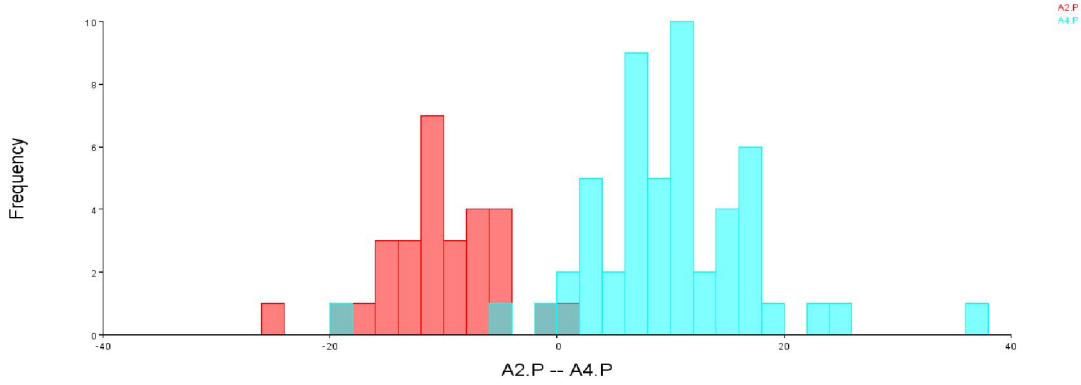
b)
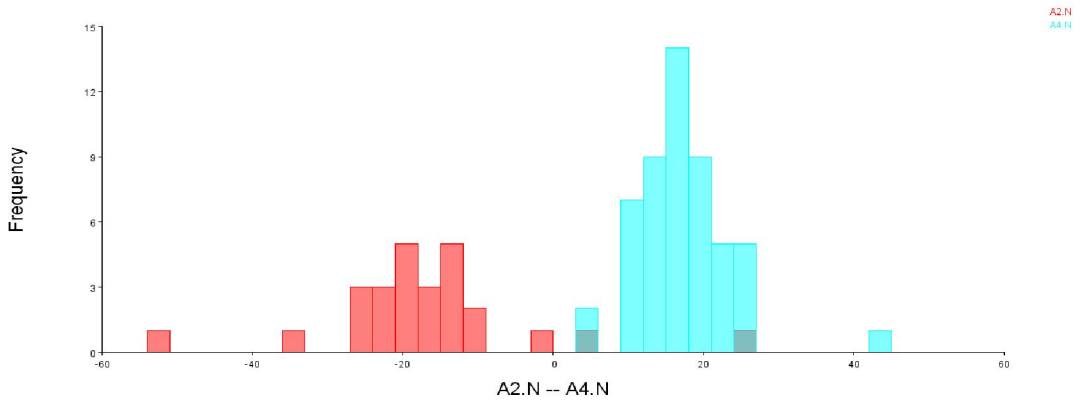
c)
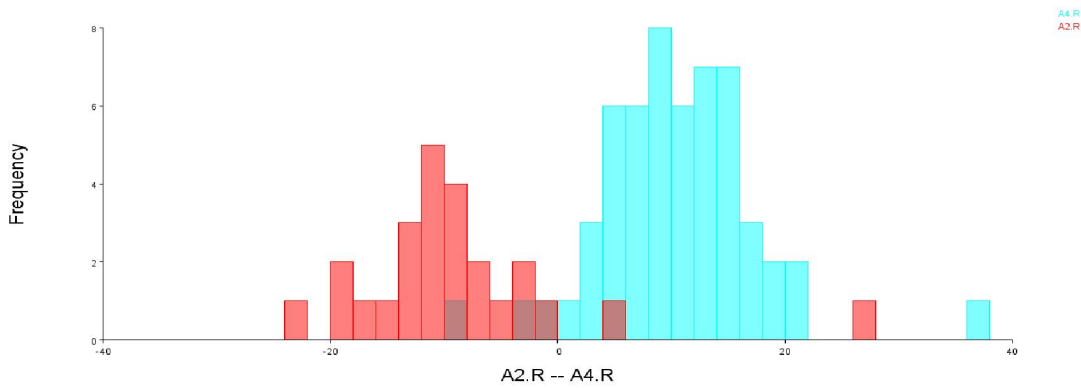
Figure 3.32: The cross-validation scores from the Discriminant function analysis (DFA) for the comparison between the assays of: **a)** the positive line, **b)** the negative line, and **c)** the random line. The low levels of the discriminant scores (on the left) represent the treatment of assay 2. The high levels of the discriminant scores (on the right) represent the treatment of assay 4. The color key is on the upper right corner.

79

3.2.3.2 Shape variation between treatments

The shape variation between the treatments of the assay dataset was investigated by using canonical variate analysis (CVA). Two separate CVA comparisons were performed, one for each assay. For assay 2, highly significant differences were detected between the positive and the random treatments, and also between the negative and the random treatments, according to both Mahalanobis and Procrustes distances. The difference between the positive and the negative treatments was highly significant for Mahalanobis distance but non-significant for Procrustes distance (table 20). Figure 3.33 represents the ordination plot of the canonical scores; it shows all the treatments to be highly overlapped, despite this, there is some separation between the random treatment and the two other treatments along the first CV axis, specifically, the individuals of the random treatment are clustered at the high levels of CV1, while the individuals of the positive and the negative treatments are clustered at the low levels of CV1. Therefore, the shape changes between the starting and the target shapes of CV1 can be interpreted as representing the shape differences between the positive and the negative treatments on one side (the starting shape) and the random treatment on the other side (the target shape).

Table 3.20: The outputs of canonical variate analysis (CVA) for the shape variation among the treatments of the assay dataset. The inputs were the residuals of the regression of procrustes coordinates on centroid size, representing the allometry-free shape variation. The analysis was performed separately for each assay, but jointly for the pairwise comparisons within each assay. The analysis was performed on a dataset containing one, left-side photograph for each individual, by using the program MorphoJ. The p values were calculated from permutation tests with 10,000 rounds. Significance levels: '***' = $p < 0.001$, '**' = $p < 0.01$, '*' = $p < 0.05$, '-' = $p \leq 0.10$.

Comparison	Mahalanobis distance	P-value for Mahalanobis distance	Procrustes distance	P-value for Procrustes distance
A 2-P, A 2-N	1.4160	0.0083**	0.0215	0.1326
A 2-P, A 2-R	1.8831	< 0.0001***	0.0302	0.002 **
A 2-N, A 2-R	1.6568	0.001 **	0.0428	0.0063**
A 4-P, A 4-N	1.1264	0.0046**	0.0308	0.2213
A 4-P, A 4-R	1.3958	< 0.0001***	0.0281	0.2911
A 4-N, A 4-R	1.4430	< 0.0001***	0.0139	0.4284

80
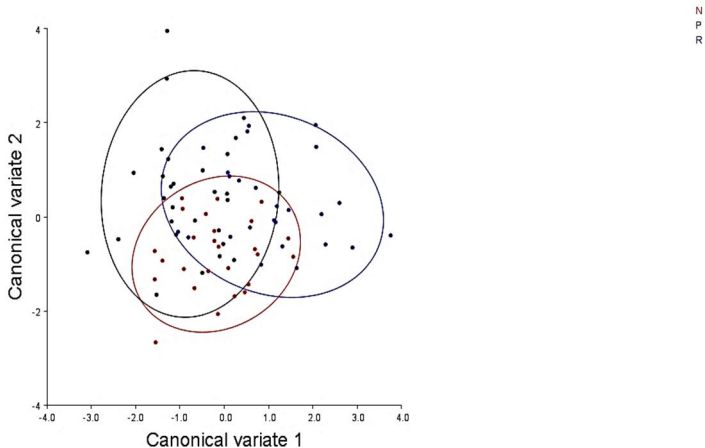
Figure 3.33: The ordination plot of canonical variate analysis (CVA) for the positive, the negative and the random treatments of assay 2. Each dot represents one specimen. Each treatment is represented by a different color; the color key is at the upper right corner. Confidence ellipses were set at a probability of 0.9.

Figure 3.34a displays the shape changes associated with CV1 (which accounts for 66% of the total variance); it shows an elongation in the standard length and the snout-eye distance, accompanied by a shortening in the caudal fin length. The dorsal and the anal fins became wider (but the change was very slight for the latter), while the caudal fin became narrower. Both the dorsal and the anal fins are shifted dorsally; as a result, the ventral side of the body became narrower relative to the dorsal side. The dorsal fin is shifted anteriorly, while the anal and the caudal fins are shifted posteriorly. Note that the changes are similar to those observed between the assays.

There is an extensive overlap at the vertical axis (CV2) of the ordination plot (figure 3.33), however, the individuals of the negative treatment are mostly found at the lower levels of CV2, while the individuals of the positive treatment are spread more widely, but they are mostly found at the higher levels of CV2. Therefore, the shape changes between the starting and the target shapes of CV2 can be interpreted as representing the differences between the negative treatment (the starting shape) and the positive treatment (the target shape). Figure 3.34b illustrates the shape changes associated with CV2 (which accounts for 34% of the total variance); it shows an elongation in the standard length and a shortening in the caudal fin length. The dorsal and the anal fins became wider, while the caudal fin became narrower. The dorsal fin is shifted anteriorly, while the anal fin is shifted posteriorly. Both the anal fin and the (anterior end of the) dorsal fin are shifted ventrally; all the above changes are similar to the ones observed in CV1. There are also anterior shifts in the dorsal and the ventral extents of the caudal fin insertion, accompanied by a posterior shift at the posterior tip of the caudal peduncle; as a result, the caudal peduncle became narrower.

81

a)
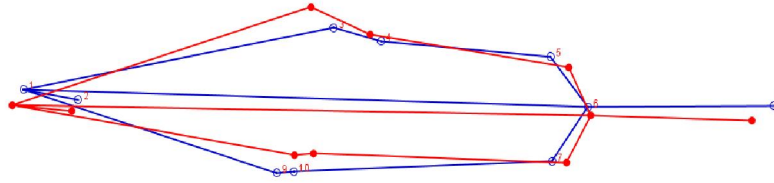
CV1

b)

CV2

Figure 3.34: Wireframe representations of the shape changes that are associated with: **a)** the first canonical variate (CV1), and **b)** the second canonical variate (CV2), of canonical variate analysis (CVA) for the treatments of assay 2. The blue color represents the starting shape, while the red color represents the target shape. The shape changes were exaggerated by a factor of 5.

