The roles of divergent selection and environmental heterogeneity in speciation of Lake Victoria cichlid fish

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Isabel S. Magalhães
von Portugal

Leiter der Arbeit:
Prof. Dr. Ole Seehausen

Institute of Ecology and Evolution
Division of Aquatic Ecology
Universität Bern

The roles of divergent selection and environmental heterogeneity in speciation of Lake Victoria cichlid fish

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Isabel S. Magalhães
von Portugal

Leiter der Arbeit:
Prof. Dr. Ole Seehausen

Institute of Ecology and Evolution
Division of Aquatic Ecology
Universität Bern

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen

Bern, 17 February 2009

Der Dekan:
Prof. Dr. Urs Feller

Contents

Chapter 1

Introduction and Synthesis

Introduction and Synthesis

This thesis wants to contribute to the understanding of the environmental conditions under which (ecological or sexual) selection can cause phenotypic divergence and speciation. It starts with an introduction to the problem of speciation in the presence of gene flow. This is followed by theoretical support and empirical evidence for the role of disruptive ecological and sexual selection as mechanisms of speciation in sympatry and for the role of environmental heterogeneity in this process. I then introduce the study system and provide a summary of the chapters. This is followed by five chapters, followed by a summary and conclusions.

1. Speciation in the presence of gene flow

One of the least well understood problems in evolutionary biology is why some taxa have speciated while others often closely related and superficially similar fail to do so. The advent of molecular techniques combined with increasingly powerful computer simulations and the development of a mathematical framework for speciation genetics (Gavrilets, 2004) have in recent years led to several paradigm shifts in speciation theory. One of these shifts is that speciation is no longer considered to be exclusively a by-product of drift and ecological adaptation in geographical isolation (Mayr, 1942). For many decades empirical evidence in animals failed to produce convincing evidence for speciation other than in geographical isolation. However, more recently empirical evidence for speciation in the absence of geographical isolation has accumulated (Bush 1994; Schluter, 2000; Via 2001; Dieckmann et al., 2004). Speciation in the presence of gene flow is now receiving a great deal of attention, and an intensive scientific debate has emerged about conditions leading to it. Two questions are central to this debate: one is how commonly and under which genetic and environmental conditions can divergent selection lead to speciation when gene flow is frequent (Nosil & Harmon, 2008; Rasanen & Hendry, 2008; Nosil et al., in press); the second is whether assortative mating evolves in response to ecological selection exerted by competition for ecological resources, competition for mating opportunities or by subtle interactions of the two. Increasingly sophisticated models of speciation by disruptive ecological and sexual selection or divergent selection along environmental gradients have been explored (Doebeli, 1996; Dieckmann & Doebeli, 1999; Kondrashov & Kondrashov, 1999; van Doorn et al., 2004; Kawata et al., 2007), but empirical testing of theory and model assumptions are lagging behind.

2. Ecological and sexual selection as mechanisms of sympatric speciation

Sympatric speciation in its strict definition is speciation in a population in which the site where an individual reproduces is independent of its site of birth (Gavrilets, 2004). As mating is random with respect to birth place, the proportion of members of the two parental populations that is exchanged per generation (i.e. migration rate) is expected to be 0.5. Sympatric speciation requires disruptive selection of some kind that places intermediate phenotypes at a fitness minimum and can generate the evolution of assortative mating. In terms of the biology, two main classes of theoretical models have emerged on sympatric speciation: models based on disruptive ecological selection involving competition for resources and models based on disruptive sexual selection.

Both types of models have been subjected to several criticisms (Arnegard & Kondrashov, 2004; Coyne & Orr, 2004; Turelli et al., 2001).

Empirical examples of sympatric speciation by sexual or ecological selection also suffer from difficulties in proving themselves as such. The most important difficulty is showing that population divergence is initiated by sexual or ecological selection in sympatry and was not preceded by an allopatric phase (Schluter, 2001). Nonetheless evidence for sympatric speciation by sexual or ecological selection is no longer negligible. However many empirical studies also illustrate that there are several alternative outcomes of ecological or sexual selection and speciation is only one of them (reviewed by Panhuis et al., 2001; Rueffler et al., 2006).

2a. Ecological selection - models -

Theoretical models of sympatric ecological speciation can be usually divided in two types: discrete-habitat models (Maynard Smith, 1966; Felsenstein, 1981; Rice, 1984; Rice, 1987; Diehl & Bush 1989; Johnson et al. 1996a; Kawecki, 1996; Kawecki, 1997; Fry, 2003; Gavrilets & Vose, 2007; Gavrilets et al., 2007) and continuous-resource models (Doebeli, 1996; Dieckmann & Doebeli, 1999; Kondrashov & Kondrashov, 1999; Doebeli & Dieckmann, 2000; Ito & Dieckmann, 2007). In the first type models sympatric populations, exploiting different niches, are most fit in their own niche and mate preferably with members of their own population. Therefore, they require the evolution of niche preference, niche adaptation and assortative mating. Sympatric speciation requires the development of at least two of these traits and there are models for all three pairwise combinations of them (reviewed by Coyne & Orr, 2004).

The second type of models assumes the simultaneous coevolution of ecological traits and assortative mating that splits the population into sympatric groups, each using different parts of an initially continuous resource distribution (Doebeli, 1996; Dieckmann & Doebeli, 1999; Kondrashov & Kondrashov, 1999; Doebeli & Dieckman, 2000; Ito & Dieckman, 2007). In these models competition between similar phenotypes produces negative frequency-dependent selection that confers higher fitness to the individuals at the extremes of the resource distribution. This can create intraspecific polymorphisms but if mating stays random, the population will not split into genetically differentiated subpopulations. Speciation is possible through the development of assortative mating, either based directly on the ecological trait that is under selection or based on a marker trait that may become associated with the ecological trait (Dieckmann & Doebeli, 1999; Kondrashov & Kondrashov, 1999). An alternative to this scenario is assortative mating exerted by females carrying genes that affect their preference for different types of the marker trait. A female's willingness to mate depends on the difference between their preference and the marker trait (Kondrashov & Kondrashov, 1999).

- Empirical data -

Good empirical examples supporting that disruptive ecological selection can initiate speciation in the absence of geographical isolation remain few. In vertebrates, examples of sympatric speciation come from cichlid fishes (Schliewen et al., 1994; Barluenga et al., 2006), palm trees (Savolainen et al., 2006) and more recently *Neospiza* finches from the Tristan da Cunha Archipelago (Ryan et al. 2007). However there is no conclusive evidence for disruptive ecological selection having initiated speciation.

threespine sticklebacks (*Gasterosteus aculeatus*) from coastal lakes in British Columbia, Canada (McPhail, 1994; Schluter 1996a; Schluter 1996b; Rundle et al. 2000; Bolnick 2004; Bolnick & Lau, 2008) and Darwin's finches (Grant & Grant, 2006; Hendry et al., 2008), on the other hand, are prominent examples of disruptive ecological selection as an important agent in divergence with gene flow. However there is no evidence for sympatric initiation of speciation in these organisms.

Other responses to disruptive ecological selection, illustrated by empirical studies, are quite abundant and include species sorting, evolution of phenotypic plasticity, single trait polymorphisms and sexual dimorphisms (reviewed by Rueffler et al,. 2006). The asymmetric lower jaws of the scale eating cichlid fish *Perissodus microlepis* (Hori, 1993) and beak size polymorphism in the African seed cracker finch *Perinestes ostrinus* (Smith, 1993) are examples of how disruptive ecological selection resulted in trophic polymorphisms rather than speciation. It is unclear if this was due to pre-existing genetic architectures of the traits under selection, or if these architectures evolved to resolve the antagonism between gene flow and disruptive selection. The difference in bill length and curvature between males and females of the humming bird *Eulampis jugularis* is an example of how ecological selection leads to sexual dimorphism (Temeles et al., 2000). The two sympatric morphs of Arctic charr *Salvelinus alpinus* from Loch Rannoch are an example of divergence in trophic morphology through phenotypic plasticity. An experimental study showed that variation in head anatomy between the two morphs is discrete and correlates with alternative feeding strategies, and although there is an underlying genetic effect influencing the phenotype, over 80% of the phenotypic differences are environmentally induced (Adams & Huntingford, 2004).

2b. Sexual selection - models -

Exclusively sexual selection is thought to be an even less likely mechanism of sympatric speciation than exclusively natural selection (Turelli et al., 2001; Kirkpatrick & Ravigne, 2002; Arnegard & Kondrashov, 2004; Coyne & Orr, 2004; Kirkpatrick & Nuismer, 2004). Many models of sympatric speciation by sexual selection involve disruptive selection exerted by female mating preferences for male traits at opposite ends of the trait value distribution (Turner & Burrows 1995; van Doorn et al. 1998; Higashi et al. 1999; Payne & Krakauer 1997; Takimoto et al. 2000). Related are models of sympatric speciation through sexual conflict (Gavrilets 2000a; Gavrilets & Waxman 2002; Gavrilets 2004) and sensory bias (Kawata et al., 2007).

Sexual conflict arises when the reproductive success of individuals of one sex reduces the success of the mating partner, leading to an evolutionary arms race between the sexes. Females may diversify genetically and split into separate clusters in order to avoid being harassed by too many compatible males. Under certain circumstances this can then lead to males also splitting into separate clusters, resulting in the formation of reproductively isolated clusters.

The term sensory bias has been applied when female preference and male trait evolve as by-product of sensory adaptation to the local environment. Divergence occurs because local environments impose selection of varying strengths and direction on the sensory systems of females and on male traits. Traditionally the sensory drive hypothesis was viewed as a case in which speciation occurs as a by-product of adaptation to geographically isolated habitats; however a very recent theoretical model suggested that

sensory drive can lead to speciation in a single heterogeneous habitat (Kawata et al., 2007). The model simulated the adaptive evolution of visual pigments that changes females` colour perception in response to an environmental light gradient. The visual pigments of females and consequently their colour perception will evolve to adapt to the local light environment and their mating decisions will be biased towards males with nuptial colouration that appears most conspicuous to the females. Speciation happened independently of male nuptial colour and visual sensitivity being initially skewed or not. Factors that influenced positively the likelihood of speciation were an intermediate steepness of the light gradient, large population size, small home ranges, small dispersal distance of the offspring and mutations causing major changes in visual sensitivity.

The major criticism to earlier models on sympatric speciation by sexual selection is that they provide no mechanism for the stabilization of morph frequencies during the speciation process and for post-speciation coexistence. Stabilization of morph frequencies is a critical problem since it requires not only maintaining genetic variation of the trait under selection but also maintaining genetic variability for mating preference. Disruptive sexual selection leading to mating preference for extreme phenotypes would rapidly deplete genetic variation, constraining speciation (Arnegard & Kondrashov, 2004). Also, if species diverge only in secondary sexual traits but not in ecological traits then according to the principle of competitive exclusion one species should outcompete the other. However, extinction can of course be prevented by ecological character displacement soon after speciation. Other criticisms surrounding models on sympatric speciation by sexual selection focus mostly on some of their assumptions, which have been considered unrealistic or restrictive: rare phenotypes or choosiness are not penalized, which may not be realistic if females prefer a rare male phenotype, or if the most attractive phenotypes are the most conspicuous ones (Higashi et al., 1999); initially symmetric allele frequencies at loci that control mating (Higashi et al., 1999; a criticism that applies to some models of ecological speciation too, e.g. Dieckmann & Doebeli, 1999); or a small number of genes controlling female preference and / or male trait being required for sympatric speciation to happen (Turner & Burrows, 1995; Payne & Krakauer, 1997; Takimoto et al. 2000; Van Doorn et al. 2001).

An increasing body of theoretical work has addressed the most critical issue, stabilization of trait polymorphisms, and demonstrated that under less stringent conditions than initially thought, sexual selection can drive and maintain polymorphisms in populations and even lead to speciation (Mikami et al., 2004; Seehausen & Schluter, 2004; van Doorn et al., 2004). According to these verbal and simulation models, the key to stabilization of polymorphisms is negative frequency-dependent selection that confers rare phenotypes an advantage. One such mechanism could be intrasexual selection if operating on the same traits as intersexual selection: if individuals direct more aggression towards rivals that phenotypically resemble themselves than towards different phenotypes, rare phenotypic varieties would enjoy a frequency-dependent selection advantage because they receive less aggression. This advantage would decrease as the rare phenotype becomes progressively more common and the number of intramorph encounters increases. An own-phenotype bias would therefore allow for the invasion of a new phenotype and also at the same time stabilize coexistence of the incipient species. Experimental studies showing male own-colour morph aggression bias (Dijkstra et al., 2005) and female own-colour morph aggression bias (Dijkstra et al., 2008), as well as

community assembly and distribution patterns (Seehausen & Schluter, 2004) in Lake Victoria cichlids provided support for these theories.

- empirical data -

Empirical evidence for sexual selection as an agent in divergence with gene flow is far less abundant than theory. Nonetheless, intraspecific sexual selection is maintaining polymorphisms in colouration in Papua New Guinea coral reef fish (Messmer et al. 2005) and African cichlids (Seehausen et al. 1999b; Turner et al. 2001; Kocher 2004). Disruptive sexual selection has also been suggested to form a barrier to gene flow, helping maintain the phenotypic distinctiveness of hybridizing species of *Heliconius* butterflies (Naisbit et al. 2001) and Lake Victoria cichlid fish (Stelkens et al., 2008; Van der Sluijs et al., 2008b). Examples of disruptive sexual selection arising as a by-product of natural selection on sensory properties has been shown to enhance reproductive isolation in threespine sticklebacks in British Columbia (Boughman 2001) and even lead to speciation in Lake Victoria cichlid fish (Seehausen et al., 2008 (chapter 4)). However, similarly to what happens with ecological selection, except for the last example (Seehausen et al., 2008 (chapter 4)), none of these examples constitutes evidence for sexual selection driving speciation.

3. Environmental heterogeneity

Spatial ecological structure in the environment has been recognized as a major factor facilitating divergent evolution between populations and eventually speciation. The interaction of divergent selection with limited migration on spatial gradients of ecologically relevant variables can result in complex geographic and genetic clusters.

"Clinal speciation" was first conceptualized by (Fisher 1930) but mathematical models only followed 50 years later (Balkau & Feldman 1973; Lande 1982). Theoretical work has simulated how spatially structured environments can lead to local adaptation and reduced gene flow. Initially, these models only assumed that forces responsible for divergence were imposed by the environment (Endler 1973; Nagylaki 1975; Nagylaki 1976; Barton 1979). Gradual spatial variation in environmental conditions may lead to gradual variation in traits involved in local adaptation. Along the environmental gradient, individuals with different variations of the trait under selection have different fitness. If individuals of similar phenotype prefer to mate with each other, the combination of the two factors may generate genetic clusters along the gradient.

More recent models of clinal speciation have added selection resulting from resource competition to spatially heterogeneous selection (Doebeli & Dieckmann, 2003; Leimar et al., 2008). The idea behind these models is that if spatial heterogeneity and resource competition each alone can lead to the formation of genetically isolated clusters of individuals, then the combination of the two will more easily lead to speciation. In these models competition generates negative frequency-dependent selection along an environmental gradient. They suggest that spatially localized interactions along environmental gradients can facilitate speciation through frequency-dependent selection and result in patterns of geographical structure between the emerging species that resemble those of geographical speciation. However, this has been the subject of some controversy (Polechova & Barton, 2005). Reanalysis of the asexual model demonstrated that clusters arise even when there is no competition. Reduced maladaptive gene flow at

the margin of the species range provides an advantage to marginal phenotypes creates an edge effect and generates discrete clusters.

Non-ecological models of speciation in spatially heterogeneous environments focus on the incompatibility of alleles generating Dobzansky-Muller type selection (Gavrilets, 1997; Gavrilets, 2000b; Gavrilets, 2004). If incompatible alleles from different populations come together, individuals carrying these combinations of alleles have zero or very low fitness generating "holes" between "ridges" of well fit genotypes in what as been called a "holey adaptive landscape".

Prominent examples of clinal morphological and genetic divergence along habitat gradients include the Caribbean lizard *Anolis roquet* (Odgen & Thorpe, 2002), the little Greenbull *Andropadus virens* in Cameroonian-forest ecotones (Smith et al., 1997; Smith et al., 2001; Smith et al., 2005) and the periwinkle *Littorina saxatilis* (reviewed by Butlin et al., 2008). These studies provide evidence for the role of habitat-specific selection in promoting population divergence, but not speciation.

Empirical studies on closely related species having hybrid zones that lie in areas of ecological transition such as the grasshoppers *Corthippus brunneus* and *Corthippus jacobsi* (Saldamando et al., 2005; Bridle et al., 2006) can be seen as examples of speciation in heterogeneous environments. However, it is sometimes difficult to distinguish if hybrid zones are the product of speciation in a heterogeneous environment or if they are the result of secondary contact of two allopatric species that have expanded their ranges (Coyne & Orr, 2004).

4. The African cichlid fish model system

African cichlid fish have become an important model system for the study of evolutionary diversification. Several reasons make them a unique system for the study of evolution (reviewed by Kocher, 2004): (1) several hundred species inhabit several of the African lakes despite indications that some lakes are geologically very young (Cohen et al., 1993; Johnson et al., 1996b; Johnson et al., 2000; Stager & Johnson, 2008), implying extremely rapid rates of speciation (MacCune & Lovejoy, 1998; Turner et al., 2001; Seehausen, 2002); (2) many of these species live sympatrically but maintain enormous diversity in morphology, colouration and behaviour (Fryer & Iles, 1972; Greenwood, 1974; Seehausen, 1996; Genner & Turner, 2005) and (3) hybrids of closely related species are viable and fertile indicating that genetic differences between cichlid species are maintained by pre-zygotic or extrinsic post-zygotic mechanisms (Capron de Caprona & Fritzsch, 1984; Seehausen et al., 1997; Van der Sluijs et al., 2008c).

I focus on the Lake Victoria haplochromine cichlids in this thesis. The Lake Victoria catchment formed approximately 400 000 years ago but since then it has dried up and refilled several times with the last dry period ending around 15 000 years ago (Johnson et al. 1996b; Stager & Johnson 2008). The cichlid mitochondrial haplotype diversity in Lake Victoria is approximately between 90 000 and 120 000 years old but the species radiation must be much more recent since lineage sorting is effectively absent (Genner et al., 2007). The lake contains an estimated 450 to 550 haplochromine cichlid species (Genner et al., 2004). Most of these must have arisen in the past 15 000 years, which makes them the fastest large-scale species radiation known. Several studies indicate that several ancestral species from the surrounding rivers and lakes may have colonized the lake, hybridized and subsequently radiated into a great number of species

differing in colouration, morphology, ecology and behaviour (Nagl et al., 2000; Seehausen et al., 2003; Seehausen 2004; Salzburger et al., 2005).

An increasing number of studies have shown that both sexual and natural selection have played a major role in the radiation of haplochromine cichlids. The high adaptability of cichlids` jaws and associated dentition plays an important role in species coexistence through resource partitioning (Bouton et al., 1999; Genner et al., 1999; Bouton et al., 2002; Genner et al., 2003). Ecological selection leading to divergence in morphology most likely played an important role in the early speciation events that generated the deeper splits in the phylogeny each cichlid radiation, because most morphological variation is partitioned between deeper clades in each radiation (Seehausen, 2000; Streelman & Danley, 2003). However, within genera numerous closely related species coexist differing little along ecological axes and having largely overlapping habitat distributions but display an amazing amount of diversity in colouration (Seehausen, 2000). Nuptial colour is one of the most compelling features of the haplochromine cichlid fishes and there is evidence that colour variation plays a central role in the evolution and maintenance of haplochromine species richness (Fryer & Iles, 1972; Seehausen et al., 1997; Seehausen, 2000; Knight & Turner, 2004; Seehausen & Schluter, 2004).

The polygynous mating system of haplochromine cichlids has been suggested to be conducive to strong sexual selection by female choice based on male colouration, which leads to speciation events (Seehausen et al., 1999a; Seehausen, 2000). Only dominant and territorial males of Lake Victoria haplochromines fully express their colouration which is used both by females to choose their mates, and by males to identify competitors, generating both intra and interspecific sexual selection on the same phenotypic trait (Seehausen & van Alphen, 1998; Maan et al., 2004; Stelkens et al., 2008; van der Sluijs et al., 2008a). Parallel to this, colour is also subjected to negative frequency-dependent intrasexual selection with individuals often, but not always, directing their aggression towards individuals with similar colour (Dijkstra et al., 2005; Dijkstra et al., 2006; Dijkstra et al., 2008). This evidence supports the hypothesis that negative-frequency dependent intrasexual selection generated by colour-based aggression bias combined with intersexual selection by colouration-based mate choice allows for the invasion of new colour phenotypes and stabilizes coexistence during and post-speciation, allowing for the unusually high levels of diversity.

Trophic morphology (Bouton et al., 1999), male nuptial colouration (Seehausen et al, 2000) and X-linked colouration (Seehausen et al., 1999) are all functional traits that characterize different species of cichlid fish but can also be observed as segregating polymorphisms within populations, and have therefore been implicated in the speciation of cichlid fish. The study of multiple populations that vary in these traits, which was done in this thesis, allowed testing the relative importance of divergent and disruptive selection on these polymorphisms during speciation with gene flow.

5. Thesis outline and summary of the chapters

Chapter 2 - Evidence for colour polymorphism as incipient stages of speciation

Stable sympatric phenotypic polymorphisms can be an incipient stage of sympatric speciation maintained largely by negative frequency-dependent selection acting on the speciation phenotypes (van Doorn et al., 2004). Seehausen et al. (1999) analysed the three colour morphs of a population of *Neochromis omnicaeruleus,* a rock-dwelling cichlid species from Lake Victoria, reported their frequencies for several years, noted the apparent lack of ecological differentiation between them (see also Seehausen & Bouton, 1997), and suggested they might be an incipient stage of sympatric speciation. Experimental studies on individuals of the same population also presented evidence for intraspecific frequency-dependent selection acting on colouration (Dijkstra et al., 2008). This makes this population of *N. omnicaeruleus* a suitable model system to study phenotypic polymorphisms and ask if disruptive selection, sexual or of other unknown source, can reduce gene flow between the colour morphs and permit differentiation in traits other than colour.

The three morphs present in the population of *N. omnicaeuruleus* from Makobe Island are commonly known as OB (black blotches on orange), WB (black blotches on white) and P (plain, it has no blotches). The OB and WB morphs are associated with dominant female determiners on the X chromosome (Seehausen et al., 1999). The effect of the female determiners can be suppressed by autosomal rescue genes, required to generate blotched males. If a blotched female mates with a plain male that lacks rescue genes the offspring sex ratio will be distorted but not if a blotched female mates with a male (plain or blotched) that possesses rescue genes. Yet OB and WB males are extremely rare in nature (< 1%), which leads to the inference that such males have some fitness disadvantage (Seehausen et al., 1999). The rarity of blotched males and the putative fitness disadvantages have been proposed to be preventing speciation from becoming complete (Seehausen et al., 1999).

I re-examined the frequency distributions of the *N. omnicaeruleus* colour types from Makobe Island and compared them with the earlier data. I then analysed morphology and nine microsatellite DNA loci of the samples from the last sampling year. Additionally, I analysed morphology and the same nine microsatellite loci in a closely related species, *Neochromis. greenwoodi* with two colour morphs (P and OB), from a nearby island, Igombe.

I found that in *N. omnicaeruleus* at Makobe Island the OB, WB and P morph frequencies have fluctuated but have not changed directionally for the last 14 years, which represent 14 generations. I found limited evidence of eco-morphological differentiation between sympatric ancestral (plain) and derived (blotched) colour morphs and pattern of non-random mating, manifested in significantly non-random genotypic assignment and in an excess of linkage disequilibrium in the plain morph, though no significant F_{ST} divergence at neutral loci between sympatric morphs. The polymorphism in *N. greenwoodi* presented a very similar pattern. Analysing sympatric and allopatric morphs together, the extent of eco-morphological differentiation correlated positively with neutral genetic differentiation, but significant F_{ST} differentiation at neutral loci occurred only between blotched colour morphs of geographically isolated populations. Frequency dependent sexual selection on the colour morphs appears to be the most plausible explanation for the maintenance of these polymorphisms in sympatry, which have characteristics of arrested stages in truly sympatric speciation. Also, the data on morphology suggest the possibility of ecological character displacement following or

parallel to divergence in colouration.

Chapter 3 - Evidence for eco-morphological divergence in heterogeneous environments

Environmental heterogeneity, generating environmental gradients of different steepness has been shown to be important in explaining diversity in colouration of rock-dwelling cichlids (Seehausen et al, 1997). In this study I investigated three trophic polymorphisms, which were described thirteen years ago (Seehausen, 1996), from three locations with different environmental structure, and putatively variable degrees of progress towards speciation.

The three trophic polymorphisms, composed of two types each, follow a similar pattern: one type is a typical algae scraper with bicuspid dentition, deeper body and large head and the second type is a more slender type, with unicupsid dentition. One of the polymorphisms was reported at Makobe island and is composed of the species *N. omnicaeruleus* a typical algae scraper and *N.* "unicuspid scraper" reported to have unicuspid dentition (Seehausen, 1996). The other polymorphism, reported at Igombe island, is very similar to the previous, except that at this island *N. greenwoodi* replaces *N. omnicaeruleus* (Seehausen, 1996). The third polymorphism was reported in the species *Neochromis* "Bihiru scraper" from Bihiru Island. *N.* "Bihiru scraper" was considered a polymorphic species, which appeared to have two discrete types that differ in tooth morphology, body and jaw shape and were called *N.* "Bihiru scraper" unicuspid type and *N.* "Bihiru scraper" bicuspid type (Seehausen, 1996). The shores of the three islands differ in water clarity, steepness of the slope and depth of the rock-sand interface.

I examined the environment around the three sites: Makobe, Igombe and Bihiru. For each population I analysed morphology, dentition and nine microsatellite DNA loci. I then tested if there was evidence for divergence in eco-morphology and at neutral loci at each of the islands and between populations from different islands. I tested for correlations between divergence in dentition, morphology and neutral loci between types from different and from the same sites. Finally I investigated associations between water depth and variation in morphology, dental characters and neutral loci. Using the species *N.* "Bihiru scraper" I also performed a common garden experiment to test if differences in morphology and dentition had any genetic basis.

I found that the replicate pairs of sympatric eco-morphs show consistent phenotype-environment correlations within islands. Islands vary in extent of eco-morph phenotypic differentiation, from single traits to more than 10 different traits. At Makobe island, which has a gentle slope, clear water and substructured environment two eco-morphologically highly distinct morphs, *N.* "unicuspid scraper" and *N. omnicaeruleus* appear to be eco-morphs within a single species, despite evidence for non-random mating coming from non-random genetic assignment. Only at this site is phenotypic variation correlated with depth distribution. At Igombe island, which has an intermediate environmental gradient to the other two islands, *N.* "unicuspid scraper" and *N. greenwoodi* diverge in some of the same eco-morphological and dental traits, but in this case the morphs are significantly differentiated at neutral genetic loci. While in *N.* "Bihiru scraper", at Bihiru island, which has a steep environmental gradient, phenotypic response to disruptive selection is only observed on one trait, tooth shape, and there is no evidence for neutral loci differentiation These examples provide evidence for disruptive

ecological selection maintaining eco-morph differentiation in dentition, and divergent selection maintaining differentiation in head and jaw shape despite high gene flow.

Chapter 4-6: The Pundamilia system: evidence for sensory drive speciation, divergent selection and phenotypic plasticity.

The sister species *Pundamilia pundamilia* and *Pundamilia nyererei* are widely and sympatrically distributed at rocky islands in Lake Victoria (Seehausen et al., 1999a). These species differ primarily in male breeding colouration, which is used in their identification. Males of both species have blackish underparts and blackish vertical bars on the flanks but *Pundamilia nyererei* males are yellow laterally and bright red dorsally, and *Pundamilia pundamilia* males are metallic grey-blue (Seehausen, 1997). Females of both species are cryptically coloured. At some islands *P.nyererei* females are more yellowish and P. *pundamilia* females are more greyish but they can be quite hard to distinguish. At places in the lake where the water transparency is high these forms are ecologically and genetically differentiated sister species; genetically isolated largely by female preference for male breeding colour and no intermediate phenotypes are present (Seehausen, 1997; Seehausen et al., 1997; Seehausen, 2008). A series of studies have shown that at such places these two species are differentiated also in colour vision (Carleton et al. 2005), behavioural response to light of different colour (Maan et al., 2006) and feeding ecology (Bouton et al., 1997). However, at other places where water transparency is low intermediate phenotypes are found, or even dominate the population (Seehausen, et al., 1997, Seehausen, 2008). At these places the two phenotypes are merely extremes in a single panmictic population. In **chapter 4**, these phenotypes were analysed at five different islands along five light gradients of different steepness. This study identifies the ecological and molecular basis of divergent evolution in the cichlid visual system, demonstrates associated divergence in male colouration and female preferences, and shows subsequent differentiation at neutral loci, indicating reproductive isolation. The hypothesis of speciation through sensory drive in these two species is supported by showing that the two species differ in water depth and photic environment and microhabitat, that the visual system of the two species has diverged in adaptation to the different photic environments and that females of each species select for more males with the more conspicuous nuptial colour in each of the environments. There are correlations between all these elements in populations living on moderately shallow to moderately steep light gradients. However, the correlations become weaker and eventually disappear altogether as the light gradient becomes steeper. Along this transect of islands with different light gradients mediated by water depth and turbidity the two species go from one species with high phenotypic variation, to two incipient species, and finally to two sister species, in what has been called a speciation or a "speciation in reverse" transect (Seehausen, 2008).

Based on the findings of **chapter 4**, **chapter 5** analyses in more detail the phenotype pair along the turbidity transect that shows the weakest bimodality in the distribution of male nuptial colouration and the lowest significant differentiation at neutral loci. In this study I tested the hypothesis that divergent or disruptive selection acting on several traits can maintain phenotypic differentiation despite gene flow. I aimed at identifying the phenotypic traits under selection. I confirmed the findings of **chapter 4** that male colouration is associated with water depth. I then analysed if the same

happened for morphology and if these traits also diverged between species. By comparing divergence in morphology and colouration with that of neutral loci I tested if divergent or disruptive selection along a water depth gradient and between species could be invoked. Finally I performed a common garden experiment to test if phenotypic differences found in the wild populations were heritable or whether they are environmentally induced. This study found that male nuptial colouration is correlated with water depth and has a bimodal distribution and that divergence at neutral loci is extremely low but significant, which is consistent with the findings of **chapter 4**. I have also found that differentiation in morphological traits in wild fish is correlated with water depth and is generally much higher than the degree of differentiation for neutral genetic makers, which is consistent with divergent selection acting on ecologically relevant morphological traits along the same gradient. However, several of the morphological differences were lost when fish were raised in a common laboratory environment. This indicates that selection is partly accommodated by heritable adaptation and partly by phenotypic plasticity, where the latter appears to accentuate differences between the incipient species.

The common garden experiment of **chapter 5** also shows that colour is a heritable trait. In **chapter 6** I confirm these findings and estimate the minimum number of genes controlling differences in red and yellow nuptial colouration in the *Pundamilia* sister species pair. I analysed 147 pictures of males of the two parental species, F_1, F_2 and backcross hybrid generations. My analysis suggests that differences in red male colouration between the species are controlled by a minimum number of two to four genes with non-additive effects. A model of one gene with dominance appears to explain differences in yellow colouration. The data also suggests that the two colours are epistatically linked. Haesler & Seehausen (2005) estimated that one to four unlinked loci controlled the differences in female preference in this sympatric species pair. From these estimates it appears that differences in male trait and female preference can be explained by relatively few loci. This would be consistent with assumptions of several models on speciation by sexual selection (Udovic, 1980; Kirkpatrick, 1982; Turner & Burrows, 1995; Takimoto et al., 2000; Van Doorn et al., 2001; Van Doorn et al., 2004;). However, most of these models also assume that the genes controlling the trait(s) maintaining reproductive isolation have additive effects, which is not consistent with my findings.

Chapter 7: Summary and conclusion

In the final chapter I summarise the main findings of the work presented in this thesis.

First I found evidence for a stable sex-linked colour polymorphism, in two species of the genus *Neochromis*, maintained despite almost complete absence of neutral genetic differentiation, although intraspecific morphs are completely sympatric. We could not find any environmental heterogeneity associated with this polymorphism, which has characteristics of an arrested stage of sympatric speciation driven by negatively frequency-dependent sexual selection.

Second, I found evidence for eco-morphological differentiation along water depth in three populations of the genus *Neochromis*, forming coexisting eco-morphs with variable extent of phenotypic and genetic differentiation between them. Variation in the

extent of phenotypic differentiation was well predicted by variation in the slope of the environmental gradients between the three islands. However, neutral genetic differentiation was weak or absent, and variation in the extent of neutral genetic differentiation, was not well predicted by the gradient slopes, nor correlated with the extent of phenotypic differentiation.

Third, in two sister species of the genus *Pundamilia* inhabiting environmental gradients of different steepness too, the greatest amount of phenotypic and genotypic divergence is found on gradients of shallow or intermediate steepness, and none at very steep gradients. This incipient species complex provided the strongest neutral genetic evidence for speciation. We obtained strong evidence that this speciation was driven through sensory drive. I also found evidence for divergent selection and adaptive phenotypic plasticity contributing to differences in eco-morphology between these species. Finally I estimated the number of genes controlling the difference between these two species in the extent of red male nuptial colouration to be between two and four plus one gene with dominance controlling presence/absence of yellow colouration. These numbers of genes are within the parameter space that allows for sympatric speciation in most models of sympatric speciation by sexual selection.

The main conclusion from my findings is that divergent selection, sexual and / or ecological, acting on several types of traits can maintain high levels of phenotypic diversity and differentiation despite high gene flow among Lake Victoria cichlids. Variation in the steepness of environmental gradients can explain much of the variation in the extent of phenotypic differentiation: better differentiation is associated with shallower gradients. Even though some populations maintain some phenotypic differentiation in complete sympatry, significant differentiation at neutral loci, was only observed between phenotype pairs inhabiting environmental gradients that were not very steep. Therefore, it appears that in Lake Victoria cichlid fish, environmental heterogeneity is often required for speciation, as evidenced by significant amounts of neutral genetic differentiation. All environmental gradients that we studied were mediated by light. The eutrophication of Lake Victoria in recent decades has lead to a massive increase of turbidity and associated loss of light-mediated habitat heterogeneity. My findings have important implications for conservation of species diversity and ecological diversity.

References

Adams, C. E. & Huntingford, F. A. 2004. Incipient speciation driven by phenotypic plasticity? Evidence from sympatric populations of Arctic charr. *Biol. J. Linn. Soc.* **81**: 611-618.

Arnegard, M. E. & Kondrashov, A. S. 2004. Sympatric speciation by sexual selection alone is unlikely. *Evolution* **58**: 222-237.

Balkau, B. J. & Feldman, M. W. 1973 Selection for Migration Modification. *Genetics* **74**: 171-174.

Barluenga, M., Stoelting, K. N., Salzburger, W., Muschick, M. & Meyer, A. 2006. Evolutionary biology - Evidence for sympatric speciation? Reply. *Nature* **444**: E13-E13.

Barton, N. H. 1979. Gene Flow Past a Cline. *Heredity* **43**: 333-339.

Bolnick, D. I. 2004. Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. *Evolution* **58**: 608-618.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**: 944-948.

Boughman, J. W. 2002. How sensory drive can promote speciation. *Trends Ecol. Evol* **17**: 571-577.

Bouton, N., Seehausen, O. & van Alphen, J. J. M. 1997. Resource partitioning among rock-dwelling haplochromines (Pisces : Cichlidae) from Lake Victoria. *Ecol. Freshw. Fish* **6**: 225-240.

Bouton, N., Witte, F. & Van Alphen, J. J. M. 2002. Experimental evidence for adaptive phenotypic plasticity in a rock-dwelling cichlid fish from Lake Victoria. *Biol. J. Linn. Soc.* **77**: 185-192.

Bouton, N., Witte, F., van Alphen, J. J. M., Schenk, A. & Seehausen, O. 1999. Local adaptations in populations of rock-dwelling haplochromines (Pisces : Cichlidae) from southern Lake Victoria. *P. Roy. Soc. B-Biol. Sci.* **266**: 355-360.

Bridle, J. R., Saldamando, C. I., Koning, W. & Butlin, R. K. 2006. Assortative preferences and discrimination by females against hybrid male song in the grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera: Acrididae). *J. of Evol. Biol.* **19**: 1248-1256.

Bush, G. L. 1994. Sympatric Speciation in Animals - New Wine in Old Bottles. *Trends Ecol. Evol* **9**: 285-288.

Butlin, R. K., Galindo, J. & Grahame, J. W. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos. T. Roy. Soc. B* **363**, 2997-3007.

Capron de Caprona, M. D. & Fritzsch, B. 1984 Interspecific Fertile Hybrids of Haplochromine Cichlidae (Teleostei) and Their Possible Importance for Speciation. *Neth. J. of Zool.* **34**: 503-&.

Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O. 2005. Colour vision and speciation in Lake Victoria cichlids of the genus Pundamilia. *Mol. Ecol.* **14**: 4341-4353.

Chapman, T., Arnqvist, G., Bangham, J. & Rowe, L. 2003. Sexual conflict. *Trends Ecol. Evol* **18**: 41-47.

Cohen, A. S., Soreghan, M. J. & Scholz, C. A. 1993. Estimating the Age of Formation of Lakes - an Example from Lake Tanganyika, East-African Rift System. *Geology* **21**: 511-514.

Coyne, J. A. & Orr, H. A. 2004. *Speciation*: Sunderland, Massachusetts, Sinauer Associates.

Dieckmann, U. & Doebeli, M. 1999. On the origin of species by sympatric speciation. *Nature* **400**: 354-357.

Dieckmann, U., Doebeli, M., Metz, J. A. J. & Tautz, D. 2004. *Adaptive Speciation*: Cambridge University Press.

Diehl, S. R. & Bush, G. L. 1989. The role of habitat preference in adaptation and speciation. In *Speciation and Its Consequences* (ed. D. Otte & J. Ender), pp. 345-365. Sunderland, MA: Sinauer Associates.

Dijkstra, P. D., Seehausen, O., Gricar, B. L. A., Maan, M. E. & Groothuis, T. G. G. 2006. Can male-male competition stabilize speciation? A test in Lake Victoria haplochromine cichlid fish. *Behav. Ecol. and Sociobiol.* **59**: 704-713.

Dijkstra, P. D., Seehausen, O. & Groothuis, T. G. G. 2005. Direct male-male competition can facilitate invasion of new colour types in Lake Victoria cichlids. *Behav. Ecol. and Sociobiol.* **58**: 136-143.

Dijkstra, P. D., Seehausen, O. & Groothuis, T. G. G. 2008. Intrasexual competition among females and the stabilization of a conspicuous colour polymorphism in a Lake Victoria cichlid fish. *P. Roy. Soc. B-Biol. Sci.* **275**: 519-526.

Dijkstra, P. D., Seehausen, O., Pierotti, M. E. R. & Groothuis, T. G. G. 2007. Male-male competition and speciation: aggression bias towards differently coloured rivals varies between stages of speciation in a Lake Victoria cichlid species complex. *J. Evol. Biol.* **20**: 496-502.

Doebeli, M. 1996. A quantitative genetic competition model for sympatric speciation. *J. Evol. Biol.***9**: 893-909.

Doebeli, M. & Dieckmann, U. 2003. Speciation along environmental gradients. *Nature* **421**: 259-264.

Doebeli, M. & Dieckmann, U. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am. Nat.* 156: 77-101.

Endler, J. A. 1973. Gene Flow and Population Differentiation. *Science* **179**, 243-250.

Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why Are There So Few Kinds of Animals. *Evolution* **35**, 124-138.

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

Fry, J. D. 2003. Multilocus models of sympatric speciation: Bush versus Rice versus Felsenstein. *Evolution* **57**: 1735-1746.

Fryer, G. & Iles, T. D. 1972. *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution.* Edinburgh: Oliver & Boyd.

Gavrilets, S. 1997. Hybrid zones with Dobzhansky-type epistatic selection. *Evolution* **51**: 1027-1035.

Gavrilets, S. 2000a. Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403**: 886-889.

Gavrilets, S. 2000b. Waiting time to parapatric speciation. *P. Roy. Soc. B-Biol. Sci.* **267**: 2483-2492.

Gavrilets, S. 2003. Perspective: Models of speciation: What have we learned in 40 years? *Evolution* **57**: 2197-2215.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species*: Princeton University press.

Gavrilets, S. & Vose, A. 2007. Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Mol. Ecol.* **16**: 2910-2921.

Gavrilets, S., Vose, A., Barluenga, M., Salzburger, W. & Meyer, A. 2007 Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Mol. Ecol.* **16**: 2893-2909.

Gavrilets, S. & Waxman, D. 2002. Sympatric speciation by sexual conflict. . *Nat. Acad. Sci. USA* **99**: 10533-10538.

Genner, M. J., Hawkins, S. J. & Turner, G. F. 2003. Isotopic change throughout the life history of a Lake Malawi cichlid fish. *J. of Fish Biol.* **62**: 907-917.

Genner, M. J., Seehausen, O., Cleary, D. F. R., Knight, M. E., Michel, E. & Turner, G. F. 2004. How does the taxonomic status of allopatric populations influence species richness within African cichlid fish assemblages? *J. of Biogeography* **31**: 93-102.

Genner, M. J. & Turner, G. F. 2005. The mbuna cichlids of Lake Malawi: a model for rapid speciation and adaptive radiation. *Fish and Fisheries* **6**: 1-34.

Genner, M. J., Turner, G. F. & Hawkins, S. J. 1999. Foraging of rocky habitat cichlid fishes in Lake Malawi: coexistence through niche partitioning? *Oecologia* **121**: 283-292.

Greenwood, P. H. 1974. The cichlid fishes of Lake Victoria, East Africa: The biology and evolution of a species flock. *Bull. Br. Mus. Nat. Hist. (Zool.) Suppl.* **6**:1-134.

Grant, P. R. & Grant, B. R. 2006. Evolution of character displacement in Drawin's finches. *Science* **313**: 224-226.

Haesler, M. P. & Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *P. Roy. Soc. B-Biol. Sci.* **272**: 237-245.

Hendry, A. P., Huber, S. K., De Leon L. F., Herrel, A. & Podos, J. 2008. Disruptive selection in a bimodal population of Drawin's finches. *P. Roy. Soc. B-Biol. Sci.*, in press.

Higashi, M., Takimoto, G. & Yamamura, N. 1999. Sympatric speciation by sexual selection. *Nature* **402**: 523-526.

Hori, M. 1993. Frequency-Dependent Natural-Selection in the Handedness of Scale-Eating Cichlid Fish. *Science* **260**: 216-219.

Ito, H. C. & Dieckmann. A new mechanism for recurrent adaptive radiations. *Am. Nat.* **170**: 96-111.

Johnson, P. A., Hoppensteadt, F. C., Smith, J. J. & Bush, G. L. 1996a. Conditions for sympatric speciation: A diploid model incorporating habitat fidelity and non-habitat assortative mating. *Evol. Ecol.* **10**: 187-205.

Johnson, T. C., Kelts, K. & Odada, E. 2000. The holocene history of Lake Victoria. *Ambio* **29**: 2-11.

Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J. W. 1996b. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**: 1091-1093.

Kawata, M., Shoji, A., Kawamura, S. & Seehausen, O. 2007. A genetically explicit model of speciation by sensory drive within a continuous population in aquatic environments. *Bmc Evolutionary Biology* **7**.

Kawecki, T. J. 1996. Sympatric speciation driven by beneficial mutations. *P. Roy. Soc. B-Biol. Sci.* **263**: 1515-1520.

Kawecki, T. J. 1997. Sympatric speciation via habitat specialization driven by deleterious mutations. *Evolution* **51**: 1751-1763.

Kirkpatrick, M. 1982. Sexual Selection and the Evolution of Female Choice. *Evolution* **36**: 1-12.

Kirkpatrick, M. & Nuismer, S. L. 2004. Sexual selection can constrain sympatric speciation. *P. Roy. Soc. B-Biol. Sci.* **271**: 687-693.

Kirkpatrick, M. & Ravigne, V. 2002. Speciation by natural and sexual selection: Models and experiments. *Am. Nat.* **159**: S22-S35.

Knight, M. E. & Turner, G. F. 2004. Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish *Pseudotropheus* zebra from Lake Malawi. *P. Roy. Soc. B-Biol. Sci.* **271**: 675-680.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Rev. Genetics* **5**: 288-298.

Kondrashov, A. S. & Kondrashov, F. A. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351-354.

Lande, R. 1982 Rapid Origin of Sexual Isolation and Character Divergence in a Cline. *Evolution* **36**: 213-223.

Leimar, O., Doebeli, M. & Dieckmann, U. 2008. Evolution of phenotypic clusters through competition and local adaptation along an environmental gradient. *Evolution* **62**: 807-822.

Maan, M., Seehausen, O., Soderberg, L., Johnson, L., Ripmeester, E. A. P., Mrosso, H. D. J., Taylor, M. I., van Dooren, T. J. M. & van Alphen, J. J. M. 2004. Intraspecific sexual selection on a speciation trait, male colouration, in the Lake Victoria cichlid *Pundamilia nyererei*. *P. Roy. Soc. B-Biol. Sci.* **271**: 2445-2452.

Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. 2006. Sensory drive in cichlid speciation. *Am. Nat.* **167**: 947-954.

MacCune A. R. & Lovejoy, N. R. 1998. The relative rates of sympatric and allopatric speciation in fishes. *Endless Forms: Species and Speciation*.

Maynard Smith, J. 1966. Sympatric speciation. *Am. Nat.* **100**: 637-650.

Mayr, E. 1942. *Systematics and the origin of species*: Columbia University Press.

Messmer, V., Jones, G. P., van Herwerden, L. & Munday, P. L. 2005. Genetic and ecological characterisation of colour dimorphism in a coral reef fish. *Env. Biol. of Fishes* **74**: 175-183.

Mikami, O. K., Kohda, M. & Kawata, M. 2004. A new hypothesis for species coexistence: male-male repulsion promotes coexistence of competing species. *Pop. Ecol.* **46**: 213-217.

Nagl, S., Tichy, H., Mayer, W. E., Takezaki, N., Takahata, N. & Klein, J. 2000. The origin and age of haplochromine fishes in Lake Victoria, East Africa. *P. Roy. Soc. B-Biol. Sci.* **267**: 1049-1061.

Nagylaki, T. 1975. Conditions for Existence of Clines. *Genetics* **80**: 595-615.

Nagylaki, T. 1976. Clines with Variable Migration. *Genetics* **83**: 867-886.

Naisbit, R. E., Jiggins, C. D. & Mallet, J. 2001. Disruptive sexual selection against hybrids contributes to speciation between Heliconius cydno and Heliconius melpomene. *P. Roy. Soc. B-Biol. Sci.* **268**: 1849-1854.

Nosil, P. & Harmon, L. J. 2008. Niche dimensionality and ecological speciation. In *Ecology and Speciation*: Cambridge University Press.

Odgen, R., Thorpe, R. S., 2002. Molecular evidence for ecological speciation in tropical habitats. . *Nat. Acad. Sci. USA* **90**: 13612-13615.

Panhuis, T. M., Butlin, R., Zuk, M. & Tregenza, T. 2001. Sexual selection and speciation. *Trends Ecol. Evol* **16**, 364-371.

Payne, R. J. H. & Krakauer, D. C. 1997. Sexual selection, space and speciation. *Evolution* 51:1-9..

Polechova, J. & Barton, N. H. 2005. Speciation through competition: A critical review. *Evolution* **59**: 1194-1210.

Rasanen, K. & Hendry, A. P. 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol. Lett.* **11**: 624-636.

Rice, W. R. 1984. Disruptive Selection on Habitat Preference and the Evolution of Reproductive Isolation - a Simulation Study. *Evolution* **38**: 1251-1260.

Rice, W. R. 1984. Selection via habitat specialization: The evolution of reproductive isolation as a correlated character. *Evol. Ecol.* **1**: 1637-1653.

Rueffler, C., Van Dooren, T. J. M., Leimar, O. & Abrams, P. A. 2006. Disruptive selection and then what? *Trends Ecol. Evol* **21**: 238-245.

Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**, 306-308.

Ryan, M. J. 1998. Sexual selection, receiver biases, and the evolution of sex differences. *Science* **281**: 1999-2003.

Ryan, P. G., Bloomer, P., Moloney, C. L., Grant, T. J. & Delport, W. 2007. Ecological speciation in South Atlantic island finches. *Science* **315**: 1420-1423.

Saldamando, C. I., Tatsuta, H. & Butlin, R. K. 2005. Hybrids between Chorthippus brunneus and C-jacobsi (Orthoptera : Acrididae) do not show endogenous postzygotic isolation. *Biol. J. Linn. Soc.* **84**: 195-203.

Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. 2005. Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *Bmc Evolutionary Biology* **5**.

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N. & Baker, W. J. 2006. Sympatric speciation in palms on an oceanic island. *Nature* **441**: 210-213.

Schliewen, U., Rassmann, K., Markmann, M., Markert, J., Kocher, T. & Tautz, D. 2001. Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Mol. Ecol.* **10**: 1471-1488.

Schluter, D. 2000. Ecology and the origin of species. *Trends Ecol. Evol*. **15**: 372-380.

Schluter, D. 1996a. Ecological causes of adaptive radiation. *Am. Nat.* **148**: 40-64.

Schluter, D. 1996b. Ecological speciation in postglacial fishes. *Philos. T. Roy. Soc. B* **351**: 807-814.

Schluter, D. 2000. Ecological character displacement in adaptive radiation. *Am. Nat.* **156**: S4-S16.

Schluter, D. & Price, T. 1993. Honesty, Perception and Population Divergence in Sexually Selected Traits. *P. Roy. Soc. B-Biol. Sci.* **253**: 117-122.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids. Taxonomy, Ecology and Distribution.* Zevenhuizen, NL: Cichlid Press.

Seehausen, O. 1997. Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid (Haplochromis nyererei) (vol 5, pg 195, 1996). *Ecol. Freshw. Fish* **6**: 59-66.

Seehausen, O. 2000. Explosive speciation rates and unusual species richness in haplochromine cichlid fishes: Effects of sexual selection. In *Advances in Ecological Research,* vol. 31, pp. 237-274.

Seehausen, O. 2002. Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. *P. Roy. Soc. B-Biol. Sci* **269**: 491-497.

Seehausen, O. 2004. Response to Schilthuizen et al.: Hybridization, rare alleles and adaptive radiation. *Trends Ecol. Evol* **19**: 405-406.

Seehausen, O. 2008. Progressive levels of trait divergence along a "speciation transect" in the Lake Victoria cichlid fish Pundamilia. In *Ecology and Evolution* (ed. R. Butlin, D. Schluter & J. R. Bridle): Cambridge University Press.

Seehausen, O. & Bouton, N. 1997. Microdistribution and fluctuations in niche overlap in a rocky shore cichlid community in Lake Victoria. *Ecol. Freshw. Fish* **6**: 161-173.

Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., Turner, G. F., van Alphen, J. J. M. & Bills, R. 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock (vol 270, pg 129, 2003). *P. Roy. Soc. B-Biol. Sci.* **270**: 2637-2638.

Seehausen, O., Mayhew, P. J. & Van Alphen, J. J. M. 1999a. Evolution of colour patterns in East African cichlid fish. *J. of Evol. Biol.* **12**: 514-534.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. . *Nat. Acad. Sci. USA* **271**: 1345-1353.

Seehausen, O. & van Alphen, J. J. M. 1998. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behav. Ecol. and Sociobiol.***42**: 1-8.

Seehausen, O., van Alphen, J. J. M. & Lande, R. 1999b. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**: 367-378.

Seehausen, O., vanAlphen, J. J. M. & Witte, F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808-1811.