For assay 4, the differences between the treatments were highly significant according to Mahalanobis distances, but were non-significant according to procrustes distance. Figure 3.35 represents the ordination plot of canonical scores for the treatments of assay 4; again, it shows extensive overlap along both the horizontal and the vertical axes. Most of the individuals of the random treatment are found at the lower levels of the horizontal axis, indicating that the shape changes associated with CV1 can be interpreted as differences between the random treatment (the stating shape) and the two other treatments (the target shape). Figure 3.36a illustrates the shape

changes that correspond with the first CV axis (which accounts for 65% of the total variance). It shows a lengthening in the snout-eye distance and the caudal fin length, accompanied by a shortening in the standard length. The basis of the dorsal, the anal and the caudal fins became wider. The anterior part of the body shows a dorsal shift relative to the posterior part. The dorsal fin is shifted anteriorly, while the anal fin is shifted posteriorly.

Despite the extensive overlap between the positive and the negative treatments, the individuals of the positive treatment were mostly clustered at the higher levels of the vertical axis, while the individuals of the negative treatments were mostly clustered at the lower levels of the vertical axis (figure 3.35), as a result, the changes between the starting and the target shapes of CV2 are interpretable as differences between the positive treatment (the starting shape) and the negative treatment (the target shape). Figure 3.36b illustrates the shape changes that are associated with the second CV axis (which accounts for 35% of the total variance). It shows a shortening in the caudal fin length. It also shows a narrowing in the base of the anal fin, and a widening in the base of the caudal fin. The dorsal fin is shifted dorsally, while the anterior extent of the dorsal fin insertion is shifted dorsally and posteriorly, resulting in an increase in the area of the dorsal side relative to that of the ventral side. The dorsal fin is shifted anteriorly, while the caudal fin is shifted posteriorly and ventrally; because of these changes, the caudal peduncle became larger relative to the anterior part of the body.
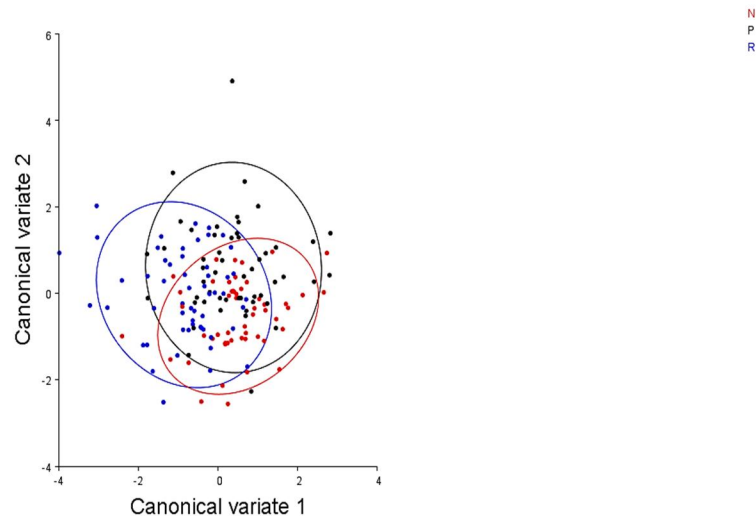

Figure 3.35: The ordination plot of canonical variate analysis (CVA) for the positive, the negative and the random treatments of assay 4. Each dot represents one specimen. Each treatment is represented by a different color; the color key is at the upper right corner. Confidence ellipses were set at a probability of 0.9.
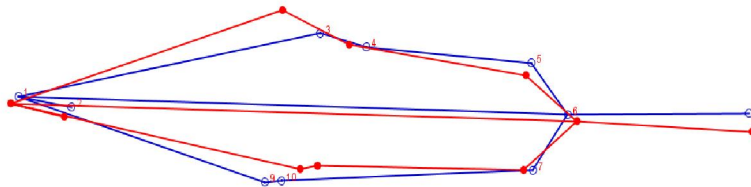
a)
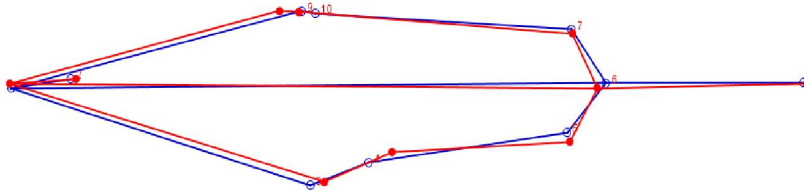
CV1

b)
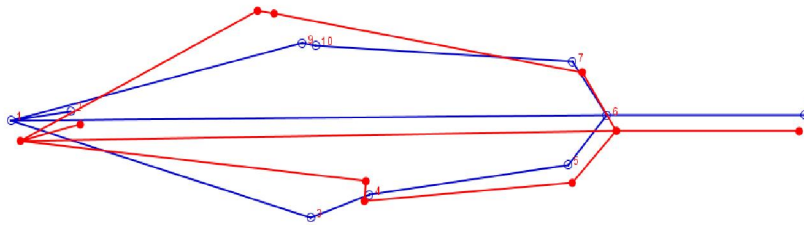
CV2

Figure 3.36: Wireframe representations of the shape changes that are associated with: **a)** the first canonical variate (CV1), and **b)** the second canonical variate (CV2), of canonical variate analysis (CVA) for the treatments of assay 4. The blue color represents the starting shape, while the red color represents the target shape. The shape changes were exaggerated by a factor of 5.

To compare the shape variation between the treatments of assay 2 to that between the treatments of assay 4, an ordination plot was produced that jointly represent all the treatment groups in the assay dataset (figure 3.37). The plot shows that the separation between the assays (along the horizontal axis) is greater than the separation between the treatments (along the vertical axis); the first canonical variate (CV1), which describes the separation between the assays, accounts for 88% of the total variation among the groups. The plot also shows that the individuals of assay 2 display

greater shape variation than the individuals of assay 4, as they were dispersed more widely along both the horizontal and the vertical axes. Finally, despite the extensive overlap between the treatments in both assays, there is more separation between the treatments of assay 2 than that between the treatments of assay 4, indicating that the shape variation among the treatments has decreased during the period between the two assays. This reduction in variation can also be seen by comparing the probability values for Procrustes distance between assay 2 treatments to these values between assay 4 treatments (Table 3.20).
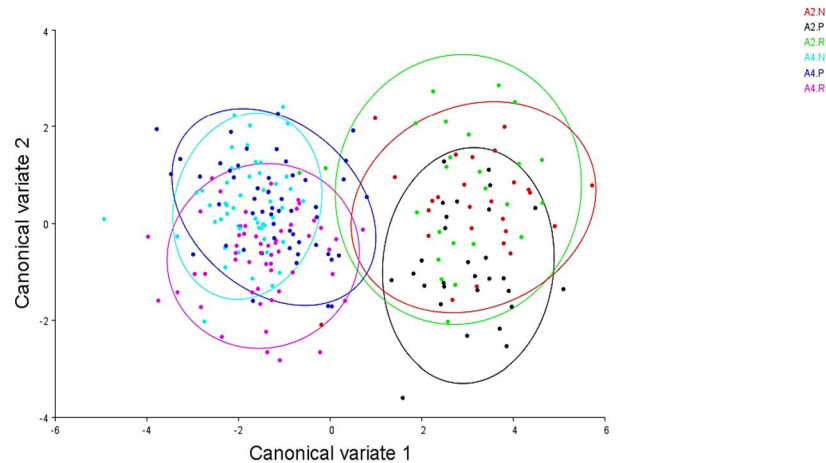
Figure 3.37: The ordination plot of canonical variate analysis (CVA) for the positive, the negative and the random treatments of the assays 2 and 4. The groups are defined by the assay and the treatment. Each dot represents one specimen. Each group is represented by a different color; the color key is at the upper right corner. Confidence ellipses were set at a probability of 0.9.

3.2.4 Trends in shape data: a synthesis

For both the harvest and the assay datasets, the statistical comparisons through time and between treatments have detected significant shape differences (tables 3.15, 3.17, 3.18 and 3.20). In order to distinguish evolutionary changes from changes due to phenotypic plasticity (or measurement error), the focus will be on the robust shape differences, that is, on the differences that occurred in the harvest populations and persisted in the assay populations under common garden conditions. In addition to describing the robust shape differences, we will also investigate the patterns of shape variation across time and between treatments, specifically, how the shape variation between the individuals and between the treatments has changed through time in the harvest and the assay experiments.

Regarding the shape changes through time, some of the observed changes were most likely plastic responses to the variation in the rearing conditions; for example, the relative length of the caudal peduncle (compared to the length of the anterior part of the body) did not change consistently through time, as it increased in the harvest experiment but decreased in the assay experiment

(figures 3.16 and 3.27). Other changes are likely to be adaptive responses to the experimental environment, as they occurred in the harvest experiment and persisted in the assay experiment. There are four shape changes that fit this description: 1) the caudal peduncle became narrower, 2) the dorsal fin became narrower, 3) the anal fin became narrower, and 4) the dorsal side of the body became narrower relative to the ventral side. All the four shape changes were replicated in both experiments (figures 3.16 and 3.27), and in each treatment line within each experiment (figures 3.19 and 3.30), thereby indicating the existence of genetic changes that underlay the observed shape changes.

The investigation of the shape changes through time within each treatment line yielded similar results to the investigation of the shape changes through time within the whole harvests/assays, indicating that the positive, the negative, and the random lines evolved in similar ways as they were adapting to the experimental environment (figures 3.19 and 3.30). Comparisons of the shape features between the treatments within each harvest/assay detected a number of significant shape differences during both stages of the selection experiment (harvest 4/ assay 2, and harvest 28/ assay 4) (tables 3.17 and 3.20). However, the observed shape differences were not consistent, neither through time (i.e. between subsequent harvests/ assays) nor across experiments (i.e. between each harvest and the corresponding assay) (see sections 3.2.2.1 and 3.2.3.2). This indicates that the positive, the negative and the random treatments did not evolve stable shape differences that consistently separate each treatment from the others, contrary to the case with the harvests and the assays. Most of the variation between the treatments was due to the differences between the random treatment on one side, and the positive and the negative treatments on the other side (in 3 out of 4 groups: harvest 28, assay 2, and assay 4) (figures 3.24, 3.33 and 3.35), indicating that the positive and the negative treatments were (in most cases) more separated from the random treatment than they are separated from each other.

For both the harvest and the assay datasets, there was more changes through time than there was differences between the treatments. For both the harvest and the assay datasets, the individuals of the early stages of the experiment (harvest 4 / assay 2) spread more widely across the shape space than did the individuals of the late stages of the experiment (harvest 28 / assay 4), indicating that the shape variation between individuals has decreased during the period between the two stages. Moreover, and for both the harvest and the assay datasets, there was more variation between the treatments at the early stages of the experiment than at the late stages of the experiment, indicating that the shape variation among the treatments has decreased during the period between the two stages. The figures 3.26 and 3.37 illustrate the above trends for the harvest and the assay datasets, respectively. Comparing the two figures reveals an additional pattern, namely, that the changes through time were greater in the harvest experiment than in the assay experiment, as there was more separation between the harvests than the separation between the assays; this indicates the effect of phenotypic plasticity, because successive harvests varied in their environmental conditions (e.g. abundance), unlike successive assays.

4. Discussion

The design of the guppy selection experiment, with its positive, negative and random treatments, renders certain outcomes very straightforward to explain; these outcomes obtain if the positive and the negative treatments diverged most from each other, leaving the random treatment somewhere in between. The actual results of our study did not show such a clear pattern, neither for color nor for shape traits. However, the results also did not match the predictions of the null hypothesis, because significant effects of the experimental treatments were observed for both color and shape traits. The actual results of color and shape analyses represent complex patterns of changes (through time) and differences (between lines) with various directions and magnitudes, and possibly with various causes as well. In explaining these patterns, the experiment will be treated as a miniaturized ecosystem or a microcosm; explanations will be presented and qualitatively assessed for their consistency with the data, with the conditions during the experiment, and with previous research on similar systems, natural and experimental. It is important to keep in mind that, because the outcomes did not conform to the transparent pattern described above, and because the design of the experiment allows multiple mechanisms to affect the outcomes, either separately or jointly, it is not possible to establish any particular explanation with certainty, since alternative explanations cannot be completely ruled out.

4.1   Color changes

In the present study, a number of color changes were observed that are likely to have genetic basis. Two patterns can be identified, one related to the orange coloration and the other to the black coloration. For the orange coloration, the numbers and the relative areas of orange spots increased in the positive line between the assays 1 and 2 (regarded as the 'early stage' of the experiment), then decreased between the assays 2 and 4 (regarded as the 'late stage' of the experiment). The available data from the harvest dataset falls within the late stage of the experiment (i.e. between the harvests 4 and 28), and it confirms that the positive line showed a decrease in orange coloration during this period. The changes in orange coloration that occurred in the random and the negative lines were either non-significant (for the early stage) or not consistent between the harvest and the assay experiments (for the late stage), suggesting that the pattern identified above was probably unique to the positive line. For black coloration, the relative area of black spots decreased during the early stage, then increased during the late stage; pairwise comparisons of whole assays show that this pattern is highly significant. Moreover, comparisons within each treatment line confirm that the decrease (during the early stage) took place in all the lines, and suggest that the increase (during the late stage) probably occurred in the positive and the random lines. Overall, these results indicate that the orange and the black coloration were evolving independently during the experiment, and that the changes in the orange coloration were unique to the positively-selected populations while the changes in the black coloration were general to the experimental populations (with the exception of the negatively-selected populations during the late stage of the experiment).

The changes in black coloration are probably genetic, since they occurred under common garden conditions. However, it is difficult to identify the mechanism underlying these changes. The populations from all the lines showed similar changes even during the early stage of the experiment, during which the population densities (that affect competition for food) varied substantially between the lines. We therefore attribute the observed changes to selection imposed by unidentified aspect of the experimental environment, one that fluctuated through time. The rest of the discussion of color changes will be focused on explaining the changes in orange coloration that occurred in the positive line.

Because of the design of the selection experiment, changes that occur in one but not the other lines are most likely caused by the harvesting regimes, as these are the only aspects of the experimental environment that varies systematically between the lines. Therefore, the positive harvesting regime is the most likely cause for the observed changes in the orange coloration. Note, however, that the causal link between the harvesting regime and the color change cannot be direct. Guppy's colors are well-known to evolve under the pressure of visually-hunting predators. Previous work has shown that various elements of guppies' color patterns, including the numbers, the relative areas, and the brightness of color patches, can change adaptively within ecological time scales, responding to predators' selection against conspicuous color patterns (see the introduction; section 1.2.2.). However, the guppy selection experiment investigates the evolutionary effect of a different kind of predation, namely, human harvesting, The key difference between harvesting and natural predation is that the former does not select against conspicuousness; in the guppy experiment, more colorful males were not inherently more vulnerable to harvesting mortality than other males; the situation is similar in commercial fisheries, where the fishing gears are typically selective for size, behavior and spatial distribution (Heino, and Godø, 2002), but not for conspicuousness (however, other forms of human harvesting are highly selective for conspicuousness, such as the collection of terrestrial snails 'Allendorf and Hard, 2009 '). Therefore, the positive harvesting regime dose not directly select for color in the same way it selects for size; a more indirect link between the harvesting regime and the male color must underlie the observed changes.

One possible explanation is that the observed changes in the orange coloration represent indirect responses to selection against large size. In the positive line, large males were preferentially harvested; selection against large males is expected to favor fast life histories, that is, life histories that are characterized by more investment on current reproduction than on future reproduction (Reznick and Ghalambor, 2005). One of the features of a fast life history (together with early maturation and slow post-maturation growth) is increased reproductive effort. The evidence for fisheries-induced evolution in reproductive effort is little, possibly because little research attention was directed to this particular aspect of the life history compared to other aspects such as maturation schedules (see the introduction; section 1.1.1.). However, the theory of life history evolution predicts that an increase in mortality rates for the large size classes will favor individuals that allocate more resources to reproduction than to growth or maintenance. Furthermore, previous research on the effect of predation on guppies' life histories confirmed this prediction; guppies from high predation locations (where large individuals suffer elevated mortality from predatory fishes) had greater reproductive allocation than guppies from low predation locations, and these differences were found to persist under common garden conditions, indicating their genetic origin

(see the introduction; section 1.2.1). The relationship between the amount of orange coloration and reproductive effort can be inferred from two aspects of the ecology of guppies: first, that, for most populations, orange coloration is attractive to females (Houde, 1987; Endler and Houde, 1995), therefore, an increase in the area of orange color improves the males' mating success. Second, that having more orange coloration is expensive, not only because brightly colored males suffer higher predation mortality, but also because they suffer higher juvenile mortality (Brooks, 2000). Because orange coloration improves reproductive success at the expense of the survival rate, an increase in orange coloration indicates a life history shift toward increased reproductive effort.

Above, we argued that: 1) selection against large size could lead to an increase in reproductive investment, and 2) for guppies, an increase in reproductive investment could manifest as an increase in orange coloration. Together, these points lead to the prediction that the males under the positive harvesting regime will be selected for increased orange coloration. Interestingly, this prediction is valid only for experimental environments, where harvesting is neutral for color; in the wild, predators select simultaneously against large size (Reznick and Ghalambor, 2005) and bright coloration (Endler, 1983). Consequently, the possible adaptive responses to size selection are constrained by color selection; the males that develop more orange coloration under high predation conditions will suffer substantially higher predation rate, most likely overwhelming any benefit from increased reproductive effort.

There is an alternative explanation for the observed changes in the orange coloration; one that also depends on the indirect effects of the harvesting regime, but it relies on a different link, namely sexual selection. In guppies, females are larger than males, consequently, selection against large size (as in the positive harvesting regime) is expected to remove females more often than males, leading to male-biased populations. Because female choice is expected to be more intense in male-biased populations (Jirotkul, 2000), the reproductive value of the orange coloration is expected to be greater, potentially leading to the evolution of more orange coloration. Previous research has documented that fishing activities can alter the population sex ratios of the exploited species, potentially leading to changes in sexual selection and mating systems. The effects of fishing on the sex ratio varies between species, as it depends on the sex-specific pattern of sexual dimorphism in size, behavior, and distribution (Rowe and Hutchings, 2003).

Natural guppy populations are mostly female-biased (Rodd and Reznick, 1997), probably because males' coloration renders them more conspicuous to predators (Reznick *et al.*, 1996a). However, extensive fluctuations in the sex ratio have been observed in wild guppy populations, including shifts from female-biased to male-biased sex ratios (Pettersson *et al.*, 2004). An increase in the percentage of males will increase the range of candidates that are available for females to choose from, as a result, the selectivity of females' choices will increase, and, consequently, the reproductive cost to the males that are not highly colorful will also increase. The expected result is an increase in the intensity of sexual selection, driven by the male-biased sex ratio. Consistent with this prediction, Jirotkul (1999; 2000) showed experimentally that, under male-biased operational sex ratio (i.e. the ratio of the sexually active males to the total number of sexually active adults), females exerted stronger preference for males with larger orange-colored area, thereby increasing the intensity of sexual selection on the males. To conclude, the above

explanation can be summarized in two points: 1) selection against large size could lead to male-biased sex ratios, and 2) male-biased sex ratios could increase the intensity of sexual selection, leading to the evolution of more orange coloration.

Above, we considered two possible explanations for the changes in the orange coloration that occurred in the positive line, the first assumes changes in the reproductive effort, induced by the harvesting regime, and the second assumes changes in the population sex ratio, also induced by the harvesting regime. Both explanations are consistent with the design of the experiment and with the relevant literature. The key observation that could distinguish between these explanations is that the orange coloration did not change consistently through time, but fluctuated, increasing during the early stage of the experiment then decreasing during the late stage. If the observed color changes reflect changes in the reproductive effort, it is expected that the intensity of selection for a faster life history should be higher during the early stage than during the late stage. If, on the other side, the color changes reflect changes in the intensity of sexual selection, it is expected that more male-biased sex ratios should be observed during the early stage than during the late stage.