Smith, T. B. 1993. Disruptive Selection and the Genetic-Basis of Bill Size Polymorphism in the African Finch Pyrenestes. *Nature* **363**: 618-620.

Smith, T. B., Wayne, R. K., Girman D. J. & Bruford, M. W. 1997. A role for ecotones in generating rainforest biodiversity. *Science* **276**: 1855-1857.

Smith, T. B., Schneider, C. J. & Holder, K. H. 2001. Refugial isolation versus eoclogical gradients. Testing alternative mechanisms of evolutionary divergence in four rainforest vertebrates. *Genetica* **112-113**: 383-398.

Smith, T. B., Calsbeek, R., Wayne, R. K., Holder, K. H., Pires, D. & Bardeleben, C. 2005. Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. *J. of Evol. Biol.* **18**: 257-268.

Stager, J. C. & Johnson, T. C. 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5-16.

Stelkens, R. B., Pierotti, M. E. R., Joyce, D. A., Smith, A. M., van der Sluijs, I. & Seehausen, O. 2008. Disruptive sexual selection on male nuptial colouration in an experimental hybrid population of cichlid fish. *Philos. T. Roy. Soc. B*. **363**: 2861-2870.

Streelman, J. T., & Danley P. D. 2003. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol* 18: 126-131.

Takimoto, G., Higashi, M. & Yamamura, N. 2000. A deterministic genetic model for sympatric speciation by sexual selection. *Evolution* **54**: 1870-1881.

Temeles, E. J., Pan, I. L., Brennan, J. L. & Horwitt, J. N. 2000. Evidence for ecological causation of sexual dimorphism in a hummingbird. *Science* **289**: 441-443.

Turelli, M., Barton, N. H. & Coyne, J. A. 2001. Theory and speciation. *Trends Ecol. Evol.* **16**: 330-343.

Turner, G. F. & Burrows, M. T. 1995. A Model of Sympatric Speciation by Sexual Selection. *P. Roy. Soc. B-Biol. Sci.* **260**: 287-292.

Turner, G. F., Seehausen, O., Knight, M. E., Allender, C. J. & Robinson, R. L. 2001. How many species of cichlid fishes are there in African lakes? *Mol. Ecol.* **10**: 793-806.

Udovic, D. 1980. Frequency-Dependent Selection, Disruptive Selection, and the Evolution of Reproductive Isolation. *Am. Nat.* **116**: 621-641.

van der Sluijs, I., van Alphen, J. J. M. & Seehausen, O. 2008a. Preference polymorphism for coloration but no speciation in a population of Lake Victoria cichlids. *Behav. Ecol.* **19**, 177-183.

Van der Sluijs, I., Van Dooren, T. J. M., Hofker, K. D., Van Alpen, J. J. M., Stelkens, R. B. & Seehausen, O. 2008b. Female mating preference functions predict sexual selection against hybrids between sibling species of cichlid fish. *Philos. T. Roy. Soc. B*.

Van der Sluijs, I., Van Dooren, T. J. M., Seehausen, O. & Van Alphen, J. J. M. 2008c. A test of fitness consequences of hybridization in sibling species of Lake Victoria cichlid fish. *J. of Evol. Biol.* **21**: 480-491.

Van Doorn, G. S., Dieckmann, U. & Weissing, F. J. 2004. Sympatric speciation by sexual selection: A critical reevaluation. *Am. Nat.* **163**: 709-725.

Van Doorn, G. S., Luttikhuizen, P. C. & Weissing, F. J. 2001. Sexual selection at the protein level drives the extraordinary divergence of sex-related genes during sympatric speciation. *P. Roy. Soc. B-Biol. Sci.* **268**: 2155-2161.

Van Doorn, G. S., Noest, A. J. & Hogeweg, P. 1998. Sympatric speciation and extinction driven by environment dependent sexual selection. *P. Roy. Soc. B-Biol. Sci.* **265**: 1915-1919.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol* **16**: 381-390.

Chapter 2

Sympatric colour polymorphisms associated with non-random gene flow in cichlid fish of Lake Victoria

Sympatric colour polymorphisms associated with non-random gene flow in cichlid fish of Lake Victoria

Isabel S. Magalhaes, Salome Mwaiko, Ole Seehausen

(Submitted to Journal of Evolutionary Biology)

Colour polymorphisms have been suggested to be incipient stages of speciation. We studied two geographically isolated species of Lake Victoria cichlids exhibiting very similar sex-linked colour polymorphisms. The ecology and behaviour of one of these species is well studied, with colour-based mating and aggression preferences. Here we ask if the resulting disruptive selection, sexual or of other unknown source, can reduce gene flow between the colour morphs and permit differentiation in traits other than colour. Over the past 14 years, the frequencies of colour morphs have oscillated, but there is no evidence for directional change, suggesting the colour polymorphism is persistent on an ecological time scale. We find limited evidence of eco-morphological differentiation between sympatric ancestral (plain) and derived (blotched) colour morphs and of non-random mating, manifested in significantly non-random genotypic assignment and in an excess of linkage disequilibrium in the plain morph. Analysing sympatric and allopatric morphs together, the extent of eco-morphological differentiation correlated positively with neutral genetic differentiation, but significant multilocus-differentiation at neutral loci occurred only between derived colour morphs of geographically isolated populations. Frequency dependent sexual selection on the colour morphs remains the most plausible explanation for the maintenance of these polymorphisms in sympatry, although our data suggests the possibility of ecological character displacement following or parallel to divergence in colouration.

Key words: cichlids, colour polymorphism, character displacement, F_{ST}, speciation

Introduction

Sympatric speciation remains a controversial issue (Coyne & Orr, 2004). Several supportive cases have been identified in lake-dwelling fish (Schliewen et al., 1994, Schliewen et al., 2001, Jonsson & Jonsson, 2001; Barluenga et al., 2006). Sympatric speciation has been defined as speciation in a population in which the site where individuals reproduce is independent of their site of birth (Gavrilets, 2004). As mating is random with respect to birth place, the proportion of members of the two diverging populations that is exchanged per generation (i.e. migration rate) is expected to be 0.5. Despite occasional claims to the contrary, population genetic theory and modelling identify only fairly narrow parameter space in which such speciation is possible. Importantly, it requires negative frequency dependent selection on the speciation phenotype to stabilize phenotypic polymorphism at the incipient stage of speciation (Dieckmann & Doebeli, 1999, Kondrashov & Kondrashov, 1999, van Doorn et al., 2004).

Colour polymorphisms have been suggested to be potential incipient stages in sympatric speciation of African cichlid fish (Holzberg, 1978, Marsh et al., 1981, Seehausen, 2000, Kocher, 2004) and in speciation in many other organisms (Gray & McKinnon 2006). There is fairly strong ecological and population genetic evidence that male nuptial colour polymorphisms can be an incipient stage of speciation in cichlid fish of Lake Victoria (Seehausen et al., 2008; Magalhaes et al., in press). In this case evolutionary divergence is driven by differentiation in microhabitat between the colour morphs, suggesting $m<0.5$, i.e. ecological parapatry. In other cases, however, colour morphs are fully sympatric and syntopic (Seehausen et al., 1999; Maan et al., 2008). Here we investigate such cases.

The colour polymorphisms investigated in our study affect predominantly females. The same polymorphisms have been identified in several species of cichlids from Lake Victoria and involve the orange blotched and the white blotched phenotypes (OB; black blotches on orange and WB; black blotches on white, respectively). Both are sex linked in all populations for which data are available (Lande et al., 2001). Sex linkage of both OB and WB has been attributed to weak physical linkage between two different major effect genes for colour and dominant female determiners, together linked to the original recessive female determiner X (Seehausen et al., 1999). Blotched individuals are female, unless additional autosomal male determiners are present that compensate the dominant female effect of the OB and WB-linked female determiners. Males of OB and WB colouration are extremely rare (< 1%) in nature in most studied populations (Seehausen et al., 1999; Lande et al., 2001; Maan et al., 2008). Species possessing OB or WB morphs usually also possess a plain (P) morph in both sexes, which has no blotches and is considered the ancestral condition, but in some species, the blotched morph has become fixed in at least the female sex (Lande et al., 2001). Theoretical studies modeling this polymorphism suggested that sympatric speciation may be possible between such morphs without spatial or ecological differentiation through the interaction of selection on sex reversal and sexual selection (Lande et al., 2001).

Most of the knowledge on these polymorphisms derives from work on a species inhabiting rocky shores of Lake Victoria, *Neochromis omnicaeruleus*. Some populations of this species possess all three female colour morphs: WB, OB and P, others have just two (generally OB and P), and again others have just P morphs. Populations that possess blotched morphs inhabit locations with relatively clear water, while at places where the

water is turbid only individuals of the P morph are found (Seehausen, 1996). This suggests that the colour polymorphism is not neutral, but that selection has to be invoked in its maintenance in clear, or in its loss in turbid water habitats. There is no evidence of either microhabitat differentiation (Seehausen & Bouton, 1997; Maan et al., 2008) or morphological differentiation between colour morphs in the best studied population at Makobe Island (Seehausen et al., 1999). Yet, investigations of the behavioural ecology of that population found selective and partially assortative mating based on the colour variation (Seehausen et al., 1999), and that females of the three colour morphs have significant own-morph aggression biases (Dijkstra et al., 2008). Seehausen et al. (1999) and Maan et al. (2008), reporting frequencies of the three *N. omnicaeruleus* colour morphs at Makobe Island for several years, found a large numerical deficiency of intermediate phenotypes suggesting nonrandom mating and/or strong disruptive selection on colouration.

Here we revisit the three colour morphs (P, OB and WB) of *N. omnicaeruleus* from Makobe Island more than a decade after the original work (Seehausen et al., 1999). We look at morph frequencies across 14 years, and analyse morphology and nine microsatellite DNA loci for samples collected in 2005. We use this data to ask whether (1) morph frequencies have changed over 14 years; and whether (2) disruptive selection reduces gene flow between the colour morphs, permitting differentiation in traits other than colour. Additionally, we extend this analysis to a population of the sister species *Neochromis greenwoodi* with two colour morphs (P and OB), from the geographically nearest other island. Finally we study patterns of gene flow and morphological differentiation between the populations of the two species on these two adjacent islands, and between their colour morphs.

Materials and Methods

Study species and sampling

Neochromis omnicaeruleus and *Neochromis greenwoodi* are two species of rock-dwelling algae scrapers widely distributed in the southern part of Lake Victoria (Fig. 1). They are allopatric sibling species with complementary geographical distributions: *N. omnicaeruleus* almost exclusively inhabits offshore islands in the open lake and western Speke Gulf while *N. greenwoodi* is most common along much of the mainland, and at the inshore islands of the Mwanza Gulf (Seehausen, 1996; Seehausen et al. 1998). There are very few localities where the two species occur together: at Juma Island (Sengerema, Tanzania) they behave like distinct species, but from Ndurwa Point near our sampling site Igombe Island (Speke Gulf, Tanzania) a phenotypically intermediate and variable population is known, which may represent a hybrid swarm. In most populations of *N. omnicaeruleus,* males are bright metallic blue while females are yellow-brown or grey with dark vertical bars. These are referred to as the plain (P) morph. Most populations are colour polymorphic and segregate the orange blotch (OB) morph and the P morph. In some populations, inhabiting sites with extremely clear water (i.e. secchi depth higher than 2.5m), a third morph occurs, white blotched (WB). The coexistence of the three morphs is currently known from five islands, all far offshore in very clear waters in Tanzania (Fig.1) and Uganda (Seehausen, unpubl. data). In *N. greenwoodi*, typical male breeding colouration is uniformly blue-black with a narrow red edge on the caudal fin,

while females are typically dark brown. The OB morph is present in many populations of this species and can even be the most abundant morph in some clear water populations, but is entirely absent from turbid water sites (Seehausen & Bouton, 1996). White blotched morphs have not been observed in this species.

We studied allopatric *N. omnicaeruleus* and *N. greenwoodi* at two adjacent rocky islands located in the entrance to the Speke Gulf in southeastern Lake Victoria, Makobe (*N. omnicaeruleus*) and Igombe (*N. greenwoodi*) (Fig. 1). The islands are separated from each other by five kilometres of 40m deep water with muddy substrate. While Makobe is separated from the mainland by five kilometres, Igombe is less than 50m from the mainland. Between October and November 2005 a total of 123 females and 45 males, and 40 females and 30 males were caught using gillnets at Makobe and Igombe islands, respectively. Only females were used for our phenotypic analyses in this manuscript because OB and WB morph males are very rare (see above), and P males can have mating preferences and sex determination genes compatible with any of the 3 female morphs, such that we can not use their colour phenotype to assign them to either of the putative incipient species. The same is true for P females but to a lesser degree (Seehausen et al., 1999). We do, however, report genetic data on the males of the P morph from both species.

Immediately after capture each fish was photographed in a custom-made water-filled photo cuvette. After the fish were humanely killed, fin clips were taken for genetic analysis. Specimens were fixed in formalin and later transferred to 70% ethanol. In figures and tables, groups of individuals from the plain morph, the OB morph and WB morph of *N. omnicaeruleus* and *N. greenwoodi* will be referred to as "No P", "No OB", "No WB", "Ng P" and "Ng OB", respectively.

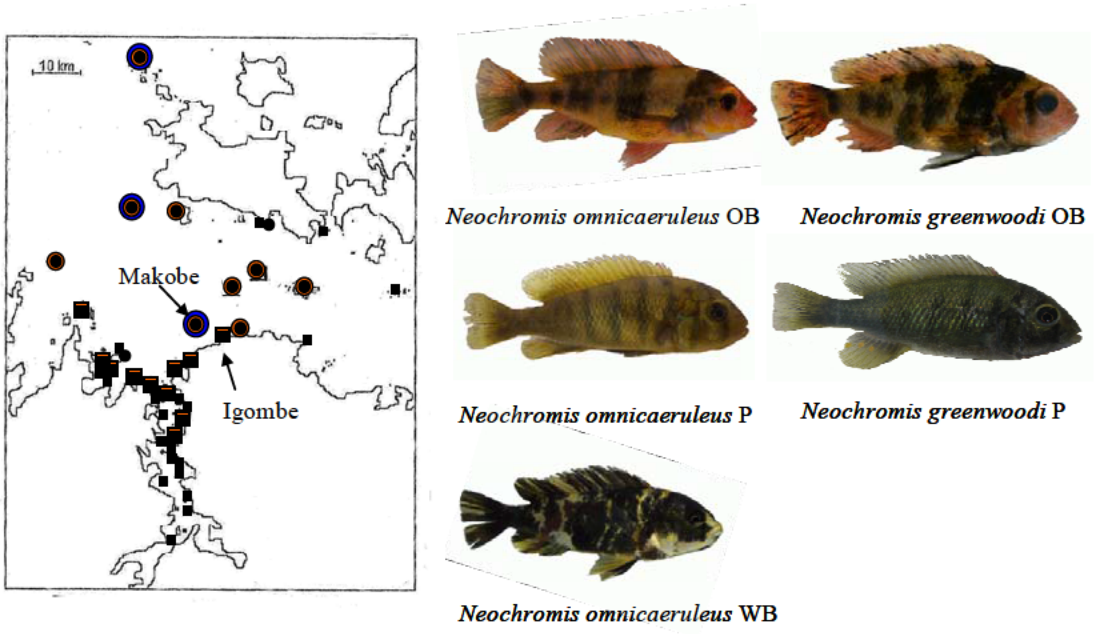

Figure1. Map of the southern part of Lake Victoria with known distributions of the colour morphs of WB (blue), OB (orange) and P (black) morphs of *N. omnicaeruleus* (circles) and *N. greenwoodi* (squares) (adapted from Seehausen, 1996). Arrows indicate the two sampling locations.

Colour analysis

We subjected 28 P, 44 OB and 30 WB *N. omnicaeruleus* and 15 P and 25 OB *N. greenwoodi* to colour analysis. We placed the captured fish in a custom-made water-filled perpex cuvette, squeezed gently between a grey PVC sheet and the front window and photographed them with a Kodak Easyshare CX7430 digital camera. All digital fish images were loaded into image analysis software (Photoshop 6.0, Adobe System Inc.) where fins and eyes were removed, the area corresponding to blotches was selected and hue was measured. Hue was quantified as the average of five hue measurements taken from five different parts of the body not covered by blotches or vertical bars (i.e., dorsum, head, lower body, caudal area and mid body) (Fig 2a). Melanic blotches can vary from brown to grey or black. We selected the area corresponding to blotches by using the "magic wand" tool in Photoshop 6.0, adjusting its tolerance so that by eye the total blotch areas were delimited. We then cut the delimited areas and created a new layer that was saved as a new image. These images were then loaded into Sigma Scan 4.0 (SPSS Inc.) where the area covered by blotches and the number of blotches, were quantified for each fish. Blotch areas and number of blotches were corrected for fish size by dividing their values by the total number of pixels occupied by the entire fish. P individuals were assigned a zero for number of blotches and size of the biggest blotch.

Four parameters were used to characterize colouration: the number of blotches, the total area covered by blotches, the size of the largest blotch and the hue of the area of the body without blotches.

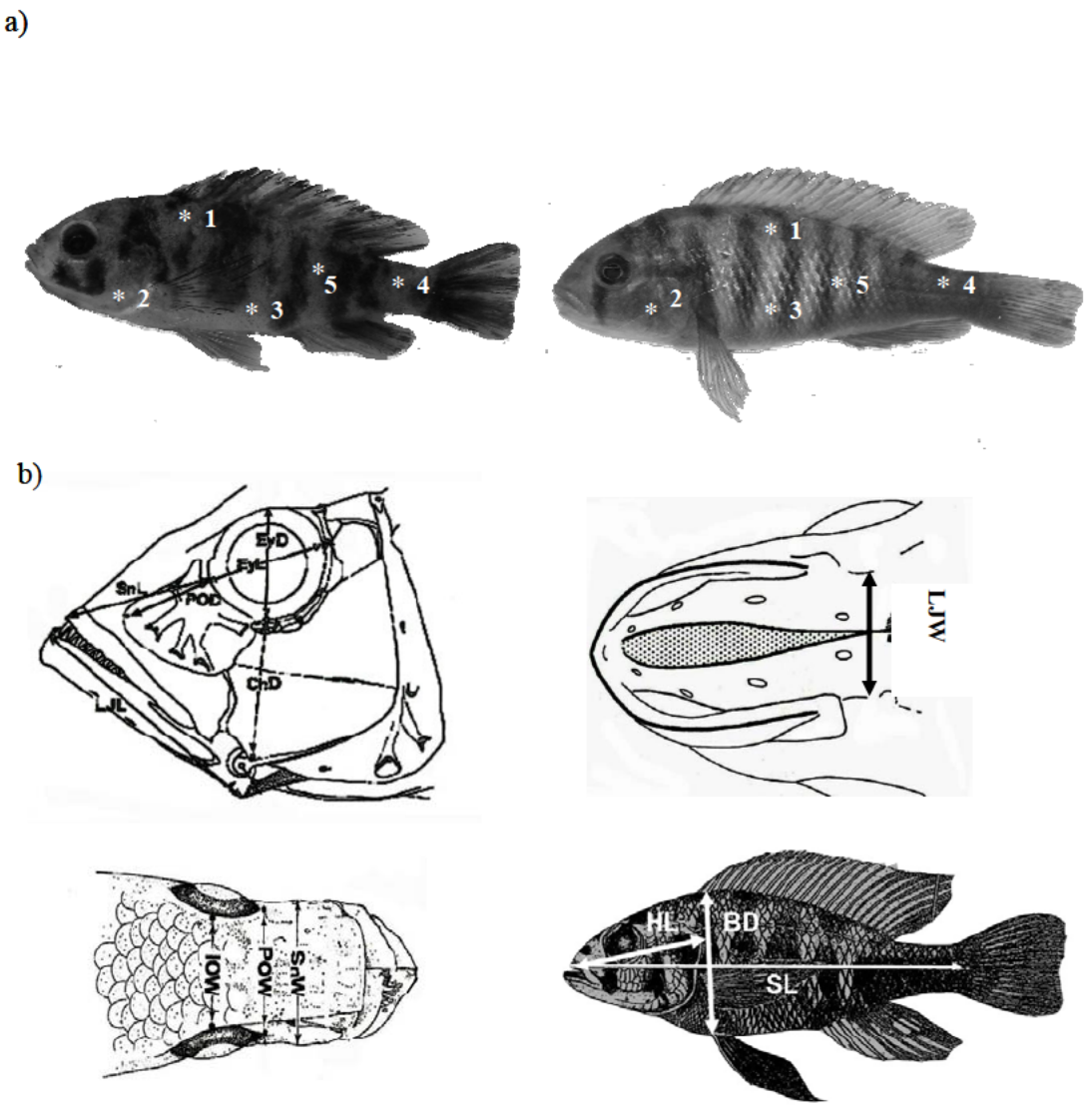

Figure 2. a) Images of a blotched and plain morph of *N. omnicaeruleus* with the five points of the fish body used to estimate the average hue of the body not covered by blotches or bars: 1)dorsum, 2)head, 3)lower body, 4)caudal area and 5)middle of the body; b) Eco-morphometric distances taken by a digital caliper (adapted from Barel *et al.*, 1977): standard length (SL), head length (HL), head width (HW), body depth (BD), lower jaw length (LJL), lower jaw width (LJW), snout length (SnL), snout width (SnW), cheek depth (ChD), pre-orbital depth (PoD), inter-orbital width (IoW), eye length (EyL), eye depth (EyD).

Morphometric analysis

Of the 164 fish, eight were distorted upon preservation and were therefore excluded from the morphometric analyses. Thirteen standard morphometric distances were taken on the remaining 156 fish, using a digital calliper (Barel et al., 1977) (Fig. 2b). Variables were log-transformed and regressed against the log-transformed standard length to adjust each morphometric measurement for size heterogeneity among individuals (Fleming & Gross, 1994). Standardized residuals from these regressions were subsequently used as variables in further analyses.

Statistical analysis

The relative abundances of *N. omnicaeruleus* colour morphs at Makobe island in 2005 were compared to data of five earlier sampling periods (1991: 154 individuals, 1993: 129, 1995: 144, 1996: 89 (Seehausen et al., 1999) and 2002: 216 (Maan et al., 2008)). For *N. greenwoodi* from Igombe Island we compared the relative abundances in 2005 with those from published data from the years 1993/1995 (n = 44 individuals) (Seehausen & Bouton, 1996). Cichlids have a generation time of one year, which means that we compared data across roughly 14 generations for *N. omnicaeruleus* and 10 to 12 generations for *N. greenwoodi*.

We estimated the expected frequencies of each colour morph for each sampling year as sample size multiplied by the relative frequency of a morph among all samples taken over all years. We used chi-square tests to test for each year separately if the observed relative abundances of colour morphs were differed from the expected values. We calculated a generalised linear model (GLM) using a quasibimodal distribution in R (https://www.r-project.org) to test for trends in time in the proportion of blotched females.

To compare colouration and morphology across colour morphs and species, we performed nested univariate analyses of variance (ANOVAs) on standard length, standardized residuals of morphometric distances and on colour measurements (two factors: morph (3 morphs) nested in species (2 species)). For variables showing a significant difference between morphs we performed Tukey *post-hoc* tests to further investigate which groups were significantly different.

Differences between and variances amongst individuals both within and among colour morphs were visualised by Principal Component Analyses (PCA). PCAs were performed on residuals of the log transformed morphometric measurements and on colour measurements separately, with principal components extracted from covariance matrices.

Microsatellite analysis

DNA was extracted from fin clips of 164 females and 75 males using a QIAGEN® (Basel, Switzerland) Biosprint^{TM} 96 extraction robot with a corresponding standard digestion protocol. Nine microsatellite loci developed for Lake Victoria cichlids (Ppun 5, Ppun7, Ppun17, Ppun21 and Ppun32 (Taylor et al., 2002) and for other haplochromine cichlids (OSU20d. OSU19T, OSU16d (Wu et al., 1999) and TmoM5 (Zardoya et al., 1996)) were chosen for population genetics analysis. PCRs for microsatellite amplification were performed using the QIAGEN Multiplex PCR kit according to the manufacturer's protocol. We combined markers into two sets for multiplexing. The first set included the five Ppun loci and the second set the other four

loci. For analysis of fragments a quantity of 1 µl of a 1:2 dilution of the PCR product was added to a volume of deionised formamide (SLS, CEQ Beckman Coulter Fulerton CA, USA) and 400 bp DNA size standard (Beckman Coulter). Denaturated fragments were resolved on an automated DNA sequencer (Beckman Coulter, CEQ 8000) and genotypes manually scored.

Genetic data analysis

We used Convert (Glaubitz, 2004) to generate allele frequency tables and transform the data sheets into input files for other programs. We used ARLEQUIN 3.11 (Excoffier et al., 2005) to test in each locus in each species and morph for departures from Hardy-Weinberg equilibrium (HWE), to calculate observed (H_O) and expected (H_E) heterozygosities and to perform tests for genotypic linkage disequilibrium. Estimates of the mutlilocus inbreeding coefficient (F_{IS}), H_O and H_E for each group were obtained using GENETIX 4.05 (Belkhir et al., 1996). Statistical significance of all tests was adjusted for multiple comparisons using sequential Bonferroni corrections (Rice, 1989).

Genetic differences between colour morphs and between species were determined as F_{ST} (Weir & Cockerham, 1984) by performing AMOVAs in ARLEQUIN. Ninety-five percent confidence intervals for F-statistics were estimated by bootstrapping over loci (10000 replicates) and deviations from zero were determined by 1000 permutations as implemented in ARLEQUIN (Excoffier et al., 2005). As a further test of differentiation we used the allele frequency-based program DOH (http://www.biology.ualberta.ca/jbrzusto/Doh.php) to calculate genetic assignment for each phenotypically assigned individual (Paetkau et al., 1995). Significance of deviation from random assignments was assessed with chi-square tests.

GENETIX 4.05 (Belkhir et al., 1996) was used to perform a Factorial Correspondence Analysis (FCA) over individuals visualising molecular variance among individuals within and between colour morphs.

Correlation between phenotypic and genetic divergence

We quantified phenotypic distances between intra- and interspecific colour morphs by computing multidimensional Euclidean distances between means of groups (colour morphs), using morphometric and colouration data separately. We computed multidimensional Euclidean distances based on the 12 morphometric distances and, separately, based on the four colour characters using SPSS (SPSS Inc., Chicago Illinois, USA). Genetic distances between colour morphs were computed as Nei's standard genetic distance (D_S) using DOH (http://www.biology.ualberta.ca/jbrzusto/Doh.php).

Relationships between genetic, morphometric and colour distances were analysed for intra- and interspecific morphs separately using generalized linear models (GLM) in SPSS 14.0 (SPSS Inc., Chicago Illinois, USA).

Results

Morph frequencies through time

We obtained 34 WB, 48 OB and 42 P-morph females of *N. omnicaeruleus* and 25 OB and 15 P-morph females of *N. greenwoodi*. In *N. omnicaeruleus* at Makobe Island frequencies of the three different colour morphs P, OB and WB oscillated between 1991 and 2005 (Fig. 3), but did not undergo significant directional change (GLM estimate: -0.0105 $\pm$0.0138, z = -0.766, P = 0.443). Chi-square tests however, revealed significant differences in colour morph frequencies between some years. In 1993, the proportion of WB morphs was significantly higher than expected (n_{total} = 129, n_{WB} = 44, χ^2 = 5.34, d.f. = 1, P = 0.021). In 2002, the proportion of WB morphs was significantly lower than expected (n_{total} = 216, n_{WB} = 39, χ^2 = 5.94, d.f. = 1, P = 0.015) (see also Maan et al., 2008). In *N. greenwoodi* at Igombe Island the OB morph was numerically dominant in both sampling periods, represented by 61% and 62.5% of the females caught in 1993/1995 and 2005 respectively. There was no significant difference between the years (χ^2 = 0.107, d.f. = 1, P = 0.743).

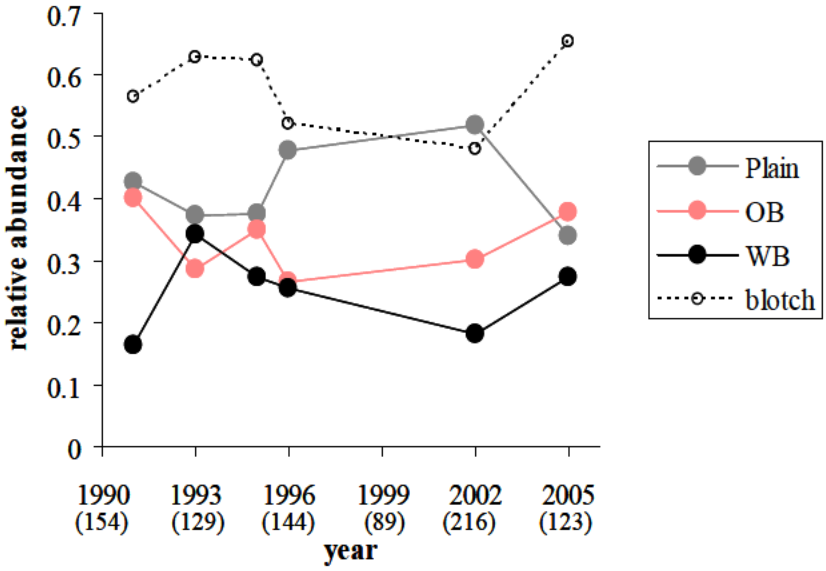

Figure 3. Relative frequencies of females of the three colour morphs separately, and of the two blotched morphs pooled together, of *N. omnicaeruleus* at Makobe Island for the period 1991- 2005.

Variation and divergence in colouration

Nested ANOVAs revealed significant differences between morphs in all measures of colour, while species differed only in two of them (Table 1). The latter was expected, given that the species display the same female colour polymorphisms except that *N. greenwoodi* lacks the WB morph. Note that the two species differ in male nuptial colouration, which was not included in this analysis.

OB morphs differed significantly in hue from both P and WB morphs, both within and between species (Fig. 4a). The WB colour morph had fewer but larger blotches than did the OB colour morph, covering a larger area of the body (Fig 4b, c, d).

Table 1. Results of nested ANOVAs for females in 4 measures of colour and 12 standard morphological measurements for morphs nested within species. Significant P-values are in bold.

	morph(species)			species		
character	d f.	*F*	*P*	d f.	*F*	*P*
hue	3	52.978	**0.000**	1	0.587	0.445
no. blotches	3	48.789	**0.000**	1	2.522	0.115
black pixels	3	318.776	**0.000**	1	114.012	**0.000**
biggest blotch	3	74.978	**0.000**	1	38.836	**0.000**
SL	3	4.072	**0.008**	1	45.408	**0.000**
HL	3	0.431	0.731	1	14.715	**0.000**
HW	3	2.263	0.085	1	1.020	0.315
BD	3	0.766	0.515	1	4.123	**0.045**
LJL	3	0.154	0.927	1	0.379	0.540
LJW	3	0.889	0.449	1	6.025	**0.016**
SnL	3	5.808	**0.001**	1	0.788	0.377
SnW	3	6.118	**0.001**	1	10.794	**0.001**
ChD	3	0.653	0.583	1	2.743	0.100
POD	3	1.310	0.275	1	1.952	0.165
IOW	3	0.330	0.804	1	6.436	**0.013**
EyL	3	0.536	0.659	1	42.569	**0.000**
EyD	3	2.406	0.071	1	55.592	**0.000**

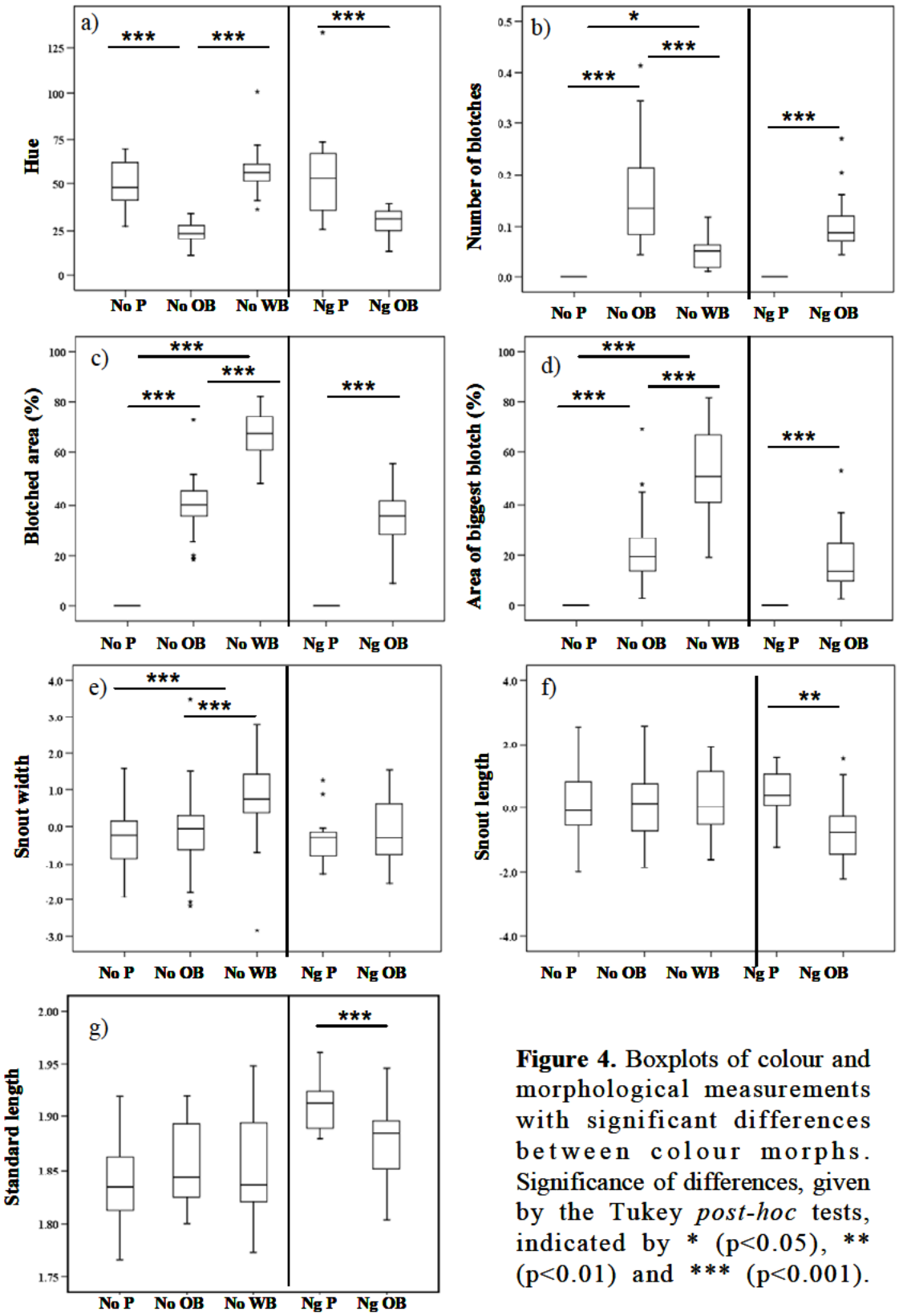

Figure 4. Boxplots of colour and morphological measurements with significant differences between colour morphs. Significance of differences, given by the Tukey *post-hoc* tests, indicated by * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

The total area covered by blotches and the size of the biggest blotch loaded very heavily on the first principal component of colour variation, which explained 51.1% of the variance. PC2 was primarily an axis of hue and number of blotches (Table 2).

Bivariate plots of colouration PC scores show that plain morphs from both species clustered around the centre of the abscissa at origin with little variation along PC1 (Fig. 5 a, b, c). This was likely a consequence of PC1 explaining predominantly variance in size and number of blotches, which are absent in P colour morphs. Interspecific female colour differences appear to be mostly due to the existence of a colour morph in *N. omnicaeruleus*, WB, which does not exist in *N. greenwoodi*. Hence *N. omnicaeruleus* simply occupied a larger area in female colour space but fully overlapped with the space occupied by *N. greenwoodi* (Fig. 5 a).

Table 2. Loadings of variables on the first three principal component axes of colouration (above horizontal line) and of morphometry and the percent variance explained by each axis. Traits with highest loadings are in bold.

trait	PC1	PC2	PC3
hue	-0.205	**0.859**	-
no blotches	0.484	**-0.725**	-
black pixels	**0.963**	0.182	-
biggest blotch	**0.895**	0.393	-
% variance explained	50.10%	36.30%	-
HL	**0.74**	-0.22	0.32
HW	**0.84**	0.11	-0.16
BD	0.41	-0.22	-0.31
LJL	0.58	0.16	0.2
LJW	0.53	0.48	-0.21
SnL	0.43	0.36	0.42
SnW	**0.65**	0.46	-0.3
ChD	0.32	0.26	0.36
POD	0.18	0.18	**0.66**
IOW	0.46	0.3	-0.45
EyL	**0.7**	**-0.61**	-0.06
EyD	0.58	**-0.72**	0.08
%variance explained	31.85	14.82	11.39

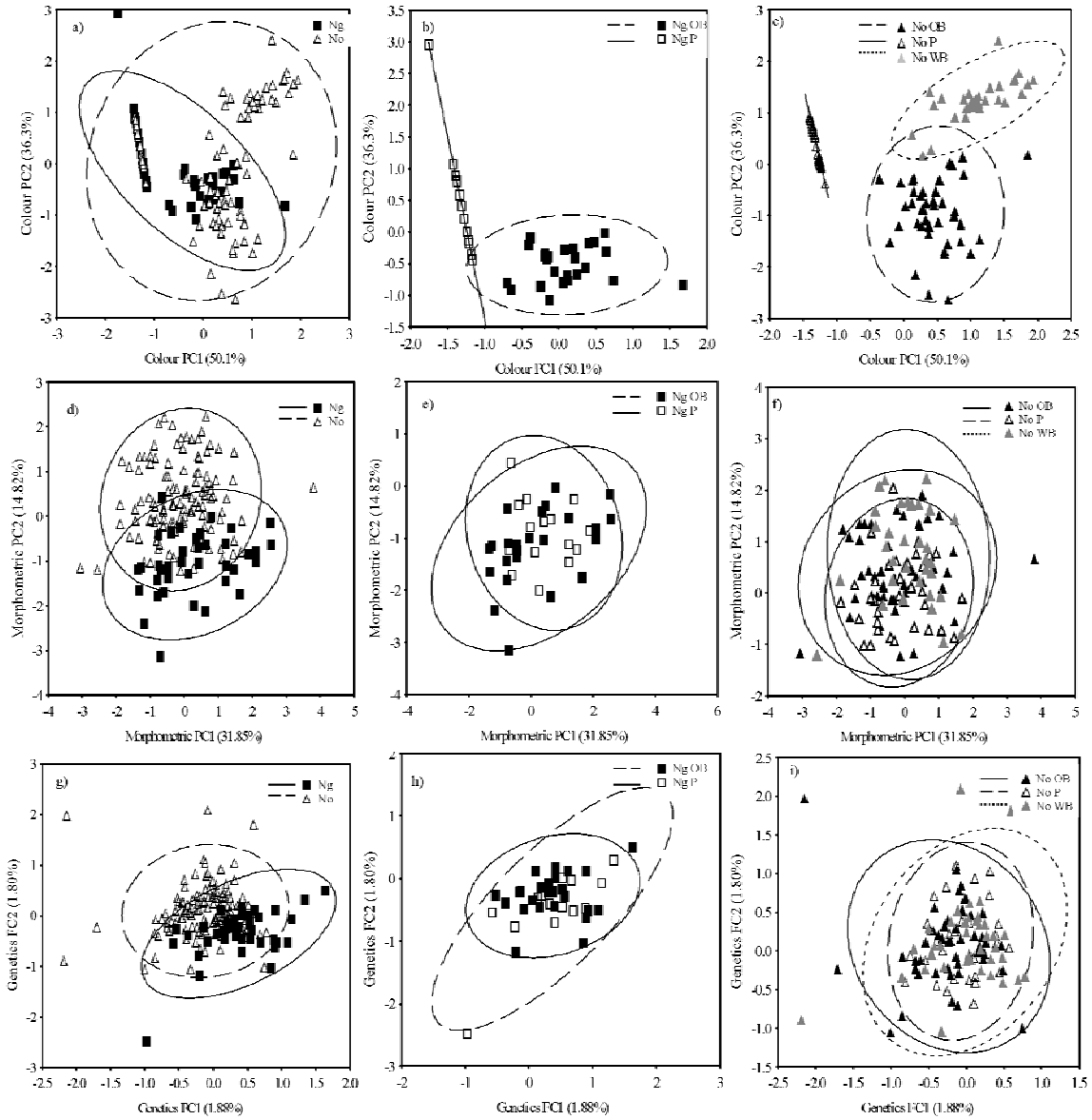

Figure 5. Top panels: bivariate plots of PC1 versus PC2 scores of colour characters for a) *N. omnicaeuruleus* (No) and *N. greenwoodi* (Ng) b) P and OB morphs of *N. greenwoodi*, c) OB, P and WB morphs of *N. omnicaeuruleus* ; Middle panels : bivariate plots of PC1 versus PC2 of morphometric characters for d) *N. omnicaeuruleus* (No) and *N. greenwoodi* (Ng), e) P and OB morphs of *N. greenwoodi*, f) OB, P and WB morphs of *N. omnicaeuruleus*; Bottom panels: FC1 versus FC2 of genotypic variation (9 microsatellite loci genotypes) for g) *N. omnicaeuruleus* (No) and *N. greenwoodi* (Ng), h) P and OB morphs of *N. greenwoodi*, i) OB, P and WB morphs of *N. omnicaeuruleus* . Ellipses represent the 95 % confidence intervals of each group.

Morphometric variation and divergence

Eight of 12 morphometric distances differed significantly between the two species (Table 1). Colour morph, nested within species, explained significant portions of variance in three morphological distances namely, standard length (SL), snout length (SnL) and snout width (SnW). Between conspecific colour morphs, *N. omnicaeruleus* WB had a significantly wider snout than *N. omnicaeruleus* OB and P (Fig. 4e), and *N. greenwoodi* P had significantly longer snout (Fig. 4f) and body (Fig. 4g) than *N. greenwoodi* OB. No morphological difference was found between OB and P morphs of *N. omnicaeuruleus*. Comparisons between heterospecific morphs revealed that *N. omnicaeruleus* WB had a significantly wider snout than all *N. greenwoodi* morphs and a significantly longer snout than *N. greenwoodi* OB, and that *N. greenwoodi* P had a significantly longer body than *N. omnicaeruleus* P.

The first principal component of morphometry explained 31.85% of the variance among fish and was mostly constructed from head length (HL) and width (HW), snout length (SnL) and eye length (EyL) (all with positive loading; Table 2). Eye width and length had the highest negative loading on PC2 which accounted for 14.82% of the total variance. Pre-orbital depth loaded heavily positively on PC3 accounting for 11.39% of the total variance among individuals. Combining the results of the Principal Component Analysis with those of the ANOVAs on univariate size-corrected morphological distances demonstrated that the distances that were most variable among individuals and also discriminated between species were head length, snout width, eye width and eye depth. Visualisations of these differences showed that most of the morphological differences between species occurred along PC2 with no differentiation along PC1 (Fig. 5d) or PC3 (Fig. 6a) and that very little morphological distinction could be made among sympatric conspecific colour morphs (Fig. 5e, f; Fig. 6 b,c).

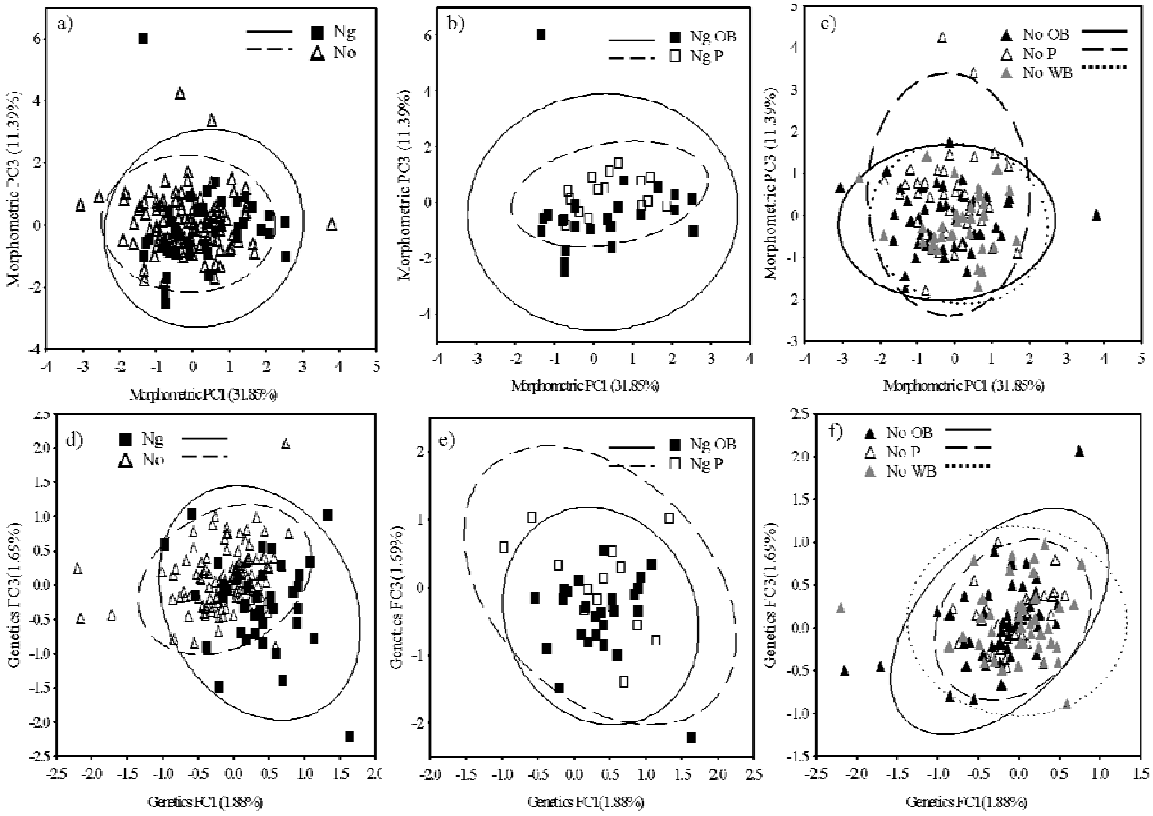

Figure 6. Top panels: bivariate plots of morphometric PC1 versus morphometric PC3 of morphological characters for d) *N. omnicaeuruleus* (No) and *N. greenwoodi* (Ng), e) P and OB morphs of *N. greenwoodi*, f) OB, P and WB morphs of *N. omnicaeuruleus*; Bottom panels: FC1 versus FC3 of genotypic variation (9 microsatellite loci genotypes) for g) *N. omnicaeuruleus* (No) and *N. greenwoodi* (Ng), h) P and OB morphs of *N. greenwoodi*, i) OB, P and WB morphs of *N. omnicaeuruleus* . Ellipses represent the 95 % confidence intervals of each group.

Patterns of gene flow and neutral genetic differentiation

Most microsatellite markers exhibited a high degree of polymorphism among the individuals sampled (Appendix 1). Of the 164 females, nine were not successfully genotyped at several loci and hence excluded. 155 females and 75 males were included in the analysis. A total of 282 alleles were found over all loci, ranging from 5 to 70 alleles per locus (mean: 23.36, SD: 11.93).

Before sequential Bonferroni correction, for *N. omnicaeruleus* females (all morphs pooled), deviations from HWE were observed at locus Ppun7 ($P = 0.015$) and seven tests out of 36 (20%) showed significant linkage disequilibrium (LD) (Appendix 1). After sequential Bonferroni correction, no departures from HWE remained significant at species level, but one test of linkage disequilibrium did. For intraspecific colour morphs, before Bonferroni correction, significant departures from Hardy-Weinberg equilibrium were observed at locus Ppun7 in OB ($P = 0.013$) and at locus OSU20d in WB ($P = 0.028$). Ppun7 deviation from HWE in OB was still significant after sequential Bonferroni correction. Out of 36 tests of LD in each morph, three tests in OB, one in WB and seven in P were significant before sequential Bonferroni correction (Table 3). Four tests were still significant, all in the P morph, after Bonferroni correction. For *N. omnicaeruleus* P-morph males, before sequential Bonferroni correction, there were no significant departures from HWE but 13 tests of LD were significant. None of these coincided with those in the female morphs. Nine of these tests were still significant after the correction.

For *N. greenwoodi* females no deviations from HWE were observed, but 14 tests showed significant linkage disequilibrium, before sequential Bonferroni correction (Table 3). Two tests were still significant after the correction. For intraspecific morphs, before sequential Bonferroni correction, significant departure from HWE was observed at locus OSU19t in the P morph ($P = 0.028$), but significance was lost after Bonferroni correction. Out of 36 tests of linkage disequilibrium in each morph, two tests in OB and 14 in P were significant before sequential Bonferroni correction. Four tests were still significant after the correction, again all in the P morph. For P-morph males, before sequential Bonferroni correction, significant departures from HWE were observed at loci OSU19t and OSU20d. OSU19t deviation from HWE was still significant after sequential Bonferroni correction. Five out of 36 tests of LD were significant before, but not after, sequential Bonferroni correction. Two of these coincided with significant tests of LD in the P-morph females.

Factorial Correspondence Analysis of the microsatellite data from females showed substantial overlap of species (Fig. 5g, Fig 6d) and of conspecific colour morphs in genotypic space (Fig.5 h, i, Fig 6 e, f). FC1 explained 1.88% of the total variance while FC2 and FC3 explained 1.80% and 1.69%, respectively.

Table 3. P-values of tests of linkage disequilibrium for 9 microsatellite loci. Significant P-values, before sequential Bonferroni correction, are in bold.

comparison	*N.omnicaeruleus females*	No OB	No WB	No P	No P males	*N.greenwoodi females*	Ng OB	Ng P	Ng P males
ppun5,osu16d	**0.028**	0.401	0.365	0.094	0.324	**0.002**	0.297	0.268	0.434
ppun5,osu19t	0.059	**0.044**	0.590	0 121	0.084	**0.006**	0.445	0.226	0.954
ppun5,osu20d	0.097	0.081	0.603	0 196	0.171	**0.007**	0.071	0.390	0.200
ppun5,TMOm5	0.141	0.211	0.687	0 208	**0.000**	**0.020**	0.923	0.312	0.774
ppun5,ppun17	0.228	0.366	0.830	0 271	**0.000**	**0.026**	0.390	**0.009**	0.785
ppun5,ppun21	0.237	0.666	0.835	0.434	**0.000**	0.100	0.078	**0.000**	**0.043**
ppun5,ppun32	0.709	0.675	0.885	0.477	**0.009**	0.477	0.255	**0.001**	**0.049**
ppun5,ppun7	0.944	0.882	0.920	0.707	0.602	0.951	0.821	**0.000**	0.449
ppun7,osu16d	**0.045**	**0.008**	0.125	0.086	0.620	**0.006**	0.962	0.189	0.604
ppun7,osu19t	0.086	0.125	0.146	0 156	**0.004**	0.060	0.706	0.248	0.926
ppun7,osu20d	0.125	0.290	0.335	0 198	0.112	0.162	**0.023**	0.273	0.058
ppun7,TMOm5	0.181	0.398	0.445	0 364	**0.000**	0.259	0.809	0.368	0.304
ppun7,ppun17	0.266	0.605	0.504	0.451	**0.000**	0.266	0.509	**0.027**	0.060
ppun7,ppun21	0.561	0.816	0.612	0.491	**0.000**	0.393	0.443	**0.002**	0.627
ppun7,ppun32	0.900	0.956	0.833	0.895	0.469	0.677	0.383	**0.002**	0.664
ppun17,osu16d	**0.026**	0.062	0.226	**0.046**	0.389	**0.001**	0.588	0.481	0.308
ppun17,osu19t	0.096	0.142	0.325	0.051	0.109	0.081	0.347	**0.040**	0.740
ppun17,osu20d	0.363	0.486	0.402	0.107	0.185	0.234	0.333	0.248	0.397
ppun17,TMOm5	0.396	0.695	0.441	0.688	**0.000**	0.286	0.570	0.894	0.411
ppun17,ppun21	0.492	0.786	0.860	0.707	**0.000**	0.462	0.521	**0.041**	0.631
ppun17,ppun32	0.911	0.861	0.903	0.769	0.368	0.502	0.141	0.544	0.121
ppun21,osu16d	0.061	0.259	**0.003**	**0.043**	0.317	**0.006**	0.838	**0.037**	0.080
ppun21,osu19t	0.284	0.350	0.062	0.575	0.092	0.145	0.931	0.143	**0.030**
ppun21,osu20d	0.379	0.547	0.773	0.644	0.462	0.153	0.514	**0.009**	0.293
ppun21,TMOm5	0.529	0.795	0.789	0.689	**0.000**	0.435	0.260	0.085	0.080
ppun21,ppun32	0.675	0.851	0.992	0.823	0.232	0.584	0.478	**0.000**	0.786
ppun32,osu16d	0.080	0.110	0.536	0.152	0.355	**0.034**	0.353	**0.041**	0.233
ppun32,osu19t	0.207	0.111	0.600	0.528	0.150	**0.045**	0.240	**0.012**	0.424
ppun32,osu20d	0.396	0.812	0.803	0.799	**0.039**	0.145	0.331	0.073	0.258
ppun32,TMOm5	0.494	0.938	0.906	0.810	0.430	0.316	0.312	0.478	0.952
osu16d,osu19t	**0.001**	0.122	0.121	**0.000**	0.120	**0.006**	0.366	0.765	0.445
osu16d,osu20d	**0.008**	0.133	0.605	**0.001**	0.834	**0.024**	0.374	0.102	**0.013**
osu16d,TMOm5	**0.039**	0.983	0.762	**0.013**	0.262	0.115	0.247	0.624	0.871
osu19t,osu20d	**0.042**	**0.018**	0.686	**0.020**	0.387	**0.001**	0.240	0.066	0.331
osu19t,TMOm5	0.064	0.348	0.871	0.122	**0.018**	**0.007**	**0.002**	0.742	**0.018**
osu20d,TMOm5	0.159	0.852	0.527	**0.001**	0.204	0.185	0.302	0.192	0.770

Neutral genetic differentiation between species was low but highly significant (F_{ST} = 0.017, P <0.001) (Table 4). Among the sympatric colour morphs, the P and OB morphs of *N. omnicaeruleus* were significantly differentiated at two out of nine loci (Appendix 1). No other significant differences were observed between sympatric colour morphs.

Table 4. Pairwise F statistics between colour morphs. F_{ST} values and significance (* p<0.05, ** p<0.01, ***p<0.001, n.s. not signicant) in the lower matrix and confidence intervals in the upper matrix.

	No P	No OB	No WB	Ng P	Ng OB
No P		0.000-0.006	-0.001-0.005	0.002-0.020	0.003-0.017
No OB	0.0034 n.s.		-0.003-(-0.001)	0.001-0.0253	0.010-0.028
No WB	0.0019 n.s.	-0.0022 n.s.		0.004-0.027	0.007-0.29
Ng P	0.0109n.s.	0.0123n.s.	0.0129n.s.		-0.009-0.012
Ng OB	0.0096n.s.	0.0186***	0.0174**	0.0016 n.s.	

Between the allopatric heterospecific morphs, the highest and significant differentiations were observed when *N. greenwoodi* OB were compared with OB (F_{ST} = 0.0186, P <0.001) and WB (F_{ST} = 0.0174, P <0.001) morphs of *N. omnicaeruleus* but there was no significant multilocus differentiation between females of the P morphs (F_{ST} = 0.011, P = 0.230) (Table 4). When P-morph males of the different species were compared, however, neutral genetic divergence between them was slightly higher than between P morph females (F_{ST} = 0.015) and was highly significant (P < 0.001). Single locus F_{ST} values had a narrow distribution both among sympatric colour morphs and among allopatric morphs, the highest F_{ST} value being 0.111 (Apendix 2).

The very different frequencies of LD in different colour morphs and the variable extents of differentiation between sympatric and between allopatric morphs suggested that gene flow between colour morphs may not be random. Specifically, these results suggested that the subpopulations phenotypically defined as P are genetically heterogeneous mixes of individuals sampled from more than one randomly mating population. To explore these patterns further, we calculated individual assignment tests between all five phenotypically and geographically defined groups of females and the two groups of males, separately (Table 5). Further, we compared allelic richness and the difference between observed and expected heterozygosities at the 9 loci between the sympatric P and blotched morphs using paired t-tests. The assignment of individuals between females (χ^2 = 62.67, d.f. = 1, P<0.0001) and between males (χ^2 = 29.89, d.f. = 1, P<0.0001) of the two species was highly significantly different from random. Out of 116 females of *N. omnicaeruleus* only four (3.4%) were assigned to *N. greenwoodi*. Yet, over a third (38%) of the individuals of *N. greenwoodi* were assigned to *N. omnicaeruleus*. Eight out of 45 (18%) males of *N. omnicaeruleus* were assigned to *N. greenwoodi* and eight out of 30 (27%) males of *N. greenwoodi* were assigned to *N. omnicaeruleus*. These results suggest asymmetric gene flow from *N. omnicaeruleus* to *N. greenwoodi*, particularly so in females.

Table 5. Contingency table with results of the assignment test based on allele frequencies. Represented are numbers of individuals of a colour morph (rows) that were assigned to their own morph and to other morphs (columns).

	No P	No OB	No WB	No P males	Ng P	Ng OB	Ng P males	total
No P	16	13	8	-	0	2	-	39
No OB	8	19	19	-	0	0	-	46
No WB	8	16	5	-	0	2	-	31
No P males	-	-	-	38	-	-	7	45
Ng P	2	2	3	-	1	6	-	14
Ng OB	5	0	3	-	3	14	-	25
Ng P males	-	-	-	7	-	-	23	30

Within *N. omnicaeruleus* females, the assignment of genotypes of the three colour morphs was also significantly different from random (χ^2 = 13.59, d.f. = 4, P = 0.008). Highly significant probability of correct assignment was observed between the colour morphs P and OB (χ^2 = 15.37, d.f.= 1, P = 0.00009), but not between OB and WB, where the tendency was that individuals of both phenotypes were assigned to OB (χ^2 =2.96, d.f.= 1, P = 0.08), and also not between P and WB (χ^2 = 1.28, d.f.= 1, P = 0.25). Whereas 19 of 46 OB females were assigned to the OB population, 19 others to WB and only 8 to P, only 16 of 39 P individuals were assigned to P, all of the others being assigned to one of the blotched phenotypes. Within *N. greenwoodi* females, the assignment of the two colour morphs was highly significantly different from random too (χ^2 =12.702, d.f. = 1, P = 0.0003), but also here, while most OB were indeed assigned to the OB population, most P individuals too were assigned to OB.

To ask if the excess of LD in the P-morph females of both species was due to the admixture within the P category of females of blotched and non-blotched ancestry, we recalculated the amount of LD after reassigning the P morph females to P or to blotch (OB and WB pooled together) according to the output from the assignment tests. In *N. omnicaeruleus* only four out of 36 tests were significant before Bonferroni correction, and none remained after Bonferroni correction. We could not perform this test on *N. greenwoodi* P because only five P morph females were assigned to P. However LD in the OB morph, with the nine P morph females assigned to OB included, was significant for three out of 36 tests before Bonferroni correction and only for one after it. These results are consistent with the hypothesis of admixture within the P category of females of blotched and non-blotched ancestry.

Consistent too with this hypothesis, there was a significantly more positive difference between observed and expected heterozygosity in the P morph than in the OB morph of *N. omnicaeruleus* (t = -3.27, d.f. = 8, P (one tailed) = 0.006) and an almost significantly more positive difference in the P morph than in the WB morph (t = -1.56, d.f. = 8, P (one tailed) = 0.07). No such trend was observed in *N. greenwoodi* at Igombe Island. Allelic richness did not differ between any of the morphs.

Relationships between phenotypic and genetic differentiation

None of the relationships between morphological distance, colour distance and genetic distance had slopes significantly different from zero, either between allopatric heterospecific or between sympatric conspecific colour morphs (Fig. 7). However, given that both slope and intercept between Euclidean distances in morphometry and microsatellite D_s were very similar for conspecific and heterospecific comparisons we additionally combined both types of comparisons into a single analysis. This resulted in a very significant correlation between genetic and morphological divergence (R^2= 0.826, $F_{1,8}$= 38.01, P<0.001).

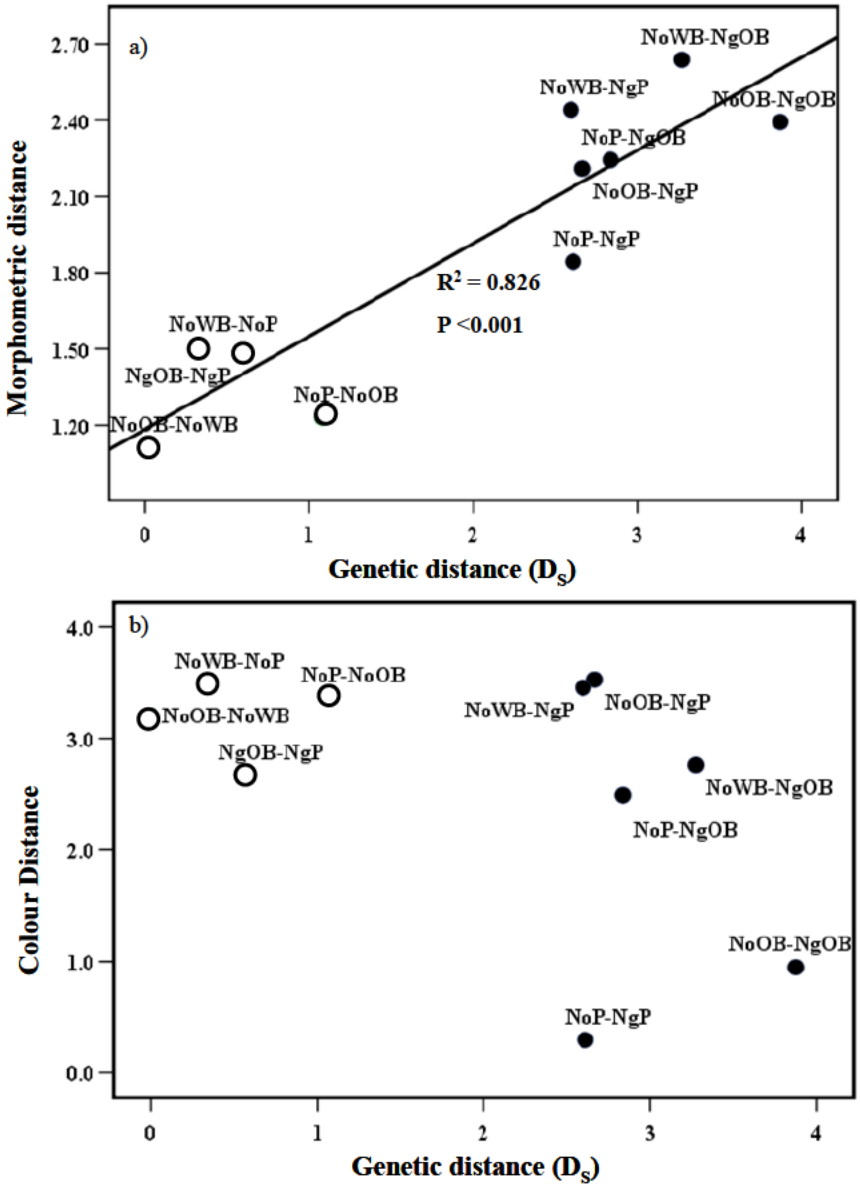

Figure 7. a) normalized Euclidean distances based on morphometric measurements versus genetic distances (D_S) and c) normalized Euclidean distance based on colour measurements versus genetic distances (D_S) for intra-and interspecific colour morphs.

Discussion

Colour polymorphisms as an arrested incipient stage of speciation

When phenotypic polymorphisms are stabilized by negative frequency dependent selection, they can theoretically provide the seed for sympatric speciation (van Doorn et al., 2004). Here we investigated genetic and morphological correlates of a truly sympatric and syntopic colour polymorphism in two cichlid species living on different islands in Lake Victoria. Over 14 years, corresponding to 14 generations, the frequencies of the three colour morphs of *N. omnicaeruleus* at Makobe Island have oscillated, but not changed directionally, suggesting that this polymorphism is not transient at ecological time scale. Using nine unlinked microsatellite loci, we found indications of non-random mating between colour morphs within and between species, and weak traces of neutral genetic differentiation between sympatric morphs at Makobe Island, consistent with the hypothesis that the polymorphism is an arrested incipient stage of sympatric speciation.

Behavioural work suggested potential for partial assortative mating between morphs of *N. omnicaeruleus* at Makobe Island (Seehausen et al., 1999). P-morph females and P-morph and blotched males had mating preferences whereas blotched females mated randomly. Blotch genes were found to be linked to dominant female determiners and together with those, linked to X. The authors hypothesized that the resulting rarity of blotched males creates selection against a preference for blotched males in blotched females, effectively arresting progress toward speciation. P males fell into two preference classes depending on their pedigree. P sons of P parents preferred P females, but P sons of blotched parents, just like blotched males, preferred blotched females as partners. Mating between blotched females and P sons of blotched females generates phenotypic P females of blotched ancestry. Being courted predominantly by P males, they are a potential source of gene flow from blotched to P morphs. Consistent with this, we found no evidence for reproductive isolation between sympatric conspecific colour morphs: F_{ST} values were very low and mostly non-significant. Consistent too with predictions based on the behavioural work though, we found a high frequency of linkage disequilibrium in P morphs of both species, and a significantly larger difference between observed and expected heterozygosity in the P morph than in the blotch morphs of *N. omnicaeruleus*. However, allelic richness did not differ between any of the morphs of either species, indicating that the excess of LD in P morphs is not due to gene flow from other species. These results hence suggest that at both islands, the P phenotype is genetically more heterogeneous than the blotched phenotypes, due to admixture of individuals of blotched and non-blotched ancestry.

The assignment tests provided further evidence of slight genetic differentiation and restricted gene flow between blotched and plain colour morphs. Whereas there were significant indications of neutral genetic differentiation between P and OB morphs of *N. omnicaeruleus*, there was also a strong trend in both species to assign a large fraction of P with highest likelihood to OB. Removing these from the P populations removed most of the linkage disequilibria that was otherwise observed in this morph. These females might hence be of blotched or mixed ancestry as predicted if blotched females breed with P males and produce phenotypically P offspring alongside phenotypically blotched offspring. The even higher levels of LD in P-morph males of *N. omnicaeruleus* further supports this scenario. A higher amount of LD would be expected because even matings

between blotched females and blotched males produce P male offspring of blotched ancestry.

Genetic and eco-morphological divergence between parapatric species

Significant genetic differentiation and eco-morphological differences were found between the two geographically distinct but nearby populations of the sister species *N. omnicaeruleus* and *N. greenwoodi*. Most of the significantly different morphological traits, such as head length, lower jaw width, snout width and inter-orbital width, are related to resource use. Although the two species occupy corresponding niches in these and other rocky shore communities, the islands Makobe and Igombe differ in physical characteristics, such as shore slope, boulder size and water clarity (Seehausen et al., 1997; Magalhaes et al., in prep). It is possible that differences in resource availability are driving morphometric differentiation between these species, as has been shown for populations within either of both species (Bouton et al., 1999). However, we found only weak, though significant, genetic divergence between females (F_{ST} = 0.012) and males (F_{ST} = 0.015) of the two species. This is unexpected, because *Neochromis* are habitat-stenotopic rock-dwelling species, and Makobe and Igombe Islands are separated by 5km of deep open water with soft bottom, supposedly an alienating habitat for stenotopic rock cichlids. This level of genetic differentiation corresponds to what has been found between sympatric hybridizing incipient species of other cichlids in Lake Victoria (Seehausen et al., 2008). It thus appears that the sister species *N. omnicaeruleus* and *N. greenwoodi,* with their complementary geographical distributions around either offshore or inshore rocky habitats respectively, should be considered parapatric incipient species with ongoing gene flow between them. This is consistent with a previous study using SINE loci (Terai et al., 2006). Interestingly, there seems to be much more gene flow from the offshore Makobe Island to the inshore Igombe Island than vice versa.

Interestingly too, genetic differentiation between the parapatric species, though significant overall, varied considerably depending on which colour morphs we compared. The multilocus F_{ST} values between the OB morph of *N. greenwoodi* and the WB and OB morphs of *N. omnicaeruleus* were significant, whereas none of the comparisons involving P-morph females were significant. This indicates that interspecific gene flow may happen predominantly through the P morph. P is presumably the ancestral morph, occurring in all populations of both species and all other *Neochromis* species (Seehausen et al., 1999). However, P-morph males of the two species were somewhat more strongly, and indeed significantly differentiated. The difference between P-morph females and males could be a consequence of the small sample size of *N. greenwoodi* P females. Alternatively it can be consequence of, as mentioned before, a larger fraction of individuals of mixed or blotched ancestry in P-morph males.

We found that the extent of neutral genetic differentiation between colour morphs within and between the species combined correlated positively with eco-morphological divergence. Eco-morphological distances and genetic distances were low between intraspecific morphs and high between interspecific morphs. This is consistent with the notion that gene flow constrains eco-morphological differentiation unless divergent selection is very strong (Rasanen & Hendry, 2008). It also suggests that for speciation to happen in response to selection on a polymorphism that is not associated with any microhabitat differentiation (syntopy) between morphs (Maan et al., 2008), geographical

parapatry may be required.

Frequency-dependent sexual selection and ecological character displacement between colour morphs

Several theoretical studies on colour polymorphims have proposed mechanisms that allow for their stabilization with completely overlapping niches (Mikami et al., 2004, Seehausen & Schluter, 2004; van Doorn et al. 2004; Gray & McKinnon, 2006). In these models, the key to stabilization of polymorphisms is negative frequency-dependent selection that confers an advantage to rare phenotypes: if individuals direct more aggression towards rivals that phenotypically resemble themselves than towards different phenotypes, rare phenotypic varieties would enjoy an advantage because they receive less aggression. This advantage would decrease as the rare phenotype becomes progressively more common and the number of intramorph encounters increases. An own-phenotype bias would therefore facilitate both, the invasion of a new phenotype, and the stable coexistence of morphs. Experimental studies using females of the three colour morphs of *N. omnicaeruleus* from Makobe Island did indeed reveal strong own-morph aggression biases in females of the three morphs (Dijkstra et al., 2008). This supports the hypothesis that in this population colouration is subject to negative frequency-dependent intrasexual selection.

There are several empirical examples showing how species that diverge in secondary sexual characters also undergo ecological character displacement (Chiba 1999; Panhuis et al., 2001; McKinnon & Rundle, 2002). The difference found in snout width between the WB and the other two sympatric colour morphs in *N. omnicaeruleus* and the difference in snout length between the OB and the P morph in *N. greenwoodi* could indicate potential for ecological character displacement. We do not know currently if these differences are phenotypically induced or genetic. However, we did find that variation in snout width was heritable in another Lake Victoria cichlid species (Magalhaes et al., in press).

In conclusion, we show that X-linked female colour polymorphisms appear relatively stable through ecological time. Colour variation seems to influence gene flow between sympatric morphs and between allopatric populations, consistent with documented behavioural mating preferences. However, differentiation between the morphs at neutral loci remains weak, suggesting these polymorphisms behave like arrested early stages of speciation. Our data suggest the possibility of ecological character displacement following or parallel to divergence in colouration.

Acknowledgements

We would like to thank Mhoja Kayeba and Mohammed Haluna for providing crucial assistance during fieldwork, Martine Maan for providing data on frequencies of colour morphs for the year of 2002, Denis Roy and Michele Pierotti for comments on an earlier version of the manuscript and the members of the eawag Fish Ecology & Evolution lab for helpful feedback and discussions. We thank the Tanzanian Fisheries Research Institute and the director of its Mwanza Centre, Mr. E.F.B. Katunzi, for hospitality and access to facilities, and the Tanzania Commission for Science and Technology for research permits. This research was supported by the Swiss National Science Foundation, Project 3100A0-106573 attributed to OS.

References

Barel, C. D. N., Van Oijen, M. J. P., Witte, F. & Wittemaas, E. L. M. 1977. Introduction to Taxonomy and Morphology of Haplochromine Cichlidae from Lake Victoria - Manual to Greenwoods Revision Papers. *Neth. J. Zool.* **27**: 333-338.

Barluenga, M., Stolting, K. N., Salzburger, W., Muschick, M. & Meyer, A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439: 719-723.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. & Bonhomme, F. 1996. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Bouton, N., Witte, F., van Alphen, J. J. M., Schenk, A. & Seehausen, O. 1999. Local adaptations in populations of rock-dwelling haplochromines (Pisces : Cichlidae) from southern Lake Victoria. *P. Roy. Soc. B-Biol. Sci.* **266**: 355-360.

Chiba, S., 1999. Character displacement, frequency-dependent selection, and divergence of shell colour in land snails Mandarina (Pulmonata). *Biol. J. Linn. Soc.* **66**: 465-479.

Coyne, J. A. & Orr, H. A. 2004. *Speciation*. Sunderland, Massachusetts, Sinauer Associates.

Dieckmann, U. & Doebeli, M. 1999. On the origin of species by sympatric speciation. *Nature* **400**: 354-357.

Dijkstra, P. D., Seehausen, O. & Groothuis, T. G. G. 2008. Intrasexual competition among females and the stabilization of a conspicuous colour polymorphism in a Lake Victoria cichlid fish. *P. Roy. Soc. B-Biol. Sci*. **275**: 519-526.

Excoffier, L., Estoup, A. & Cornuet, J. M. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727-1738.

Fleming, I. A. & Gross, M. R. 1994. Breeding Competition in a Pacific Salmon (Coho, Oncorhynchus-Kisutch) - Measures of Natural and Sexual Selection. *Evolution* 48: 637-657.

Gavrilets, S. 2004. *Fitness lanscapes and the origin of species*. Princeton University press.

Glaubitz, J. C. 2004. CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol. Ecol. Notes* **4**: 309-310.

Gray SM, McKinnon JS, 2006. Linking colour polymorphism maintenance and speciation. *Trends Ecol. Evol*. **22**: 71-79.

Holzberg, S. 1978. Field and Laboratory Study of Behavior and Ecology of Pseudotropheus Zebra (Boulenger), an Endemic Cichlid of Lake Malawi (Pisces-Cichlidae). Z. Zool. Syst. Evol. 16: 171-187.

Jonsson, B., Jonsson, N. 2001. Polymorphism and speciation in Arctic charr. *J. Fish Biol*. **58**: 605-638.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Rev. Gen*. **5**: 288-298.

Kondrashov, A. S. & Kondrashov, F. A. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351-354.

Lande, R., Seehausen, O. & van Alphen, J. J. M. 2001. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* **112**: 435-443.

Maan, M. E., Eshuis, B., Haesler, M. P., Schneider, M. V., van Alphen, J. J. M. & Seehausen, O. 2008. Color polymorphism and predation in a Lake Victoria cichlid fish. *Copeia* **3**: 621-629.

Magalhaes, I. S., Mwaiko, S., Schneider, M. V. & Seehausen, O. 2008. Divergent selection and phenotypic plasticity duirng incipient speciation in Lake Victoria cichlid fish. *J. Evol. Biol*., in press.

Marsh, A. C., Ribbink, A. J. & Marsh, B. A. 1981. Sibling Species Complexes in Sympatric Populations of Petrotilapia Trewavas (Cichlidae, Lake Malawi). *Zoo. J. Linn. Soc.* **71**: 253-264.

McKinnon, J. S. & Rundle, H. D. 2002. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* **17**: 480-488.

Mikami, O. K., Kohda, M. & Kawata, M. 2004. A new hypothesis for species coexistence: male-male repulsion promotes coexistence of competing species. *Popul. Ecol*. **46**: 213-217.

Panhuis, T. M., Butlin, R., Zuk, M. & Tregenza, T. 2001. Sexual selection and speciation. *Trends Ecol. Evol*. **16**: 364-371.

Rasanen, K. & Hendry, A. P. 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol. Lett*. **11**: 624-636.

Rice, W. R. 1989. Analyzing Tables of Statistical Tests. Evolution 43: 223-225.

Schliewen, U., Tautz, D. & Pääbo, S. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**: 306-308.

Schliewen, U., Rassmann, K., Markmann, M., Markert, J., Kocher, T. & Tautz, D. 2001. Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Mol. Ecol*. **10**: 1471-1488.

Seehausen, O. & Bouton, N. 1996. Polychromatism in rock dwelling Lake Victoria cichlids: types, distribution and observations on their genetics. In: *The Cichlids Yearbook*, Vol. 6 (Konings, A. E., ed.). pp. 36-45. Cichlid Press, Germany.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids. Taxonomy, Ecology and Distribution.* Zevenhuizen, NL: Cichlid Press.

Seehausen, O. & Bouton, N. 1997. Microdistribution and fluctuations in niche overlap in a rocky shore cichlid community in Lake Victoria. *Ecol. of Freshw. Fish* **6**: 161-173.

Seehausen, O., vanAlphen, J. J. M. & Witte, F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808-1811.

Seehausen, O**.**, Lippitsch, E., Bouton, N., Zwennes, H., 1998. Mbipi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyol. Explor. Freshw*. **9**:129-228.

Seehausen, O., van Alphen, J. J. M. & Lande, R. 1999. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett*. **2**: 367-378.

Seehausen, O., 2000. Explosive speciation rates and unusual species richness in haplochromine cichlid fishes: Effects of sexual selection. *Adv. Ecol. Res*. **31**: 237-274.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *P. Roy. Soc. B-Biol. Sci.* **271**: 1345-1353.

Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., van der Sluijs, I., Schneider, M. V., Maan, M. E., Tachida, H., Imai, H. & Okada, N. 2008. Speciation through sensory drive in cichlid fish. *Nature* **455**: 620-626.

Taylor, M. I., Meardon, F., Turner, G., Seehausen, O., Mrosso, H. D. J. & Rico, C. 2002. Characterization of tetranucleotide microsatellite loci in a Lake Victorian, haplochromine cichlid fish: a *Pundamilia pundamilia* x *Pundamilia nyererei* hybrid. *Mol. Ecol. Notes* **2**: 443-445.

Terai, Y., Seehausen, O., Sasaki, T., Takahashi, K., Mizoiri, S., Sugawara, T., Sato, T., Watanabe, M., Konijnendijk, N., Mrosso, H. D. J., Tachida, H., Imai, H., Shichida, Y. & Okada, N. 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *Plos Biol*. **4**: 2244-2251.

van Doorn, G. S., Dieckmann, U. & Weissing, F. J. 2004. Sympatric speciation by sexual selection: A critical reevaluation. *Am. Nat*. **163**: 709-725.

Weir, B. S. & Cockerham, C. C. 1984. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**: 1358-1370.

Wu, L., Kaufman, L. & Fuerst, P. A. 1999. Isolation of microsatellite markers in Astatoreochromis alluaudi and their cross-species amplifications in other African cichlids. *Mol. Ecol*. **8**: 895-897.

Zardoya, R., Vollmer, D. M., Craddock, C., Streelman, J. T., Karl, S. & Meyer, A. 1996. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *P. Roy. Soc. B-Biol. Sci.* **263**: 1589-1598.

Appendix 1. Allelic variability at 9 microsatellite loci. Number of samples analysed in each group (N), number of alleles at each locus (A), allelic richness (AR), observed heterozygosity (HO) and expected heterozygosity (HE) at each locus. Significant differences between H_O and H_E are in bold, with * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

		No females	No OB	No WB	No P	No P males	Ng females	Ng OB	Ng P	Ng P males
	N	116	46	31	39	45	39	25	14	30
Ppun5	A	22	18	17	18	19	21	18	13	17
	AR	17.693	11.815	12.066	11.652	13.672	21.000	13.274	13.000	13.167
	HO	0.991	0.978	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	HE	0.929	0.931	0.932	0.928	0.939	0.944	0.949	0.944	0.940
Ppun7	A	24	21	20	20	20	19	18	9	21
	AR	19.546	12.525	13.405	12.435	13.673	19.000	12.896	10.000	15.149
	HO	**0.974***	0.957	**1.000***	0.974	0.854	1.000	1.000	1.000	1.000
	HE	**0.938**	0.932	**0.949**	0.934	0.939	0.929	0.941	0.896	0.944
Ppun17	A	15	13	10	14	15	14	12	8	12
	AR	12.414	9.069	8.789	10.536	11.281	13.864	8.903	7.746	9.855
	HO	0.897	0.870	0.903	0.923	0.902	0.919	0.920	0.917	0.862
	HE	0.905	0.894	0.889	0.920	0.911	0.886	0.880	0.859	0.899
Ppun21	A	23	20	18	18	17	19	14	13	18
	AR	18.648	12.034	11.977	12.939	12.523	19.000	11.397	13.000	14.140
	HO	0.940	0.870	0.968	1.000	0.854	0.889	0.920	0.818	0.862
	HE	0.936	0.930	0.931	0.946	0.929	0.941	0.924	0.957	0.945
Ppun32	A	8	8	6	5	6	6	6	5	5
	AR	6.650	5.460	5.222	3.563	4.687	6.000	4.702	5.000	4.898
	HO	0.759	0.783	0.742	0.744	0.683	0.694	0.720	0.636	0.586
	HE	0.719	0.729	0.756	0.668	0.651	0.682	0.680	0.714	0.705
osu16d	A	34	29	24	23	28	25	20	20	21
	AR	25.900	14.935	14.319	13.736	16.384	24.755	13.625	17.498	15.303
	HO	0.956	0.935	0.968	0.972	0.977	0.973	0.958	1.000	0.867
	HE	0.954	0.958	0.952	0.949	0.956	0.950	0.949	0.975	0.955
osu19t	A	34	25	22	25	26	22	18	14	24
	AR	23.865	12.929	13.461	13.355	14.757	22.000	11.654	12.829	14.982
	HO	0.886	0.804	0.968	0.919	0.909	0.861	0.913	0.769	**0.667*****
	HE	0.933	0.931	0.944	0.932	0.931	0.890	0.875	0.929	**0.930**
osu20d	A	70	44	38	40	45	36	32	20	31
	AR	41.845	18.084	18.255	17.357	20.555	35.402	14.674	17.365	19.161
	HO	0.921	0.935	**0.871***	0.946	0.977	0.946	0.958	0.923	**0.900***
	HE	0.98	0.98	**0.981**	0.974	0.978	0.959	0.945	0.969	**0.975**
TMO5	A	29	26	21	23	29	31	23	17	22

	AR	23.599	13.891	12.508	14.550	17.157	30.539	15.168	14.483	15.441
	HO	0.904	0.913	0.839	0.946	0.955	0.973	0.957	1.000	0.966
	HE	0.946	0.943	0.931	0.958	0.964	0.960	0.961	0.947	0.953
Total	HO	0.914	0.894	0.918	0.936	0.901	0.917	0.927	0.896	0.857
	HE	0.911	0.905	0.904	0.900	0.900	0.892	0.882	0.872	0.901
	Fis	0.002	0.023	0.001	-0.027	0.011	-0.014	-0.031	0.016	0.066

Appendix 2. Single locus and multilocus F_{ST} values -between intra- and interspecific colour morphs. Significant P-values are in bold. Significance of differences indicated by * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

comparison	ppun5	ppun7	ppun17	ppun21	ppun32	osu16d	osu19t	osu20d	TMOm5	multilocus
No OB-No WB	-0.004	-0.001	-0.004	-0.002	-0.004	-0.001	-0.004	0.002	-0.001	-0.002
No P-No OB	-0.002	0.006	**0.012***	0.001	0.008	0.004	-0.005	**0.005***	0.003	0.003
No P-No WB	-0.005	-0.005	0.008	-0.002	0.015	0.001	0.000	0.001	0.005	0.002
Ng P-Ng OB	-0.007	0.003	0.032	0.013	-0.028	-0.018	-0.012	0.012	0.006	0.002
No OB-Ng P	-0.006	**0.052*****	0.013	-0.006	0.033	0.001	0.010	**0.011***	0.010	0.012
No OB-Ng OB	-0.001	**0.015***	**0.032*****	0.011	**0.046*****	**0.014*****	**0.031*****	**0.020*****	0.005	**0.019*****
No WB-Ng P	0.001	**0.023***	0.019	-0.008	0.047	0.000	0.014	0.007	**0.02***	0.013
No WB-Ng OB	0.000	0.002	**0.030****	0.010	**0.056*****	0.009	**0.035*****	**0.020*****	0.003	**0.017****
No P-Ng OB	-0.002	0.001	0.014	0.003	0.009	0.004	**0.032*****	**0.020*****	0.000	0.010
No P-Ng P	-0.005	**0.030***	**0.030****	-0.006	0.007	-0.007	0.011	**0.018***	**0.014***	0.011
No males -Ng males	0.000	0.002	0.003	0.010	**0.111*****	0.006	**0.025*****	**0.007****	0.001	**0.015*****

Chapter 3

Eco-morphological but not genetic differentiation within cichlid fish populations correlates with the slope of resource gradients

Eco-morphological but not genetic differentiation within cichlid fish populations correlates with the slope of resource gradients

I. S. Magalhaes, B. Lundsgaard-Hansen, S. Mwaiko, O. Seehausen

Abstract

Theoretical models of ecological speciation in clines suggest environmental structure plays a major role in determining the extent of population divergence. Using this framework we investigated polymorphic populations of Lake Victoria cichlids living along environmental gradients that varied in slope and linearity. We sampled populations from three islands with replicate pairs of putative trophic morphs in the genus *Neochromis*, quantified morphology and dentition and typed population samples at 9 different microsatellite loci. Replicate pairs varied in extent of eco-morphological differentiation between one trait and more than 10 different traits. This variation was well predicted by variation in steepness and linearity of the resource gradient: the less steep and the more linear the gradient, the stronger the differentiation. A common garden experiment supported heritability of some of these traits. However, neutral genetic differentiation is less well predicted by the environment, and does not correlate well with morphological differentiation. We find evidence for deviations from random mating between sympatric morphs, but significant differentiation only at the island with intermediate resource gradient and intermediate eco-morphological differentiation. The eco-morphologically most strongly and the least strongly differentiated morph pairs are both undifferentiated at neutral loci. Divergent and disruptive selection, perhaps aided by genetic architecture of eco-morphological traits, seem sufficient to maintain eco-morph differentiation in the face of nearly unrestricted gene flow.

Key words: cichlids, ecomorphs, morphology, dentition, environmental gradient, divergent selection, ecological speciation

Introduction

Ecologically relevant structure in the environment is as a major factor contributing to adaptive differentiation and eventually speciation between populations. Recent models on local adaptation along environmental gradients have suggested that spatially localized ecological interactions such as competition for resources in conjunction with environmental gradients can facilitate divergence of populations (Case & Taper, 2000); (Doebeli & Dieckmann, 2003). In these models competition along an environmental gradient generates negative frequency-dependent selection. They suggest that ecological contact may, in fact, be the driving force for divergence and that environmental gradients of intermediate steepness that allow interactions between individuals of the diverging populations will more easily lead to reproductive isolation than environmental gradients alone.

The Caribbean lizard *Anolis roquet* (Ogden & Thorpe, 2002) and the little Greenbull *Andropadus virens* in Cameroonian-forest ecotones (Smith et al., 1997; Smith et al., 2005) are prominent examples of morphological and genetic divergence along environmental gradients. Studies like these also support the view that natural selection caused by changes in ecology due to habitat range expansion can play an important role in adaptive divergence.

Lake Victoria cichlids are another classical example of fast evolution, with 500 endemic species emerging from just a few ancestors (Turner et al., 2001; Seehausen et al., 2003) in the last 15 000 years (Johnson et al., 1996; Stager & Johnson, 2008). In particular the rock-dwelling cichlids of this lake are known for their high diversity in morphology, colouration and behaviour (Kocher, 2004). These fish live on isolated rocky patches along the coast and around offshore islands, which vary a lot in important environmental variables such as water clarity and light penetration, rocky substrate structure, and inclination of the shore slope. Environmental gradients of different steepness generated by variation in water clarity and shore slope can affect the completeness of speciation and places with clear water have higher species diversity (Seehausen et al., 1997; Seehausen et al., 2008 (chapter 4)). Because the majority of the island assemblages contain species with very different colouration but apparently little or no morphological differentiation, most studies have focused on the origin and maintenance of colour polymorphisms (Seehausen & Van Alphen, 1999; Seehausen et al., 1999 ; Seehausen & Schluter, 2004). Due to the rarity of intra-population morphological polymorphisms (Seehausen, 1996; Seehausen et al., 1998), less research had until now been devoted to the possibility of trophic polymorphisms being involved in speciation. In this study we investigated three trophic polymorphisms, which were described thirteen years ago (Seehausen, 1996), from three locations with different environmental structure, and putatively variable degrees of progress towards speciation.

These polymorphisms were reported within populations and between sympatric incipient species of the genus *Neochromis*, rock-dwelling, algae scraper cichlids with a wide distribution around rocky islands all over Lake Victoria. The three trophic polymorphisms, composed of two morphs each, appear to follow a repetitive pattern: one morph is a typical algae scraper with bicuspid teeth in the outer tooth row, deeper body and large head, whereas the second morph is more slender, with unicuspid teeth in the outer tooth row. One of the polymorphisms was reported at Makobe Island and is composed of the species *N. omnicaeruleus* a typical algae scraper and *N.* "unicuspid

scraper" reported to have unicuspid dentition (Seehausen, 1996). The other polymorphism, reported at Igombe island, is very similar to the previous, except that at this island *N. greenwoodi* replaces *N. omnicaeruleus* (Seehausen, 1996). The third polymorphism was reported within the species *Neochromis* "Bihiru scraper" from Bihiru Island. *N.* "Bihiru scraper" was considered a polymorphic species, which appeared to have two discrete morphs that differ in tooth morphology, body and jaw shape and were called *N.* "Bihiru scraper" unicuspid type and *N.* "Bihiru scraper" bicuspid type (Seehausen, 1996). Seehausen (1996) considered the Bihiru Island case more likely an intraspecific polymorphism, but the other two interspecific polymorphisms because in the latter two, but not in the first, morphology was associated with some differences in male colouration. The shores of the three islands differ in water clarity, steepness of the slope, substrate and depth of the rock-sand interface.

The differences in environmental structure around these islands and the potential existence of ecological morphs in each of them allowed us to test the effects of environmental structure on eco-morphological and genetic variation and differentiation. We set out to test the hypothesis that divergent ecological selection, mediated by gradients along water depth, can maintain phenotypic polymorphisms. The three different islands, Makobe, Igombe and Bihiru, are separated by large habitat discontinuities and thereby can reasonably be considered as independent replicates. In particular, we ask the following questions: 1) can we demonstrate eco-morphological and genetic variation and differentiation within island populations? 2) can environmental factors such as water depth, steepness of the shore slope and environmental structure explain the extent of sympatric differentiation?

In order to answer these questions we examined the environments, and analysed how variation in morphology, dentition and nine microsatellite loci was partitioned between and within the three islands and the two morphs described from each island. We then tested for correlations between divergence in dentition, external morphology and neutral loci between morphs within and between islands. Finally we investigated associations between water depth and variation in morphology, dentition and neutral loci. Using the species *N.* "Bihiru scraper" we also performed a common garden experiment to test if differences in morphology and dentition had any genetic basis.

Materials and Methods

Study species: geographical distribution and description

Neochromis omnicaeruleus Seehausen & Bouton, 1998 and *Neochromis greenwoodi* Seehausen & Bouton, 1998 are two described species of rock-dwelling algae scrapers widely distributed in Lake Victoria (Fig. 1). They are largely allopatric sibling species with complementary and adjacent distributions: *N. omnicaeruleus* inhabits almost exclusively offshore islands in the open lake and western Speke Gulf while *N. greenwoodi* is most common along much of the mainland and at the inshore islands of the Mwanza Gulf (Seehausen, 1996). There are very few localities where the two species occur together: at Juma and Mafwinki Islands (Sengerema, Tanzania) they behave like distinct species, but at Ndurwa Point (Speke Gulf, Tanzania) hybrids of the two species appear to exist in a predominantly *N. omnicaeruleus* -like population. *N.* "unicuspid scraper" is known from four locations and coexists with *N. greenwoodi* at two of these locations (Igombe Island, Bwiro Point) and with *N. omnicaerueleus* at the other two

(Makobe Island, Ndurwa point). *N.* "Bihiru scraper" is known only from Bihiru Island. *N. greenwoodi* also occurs at this island, however, it is rare and phenotypically very distinct from *N.* "Bihiru scraper".

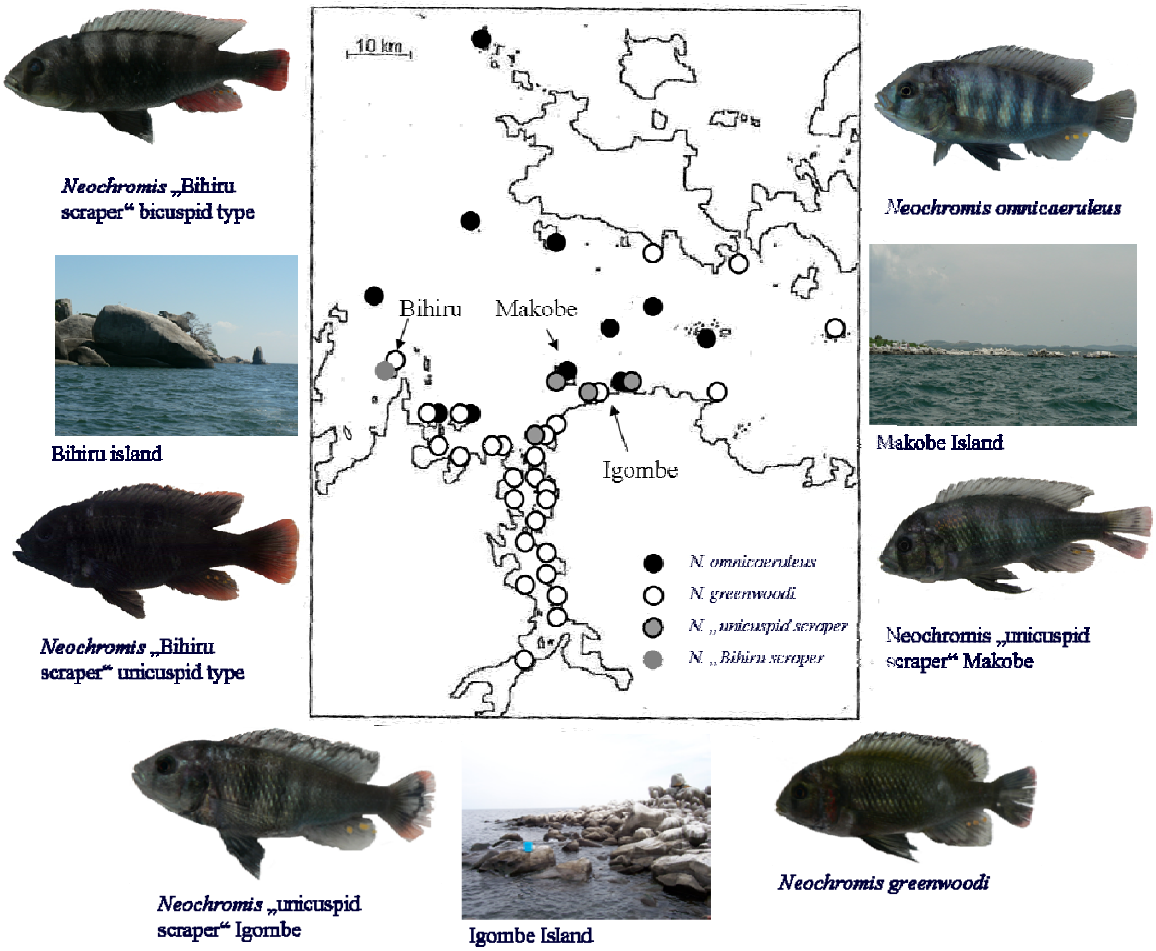

Figure 1. Map of the southern part of Lake Victoria showing known distributions of the species (adapted from Seehausen, 1996). Arrows indicate the three sampling locations: Makobe, Igombe and Bihiru islands. Around the map are pictures of the shores of each island and of the phenotypes studied at each location.

Neochromis omnicaeruleus is a typical algae scraper with a relatively large head, deep body and a specialized dentition of densely spaced bicuspid teeth in the outer tooth rows and many inner tooth rows (Seehausen et al., 1998). They have characteristic bright metallic blue or yellow nuptial colouration while females are yellow-brown; both sexes have dark vertical bars. Most populations are polymorphic for x-linked colouration, and segregate for one or two colour morphs additionally to those above: the orange blotch (OB) and the white-blotched (WB) morph. While females of the blotch morphs are abundant, blotched males are extremely rare (<1%) (Seehausen et al., 1999; Maan et al., 2008) and we had none included in this analysis.

N. greenwoodi are ecologically similar to *N. omnicaeruleus* but have a typical breeding colouration that is uniformly black with a narrow red caudal fin edge while

females are typically dark brown. They also share a similar sex-linked colour polymorphism with *N. omnicaeuruleus,* but only the OB morph is found in some populations of this species, including that from Igombe Island. The blotch polymorphism present in these two species was analysed separately in a different context (chapter 2).

N. "unicuspid scraper" is an undescribed species (Seehausen, 1996) whose name derives from the fact that adult fish have widely spaced unicuspid teeth, an uncommon feature in *Neochromis*, that is otherwise known within this genus only from the unicuspid morph of *N.* "Bihiru scraper". Its morphology, breeding colouration and micro-distribution is also different from that of the sympatric populations of *N. omnicaeruleus* and *N. greenwoodi*: *N.* "unicuspid scraper" has a relatively slender body and inhabits slightly deeper sections of the shores. At Igombe Island males are either light blue resembling *N. omnicaeruleus* from Makobe, or black resembling the sympatric *N. greenwoodi*. At Makobe island, males are greyblue, and much less vividly coloured than the sympatric *N. omnicaeruleus* (Seehausen, 1996). *N.* "unicuspid scraper" shares the OB morph with the other two species, though the OB morph is only known from the locations where the species coexists with *N. greenwoodi* (Seehausen & van Alphen, 1999). Seehausen (1996) tentatively considered *N.* "unicuspid scraper" as specifically differentiated from their obviously closely related sympatric congeners, but expressed doubts.

N. "Bihiru scraper" is another undescribed species, only known from Bihiru Island (Seehausen 1996). It inhabits a range of micro-habitats and the full depth range from less than two meters to at least nine meters of depth. The species appears to exhibit a fairly discrete polymorphism, with two trophic morphs that vary in tooth morphology, body and jaw shape (Seehausen, 1996). One type, the "bicuspid type" tends to have a shorter lower jaw, more inner tooth rows and bicuspid teeth present in the outer tooth rows in adults, similarly to other typical algae scrapers such as *N. omnicaeruleus* and *N. greenwoodi*. The second type, the "unicuspid type", resembles *N.* "unicuspid scraper" and has been described as having a longer lower jaw, fewer inner teeth rows and unicuspid teeth in the outer tooth rows in adult fish. However none of these traits had been quantified and the existence of two discrete morphs as opposed to a continuously variable population remained to be confirmed. Seehausen (1996) considered the morphs conspecific because no consistent differences was observed in male coloration, and because no morphological polymorphism could be detected in females.

In figures and tables, *N. omnicaeruleus*, *N. greenwoodi*, *N.* "unicuspid scraper" from Makobe and Igombe and N. "Bihiru scraper" bicuspid morph and unicuspid morph will be referred to as "N.o.", "N.g.", "N.us Ma", "N.us Ig", "N.Bs bic" and "N.Bs uni" respectively. In the remainder of the paper we refer to the two morphs or species within each island simply as "morphs".

Sampling

With the exception of blotched morphs, rock-dwelling cichlid females tend to be cryptically coloured making identification to species level among very closely related species very difficult. Therefore only males were used for this study. We targeted adults because eco-morphological traits, including tooth shape, are subject to ontogenetic change. Every male was photographed upon capture in a custom made perspex photographic cuvette. Prior assignment of males to morph/species was done by visual

inspection of individual photographs, using a combination of male nuptial colouration, head and body profile. No morphological measurements or analysis of the dentition, nor genetics were done prior to the assignment of individuals to a morph/species.

By angling and gill-netting we collected a total of 174 males at the three islands: 45 *N. omnicaeruleus* and 20 *N.* "unicuspid scraper" at Makobe Island, 30 *N. greenwoodi* and 17 *N.* "unicupid scraper" at Igombe Island, and 62 *N.* "Bihiru scraper", 29 assigned to the bicuspid morph and 33 to the unicuspid morph. Immediately after capture males were photographed in a standard photo cuvette. After humanely killing the fish by overdosing an anaesthetic we took fin clips for genetic analysis and the specimens were fixed in 4% formaldehyde solution. After transporting to the laboratory, we transferred the specimens in three steps to ethanol concentrations of 30%, 50% and finally 70% for storage.

In order to correlate eco-morphological and genetic variation with depth of capture, we recorded for every individual the water depth at which they were collected. We measured the length of the angling line or recorded the depth at which nets were set. At the same time, we recorded the type of substrate (rock or sand) to estimate the depth of the rock-sand interface. We also recorded water transparency (secchi depth) for each sampling day and calculated the inclination of the slope.

Phenotypic analysis

To analyse morphology we took thirteen standard morphometric distances using a digital calliper (Barel et al., 1977) (see names of distances in Table 3). Variables were log-transformed and regressed against the log-transformed standard length to adjust each morphometric measurement for size heterogeneity among individuals (Fleming & Gross, 1994). Standardized residuals from these regressions were used as variables for further analyses.

Haplochromine oral tooth shape is subject to ontogenetic changes. Subadult Lake Victoria haplochromines generally have bicuspid teeth. During further ontogeny, these can change to weakly bicuspid and unicuspid through gradual reduction in size of the minor cusp. The extent to which this happens varies between species. In most populations of *Neochromis* teeth stay bicuspid, however not so in the unicuspid morphs that we studied here. The upper jaw teeth are arranged in multiple rows, with algae scrapers having typically a very large number of rows. For each individual we counted the number of uni-, bi- and tricuspid teeth in the outermost tooth row of the upper jaw and the number of tooth rows in the upper jaw. We calculated the percentage of unicuspid teeth by dividing the number of unicuspid teeth by the total number of teeth in the outer most row of the upper jaw.

Statistical analysis of phenotypic variation and differentiation

Morphological differences between individuals both within and among morphs and islands were visualised by Principal Component Analyses on residuals of the log transformed morphological distances. Principal components were extracted from covariance matrices.

We used two complementary methods to test for the signature of disruptive selection acting on eco-morphology. First, as visual representations we plotted frequency histograms of the first three Principal Components, percentage of unicuspid teeth and

number of tooth rows for each island separately. Then, using discmixtureprogs v0.4 (Brewer, 2003) we estimated the Akaike Information Criteria (AIC) corrected for sample size (AIC_c; Burnham & Anderson, 2002) and tested statistically for each island separately, if our sample was better represented by a single normal distribution or by a mixture of two normal distributions. We then estimated AIC_c differences (ΔAIC_c) as AIC_c for the single normal distribution minus AIC_c for the fitted mixture of two distributions. We followed established guidelines (Burnham & Anderson, 2002) and used the interpretation of (Hendry et al., 2006): $\Delta AIC_c < -8$ is considered as strong support for a single normal distribution, $-8 < \Delta AIC_c < -5$ as moderate support for a single normal distribution, $-5 < \Delta AIC_c < 5$ as roughly equivalent support for a single normal distribution or a mixture of two normal distributions, $5 < \Delta AIC_c < 8$ as moderate support for a mixture of two normal distributions and $\Delta AIC_c > 8$ is considered as strong support for a mixture of two normal distributions.