A key feature of the experimental environment that affected the changes in all the lines and in all the traits is the population density. During the early stage of the experiment, there was more variation in population densities between the lines than during the late stage. The negative treatment had higher densities than the two other treatments, therefore, competition for food (food levels were equal for all the populations) was more severe in the negative treatment. During the late stage of the experiment, population densities were reduced and controlled, therefore, all the lines had similar densities in this stage, and the only difference between their environments was the harvesting regimes. Importantly, the positive line showed a decrease in the length at maturation during the early stage, consistent with the occurrence of selection for a fast life history imposed by the harvesting regime. During the late stage, the length at maturation started increasing, indicating that the harvesting regime had less selective effect during this stage; the likely reason for this is that the population densities have dropped to a level at which the increased availability of food, and the subsequent reduction in competition, have compensated for harvesting mortality (Diaz Pauli, unpublished data). Overall, the observed changes in maturation schedules suggest changes in the intensity of harvesting selection that are consistent with the observed changes in the orange coloration. Therefore, the first explanation is consistent with observations.

Regarding the second explanation, the data on the sex ratio changes through time are not consistent with this explanation: first, during the early stage of the experiment, the positive line did not have higher percentage of males than the other lines. Second, the percentage of males have decreased through time in the random and the negative line, but not in the positive line, which maintained roughly similar values throughout the experiment (Diaz Pauli, unpublished data). Therefore, the observations do not agree with the scenario in which the positive harvesting regime caused male-biased sex ratio during the early stage of the experiment, followed by a reduction in the male-bias during the late stage. Note, however, that at this point, the sex ratio data is not fully analyzed; the presented analysis is only preliminary.

4.2   Shape changes

In the present study, geometric morphometrics analysis revealed various differences in the shape features between the experimental groups; including changes through time and differences between the treatments. Both the nature and the magnitude of the observed differences were dependent on the 'dimension' of the comparison. For the comparisons through time, robust shape differences were observed that persisted across the lines (i.e. the positive, the negative and the random lines) and also across the experiments (i.e. the harvest and the common garden experiments). These differences include narrowing in the caudal peduncle, in the dorsal and the anal fins, and in the dorsal side of the body for the males of the late generation (i.e. harvest 28 and assay 4) compared to the males of the early generation (i.e. harvest 4 and assay 2). Moreover, Different generations were statistically distinguishable by their shape features, indicating considerable change through time. For the comparisons between the treatments, the observed differences were not robust, since they neither persisted through time nor replicated across the experiments. Moreover, the treatments showed extensive overlap in their shape features, indicating that the differences, though statistically significant, were slight in magnitude. Below, we first discuss the evolutionary significance of fish shape, with special emphasis on poeciliids; in the light of the that discussion, the shape differences between the experimental groups will be explained, starting with the differences through time then the differences between the treatments.

The shape traits of fishes are known to affect various aspects of their fitness, among which the most prominent is swimming performance (Domenici, 2003). Previous research has shown that the shape features, such as the depth of the caudal peduncle, the caudal fin size, and the body fineness ratio (i.e. body length divided by maximum body depth; a measure of streamlining) have strong and predictable effects on swimming ability. Moreover, research also revealed that swimming ability has strong and predictable effects on performance in key fitness-determinant activities, such as foraging, predator avoidance and mate acquisition (Langerhans and Reznick, 2009). These fundamental relationships between body design, swimming ability, and fitness can induce adaptive phenotypic divergence between populations of fishes if two conditions are met: first, that there is a tradeoff between the different components of swimming performance, such that not all of them can be maximized simultaneously, and second, that the relative fitness value of these components varies through time and space (e.g. due to an environmental gradient) (Langerhans, 2009). Under these conditions, divergence in body design is expected, because overall individual's performance will be optimized by different designs depending on the local selection regimes.

There is a considerable and growing literature on adaptive, inter-population shape divergence in fishes; a number of tradeoffs involving body shape were identified, together with a number of environmental factors (both biotic and abiotic) that affect the optimal solution for these tradeoffs (Langerhans and Reznick, 2009). The tradeoff between steady and unsteady swimming has been investigated in a number of fish taxa, including Poeciliids, and was documented to cause phenotypic divergence in the wild (Langerhans and DeWitt, 2004). Steady swimming refers to prolonged swimming activities that maintains constant speed and direction; this mode of swimming is used for cruising in open water, searching for food and mates, as well as during the

antagonistic interactions between conspecifics (e.g. chasing). Unsteady swimming refers to swimming movements that are characterized by repeated changes in velocity and direction, such as fast-start escapes and rapid turns, and are typically employed in predatory and anti-predatory behaviors, and also during navigation in structurally complex habitats (e.g. coral reefs). Fast start escapes refer to the sudden, short bursts of high-energy swimming movements that are employed in escaping predators. Both steady and unsteady swimming depend on body shape, however, the shape features that increase the efficiency of each swimming mode tend to decrease the efficiency of the other mode, thereby leading to a fundamental design tradeoff. Steady swimming is most efficient (in terms of energy expenditure and endurance), if the body has a streamlined shape, with deep anterior body and narrow caudal peduncle, as this body shape increases thrust while reducing drag and energy loss. Fast-start escapes, on the other side, are most efficient (in terms of speed) when the anterior body is narrow and the caudal peduncle is deep, as this body shape increases thrust and stability during rapid swimming movements (Langerhans *et al*., 2004; Langerhans, 2009; Langerhans and Reznick, 2009). Due to this biomechanical constraint, it is expected that natural selection on the shape can improve one swimming mode only at the expense of the other.

A straightforward prediction that follows from the above considerations is that conspecific populations of fishes that experience different levels of predation and competition will evolve body shapes that maximize different swimming modes, leading to adaptive phenotypic divergence (Langerhans and Reznick, 2009). Previous research on the family poeciliidae supported this prediction, especially regarding the effect of predation. Langerhans and DeWitt (2004) found that the high-predation populations of three poeciliid species: *Poecilia reticulata*, *Gambusia affinis*, and *Brachyrhaphis rhabdophora* have converged on elongated body shapes and larger caudal peduncles compared to conspecific, low-predation populations. Further research on *Poecilia reticulata* (Hendry *et al*., 2006) and *Gambusia affinis* (Langerhans *et al*., 2004) showed that high predation populations also had smaller heads than their low-predation conspecifics. The design tradeoff between steady and unsteady swimming was investigated in more detail in *Gambusia affinis* (Langerhans, 2009); high-predation populations (which are also low-density populations) had faster burst swimming, and low predation populations (which are also high-density populations) had greater endurance during steady swimming. These results indicate that the observed inter-population differences in the shape features and the swimming performance are most likely caused by differences in the relative importance of predation versus competition between high- and low-predation habitats.

The tradeoff between steady and unsteady swimming is 'internal', in the sense that both are aspects of the same biological function, namely swimming. Other tradeoffs are 'external', that is, swimming performance (in both steady and unsteady swimming) is constrained by performance at some other biological function; the examples include reproductive allocation and growth rate (Ghalambor *et al*., 2003). In *Poecilia reticulata*, females in high-predation populations are selected for increased reproductive allocation, therefore, they have more embryos developing in their bodies compared to females from low-predation populations. Under high predation conditions, females are also selected for increased escaping performance. The two selective pressures are conflicting, because: 1) the greater weight of the developing embryos will increase drag, thereby decreasing acceleration, and 2) the increased reproductive allocation will reduce the resources

available for muscular tissues. Consistence with the tradeoff hypothesis, high-predation females were found to achieve faster escape speeds but at higher energetic costs, especially as they approach parturition (Ghalambor *et al*., 2004). In the Atlantic silverside, populations at the northern latitudes experience lower temperatures and a shorter growth season than their southern counterparts; because larger individuals have higher rates of survival during winter, selection favors enhanced growth. However, because allocating more resources to growth could reduce the resources available for the maintenance of muscular tissues, it is predicted that the northern populations might have lower escape speeds and, therefore, lower rates of surviving predator attacks, compared to the southern populations; both predictions were confirmed experimentally (Billerbeck *et al*., 2001; Lankford *et al*., 2001).

In the present study, guppies were maintained for multiple generations under experimental conditions that differ substantially from their original habitat (which is a low-predation site in the Yarra river in Trinidad). Two of the differences between the ancestral and the derived conditions could have a strong effect on body shape, either as selective agents for evolutionary change, or as environmental cues for phenotypic plasticity: first, natural predation does not occur, and the experimental harvesting that simulates predation is not selective for swimming performance; all the individuals from a given size class have the same probability of being harvested, regardless to their ability for fast-start escapes. As a result, harvesting mortality has no direct selective effect on (allometry-independent) shape features. Second, guppy populations were kept in small habitats (relative to the size of their original habitat; Diaz Pauli, personal communication), therefore, the number of individuals per water volume must have been higher than that in the original habitat. Consequently, intraspecific interactions must have occurred more frequently, and competition for food and mates is likely to be more intense, compared with the original habitat. Taken together, these differences suggest that the balance between predation and competition is shifted strongly toward competition as the most important selective force under the experimental conditions. Therefore, it is predicted that guppies will evolve (or show a plastic response) toward more efficient sustained swimming by developing more streamlined bodies.

The results of the shape analysis were diverse and quite complex, most likely due to phenotypic plasticity and demographic histories (see below). despite this, some of the observed shape changes agree with our prediction. Specifically, the males of harvest 28 had narrower caudal peduncle than the males of harvest 4, resulting in a more elongated posterior part (figure 3.16). The landmark configuration was not sensitive to the depth of the anterior part, however, it could be inferred from the lack of change in the snout-eye distance that there was no similar elongation in the anterior part. Therefore, the overall shape change in the harvest experiment is that the posterior part became narrower relative to the anterior part, that is, the body became more streamlined. The narrowing of the caudal peduncle persisted under common garden conditions, as shown by the comparison between the males of assay 4 and assay 2 (figure 3.27). This convergence between the harvest and the assay experiments indicates that the observed shape change has a genetic component, and therefore represents an evolutionary change (but it has a plastic component as well; see below). Taken in the light of the biomechanical tradeoff between steady and unsteady swimming (Langerhans and Reznick, 2009), this result predicts that the guppies in the late stages of the experiment evolved more efficient sustained swimming at the expense of less efficient unsteady

swimming; if this is the case, late-stage guppies should have lower escape speeds and lower survival rates during encounters with predators (lower relative to early-stage guppies as well as wild guppies from both high-predation and low-predation habitats). It would be interesting to test this further prediction by conducting a comparison of escaping performance and escaping success between the late-stage guppies and wild guppies that belong to the same original site in the Yarra river.

The other robust shape changes that occurred during the experiment include the narrowing that occurred in the dorsal fin, in the anal fin and also in the dorsal side of the body (figures 3.16 and 3.27). These changes were genetic, therefore, they were most likely selected for by some aspects of the experimental environment, but the identity of the selective agent is unclear. We suspect that these changes, especially the narrowing of the fins, might be due to the reduced locomotory demands in the aquaria (e.g. the lower flow rate, and the limited spatial distribution of the resources). To our knowledge, no previous study on guppy's shape has reported similar changes.

Some of the changes in the shape of the caudal peduncle were replicated in both the harvest and the assay experiments (i.e. the narrowing). Other changes, however, were unique to each experiment. In the harvest experiment, the caudal peduncle became narrower/more elongated relative to the anterior part of the body (figure 3.16), while in the assay experiment, the whole body became narrower/more elongated (figure 3.27). The difference between assay 4 and assay 2 is genetic, and its genetic base must be the same as the one underlying the difference (in the shape of the caudal peduncle) between harvest 28 and harvest 4. This is because the guppies in the assay experiment were not allowed, by the design of the experiment, to evolve independently, but only to reflect the genetic differences between their grandparents that were collected from the harvest populations at different points in time (see chapter 2; section 2.1.3). Therefore, the difference between the shape changes that occurred in the harvest and the assay experiments is not due to a genetic difference between the respective populations, but were caused by their plastic responses to their respective rearing conditions. The environmental conditions during the development differed substantially between the two experiments; compared with the harvest guppies, the assay guppies are reared in substantially smaller aquaria (2-litre aquaria compared to 400-litre aquaria), and experienced substantially less social interaction (they were kept in isolation, while the harvest guppies are maintained in large populations where competition, cannibalism and courtship occur frequently). As a result, guppies in the harvest and the assay experiments were exposed to different environmental cues, leading to the more extreme shape change observed in the assay experiment. We conclude that some of the shape changes that were observed between the harvests 4 and 28, and between the assays 2 and 4, indicate genetic changes in the sensitivity to environmental cues, that is, in the pattern of the genotype-by-environment interaction (Herron and Freeman, 2014).

Previous research on guppies identified an important role for phenotypic plasticity in determining body shape. Research also revealed the existence of genetic variation between guppy populations in their sensitivity to rearing conditions (i.e. in phenotypic plasticity), consistent with the findings of the present study. Burns *et al.* (2009) compared the body shapes of wild-collected with lab-reared guppies from multiple populations; they found that, for some populations, the shape of the caudal peduncle differed substantially between wild-collected and lab-reared guppies from the

same population. For other populations, rearing conditions had little effect on the shape of the caudal peduncle. Dowdall *et al*. (2012) showed that phenotypic plasticity in head shape and other traits varies between populations, and that it is adaptive; all the guppies reared in the presence of predator cues developed predator-associated phenotypes (including a fusiform head and a lower position of the eye), regardless to their population of origin. However, when reared in the absence of predator cues, guppies from high-predation populations developed high-predation phenotypes, and those from low-predation populations developed low-predation phenotypes.

Above, we discussed the adaptive value of the streamlined shapes that developed in the harvest experiment, in the light of the tradeoff between steady and unsteady swimming. The additional plastic response that occurred in the assay experiment might as well be adaptive. A symmetrical narrowing/elongation for the entire body does not lead to streamlining, and therefore does not improve the efficiency of steady swimming. Note, however, that under the conditions of the assay experiment, efficiency in steady swimming, as well as efficiency in unsteady swimming, are redundant; this is because neither competition nor predation take place under these conditions. Therefore, the symmetrical narrowing of the body might serve another function, namely, to divert resources from muscular tissues toward another task; if this is the case, the (potentially adaptive) plastic response would be caused by an external tradeoff between swimming performance (including both steady and unsteady swimming) and another aspect of the individual's performance. Further investigation is needed in order to confirm that the plastic shape change is adaptive, and to identify the target of the resource allocation (the aspect of performance that was improved at the expense of swimming performance).

During the experiment, harvest guppies were subjected to different mortality regimes: a positive harvest in which large individuals were preferentially removed, a negative harvest in which small individuals were preferentially removed, and a random harvest in which the removal of individuals was non-preferential. It is expected that, in the positive line, the higher mortality rate for larger, older individuals will induce adaptive changes toward smaller size, earlier maturation and greater early-life reproductive effort for both sexes, and slower post-maturation growth for females (see chapter 1; section 1.1.1); the shift toward greater investment in early reproduction versus future reproduction might induce further shifts in the optimal solutions for the tradeoffs related to body design, specifically, resources might be reallocated from muscular development toward gonadal development, embryos' development, or courtship activity. A predicted outcome would be a narrowing of the body in the positive line. Because large body and caudal fin sizes are attractive to females (Reynolds and Gross, 1992), it is also predicted that males in the positive line might compensate for their smaller size by evolving a larger caudal fin so that they look bigger, in order to retain sexual attractiveness at a lower cost.

Regarding the body shape, the positive, the negative and the random lines of the selection experiment did not diverge in the predicted way. For both the harvest and the assay experiments, the lines showed highly similar shape changes through time (e.g. the narrowing of the caudal peduncle) (figures 3.19 and 3.30). This indicates that the selection by the experimental environment (which was similar across all the lines) overwhelmed selection by the experimental treatment (which varied across the lines). Although this outcome was not predicted, it can be

straightforwardly explained; The absence of natural predators (that preferentially kill the slow-escaping guppies), and the subsequent increase in the advantage of steady swimming compared to unsteady swimming, exert direct selection on the body shape, and therefore should induce a greater evolutionary response than the indirect selection on the body shape by the harvesting regimes, which was mediated by their direct selection on the life-history traits.

Although the experimental lines did not diverge as predicted, they did diverge in their body shapes. For all the experiments and the generations (harvest 4 and harvest 28 / assay 2 and assay 4), all the treatments were significantly different from each other according to Mahalanobis distance, and some pairs of treatments were also significantly different according to Procrustes distance. The pattern of the divergence between the treatments had three characteristic features that are relevant to the explanation of this divergence: first, at the beginning of the selection experiment, all the treatments had similar body shapes (Jayawickrama, 2013); the differences between the treatments accumulated during the first stage of the experiment, and then decreased during the second stage. As a result, the treatments were less separated by the end of the experiment than they were during the middle (figures 3.26 and 3.37). Second, none of the treatments evolved characteristic shape features (e.g. narrower caudal peduncle) that consistently distinguish one treatment from the others through time and across experiments. Third, most of the differences between the treatments were qualitatively similar to the differences between the generations but were of lower magnitudes; these include differences in the dimensions of the caudal peduncle, the elongation of the body, the width of dorsal and the anal fins, and the relative area of the dorsal and the ventral sides.

There is a parsimonious explanation which accounts for all the observed features of the divergence between the treatments, without assuming any further selective pressures other than that responsible for the divergence between the generations. The basic idea is that the guppies in all the lines were evolving from the same starting shape (that of the founding population), and toward the same target shape (one with a narrower caudal peduncle, narrower dorsal and anal fins, and narrower dorsal side, compared to the starting shape). In all the lines, the body shapes of the guppies were selected for improved competitive ability (i.e. sustained swimming) at the expense of lower escaping ability (i.e. burst swimming). The shape differences between the lines are not because different selective pressures are taking place at each line, but, rather, because selection is acting at different rates in the different lines. Population densities fluctuated through time, and varied between the lines; the random treatment had the lowest densities during most of the experiment, except near the end. The negative treatment showed a substantial increase in density during the first half of the experiment, then decreased substantially during the second half. The positive treatment increased then decreased in absolute numbers, but it had intermediate densities (compared to the two other treatments) during most of the experiment (Diaz Pauli, unpublished data). The intensity of selection for enhanced competitive ability is expected to vary with population density (i.e. with the abundance of potential competitors), therefore, there are good reasons to expect selection on the shape to act in the same direction but at different rates in the three lines; if this is the case, the shape differences between the lines can be explained by their independent demographic histories. If the lines were evolving by different rates toward the same target shape, it is easy to see that a snapshot taken at an early stage of the experiment (when the fast-evolving populations have changed considerably while the slow-evolving populations have

changed only slightly) will detect more shape variation between the lines than a snapshot taken at the end of the experiment (when all the populations, whether fast- or slow-evolving, have had sufficient time to respond to selection). The absence of a consistent, line-specific shape features is to be expected if the life-history changes induced by the experimental treatments did not affect the shape features. The similarity between the changes through time and the differences between the lines is to be expected given that the latter is a by-product of the former, that is, the differences between the lines are the consequences of the different rates by which adaptation to the common experimental conditions took place in each line.

4.3   Concluding remarks

Above, we reported a number of potential evolutionary changes in the morphology of male guppies that occurred during the selection experiment. What remained unexplored is the relevance of these findings to the broader issues that motivated the guppy selection experiment and guided its design. Though interesting in themselves, the true value of our experimental findings lies in what they reflect about evolution in the wild. Here, we investigate the lessons (or, more accurately, the hints) provided by our study, focusing on the topics of fisheries-induced evolution, and contemporary evolution in general.

Consistent with previous studies (e.g. Walsh *et al*., 2006; Nannini *et al*., 2011), the present study suggests that, in addition to the traits that are directly subjected to fishing selection (i.e. size, vulnerability), other traits can also evolve as a result of extensive fishing, driven by indirect selection that can be mediated either by the life history changes induced by fishing (probably the cause of the observed changes in orange coloration) or the ecological effects of fishing (probably the cause of the observed changes in the shape features). The tradeoffs involving life history traits (e.g. between growth and reproductive effort) have received most research effort, because they are viewed as the main target of fishing selection (Heino *et al*., 2015). Although this might be true, the present study (together with earlier selection experiments on the guppy and the largemouth bass; see the introduction: section 1.1.1.2.) suggests that other evolutionary tradeoffs might be sensitive to the unnatural conditions imposed by fishing. The color traits of guppies represent a tradeoff between sexual selection and natural selection (Endler, 1983), while the shape traits represent a tradeoff between steady and unsteady swimming (Langerhans and Reznick, 2009); both tradeoffs are not directly affected by size-selective harvesting, despite this, both color and shape have shown evolutionary changes during the selection experiment. The specific changes observed during the experiment may not be directly relevant to the harvested species in the wild, since the observed changes were heavily influenced by the conditions in the experimental environment. However, the general possibilities suggested by the experiment that reproductive effort (reflected by male color) and swimming performance (reflected by male shape) can evolve under harvesting selection may apply to wild populations; both possibilities were previously discussed as potential consequences of harvesting (Reznick and Ghalambor, 2005; Walsh *et al*., 2006). Moreover, the possibility that sexually-selected traits such as color might undergo fisheries-

induced evolution is relevant to the assessment of the potential consequences of fishing on the mating systems of the harvested species (Rowe and Hutchings, 2003).