To look at how differences in morphology and dentition are partitioned between the three islands and the two morphs at each island, we performed nested univariate analyses of variance (ANOVAs) on the number of tooth rows, percentage of unicuspid teeth, standard length and on standardized residuals of 12 morphological distances (two factors: morphs (2 morphs) nested in islands (3 islands)). For traits showing significant differences between morphs we then performed Tukey *post-hoc* tests to further investigate which groups were significantly different. We performed the same type of analysis on the first three morphological Principal Components.

DNA extraction and microsatellite amplification

We extracted DNA from fin tissue of 174 individuals using a QIAGEN® (Basel, Switzerland) Biosprint^{TM} 96 extraction robot with a corresponding standard digestion protocol. Nine microsatellite loci developed for lake Victoria cichlids (Ppun 5, Ppun7, Ppun17, Ppun21 and Ppun32 (Taylor et al., 2002) and for other African cichlids (OSU20d, OSU19T, OSU16d (Wu et al., 1999) and TmoM5 (Zardoya et al., 1996)), were chosen for population analysis. We used two different sets of markers for multiplexing, avoiding overlapping allele ranges. The first set included the five Ppun loci and the second set the other four loci. We used the QIAGEN Multiplex PCR kit for PCR amplification according to the manufacturer's protocol and PCR program. A quantity of 1 µl of a 1:2 dilution of the PCR was added to a volume of sample loading solution (deionised formamide) and 400 bp DNA size standard for analysis of fragments up to 400 nucleotides (Beckman Coulter). Denaturated fragments were resolved on an automated DNA sequencer (Beckman Coulter, CEQ 8000). We then determined genotypes manually.

Analysis of neutral genetic variation and divergence

We used Convert (Glaubitz, 2004) to create input files for other programs and to create allele frequency tables. Genotypes were checked for scoring errors that might be attributable to stutter products, large allele dropout, or the presence of null alleles, using MICRO-CHECKER v.2.2 (Van Oosterhout et al., 2004). We used ARLEQUIN 3.11 (Excoffier et al., 2005) to test each locus in each population for departure from Hardy-Weinberg equilibrium (HWE), to calculate observed (H_O), expected (H_E) heterozygosities and to perform tests for genotypic linkage disequilibrium (LD). Estimates of the

multilocus inbreeding coefficient (F_{IS}), H_O and H_E for each group using Genetix 4.05 (Belkhir et al., 1996). Statistical significance in the above tests was adjusted for multiple comparisons using sequential Bonferroni adjustments (Rice, 1989). To explore patterns of genetic mixing between morphs from the same island, we used paired t-tests on the allelic richness and the difference between observed and expected heterozygozities at the nine loci.

To visualize the partitioning of molecular variance among individuals within and between islands we performed a Factorial Correspondence Analysis over individuals in Genetix 4.05 (Belkhir et al., 1996).

We performed model based assignment tests on genotypes from all islands combined, as implemented in the program STRUCTURE 2.2 (Pritchard et al., 2000). MCMC simulations were run with 500000 replicates and a burn-in of 50000 replicates for K (number of populations) = 1 to 6 and applying the admixture model, in which individuals may share portions of the genome assigned to more than one population as a result of mixed ancestry (Pritchard et al., 2000).

We used F_{ST} values as measures of neutral genetic differentiation between morphs from the same and from different islands and also to estimate divergence between islands. Single locus and multilocus F_{ST} values and corresponding 95 % confidence intervals were estimated with 10'000 bootstrap replicates (Weir & Cockerham, 1984) using ARLEQUIN 3.11 (Excoffier et al., 2005). We also compared single locus F_{ST} distributions using paired t-tests.

As a further test of differentiation we used the allele frequency-based program DOH (http://www.biology.ualberta.ca/jbrzusto/Doh.php) to calculate genetic assignment for each phenotypically assigned individual (Paetkau et al., 1995). We performed an overall assignment test with all morphs from all islands and three separate assignment tests for the assignment of individuals within each island. Significance of deviation from random assignments was assessed with chi-square tests.

Correlation between phenotypic and genetic divergence

We tested if gene flow might be constraining eco-morphological differentiation by testing for correlations between phenotypic and genetic distances. We computed multidimensional Euclidean distances between means of groups (morphs) from the same island and populations from different islands for morphological distances and dental characters separately. Genetic distances between morphs and populations were computed as Nei's standard genetic distance (D_S) using DOH. Relationships between phenotypic distances and between phenotypic and genetic distances were analysed separately for sympatric morphs and allopatric populations and for both combined with generalized linear models (GLM) in SPSS 14.0 (SPSS Inc.).

The distribution of phenotypic and genetic variation over water depth

We tested for an association of male phenotype with water depth at capture point within each island, using Chi-square tests. Additionally, we calculated Spearman-rank correlations between capture depth and number of tooth rows, percentage of unicuspid teeth, the first three Principal Components of morphology.

To test for effects of water depth on neutral genetic variation we used an individual-based approach. Matrices of capture depth differences and genetic distances (â)

(Rousset, 2000) between individuals were computed for each island separately using the program SPAGeDI (Hardy & Vekemans, 2002). We then tested the significance of relationships between the two matrices with a Mantel test based on 1,000 permutations as implemented in ARLEQUIN (Excoffier et al., 2005).

Common garden experiment and analysis of laboratory-bred fish

We conducted a common garden experiment in our laboratory aquarium system on *Neochromis* "Bihiru scraper". We used this data to test if the phenotypic differences between morphs observed in the wild were heritable. Given the close genetic relatedness of the four species used in this study, and the close resemblance of the phenotypic polymorphisms, we believe that results will be representative of the other species too.

Breeders were 15 males and 9 females collected at Bihiru Island in November 2005. Individuals were kept in separate tanks at 24 ± 1°C and 12L: 12D daylight cycle. Tanks were part of a 30,000 litres recirculation system with mechanical, biological, and UV filter. All fish were fed twice a day with dry commercial cichlid pellets in the morning and a fresh blend of mashed shrimps and peas, supplemented with vitamins and blue-green algae in the afternoon. Males and females were tagged individually with passive integrated transponders. The breeding was done in three experimental tanks, each containing one male and three females. As soon as one female was brooding the male was removed from the tank and replaced by another male. *Neochromis* spp. are female mouth-brooders and we removed brooding females with their incubated clutches from the breeding tank. Once a female had raised a clutch to free-swimming stage (ca. 2 weeks), she was put back into an experimental tank with a different male. It was not possible to distinguish the females of the two putative morphs (Seehausen, 1996). In order to control for variation in the genotype of the dam, each female was crossed with two different males. Families of dams that bred only with one sire were excluded from further analysis. During the first two months of 2006, six F1 half-sib families from three different dams and six different sires were produced. Each family was raised in a tank on its own. We analysed at total of 40 males from the 6 families, with between 3 and 19 males >12 months.

Parental individuals were not available for phenotypic analysis because, in order to complete the breeding scheme, we failed to preserve them in time before they showed strong phenotypic effects of aging. To test for family effects on the phenotypic variation observed among the experimental families, nested analysis of variance (ANOVA) was carried out (two factors: full-sib family (six families) nested in half-sib family (three families)) for the number of tooth rows, the percentage of unicuspid teeth in the outer tooth row of the upper jaw, standard length and standardized residuals of morphological distances. Additionally we performed the same type of analysis on the first three Principal Components of morphometric distances. For variables showing a significant difference between half-sib families or full-sib families within half-sib families we performed Tukey *post-hoc* tests to investigate which groups were significantly different.

Results

Environmental data

The three islands, Makobe, Igombe and Bihiru Islands have different water clarity, shore slopes and depths of the rock-sand interface (Table 1). Makobe Island has the most gently sloping shore and the highest water transparency, which together create the shallowest environmental gradient amongst the three islands. The shore of Igombe Island is in many ways similar to Makobe but has a slightly steeper environmental gradient. This island has a steeper shore slope, the rock-sand interface is at greater depth and probably due to its proximity to the mainland, water transparency is lower. On the other hand, the shore of Bihiru Island is extremely steep which, despite the clarity of the water, creates a steep environmental gradient.

Table 1 - Geographical and habitat information and number of individuals collected (N) per each island.

island	latitude	longitude	slope	rock-sand interface (metres)	secchi depth	N
Igombe	02°23' S	32°37'E	17.9°	4-5	216 cm	47
Makobe	02°22' S	32°39'E	6.6°	5-7	287 cm	65
Bihiru	02°22' S	32°55'E	36.8°	4-9	248 cm	62

Phenotypic Variation

Percentage of unicuspid teeth had clear bimodal distribution in populations from Makobe and Bihiru islands. Number of tooth rows was bimodally distributed at Makobe, but not at the other islands. None of the Principal Components had a bimodal distribution (Fig. 2).

Percentage of unicuspid teeth and number of tooth rows varied greatly within and among islands. Individuals ranged from having no unicuspid teeth to having almost exclusively unicuspid teeth. At Igombe Island no individuals were found having only unicuspid teeth and at Bihiru Island all individuals had some unicuspid teeth. Percentage of unicuspid teeth showed a strongly bimodal frequency distribution at Makobe and Bihiru Islands. Strong bimodality in these samples was evident from strong discontinuities in the frequency histograms. ΔAICc values of 123.14, 65.72 and 68.61 for Makobe, Igombe and Bihiru populations respectively, were considered strong support for a mixture of two normal distributions (ΔAICc>8) at each island.

Tooth rows ranged from two to nine, but individuals with the extreme low and high numbers of rows were found only at Makobe Island. At Igombe Island tooth rows ranged between three and seven and at Bihiru only between three and six. The distribution of the number of tooth rows showed discontinuities and deviations from normality in the population of Makobe Island, but not in the other two populations. While there was moderate support for a mixture of two normal distributions at Makobe ($5<\Delta$AICc<8), there was no support for a mixture of two normal distribution as opposed to one single normal distribution at either Igombe or Bihiru island ($-5<\Delta$AICc<5).

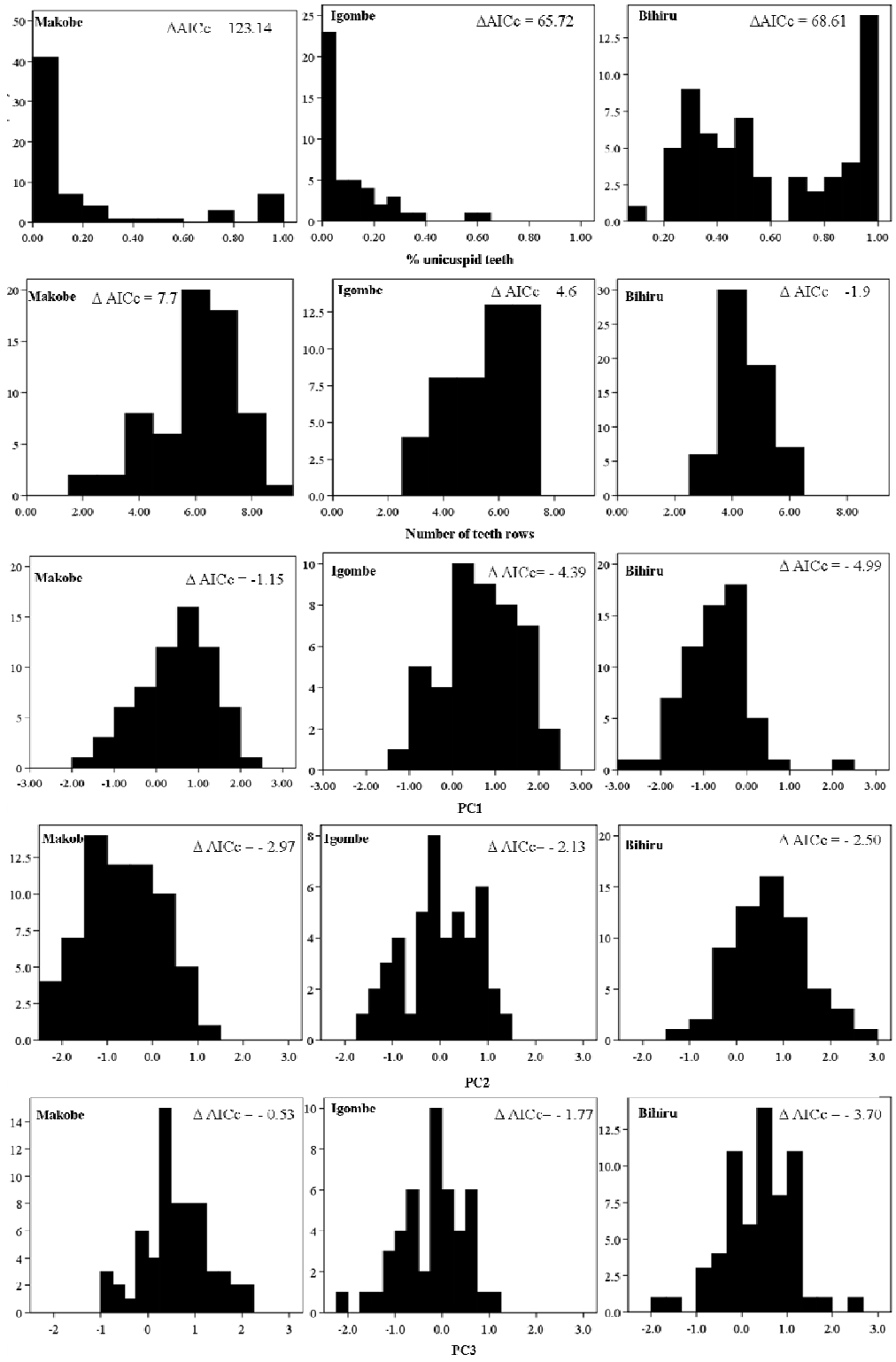

Figure 2. Frequency distribution and ΔAIC_c values for (from top to bottom) percentage of unicuspid teeth, number of teeth rows and morphometrics Principal Components one, two and three for populations from the three islands separately.

Principal component analysis identified three major axes of morphological variation (axes that explain more than 10% of the variation). PC1 explained 28.23% of the variation and can be interpreted as a measure of head width, body depth, eye length and eye width (Table 2). PC2 explained 18.57% of the variation and was loaded heavily by snout length. PC3, which explained 13.96% of the variation, was loaded mostly by cheek depth. ΔAICc values for PC1, PC2 and PC3 were all between -5 and 5 for all populations, indicating equivalent support for a mixture of two normal distributions or one single normal distribution.

Table 2- Loadings of morphological variables on the first three principal component axes and the percent variance explained by each axis.

Trait	PC1	PC2	PC3
Head Length	0.174	0.735	-0.42
Head Width	**0.751**	0.101	0.2
Body Depth	**0.78**	-0.191	-0.138
Lower Jaw Length	-0.139	0.627	-0.051
Lower Jaw Width	0.477	0.165	0.468
Snout Length	-0.031	**0.846**	0.198
Snout Width	0.612	0.063	0.392
Cheek Depth	0.112	0.342	**-0.674**
Pre-Orbital Depth	0.129	0.521	0.489
Inter-Orbital Width	0.618	-0.023	0.362
Eye Length	**0.768**	-0.108	-0.318
Eye Width	**0.785**	-0.037	-0.282
Percentage variance	28.233	18.571	13.958

Phenotypic divergence

The nested ANOVAs revealed significant differences between number of tooth rows and percentage of unicuspid teeth between morphs from different islands (Table 3; Fig 4a). Within islands number of tooth rows and percentage of unicuspid teeth were significantly different between *N. omnicaeruleus* and *N.* "unicuspid scraper" (Makobe island) and between *N. greenwoodi* and *N.* "unicuspid scraper" (Igombe island) (Fig.3 a, b).

N. "unicuspid scraper" generally had a smaller number of tooth rows and a higher percentage of unicuspid teeth than *N. omnicaeruleus* and *N. greenwoodi* (Fig. 4 b, c). At Bihiru island, the island with the steepest gradient on the other hand, the two putative morphs of *N.* "Bihiru scraper" were significantly differentiated only in the percentage of unicuspid teeth (Fig. 3b, Fig. 4d).

Table 3 - Results of nested ANOVAs for males in number of tooth rows, percentage of unicuspid teeth, 13 standard morphological distances and the first three PCs based on those morphological distances for morphs nested within islands. Significant P-values are in bold.

	morphs (island)			island		
character	df	F	Sig.	df	F	Sig.
tooth rows	3	32.954	**0.000**	2	18.398	**0.000**
% unicuspid teeth	3	26.017	**0.000**	2	54.720	**0.000**
Standard Length	3	0.824	0.482	2	5.826	**0.004**
Head Length	3	5.451	**0.001**	2	26.427	**0.000**
Head Width	3	13.395	**0.000**	2	7.893	**0.001**
Body Depth	3	29.338	**0.000**	2	42.782	**0.000**
Lower Jaw Length	3	15.086	**0.000**	2	16.092	**0.000**
Lower Jaw Width	3	6.738	**0.000**	2	1.425	0.244
Snout Length	3	2.302	0.079	2	26.299	**0.000**
Snout Width	3	6.332	**0.000**	2	6.094	**0.003**
Cheek Depth	3	2.450	0.065	2	14.131	**0.000**
Pre-Orbital Depth	3	0.270	0.847	2	2.569	0.080
Inter-Orbital Width	3	7.420	**0.000**	2	15.202	**0.000**
Eye Length	3	13.202	**0.000**	2	28.998	**0.000**
Eye Width	3	6.165	**0.001**	2	37.926	**0.000**
PC1	3	28.692	**0.000**	2	41.484	**0.000**
PC2	3	1.741	0.161	2	39.604	**0.000**
PC3	3	1.439	0.233	2	11.405	**0.000**

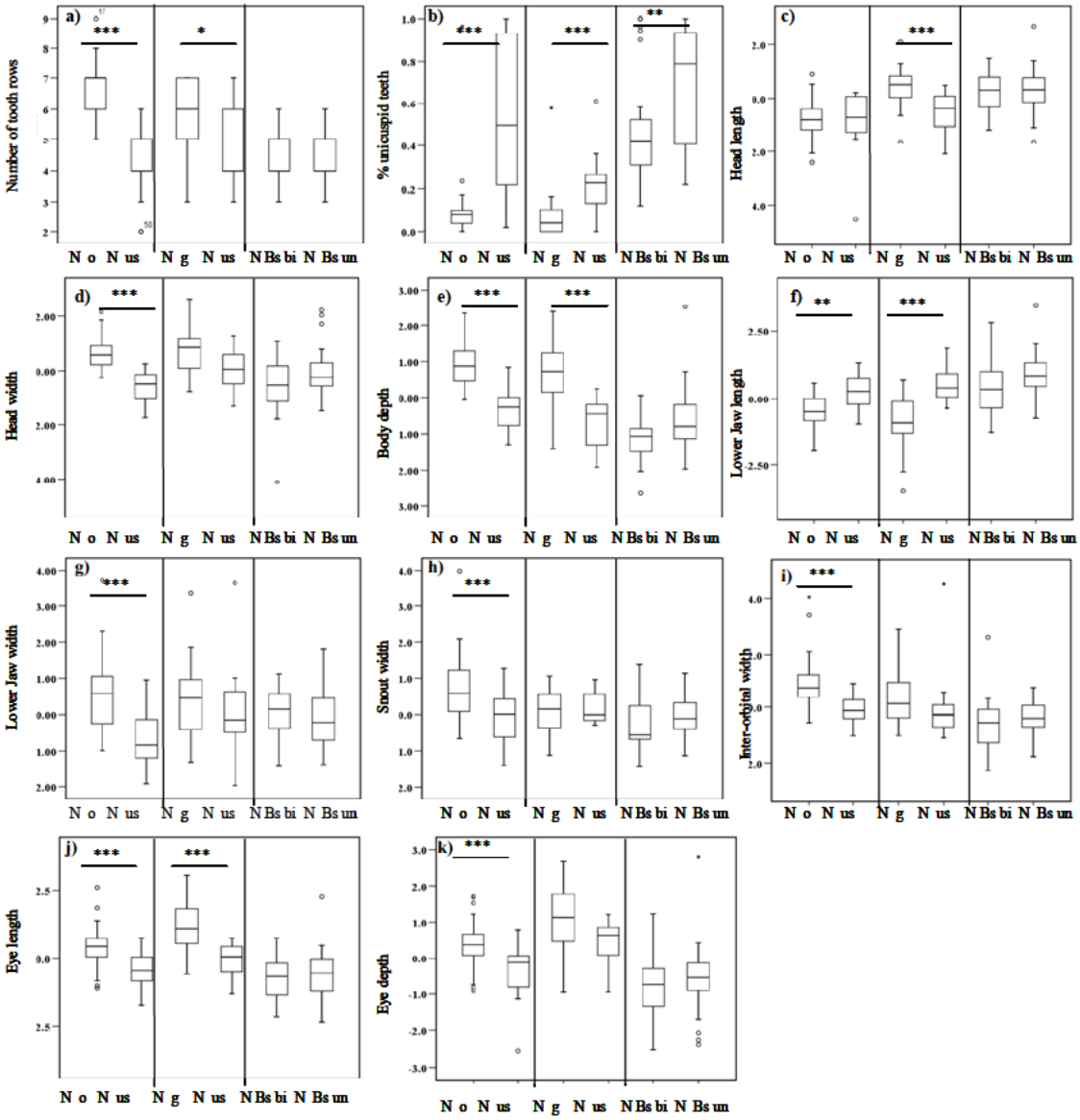

Figure 3. Boxplots of dentition and of morphological distances with significant differences between morphs. Significance of the differences indicated by * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

Eleven out of the 13 morphological distances differed significantly between islands (Table 3). Aside from four morphological distances, namely standard length, snout length, cheek depth and pre-orbital length, morphs nested within islands explained significant portions of the variance in all traits too. However there were clear differences in the number of morphological distances that were diverged between morphs from the same island (Fig. 3). At Makobe Island (shallowest gradient slope) 8 of 13 morphological distances were significantly different between *N. omnicaeruleus* and *N.* "unicuspid scraper". At Igombe Island 4 of 13 morphological distances were significantly different between *N. greenwoodi* and *N.* "unicuspid scraper". At Bihiru Island (steepest gradient slope), the two putative morphs of *N.* "Bihiru scraper" were not significantly different in any of the 13 morphological distances.

When variation in individual morphological distances was reduced to variation in Principal Components, all three PCs were significantly different between islands (Table 3; Fig. 4e), but only PC1 was significantly different between morphs within two islands: Makobe ($p < 0.0001$) and Igombe ($p < 0.0001$) (Figs 4 f, g, h). Here again, the magnitude of phenotypic differentiation followed the variation in gradient slope steepness.

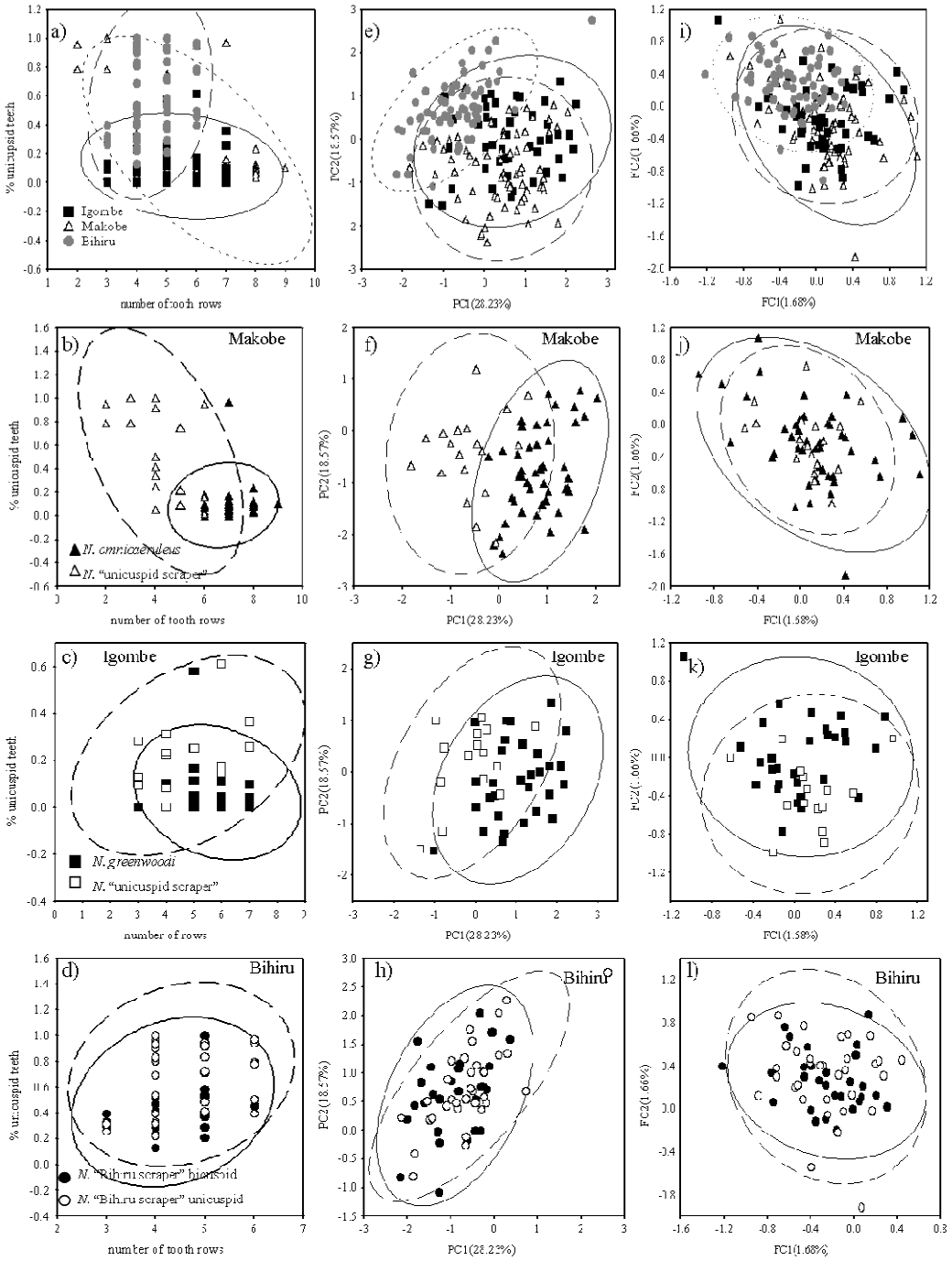

Figure 4. Panels on the left (a, b, c, d): bivariate plots of number of tooth rows and percentage of unicuspid teeth. Panels in the middle (e, f, g, h): bivarite plots of morphometrics Principal Components one and two. Panels on the right (i, j, k, l): bivariate plots of Factorial Component one and two.

Neutral genetic variation within morphs

Most microsatellite markers exhibited high polymorphism (Appendix 1). Of the 174 individuals, five were not successfully genotyped at several loci, 169 individuals were hence included in the analysis. A total of 287 alleles were found, ranging from 8 to 72 per locus.

Before Bonferroni correction, for the pooled Makobe population (*N. omnicaeruleus* and *N.* "unicuspid scraper" pooled), no deviations from Hardy-Weinberg equilibrium (HWE) were observed but 13 tests out of 36 showed significant linkage disequilibrium (LD) (Appendix 2). After sequential Bonferroni 11 tests of LD remained significant. When the two morphs were analysed separately, before sequential Bonferroni no deviations from HWE were observed but 13 and 5 tests showed significant linkage disequilibrium in *N. omnicaeruleus*, and *N.* "unicuspid scraper" respectively. After Bonferroni correction 9 tests in *N. omnicaeruleus* remained significant.

For the pooled Igombe population (*N. greenwoodi* and *N.* "unicuspid scraper" pooled), two deviations from HWE were observed and 12 tests showed significant LD, before Bonferroni correction (Appendix 2). After the correction only one test of LD remained significant and one locus (OSU19t) still deviated from HWE. The latter may be the result of a technical artefact since this was the only locus for which Micro-Checker detected the possibility for null alleles. When *N. greenwoodi* and *N.* "unicuspid scraper" were analysed separately, before sequential Bonferroni correction, departures from HWE were observed at locus OSU19t and OSU20d in *N. greenwoodi*, LD was observed in 5 tests in *N. greenwoodi,* and in 10 tests in *N.* "unicuspid scraper". No significant LD was observed after sequential Bonferroni correction but OSU19t still deviated from HWE in *N. greenwoodi*.

One locus (OSU19t) deviated from HWE in *N.* "Bihiru scraper" and 3 tests out of 36 showed significant linkage disequilibrium, before sequential Bonferroni correction. No significant LD nor any deviations from HWE were observed after correction (Appendix 2). No deviations from HWE were observed, but nine tests of LD were significant in the bicuspid morph and four in the unicuspid morph, before sequential Bonferroni correction, but no significant LD nor were any deviations from HWE observed after correction.

The paired t-tests indicated that the difference between the observed and the expected heterozygozities was significantly more positive in *N.* "unicuspid scraper" than in *N. greenwoodi* at Igombe ($t = -3.01$, d.f. = 8, P (two-tailed) = 0.017), and significantly more positive in the bicuspid than in the unicuspid morph of *N.* "Bihiru scraper" at Bihiru ($t = 2.54$, d.f. = 8, P (two-tailed) = 0.034). No such trend was observed at Makobe. Allelic richness did not differ between any of the morphs.

Neutral genetic divergence between morphs and islands

A Factorial Correspondence Analysis of microsatellite data revealed substantial overlap in genotypic space of pooled island populations (Fig. 4i) and larger overlap even between morphs within each island (Fig.4 j, k, l). FC1 explained 1.68% of the total variance while FC2 explained 1.66%.

The population assignment test performed with STRUCTURE 2.2 (Pritchard et al., 2000) failed to recognize any genetic structure between the 6 sampled units (morphs * islands) (K=1; estimated -ln probability of data = -8 917.41; p =0.993). Nonetheless, plots

for K=6 showed differences between the assignment of individuals from Bihiru Island and individuals from the other two islands, but no obvious other differences (Fig. 5).

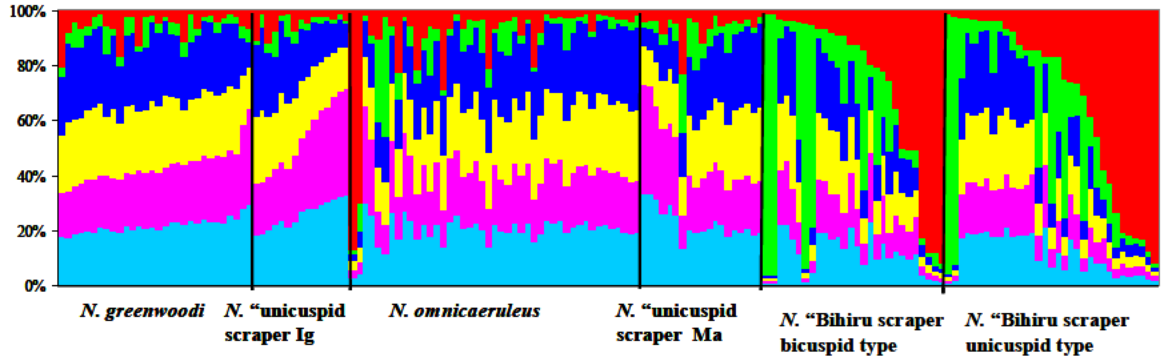

Figure 5. Bayesian assignment probabilities of individuals to lineages, where each vertical line represents an individual and colours indicate the proportion of an individual's genotype assigned to a particular lineage for populations of the three islands divided Within morphs, individuals are ordered by an increasing probability of their genotype being assigned to the unicuspid morph of their island (pink at Igombe, dark blue at Makobe and red at Bihiru).

Between islands neutral genetic divergence was low but always highly significant ($P<0.0001$). The lowest level of between-island neutral genetic divergence was between Bihiru and Makobe islands ($F_{ST} = 0.0128$), followed by divergence between Igombe and Makobe ($F_{ST} = 0.0139$) and finally the highest divergence was between the populations of Bihiru and Igombe ($F_{ST} = 0.021$). This trend coincides with the position of islands in geographical space.

F_{ST} values between morphs from the same island were extremely low and non-significant at Bihiru ($F_{ST} = 0.0027$) and Makobe ($F_{ST} = 0.0032$), but higher ($F_{ST} = 0.0176$) and significant ($p<0.001$) at Igombe (Table 4). Single locus F_{ST} values were generally low and had narrow distributions (Appendix 3). However, the single locus F_{ST} distribution between the two morphs from Igombe was significantly different, with a higher mean, from that between the two morphs from Bihiru ($t = 2.61$, d.f. = 8, P (two-tailed) = 0.031).

Table 4 - Pairwise F-statistics (F_{ST}) between morphs within (bold) and between islands. F_{ST} values and significance (* $p<0.05$, ** $p<0.01$, ***$p<0.001$, n.s. not signicant) in the lower matrix and confidence intervals in the upper matrix.

	No	N.us Ma	Ng	N.us Ig	N.Bs bic	N.Bs uni
No		**0.000-0.010**	0.003-0.037	0.011-0.039	0.007-0.033	0.006-0.028
N.us Ma	**0.00325 n.s.**		0.005-0.032	0.009-0.039	0.008-0.021	0.006-0.019
Ng	0.01553***	0.01749***		**0.007-0.029**	0.011-0.046	0.008-0.037
N.us Ig	0.02452***	0.02082***	**0.0176*** **		0.017-0.051	0.017-0.050
N.Bs bic	0.01676***	0.01357***	0.02641***	0.02851***		**0.000-0.009**
N.Bs uni	0.01358***	0.01189**	0.02132***	0.03021***	**0.00274 n.s.**	

As there was no significant differentiation between the trophic morphs at Bihiru Island, but the population presented a highly bimodal distribution for the percentage of unicuspid teeth, we also calculated the F_{ST} between the group of individuals with more than 50% unicuspid teeth (n = 32) and that with less than 50% unicuspid teeth (n = 29). This F_{ST} value was even lower than the original one (F_{ST} =0.00084) and not significant either (P = 0.360).

The overall assignment test revealed a non-random distribution between morphs (Table 5) ($\chi 2$ = 215.95, d.f. = 25, P <0.001).

Within islands, assignment of genotypes of *N. omnicaeruleus* and *N.* "unicuspid scraper" from Makobe were significantly different from random ($\chi 2$ = 7.928, d.f. = 1, P =0.005), as were those for *N. greenwoodi* and *N.* "unicuspid scraper" from Igombe ($\chi 2$ = 15.40, d.f. = 1, P =0.00008) (Fig. 6). However, this was in both cases mainly due to largely correct assignment of the individuals of the bicuspid morphs. The individuals of *N.* "unicuspid scraper" were in both cases more randomly assigned with only a weak bias towards assignment to *N.* "unicuspid scraper". At Bihiru Island the distribution of genotypes from the two putative morphs was not significantly different from random even though a weak trend for correct assignment was detectable ($\chi 2$ = 0.754, d.f. = 1, P =0.098).

Table 5 - Contingency table with results of the assignment test based on allele frequencies. Represented are numbers of individuals of a morph that were assigned to their own morph or to other morphs (columns). Assignments of genotypes to their own phenotype in bold.

	No	N.us Ma	Ng	N.us Ig	N.Bs bic	N.Bs uni	total
No	**30**	2	4	2	2	5	45
N.us Ma	6	**7**	2	1	1	2	19
Ng	8	2	**17**	1	0	2	30
N.us Ig	3	1	3	**5**	0	3	15
N.Bs bic	4	1	0	0	**13**	10	28
N.Bs uni	6	0	1	0	9	**17**	33

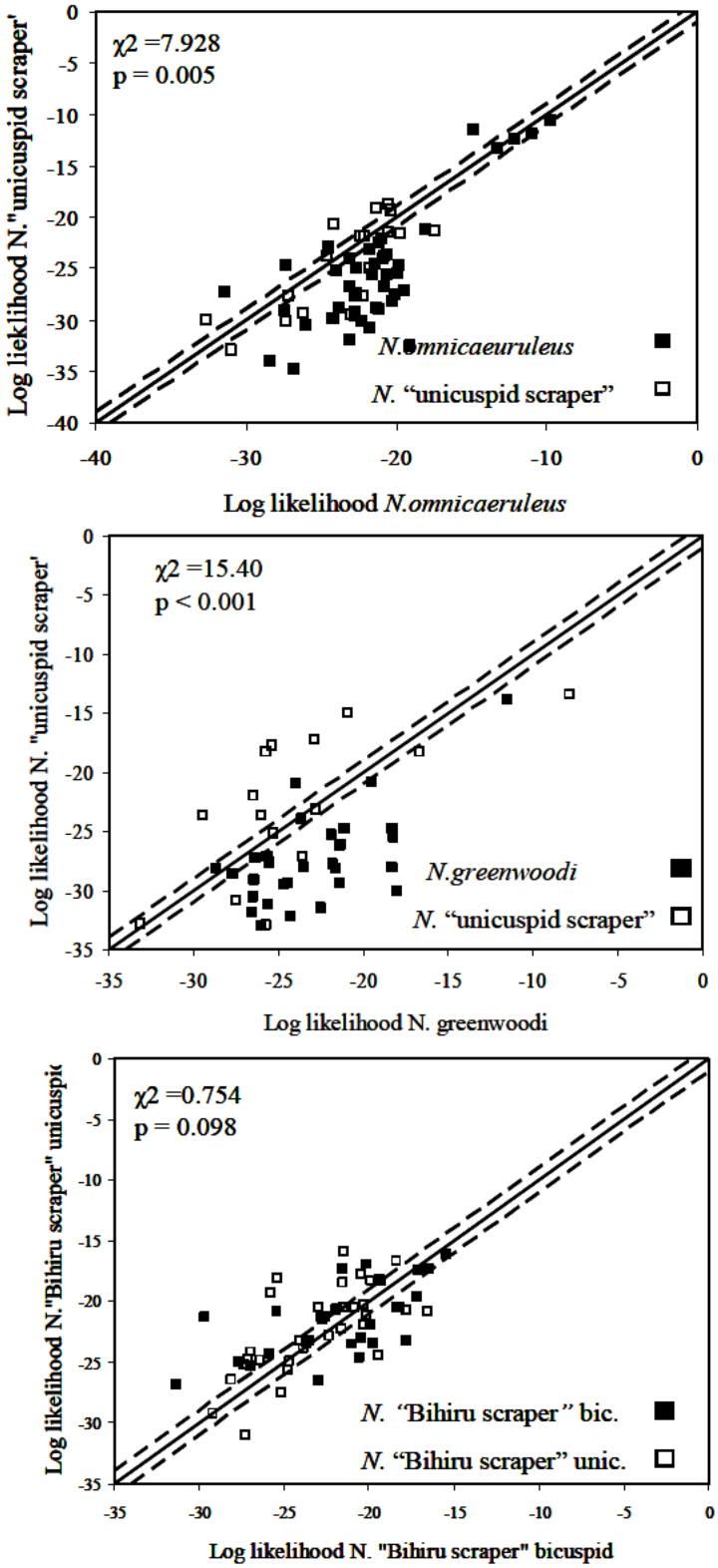

Figure 6. Plots of likelihood of assignment tests showing the likelihoods of individuals belonging to each of the groups compared.

Neutral genetic divergence between populations from different islands ranged from F_{ST} =0.012 (between the unicuspid morph of *N.* "Bihiru scraper" and *N*. "unicuspid scraper" from Makobe) to F_{ST} =0.0302 (between the unicuspid morph of *N.* "Bihiru scraper" and *N*. "unicuspid scraper" Igombe) (Table 4). There were significantly negative differences in single locus F_{ST} distributions between the morphs from Makobe and from Bihiru and single locus F_{ST} distributions between *N.* "unicuspid scraper" from Igombe and the two morphs from Bihiru (Appendix 4).

No correlation between phenotypic and genetic differentiation

They relationship between distances for morphological characters and for dental characters were significant for morphs from different islands ($F_{1,3}$ = 0.251, P = 0.026), but not for morphs from the same island ($F_{1,12}$ = 6.865, P = 0.704) (Fig. 7a).

The relationships between distance in dental traits and genetic distance did not have slopes significantly different from zero either (morphs from the same island: $F_{1,3}$ = 0.057, P = 0.850; from different islands: $F_{1,12}$ = 0.577, P = 0.465) (Fig. 7b). The relationships between morphological and genetic distances were not significant for morphs from the same island ($F_{1,3}$ = 0.099, P = 0.759) or for morphs from different islands ($F_{1,12}$ = 1.414, P = 0.445) (Fig. 7c).

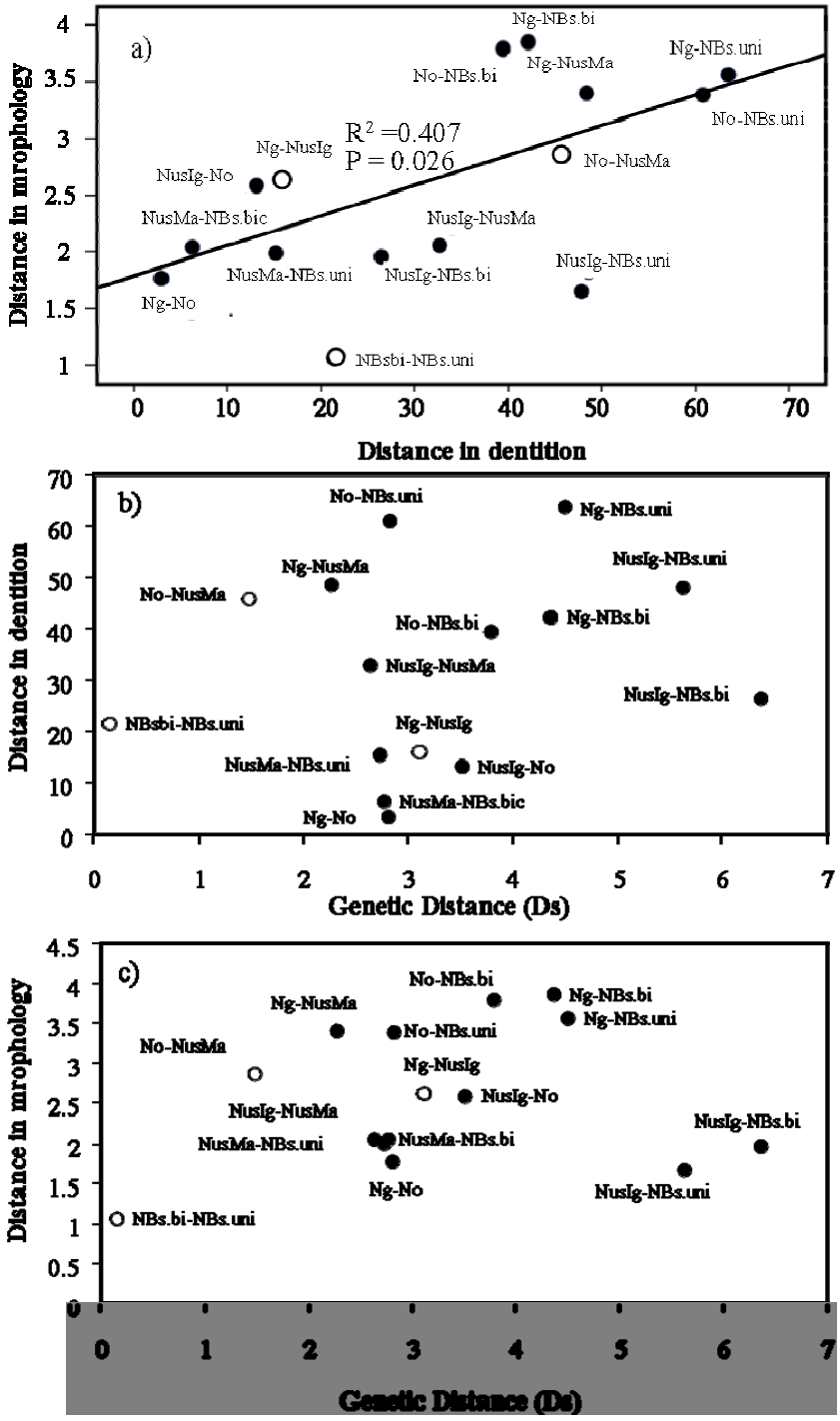

Figure 7. a) normalised Euclidean distances of morphological characters versus genetic distances (D_S) and b) normalised Euclidean distances of dentition characters versus genetic distances for intra- and interspecific morphs.

The distribution of phenotypic and genetic variation over water depth

At the island with the shallowest gradient slope, Makobe Island, phenotypic variation was significantly correlated with water depth: *N. omnicaeruleus* were caught at shallower depth than individuals identified as *N.* "unicuspid scraper", morphological PC1 and the number of tooth rows were significantly negatively correlated with water depth, and the percentage of unicuspid teeth was significantly positively correlated with water depth (Fig. 8). Neither significant associations of eco-morph assignment with water depth, nor significant correlations between phenotype and water depth were found at Igombe or Bihiru islands, but at Igombe a weak negative trend existed in the relationship between number of tooth rows and depth, and at Bihiru an almost significant trend for a positive association of eco-morph with water depth (unicuspid morph occurring deeper). All significant relationships and all trends were in the direction predicted by the resource distribution, i.e. algae scraping morphologies (many tooth rows, few unicuspid teeth, high PC1 score) dominating in shallower water. After aligning their signs by the direction of the prediction, the mean of all correlation coefficients was significantly different from zero (one-sample t-test, $t=3.15$, 8df, $p=0.014$). To test if genetic / developmental non-independence of the traits analysed might be behind the results obtained we performed Correlations performed between the three traits (number of rows, percentage of unicuspid teeth and PC1 scores), for the Bihiru population, were not significant ($P > 0.05$), indicating that these are genetically / developmental independent traits and, therefore the three trends measured for every island are statistically independent tests of divergent adaptation.

At Igombe Island, water depth at capture had a significant effect on genetic structure: pairwise genetic distances among individuals were significantly positively correlated with the depth differential between them ($r = 0.1744$, $p = 0.015$), although depth differential only explained 3% of the variance in genetic distance. At Makobe the same effect was close to significant ($r = 0.065$, $p = 0.053$), whereas no trend existed at Bihiru ($r = 0.0012$, $p = 0.491$). At both these islands, depth differential explained less than 1% of the variance in genetic distance.

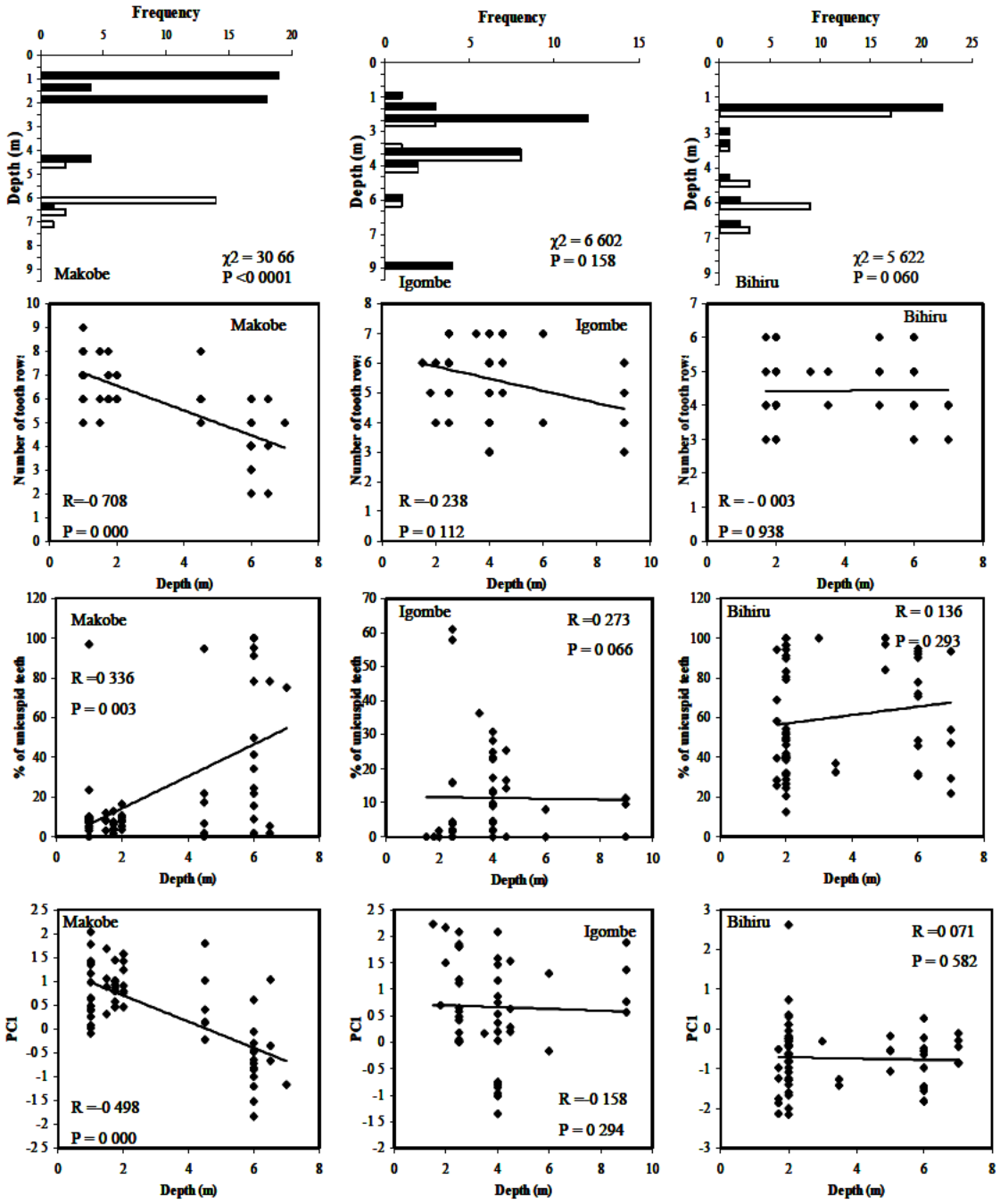

Figure 8. Depth distribution in meters of individuals by a) morph, b) Principal component one scores, b) Principal Component 2 scores and d) Principal Component 3 scores. Correlation coefficient (R) between depth and each variable and its significance (P-value) are also indicated.

Common garden experiment

There was significant variation for several morphological distances and dental characters between unrelated families and also between the two half-sib broods nested within each of the families (Table 6, Fig. 9). Unrelated families differed significantly in the number of tooth rows and in four morphological distances, namely standard length, lower jaw length, eye length and eye depth, as well as in PC1. Between half-sib broods, significant differences were found in the percentage of unicuspid teeth, and three morphological distances: head length, body depth and lower jaw length.