For both color and shape features, the selective effects of the harvesting regimes interacted with their demographic consequences. For male color, the fluctuations in the orange coloration that occurred in the positive line were explained as responses to the temporal variation in abundance levels (and therefore in the intensity of competition for resources). For male shape, the shape differences between the lines were explained as consequences of the different rates by which adaptation to the experimental environment took place in each line; these different rates were, in turn, explained as consequences of the unique demographic history (i.e. abundance levels through time) of each line. These results suggest that the demographic and the evolutionary effects of fishing should not always be viewed as alternative explanatory hypotheses, as is often the case in the debates on fisheries-induced evolution (e.g. Kuparinen and Merila, 2007). The radical demographic consequences of fishing on the exploited species (i.e. reductions in population sizes, changes in population sex ratios) can induce evolutionary changes in traits related to competitive ability, mating success and other traits as well.

As noted earlier, the unique feature of the guppy selection experiment is that its complex setup allows for unanticipated responses to occur and alter the predicted responses to the experimental treatments. The most interesting of these unanticipated responses was observed for the shape traits, which was the narrowing of the caudal peduncle that occurred through time. Because this change took place in all the lines, it reflects adaptation to the experimental conditions rather than to the experimental treatments, and was interpreted as an adaptive shift toward improving the competitive ability at the expense of the escaping ability. This shift occurred because, with respect to the shape traits, the experimental environment simulated the effect of the release from predation pressure (Note that the harvesting regimes were not selective for escaping performance, and therefore shape). Similar shift toward narrower caudal peduncle (in addition to changes in trophic morphology and foraging performance) was observed by Palkovacs *et al*. (2011) in their study of the evolution of trophic traits in high-predation guppies that were introduced to low-predation habitat. Consistent with the present study, they concluded that the release from predation pressure led to an enhancement in trophic performance due to the increase in the relative importance of competitive intra-specific interactions. The results of our study and that of Palkovacs *et al*. (2011), point to a relatively overlooked aspect of fisheries-induced evolution, namely, that the harvesting of predatory fishes can induce evolutionary changes not only in the harvested species themselves, but also in their prey. The expected direction of the changes is toward an increase in the resource-acquisition ability of the prey species and, consequently, in their impact on the lower trophic levels in the community.

A characteristic feature of the observed differences between the lines, both in color and in shape, is that they were not persistent, but fluctuating. For the positive line, the amount of orange coloration increased, then decreased, during the experiment. Likewise, the shape differences between the lines increased, then decreased, during the experiment. As a result, the separation between the lines was greater during the middle of the experiment than it was by the end. This indicates that the observed differences were due to rapid fluctuating selection rather than to

sustained directional selection. In the studies of contemporary evolution, episodes of directional selection (e.g. due to species introductions) tend to attract more attention (e.g. Stockwell *et al*., 2003), because of their ability to induce fixed changes, thereby contributing to macro-evolutionary patterns such as speciation. Recently, however, there has been a growing realization that fluctuating selection might also be important, as it can contribute to the eco-evolutionary dynamics that takes place during contemporary time scales, with the potential to stabilize or disturb biological communities. Thus, rapid fluctuating selection could have an important ecological effect, even though it does not contribute to long-term evolutionary change (Thompson, 1998; Schoener, 2011). We recommend that future studies of fisheries-induced evolution should explore the possibility of evolution due to fluctuating selection pressures that are induced by the demographic effects of fishing activities; these investigations can complement the existent body of research on the stable, potentially long-term consequences of fishing (e.g. reduction in the age and size at maturity).

References:

Adams, D. C., Rohlf, F. j. and Slice, D. E. 2004. Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology* **71**: 5-16.

Alexander, H. J., Taylor, J. S., Wu, S. S-T. and Breden, F. 2006. Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple temporal and spatial scales. *Evolution* **60**: 2352-2369.

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**: 9987-9994.

Ariyomo, T. O., Carter, M. and Watt, P. J. 2013. Heritability of boldness and aggressiveness in the zebrafish. *Behavioral genetics* **43**:161–67.

Baquero, F. and Blazquez, J. 1997. Evolution of antibiotic resistance. *Trends in Ecology and Evolution* **12**: 482-487.

Barot, S., Heino, M., O'Brien, L. and Dieckmann, U. 2004. Long-term trend in the maturation reaction norm of two cod stocks. *Ecological Applications* **14**: 1257-1271.

Billerbeck, J. M., Lankford, T. E. and Conover, D. O. 2001. Evolution of intrinsic growth and energy acquisition rates. I. Trade-offs with swimming performance in *Menidia menidia*. *Evolution* **55**: 1863-1872.

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**: 2919-2922.

Bookstein, F. L. 1991. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, Cambridge.

Brooks, R. 2000. Negative genetic correlation between male sexual attractiveness and survival. *Nature* **406**: 67-70.

Brooks, R. and Endler, J. A. 2001. Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). *Evolution* **55**: 1002-1015.

Brown, C. J., Hobday, A. J., Zeigler, P. E. and Welsford, D. C. 2008. Darwinian fisheries science needs to consider realistic fishing pressures over evolutionary time scales. *Marine Ecology Progress series* **369**: 257-266.

Burns, J. G. 2008. The validity of three tests of temperament in guppies (*Poecilia reticulata*). *Journal of Comparative Psychology* **122**: 344-356.

Burns, J. G., Nardo, P. D. and Rodd, F. H. 2009. The role of predation in variation in body shape in guppies *Poecilia reticulata*: a comparison of field and common garden phenotypes. *Journal of Fish Biology* **75**: 1144-1157.

Chappell, M. and Odell, J. 2004. Predation intensity does not cause microevolutionary change in maximum speed or aerobic activity in Trinidadian guppies (*Poecilia reticulata* Peters). *Physiological and Biochemical Zoology* **77**: 27-38.

Chervet, N., Zottl, M., Schurch, R., Taborsky, M. and Heg, D. 2011. Repeatability and heritability of behavioural types in a social cichlid. *International Journal of Evolutionary Biology*, doi: 10.4061/2011/321729.

Conover, D. O. and Baumann, H. 2009. The role of experiments in understanding fishery-induced evolution. *Evolutionary Applications* **2**: 276-290.

Conover, D. O. and Munch, S. B. 2002. Sustaining fisheries yields over evolutionary time scales. *Science* **297**: 94-96.

Conover, D. O., Munch, S. B. and Arnott, S. A. 2009. Reversal of evolutionary downsizing caused by selective harvest of large fish. *Proceedings of the Royal Society B* **276**: 2015-2020.

Conover, D. O., Arnott, S. A., Walsh, M. R. and Munch, S. B. 2005. Darwinian fishery science: lessons from the Atlantic silverside (*Menidia menidia*). *Canadian Journal of fisheries and Aquatic Science* **62**: 730-737.

Cook, S. J., Suski, C. D., Ostrand, K. G., Wahl, D. H. and Philipp, D. P. 2007. Physiological and behavioral consequences of long-term artificial selection for vulnerability to recreational angling in a teleost fish. *Physiological and Biochemical Zoology* **80**: 480-490.

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**: 952-954.

Darwin, C. 1958. *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. Murray, London.

De Roos, A. M., Boukal, D. S. and Persson, L. 2006. Evolutionary regime shifts in age and size at maturation of exploited fish stocks. *Proceedings of the Royal Society B* **273**: 1873-1880.

Devine, J. A., Wright, P. J., Pardoe, H. and Heino, M. 2012. Comparing rates of contemporary evolution in life-history traits for exploited fish stocks. *Canadian Journal of Fisheries and Aquatic Sciences* **69**:1105-1120.

Diaz Pauli, B. 2012. Contemporary evolution caused by fisheries: contribution from experimental studies. Ph.D. thesis, University of Bergen, Norway.

Diaz Pauli, B. and Heino, M. 2014. What can selection experiments teach us about fisheries-induced evolution? *Biological Journal of the Linnean Society* **111**: 485-503.

Diaz Pauli, P., Wiech, M., Heino, M. and Utne-Palm, A. C. 2015. Opposite selection on behavioral types by active and passive fishing gear. *Journal of Fish Biology* **86**: 1030-1045.

Dieckmann, U. and Heino, M. 2007. Probabilistic maturation reaction norms: their history, strengths, and limitations. *Marine Ecology Progress Series* **335**: 253-269.

Domenici, P. 2003. Habitat, body design and swimming performance of fish. In: *vertebrate biomechanics and evolution*, Bels, V. L., Gasc, J. P. and Casinos, A. (eds.). BIOS Scientific Publishers, Oxford.

Dowdall, J. T., Handelsman, C. A., Reznick, D. N. and Ghalambor, C. K. 2012. Local adaptation and the evolution of phenotypic plasticity in Trinidadian guppies (*Poecilia reticulata*). *Evolution* **66**: 3432-3443.

Drake, M. T., Claussen, J. E., Philipp, D. P. and Pereira, D. L. 1997. A comparison of bluegill reproductive strategies and growth among lakes with different fishing intensities. *North American Journal of Fisheries Management* **17**: 496-507.

Dunlop, E. S., Heino, M. and Dieckmann, U. 2009. Eco-genetic modeling of contemporary life-history evolution. *Ecological Applications* **19**: 1815–1834.

Edeline, E., Carlson, S. M., Stige, L. C., Winfield, I. J., Fletcher, J. M., James, J. B., 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**: 15799-15804.

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**:1-25.

Endler, J. A. 1978. A predator's view of animal color patterns. *Evolutionary Biology* **11**: 319-364.

Endler, J. A. 1980. Natural selection on colour patterns in *Poecilia reticulata*. *Evolution* **34**: 76-91.

Endler, J. A. 1982. Convergent and divergent effects of natural selection on color patterns in two fish faunas. *Evolution* **36**: 178-188.

Endler, J. A. 1983. Natural and sexual selection on color pattern in poeciliid fishes. *Environmental Biology of Fishes* **9**: 173-190.

Endler, J. A. 1991. Variation in the appearance of guppy color patterns to guppies and their predators under different visual conditions. *Vision research* **31**: 587-608.

Endler, J. A. and Houde, A. E. 1995. Geographic variation in female preferences for male traits in *Poecilia reticulata*. *Evolution* **49**: 456-468.

Endler, J. A. and Mielke, P. W. 2005. Comparing entire colour patterns as birds see them. *Biological Journal of the Linnean society* **86**: 405-431.

FAO (Food and Agriculture Organization of the United Nations). 1999. The state of world fisheries and Aquaculture in 1998. Food and Agriculture Organization of the United Nations, Rome.

Fenberg, P. B. and Roy, K. 2008. Ecological and evolutionary consequences of size-selective fishing: how much do we know? *Molecular Ecology* **17**: 209-220.

Ghalambor, C. K., Reznick, D. N. and Walker, J. A. 2004. Constraints on adaptive evolution: the functional tradeoff between reproduction and fast-start swimming performance in the Trinidadian guppy (*Poecilia reticulata*). *American Naturalist* **164**: 38-50.

Ghalambor, C. K., Walker, J. A. and Reznick, D. N. 2003. Multi-trait selection, adaptation, and constraints on the evolution of burst swimming performance. *Integrative and Comparative Biology* **43**: 431-438.

Godin, J. J. and McDonough, H. E. 2003. Predator preference for brightly colored males in the guppy: a viability cost for a sexually selected trait. *Behavioral Ecology* **14**: 194-200.

Gordon, S. P., Reznick, D., Arendt, J. D., Roughton, A., Ontiveros Hernandez, M. N., Bentzen, P. and Lopez-Sepulcre, A. 2015. Selection analysis on the rapid evolution of a secondary sexual trait. *Proceedings of the Royal Society B*, doi: 10.1098/rspb.2015.1244.

Gotanda, K. M. and Hendry, A. P. 2014. Using adaptive traits to consider potential consequences of temporal variation in selection: male guppy colour through time and space. *Biological Journal of the Linnean Society* **112**: 108-122.

Grether, G. F. 2000. Carotenoid limitation and mate preference evolution: a test for the indicator hypothesis in guppies (*Poecilia reticulata*). *Evolution* **54**: 1712-1724.

Grether, G. F., Cummings, M. E. and Hudon, J. 2005. Countergradient variation in the sexual coloration of guppies (*Poecilia reticulata*): drosopterin synthesis balances carotenoid availability. *Evolution* **59**: 175–188.

Griffith, S. C., Parker, T. H. and Olson, V. A. 2006. Melanin- vs. carotenoid-based sexual signals: is the difference really so black and red? *Animal Behaviour* **71**: 749–763.

Hamley, J. M. 1975. Review of gillnet selectivity. *Journal of the Fisheries Research Board of Canada* **32**: 1943-1969.

Haskins, C. P., Haskins, E. F., McLaughlin, J. J. A. and Hewitt, R. E. 1961. Polymorphism and population structure in *Lebistes reticulatus*, a population study. In: *Vertebrate speciation*, W F Blair (ed.). University of Texas Press, Austen.

Heino, M. and Dieckmann, U. 2009. Fisheries-induced evolution. In: *Encyclopedia of Life Sciences (ELS)* John Wiley & Sons, Ltd, Chichester.

Heino, M. and Godø, O. R. 2002. Fisheries-induced selection pressures in the context of sustainable fisheries. Interim report, International Institute for Applied Systems Analysis (IIASA).

Heino, M. and Kaitala, V. 1999. Evolution of resource allocation between growth and reproduction in animals with indeterminate growth. *Journal of Evolutionary Biology* **12**: 423-429.

Heino, M., Diaz-Pauli, B. and Dieckmann, U. 2015. Fisheries-induced evolution. *Annual Review of Ecology, Evolution and Systematics* **46**: 461-480.

Hemmer-Hansen, J., Therkildsen, N. O., Meldrup, D. and Nielsen, E. E. 2014. Conserving marine biodiversity: insights from life-history trait candidate genes in Atlantic cod (*Gadus morhua*). *Conservation genetics* **15**: 213-228.

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**: 20-29.

Hendry, A. P., Kelly, M. L., Kinnison, M. T. and Reznick, D. N. 2006. Parallel evolution of the sexes? Effects of predation and habitat features on the size and shape of wild guppies. *Journal of Evolutionary Biology* **19**: 741-754.

Herron, J. C. and Freeman, S. 2014. *Evolutionary Analysis*, fifth edition. Pearson Education, Boston.

Hilborn, R. 2006. Faith-based fisheries. *Fisheries* **31**: 554-555.

Hilborn, R. and Minte-Vera, C. V. 2008. Fisheries-induced changes in growth rates in marine fisheries: are they significant? *Bulletin of Marine Science* **83**: 95-105.

Houde, A. E. (1987). Mate choice based upon naturally occurring color-pattern variation in a guppy population. *Evolution* **41**: 1-10.

Houde, A. E. and Endler, J. A. 1990. Correlated evolution of female mating preferences and male color patterns in the guppy *Poecilia reticulata*. *Science* **248**: 1405-1408.

Houde, A. E. and Torio, A. J. 1992. Effect of parasite infection on male color pattern and female choice in guppies. *Behavioral Ecology* **3**: 346-351.

Hutchings, J. A. and Rowe, S. 2008. Consequences of sexual selection for fisheries-induced evolution: an exploratory analysis. *Evolutionary Applications* **1**: 129-136.

Hutchings, J. A. and Frasier, D. J. 2008. The nature of fisheries- and farming-induced evolution. *Molecular Ecology* **17**: 294-313.

Izyumov, Y. G., Gerasimov, Y. V. and Lapshin, O. M. 2002. The effect of fishing on the walleye pollock phenotype composition. Theme session on: The effect of fishing on the genetic composition of living marine resources Y: 09, The International Council for the Exploration of the Sea (ICES).

Jachmann, H., Berry, P. S. M. and Immae, H. 1995. Tusklessness in African elephants: A future trend. *African Journal of Ecology* **33**: 230-235.

Jayawickrama, R. 2013. Consequences of size-dependent harvest in guppy: study of morphological traits and their heritability. Master's Thesis, University of Bergen, Norway.

Jirotkul, M. 1999. Operational sex ratio influences female preference and male-male competition in guppies. *Animal Behavior* **58**: 287-294.

Jirotkul, M. 2000. Operational sex ratio influences the opportunity for sexual selection in guppies. *Journal of Fish Biology* **56**: 739-741.

Jørgensen, C., Enberg, K., Dunlop, E. S., Arlinghaus, R., Boukal, D. S., Brander, K., Ernande, B., Gårdmark, A. G., Johnston, F., Matsumura, S., Pardoe, H., Raab, K., Silva, A., Vainikka, A., Dieckmann, U., Heino, M. and Rijnsdorp, A. 2007. Managing evolving fish stocks. *Science* **318**: 1247-1248.

Karim, N., Gordon, S. P., Schwartz, A. K. and Hendry, A. P. 2007. This is not a déjà vu all over again: male guppy colour in a new experimental introduction. *Journal of Evolutionary Biology* **20**: 1339-1350.

Kasperski, W. and Kozlowski, J. 1993. Effect of exploitation on size at maturity in laboratory populations of guppies *Poecilia reticulata* (Peters). *Acta Hydrobiologica* **35**: 65-72.

Kemp, D. J., Reznick, D. N. and Grether, G. F. 2008. Ornamental evolution in Trinidad guppies (*Poecilia reticulata*): insights from sensory processing-based analysis of entire colour patterns. *Biological Journal of the Linnean Society B* **95**: 734-747.

Klingenberg, C. P. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* **11**: 353-357. Available on: http://www.flywings.org.uk/morphoj_page.htm.

Klingenberg, C. P. 2013. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. *The Italian Journal of Mammalogy* **24**: 15-24.

Kodric-Brown, A. 1989. Dietary carotenoids and male mating success in the guppy: an environmental component to female choice. *Behavioral Ecology and Sociobiology* **25**: 393-401.

Kraak, S. B. M. 2007. Does the probabilistic maturation reaction norm approach disentangle phenotypic plasticity from genetic change? *Marine Ecology Progress series* **335**: 295-300.

Kuparinen, A. and Merila, J. 2007. Detecting and managing fisheries-induced evolution. *Trends in Ecology and Evolution* **22**: 652-659.

Langerhans, R. B. 2009. Trade-off between steady and unsteady swimming underlies predator-driven divergence in *Gambusia affinis*. *Journal of Evolutionary Biology* **22**: 1057-1075

Langerhans, R. B. and DeWitt, T. J. 2004. Shared and unique features of evolutionary diversification. *American Naturalist* **164**: 335-349.

Langerhans, R. B. and Reznick, D. N. 2009. Ecology and evolution of swimming performance in fishes: predicting evolution with biomechanics. In: *Fish locomotion: an etho-ecological perspective*, Domenici, P. and Kapoor, B. G. (eds.). Science publishers, Enfield.

Langerhans, R. B., Layman, C. A., Shokrollahi, A. M. and DeWitt, T. J. 2004. Predator-driven phenotypic diversification in *Gambusia affinis*. *Evolution* **58**: 2305-2318.

Lankford, T. E., Billerbeck, J. M. and Conover, D. O. 2001. Evolution of intrinsic growth and energy acquisition rates. II. Trade-offs with vulnerability to predation in *Menidia menidia*. *Evolution* **55:** 1873-1881.

Law, R. 2000. Fishing, selection and phenotypic evolution. *ICES Journal of Marine Science* **57**: 659-668.

Lele, S. R., and Richtsmeier, J. T. 2001. *An invariant approach to statistical analysis of shapes*. Chapman & Hall/CRC, Boca Raton.

Magurran, A. E., Seghers, B. H., Carvalho, G. R. and Shaw, P. W. 1992. Behavioral consequences of an artificial introduction of guppies (*Poecilia reticulata*) in North Trinidad: evidence for the evolution of antipredator behavior in the wild. *Proceedings of the Royal Society of London B* **248**: 117-122.