Table 6 - Results of nested ANOVAs on the number of tooth rows, percentage of unicuspid teeth, 13 standard morphological distances and the first three PCs based on those morphological distances for males of three unrelated families and two full-sib broods nested in each. Significant P-values are in bold.

	full-sib brood (family)			family		
character	df	F	Sig.	df	F	Sig.
tooth rows	3	1.818	0.162	2	9.482	**0.001**
% unicuspid teeth	3	6.455	**0.001**	2	2.055	0.143
morphometrics PC1	3	1.865	0.154	2	3.816	**0.032**
morphometrics PC2	3	1.435	0.249	2	0.759	0.476
morphometrics PC3	3	2.666	0.063	2	2.230	0.123
Standard Length	3	2.665	0.063	2	5.105	**0.011**
Head Length	3	8.018	**0.000**	2	1.273	0.293
Head Width	3	0.710	0.553	2	0.579	0.566
Body Depth	3	3.226	**0.034**	2	1.833	0.175
Lower Jaw Length	3	3.362	**0.029**	2	4.133	**0.024**
Lower Jaw Width	3	0.958	0.423	2	0.334	0.718
Snout Length	3	0.735	0.538	2	1.788	0.182
Snout Width	3	0.398	0.756	2	0.066	0.936
Cheek Depth	3	0.641	0.593	2	0.056	0.945
Pre-Orbital Depth	3	0.903	0.449	2	0.888	0.420
Inter-Orbital Width	3	0.758	0.526	2	0.707	0.500
Eye Length	3	1.326	0.282	2	6.638	**0.004**
Eye Width	3	1.310	0.287	2	20.032	**0.000**

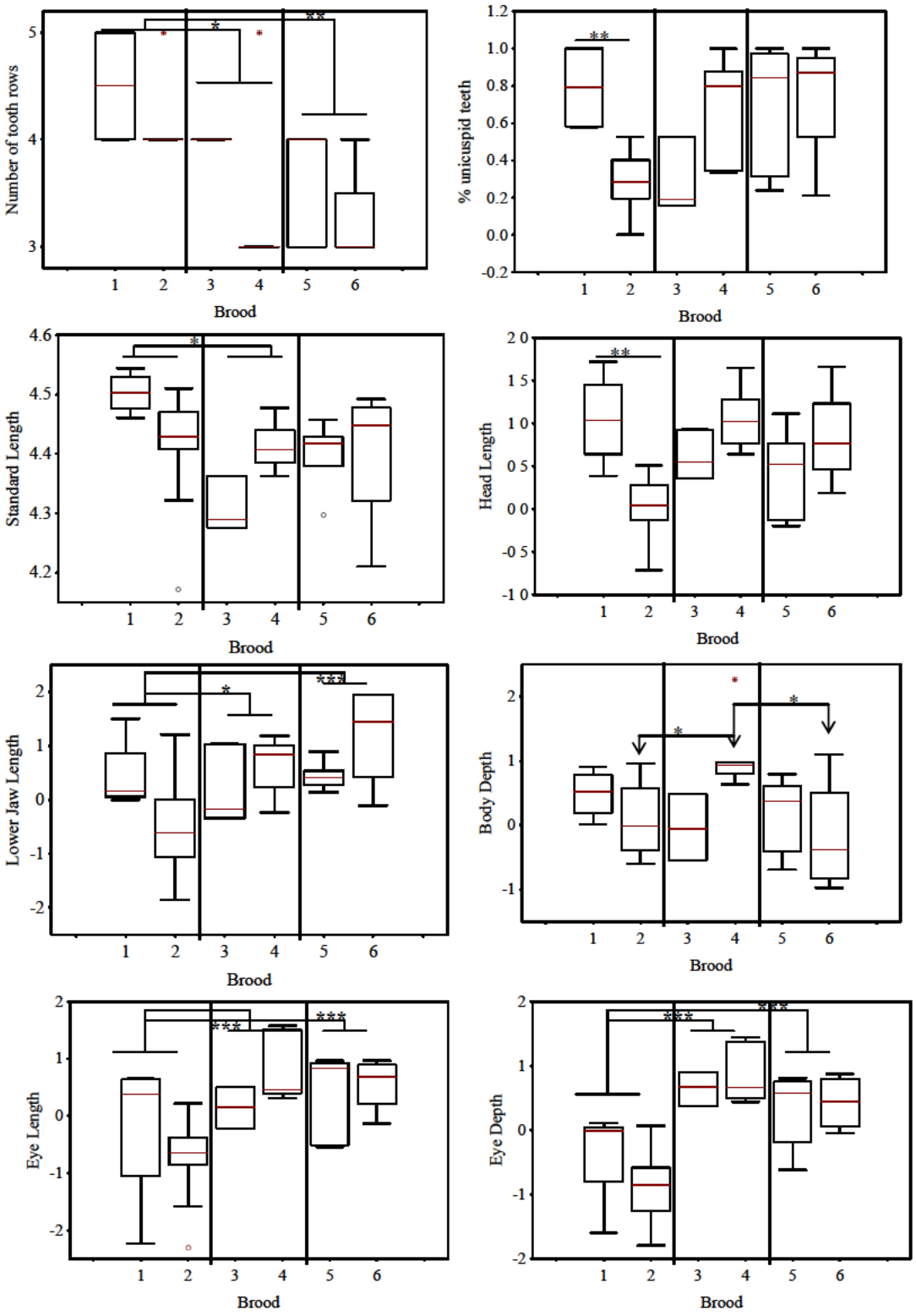

Figure 9. Boxplots of dental characters and morphological distances with significant differences between families and between broods within families. Significance of the differences indicated by * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

Discussion

In this study we investigated three island populations of benthivorous cichlid fish in Lake Victoria that exhibit similar eco-morphological polymorphisms, but vary in the extent of phenotypic and genetic differentiation of the morphs. We relate this variation in the extent of differentiation to variation in the structure of the island environment, specifically the near-shore slope and structure of the lake floor. Water depth and substrate-exposure to sunlight strongly affect the abundance and composition of food resources on the rocky substrate, and the sloping lake floor therefore constitutes an important dietary resource gradient. Across all islands, individual phenotypic variation within islands was significantly correlated with water depth in the direction predicted by adaptation to depth-related variation in dietary resources. To our knowledge, this is only the second demonstration of a phenotype-environment correlation within populations of haplochromine cichlid fish, and the first one involving trophic morphology. The other one is the relationship between nuptial colour, opsin genes and water depth in the genus *Pundamilia* (Seehausen et al., 2008 (chapter 4)).

Dietary resource gradients can be sources of strong divergent selection, and the predicted evolutionary response to such selection is the formation of clines in feeding-related characters. The strength of selection and the extent of dispersal and gene flow between demes along the gradient together predict the magnitude of the adaptive divergence. Theory predicts that some dispersal may increase the effectiveness of selection over that in situations without any gene flow because the interaction between locally adapted resident and immigrant genotypes can lead to disruptive selection along the resource gradient (Doebeli & Dieckmann, 2003; Gavrilets, 2004). However, larger amounts of gene flow can easily cancel the effects of selection and constrain adaptive divergence (Rasanen & Hendry, 2008).

Each of the islands studied has a different combination of water clarity, shore slope steepness and rocky substrate, which generates environmental gradients that differ in steepness and in spatial linearity. High water clarity and a gently sloping floor make the transition from sunlight-rich primary-production-dominated shallow water to less well illuminated deeper water long and gradual. Small boulders generate a locally homogeneous light environment and a spatially linear gradient of light-dependent resources, whereas large boulders create a mosaic of light habitats. Makobe is the most gently sloped island with the highest water transparency and the smallest boulder size. Together these factors make this the most gentle and most linear resource gradient.

Igombe Island is in many ways similar to Makobe but has a slightly steeper and less linear environmental gradient. This island has a steeper shore slope and probably due to its proximity to mainland, less clear water. The rocks are also slightly larger. The shore of Bihiru Island slopes steeply and is composed of very large rock boulders, some of which start at 10 meters below surface and emerge from the water surface projecting shadows over the shore that darken the environment. Taken together, these features create a patchy mosaic-like habitat with locally very steep gradients.

The within-island gradients from the surface to the deep end of the rocky habitat were between roughly 20 and 200m long. Individual dispersal distances of *Neochromis* are not known, but non-territorial fish can occasionally be observed cruising distances of 10m and more in a matter of minutes. Hence, our gradients certainly permit relatively high amounts of dispersal between habitats. As expected under such circumstances, eco-

morphs broadly overlapped in their depth distribution at all islands, implying that opportunity for frequency dependent competition, generating disruptive selection would exist at each. The Makobe environment would be most conducive, the Bihiru environment would be least conducive of adaptive divergence, and Igombe is intermediate but more similar to Makobe. At Makobe, the two eco-morphs are significantly differentiated in many eco-morphological traits (tooth shape, tooth row number, 8 of 13 morphometric distances, PC1) and eco-morphological variation is significantly correlated with water depth at the level of individuals. At Igombe the morphs are differentiated in fewer morphological traits (tooth shape, tooth row number, 4 of 13 morphometric distances, PC1) and phenotypic variation exhibits only non-significant trends of correlation with water depth. At Bihiru island the eco-morphs differ only in tooth cusp shape. Tooth cusp shape is also the only trait exhibiting any (though non-significant) trend of correlation with water depth there.

At neutral loci, the genetic differentiation between the morphs is the weakest at Bihiru, as indeed predicted. However, surprisingly, the phenotypically strongly differentiated eco-morphs at Makobe are barely more strongly differentiated at neutral loci than those at Bihiru (F_{ST} 0.0033 versus 0.0027), whereas the morphs from Igombe are clearly more strongly differentiated at neutral loci than both others (F_{ST} 0.0176). It hence appears that while eco-morphological differentiation among our three replicate eco-morph pairs is well predicted by the slope of the resource gradient, such variation does not strongly correlate with neutral genetic differentiation. In fact, morphological differentiation involving many different traits occurs at Makobe in the absence of neutral genetic differentiation.

Proximate and ultimate causes of phenotypic differentiation

All trends of eco-morphological divergence along the water depth gradient were in the direction predicted by adaptation in response to divergent selection along the resource gradient: more bicuspid teeth, more tooth rows and higher PC1 scores in shallow water. Head width, inter-orbital width, snout width, lower jaw width, body depth and eye size all loaded highly positively on PC1. All of these traits, perhaps with the exception of eye size, are known to be positively functionally correlated with feeding on firmly attached epilithic algae (Barel, 1983). Bicuspid tooth shape and number of tooth rows are direct adaptations to a feeding regime relying on scraping attached algae from hard substrates (Witte & Van Oijen 1990). Evidence for disruptive selection, however was found only in the cusp shape composition of outer row oral teeth.

If differentiation in these traits observed between eco-morphs in the wild were just due to phenotypic plasticity, we predicted no differences between families raised in a common environment. Instead, we found that half-sib families showed significant differences in the percentage of unicuspid teeth, body depth, head length, and lower jaw length, indicating heritable differences in these traits between the sires. Unrelated families differed significantly in number of tooth rows, lower jaw length, eye length and eye depth, indicating either heritable differences between parents or maternal effects. Hence, we consider it unlikely that the eco-morphological differentiation in the wild is just phenotypic response to the resource gradient.

Large variation in cichlid fish tooth cusp shape can be controlled by a single gene with major effect (Albertson et al., 2003; Streelman & Albertson, 2006). If applicable to

Neochromis, this would explain how considerable differences between eco-morphs can be maintained in tooth shape even in a panmictic population. Heritability of skull and jaw shape has been shown to be high too for Lake Malawi cichlids, with a small number of genes with additive effects controlling individual bony elements (Albertson et al., 2003). Studies on morphological differences between Lake Victoria cichlid species also support the idea that morphological differences between species are heritable (Bouton et al., 1999; Magalhaes et al., in press (chapter 5)), although they provide evidence for adaptive phenotypic plasticity in several morphological traits too.

Weak neutral genetic differentiation

The Bayesian assignment performed in STRUCTURE suggested existence of just one single population across major dispersal barriers. However, for weak population structure and low number of markers Structure is known to often fail to detect existing population structure (Evanno et al., 2005). The F_{ST} between populations from different islands were low but highly significant, indicating reduced migration between islands. Considering that Bihiru island is at 30 Kms of distance from Makobe, the F_{ST} values ranging from 0.012 to 0.018 between populations from Makobe and Bihiru is surprisingly low. Given that stenotopic rock habitat specialists, such as *Neochromis,* are believed to not readily cross stretches of soft substrate or open water, such weak differentiation is remarkable, and contrasts with what has often been reported from rocky shore cichlids in Lake Malawi (reviewed in Kocher 2004). Not only are F_{ST} values between distant populations lower in *Neochromis* than in ecologically similar cichlids of Lake Malawi, but they are lower also than those between fully sympatric sibling species of the closely related genus *Pundamilia* at Makobe Island, and rather resemble the weak differentiation observed between frequently hybridizing sympatric incipient species of *Pundamilia* at turbid water sites (Seehausen et al., 2008 (chapter 4)).

F_{ST} values between eco-morphs within islands were similarly low or lower and, except at Igombe, multilocus F_{ST} were non-significant, indicating that different eco-morphs within islands exchange genes similarly or more often than same eco-morphs from distant islands. Hence, a classical allopatric speciation scenario cannot explain the origin of sympatrically occurring eco-morphs. The highest and significant multilocus F_{ST} was observed between *N. greenwoodi* and *N.* "unicuspid scraper" at Igombe island. This F_{ST} value suggests that *N. greenwoodi* and *N.* "unicuspid scraper" at Igombe mate largely assortatively, and are ecological incipient species. The high frequency of significant linkage disequilibrium between multiple microsatellite loci when all genotypes from Igombe were pooled support that there is more than one incipient species present. However, while the amount of linkage within *N. greenwoodi* alone was much lower, the amount of linkage within *N.* "unicuspid scraper" was almost as high as in the pooled sample. Additionally, the difference between observed and expected heterozygosity was significantly larger in *N.* "unicuspid scraper" than in *N. greenwoodi*. All this implies more gene flow into *N.* "unicuspid scraper" than into *N. greenwoodi*.

At Makobe Island, despite the extremely low and non-significant multilocus F_{ST} value, the assignment test detected significantly non-random assignment of genotypes between the eco-morphs. Whereas almost all *N. omnicaeruleus* were correctly assigned, nearly half of the *N.* "unicuspid scraper" were assigned to *N. omnicaeruleus*, resembling the situation at Igombe. Similar too to Igombe, when genotypes of *N. omnicaeruleus* and

N. "unicuspid scraper" were pooled, there was a large amount of significant linkage disequilibrium. In contrast to Igombe, however, when the two were analysed separately *N. omnicaeruleus* had a much higher amount of linkage disequilibrium than *N.* "unicuspid scraper". This could be because the population of blue *N. omnicaeruleus* males is a mix of genotypes belonging to different subpopulations of partially assortatively mating female colour morphs (Seehausen et al., 1999; Magalhaes et al., submitted (chapter 2)). Either way, our data indicate that the eco-morphs at Makobe mate at best only incompletely assortatively and that ecological speciation has not proceeded sufficiently much to detect significant neutral population differentiation with our nine microsatellite loci. The level of differentiation resembles that between female colour morphs of *N. omnicaeruleus* (Magalhaes et al., submitted (chapter 2)) and is an order of magnitude below that between the *Pundamilia* species with different male colours at the same island.

The low and non-significant F_{ST} value, the assignment test, and the low amount of linkage all suggest that the tooth-shape morphs of *N.* "Bihiru scraper" belong to a single panmictic population, consistent with the environmental and morphological data.

Conclusion

Studying three replicate pairs of sympatric eco-morphs in the Lake Victoria cichlid genus *Neochromis*, we show consistent phenotype-environment correlations within islands. Islands vary in extent of eco-morph phenotypic differentiation, from single traits to more than 10 different traits. This variation is well predicted by variation in steepness and linearity of the resource gradient. The less steep and the more linear the gradient, the stronger the eco-typic differentiation, as predicted by theory. However, neutral genetic differentiation, which also varies between islands, is not as well predicted by the environment, and does not correlate well with eco-typic differentiation. We find evidence for deviations from random mating between sympatric morphs, consistent with incipient ecological speciation, but this has led to significant multilocus differentiation only at the island with intermediate resource gradient and intermediate eco-morphological differentiation. In fact, the eco-morphologically most strongly and the least strongly differentiated morph pairs both reveal absence of multilocus significant differentiation at microsatellite loci. Divergent and disruptive selection, perhaps aided by genetic architecture of eco-morphological traits, seems sufficient to maintain eco-morph differentiation in the face of nearly unrestricted gene flow. Progress towards ecological speciation, measured as neutral genetic differentiation, varies from nil to significant. It is in all cases less than what has been demonstrated at the same islands between the closely related species of *Pundamilia* with different male nuptial coloration (Seehausen et al., 2008) and is similar to what has been demonstrated between x-linked colour morphs within *N. omnicaeruleus* at Makobe Island (Magalhaes et al., submitted (chapter 2)).

Acknowledgements

We would like to thank Mhoja Kayeba and Mohammed Haluna for providing crucial assistance in the field, Arjun Sivasundar, Etienne Bezault and Martine Maan for the very useful discussions and comments on the manuscript and the members of the Eawag Fish Ecology & Evolution lab for helpful feedback and discussions. We thank the Tanzanian Fisheries Research Institute and the director of the Mwanza Centre, Mr. E.F.B. Katunzi, for hospitality and access to facilities, and the Tanzania Commission for Science and Technology for research permits. This research was supported by the Swiss National Science Foundation, Project 3100A0-106573 attributed to OS.

References

Albertson, R. C., Streelman, J. T., Kocher, T. D. & Yelick, P. 2003. Genetic basis of adaptive shape differences in the cichlid head. J. of heredity 94: 291-301.

Barel, C. D. N 1983. Towards a constitutional morphology of cichlid fishes (Teleostei, Perciformes). *Neth. J. of Zool*. 33: 357-424.

Barel, C. D. N., Vanoijen, M. J. P., Witte, F., Wittemaas E. L. M. 1977. Introduction to Taxonomy and Morphology of Haplochromine Cichlidae from Lake Victoria - Manual to Greenwoods Revision Papers. *Neth. J. Zool.* **27**, 333-338.

Belkhir K, Borsa P, Chikhi L, Raufaste N & Bonhomme, F. 1996. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Bouton, N., Witte, F., van Alphen, J. J. M., Schenk, A. & Seehausen, O. 1999. Local adaptations in populations of rock-dwelling haplochromines (Pisces: Cichlidae) from southern Lake Victoria. *P. Roy. Soc. B-Biol. Sci.* **266**: 355-360.

Brewer, M. J. 2003. Discretisation for inference on normal mixture models. *Statistics and Computing* **13**: 209-219.

Burnham, K. P. & Anderson, D. R. 2002. *Model selection and Multimodel inference,* New York Springer.

Case, T. J. & Taper, M. L. 2000. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. *Am. Nat.* **155**: 583-605.

Doebeli, M. & Dieckmann, U. 2003. Speciation along environmental gradients. *Nature* **421**, 259-264.

Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**: 2611-2620.

Excoffier, L., Estoup, A. & Cornuet, J. M. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727-1738.

Fleming, I. A. & Gross, M. R. 1994. Breeding Competition in a Pacific Salmon (Coho, Oncorhynchus-Kisutch) - Measures of Natural and Sexual Selection. Evolution 48: 637-657.

Gavrilets, S. 2004. *Fitness lanscapes and the origin of species*: Princeton University press.

Glaubitz, J. C. 2004. CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol. Ecol. Notes* **4** : 309-310.

Hardy, O. J. & Vekemans, X. 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 4: 618-620.

Hendry, A. P., Grant, P. R., Grant, B. R., Ford, H. A., Brewer, M. J. & Podos, J. 2006. Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. *P. Roy. Soc. B-Biol. Sci.***273**: 1887-1894.

Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J. W. 1996. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**: 1091-1093.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nat. Rev. Genetics* **5**: 288-298.

Maan, M. E., Eshuis, B., Haesler, M. P., Schneider, M. V., van Alphen, J. J. M. & Seehausen, O. 2008. Color polymorphism and predation in a Lake Victoria cichlid fish. *Copeia* **3**: 621-629.

Ogden, R. & Thorpe, R.S. 2002. Molecular evidence for ecological speciation in tropical habitats. *Proc. Nat. Acad. Sci. USA* **99**: 13612-13615.

Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. 1995. Microsatellite Analysis of Population-Structure in Canadian Polar Bears. *Mol. Ecol.* **4**: 347-354.

Pritchard, J. K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Rasanen, K. & Hendry, A. P. 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol. Lett.* **11**: 624-636.

Rice, W. R. 1989. Analyzing Tables of Statistical Tests. *Evolution* **43**: 223-225.

Rousset, F. 2000. Genetic differentiation between individuals. *J. Evol. Biol.* 13: 58-62.

Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., van der Sluijs, I., Schneider, M. V., Maan, M. E., Tachida, H., Imai, H. & Okada, N. 2008. Speciation through sensory drive in cichlid fish. *Nature* **455**: 620-626.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. Nat. Acad. Sci. USA* **271**: 1345-1353.

Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., Turner, G. F., van Alphen, J. J. M. & Bills, R. 2003. Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock. *P. Roy. Soc. B-Biol. Sci* 270: 129-137.

Seehausen, O. & van Alphen, J. J. M. 1999. Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol. Lett.* 2: 262-271.

Seehausen, O., van Alphen, J. J. M. & Lande, R. 1999. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**, 367-378.

Seehausen, O., Bouton N. 1998. The community of rock-dwelling cichlids in Lake Victoria. Bonner zoologische Beiträge, 47:301-311.

Seehausen, O**.**, Lippitsch, E., Bouton, N. 6 Zwennes, H., 1998. Mbipi, the rock-dwelling cichlids of Lake Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyol. Explor. Freshw*. **9**:129-228.

Seehausen, O. & Bouton, N. 1997. Microdistribution and fluctuations in niche overlap in a rocky shore cichlid community in Lake Victoria. *Ecol. of Freshw. Fish* **6**: 161-173.

Seehausen, O., van Alphen, J. J. M. & Witte, F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808-1811.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids. Taxonomy, Ecology and Distribution.* Zevenhuizen, NL: Cichlid Press.

Smith, T. B., Calsbeek, R., Wayne, R. K., Holder, K. H., Pires, D. & Bardeleben, C. 2005. Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. *J. Evol. Biol.* **18**: 257-268.

Smith, T. B., Wayne, R. K., Girman D. J. & Bruford, M. W. 1997. A role for ecotones in generating rainforest biodiversity. *Science* **276**: 1855-1857.

Stager, J. C. & Johnson, T. C. 2008. The late Pleitocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5-16.

Streelman, J. T. & Albertson, R. C. 2006. Evolution of the novelty in the cichlid dentition. *J. of Exp. Zool*. 306: 216-226.

Taylor, M. I., Meardon, F., Turner, G., Seehausen, O., Mrosso, H. D. J. & Rico, C. 2002. Characterization of tetranucleotide microsatellite loci in a Lake Victorian, haplochromine cichlid fish: a *Pundamilia pundamilia* x *Pundamilia nyererei* hybrid. *Mol. Ecol. Notes* **2**: 443-445.

Turner, G. F., Seehausen, O., Knight, M. E., Allender, C. J. & Robinson, R. L. 2001. How many species of cichlid fishes are there in African lakes? *Mol. Ecol.* **10**: 793-806.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**: 535-538.

Weir, B. S. & Cockerham, C. C. 1984. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**: 1358-1370.

Witte, F., Barel, . D. N. & Hoogerhoud, R. J. C. 1990. Phenotypic plasticity of anatomical structures and its ecomorphological significance. *Neth. J. Zool.* **40**: 278-298.

Wu, L., Kaufman, L. & Fuerst, P. A. 1999. Isolation of microsatellite markers in *Astatoreochromis alluaudi* and their cross-species amplifications in other African cichlids. *Mol. Ecol.* **8**: 895-897.

Zardoya, R., Vollmer, D. M., Craddock, C., Streelman, J. T., Karl, S. & Meyer, A. 1996. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *P. Roy. Soc. B-Biol. Sci.* **263**: 1589-1598.

Appendix 1. Allelic variability at 9 microsatellite loci. Number of samples analysed in each group (N), number of alleles at each locus (A), observed heterozygosity (H_O) and expected heterozygosity (H_E) at each locus. F_{IS}, H_O and H_E for multilocus. Significant differences between H_O and H_E are in bold with * (p<0.05), ** (p<0.01), *** (p<0.001).

		N. Igombe	N. green	N. uni Ig	N. Makobe	N. omni	N. uni Ma	N. Bh sc	N. bh sc bic	N. bh sc uni
	N	45	30	15	63	44	19	61	28	33
Ppun5	A	19	17	11	21	19	14	23	19	21
	HO	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.96	1.00
	HE	0.94	0.94	0.89	0.93	0.94	0.93	0.94	0.94	0.94
Ppun7	A	18	21	15	21	20	17	22	17	20
	HO	1.00	1.00	1.00	0.90	0.85	1.00	0.95	0.93	0.97
	HE	0.95	0.94	0.94	0.94	0.94	0.95	0.94	0.91	0.94
Ppun17	A	22	12	14	16	15	13	15	13	15
	HO	0.86	0.86	0.86	0.90	0.90	0.89	0.87	0.96	0.79
	HE	0.90	0.90	0.90	0.91	0.91	0.91	0.91	0.91	0.90
Ppun21	A	20	18	15	18	17	14	20	19	18
	HO	0.91	0.86	1.00	0.85	0.85	0.84	0.89	0.93	0.85
	HE	0.94	0.94	0.93	0.93	0.93	0.91	0.93	0.94	0.92
Ppun32	A	5	5	3	7	6	5	5	5	5
	HO	0.63	0.59	0.71	0.70	0.68	0.74	0.69	0.71	0.67
	HE	0.70	0.70	0.68	0.66	0.65	0.70	0.63	0.63	0.63
osu16d	A	25	21	14	30	28	18	30	19	28
	HO	**0.89***	0.87	0.93	0.98	0.98	1.00	0.95	0.96	0.94
	HE	**0.96**	0.95	0.94	0.95	0.96	0.93	0.95	0.94	0.96
osu19t	A	30	24	15	30	26	16	25	19	18
	HO	**0.73*****	**0.67*****	0.87	0.90	0.91	0.89	**0.80***	0.88	0.73
	HE	**0.93**	**0.93**	0.92	0.94	0.93	0.93	**0.86**	0.84	0.87
osu20d	A	37	31	18	55	45	25	43	30	32

	HO	0.93	**0.90***	1.00	0.95	0.98	0.89	0.90	0.93	0.88
	HE	0.97	**0.97**	0.97	0.98	0.98	0.97	0.96	0.97	0.96
TMO5	A	25	22	13	31	29	22	26	18	22
	HO	0.91	0.97	0.80	0.95	0.95	0.94	0.97	0.96	0.97
	HE	0.93	0.95	0.81	0.96	0.96	0.96	0.93	0.89	0.94
Average	HO	0.87	0.86	0.91	0.90	0.90	0.91	0.89	0.92	0.87
	HE	0.90	0.90	0.86	0.91	0.90	0.89	0.89	0.87	0.88
	FIS	0.05	0.07	-0.02	0.01	0.01	0.00	0.01	-0.03	0.04

Appendix 2. P-values of the tests of linkage disequilibrium for nine mircosatellite loci. Significant P-values, before sequential Bonferroni correction, are in bold.

	N. Makobe	N. omni	N. uni Ma	N. Igombe	N. green	N. uni Ig	N. Bh sc	N. bh sc bic	N. bh sc uni
Ppun5,ppun7	**0.021**	**0.009**	0.069	**0.047**	**0.049**	0.112	0.371	**0.016**	0.472
Ppun5,ppun17	**0.001**	**0.000**	0.773	0.113	0.774	0.121	0.247	0.880	0.194
Ppun5,ppun21	**0.000**	**0.000**	0.327	0.222	0.785	**0.033**	0.221	0.354	0.399
Ppun5,ppun32	**0.000**	**0.000**	0.974	**0.002**	**0.043**	0.133	0.385	0.264	0.634
Ppun5,osu16d	0.409	0.324	0.513	0.790	0.434	0.932	0.328	0.571	0.691
Ppun5,osu19t	0.089	0.084	0.272	0.733	0.954	0.069	0.324	0.646	0.105
Ppun5,osu20d	0.121	0.171	0.191	**0.029**	0.200	0.056	0.216	0.623	0.857
Ppun5,TMOm5	0.448	0.602	**0.014**	0.105	0.449	0.515	0.188	0.068	0.611
Ppun7,ppun17	**0.000**	**0.000**	0.727	**0.012**	0.304	**0.016**	0.086	**0.019**	0.212
Ppun7,ppun21	**0.000**	**0.000**	0.182	**0.017**	0.060	**0.008**	0.514	0.490	0.346
Ppun7,Ppun32	**0.000**	**0.000**	0.762	0.131	0.627	0.207	0.432	0.477	**0.030**
Ppun7,osu16d	0.754	0.620	0.125	0.610	0.604	0.321	0.095	0.088	0.597
Ppun7,osu19t	**0.007**	**0.004**	**0.017**	0.705	0.926	**0.040**	0.089	**0.021**	0.782
Ppun7,osu20d	0.085	0.112	0.176	**0.031**	0.058	**0.029**	0.936	0.999	0.254
Ppun7,TMOm5	0.716	0.469	0.825	0.212	0.664	**0.011**	0.088	**0.032**	0.816
Ppun17,ppun21	**0.000**	**0.000**	0.767	0.365	0.411	0.256	0.241	0.782	**0.018**
Ppun17,ppun32	**0.000**	**0.000**	0.185	**0.025**	0.631	0.163	0.124	**0.018**	0.430
Ppun17,osu16d	0.138	0.389	0.231	0.553	0.308	0.804	0.566	0.216	0.739
Ppun17,osu19t	0.121	0.109	0.236	0.873	0.740	0.219	**0.037**	0.198	**0.013**
Ppun17,osu20d	0.464	0.185	0.176	0.496	0.397	0.060	0.161	0.506	**0.007**
Ppun17,TMOm5	0.742	0.368	0.805	0.073	0.121	**0.019**	0.712	0.226	0.408
Ppun21,Ppun32	**0.000**	**0.000**	0.627	**0.008**	0.080	**0.003**	0.604	0.621	0.852
Ppun21,osu16d	0.142	0.317	**0.004**	0.206	0.080	0.438	0.178	0.964	0.359
Ppun21,osu19t	0.081	0.092	0.432	**0.042**	**0.030**	0.175	0.247	0.825	0.110
Ppun21,osu20d	0.327	0.462	0.061	0.182	0.293	0.530	0.105	0.479	0.178
Ppun21,TMOm5	0.484	0.232	0.679	0.478	0.786	0.640	0.655	**0.021**	0.641
Ppun32,osu16d	0.332	0.355	0.174	0.178	0.233	0.468	0.907	0.838	0.921
Ppun32,osu19t	0.176	0.150	0.140	0.717	0.424	0.065	0.082	**0.047**	0.253
Ppun32,osu20d	0.336	**0.039**	0.957	**0.019**	0.258	0.692	0.757	0.863	0.378
Ppun32,TMOm5	0.573	0.430	0.194	0.560	0.952	0.648	0.088	0.431	0.221
osu16d,osu19t	0.080	0.120	0.514	0.141	0.445	**0.046**	0.940	0.228	0.837

osu16d,osu20d	**0.047**	0.834	0.113	**0.006**	**0.013**	0.156	0.578	0.300	0.757
osu16d,TMOm5	0.866	0.262	0.937	0.864	0.871	0.288	**0.029**	0.102	0.226
osu19t,osu20d	0.117	0.387	**0.010**	0.229	0.331	0.058	**0.004**	0.199	0.058
osu19t,TMOm5	**0.000**	**0.018**	0.198	**0.000**	**0.018**	**0.006**	0.233	**0.004**	0.102
osu20d,TMOm5	0.467	0.204	**0.023**	0.694	0.770	0.060	0.055	**0.029**	0.175

Appendix 3. Single locus and multilocus F_{ST} values between morphs. Significant P-values are in bold. Significance of differences is indicated by * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

comparison	ppun5	ppun7	ppun17	ppun21	ppun32	osu16d	osu19t	osu20d	TMOm5	multilocus
Ng-N.us Ig	**0.027****	0.011	0.015	0.003	0.022	**0.022***	0.008	0.001	**0.055*****	**0.0176*****
No-N.us Ma	-0.010	0.006	0.012	0.016	-0.012	0.007	0.009	**0.013*****	-0.003	0.004
N.Bs bic-N.Bs uni	-0.004	**0.025*****	0.008	-0.004	-0.015	0.000	-0.006	0.001	**0.011***	0.003
No-Ng	0.000	0.002	0.003	0.010	**0.111*****	0.006	**0.025*****	**0.007****	0.001	**0.016*****
No-N.us Ig	**0.023****	0.005	0.007	**0.025****	0.025	**0.014***	0.013	**0.012*****	**0.078*****	**0.025*****
No-N.Bs bic	0.007	**0.020****	0.007	0.006	-0.004	0.007	**0.066*****	**0.011*****	0.035***	**0.017*****
No-N.Bs uni	-0.004	0.013	**0.022****	**0.015*****	0.005	0.005	**0.060*****	**0.014*****	0.008	**0.014*****
N.us Ma-Ng	-0.006	0.007	**0.027*****	**0.016***	**0.076*****	0.009	**0.035*****	0.003	0.004	**0.017*****
N.us Ma-N.us Ig	0.026	-0.010	**0.036*****	**0.031****	0.011	0.015	0.018	0.002	**0.075****	**0.021*****
N.us Ma-N.Bs bic	0.001	0.007	**0.017***	0.023	-0.006	**0.022*****	**0.030***	0.009	**0.023***	**0.014*****
N.us Ma-N.Bs uni	-0.006	0.000	**0.020***	0.024	0.003	**0.019****	**0.031*****	0.013	0.011	**0.012****
Ng-N.Bs bic	0.006	**0.021***	**0.022****	0.002	**0.076*****	**0.015****	**0.08*****	0.003	**0.031*****	**0.026****
Ng-N.Bs uni	-0.001	0.011	0.010	**0.011***	**0.07*****	**0.011***	**0.074*****	**0.012***	**0.016***	**0.021*****
N.us Ig-N.Bs bic	**0.039*****	0.005	**0.039*****	0.012	0.004	**0.037*****	**0.094*****	**0.017****	**0.027***	**0.029*****
N.us Ig-N.Bs uni	0.020	0.003	**0.045****	0.027*	0.009	**0.022****	**0.084*****	**0.013****	**0.067*****	**0.030*****

Appendix 4. P-values (two - tailed) of t-tests comparing single locus distributions of morphs from different islands. Significant P-values are in bold.

	Ng-N.us Ig	No-N.us Ma	N.Bs bic-N.Bs uni	No-Ng	No-N.us Ig	No-N.Bs bic	No-N.Bs uni	N.us Ma-Ng	N.us Ma-N.us Ig	N.us Ma-N.Bs bic	N.us Ma-N.Bs uni	Ng-N.Bs bic	Ng-N.Bs uni	N.us Ig-N.Bs bic	N.us Ig-N.Bs uni
Ng-N.us Ig	-														
No-N.us Ma	0.122	-													
N.Bs bic-N.Bs uni	**0.031**	0.595	-												
No-Ng	0.998	0.345	0.293	-											
No-N.us Ig	0.336	0.086	**0.038**	0.776	-										
No-N.Bs bic	0.922	0.109	0.074	0.954	0.600	-									
No-N.Bs uni	0.784	0.072	0.104	0.841	0.536	0.668	-								
N.us Ma-Ng	0.945	0.173	0.154	0.898	0.777	0.890	0.708	-							
N.us Ma-N.us Ig	0.421	0.079	0.047	0.780	0.930	0.586	0.502	0.764	-						
N.us Ma-N.Bs bic	0.552	**0.015**	0.044	0.775	0.311	0.556	0.769	0.637	0.265	-					
N.us Ma-N.Bs uni	0.514	**0.017**	0.108	0.692	0.309	0.500	0.553	0.505	0.266	0.632	-				
Ng-N.Bs bic	0.367	0.069	0.055	0.258	0.623	0.249	0.153	0.144	0.670	0.210	0.153	-			
Ng-N.Bs uni	0.633	0.103	0.099	0.491	0.912	0.491	0.301	0.382	0.943	0.360	0.258	0.127	-		
N.us Ig-N.Bs bic	0.289	**0.023**	0.028	0.478	0.531	0.059	0.039	0.381	0.510	0.060	0.051	0.854	0.532	-	
N.us Ig-N.Bs uni	0.162	**0.017**	**0.017**	0.416	0.314	**0.023**	**0.034**	0.305	0.241	**0.027**	**0.030**	0.722	0.439	0.773	-

Chapter 4

Speciation through sensory drive in cichlid fish

Speciation through sensory drive in cichlid fish

Ole Seehausen, Yohey Terai, Isabel S. Magalhaes, Karen L. Carleton, Hillary D. J. Mrosso, Ryutaro Miyagi, Inke van der Sluijs, Maria V. Schneider, Martine E. Maan, Hidenori Tachida, Hiroo Imai & Norihiro Okada

(*Nature* 455: 620-626)

Abstract

Theoretically, divergent selection on sensory systems can cause speciation through sensory drive. However, empirical evidence is rare and incomplete. Here we demonstrate sensory drive speciation within island populations of cichlid fish. We identify the ecological and molecular basis of divergent evolution in the cichlid visual system, demonstrate associated divergence in male colouration and female preferences, and show subsequent differentiation at neutral loci, indicating reproductive isolation. Evidence is replicated in several pairs of sympatric populations and species. Variation in the slope of the environmental gradients explains variation in the progress towards speciation: adaptation occurs on all but the steepest gradients. This is the most complete demonstration so far of speciation through sensory drive without geographical isolation. Our results also provide a mechanistic explanation for the collapse of cichlid fish species diversity during the anthropogenic eutrophication of Lake Victoria.

Introduction

The sensory drive hypothesis for speciation (Schluter et al, 1993; Boughman, 2002) predicts that adaptation in sensory and signalling systems to different environments in allopatry may cause premating isolation on secondary contact of populations. Recent theoretical work suggested that sensory drive can lead to the evolution of colour polymorphisms (Gray & McKinnon, 2007; Chunco et al, 2007) and speciation (Kawata et al, 2007) even in the absence of geographical isolation when the light environment is heterogeneous. However, the only case of sympatric sister species, in which assortative mating has been shown to be facilitated by sensory drive, were sticklebacks in British Columbia (Boughman, 2001). Here we provide ecological, population genetic and molecular evidence for each of the predictions of sensory drive speciation (Boughman, 2002)in sympatric cichlid fish inhabiting light gradients in Lake Victoria (East Africa).

Lake Victoria is spatially highly heterogeneous in water clarity and ambient light (Levring & Fish, 1956; Seehausen et al, 1997) and there is much evidence that the cichlid visual system has been under strong diversifying selection during the adaptive radiation of cichlids into several hundred species in Lake Victoria (Terai et al, 2006). Vertebrate visual pigments consist of a light-absorbing component, the chromophore, and a protein moiety, the opsin (Shichida, 1999). Spectral sensitivity is determined by the chromophore (A1 or A2 pigments), and by its interaction with the amino acid residues lining the retinal-binding pocket of the opsin in which the chromophore lies (Yokoyama et al, 2000). Eight different visual pigments have been found in all haplochromine cichlids (Carleton & Kocher, 2001; Terai et al, 2002; Parry et al, 2005), but only a subset of these are expressed in any individual species (Carleton & Kocher, 2001; Parry et al, 2005; Carleton et al, 2008). Several *Pundamilia* species from Lake Victoria expressed the same complement of four opsin genes: short-wavelength-sensitive opsin gene 2a (*SWS2A*, λ_{max} ~455 nm) in single cones; rhodopsin-like (*RH2*, λ_{max} ~528nm) and long-wavelength-sensitive opsin gene (*LWS*, λ_{max} ~565 nm) in double cones; and rhodopsin (*RH1*, λ_{max} ~505 nm) in rods (Carleton et al, 2005). Of these, the *LWS* opsin gene is by far the most variable among Lake Victoria cichlids (Terai et al, 2002; Spady et al, 2005), with sequence variation being five times greater than in Lake Malawi cichlids despite a tenfold greater age of the latter species flock (Genner et al, 2007).

Female Lake Victoria cichlids have mating preferences for conspicuously coloured males (Maan et al, 2004). Perception of conspicuousness is influenced by ambient and background light, signal transmission, receiver sensitivity and higher level processing (Boughman, 2002; Endler, 1993). Sympatric pairs of closely related cichlid species, one with red and one with blue nuptial colouration (Fig.1 and Supplementary Fig. 3), are common in Lake Victoria (Seehausen et al, 1997). Visual pigments have been compared for three pairs, and behavioural light detection thresholds measured in two. In each pair, the red species has its LWS λ_{max} at a longer wavelength (Carleton et al, 2005; Vandermeer et al, 1995), with a lower detection threshold for red but a higher one for blue light (Smit & Anker, 1997). These observations are consistent with a role for sensory drive in speciation, whereby interaction between ambient light, natural-selection-driven divergence of visual sensitivities and sexual selection for conspicuous male colours leads to the fixation of different male colours (Schluter et al., 1993; Boughman, 2002; Carleton et al., 2005; Maan et al., 2006).

Examining the role of environmental gradients in speciation requires tests at replicate gradients, as is recognized both in evolutionary ecology (Endler, 1973; Schluter & Nagel, 1995; Nosil et al., 2008) and in population genomics (Stinchcombe & Hoekstra, 2007). A recent model of clinal speciation through sensory drive (Kawata et al., 2007), as well as other models of clinal speciation (Doebeli & Dieckmann, 2003; Gavrilets, 2004; Leimar et al, 2008), predicts the greatest probability of speciation on gradients of intermediate slope. There, migration rates are sufficiently low to be compensated for by selection, but are sufficiently high to generate significant migration load (Nosil et al., 2005) and intermediate genotypes with a poor fit to the local environment. Migration load and reduced fitness of intermediate genotypes lead to disruptive selection, which may be required for the evolution of assortative mating through reinforcement-like mechanisms (Doebeli & Dieckmann, 2003; Gavrilets, 2004; Leimar et al., 2008). Previously we demonstrated adaptive evolution in the *LWS* opsin gene of the Lake Victoria cichlid fish *Neochromis greenwoodi* and *Mbipia mbipi* along very shallow gradients of light colour mediated by variation in turbidity between islands (Terai et al., 2006). *LWS* genotype frequencies and male colour morph frequencies formed correlated clines, but, even though populations at opposite ends of one gradient fixed different *LWS* alleles, all populations retained polymorphism for colour, indicating that speciation remained incomplete (Terai et al., 2006).

Here we investigate populations of cichlid fish living on light gradients primarily mediated by water depth within islands in Lake Victoria. *Pundamilia pundamilia* and *Pundamilia nyererey* (Seehausen, 1996) (Fig. 1a and Supplementary Fig. 3) are geographically fully sympatric. Within islands, they have narrowly parapatric depth ranges. Where they are phenotypically well differentiated, *P. pundamilia* has blue–grey male nuptial colouration whereas *P. nyererei* nuptial males are yellow with a bright crimson-red dorsum. Females of both are cryptically yellowish and have mating preferences for the nuptial colouration of conspecific males (Seehausen & van Alphen, 1998; Stelkens et al, 2008). The red *P. nyererei* occurs at greater mean water depths, in more red-shifted ambient light than the blue *P. pundamilia* (Maan et al, 2006). *P. nyererei* have a lower threshold for the detection of red light, whereas *P. pundamilia* possess a lower threshold for detection of blue light (Maan et al, 2006). Earlier we found that red and blue fish tended to possess different alleles at the *LWS* opsin gene locus (Carleton et al, 2005). Here we fully develop this system to test predictions of sensory drive speciation.

If sensory drive caused speciation into a red and a blue species, we expected to find: (1) variation in the *LWS* opsin sequence at amino acid positions where they shift λ_{max}; (2) an association of such sequence variation with water depth, such that more red-shifted alleles occur at greater depth; and (3) an association of *LWS* alleles with the predominant male nuptial colouration of a population, such that populations with predominantly red-shifted opsin alleles have predominantly red males. Furthermore, if disruptive selection was required to complete speciation through the evolution of assortative mating, we predicted that the strongest associations between *LWS* alleles, water depth and colour occur on intermediate light slopes (prediction 4). For testing prediction (4), we compared the data from the depth-mediated gradients of this study with data we had collected earlier on populations occupying the same depth at different islands with different turbidities (Terai et al., 2006) (see Supplementary Information).

Methods

Ambient light gradients and water clarity

Water transparency was measured using a white Secchi disk. Ambient, absorbance and transmittance light spectra between 400 nm and 750 nm were measured every metre between the surface and 3 m water depth with an Ocean Optics PS 1000 spectrophotometer and an optical fibre (100 μm), using SpectraWin 4.16 software (Avantes). Measurements were taken in the shade, between 8:50 and 9:00 in the morning. We calculated at every depth the 'transmittance orange ratio', which is a property of the water unaffected by variation in solar irradiance, as the ratio of transmittance in the 550–700 nm range (yellow, orange, red) over the total visible range (400–700 nm). The steepness of the light gradient, the 'light slope', was calculated by regressing the transmittance orange ratio against the mean distance (m) from the shore, measured along the lake floor in three transects for every island. The turbidity-mediated between-island light slopes were calculated by regressing the transmittance orange ratio measured at every island at 2 m water depth against the distance (m) from the clear water end of each gradient. The light differential was measured for both types of gradients as the difference between the transmittance orange ratios at the end points of a gradient. The largest possible value is 0.5, which is given when there is no longer any detectable blue light at the deep end of a gradient (transmittance orange ratio = 1 (that is, orange is the only transmitted light); whereas at the surface the full amounts of both blue and orange light are present (that is, transmittance orange ratio = 0.5)).

Frequency and depth distribution of male colouration

Males were collected by angling and gill nets in April and August 2001, February 2003, and January and May 2005. Photos were taken of 11 (Marumbi), 241 (Luanso), 64 (Kissenda), 34 (Python) and 130 (Makobe) males in breeding dress—480 in total—immediately on capture in specially designed photographic cuvettes. Photos were scored on a 5-point (0–4) colour phenotype scale by two to five independent observers, and the mean value was used (Van der Sluijs et al, 2007) (Fig. 1). Phenotype scoring of different observers was very similar (Spearman correlations between 0.605 and 0.729, $P < 0.05$). Linear regressions with a quadratic term were fitted to the log-transformed counts of the colour phenotypes from each island separately using *R* (Venables, 2002). Frequency distributions were compared between islands by *G*-tests.

Water depth was measured and recorded to the nearest 0.5 m for each of 960 males. The association between phenotype and water depth was tested for each island separately using ANOVA tests. These males were assigned to colour classes in the field, and only three robust classes were used: blue, intermediate and red (corresponding to classes 0 + 1, 2 and 3 + 4). *G*-tests were performed to compare depth distributions between islands. The curve-fitting procedure in SPSS (SPSS Inc. 2005) was used to quantify the relationship between strength of association (*F*-value) and steepness of the light slope.

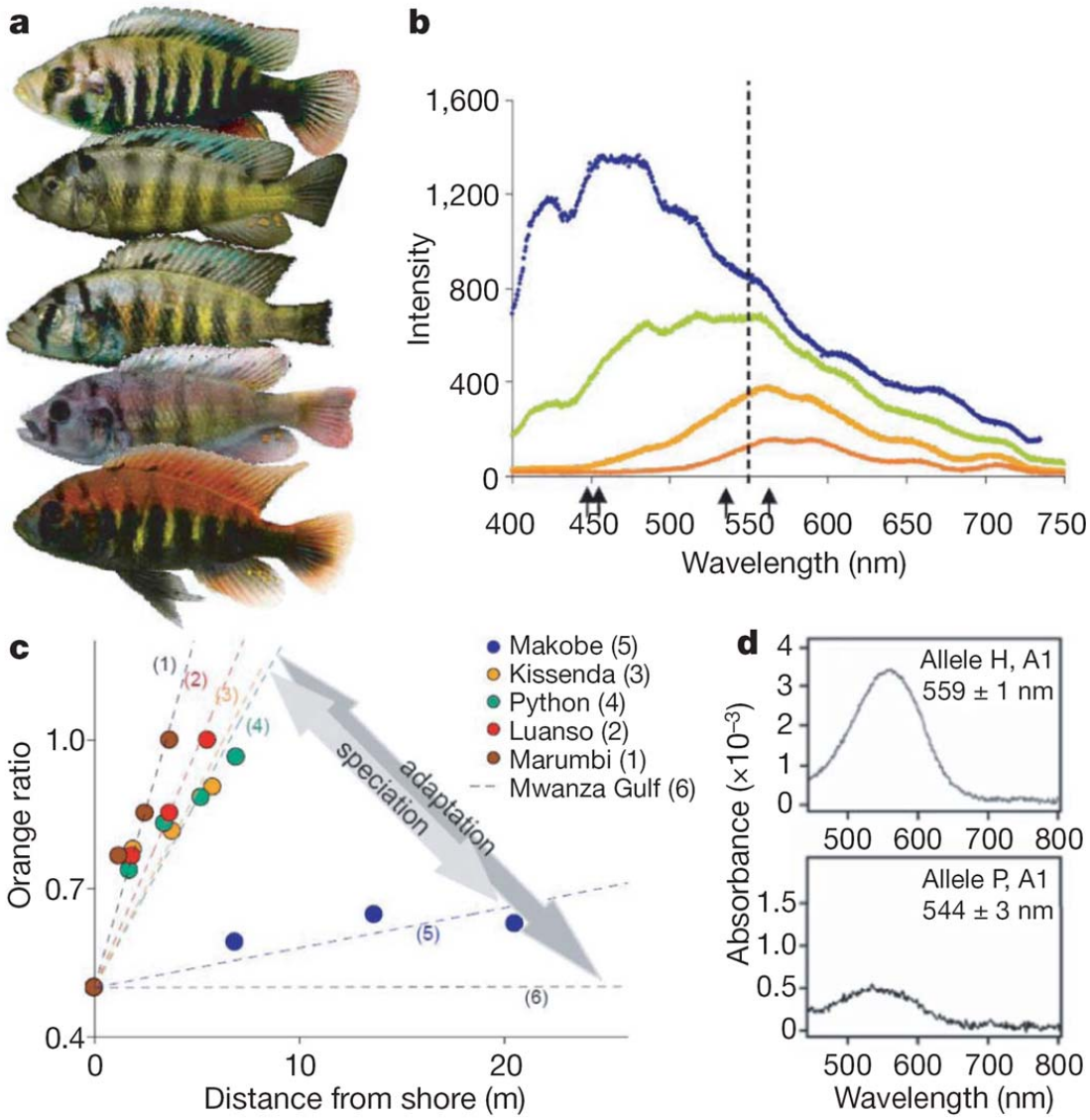

Figure 1 Male phenotypes, light gradients and *LWS* opsin absorbance. a, Variation in male nuptial colouration. Five phenotype classes from 0 ('blue', typical *P. pundamilia*; top) to 4 ('red', typical *P. nyererei*; bottom). **b**, An example of a moderately steep light gradient (Python island): surface light spectrum (blue) and three subsurface light spectra measured at 0.5 m (green), 1.5 m (orange) and 2.5 m (red) water depth. Line through 550 nm indicates divide used to calculate the transmittance orange ratio. Arrows indicate peak absorbance of two opsin pigments: main allele groups at *LWS* opsin locus (544 nm and 559 nm) and known range of peak absorbance at *SWS2A* locus (Carleton et al, 2005). **c**, Slopes of seven different light gradients. The lines for two shallow gradients overlay each other and are together labelled 'Mwanza Gulf'. For this line, the *x*-axis represents the distance from clear water (rather than from shore). Significant differentiation in opsin genes was observed on all gradients with slopes equal to or shallower than the Kissenda (orange) line, but speciation was observed only on gradients with slopes between the Kissenda (orange) and the Makobe (blue) lines. The dark grey arrow indicates region with divergent adaption at *LWS* opsin gene, and the light grey arrow indicates region with speciation. **d,** Absorption spectra of *LWS* pigments evaluated by the dark–light difference spectra (Terai et al, 2006). The *LWS* pigments were reconstituted from the H allele with A1 retinal (top) and from the P allele with A1 retinal (bottom). λ_{max} values (with standard errors) are indicated.

LWS absorption spectra

In vitro mutagenesis of *LWS* for construction of the sequence of P alleles, expression, reconstitution, purification and measurement were performed as described previously (Terai et al, 2006) with minor modifications. We measured absorption spectra of reconstituted pigments before and after irradiation with light (>490 nm). On the basis of the λ_{max} values determined by 3 independent difference spectra calculated from the measurements using independent preparations, we determined the absorption maximum values for each allele with standard errors.

Population genetics of neutral loci

DNA was extracted from fin tissue of 305 individuals (Marumbi 13, Luanso 61, Kissenda 59, Python 84, Makobe 88) and amplified using 11 microsatellite primers developed for these or other haplochromine species (see Supplementary Methods). We used Arlequin (Excoffier et al, 2005) to calculate observed and expected heterozygosities, to test for significance of departure from Hardy–Weinberg equilibrium for each locus in each population (1 million MCMC permutations), and for significant deviations from linkage equilibrium (10.000 permutations). After sequential Bonferroni correction (Rice, 1989), 3 out of 55 tests revealed significant deviations from Hardy–Weinberg equilibrium (1 locus each in *P. pundamilia* and *P. nyererei* from Makobe, 1 in *P. pundamilia* from Kissenda), and 2 tests of linkage equilibrium were significant: 1 in *P. pundamilia* from Python island and 1 in *P. pundamilia* from Kissenda island. Because there was no indication of any consistent linkage disequilibrium across populations between any pair of loci, all loci were retained for subsequent analysis. Molecular variance among individuals within and between phenotype groups was visualized in a factorial correspondence analysis performed over individuals in Genetix 4.05 (Belkhir et al, 1996). F_{ST} estimates and their significance were calculated over 100 permutations, as implemented in Arlequin (Excoffier et al, 2005).

Population genetics of opsin genes

Determination of the *LWS* gene was as described previously (Terai et al, 2002). We determined the sequences of exons 2–5 of *LWS* (872 bp), which encode the transmembrane region, from 263 individuals (526 haplotypes): Marumbi (12 individuals; 24 haplotypes), Luanso (27; 54), Kissenda (62; 124), Python (90; 180) and Makobe (72; 144). Additionally, we sequenced exons 2–5 of several hundred individuals of other species of Lake Victoria cichlids (Supplementary Fig. 4). Determination of the *SWS2A* gene is described in Supplementary Methods. We sequenced exons 1–5 (including introns) from males of Makobe (16 *P. pundamilia* and 17 *P. nyererei*) and Kissenda (20 blue and 20 red males). F_{ST} values for *LWS* and *SWS2A* sequences were calculated using DnaSP 4.0 (Rozas et al, 2003). The *SWS2A* sequence (1,930 bp) was split into two putative alleles for the analysis.

Molecular signature of selection on LWS

Determination of the *LWS* flanking sequences and the tests for detection of selection were performed as described previously (Terai et al, 2006) with minor modifications. The *LWS* gene and its 5 kb upstream and 3.5 kb downstream flanking sequences (total 10.5 kb) were amplified by long PCR (Terai et al, 2006) from 10 red and

9 blue males. To reflect the approximate frequencies of *LWS* alleles in the two phenotype populations, we included one heterozygous (H/P) individual of each nuptial colour. The McDonald test (McDonald, 1998) was calculated with the recombination parameter set to 2, 4, 10, 32 and 1,000 replicates.

Female mating preferences

We conducted laboratory two-way mate choice assays as described elsewhere (Haesler & Seehausen, 2005). Each female was tested in at least 5 trials with 5 different male pairs. A *G*-test was used to compare the frequency distributions of mating preferences between islands.

Results

Light, depth and colour

We examined depth-mediated light gradients at five islands. The light climate of Lake Victoria is dominated by effects of particulate (non-phytoplankton) matter, selectively absorbing and scattering light of short wavelengths (Okullo et al, 2007), causing successive shifts of ambient light towards longer wavelengths with increasing water depth (this study), and also with increasing turbidity (earlier study) (Levring & Fish, 1956; Seehausen et al, 1997). The rate at which ambient light changes with increasing depth is positively correlated with turbidity (Seehausen et al, 1997) (difference between islands in this study). The cichlids we study feed and breed right above and within the rocky substrate. We characterize depth-associated light gradients in their habitat by the change in the 'transmittance orange ratio' that occurs per metre as one moves outwards from the shore into the lake along the lake floor (the 'light slope', see Methods and Fig. 1b). Steeper slopes occur with more turbid water and steeper shores (Table 1). The steepest light slopes occurred at the most turbid sites, Marumbi and Luanso (Table 1 and Fig. 1c). Intermediate slopes occurred at Kissenda and Python islands, and the shallowest slope at Makobe island. The latter was still steeper than all the turbidity-mediated light slopes of earlier work. The size of the light differential between the ends of the gradients was similar between the five depth-mediated gradients, and larger than on the turbidity-mediated gradients (Table 1 and Supplementary Table 1).

Table 1. The five environmental gradients of this study

Island	Water clarity (cm Secchi)	Shoreline slope	Light slope	Light differential
Marumbi island	53 ± 8	0.82 ± 0.15	1.4×10^{1}	0.50
Luanso island	50 ± 10	0.54 ± 0.05	9.6×10^{2}	0.50
Kissenda island	78 ± 24	0.52 ± 0.12	7.9×10^{2}	0.50
Python island	96 ± 21	0.58 ± 0.24	7.6×10^{2}	0.50
Makobe island	225 ± 67	0.15 ± 0.04	8×10^{3}	0.35

Mapping the microdistribution of phenotypes on the five depth-mediated gradients using data from 960 males (Fig. 2a) revealed significant differences between islands. It showed the absence of any association between colour and ambient light (water depth) at Marumbi and Luanso (analysis of variance, ANOVA: df = 2, F = 1.1, P = 0.3, and df = 2, F = 0.3, P = 0.7, respectively), but significant associations at all other sites (ANOVA: df = 2 (Kissenda), df = 1 (Python, Makobe), F > 50, P < 0.0001), and increasing strength of association with decreasing light slope (F ratio against slope, logarithmic regression, df = 4, R^2 = 0.87, P = 0.021; Fig. 3). Blue phenotypes are associated with shallow waters (<3 m) in all locations, whereas red phenotypes occur in shallow waters only on the steepest gradients, and become restricted to greater depths with decreasing light slope. Frequency distributions of male nuptial colour phenotypes differ significantly between islands too (Fig. 2b). Distributions are unimodal and skewed towards blue on the two steepest gradients. They are bimodal with few intermediates on gradients of intermediate slope, and consist of two discrete classes, blue and red, on the shallowest within-island gradient.

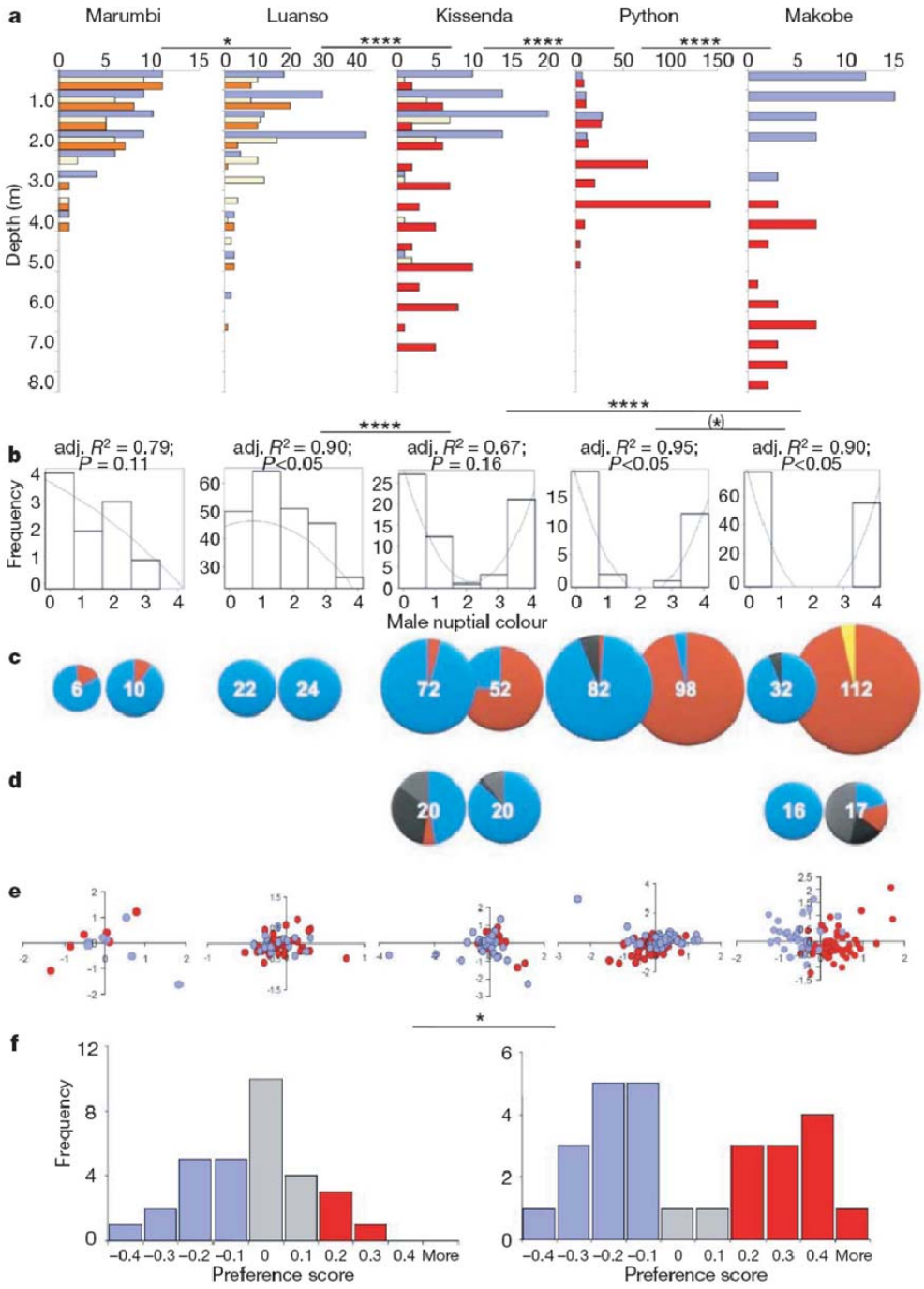
a
Marumbi
Luanso
Kissenda
Python
Makobe
Depth (m)
b
adj. R^2 = 0.79; P = 0.11
adj. R^2 = 0.90; P<0.05
adj. R^2 = 0.67; P = 0.16
adj. R^2 = 0.95; P<0.05
adj. R^2 = 0.90; P<0.05
Frequency
Male nuptial colour
c
6
10
22
24
72
52
82
98
32
112
d
20
20
16
17
e
f
Frequency
Preference score
Preference score

Figure 2 Ecological, phenotypic, genetic and behavioural divergence between blue and red *Pundamilia* nuptial phenotypes at five islands. All data for the same island are presented in the same column. Significant differences between islands indicated by asterisks (all tests two-tailed): $*P < 0.05$, $****P < 0.0001$, $(*)P < 0.1$. **a**, Depth distributions of male nuptial colour phenotypes. Blue bars, blue; pale yellow bars, intermediate; and orange or red bars, red (orange if dominated by class 3; red if dominated by class 4). Significance levels of differences between islands in the divergence between red and blue reported as *P* values of *G*-tests. **b,** Frequency distributions of male nuptial colour phenotypes (see Fig. 1a and text). Lines are quadratic fits; R^2 and significance levels indicated. Significance levels of differences between islands reported as *P* values of *G*-tests. **c**, Frequencies of functional allele groups at the *LWS* opsin gene by island and male colour (left, blue; right, red). Numbers report sample sizes of completely sequenced haplotypes. For Marumbi and Luanso islands, only the haplotypes of those individuals included that could be assigned to 'blueish' and 'reddish' phenotypes (24 and 54 haplotypes were sequenced from Marumbi and Luanso, respectively). Fish from Marumbi were divided into classes 0 + 1 and classes 2 + 3. Fish from Luanso were divided into classes 0 + 1 and 2–4. At all other islands, only fish of phenotype classes 0 and 4 were included. Alleles of the P group shown in blue, alleles of the H group in red, M3 alleles in yellow, and other alleles in grey. **d**, Allele frequencies at *SWS2A* opsin gene by island and nuptial colour class. The *SWS2A* P allele shown in blue, the N allele in red, other alleles in black, and alleles not determined in grey. **e**, Individuals plotted on first and second axes of a factorial correspondence analysis of genetic variance calculated from 11 unlinked microsatellite loci. Colours indicate pooled male nuptial colour classes as described in **c**. **f**, Histograms of female mating preferences at Luanso island (Van der Sluijs et al, 2007 (left) and Python island (Haesler et al, 2005 (right, includes new data). Blue: preference classes in which most females had statistically significant individual preferences for blue males, red: preference classes in which most females had significant preferences for red males, grey: preference classes in which females had no significant mating preference. Significance level of the difference in the frequency distributions between the two islands reported as *P* values of a *G*-test.

LWS gene variation, light and colour

We observed 13 polymorphic sites (3 synonymous, 10 nonsynonymous) among the *LWS* sequences (Supplementary Table 6). Three nonsynonymous substitutions occurred at high frequencies. From the bovine rhodopsin crystal structure (Palczewski et al, 2000) we inferred that two of these variable amino acid positions, 216 (nucleotide site 647) and 275 (823 and 824), are located in or near the retinal-binding pocket. The third one was position 230 (688), one of the tuning sites of human red/green opsin (Asenjo et al, 1994). Focusing on these three positions, we divided alleles into three groups described previously (Terai et al, 2006): the H group (all alleles with 216Y, 230A, 275C), the P group (216F, 230T, 275I) and the M3 group (216Y, 230T, 275I). M3 alleles can be considered recombinants or intermediate between H and P alleles. H and P alleles differed in only the 3 amino acid positions 216, 230 and 275. Substitutions at the other 7 nonsynonymous sites were rare and resulted in other allele variants (Supplementary Table 5).

We reconstituted the LWS pigments from P alleles *in vitro* with A1-derived retinal, and measured their absorption spectra, as previously done for the H alleles (Terai et al, 2006) (Fig. 1d). The peak spectral sensitivity (λ_{max}) of the A1 pigment of the P allele was blue-shifted by 15 nm relative to the H allele. The λ_{max} values of cone outer segments expressing either P or H pigments were previously measured by microspectrophotometry, reporting too that P pigments were blue-shifted relative to H pigments (Carleton et al, 2005). Hence, the absorption spectra of P and H alleles seem to be adapted to shallower and deeper water light environments in Lake Victoria, respectively (Fig 1b, d), supporting prediction (1).

Light gradients with slopes steeper than 0.09 were inhabited by populations with one or two different *LWS* alleles, whereas up to six different alleles were present on less steep gradients (Table 1 and Supplementary Tables 2 and 6). On these gradients of steepness $0.008 \leq x \leq 0.09$, H alleles were strongly associated with red nuptial colouration ($\chi^2 > 66$, $df = 1$, $P < 0.0001$; Spearman Correlation Coefficients 0.74, 0.91 and 1, respectively for slopes 0.079, 0.076 and 0.008; $P < 0.0001$), and were rare in blue phenotypes (Fig 2c, Supplementary Table 6 and Supplementary Information), supporting prediction (3).

A strong association between *LWS* alleles and water depth emerges from these results, supporting prediction (2): at Marumbi and Luanso islands, most individuals reside in waters less than 3 m deep. P alleles strongly dominate. At all other islands, only the blue phenotype is confined to depths less than 3 m, and P alleles predominate among these fish, even where gene exchange with the red phenotype is frequent (see later). The sweep to high frequency of H alleles in the red phenotype is associated with shifting larger fractions of the population to depths beyond 3 m. At Kissenda island, 75% of the *LWS* alleles of the red population belong to the red-shifted H group. The proportion of H alleles, residing in red individuals, increases to Python island and further to Makobe island, associated with successively increasing fractions of the red population living in deep water (Fig. 2a versus 2c). Red and blue phenotypes were highly significantly differentiated at the *LWS* locus at Kissenda, Python and Makobe islands (F_{ST} : Fixation index 0.65, 0.83, 0.89), but neither at Luanso nor at Marumbi islands (F_{ST} 0.00).

Gene flow at neutral loci

A sensory drive model of speciation predicts that the rate of divergence at the opsin loci should exceed the rate of divergence at neutral loci. Our data are fully consistent with this prediction (Table 2).

Table 2. Pairwise F_{ST} statistics between sympatric phenotypes

Island	Marumbi island	Luanso island	Kissenda island	Python island	Makobe island
Light slope	0.144	0.096	0.079	0.076	0.008
F_{ST} at *LWS* opsin locus					
LWS	0.000	0.000	**0.648**	**0.826**	**0.890**
F_{ST} at microsatellite loci					
Ppun21	0.000	0.000	0.010	0.006	**0.023**
Ppun7	0.000	0.000	0.003	**0.023**	**0.013**
Ppun5	0.000	0.002	0.002	0.000	0.010
Ppun32	0.041	0.005	0.000	0.016	**0.080**
Ppun17	0.000	0.000	0.000	**0.046**	**0.027**
OSU16d	0.017	0.002	0.011	0.006	**0.020**
OSU20d	0.002	0.000	**0.040**	0.004	**0.008**
OSU19t	0.000	0.022	0.013	0.013	**0.032**
TMO5	0.000	0.013	0.000	0.012	0.010
Pzeb3	0.000	0.000	0.002	0.048	0.107
Pzeb5	0.000	0.000	0.024	0.024	0.049
Multilocus (11µsats)	0.000	0.002	**0.010**	**0.014**	**0.026**

Significant F_{ST} ($P < 0.05$) are shown in bold.

Pairwise F_{ST} between sympatric blue and red phenotypes estimated from 11 microsatellite loci reveal no differentiation at Marumbi or Luanso islands (Fig. 2e), consistent with the unimodal frequency distributions of male nuptial colour variants and the absence (Marumbi) or rarity of really red males. Pairwise F_{ST} at all other islands suggest significant, albeit weak, differentiation, consistent with the strongly bimodal frequency distributions of male nuptial colour variants and the emergence of the really red phenotype at those islands. Whereas F_{ST} at the *LWS* locus jumps from 0 at Marumbi and Luanso to 0.65 at Kissenda, F_{ST} at microsatellite loci increases gradually and much more slowly (Figs 2c, e and 3). The number of loci carrying the signature of differentiation increases steadily from Marumbi and Luanso (0 out of 11) to Makobe island (8 out of 11; Table 2), consistent with the successive disappearance of intermediate phenotypes.

With the exception of Makobe island, all microsatellite F_{ST} among sympatric red and blue phenotypes are smaller than F_{ST} between any two allopatric populations of the blue phenotype, and 7 out of 10 of the red phenotype (Supplementary Fig. 1a and Supplementary Table 3). Even the largest between-phenotype F_{ST} at Makobe is smaller than most within-phenotype F_{ST} between islands. This suggests either more gene flow or more recent divergence between phenotypes within islands than between island populations of the same phenotype. It implies either parallel maintenance of phenotypic differentiation in the face of gene flow, or parallel sympatric speciation. All H alleles as well as the most frequent P allele are shared with several distantly related cichlid species (Supplementary Fig. 4). The two *Pundamilia* H alleles are the most frequent H alleles in those distantly related species too. Either red *Pundamilia* populations acquired these alleles once or multiple times from other species through introgressive hybridization, or the shared ancestor of red and blue *Pundamilia* possessed all the P and H alleles. In either scenario, the H and P allele split must pre-date the origin of the blue and red *Pundamilia* species.

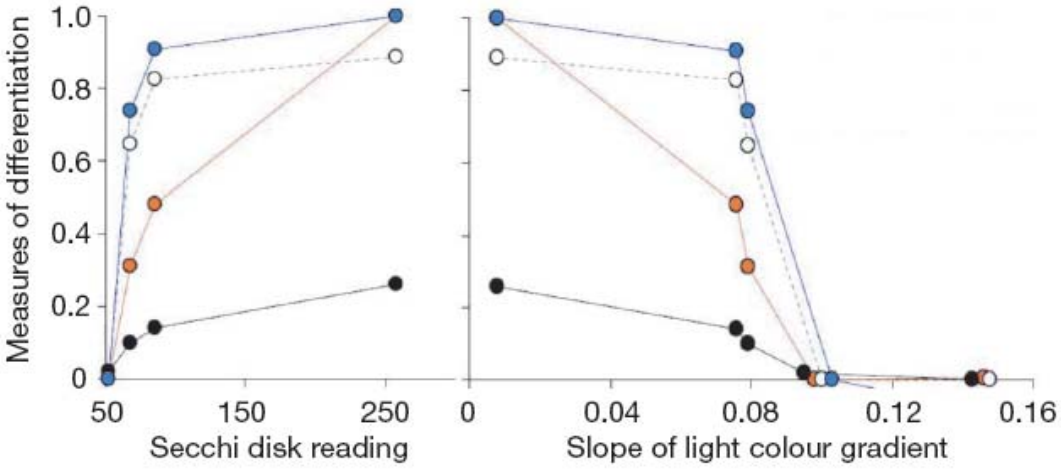

Figure 3. Measures of differentiation between sympatric *Pundamilia* phenotypes plotted against water transparency (left) and light slope (right). Blue symbols and line: Spearman rank correlations between colour and *LWS* genotype (best fit to water clarity $R^2 = 0.79$, $P = 0.045$, d$f = 4$; best fit to light slope $R^2 = 0.69$, P(one-tailed) = 0.042, d$f = 4$.). Open symbols and dashed line: *LWS* F_{ST} between red and blue phenotypes (best fit to water clarity, $R^2 = 0.79$, $P = 0.044$, d$f = 4$; best fit to light slope, $R^2 = 0.65$, P(one-tailed) = 0.045, d$f = 4$). Filled orange symbols and orange line: association between colour and water depth (ANOVA F ratios (normalized to range 0–1 for display); best fit to water clarity $R^2 = 0.99$, $P = 0.000$; best fit to light slope $R^2 = 0.87$, $P = 0.021$; both d$f = 4$). Filled black symbols and black line: microsatellite F_{ST} (multiplied by 10 for display) between red and blue phenotypes (best fit to water clarity, $R^2 = 0.99$, $P = 0.000$; best fit to light slope, $R^2 = 0.90$, P(one-tailed) = 0.013; both d$f = 4$).

Selection on the LWS gene

We analysed sequences up- and down-stream of *LWS* in a population (Python) that exhibits strong divergence in *LWS* but only weak differentiation at microsatellite loci. Sliding-window F_{ST} analysis revealed at least 6 times greater divergence in the *LWS* gene exons and in 2 kilobases (kb) of upstream sequence ($F_{ST} > 0.8$; Supplementary Fig. 2a) than in the downstream sequences ($F_{ST} < 0.15$), and more than 50 times greater divergence than at microsatellite loci (Table 2). Together with results of McDonald tests[38] and HKA tests (Hudson et al, 1987) (Supplementary Table 4 and Supplementary Information), this is consistent with a recent selective sweep in the red species, associated with increased presence in a red-shifted environment.

Divergence in the SWS2A opsin gene

We sequenced the *SWS2A* opsin gene at two islands to test for divergence at the short-wavelength end of the light spectrum. Out of 10 variable nucleotide positions, 5 were synonymous and 5 were located in introns (Supplementary Table 7). At Kissenda, the *SWS2A* sequences were variable in both phenotypes, and differentiated between them (F_{ST} 0.1, $P < 0.01$). At Makobe, a single *SWS2A* sequence variant was almost fixed in *P. pundamilia*, and the species were more strongly differentiated, although not as strongly as in *LWS* (F_{ST}: 0.437, $P < 0.001$; Fig. 2d).

Female mating preferences

Experiments and field data suggest that female *Pundamilia* use male colour as an important mate choice cue (Maan et al, 2004, Seehausen & van Alphen, 1998; Stelkens et al, 2008). Most wild and laboratory-bred Python island females prefer either blue or red males, but laboratory-bred F_1-hybrid females, most laboratory-bred F_2-hybrid females and most Luanso females have no preference between blue and red males (Haesler & Seehausen, 2005; Van der Sluijs et al, 2007). Combining published data (Haesler & Seehausen, 2005; Van der Sluijs et al, 2007) with previously unpublished data for 11 females from Python island, we find that the frequency distributions of female mating preferences differ between the islands (*G*-test, $P = 0.02$), roughly resembling those of male nuptial colour (compare Fig. 2f with 2b). The distribution at Luanso (38 females) had a single mode on no preference, and a skew towards blue preference. The distribution at Python (27 females) was bimodal.

We analysed Python island non-hybrid and laboratory-bred F_2 hybrid females to ask whether the *LWS* genotype directly determines mating preference. For non-hybrids and hybrids combined, we observed a significant association between individual *LWS* genotype and mating preference ($\chi^2 = 22$, $df = 10$, $P = 0.02$, 1,000 randomizations). However, this relationship was not significant when restricted to F_2 hybrid females (χ^2 10.2, $df = 6$, $P = 0.07$, 1,000 randomizations). Hence, variation in the *SWS2A-SWS2B-LWS* chromosomal region alone does not strongly predict visual mating preferences in a laboratory environment: some component of mating preference seems independent of it, consistent with biometric estimates that implied that the difference in mating preferences between *P. pundamilia* and *P. nyererei* was due to more than one factor (Haesler & Seehausen, 2005). Modelling light detection, using solar spectrum, water transmission, *Pundamilia* colour patch reflection and *Pundamilia* visual pigment absorption, suggested that a λ_{max} shift of 4 nm towards longer wavelengths causes a 10% increase in quantum catch for a fish looking at a red patch (Carleton et al, 2005). It seems probable that, in interaction with ambient light in the natural environment, the opsin genotype more strongly determines mating preference than it does under standard laboratory light conditions.

Discussion

Our data on ambient light colour, male nuptial colour, visual pigment ˜ $_{max}$ and female mating preference indicate sensory drive speciation, which occurred or is maintained by selection without geographical isolation. However, we only observed this under a restricted range of environmental conditions. At all sites with moderately shallow to moderately steep light gradients, two differentiated populations emerged with strong associations between water depth, *LWS* alleles, colouration and preferences (Fig. 3). Strong bimodalities in the quantitative traits colour and preference, strong heterozygote deficiencies at the *LWS* opsin gene, and differentiation at microsatellite loci clearly indicate speciation initiated by strong selection on *LWS*. Very steep light gradients, in contrast, were inhabited by single randomly mating populations that showed little variation in *LWS*, even though they contained some variation in colour and mating preference.

The following sensory drive speciation scenario is fully consistent with our data. First, divergent natural selection between light regimes at different water depths acts on

LWS. Second, sexual selection for conspicuous colouration is also divergent because perceptual biases differ between light regimes. Third, their interaction generates initial deviation from linkage equilibrium between *LWS* and nuptial colour alleles as observed on all but the steepest gradients. Fourth, subsequent disruptive selection due to reduced fitness of genotypes with a mismatch between *LWS* and colour alleles causes speciation, perhaps involving reinforcement-like selection for mating preferences, whereby male nuptial colour may serve as a marker trait for opsin genotype.

The strong association between *LWS* alleles and male nuptial colouration with few or no mismatch genotypes in sympatric species pairs is not restricted to *P. pundamilia* and *P. nyererei* (Table 3, Supplementary Fig. 3 and Supplementary Information). In contrast with these results, we did not find any such discontinuities in the frequency distribution of opsin genotypes along very shallow (between-island) gradients investigated earlier (Terai et al, 2006)—that is, intermediate *LWS* genotypes predominated in large sections of each gradient. This suggested the presence of divergent selection but the absence of disruptive selection (or the absence of an evolutionary response to disruptive selection). This is consistent with the low migration load predicted from the very small difference in ambient light that migrants between adjacent islands experience (Supplementary Table 1). Despite positive correlations between frequencies of *LWS* alleles and male nuptial colour morphs, and complete fixation of different *LWS* alleles between some populations, speciation as would be indicated first, by strong association between *LWS* and colour and, second, by genotypic and phenotypic discontinuities was not observed on these gradients. This may be due to a difference between the taxa that we studied, but it may also imply that speciation requires disruptive selection, and hence migration and gene flow between habitats(Kawata et al, 2007; Doebeli & Dieckmann, 2003; Gavrilets, 2004; Leimar et al, 2008; Nosil et al, 2003). In contrast, when migration exceeds selection, divergence cannot occur either (Nosil & Crespi, 2004; Rasanen & Hendry, 2008). This explains the absence of speciation on the steepest of our gradients.

Table 3. *LWS* opsin allele-group frequency (%) and male nuptial colouration in species of *Pundamilia*

Species	Population	Male nuptial colour type	P	M3	H	Others	*n*†
P. "Luanso"	Luanso island	Predominantly blue	100	0	0	0	54
P. "Marumbi"	Marumbi island	Predominantly blue	92	0	8	0	24
P. pundamilia	Makobe island	Blue	94	0	0	6	32
P. pundamilia	Igombe island	Blue	83	17	0	0	6
P. pundamilia-like*	Kissenda island	Blue	96	0	4	0	70
P. pundamilia-like*	Python island	Blue	90	1	4	5	82
P. azurea **	Ruti island	Blue	100	0	0	0	6
P. nyererei-like*	Kissenda island	Red dorsum	25	0	75	0	52
P. nyererei-like*	Python island	Red dorsum	4	0	96	0	98
P. nyererei	Makobe island	Red dorsum	0	4	96	0	112
P. igneopinnis	Igombe island	Red dorsum	0	0	100	0	6
P. "red head" **	Zue island	Red chest	0	0	100	0	6
Total							548

* Hybridizing populations (this study).
** Carleton et al, 2005.
† *n* represents *n* haplotypes sequenced.

Our results are relevant to conservation because they provide a mechanistic explanation for the collapse of cichlid fish species diversity during the anthropogenic eutrophication of Lake Victoria (Seehausen et al, 1997). Eutrophication changes the slope of environmental light gradients, and, by steepening them, potentially moves sites from the region in parameter space that is permissive of species coexistence into the region that is not. We hope these results help focus attention of biodiversity conservation efforts in Lake Victoria and other lakes to issues of water quality.

References

Asenjo, A. B., Rim, J. & Oprian, D. D. 1994. Molecular Determinants of Human Red/Green Color Discrimination. *Neuron* **12**: 1131-1138.

Belkhir K., Borsa P., Chikhi L., N., R. & F., Bonhomme 1996-2004. Genetix Version 4.05 for Windows Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Boughman, J. W. How sensory drive can promote speciation.2002. *Trends Ecol. Evol*. **17**: 571-577.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**: 944-948.

Carleton, K., Spady, T. C., Streelman, J. T., Kidd, M. R.., McFarland, W. N., & Lowe, E. R. 2008. Visual sensitivities tuned by heterochronic shifts in opsin gene expression. *BMC Evol. Biol.*22.

Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O.2005. Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Mol. Ecol.* **14**: 4341-4353.

Carleton, K. L. & Kocher, T. D. 2001. Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Mol. Biol. and Evol*. **18**: 1540-1550.

Chunco, A. J., McKinnon, J. S. & Servedio, M. R. 2007. Microhabitat variation and sexual selection can maintain male color polymorphisms. Evolution 61, 2504-2515.

Doebeli, M. & Dieckmann, U.2003. Speciation along environmental gradients. *Nature* **421**: 259-264.

Endler, J. A. 1973. Gene Flow and Population Differentiation. Science 179: 243-250.

Endler, J. A. 1993. Some General-Comments on the Evolution and Design of Animal Communication-Systems. *Phil. Trans. Roy. Soc. B*.**340**: 215-225.

Excoffier, L., Laval, G. & Schneider, S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. . *Evolutionary Bioinformatics Online* **1**: 47-50.

Gavrilets, S. 2004. Fitness lanscapes and the origin of species. Princeton University Press.

Genner, M. J. et al. 2007. Age of cichlids: New dates for ancient lake fish radiations. *Mol. Biol. and Evol*. **24**: 1269-1282.

Gray, S. M. & McKinnon, J. S. 2007. Linking color polymorphism maintenance and speciation. Trends in Ecology & Evolution **22**: 71-79.

Haesler, M. P. & Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Pr. Roy. Soc. B-Biol. Sci.* **272**: 237-245.

Hudson, R. R., Kreitman, M. & Aguade, M. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* **116**: 153-159.

Kawata, M., Shoji, A., Kawamura, S. & Seehausen, O. 2007 A genetically explicit model of speciation by sensory drive within a continuous population in aquatic environments. BMC Evolutionary Biology 7: 99.

Leimar, O., Doebeli, M. & Dieckmann, U. 2008. Evolution of phenotypic clusters through competition and local adaptation along an environmental gradient. *Evolution* **62**: 807-822.

Levring, T. & GR, F. 1956. The penetration of light in some tropical East African waters. *Oikos* **7**: 98-109.

Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. 2006. Sensory drive in cichlid speciation. *Am. Nat*. **167**: 947-954.

Maan, M. E. et al. 2004. Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid Pundamilia nyererei. *Pr. Roy. Soc. B-Biol. Sci.* **271**: 2445-2452.

McDonald, J. H. 1998. Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. and Evol*. **15**: 377-384.

Nosil, P., Egan, S. R. & Funk, D. J. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**: 316-336.

Nosil, P., Vines, T. H. & Funk, D. J. 2005. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution **59**: 705-719.

Nosil, P. & Crespi, B. J. 2004. Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in Timema cristinae walking-sticks. *Evolution* **58**: 102-112.

Nosil, P., Crespi, B. J. & Sandoval, C. P. 2003. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Pr. Roy. Soc. B-Biol. Sci.* 270: 1911-1918.

Okullo, W. et al. 2007. Parameterization of the inherent optical properties of Murchison Bay, Lake Victoria. *Applied Optics* **46**: 8553-8561.

Palczewski, K. et al. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**: 739-745.

Parry, J. W. L. et al. 2005. Mix and match color vision: Tuning spectral sensitivity by differential opsin gene expression in Lake Malawi Cichlids. *Current Biol*. **15**: 1734-1739.

Rasanen, K. & Hendry, A. 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. Ecol. Lett. **11**: 624-636.

Rice, W. R. 1989. Analyzing Tables of Statistical Tests. *Evolution* **43**: 223-225.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Schluter, D. & Nagel, L. M. 1995. Parallel Speciation by Natural-Selection. *Am. Nat*. **146**: 292-301.

Schluter, D. & Price, T.1993. Honesty, Perception and Population Divergence in Sexually Selected Traits. *P. Roy. Soc. B-Biol. Sci.* **253**: 117-122.

Seehausen, O. & van Alphen, J. J. M. 1998. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (Haplochromis nyererei complex). *Behav. Ecol. Sociobiol* **42**: 1-8.

Seehausen, O., vanAlphen, J. J. M. & Witte, F.1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808-1811.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids, 1996. Taxonomy, Ecology and Distribution*. Zevenhuizen, NL: Cichlid Press.

Shichida, Y. 1999. The Retinal Basis of Vision: Visual pigment: photochemistry and molecular evolution (ed. J, T.) (Elsevier Science).

Smit, S. A. & Anker, G. C. 1997. Photopic sensitivity to red and blue light related to retinal differences in two zooplanktivorous haplochromine species (Teleostei, Cichlidae). *Neth. J. of Zool*. **47**: 9-20.

Spady, T. C. et al. 2005. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol. Biol. and Evol*. **22**: 1412-1422.

Stinchcombe, J. T. & Hoektsra, H. E. 2007. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*. **100**: 158-170.

Stelkens, R. B., Pierotti, M.E.R., Joyce, D.A., Smith, A.M., Van der Sluijs, I., Seehausen, O. 2008. Female mating preferences facilitate disruptive sexual selection on male nuptial colouration in hybrid cichlid fish. *Phil.Trans.Roy.Soc. B*. **363**: 2861-2870.

Terai, Y. et al. Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids 2006. *Plos Biology* **4**: 2244-2251.

Terai, Y., Mayer, W. E., Klein, J., Tichy, H. & Okada, N 2002. The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. *P. Nat. Acad. Sci. USA* **99**: 15501-15506.

Vandermeer, H. J., Anker, G. C. & Barel, C. D. N. 1995. Ecomorphology of Retinal Structures in Zooplanktivorous Haplochromine Cichlids (Pisces) from Lake-Victoria. *Env. Biol. of Fishes* **44**: 115-132.

Van der Sluijs, I., Alphen, J.J.M., Seehausen, O. 2007. Preference polymorphism for coloration but no speciation in a population of Lake Victoria cichlids. *Behav. Ecol*. 19: 177-183.

Venables, W. N. & Ripley, B. D 2002. *Modern applied statistics with S*. New York: Springer-Verlag.

Yokoyama, S., Blow, N. S. & Radlwimmer, F. B, 2000. Molecular evolution of color vision of zebra finch. *Gene* **259**: 17-24.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the Tanzania Fisheries Research Institute for research permissions, hospitality and logistical support; M. Kayeba, M. Haluna, S. Mwaiko, M. Haesler and E. Burgerhout for help with data and fish collection; H. Araki, L. Excoffier, L. Harmon, B. Ibelings, I. Keller, T. Kocher, P. Nosil, M. Pierotti, A. Sivasundar and O. Svensson for comments on the manuscript; and M. Kawata, J. J. M. van Alphen, K. Young, R. Stelkens and E. Bezault for discussion. This work was supported by Swiss National Science Foundation project 3100A0-106573 (to O.S.), and by the Ministry of Education, Culture, Sports, Science and Technology of Japan (to N.O.).

Author Contributions

O.S. conceived and designed the study, collected, photographed and identified fish, measured light and shore slopes, supervised field work, conducted the hybridization experiments, supervised microsatellite analyses and mate choice experiments, did the statistical data analyses and the writing. Y.T. designed experiments on opsins, did most of the laboratory work and data analysis on opsins and contributed to writing, I.S.M. collected depth distribution data and did all microsatellite analyses. K.L.C. determined *LWS* sequences from experimental females and contributed to writing. H.D.J.M. collected depth distribution, light data, and fish. R.M. determined *LWS* and *SWS2A* sequences with Y.T. I.v.d.S. collected fish and conducted mate choice experiments. M.V.S. helped with the microsatellite analysis. M.E.M. collected fish and measured light. H.T. performed analysis of selection pressure with Y.T. H.I. measured opsin pigment absorbance with Y.T. N.O. designed and supervised the laboratory work on opsins and contributed to the writing.

SUPPLEMENTARY MATERIAL

LWS gene variation, depth and colour

H alleles were strongly associated with red nuptial colouration and were rare in blue phenotypes: 1 copy each in 3 *LWS* heterozygotes out of 36 blue individuals from Kissenda Island, 1 copy each in 3 *LWS* heterozygotes out of 41 blue fish from Python Island, no copy in 16 blue fish from Makobe Island (Fig. 2C, table S5). On the steepest gradient inhabited by two phenotypically differentiated populations (Kissenda Island, slope 0.079), 25% of the alleles residing in red phenotypes were blue-shifted P alleles (4 homozygotes and 5 heterozygotes among 26 individuals). This suggests recent introgression of LWS alleles from blue into red phenotypes, or a recent onset of divergence of the red population from the blue population. At Python Island (slope 0.076) only 5% of the LWS alleles residing in red phenotypes were P alleles (1 copy each in 4 LWS heterozygotes out of 41 fish), suggesting less gene flow, stronger divergent selection or earlier onset of divergence. Finally, no P alleles were observed in 56 red phenotypes at Makobe Island (slope 0.008). However, all 4 heterozygotes in this population carried one copy of a recombinant M3 allele, consistent with gene flow in the past.

LWS absorption spectra

Additional to H and P alleles, we also measured two intermediate alleles between H and P as follows: H allele with 275I: 557 nm. H allele with 230T and 275I (defined as M3 allele) 545-550 nm. The 15 nm shift between H and P pigments was mainly caused by A230T, and the other replacements (Y216F and C275I) may have synergistic effects.

Selection on the LWS gene

Sliding window analysis revealed fairly high levels of polymorphism in blue, but very low levels in red males (Fig. S2B). Among red but not among blue males there was significant within-sequence heterogeneity in the ratio of polymorphism to divergence (McDonald test[1], p=0.001, and p>0.58 respectively). The ratio of polymorphism in red to divergence between red and blue was significantly smaller in the -1000bp region (*LWS* exons 1 and 2 plus 1 kbp upstream, a possible promotor region) than in the downstream flanking region (HKA test[2], p<0.05; Fig. S2B, table S3). All this is consistent with a recent selective sweep in the red population, associated with increased presence in deep waters with relatively red-shifted ambient light.

Integrated analysis on effects of gradient slope

In a single combined analysis of the data from this and from our previous study of two other species in the same part of Lake Victoria, we found that progress towards speciation, as indicated by the association between colour and *LWS* genotype, was related by a positive linear term to the magnitude of the light gradient differential, while the residual variation was related to the slope of the gradient by a negative quadratic term, consistent with prediction (4): speciation most advanced on intermediate slopes.

13 gradient models: The relationship between the colour * *LWS* genotype correlation and gradient slope was best approximated by a negative quadratic term, i.e. speciation was most advanced on gradients with intermediate slope (p (one-tailed) 0.032, adjusted R^2 0.31, 12 d.f.). The relationship with gradient differential was significantly

approximated by a positive linear term, i.e. speciation was more advanced on gradients with larger differential (*p* (one-tailed) 0.037, adjusted R^2 0.20, 12 d.f.). However, it was even better approximated by a negative quadratic term (*p* (one-tailed) 0.030, adjusted R^2 0.32, 12 d.f.). The residual variation in the colour * *LWS* genotype correlation, both from the linear and the quadratic regression against gradient differential, was best explained by negative quadratic terms (residuals from linear regression: *p* (one tailed) 0.025, adjusted R^2 0.34, 12 d.f.; residuals from quadratic regression: *p* (one tailed) 0.020, adjusted R^2 0.37, 12 d.f). Multiple regression identified a model with both predictors, gradient slope (coefficient -12.4, *p* 0.000) and gradient differential (coefficient +3.4, *p* 0.000) together as best explaining variation in the correlation between colour and *LWS* genotype. This model was highly significant (*p* (one-tailed) 0.000, 12 d.f.).

7 gradient models: The relationship between the colour * *LWS* genotype correlation and gradient slope was again best approximated by a negative quadratic term, i.e. speciation was most advanced on gradients with intermediate slope, but this remained statistically non-significant (*p* (one-tailed) 0.09, adjusted R^2 0.36, 6 d.f.). The colour * *LWS* genotype correlation was not significantly related to the gradient differential (linear model: *p* (one tailed) 0.30, 6 d.f.; quadratic model: *p* (one tailed) 0.40, 6 d.f.), neither was the residual variation significantly explained by gradient slope (linear model: *p* (one tailed) 0.19, 6 d.f.; quadratic model: *p* (one tailed) 0.13, 6 d.f.). The multiple regression approach identified a model with both, gradient slope (coefficient -11.5, *p* (one-tailed) 0.033) and gradient differential (coefficient +2.8, *p* (one-tailed) 0.08) included as best explaining variation in the correlation between colour and *LWS* genotype, but the model remained non-significant (*p* (one-tailed) 0.067, 6 d.f.).

Association between opsin genotype and female mating preference

To examine the relationship between opsin genotype and mating preference we sequenced exons 2 to 5 of both *LWS* alleles of 35 laboratory-bred females with known mating preferences: 4 *P. pundamilia*, 6 *P. nyererei* (Python Island populations), and 25 *P. pundamilia* x *P. nyererei* F2 hybrid females. We also sequenced the *SWS2A* gene of 18 of these, 3 *P. pundamilia*, 3 *P nyererei*, and 12 F2 hybrids. *P. pundamilia* and *P. nyererei* differed by 4 substitutions in *SWS2A* positions 354, 1018, 1613, 2001. All four sites were differentiated in the same direction at Makobe Island (Figure 2D). We denote the *P. pundamilia*-derived allele 1018A/1613T as S2A-P, and the *P. nyererei*-derived allele 1018G/1613G as S2A-N. Other alleles can be interpreted as recombinants between these. *LWS* and *SWS2A* are about 12.4 kbp apart on the same chromosome (with the *SWS2B* opsin between them). The opsin haplotype segregated in the expected 1:2:1 ratio in the F2 hybrids. As expected, we found *LWS*-P alleles associated with *SWS2A*-P alleles, and *LWS*-H alleles with *SWS2A*-N alleles. However, 3 *LWS* heterozygotes among the F2 hybrids were homozygous for *SWS2A*-N alleles, suggesting some recombination. All 4 *P. pundamilia* females were homozygous for *LWS*-P alleles, and the 3 for which *SWS2A* was sequenced, were homozygous for the P-haplotype of the entire opsin gene battery. Two had significant mating preferences for blue males, the other two showed the same trend but were not significant. All 6 *P. nyererei* females had significant mating preferences for red males. One of them was heterozygous, with one *LWS*-H and one *LWS*-P allele. The other five where homozygous for *LWS*-H alleles, and the 3 for which *SWS2A* was sequenced, were homozygous for the entire H/N-haplotype (Table S4). The only hybrid with significant preference for blue was homozygous for the P-haplotype, but two other

P-haplotype homozygotes had no significant preferences. Heterozygotes and N-haplotype homozygotes had either no significant preference or a preference for red males. Two of the recombinant genotypes (heterozygous at LWS, but N-type homozygous at SWS2A) had a significant preference for red males.

LWS in other red/blue sister species

We sequenced both copies of the complete *LWS* gene in the sympatric *P. igneopinnis* (red- dorsum males) and *P. pundamilia* (blue males; Fig. S3) from Igombe Island. Another sympatric species pair was previously studied: *P.* "red head" (red-chested males) and *P. azurea* (blue males)[3]. All 12 haplotypes of species with red males belonged to the red-shifted H-allele group. Of 12 haplotypes of the blue-male species, 11 belonged to the blue-shifted P-allele group, and one was an M3-allele (table 3).

SUPPLEMENTARY METHODS

Male colour scale

Blue is scored as 0; 1 is a yellow flank but no red, spiny part of dorsal fin is blue; 2 is yellow flank with some red on the flank along the upper lateral line, spiny dorsal fin is blue; 3 is yellow flank with a partially red dorsum upwards from the upper lateral line, but a grey body crest and largely blue spiny dorsal fin; 4 is yellow flank with a completely red dorsum between the upper lateral line and the body crest, red spiny dorsal fin

Microsatellite primers and PCR

DNA was extracted using a QIAGEN® (Basel, Switzerland) extraction robot. The 11 microsatellite primers were developed for these (Ppun 5, Ppun7, Ppun17, Ppun32 (Taylor et al, 2002)) or other haplochromine species (OSU20d, OSU19T, OSU16d, (Wu et al. 1999), TmoM5 (Zardoya et al 1996), Pzeb3, Pzeb5, (van Oppen, 1997)). QUIAGEN Multiplex PCR kit for PCR amplification was used according to the manufacturer's protocol. PCR were carried out in 10 µl reaction volumes containing 5 µl QUIAGEN Multiplex PCR Master mix, 3 µl ddH2O and 1 µl primer mix (2pmol/ µl each primer). The thermocycler profile started with an initial denaturation step at 95 C for 15 min, followed by 30 cycles of 30 sec at 94, 90 sec at TA, 90 sec at 72 and 10 min at 72 C. Denaturated fragments were resolved on an automated DNA sequencer (Beckman coulter, CEQ 8000) with a 400 bp size standard. Genotypes were checked for stutter products, large allele dropout, or null alleles using Micro-Checker v.2.2 (van Oosterhout et al, 2004).

Determination of SWS2A opsin gene sequences

The DNA fragment including the *SWS2A* gene (from exon 1 to exon 5) was amplified by PCR using primers SWS2A_F1 and SWS2A_R1 with genomic DNA (~50 ng) as templates. Amplifications were carried out in the PTC-100 Programmable Thermal Controller (MJ Research). The PCR program consisted of a denaturation step for 3 min at 94 ˚C, followed by 30 cycles, each cycle consisting of 1 min denaturation at 94 ˚C, 1 min annealing at 55 ˚C, 3 min extension at 72˚C. The amplification product was then used as a template to amplify and sequence two overlapping fragments using the primers for upstream (SWS2A_F1, SWS2A_R2) and downstream (SWS2A_F2, SWS2A_R1) regions. These PCR products were purified and determined by direct sequencing with four primers described above. Once determined, the sequences were connected using the program GENETYX-MAC Ver. 10.1. The sequences of primers are as follows: SWS2A_F1 (5'ATGAAGGGTAAACGTGA TATGGA3'), SWS2A_F2 (5'CACCACAA ACAACAAATACAACAA3'), SWS2A_R1 (5'AGGCCCGACTTTGGAAACTTC3'), SWS2A_R2 (5'AAAAGATAATCGTGGTC AAAGGAA3').

Integrated analysis on effects of gradient slope

Terai et al. (2006) found clinal divergent adaptation at the *LWS* locus in *Neochromis greenwoodi* and *Mbipia mbipi* on long and shallow light gradients. Each of these gradients encompassed a series of islands, and can be considered a linear composite of many shorter gradients. Here we analysed these data in two ways. First, we quantified the slope of the light gradient along each series of islands as the difference in orange ratio

between the islands at the gradient ends, measured at 2m depth, divided by the waterway distance between them. We calculated the association between *LWS* and nuptial colour as the Spearman rank correlation between *LWS* genotype and colour, pooling the two populations from the opposite ends of the gradient. We refer to the results of this analysis as the "7 gradient models" (5 depth-mediated gradients + 2 turbidity-mediated gradients).

Second, we quantified the slope of the light gradient for every pair of linearly adjacent islands separately, as long as we had at least 5 (*Mbipia*: 3 gradients) or >5 (*Neochromis*: 5 gradients) genotypes from each island. For this analysis we calculated the correlations between *LWS* genotype and red/blue colour for all 8 pairs of each two populations. All but four islands (the end points of each composite gradient) were used to calculate two gradients (one up, one down). The slopes of the "upward" and "downward" gradients to and from the same island are not strictly independent statistically, but the form of dependence introduces a conservative bias to our analysis (i.e. the two gradients will be influenced in opposite directions by the value of the island that makes one end point of both). We refer to the results of this analysis as the "13 gradient models" (5 depth-mediated gradients + 8 turbidity-mediated gradients).

Whereas the gradient differential did not differ much between the 5 depth-mediated gradients, it was generally smaller on the turbidity-mediated between-island gradients (table 1). To test for effects of gradient slope, while controlling for effects of gradient differential, we analyzed the residual variation in the correlation between colour and *LWS* after regressing against gradient differential. We also calculated multiple regressions with backward elimination of variables, using gradient slope and gradient differential as independent variables.