Majerus, M. E. N. 1998. *Melanism: evolution in Action*. Oxford University Press, Oxford.

Marshall, C. T. and Browman, H. I. 2007. Disentangling the causes of maturation trends in exploited fish populations. *Marine Ecology Progress Series* **335**:249–51.

Marshall, C. T. and McAdam, B. J. 2007. Integrated perspectives on genetic and environmental effects on maturation can reduce potential for errors of inference. *Marine ecology progress series* **335**: 301-310.

Millar, N. P., Reznick, D. N., Kinnison, M. T. and Hendry, A. P. 2006. Disentangling the selective factors that act on male colour in wild guppies. *Oikos* **113**: 1-12.

Mousseau, T. A. and Roff, D. A. 1987. Natural selection and the heritability of fitness components. *Heredity* **59**: 181-197.

Nannini, M. A., Wahl, D. H., Philipp, D. P. and Cooke, S. J. 2011. The influence of selection for vulnerability to angling on foraging ecology in largemouth bass *Micropterus salmoides. Journal of Fish Biology* **79**: 1017-1028.

Olsen, E. M. and Moland, E. 2011. Fitness landscape of Atlantic cod shaped by harvest selection and natural selection. *Evolutionary Ecology* **25**: 695-710.

Olsen, E. M., Heino, M., Lilly, G. R., 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**: 932-935.

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**: e18879.

Palumbi, S. R. 2001. Humans as the world's greatest evolutionary force. *Science* **293**: 1786-1790.

Pandolfi, J. M. 2009. Evolutionary impacts of fishing: overfishing's 'Darwinian debt'. *Biology Reports*, doi: 10.3410/Bl-43.

Pardoe, H., Vainikka, A., Thordarson, G., Marteinsdottir, 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**: 1719-173.

Pettersson, L. B., Ramnarine, I. W., Becher, S. A., Mahabir, R. and Magurran, A. E. 2004. Sex ratio dynamics and fluctuating selection pressures in natural populations of the Trinidadian guppy, *Poecilia reticulata. Behavioral Ecology and Sociobiology* **55**: 461-468.

Philipp, D. P., Cooke, S. J., Claussen, J. E., Koppelman, J., Suski, C. D. and Burkett, D. 2009. Selection for vulnerability to angling in largemouth bass. *Transactions of the American Fisheries Society* **138**: 189-199.

Pilastro, A., Simonato, M., Bisazza, A. and Evans, J. P. 2004. Cryptic female preference for colorful males in guppies. *Evolution* **58**: 665-669.

Quinn, T. P., McGinnity, P. and Cross, T. F. 2006. Long-term declines in body size and shifts in run timing of Atlantic salmon in Ireland. *Journal of Fish Biology* **68**:1713–1730.

Quinn, T. P., Hodgson, S., Flynn, L., Hilborn, R. and Rogers, D. E. 2007. Directional selection by fisheries and the timing of sockeye salmon (*Oncorhynchus nerka*) migrations. *Ecological Applications* **17**:731–39.

Rasband, W. S. 1997. ImageJ: Image processing and analysis in Java. National institute of mental health, Bethesda, Maryland. Available on: rsb.info.nih.gov/ij/.html.

Ratner, S. and Lande, R. 2001. Demographic and evolutionary consequences to selective harvesting in populations with discrete generations. *Ecology* **82**: 3093-3104.

R development Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. Available on: https://www.r-project.org/

Redpath, T. D., Cooke, S. J., Aringhaus, R., Wahl, D. H. and Philipp, D. P. 2009. Life-history traits and energetic status in relation to vulnerability to angling in an experimentally selected teleost fish. *Evolutionary Applications* **2**: 312-323.

Regier, H. A. 1969. Fish size parameters useful in estimating gillnet selectivity. *Progressive Fish Culturist* **31**: 57-59.

Reis, E. G. and Pawson, M. G. 1999. Fish morphology and estimating selectivity by gillnets. *Fisheries Research* **39**: 263-273.

Reynolds, J. D. and Gross, M. R. 1992. Female mate preference enhances offspring growth and reproduction in a fish, *Poecilia reticulata. Proceedings of the Royal Society of London B* **250**: 57-62.

Reznick, D. N. 1997. Life-history evolution in guppies (*Poecilia reticulata*): guppies as a model for studying the evolutionary biology of aging. *Experimental Gerontology* **32**: 245-258.

Reznick, D. N. and Bryga, H. 1987. Life-history evolution in guppies. 1. Phenotypic and genotypic changes in an introduction experiment. *Evolution* **41**: 1370-1385.

Reznick, D. N. and Bryga, H. 1996. Life-history evolution in guppies (*Poecilia reticulata*: Poeciliidae). V. Genetic basis of parallelism in life histories. *American Naturalist* **147**: 339-359.

Reznick, D. and Endler, J. A. 1982. The impact of predation on life history evolution in Trinidadian guppies (*Poecilia reticulata*). *Evolution* **36**: 160-177.

Reznick, D. N. and Ghalambor, C. K. 2005. Can commercial fishing cause evolution? Answers from guppies (*Poecilia reticulata*). *Canadian Journal of Fisheries and Aquatic Sciences* **62**: 791-801.

Reznick, D. A., Bryga, H. and Endler, J. A. 1990. Experimentally induced life-history evolution in a natural population. *Nature* **346**: 357-359.

Reznick, D. N., Rodd, F. H. and Cardenas, M. 1996b. Life-history evolution in guppies (*Poecilia reticulata*: Poeciliidae). IV. Parallelism in life-history phenotypes. *American Naturalist* **147**: 319-338.

Reznick, D. N., Butler, M. J., Rodd, F. H. and Ross, P. 1996a. Life-history evolution in guppies (*Poecilia reticulata*). 6. Differential mortality as a mechanism for natural selection. *Evolution* **50**: 1651-1660.

Reznick, D. N., Shaw, F. H., Rodd, F. H. and Shaw, R. G. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**: 1934-1937.

Ricker, W. E. 1981. Changes in the average size and average age of pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences* **38**: 1636-1656.

Rijnsdorp, A. D. 1993. Fisheries as a large-scale experiment on life-history evolution: disentangling phenotypic and genetic effects in changes in maturation and reproduction of North Sea Plaice *Pleuronectes platessa* L. *Oecologia* **96**: 391-401.

Rijnsdorp, A. D., Grift, R. E. and Kraak, S. B. M. 2005. Fisheries-induced adaptive change in reproductive investment in North Sea plaice (*Pleuronectes platessa*)? *Canadian Journal of Fisheries and Aquatic Sciences* **62**:833–843.

Rodd, F. H. and Reznick, D. N. 1997. Variation in the demography of guppy populations: the importance of predation and life histories. *Ecology* **78**: 405-418

Roff, D. A. 1997. *Evolutionary quantitative genetics*. Chapman & Hall, New York.

Rohlf, F. J. 2004. tpsDig2, digitize landmarks and outlines, version 2.0. Department of Ecology and Evolution, State University of New York at Stony Brook. Available on: http://life.bio.sunysb.edu/morph/.

Rohlf, F. J. 2009. tpsUtility program. Department of Ecology and Evolution, State University of New York at Stony Brook. Available on: http://life.bio.sunysb.edu/morph/.

Rothschild, M. 1975. Remarks on carotenoids in the evolution of signals. In: *Coevolution of animals and plants*. Gilbert, L. E. and Raven, P. H. (eds.). University of Texas Press, Austin.

Rowe, S. and Hutchings, J. A. 2003. Mating systems and the conservation of commercially exploited marine fish. *Trends in Ecology and Evolution* **18**: 567-572.

Ruell, E. W., Handelsman, C. A., Hawkins, C. L., Sofaer, H. R., Ghalambor, C. K. and Angeloni, L. 2013. Fear, food and sexual ornamentation: plasticity of colour development in Trinidadian guppies. *Proceedings of the Royal Society B* **280**: 20122019.

Schoener, T. W. 2011. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* **331**: 426-429.

Smith, P. J., Francis, R. I. C. C. and McVeagh, M. 1991. Loss of genetic diversity due to fishing pressure. *Fisheries Research* **10**: 309-316.

Stockwell, C. A., Hendry, A. P. and Kinnison, M. T. 2003. Contemporary evolution meets conservation biology. *Trends in Ecology and Evolution* **18**: 94-101.

Sutter, D. A. H., Suski, C. D., Philipp, D. P., Klefoth, T., Wahl, D. H., Kersten, P., Cooke, S. J. and Aringhaus, R. 2012. Recreational fishing selectively captures individuals with the highest fitness potential. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 20960-20965.

Swain, D. P., Sinclair, A. F. and Hanson, J. M. 2007. Evolutionary response to size-selective mortality in an exploited fish population. *Proceedings of the Royal Society B* **274**: 1015–22.

Thompson, J. N. 1998. Rapid evolution as an ecological process. *Trends in Ecology and Evolution* **13**: 329-332.

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 of the Royal Society B* **276**: 2195-2208.

Trippel, E. A. 1995. Age at maturity as a stress indicator in fisheries. *Bioscience* **45**: 759-771.

Valeri, C. J., Cole, T. M., Lele, S. R. and Richtsmeier, J. T. 1998. Capturing data from three-dimensional surfaces using fuzzy landmarks. *American Journal of Physical Anthropology* **107**:113-124.

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**:85–93.

Van Wijk, S. J., Taylor, M. I., Creer, S., Dreyer, C., Rodriguez, 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**: 181-187.

Walker, J. A. 1997. Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (gasterosteidae) body shape. *Biological Journal of the Linnean Society B* **61**: 3-50.

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**: 142-148.

Webb, P. W. 1984. Body form, locomotion, and foraging in aquatic vertebrates. *American Zoologist* **24**: 107-120.

Webster, M. and Sheets, H. D. 2010. A practical introduction to landmark-based geometric morphometrics. *The paleontological society papers* **16**: 163-188.

Weese, D. J., Gordon, S. P., Hendry, A. P. and Kinnison, M. T. 2010. Spatiotemporal variation in linear natural selection on body color in wild guppies (*Poecilia reticulata*). *Evolution* **64**: 1802-1815.

Winemiller, K. O., Leslie, M. and Roche, R. 1990. Phenotypic variation in male guppies from natural inland populations: an additional test of Haskins' sexual selection/predation hypothesis. *Environmental Biology of fishes* **29**: 179-191.