Table S1. The ten environmental gradients studied previously (Terai et al. 2006)

Island	water clarity (cm Secchi)	light slope	light differential
Mwanza Gulf transect 1	49-225	$4*10^{-6}$	0.21
Mwanza Gulf transect 2	96-225	$4*10^{-6}$	0.14
Marumbi-Python	49-96	$5*10^{-6}$	0.08
Python-Bwiru	96-208	$6*10^{-6}$	0.12
Bwiru-Igombe	208-190	$6*10^{-7}$	0.01
Igombe-Makobe[§]	190-225	$5*10^{-6}$	0.03
Makobe-Namatembi	225-180	$1*10^{-6}$	0.04
Python-Hippo	96-117	$3*10^{-6}$	0.04
Hippo-Igombe	117-190	$5*10^{-6}$	0.09

[§] this gradient was studied in two different species

Table S2. Nucleotide diversity (Pi), allelic richness and alleles in the *LWS* opsin gene

	Marumbi	Luanso	Python	Kissenda	Makobe
Pi	0.00134	0	0.0034	0.00283	0.00269
n alleles	2	1	6	5	5
P. nyererei			HI, HII, PI	HI, HII, PI, PIV	HI, HII, M3I, M3III
P. pundamilia			HI, HII, PI, PIV, PV, PVI	HI, HII, PI	PI
P. "hybrid"	PI, HI	PI		PI	

Table S3. Multilocus Fst (estimate below diagonal; significance above diagonal: + = $p<0.05$) between any two islands and nuptial phenotypes of *Pundamilia*, based on 11 unlinked microsatellite loci. Comparisons between sympatric phenotypes are bold and italic, comparisons between allopatric populations of the same phenotype are bold and in the colour of the male phenotype.

	redMa	BlueMa	redPy	bluePy	redKs	blueKs	redLu	blueLu	redMr	blueMr
redMa		±	+	+	+	+	+	+	+	+
BlueMa	***0.026***		+	+	+	+	+	+	+	+
redPy	**0.007**	0.025		±	+	+	+	+	+	+
bluePy	0.013	**0.035**	***0.014***		+	+	+	+	+	+
redKs	**0.011**	0.030	**0.013**	0.017		±	+	+	+	+
blueKs	0.013	**0.024**	0.018	**0.019**	***0.010***		+	+	+	+
redLu	**0.030**	0.055	**0.037**	0.030	**0.032**	0.033		-	-	+
blueLu	0.038	**0.060**	0.045	**0.038**	0.038	**0.039**	***0.002***		-	+
redMr	**0.024**	0.049	**0.035**	0.020	**0.034**	0.046	**0.003**	0.000		-
blueMr	0.027	**0.058**	0.038	**0.035**	0.042	**0.047**	0.023	**0.029**	***0.000***	

Table S4. Results of HKA test for heterogeneity between regions.

Reported are the *p* values for the comparisons of the ratio of polymorphism within a species to divergence from the sister species between the *LWS* gene regions

focal species	sister species	*LWS* gene regions				
		upstream--1000	upstream-*LWS* gene	upstream-downstream	-1000-downstream	*LWS* gene-downstream
P. nyererei	*P. pundamilia*	0.433	0.817	0.157	0.049*	0.095
P. pundamilia	*P. nyererei*	0.582	0.771	0.950	0.729	0.882

Table S5. Opsin genotype and female mating preference in laboratory preference assays

	female mating preference		
***LWS* and *SWS2A* genotype**	**blue**	**none**	**red**
P. pundamilia (*LWS*-P/P *SWS2A*-P/P)	2	1	0
F2 (LWS-P/P *SWS2A*-P/P)	1	2	0
F2 (LWS-P/H *SWS2A*-P/N)	0	4	0
F2 (LWS-P/H *SWS2A*-N/N)	0	1	2
F2 (LWS-H/H *SWS2A*-N/N)	0	2	0
P. nyererei (*LWS*-H/H *SWS2A*-N/N)	0	0	3

Table S6. Alignments of *LWS* sequences. Only polymorphic sites are shown.

Marumbi Island

reddish males

	66688	
nuc. sites	44822	
	57834	
syn/non-syn	snnnn	allele type
alleleH	CAGTG	
Mr064A1	GTAAT	P
Mr064A2	GTAAT	P
Mr126A1	GTAAT	P
Mr126A2	GTAAT	P
Mr127A1	GTAAT	P
Mr127A2	GTAAT	P
Mr40A1.txt	.....	H
Mr40A2.txt	GTAAT	P
Mr43A1	GTAAT	P
Mr43A2	GTAAT	P

blueish males

Mr108A1	GTAAT	P
Mr108A2	GTAAT	P
Mr42A1.txt	.....	H
Mr42A2.txt	GTAAT	P
Mr69A1	GTAAT	P
Mr69A2	GTAAT	P

males that could not be assigned to either colour

Mr10_PLoSA1	GTAAT	P
Mr10_PLoSA2	GTAAT	P
Mr12_PLoSA1	GTAAT	P
Mr12_PLoSA2	GTAAT	P
Mr40_PLoSA1	GTAAT	P
Mr40_PLoSA2	GTAAT	P
Mr41_PLoSA1	GTAAT	P
Mr41_PLoSA2	GTAAT	P

Luanso Island

reddish males

	66688	
nuc. sites	44822	
	57834	
syn/non-syn	snnnn	allele type
alleleH	CAGTG	
Lu001A1	GTAAT	P
Lu001A2	GTAAT	P
Lu061A1	GTAAT	P
Lu061A2	GTAAT	P
Lu102A1.txt	GTAAT	P
Lu102A2.txt	GTAAT	P
Lu142A1.txt	GTAAT	P
Lu142A2.txt	GTAAT	P
Lu150A1.txt	GTAAT	P
Lu150A2.txt	GTAAT	P
Lu181A1.txt	GTAAT	P
Lu181A2.txt	GTAAT	P
Lu185A1.txt	GTAAT	P
Lu185A2.txt	GTAAT	P
Lu188A1.txt	GTAAT	P
Lu188A2.txt	GTAAT	P
Lu192A1.txt	GTAAT	P
Lu192A2.txt	GTAAT	P
Lu195A1.txt	GTAAT	P
Lu195A2.txt	GTAAT	P
Lu209A1.txt	GTAAT	P
Lu209A2.txt	GTAAT	P
Lu230A1.txt	GTAAT	P
Lu230A2.txt	GTAAT	P

blueish males

Lu069A1.txt	GTAAT	P
Lu069A2.txt	GTAAT	P
Lu110A1.txt	GTAAT	P
Lu110A2.txt	GTAAT	P
Lu149A1.txt	GTAAT	P
Lu149A2.txt	GTAAT	P
Lu157A1.txt	GTAAT	P
Lu157A2.txt	GTAAT	P
Lu208A1.txt	GTAAT	P
Lu208A2.txt	GTAAT	P
Lu229A1.txt	GTAAT	P
Lu229A2.txt	GTAAT	P
Lu236A1.txt	GTAAT	P
Lu236A2.txt	GTAAT	P
Lu238A1.txt	GTAAT	P
Lu238A2.txt	GTAAT	P
Lu82A1.txt	GTAAT	P
Lu82A2.txt	GTAAT	P
Lu89A1.txt	GTAAT	P
Lu89A2.txt	GTAAT	P
Lu95A1.txt	GTAAT	P
Lu95A2.txt	GTAAT	P

males that could not be assigned to either colour

Lu21_PLoSA1	GTAAT	P
Lu21_PLoSA2	GTAAT	P
Lu30_PLoSA1	GTAAT	P
Lu30_PLoSA2	GTAAT	P
Lu7_PLoSA1	GTAAT	P
Lu7_PLoSA2	GTAAT	P
Lu9_PLoSA1	GTAAT	P
Lu9_PLoSA2	GTAAT	P

Kissenda Island

P. nyererei-like

	466688	
nuc. sites	644822	
	557834	
syn/non-syn	ssnnnn	allele type
alleleH	TCAGTG	
Ks111HA1.txt	.....	H
Ks111HA2.txt	.....	H
Ks112HA1.txt	.....	H
Ks112HA2.txt	.....	H
Ks113HA1.txt	.....	H
Ks113HA2.txt	.....	H
Ks114HA1.txt	.....	H
Ks114HA2.txt	.GTAAT	P
Ks115HA1.txt	C.....	H
Ks115HA2.txt	C.....	H
Ks116HA1.txt	.....	H
Ks116HA2.txt	.....	H
Ks117HA1.txt	.GTAAT	P
Ks117HA2.txt	.GTAAT	P
Ks118HA1.txt	.....	H
Ks118HA2.txt	.....	H
Ks119HA1.txt	.....	H
Ks119HA2.txt	.....	H
Ks120HA1.txt	.GTAAT	P
Ks120HA2.txt	.GTAAT	P
Ks121HA1.txt	.....	H
Ks121HA2.txt	.GTAAT	P
Ks122HA1.txt	.....	H
Ks122HA2.txt	CGTAAT	P
Ks123HA1.txt	.GTAAT	P
Ks123HA2.txt	.GTAAT	P
Ks124HA1.txt	.....	H
Ks124HA2.txt	C.....	H
Ks125HA1.txt	.....	H
Ks125HA2.txt	.....	H
Ks126HA1.txt	.....	H
Ks126HA2.txt	.....	H
Ks127HA1.txt	.....	H
Ks127HA2.txt	C.....	H
Ks128HA1.txt	.....	H
Ks128HA2.txt	.....	H
Ks129HA1.txt	.....	H
Ks129HA2.txt	.....	H
Ks18A1.txt	.....	H
Ks18A2.txt	.GTAAT	P
Ks37A1.txt	C.....	H
Ks37A2.txt	C.....	H
Ks3HA1.txt	.....	H
Ks3HA2.txt	.....	H
Ks41A1.txt	C.....	H
Ks41A2.txt	C.....	H
Ks53A1.txt	.GTAAT	P
Ks53A2.txt	.GTAAT	P
Ks8HA1.txt	.....	H
Ks8HA2.txt	.....	H

```
Ks9A1.txt        C.....   H
Ks9A2.txt        CGTAAT   P
```

P. pundamilia-like

```
                 466688
 nuc. sites      644822
                 557834
syn/non-syn      ssnnnn  allele type
alleleH          TCAGTG
Ks11HA1.txt      .GTAAT   P
Ks11HA2.txt      .GTAAT   P
Ks12HA1.txt      .GTAAT   P
Ks12HA2.txt      .GTAAT   P
Ks14HA1.txt      .GTAAT   P
Ks14HA2.txt      .GTAAT   P
Ks15HA1.txt      .GTAAT   P
Ks15HA2.txt      .GTAAT   P
Ks16A1.txt       .GTAAT   P
Ks16A2.txt       .GTAAT   P
Ks17HA1.txt      C.....   H
Ks17HA2.txt      .GTAAT   P
Ks18HA1.txt      .GTAAT   P
Ks18HA2.txt      .GTAAT   P
Ks19HA1.txt      C.....   H
Ks19HA2.txt      .GTAAT   P
Ks1A1.txt        .GTAAT   P
Ks1A2.txt        .GTAAT   P
Ks1_PLoSA1.txt   .GTAAT   P
Ks1_PLoSA2.txt   .GTAAT   P
Ks20HA1.txt      .GTAAT   P
Ks20HA2.txt      .GTAAT   P
Ks21HA1.txt      .GTAAT   P
Ks21HA2.txt      .GTAAT   P
Ks22HA1.txt      .GTAAT   P
Ks22HA2.txt      .GTAAT   P
Ks23HA1.txt      .GTAAT   P
Ks23HA2.txt      .GTAAT   P
Ks24HA1.txt      .GTAAT   P
Ks24HA2.txt      .GTAAT   P
Ks25HA1.txt      .GTAAT   P
Ks25HA2.txt      .GTAAT   P
Ks26HA1.txt      .GTAAT   P
Ks26HA2.txt      .GTAAT   P
Ks27HA1.txt      .GTAAT   P
Ks27HA2.txt      .GTAAT   P
Ks28HA1.txt      .GTAAT   P
Ks28HA2.txt      .GTAAT   P
Ks29HA1.txt      .GTAAT   P
Ks29HA2.txt      .GTAAT   P
Ks29_PLoSA1.txt  .GTAAT   P
Ks29_PLoSA2.txt  .GTAAT   P
Ks2A1.txt        .GTAAT   P
Ks2A2.txt        .GTAAT   P
Ks2HA1.txt       .GTAAT   P
Ks2HA2.txt       .GTAAT   P
Ks2_PLoSA1.txt   .GTAAT   P
Ks2_PLoSA2.txt   .GTAAT   P
Ks31HA1.txt      .GTAAT   P
Ks31HA2.txt      .GTAAT   P
Ks32HA1.txt      .GTAAT   P
Ks32HA2.txt      .GTAAT   P
Ks33HA1.txt      ......   H
Ks33HA2.txt      .GTAAT   P
Ks34HA1.txt      .GTAAT   P
Ks34HA2.txt      .GTAAT   P
Ks40HA1.txt      .GTAAT   P
Ks40HA2.txt      .GTAAT   P
Ks41HA1.txt      .GTAAT   P
Ks41HA2.txt      .GTAAT   P
Ks4HA1.txt       .GTAAT   P
Ks4HA2.txt       .GTAAT   P
Ks5HA1.txt       .GTAAT   P
Ks5HA2.txt       .GTAAT   P
Ks6HA1.txt       .GTAAT   P
Ks6HA2.txt       .GTAAT   P
Ks7HA1.txt       .GTAAT   P
Ks7HA2.txt       .GTAAT   P
Ks8_PLoSA1.txt   .GTAAT   P
Ks8_PLoSA2.txt   .GTAAT   P
```

intermediate phenotype

```
              66688
 nuc. sites   44822
              57834
syn/non-syn   snnnn  allele type
alleleH       CAGTG
Ks1HA1.txt    GTAAT   P
Ks1HA2.txt    GTAAT   P
```

Python Island
P. nyererei-like

	466688	
nuc. sites	644822	
	557834	
syn/non-syn	ssnnnn	al. type
alleleH	TCAGTG	
Nye f 17.2.95 11(A9)A1.txt	C.....	H
Nye f 17.2.95 11(A9)A2.txt	.....	H
Nye f 17.2.95 12(A10)A1.txt	C.....	H
Nye f 17.2.95 12(A10)A2.txt	C.....	H
Nye m Py27.3.95 1(C2)A1.txt	C.....	H
Nye m Py27.3.95 1(C2)A2.txt	C.....	H
Nye m Py27.3.95 2(C4)A1.txt	.....	H
Nye m Py27.3.95 2(C4)A2.txt	.....	H
Nye m Py27.3.95 3(C6)A1.txt	.....	H
Nye m Py27.3.95 3(C6)A2.txt	.....	H
Nye m Py27.3.95 4(C3)A1.txt	C.....	H
Nye m Py27.3.95 4(C3)A2.txt	.....	H
Nye m Py27.3.95 5(C5)A1.txt	C.....	H
Nye m Py27.3.95 5(C5)A2.txt	.....	H
Nye m Py27.3.95(C1)A1.txt	C.....	H
Nye m Py27.3.95(C1)A2.txt	.....	H
Py 17.2.95 8A1.txt	.....	H
Py 17.2.95 8A2.txt	.GTAAT	P
Py004A1.txt	C.....	H
Py004A2.txt	.....	H
Py018A1.txt	.....	H
Py018A2.txt	.GTAAT	P
Py034A1.txt	C.....	H
Py034A2.txt	C.....	H
Py035A1.txt	.....	H
Py035A2.txt	.GTAAT	P
Py036A1.txt	.....	H
Py036A2.txt	.....	H
Py045A1.txt	C.....	H
Py045A2.txt	.....	H
Py053A1.txt	C.....	H
Py053A2.txt	C.....	H
Py054A1.txt	C.....	H
Py054A2.txt	C.....	H
Py055A1.txt	.....	H
Py055A2.txt	.....	H
Py066A1.txt	C.....	H
Py066A2.txt	.....	H
Py069A1.txt	.....	H
Py069A2.txt	.....	H
Py071A1.txt	C.....	H
Py071A2.txt	.....	H
Py076A1.txt	C.....	H
Py076A2.txt	.....	H
Py080A1.txt	C.....	H
Py080A2.txt	C.....	H
Py081A1.txt	C.....	H
Py081A2.txt	.....	H
Py082A1.txt	.....	H
Py082A2.txt	.....	H
Py099A1.txt	C.....	H
Py099A2.txt	C.....	H
Py100A1.txt	.....	H
Py100A2.txt	.....	H
Py101A1.txt	C.....	H
Py101A2.txt	.....	H
Py102A1.txt	.....	H
Py102A2.txt	.....	H
Py103A1.txt	C.....	H
Py103A2.txt	.....	H
Py126 12.1.96(C8)A1.txt	C.....	H
Py126 12.1.96(C8)A2.txt	.....	H
Py15 17.2.95(C1)A1.txt	C.....	H
Py15 17.2.95(C7)A2.txt	.....	H
Py16A1.txt	C.....	H
Py16A2.txt	.....	H
Py17.2.95 1(B1)A1.txt	C.....	H
Py17.2.95 1(B1)A2.txt	.....	H
Py17.2.95 10(B10)A1.txt	.....	H
Py17.2.95 10(B10)A2.txt	.....	H
Py17.2.95 2(B2)A1.txt	C.....	H
Py17.2.95 2(B2)A2.txt	.....	H
Py17.2.95 3(B3)A1.txt	C.....	H
Py17.2.95 3(B3)A2.txt	C.....	H
Py17.2.95 4(B4)A1.txt	C.....	H
Py17.2.95 4(B4)A2.txt	.....	H
Py17.2.95 5(B5)A1.txt	C.....	H
Py17.2.95 5(B5)A2.txt	.....	H
Py17.2.95 6(B6)A1.txt	.....	H
Py17.2.95 6(B6)A2.txt	.....	H
Py17.2.95 7(B7)A1.txt	C.....	H
Py17.2.95 8(B7)A2.txt	.....	H
Py17.2.95 9(B9)A1.txt	.....	H
Py17.2.95 9(B9)A2.txt	.....	H
Py1A1.txt	C.....	H
Py1A2.txt	.....	H
Py24A1.txt	C.....	H
Py24A2.txt	.....	H
Py25A1.txt	C.....	H
Py25A2.txt	.....	H
Py26A1.txt	.GTAAT	P
Py26A2.txt	C.....	H
Py2A1.txt	C.....	H
Py2A2.txt	C.....	H
Py4A1.txt	C.....	H
Py4A2.txt	.....	H
Py55A1.txt	.....	H
Py55A2.txt	.....	H

P. pundamilia-like

	114455666888	
nuc. sites	451623448224	
	170595578344	
syn/non-syn	snnsnnsnnnnn	allele type
alleleH	AGCTGGCAGTGA	
Py001A1.txt	......GTAAT.	P
Py001A2.txt	.A......AATG	other
Py003A1.txt	......GTAAT.	P
Py003A2.txt	......GTAAT.	P
Py005A1.txt	......GTAAT.	P
Py005A2.txt	......GTAAT.	P
Py010A1.txt	......GTAAT.	P
Py010A2.txt	......GTAAT.	P
Py012A1.txt	......GTAAT.	P
Py012A2.txt	......GTAAT.	P
Py013A1.txt	......GTAAT.	P
Py013A2.txt	......GTAAT.	P
Py019A1.txt	......GTAAT.	P
Py019A2.txt	......GTAAT.	P
Py065A1.txt	......GTAAT.	P
Py065A2.txt	......GTAAT.	P
Py067A1.txt	G.....GTAAT.	P
Py067A2.txt	...C........	H
Py070A1.txt	......GTAAT.	P
Py070A2.txt	......GTAAT.	P
Py073A1.txt	......GTAAT.	P
Py073A2.txt	....	H
Py074A1.txt	......GTAAT.	P
Py074A2.txt	......GTAAT.	P
Py077A1.txt	......GTAAT.	P
Py077A2.txt	......GTAAT.	P
Py078A1.txt	......GTAAT.	P
Py078A2.txt	......GTAAT.	P
Py083A1.txt	....TT...AT.	other
Py083A2.txt	......GTAAT.	P
Py093A1.txt	......GTAAT.	P
Py093A2.txt	......GTAAT.	P
Py094A1.txt	......GTAAT.	P
Py094A2.txt	......GTAAT.	P
Py096A1.txt	......GTAAT.	P
Py096A2.txt	..G...GTAAT.	P
Py106A1.txt	......GTAAT.	P
Py106A2.txt	......GTAAT.	P
Py114A1.txt	......GTAAT.	P
Py114A2.txt	......GTAAT.	P
Py115A1.txt	......GTAAT.	P
Py115A2.txt	......GTAAT.	P
Py117_PLoSA1.txt	......GTAA..	other
Py117_PLoSA2.txt	......G..A..	other
Py120A1.txt	......GTAAT.	P
Py120A2.txt	......GTAAT.	P
Py121A1.txt	......GTAAT.	P
Py121A2.txt	......GTAAT.	P
Py122_PLoSA1.txt	......GTAAT.	P
Py122_PLoSA2.txt	......GTAAT.	P
Py123_PLoSA1.txt	......GTAAT.	P
Py123_PLoSA2.txt	....TTG..AT.	other
Py136A1.txt	....	H
Py136A2.txt	...C..GTAAT.	P
Py142A1.txt	......GTAAT.	P
Py142A2.txt	......GTAAT.	P
Py167A1.txt	......GTAAT.	P
Py167A2.txt	......GTAAT.	P
Py205A1.txt	......GTAAT.	P
Py205A2.txt	......GTAAT.	P
Py206A1.txt	......GTAAT.	P
Py206A2.txt	......GTAAT.	P
Py207AA1.txt	......GTAAT.	P
Py207AA2.txt	......GTAAT.	P
Py207BA1.txt	......GTAAT.	P
Py207BA2.txt	......GTAAT.	P
Py208AA1.txt	......GTAAT.	P
Py208AA2.txt	......GTAAT.	P
Py208BA1.txt	......GTAAT.	P
Py208BA2.txt	......GTAAT.	P
Py20_PLoSA1.txt	......GTAAT.	P
Py20_PLoSA2.txt	......GTAAT.	P
Py58_PLoSA1.txt	......GTAAT.	P
Py58_PLoSA2.txt	......GTAAT.	P
Py59_PLoSA1.txt	......GTAAT.	P
Py59_PLoSA2.txt	......GTAAT.	P
Py63_PLoSA1.txt	......GTAAT.	P
Py63_PLoSA2.txt	......GTAAT.	P
Py64_PLoSA1.txt	......GTAAT.	P
Py64_PLoSA2.txt	......GTAAT.	P
Py65_PLoSA1.txt	......GTAAT.	P
Py65_PLoSA2.txt	......GTAAT.	P

Makobe Island

P. nyererei

	4688	
nuc. sites	6822	
	5834	
syn/non-syn	snnn	allele type
alleleH	TGTG	
Ma101HA1.txt	C...	H
Ma101HA2.txt	C...	H
Ma102HA1.txt	C...	H
Ma102HA2.txt	C...	H
Ma10HA1.txt	C...	H
Ma10HA2.txt	C...	H
Ma11HA1.txt	....	H
Ma11HA2.txt	....	H
Ma120HA1.txt	....	H
Ma120HA2.txt	C...	H
Ma121HA1.txt	....	H
Ma121HA2.txt	C...	H
Ma122HA1.txt	....	H
Ma122HA2.txt	....	H
Ma123HA1.txt	....	H
Ma123HA2.txt	....	H
Ma124HA1.txt	....	H
Ma124HA2.txt	....	H
Ma125HA1.txt	....	H
Ma125HA2.txt	C...	H
Ma126HA1.txt	C...	H
Ma126HA2.txt	C...	H
Ma127HA1.txt	....	H
Ma127HA2.txt	C...	H
Ma12HA1.txt	....	H
Ma12HA2.txt	C...	H
Ma17A1.txt	C...	H
Ma17A2.txt	C...	H
Ma196HA1.txt	....	H
Ma196HA2.txt	C...	H
Ma197HA1.txt	....	H
Ma197HA2.txt	C...	H
Ma1HA1.txt	C...	H
Ma1HA2.txt	C...	H
Ma202HA1.txt	....	H
Ma202HA2.txt	C...	H
Ma20A1.txt	....	H
Ma20A2.txt	....	H
Ma225HA1.txt	C...	H
Ma225HA2.txt	C...	H
Ma26A1.txt	....	H
Ma26A2.txt	C...	H
Ma27A1.txt	....	H
Ma27A2.txt	C...	H
Ma28A1.txt	....	H
Ma28A2.txt	....	H
Ma29A1.txt	....	H
Ma29A2.txt	....	H
Ma2HA1.txt	....	H
Ma2HA2.txt	C...	H
Ma341A1.txt	....	H
Ma341A2.txt	C...	H
Ma342A1.txt	....	H
Ma342A2.txt	C...	H
Ma343A1.txt	....	H
Ma343A2.txt	C...	H
Ma344A1.txt	C...	H
Ma344A2.txt	C...	H
Ma36A1.txt	....	H
Ma36A2.txt	....	H
Ma37A1.txt	....	H
Ma37A2.txt	....	H
Ma38A1.txt	....	H
Ma38A2.txt	C...	H
Ma39A1.txt	....	H
Ma39A2.txt	....	H
Ma3HA1.txt	....	H
Ma3HA2.txt	....	H
Ma4HA1.txt	....	H
Ma4HA2.txt	CAAT	M3
Ma53HA1.txt	....	H
Ma53HA2.txt	C...	H
Ma54HA1.txt	.AAT	M3
Ma54HA2.txt	C...	H
Ma55HA1.txt	....	H
Ma55HA2.txt	C...	H
Ma56HA1.txt	....	H
Ma56HA2.txt	C...	H
Ma57HA1.txt	....	H
Ma57HA2.txt	C...	H
Ma58HA1.txt	C...	H
Ma58HA2.txt	C...	H
Ma59HA1.txt	C...	H
Ma59HA2.txt	C...	H
Ma5HA1.txt	C...	H
Ma5HA2.txt	.AAT	M3
Ma60HA1.txt	C...	H
Ma60HA2.txt	C...	H
Ma61HA1.txt	C...	H
Ma61HA2.txt	CAAT	M3
Ma62HA1.txt	C...	H
Ma62HA2.txt	C...	H
Ma64HA1.txt	....	H
Ma64HA2.txt	....	H
Ma6HA1.txt	....	H
Ma6HA2.txt	C...	H
Ma7HA1.txt	C...	H
Ma7HA2.txt	C...	H
Ma8HA1.txt	C...	H
Ma8HA2.txt	C...	H
Ma91HA1.txt	C...	H
Ma91HA2.txt	C...	H
Ma92HA1.txt	....	H
Ma92HA2.txt	....	H
Ma93HA1.txt	....	H
Ma93HA2.txt	....	H
Ma94HA1.txt	....	H
Ma94HA2.txt	C...	H

```
Ma95HA1.txt      C...   H
Ma95HA2.txt      C...   H
Ma9HA1.txt       C...   H
Ma9HA2.txt       C...   H
```

P. pundamilia

```
                 6667888
nuc. sites       4489224
                 5784344
 syn/non-syn     snnnnnn  allele type
alleleH          CAGGTGA
Ma212HA1.txt     ..AAATG  other
Ma212HA2.txt     GTA.AT.   P
Ma214HA1         GTA.AT.   P
Ma214HA2         GTA.AT.   P
Ma215HA1         GTA.AT.   P
Ma215HA2         GTA.AT.   P
Ma216HA1         GTA.AT.   P
Ma216HA2         GTA.AT.   P
Ma217HA1         GTA.AT.   P
Ma217HA2         GTA.AT.   P
Ma3411_PLoSA1    GTA.AT.   P
Ma3411_PLoSA2    GTA.AT.   P
Ma341_PLoSA1     GTA.AT.   P
Ma341_PLoSA2     GTA.AT.   P
Ma40HA1.txt      ..AAATG  other
Ma40HA2.txt      GTA.AT.   P
Ma5893_PLoSA1    GTA.AT.   P
Ma5893_PLoSA2    GTA.AT.   P
Ma5_PLoSA1       GTA.AT.   P
Ma5_PLoSA2       GTA.AT.   P
Ma66HA1          GTA.AT.   P
Ma66HA2          GTA.AT.   P
Ma67HA1          GTA.AT.   P
Ma67HA2          GTA.AT.   P
Ma68HA1          GTA.AT.   P
Ma68HA2          GTA.AT.   P
Ma75HA1          GTA.AT.   P
Ma75HA2          GTA.AT.   P
Ma76HA1          GTA.AT.   P
Ma76HA2          GTA.AT.   P
Ma92_PLoSA1      GTA.AT.   P
Ma92_PLoSA2      GTA.AT.   P
```

Table S7. Alignments of *SWS2A* sequences. Only polymorphic sites are shown.

Kissenda Island

P. pundamilia-like

```
                 11112
               2302670
               2515190
               8483311
               ssisiis
Ks112H.txt     CCGTTGG
Ks111H.txt     YYRYK.R
Ks3H.txt       ..R....
Ks113H.txt     YYAN.R.
Ks114H.txt     YY.....
Ks115H.txt     ...YG.A
Ks116H.txt     ..ACG.A
Ks117H.txt     ..AC...
Ks118H.txt     ....K.R
Ks119H.txt     ..RY.R.
Ks120H.txt     ..AC.R.
Ks121H.txt     ..R....
Ks122H.txt     YYAC...
Ks123H.txt     ..AC.A.
Ks124H.txt     YYRYK.R
Ks125H.txt     ..RY.R.
Ks126H.txt     .YRYK.R
Ks8H.txt       ....K.R
Ks9.txt        ..RY...
Ks18.txt       ..AC...
```

P. nyererei-like

```
Ks12H.txt      ..AC...
Ks14H.txt      YYAC...
Ks15H.txt      ..AC...
Ks16.txt       ..AC.A.
Ks17H.txt      ..RYK.R
Ks18H.txt      YYRYK.R
Ks19H.txt      ..RY.R.
Ks1H.txt       ..AC.R.
Ks2.txt        ..AY...
Ks20H.txt      ..AC.R.
Ks21H.txt      YYAC...
Ks22H.txt      ..A....
Ks23H.txt      ..AC.R.
Ks24H.txt      ..AY.R.
Ks2H.txt       ..AC.R.
Ks4H.txt       ..AY...
Ks5H.txt       ..AC.R.
Ks6H.txt       ..AC...
Ks7H.txt       YYAC...
Ks1.txt        ..AY...
```

Makobe Island

P. pundamilia

```
                 11111112
               2301246770
               2510511990
               8481383151
               SSiiSSiiiS
Ma214H.txt     CCACCATATG
Ma212H.txt     YY.....R..
Ma215H.txt     ..........
Ma216H.txt     ..........
Ma217H.txt     ..........
Ma3411.txt     ..........
Ma341.txt      ..........
Ma40H.txt      .......R..
Ma5893.txt     ..........
Ma5.txt        ..........
Ma66H.txt      ..........
Ma67H.txt      ..........
Ma68H.txt      ..........
Ma75H.txt      ..........
Ma76H.txt      ..........
Ma92.txt       ..........
```

P. nyererei

```
Ma101H.txt     ..R.T.GG.A
Ma102H.txt     YY.S...R..
Ma10H.txt      ..R.Y.KR.R
Ma11H.txt      ..G.T.KG..
Ma120H.txt     ..RSY.KR.R
Ma121H.txt     YYR.Y.KG.R
Ma122H.txt     YYR.Y.KGKN
Ma123H.txt     ..G.T.KG.R
Ma124H.txt     ..G.T.KGKR
Ma125H.txt     ..G.T.KG.R
Ma127H.txt     YYR.Y.KG.R
Ma12H.txt      YY.....R..
Ma17.txt       ..R.Y.KR.R
Ma196H.txt     YT.....G..
Ma1H.txt       ..RSYR.R..
Ma2H.txt       ..R...KG.R
Ma126H.txt     YYR.Y.KG.R
```

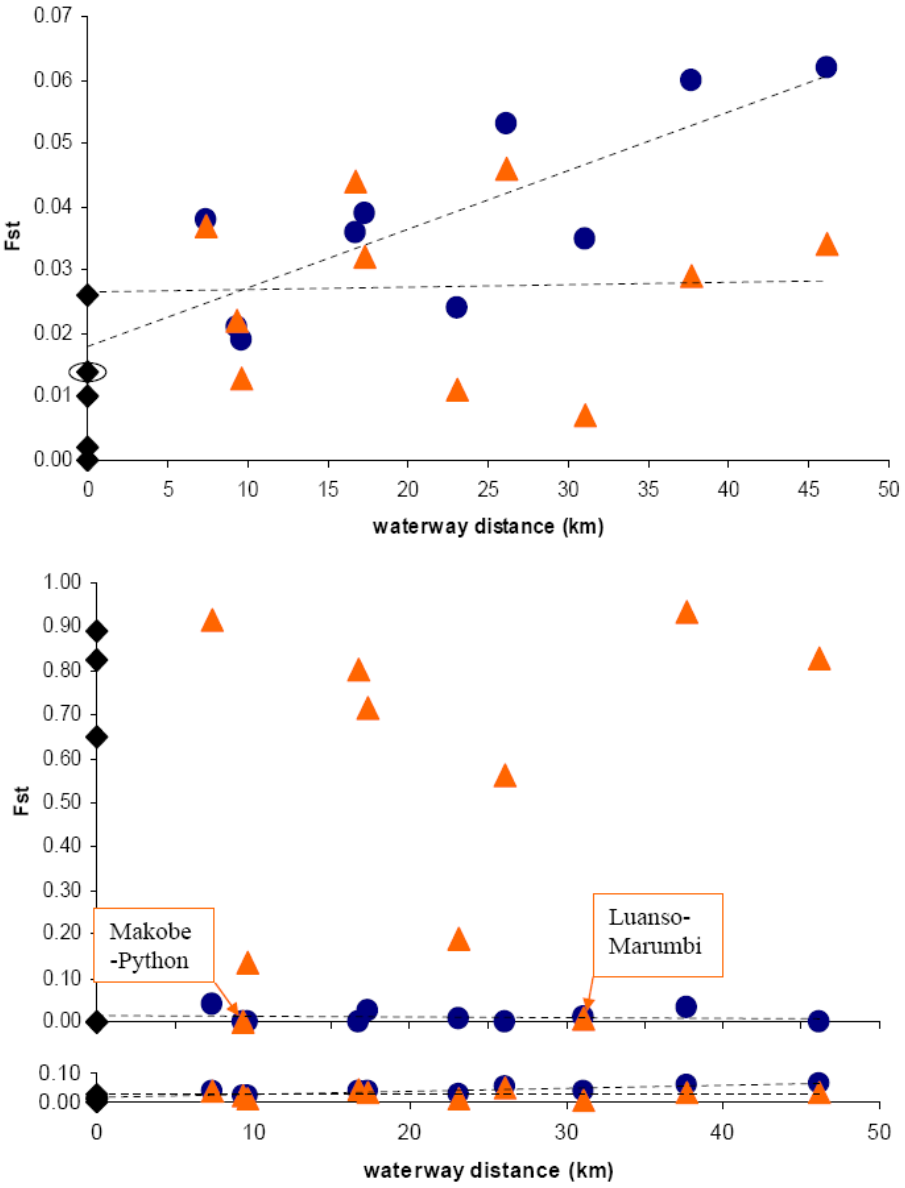

Figure S1. Contrasting patterns of population genetic divergence at neutral and adaptive loci.
A: Pairwise Fst calculated from allele frequencies at 11 microsatellite loci for all pairs of geographically separated populations of the red (*P. nyererei* and *P. nyererei*-like) phenotype (orange triangles), all pairs of geographically separated populations of the blue (*P. pundamilia* and *P. pundamilia*-like) phenotype (blue circles), and all sympatric pairs of blue and red phenotypes (black diamonds) plotted against waterway distance between localities. Circle around black diamond identifies the pair of sympatric phenotypes studied for Figure S2. Significant isolation-by-distance was found among populations of the blue phenotype (Mantel test $p<0.05$), but not among populations of the red phenotype ($p>0.05$; Fig. S1A).
B: The same as (A) but for allele frequencies at the *LWS* locus and at the 11 microsatellite loci (bottom) drawn to the same Y-axis scale. The two red-red comparisons between populations that have similar extents of introgression from the sympatric blue phenotypes are indicated (little introgression: Makobe vs Python; much introgression: Marumbi vs Luanso; Fig. 2c). All other red-red comparisons are between populations with very different extent of introgression, which explains the high Fst. For the three populations with significant LWS FST between red and blue phenotypes, all LWS FST are 1-2 orders of magnitude larger than microsatellite FST (table 2). In contrast, LWS FST between allopatric populations of the same phenotype are no larger than microsatellite FST, except between populations of the red phenotype that differed in the extent of introgression from sympatric blue phenotypes.

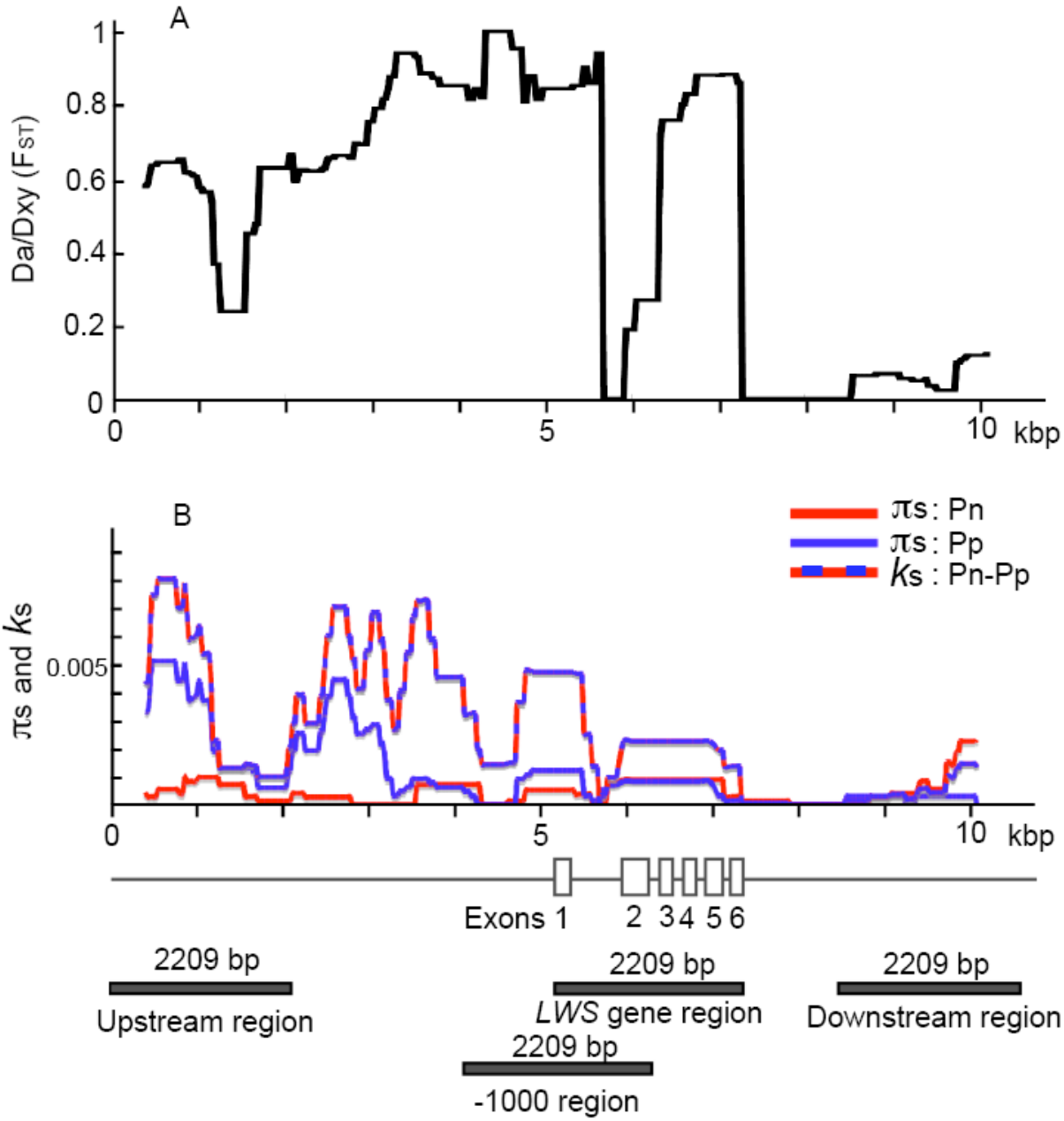

Figure S2. Detection of Selection Pressure on the *LWS* Gene. The genome structure of the *LWS* gene and its flanking region and (top) sliding-window analysis of Fst between blue and red male phenotypes from Python Island, and (bottom) sliding-window analysis of silent polymorphism (Π_s) in *P. pundamilia*-like (blue line) and *P. nyererei*-like (red line) males, and silent divergence (k_s) between them (red and blue line). Π_s and k_s were calculated for segments of 700 bp in 25-bp intervals. The blocks under the genome structure of *LWS* indicate the four regions used in HKA tests (Table S3). The *LWS* gene region is defined as the sequence between initiation codon and stop codon (2205 bp) of *LWS*. The up- and downstream regions are defined as the 5' and 3' ends of sequences of the same length as the *LWS* gene region. The -1000 region is defined as the sequence between 1000 bp upstream from initiation codon and stop codon.

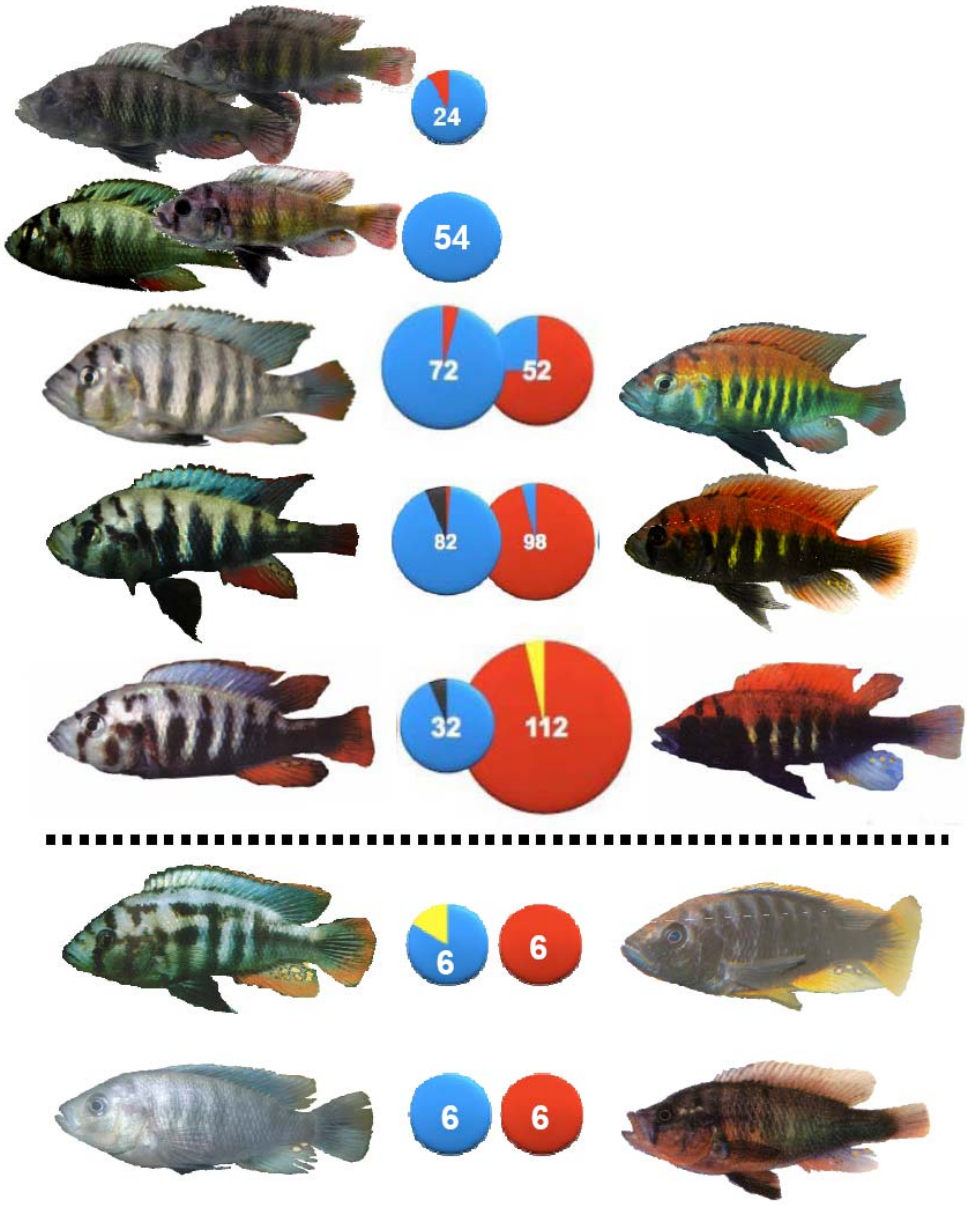

Figure S3. The male nuptial phenotypes and *LWS* allele group frequencies of all studied sympatric species and population pairs of *Pundamilia.* The hatched line separates the *P. pundamilia/P. nyererei* complex from other species pairs. Left column: blue types, right column: red dorsum and red chest types. From top: representative males of blue and reddish phenotypes in the variable but predominantly blue *Pundamilia* population from Marumbi Island; representative males of blue and reddish phenotypes in the variable but predominantly blue *Pundamilia* population from Luanso Island; representative males of the blue and red incipient species from Kissenda Island; representative males of the blue and red incipient species from Python Island; representative males of *P. pundamilia* and *P. nyererei* from Makobe Island; representative males of *P. pundamilia* and *P. igneopinnis* from Igombe Island; representative males of *P. azurea* and *P.* sp. "red head" (red chest type) from Ruti and Zue Island respectively, these species are sympatric and syntopic at Mabibi Islands, situated between Zue and Ruti Islands. Numbers in the allele frequency pies are the number of haplotypes sequenced.

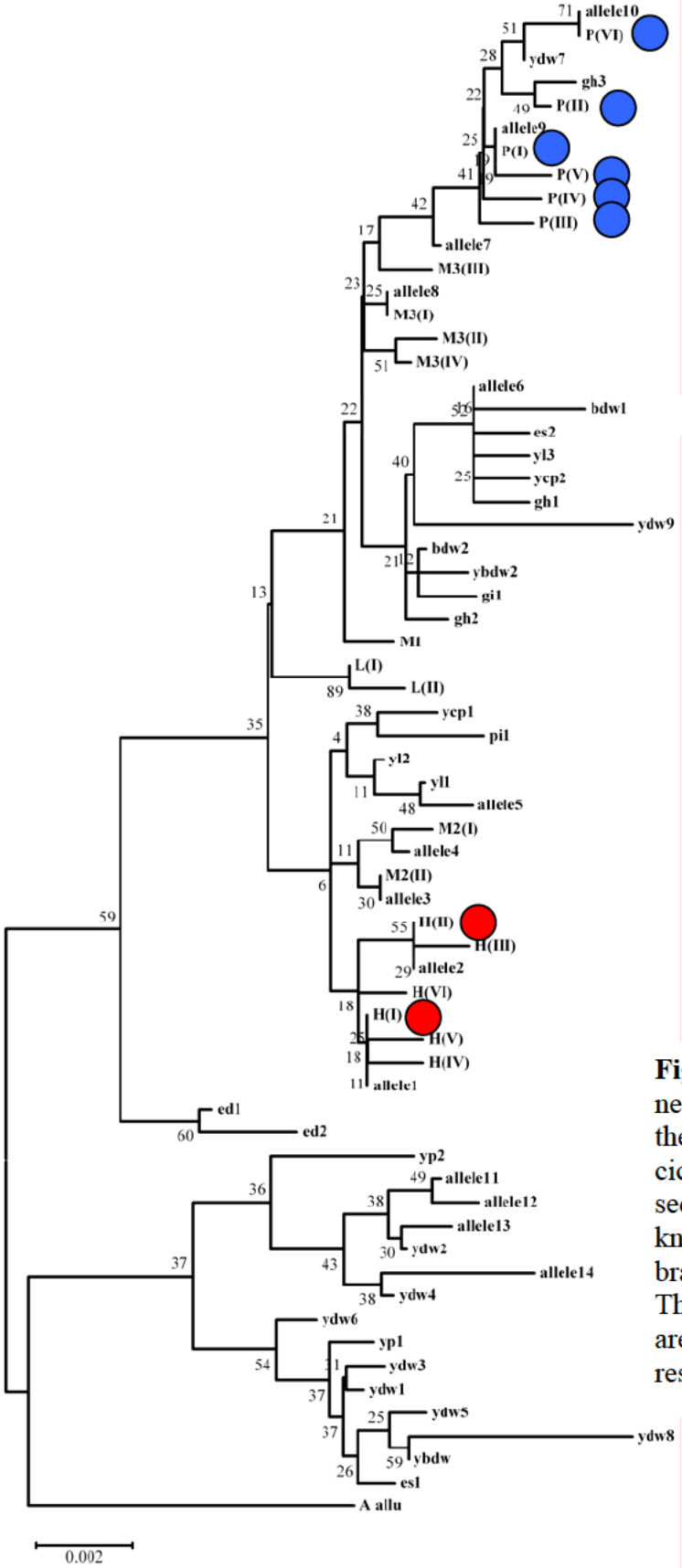

Figure S4a. Unrooted neighbourjoining genealogy of the *LWS* alleles of Lake Victoria cichlid fish built from 1098 sequences, including all currently known alleles. Numbers on branches are % bootstrap support. The *Pundamilia* H and P alleles are highlighted red and blue respectively.

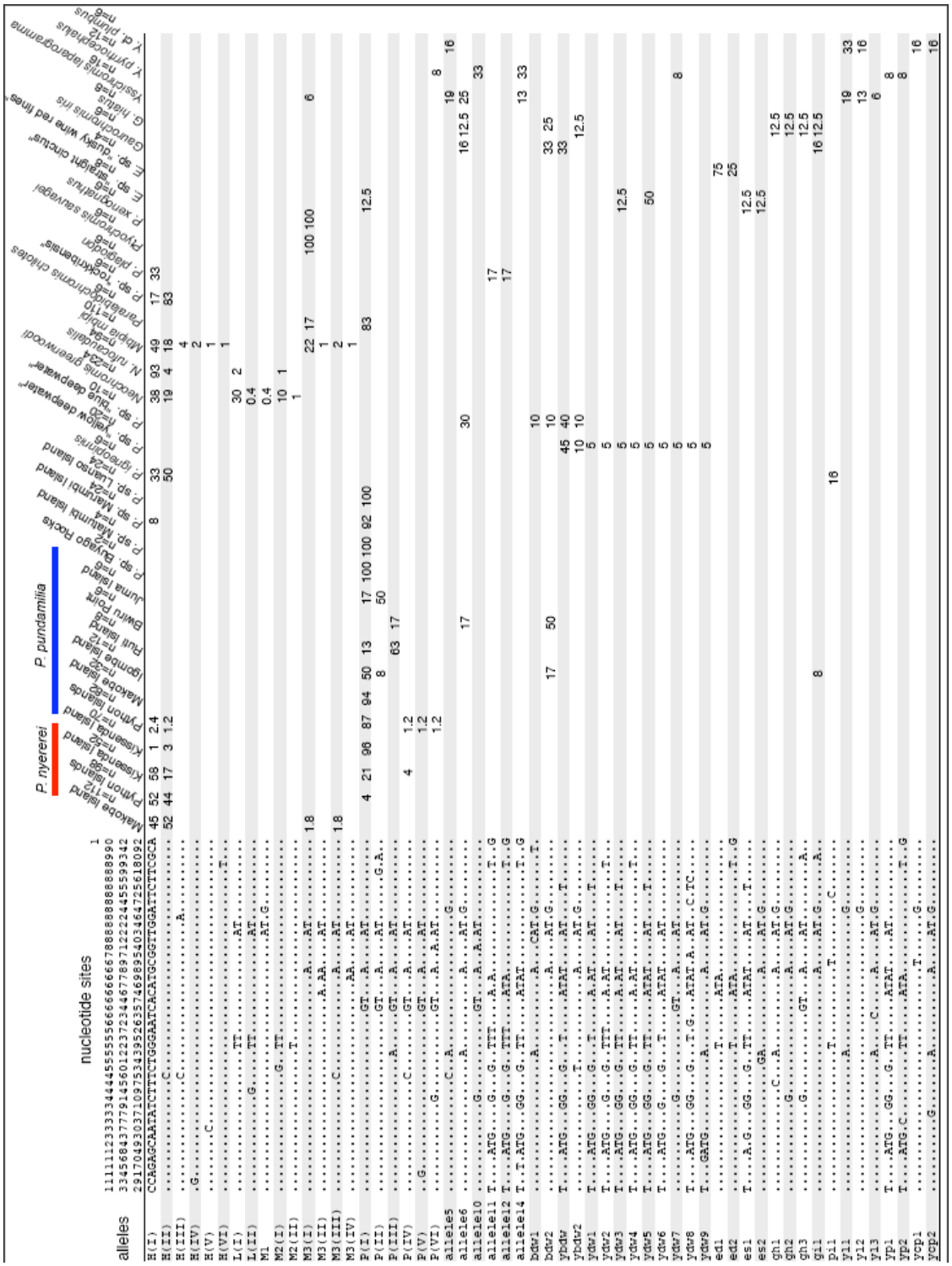

Figure S4b. An alignment of all informative sites of alleles and the frequencies (%) of each allele in *Pundamilia* and in various other species of Lake Victoria cichlids. The sample size is reported for each species under the species name and refers to the number of haplotypes (total n = 1098). Dots indicate where nucleotides are identical with those in the top line. All *Pundamilia* H alleles and the most frequent *Pundamilia* P allele are shared with several only distantly related species. Four P alleles are so far known only from *Pundamilia*. Importantly, the two different *Pundamilia* H alleles are also the most common H alleles in other unrelated species. Either red *Pundamilia* populations acquired these alleles once or multiple times from other species through introgressive hybridization, or the shared ancestor of red and blue *Pundamilia* possessed all the P and H alleles, which would then have been sorted by selection. In either scenario, the major *LWS* alleles must be older than the red and blue *Pundamilia* species.

Additional References for Supplementary Information

Taylor MI, Meardon F, Turner G, et al. (2002) Characterization of tetranucleotide microsatellite loci in a Lake Victorian, haplochromine cichlid fish: a *Pundamilia pundamilia* x *Pundamilia nyererei* hybrid. *Molecular Ecology Notes* 2, 443-445.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4, 535-538.

Van Oppen MJH, Turner GF, Rico C, et al. (1997) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 264, 1803-1812.

Wu L, Kaufman L, Fuerst PA (1999) Isolation of microsatellite markers in Astatoreochromis alluaudi and their cross-species amplifications in other African cichlids. *Molecular Ecology* 8, 895-897.

Zardoya R, Vollmer DM, Craddock C, et al. (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proceedings of the Royal Society of London Series B-Biological Sciences* 263, 1589-1598.

Chapter 5

Divergent selection and phenotypic plasticity during incipient speciation in Lake Victoria cichlid fish

Divergent selection and phenotypic plasticity during incipient speciation in Lake Victoria cichlid fish

Isabel S. Magalhaes, Salome Mwaiko, Maria Victoria Schneider, Ole Seehausen

(*Journal of Evolutionary Biology*, in press)

Divergent selection acting on several different traits that cause multidimensional shifts are supposed to promote speciation, but the outcome of this process is highly dependent on the balance between the strength of selection versus gene flow. We sampled a pair of sister species of Lake Victoria cichlids at a location where they hybridize and tested the hypothesis that divergent selection acting on several traits can maintain phenotypic differentiation despite of gene flow. To explore the possible role of selection we tested for correlations between phenotypes and environment and compared phenotypic divergence (P_{ST}) with that based on neutral markers (F_{ST}). We found indications for disruptive selection acting on male breeding colour and divergent selection acting on several morphological traits. By performing common garden experiments we also separated the environmental and the heritable components of divergence and found evidence for phenotypic plasticity in some morphological traits contributing to species differences.

Key words: cichlids, divergent selection, P_{ST}, Q_{ST}, F_{ST}, phenotypic plasticity

Introduction

Understanding how biological diversity is generated and maintained is a central issue to both evolutionary and conservation biology. Evolutionary forces shaping genetic variation are mutation, migration, genetic drift and selection. Theory predicts that even low levels of gene flow between populations will prevent neutral divergence (Slatkin, 1987) but that divergent or disruptive selection can generate and maintain adaptive divergence despite gene flow (Dobzhansky, 1970; Dieckmann & Doebeli, 1999; Higashi et al., 1999; Kondrashov & Kondrashov, 1999; Drossel & McKane, 2000; Gavrilets, 2004). An increasing number of empirical studies suggest that divergent selection can indeed lead to speciation with gene flow (Schluter, 1996a, b; Lu & Bernatchez, 1999; Rundle et al., 2000; Naisbit et al., 2001; Boughman, 2001; Schliewen et al., 2001; Emelianov et al., 2004; Bolnick, 2004; Rogers & Bernatchez, 2007; Grant & Grant, 2008). Under which conditions and how commonly divergent selection despite gene flow leads to complete speciation is a debated issue though (Nosil & Harmon, in press). Very recently it has been suggested that two different factors may explain variation in the completeness of ecological speciation: (1) the intensity of divergent selection on one trait or (2) the multifariousness of selection, i.e. the number of genetically independent traits that are under selection (Nosil & Sandoval, 2008; Nosil et al., in press).

Utilizing this theoretical framework we studied a pair of sister species of Lake Victoria cichlids at a location where they hybridize to investigate the potential role of disruptive or divergent selection in the maintenance of multivariate phenotypic differentiation despite gene flow. The sister species *Pundamilia pundamilia* and *Pundamilia nyererei* are widely and sympatrically distributed at rocky islands in Lake Victoria (Seehausen & Van Alphen, 1999). The species differ primarily in male breeding colouration, which is used in their identification. Males of both species have blackish underparts and blackish vertical bars on the flanks but *Pundamilia nyererei* males have crimson red dorsum and dorsal fin and yellow flanks, whereas *Pundamilia pundamilia* males have blue-grey dorsum, a blue dorsal fin and blue-grey flanks (Seehausen, 1997). Females of both species are cryptically coloured and either more yellowish (*P. nyererei*) or more greyish (*P. pundamilia*). At places in the lake where the water transparency is high and the light environment heterogeneous, these phenotypes are ecologically and genetically differentiated sister species, divergent in several dimensions: male breeding colour (Seehausen et al., 1997; Seehausen, 2008), opsin genes that affect colour vision (Carleton et al., 2005), depth distribution (Seehausen, 1997; Seehausen et al. 1998), photic environment (Maan et al., 2006) and feeding ecology (Bouton et al., 1997). However, at other places where the water transparency is low intermediate colour phenotypes are found or even dominate the population (Seehausen et al., 1997; Seehausen, 2008). At these places the two phenotypes are merely extremes in a single panmictic population. A recent study analysed these phenotypes at five different islands along five light gradients of different steepness (Seehausen et al., 2008). The study found that in populations living on moderately shallow to moderately steep light gradients there were strong correlations between ambient light colour, male nuptial colour, visual pigments and female mating preference for male nuptial colour, which provided evidence for speciation through sensory drive. The correlations become weaker and eventually disappear altogether as the light gradient becomes steeper. Along this turbidity transect the two species go from being one species with phenotypic variation to two incipient

species, and finally two sister species, in what has been called a speciation or a "speciation in reverse" transect (Seehausen et al., 1997; Seehausen et al., 2008).

Here we analysed in detail the incipient species from the speciation transect that showed the weakest yet significant bimodality in the distribution of male nuptial colouration and the lowest significant differentiation at neutral loci (Seehausen et al., 2008) and investigated traits not studied so far in this population. We investigated neutral genetic differentiation for two sampling years, analysed morphological and colour variation for one year and investigated if differences found among wild fish are heritable or phenotypically induced. Specifically, we ask the following questions: (1) Is divergence in male colouration associated with differences in water depth? (2) Is there any evidence for differences in morphology along water depth and between colour phenotypes? (3) Is divergence in quantitative phenotypic traits larger than expected from divergence at neutral loci and can we invoke divergent or disruptive selection? (4) Are the phenotypic differences heritable or are they environmentally induced? Finding eco-morphological divergence between the two colour phenotype classes despite low genetic differentiation would support the hypothesis that divergent selection acting on several independent traits is sufficiently strong to maintain differences in the presence of gene flow.

We first investigated associations between water depth, male colouration and morphology. Then we compared differentiation in phenotypic traits as defined by P_{ST}, the phenotypic surrogate for Q_{ST} (Spitze, 1993), to differentiation at neutral marker genes (F_{ST}; Wright, 1951) between groups of male defined by nuptial colour or water depth. Finally we performed a common garden experiment to ask (1) if phenotypic differentiation could be explained by environmentally induced plasticity and (2) if phenotypic variation could be explained by segregation of alleles with complete dominance at a single locus.

Material and Methods

Study area and sampling

We collected samples at Kissenda Island, located in the western part of the Mwanza Gulf, in southern Lake Victoria, Tanzania. The water around the island is turbid, with an average secchi depth of 80 cm. With increasing water depth, the ambient light spectrum around this island shifts towards longer wave lengths, resulting in an increasingly red-shifted environment (Seehausen et al., 2008). Phenotypically distinctly differentiated red and blue phenotypes are common around this island, with intermediate forms occurring at lower frequencies.

Using hook and line and gill nets, we randomly collected males in 2 years, 2000 and 2005. We took fin clips for genetic analysis in both years and collected whole fish, took photographs and recorded water depth at which individuals were collected in 2005. We collected 59 males in 2000 and 140 males in 2005. Water depth at which individuals were collected was recorded by measuring the length of the angling line (vertically down into rocky crevices) or by recording the depth of the sinker line of gill nets. Immediately after capture males were photographed in a standard photo cuvette for colour analysis. Fin clips were taken for genetic analysis and after sedating the fish on MS222, they were fixed in 4% formaldehyde solution. After transfer to Switzerland the fish were transferred in 3 steps to 30%, 50% and 70% ethanol for storage.

In 2005 we also collected live fish, which were transferred to aquaria in Switzerland for the common garden experiment.

Phenotypic analysis

Colour analysis

For photography, males were placed in a perpex cuvette with water immediately upon capture, and squeezed gently between a grey PVC sheet and the front window. A total of 118 wild males were photographed using a digital camera. Individuals with no picture taken and those that did not show full nuptial colouration were excluded from further analysis. Lab bred individuals were photographed in the holding tanks in full nuptial colouration. Once the most colourful, dominant male was photographed, it was taken out of the holding tank so that another male would become dominant and colour up fully. Males were assigned to a colour class on a 5-point scale (Fig. 1a) (Dijkstra et al., 2006; van der Sluijs et al., 2008). The photos were scored independently by two different observers. The correlation of the phenotype scoring of the two observers was calculated and the average of both scores was used for each fish in further analysis.

Morphological analysis

Using a digital calliper, we took thirteen standard morphometric distances that are powerful in detecting eco-morphological variation among haplochromine cichlids (Barel et al., 1977) (Fig. 1b). We used the univariate residual method to adjust each morphometric distance for size heterogeneity among individuals (Fleming & Gross, 1994): raw data were log transformed and used to establish regression lines describing the relationship between each character and standard length. Standardized residuals from these regressions were used as variables for further analyses. Multivariate morphological variation among individuals was reduced to Principal Components using SPSS 14.0 (SPSS Inc., Chicago, Illinois, USA). The relative percentage of the variance in each variable that loaded to each PC axis was estimated by first squaring the component loadings, then multiplying them by the percentage of variance explained by an axis and finally dividing this value by the sum of the percentages of variance in a trait that load to the first three PC axes.

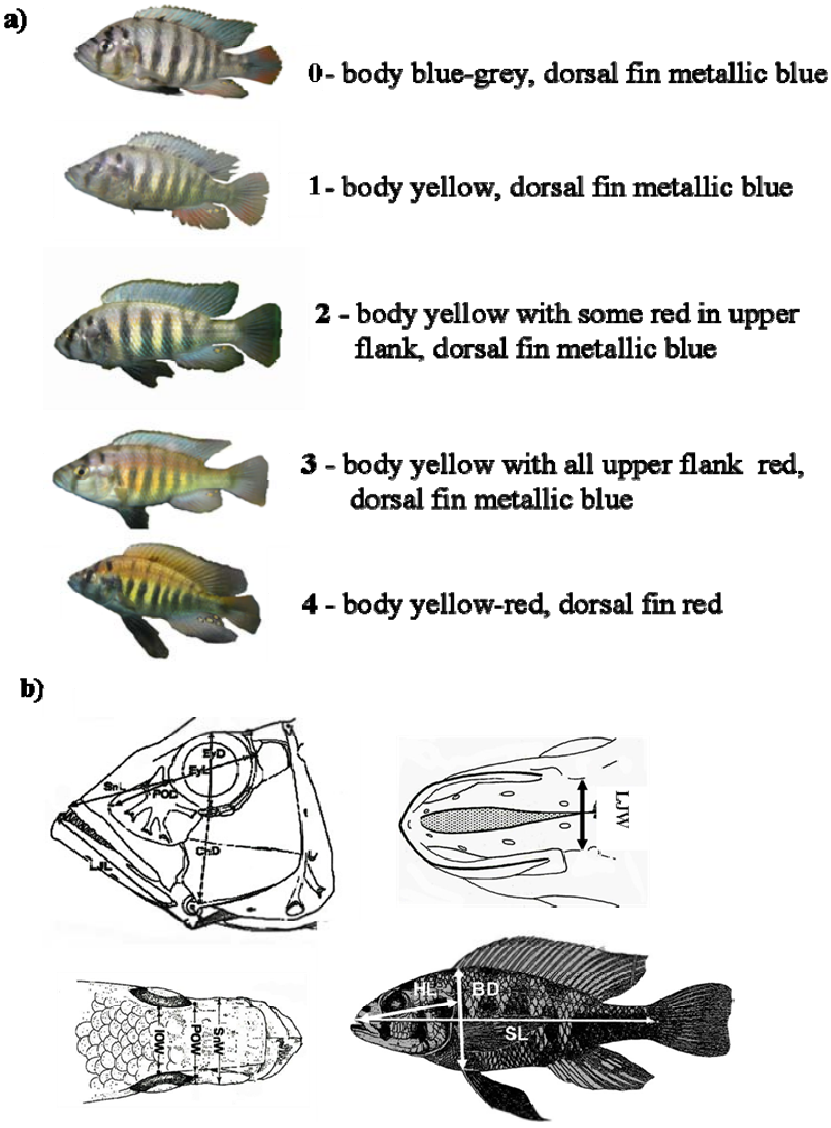

Figure1. a) Scale of male nuptial colour : blue or grey flanks with spiny part of dorsal fin blue is scored as zero, one is a yellow flank but no red, spiny part of dorsal fin is blue, two is yellow flank with some red on the flank along the upper lateral line, spiny dorsal fin is blue, three is yellow flank with a partially red dorsum upwards from the upper lateral line, but a grey body crest and largely blue spiny dorsal fin, four is yellow flank with a completely red dorsum between the upper lateral line and the body crest, red spiny dorsal fin (after van der Sluijs et al., 2008); b) Eco-morphological distances taken by a digital calliper (adapted from Barel et al., 1977): standard length (SL), head length (HL), head width (HW), body depth (BD), lower jaw length (LJL), lower jaw width (LJW), snout length (SnL), snout width (SnW), cheek depth (ChD), pre-orbital depth (PoD), inter-orbital width (IoW), eye length (EyL), eye depth (EyD).

Statistical analysis

As visual representations frequency histograms were plotted for the three PCs of eco-morphological variation and for male nuptial colour scores. Using SPSS we tested if Principal Components one, two, three and colour scores had a normal distribution and fitted alternative regression models to the frequency histograms. Additionally we also used discmixtureprogs v0.4 (Brewer, 2003) and estimated Akaikes Information Criteria (AIC) corrected for sample size (AIC_c; Burnham & Anderson, 2002), which tested statistically whether each sample was better represented by a single normal distribution or by a mixture of two normal distributions. We then estimated AIC_c differences (ΔAIC_c) as AIC_c for the single normal distribution minus AIC_c for the fitted mixture of two distributions. We followed established guidelines for the interpretation of ΔAIC_c (Burnham & Anderson, 2002) and used the interpretation of Hendry et al. (2006): $\Delta AIC_c < -8$ is considered as strong support for a single normal distribution, $-8 < \Delta AIC_c < -5$ as moderate support for a single normal distribution, $-5 < \Delta AIC_c < 5$ as roughly equivalent support for a single normal distribution or a mixture of two normal distributions, $5 < \Delta AIC_c < 8$ as moderate support for a mixture of two normal distributions and $\Delta AIC_c > 8$ is considered as strong support for a mixture of two normal distributions.

DNA extraction and microsatellite amplification

DNA was extracted from fin tissue of 59 individuals collected in 2000 and of 107 individuals collected in 2005 using a QIAGEN® (Basel, Switzerland) Biosprint 96™ extraction robot with the corresponding standard extraction method. We analyzed ten microsatellite loci developed for these species (Ppun 5, Ppun7, Ppun17 and Ppun32; (Taylor et al., 2002)) and for other African cichlid species (OSU20d, OSU19T, OSU16d, (Wu et al., 1999), TmoM5 (Zardoya et al., 1996) and Pzeb3, Pzeb5, (Van Oppen et al., 1997)). Two different sets of markers were used for multiplexing, avoiding overlapping allele ranges. The first set included loci Ppun5, Ppun7, Ppun17, Ppun32, TmoM5 and Pzeb5 and the second set included loci OSU20d, OSU19T, OSU16d and Pzeb3. We used the QIAGEN Multiplex PCR kit for PCR amplification according to the manufacturer's protocol and PCR program. A quantity of 1 µl of a 1:2 dilution of the PCR was added to a volume of sample loading solution (deionised formamide) and 400 bp DNA size standard for analysis of fragments up to 400 nucleotides (Beckman coulter, Fullerton, CA, USA). Denaturated fragments were resolved on an automated DNA sequencer (Beckman coulter, CEQ 8000). We then determined genotypes manually.

Estimation of divergence

Tests for phenotypic and genetic differentiation of individuals collected in 2005 were performed dividing the data in two different ways. In the first analysis we divided the entire sample by catching depth, creating two depth classes, from zero to two meters and below two meters of depth. The two meters breakpoint divided the sample into two groups of similar size. The group of individuals caught between zero and two meters and the group of individuals caught bellow 2 meters of depth will be referred to as "shallow" and "deep" respectively. Second, we divided the sample by male nuptial colour (colour classes 0+1 versus 3+4). This division represented the colouration of the two sister species at islands where they are well differentiated. Therefore groups of colour classes

0+1 and 3+4 will be referred to as ”blue” and ”red” respectively. The single male of colour class 2 in our sample was excluded from this analysis.

Information on depth distribution of individuals collected in 2000 was not available, so tests of genetic differentiation and divergence of these individuals was only done for when samples were divided by male nuptial colour. Therefore groups will also be referred to as “blue” and “red” respectively.

Neutral genetic variation and divergence

CONVERT (Glaubitz, 2004) was used to create input files for other programs and to create allele frequency tables. FSTAT was used to calculate allelic richness at each locus (Goudet, 1995). We used ARLEQUIN 3.11 (Excoffier et al., 2005) to test each locus in each population for departure from Hardy-Weinberg equilibrium, to calculate observed (H_O) and expected (H_E) heterozygosities and to perform tests for genotypic linkage disequilibrium. Statistical significance in the above tests was adjusted for multiple comparisons using sequential Bonferroni adjustments (Rice, 1989). Multilocus and single locus F_{ST} estimates (Wright, 1951) were calculated over 1000 permutations, as implemented in ARLEQUIN 3.11 (Excoffier et al., 2005). We used the allele frequency-based program DOH (http://www2.biology.ualberta.ca/jbrzusto/Doh.php) to calculate genetic assignment for each phenotypically assigned individual (Paetkau et al., 1995). Significance of deviation from random assignments was assessed using 2x2 contingency tests.

Model based assignment tests were performed too, as implemented in the computer program STRUCTURE 2.2 (Pritchard et al., 2000). MCMC simulations were run with 500000 replicates and a burn-in of 50000 replicates for K (number of populations) = 1 to 4 and applying the admixture model, in which individuals may share portions of the genome assigned to more than one population as a result of mixed ancestry (Pritchard et al, 2000). Finally we performed a hierarchical AMOVA, implemented in ARLEQUIN (Excoffier et al., 2005) to investigate how genetic variability was distributed between species and years. Phenotypes were nested inside years.

Phenotypic divergence

We used independent samples *t*-tests to compare mean PC scores between depth groups and male colour groups and to compare mean male colouration between depth groups. We performed a discriminant function analysis to test for differences between groups. We also performed analyses of variance to estimate the divergence in phenotypic traits and to obtain the components of phenotypic variances for the calculation of the P_{ST} values. P_{ST} values are equivalent to Q_{ST} values (Spitze, 1993) but can be influenced by environmental and nonadittive genetic effects.

The theory that population structure can be inferred by comparing the genetic differentiation of presumably selected traits with that of presumably neutral molecular markers (Wright, 1951; Lande, 1992; Spitze, 1993) predicts that the genetic differentiation of a neutral polygenic trait is identical to the genetic differentiation of a single, neutral locus. Hence the genetic differentiation of neutral markers represents the null hypothesis for the expected amount of differentiation of quantitative traits due to

migration or drift alone. If for a given trait the value of Q_{ST}, or in its absence its surrogate $P_{ST,}$ significantly exceeds F_{ST}, divergent selection is invoked, whereas similar values are consistent with neutral evolution and significantly smaller values of Q_{ST} than F_{ST} suggest purifying selection (Merilä & Crnokrak, 2001; McKay & Latta, 2002). The approach relies on a number of assumptions, the most important being that phenotypic variance is mostly additive genetic variance. Several recent studies have pointed out that differences in F_{ST} and Q_{ST} can arise due to biased estimations of differentiation, which can occur for both quantitative traits and molecular markers (Hendry 2002; O'Hara & Merilä, 2005; Goudet & Buchi, 2006; Leinonen et al., 2008; Whitlock, 2008). However, although we are aware of its constraints and caveats, the Q_{ST}/P_{ST} - F_{ST} comparison is a very useful method and perhaps the best currently available, for the inference of selection on phenotypes, other than by experimentation.

P_{ST} values were quantified as the proportion of variance in quantitative traits attributable to differences among populations ($P_{ST} = \sigma^2_{gb}/\sigma^2_{gb}+2(h^2\sigma^2_{gw})$), where σ^2_{gb} and σ^2_{gw} are the among-population and within-population variance components for a phenotypic trait respectively and h^2 is the heritability. As commonly done on studies on morphological traits (Merilä, 1997; Leinonen et al., 2006), we have assumed a heritability of 0.5, which means that environmental and nonadditive effects each account for half of the phenotypic variation. Because of the non-independence of the morphological structures and the associated morphometric distances taken, P_{ST} values for morphology were estimated from Principal Components. Confidence intervals were estimated by bootstrapping over individuals, using code in R (http://www.r-project.org/).

P_{ST} values were contrasted with multilocus F_{ST} values and considered significantly different when their 95% confidence intervals did not overlap.

Additionally we used an individual-based approach to test for a correlation between neutral genetic variation and morphological divergence. We computed morphological divergence between individuals as multidimensional Euclidean distances based on the residuals of the morphological measurements. Matrices of genetic distances ($\hat{a}$) (Rousset, 2000) were computed using the program SPAGeDI (Hardy & Vekemans, 2002). The relationship between the matrices was then tested using a Mantel test based on 1,000 permutations implemented in ARLEQUIN (Excoffier et al., 2005).

The distribution of phenotypic and genetic variation over water depth

We tested for an effect of water depth on the distribution of male colour phenotypes and on morphology of individuals collected in 2005 by performing Spearman-rank correlations between depth and the five colour classes and between depth and Principal Components one, two and three of morphology.

To test for an effect of water depth on neutral genetic variation and morphological divergence we used an individual-based approach equivalent to testing for isolation-by-distance. Depth distances between individuals were computed with the program SPAGeDI (Hardy & Vekemans, 2002). We tested the relationships between depth distances and genetic distances ($\hat{a}$) (Rousset, 2000) and between depth distances and multidimensional Euclidean distances with Mantel tests based on 1,000 permutations as implemented in ARLEQUIN (Excoffier et al., 2005).

Common garden experiment and analysis of lab bred individuals

Individuals of the parental generation were collected at Kisssenda Island in November 2005. Individuals were assigned to either incipient species on the basis of phenotype, and were kept in separate tanks at 24 ± 1°C and 12L: 12D daylight cycle. All fish were fed twice a day with dry commercial cichlid pellets in the morning and fresh shrimps and peas in the afternoon. Males and females within each incipient species population were allowed to interact and mate freely. During the first two months of 2006 six F1 families were produced. Families were raised in separate tanks. *Pundamilia* are female mouth-brooders and we removed brooding females with their broods from the breeding tank. Each brood had a different wild caught mother. To assign broods to a sire, we assigned paternity by genotyping fry, their mother, and all possible sires at five microsatellite loci. We used 1 dinucleotide (Ppun32) and four tetranucleotide (Ppun5, Ppun7, Ppun17, Ppun21) loci, with specific primers designed for *P. pundamilia* and *P. neyererei* (Taylor et al., 2002). Each brood had a different wild caught father. Hence every brood was a different and unrelated family. Ten *P. nyererei* males from three families and 25 *P. pundamilia* males from three families were raised to maturity (at least 12 months of age). Parents and F1 males from families of *P. pundamilia* and *P. nyererei* will be referred to as "blue" and "red" respectively.

Parental individuals were not available for phenotypic analysis because we could not preserve them before they showed strong phenotypic effects of aging. To detect possible family effects on the phenotypic variation observed among the experimental families, a nested ANOVA was carried out (two factors: family (6 families) nested in species (*P. pundamilia* and *P. nyererei*)). To separate environmental from heritable components of phenotypic variance between the incipient species, we compared phenotypic variance between males of "blue" and "red" raised in a common garden environment with that observed between the males of these populations in the wild. We used these data to address the question if adaptive phenotypic plasticity contributes to the divergence found between the wild populations.

To test if differences between the P_{ST} values of wild and common garden experimental fishes could be explained by sampling error, we resampled 50 times from the two wild populations to the sample size of the parents of the common garden experimental fish (i.e. 6 individuals in each species). We made pairwise comparisons for the first three Principal Components between the resampled populations (25 of each species) obtaining 625 P_{ST} values for each trait. We then compared the P_{ST} value from our common garden experiment to the distribution of P_{ST} values from wild fish. To see if the variance observed within the common garden populations was due to accidental crossing of the incipient species we compared the variances of the wild populations with the variance of the common garden populations. An increase in variance of lab bred populations might suggest mixing of the incipient species during breeding. If phenotypic differences observed in the wild have a genetic basis, differences should be maintained when individuals are raised in a common environment. On the other hand, if phenotypic plasticity was the cause behind the observed differences we expected that differences would collapse when fish were raised in a common environment.

Results

Phenotypic variation

Male nuptial colouration

We assigned 31 individuals to colour class zero, 26 to class one, one individual to class two, nine to class three and 40 individuals to class four. Strong bimodality in male nuptial colouration of the wild population was evident from obvious discontinuity in the frequency histogram of colour scores (Fig. 2a). A cubic regression made a very good (r^2 = 0.945), yet not significant fit to the data (P = 0.296). A ΔAICc value of 120.47 (ΔAICc = 357.83 - 237.36) was considered as strong support for a mixture of two normal distributions (ΔAICc > 8).

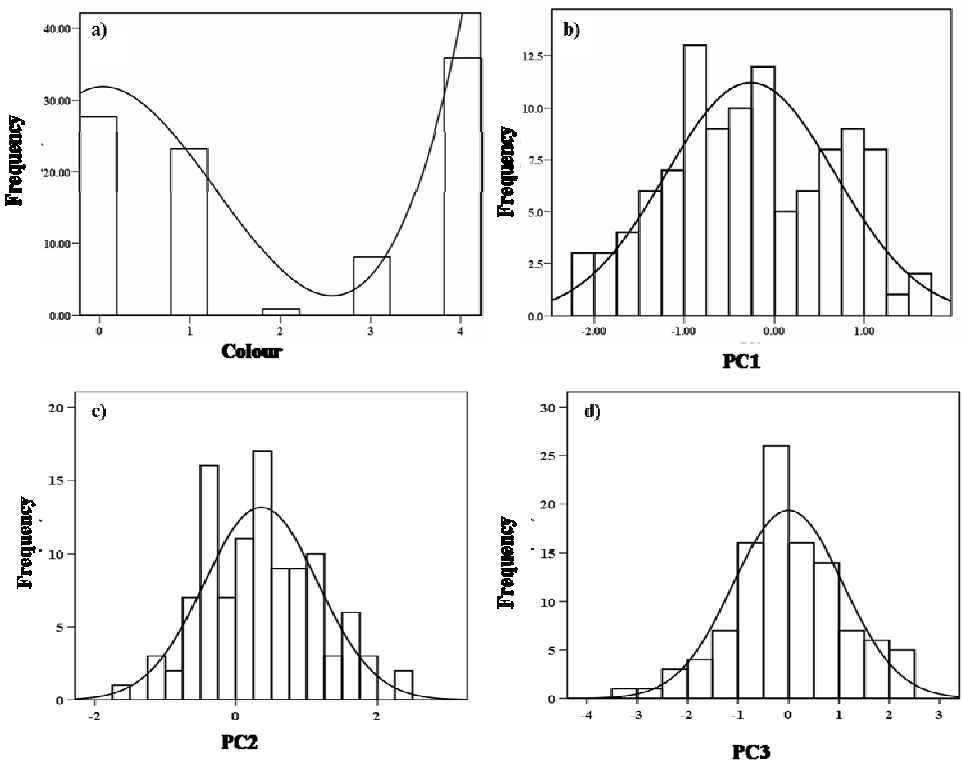

Figure 2. Frequency distributions of a) colour scores of wild caught individuals with a cubic regression fitted to it; b) Principal Component 1, c) Principal component 2 and d) Principal Component 3 of wild caught individuals with a normal distribution fitted to it.

Morphology

The Principal Component analysis identified three major axes of morphological variation (axes that explain more than 10% of the variation) (Table 1). Head length, lower jaw length, snout length, snout width and cheek depth all had heavy positive loadings on PC1, eye length and depth had heavy negative loadings on PC2 and body depth had a heavy positive loading on PC3. When comparing the percentage of variance in each trait that load to each PC, seven out of twelve traits had more than 80% of their variance loading to PC1. Pre-orbital depth was the only trait whose variance loaded heavily to PC2 (79%) and body depth was the only trait with its variance loading mostly to PC3 (58%). Eye length and eye depth had both close to 50% of their variance loading to PC1 and to PC2.

Table 1. Loadings of morphological variables on the first three principal component axes, the percent variance explained by each axis, and the relative percent variance in each variable that load to each PC axis.

Trait	PC1	PC2	PC3	PC1	PC2	PC3
Head Length	0.85	0.16	-0.26	95.99	1.55	2.47
Head Width	0.60	0.47	0.27	74.92	20.91	4.17
Body Depth	0.35	0.15	0.82	38.60	3.23	58.18
Lower jaw Length	0.69	0.33	-0.16	89.38	9.30	1.32
Lower jaw Width	0.59	-0.31	-0.22	85.93	10.79	3.28
Snout Length	0.71	0.36	-0.41	82.75	9.68	7.58
Snout Width	0.74	-0.03	0.19	98.15	0.07	1.78
Cheek Depth	0.70	-0.32	-0.08	91.02	8.65	0.33
Pre-orbital Depth	0.19	0.61	-0.16	17.00	79.69	3.31
Inter-orbital Depth	0.60	0.14	0.32	90.67	2.24	7.08
Eye Length	0.56	-0.63	-0.02	63.45	36.52	0.02
Eye Depth	0.48	-0.73	0.07	48.60	51.12	0.28
% variance	37.51	17.06	10.30			

Genetic and Phenotypic divergence

Neutral Genetic variation and divergence

Most microsatellite markers exhibited a high degree of polymorphism (Appendix 1). A total of 210 alleles were found in the 166 individuals, ranging from 2 to 44 per locus.

For the sampling year of 2000, there was one locus deviating from HWE in the population ”red” and five loci deviated from HWE in the population ”blue”. Two out of 45 tests of LD were significant in the population “red” and 20 were significant in the population “blue”. After sequential Bonferroni correction three loci were in Hardy-Weinberg disequilibrium and four tests showed linkage disequilibrium, all in the population “blue”.

For the sampling year of 2005, before sequential Bonferroni correction, deviations from HWE were observed at one locus in one population, “red”. One test out of 45 showed significant linkage disequilibrium in the same population. When populations were divided by depth, significant departures from HWE were observed at 2 loci and only in the population “deep”. Now five tests showed linkage disequilibrium in the population “deep”. After sequential Bonferroni correction no significant departures from HWE or linkage equilibrium were detected for the sampling year of 2005.

Differentiation at neutral genetic markers between depth classes was low (F_{ST} = 0.001) and not significant (P = 0.25). It was low between nuptial colour classes too but it was significantly different from zero, both for the year 2000 (F_{ST} = 0.011, P = 0.01) and for the year 2005 (F_{ST} = 0.009, P = 0.008). Single locus F_{ST} values had a very narrow distribution and few were significant, but single locus F_{ST} values were generally higher between nuptial colour classes than between depth ranges (Table 2).

Table 2. Pairwise single locus and multilocus F_{ST} statistics between groups made based on depth and on colour for the sampling year of 2005 and by colour for the year 2000. Significant F_{ST} values (P<0.05) are shown in bold face.

locus	depth (2005)	colour (2005)	colour (2000)
Ppun5	0.001	0.011	0.003
Ppun7	0.000	-0.001	0.004
Ppun17	0.006	**0.016**	-0.002
Ppun32	-0.007	0.004	-0.008
OSU16d	0.003	**0.008**	0.011
OSU19t	**0.008**	0.020	**0.015**
OSU20d	0.022	**0.009**	**0.040**
Pzeb3	-0.004	0.019	0.000
Pzeb5	-0.006	-0.005	0.027
TmoM5	-0.001	0.004	0.001
multilocus	0.004	**0.009**	**0.011**

STRUCTURE 2.2 (Pritchard et al., 2000) identified the most likely number of populations (K) in our sample of two incipient species and two years as being 1 (estimated -ln probability of data = -7632.3; P > 0.99). Nonetheless, plots for K = 2 showed clear differences between individuals phenotypically assigned to "blue" and those phenotypically assigned to "red" , but no differences between individuals of the same phenotype collected in different years (Fig. 3). The hierarchical analysis of molecular variance (AMOVA) confirmed this result: there were significant differences between phenotypes within years (P = 0.00) but no differences between years (P = 0.65).

The assignment tests corroborated the F_{ST} values. Genetic assignment to classes defined by colour was significantly different from random, but assignment to classes defined by water depth was not (Fig. 4).

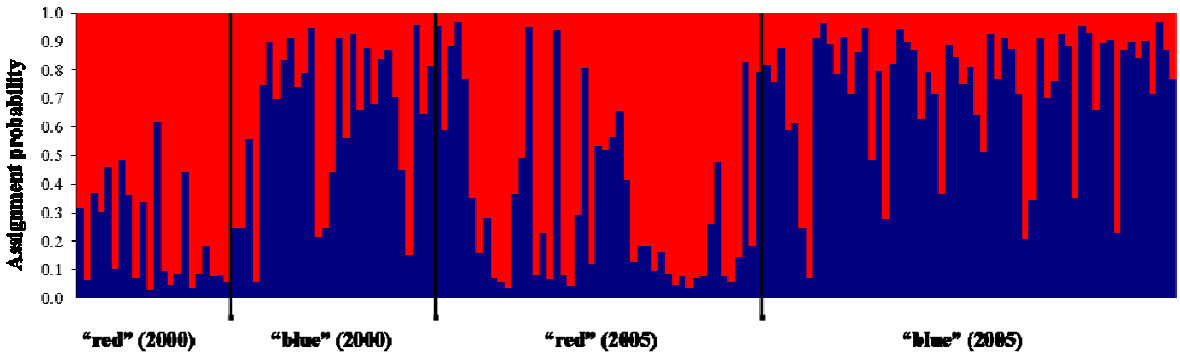

Figure 3. Bayesian assignment probabilities of individuals to lineages estimated using STRUCTURE 2.2. Each bar represents an individual and colours indicate the proportion of an individual's genotype assigned to a particular population. Assignment proportions differ between the phenotype classes but not between sampling years.

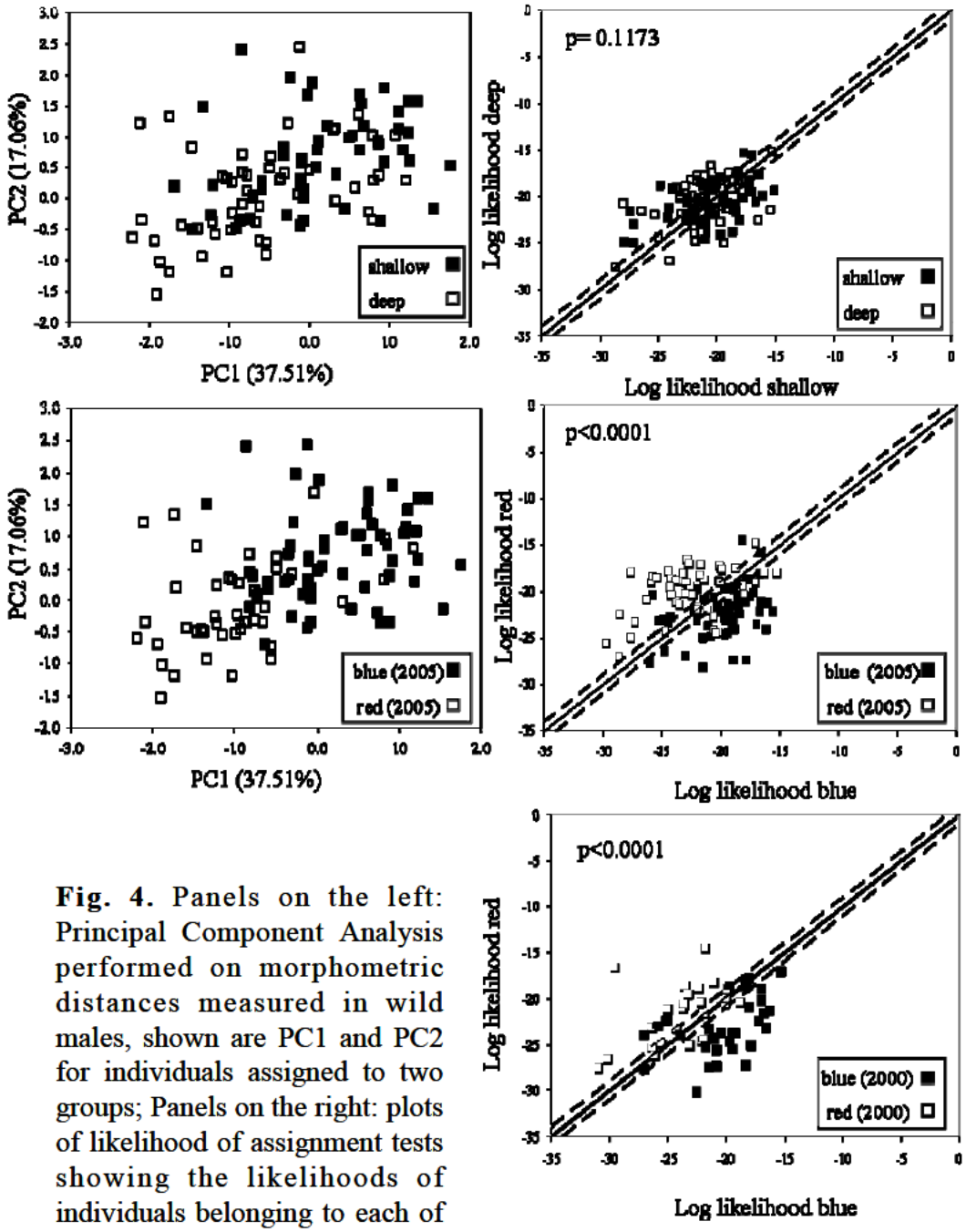

Fig. 4. Panels on the left: Principal Component Analysis performed on morphometric distances measured in wild males, shown are PC1 and PC2 for individuals assigned to two groups; Panels on the right: plots of likelihood of assignment tests showing the likelihoods of individuals belonging to each of the groups compared. From top to bottom: groups based on depth distribution from year 2005 and on colour from year 2005 and from 2000. P-values shown on the plots are the probabilities that the observed assignments were the result of chance. The solid diagonal line indicates the equal probability of being assigned to one population or the other. Points outside the dotted lines are assigned to one population with more than 10 times the likelihood as the other one.

Phenotypic divergence

Independently of how groups were formed mean PC scores and male nuptial colouration were all significantly different (Table 3).

Table 3. *T*-tests comparing mean PC scores of groups based on male colouration and comparing mean PC scores and male colouration of groups based on depth range.

		t	df	Sig.(2-tailed)	Mean Difference	Std. Error	2.50%	97.50%
colour	PC1	8.052	104	0.000	1.175	0.146	0.885	1.464
	PC2	5.143	104	0.000	0.730	0.142	0.448	1.011
	PC3	-4.37	104	0.000	-0.867	0.198	-1.260	-0.474
depth	PC1	4.003	104	0.000	0.685	0.171	0.346	1.025
	PC2	3.567	104	0.001	0.529	0.148	0.235	0.824
	PC3	-2.319	104	0.022	-0.483	0.208	-0.896	-0.070
	colour	-5.584	104	0.000	-1.601	0.287	-2.170	-1.033

However, differentiation along morphometrics PC1 and PC2 was weaker when individuals were grouped by water depth than when they were grouped by male nuptial colouration (Fig. 4). The discriminant function analyses confirmed these results: groups were significantly different both based on depth range (Wilk´s lambda = 0.763, P = 0.02) and based on male nuptial colour (Wilk´s lambda = 0.303, P = 0.00). However 91.5% of the originally grouped individuals were correctly classified when individuals were grouped by colour, whereas the same figure was 77.6% when individuals were grouped by depth. The most important variables contributing to the differences between groups were head length, snout length and lower jaw length.

Phenotypic differentiation (P_{ST}) both between depth groups and between colour groups was higher than the neutral genetic marker differentiation (Table 4, Fig. 5). Confidence intervals, however, were large for all traits. When analysis was done by depth group (Fig. 5a), male nuptial coloration had the highest P_{ST} value (P_{ST} = 0.208), followed by eco-morphometric PC1 (P_{ST} = 0.135), PC2 (P_{ST} = 0.113) and PC3 (P_{ST} = 0.049). The confidence intervals for male colouration, PC1 and PC2 laid outside the confidence interval of the multilocus microsatellite-derived F_{ST} value and were also higher than the highest of the 10 single locus F_{ST} values, indicating that differentiation along these axes of phenotypic variation was higher than expected by drift alone. Differentiation in eco-morphometric PC3 was close to the expectation from drift alone.

Table 4. P_{ST} and F_{ST} values and their 95% confidence intervals for the wild caught populations divided by depth, by colouration and for individuals from the common garden experiment.

	P_{ST} colour	P_{ST} PC1	P_{ST} PC2	P_{ST} PC3	F_{ST}
depth	0.208	0.135	0.113	0.049	0.001
	0.087-0.377	0.038-0.27	0.024-0.241	0.001-0.146	0.000-0.003
colour	-	0.375	0.183	0.156	0.009
	-	0.240-0.514	0.072-0.321	0.055-0.303	0.004-0.013
lab bred	0.525	0.323	0.001	0.049	-

All P_{ST} values were larger when populations were grouped by male nuptial colouration, but the sequence of relative magnitudes was unchanged (Fig. 5b). It is not surprising that P_{ST} in male nuptial colouration was close to 1, because this trait was used to make the groups. Nonetheless, we show its P_{ST} for comparative purposes with the common garden populations. Among the morphometrics PCs, PC1 had again the highest value (P_{ST} = 0.375), followed by PC2 (P_{ST} = 0.183) and PC3 (P_{ST} = 0.156). All confidence intervals laid outside those of the multilocus F_{ST} value and were higher than the highest single locus F_{ST} value, indicating that differentiation along several axes of morphological variation was higher than expected by drift alone.

Variation in neutral loci had no effect on morphological structure. Pairwise genetic distances among individuals were not significantly correlated with morphological distances (r = 0.046, P = 0.12).

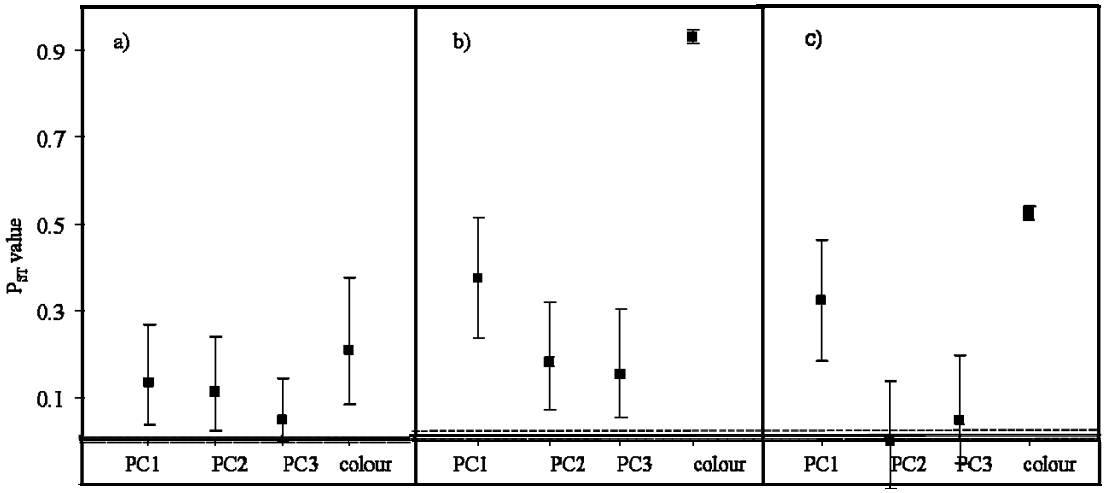

Figure 5. Comparison of F_{ST} values calculated from 10 microsatellite loci with P_{ST} values (with their 95% confidence intervals) calculated from Principal components and from colour classes between fish from "shallow" and "deep" waters (panel a), wild caught "blue" and "red" (panel b), and lab bred "blue" and "red" (panel c). Symbols from left to right in all panels refer to PC1, PC2 and PC3 and male colouration. The solid lines indicate the multilocus F_{ST} value. The dotted lines indicate the 95% confidence intervals of the F_{ST} value.

The distribution of phenotypic and genetic variation over water depth

Male colouration was significantly correlated with water depth (Fig. 6a). Individuals with bluish colouration were caught at shallower depth, whereas individuals with reddish nuptial colouration were distributed over the entire depth range. Principal Components one and two in eco-morphology showed a significant negative correlation with water depth, while PC3 showed a significant positive correlation (Fig. 6b, c, d). Individuals with lower PC1 and PC2 scores and higher PC3 scores tended to be from the lower end of the water depth range.

Variation in water depth clearly caused genetic and morphological structure. Pairwise genetic distances among individuals were significantly correlated with the depth differential between them ($r = 0.190$, $P < 0.001$) and so were morphological distances ($r=0.219$, $P < 0.001$). However, depth differential between individuals only explained 3.5% of the genetic distance and 4.7 % of the morphological distance.

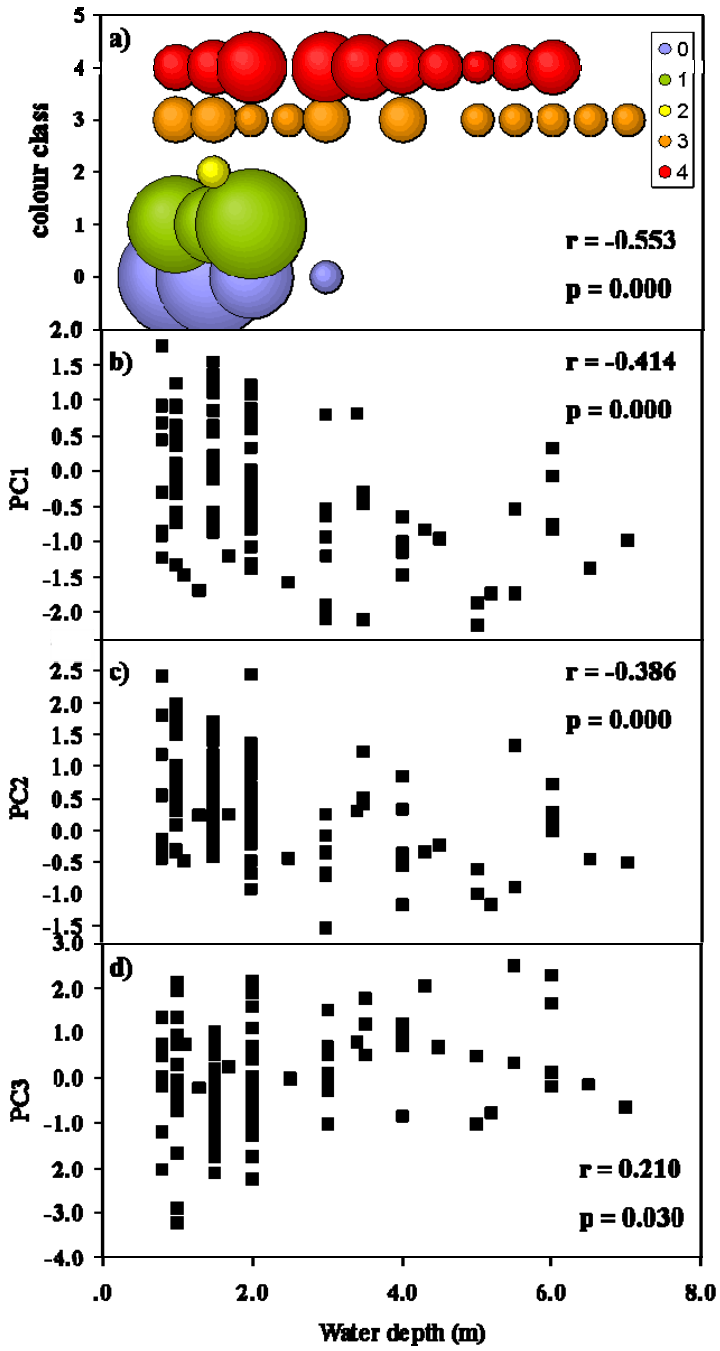

Figure 6. Depth distribution (in meters) of individuals by a) colour class; b) Principal Component 1 scores; c) Principal Component 2 scores d) Principal Component 3 scores.

Common garden experiments

Family did not explain any significant fraction of the variance in either colour ($F_{4,29} = 0.726$, $P = 0.58$) or morphological PC1 ($F_{4,29} = 0.24$, $P = 0.91$) but had significant effects on morphological PC2 ($F_{1,4} = 2.78$, $P = 0.05$) and PC3 ($F_{4,29} = 3.63$, $P = 0.02$). Parental phenotype class on the other hand, was the main source of variation in male nuptial colour ($F_{1,29} = 28.78$, $P = 0.00$) and in morphological PC1 ($F_{1,29} = 12.99$, $P = 0.00$) and of some of the variation in PC3 ($F_{1,29} = 7.37$, $P = 0.01$). That neither parental phenotype class nor family explained much of the variation in PC2 indicates low heritability (Table 5).

Table 5. Nested ANOVA performed on colour, PC1, PC2 and PC3 of lab bred individuals of *Pundamilia* with brood nested inside species. Significant values ($P<0.05$) are shown in bold face.

Source of variation	Dependent Variable	Type III Sum of Squares	df	Mean Square	F-ratio	Sig.	Noncent. Parameter	Observed Power
Family (species)	colour	2.798	4	0.699	0.726	0.581	2.904	0.206
	PC1	0.412	4	0.103	0.238	0.915	0.952	0.094
	PC2	4.081	4	1.020	2.778	0.050	11.112	0.688
	PC3	4.928	4	1.232	3.633	**0.016**	14.531	0.815
Species	colour	27.738	1	27.738	28.788	**0.000**	28.788	0.999
	PC1	5.620	1	5.620	12.991	**0.001**	12.991	0.936
	PC2	0.155	1	0.155	0.423	0.521	0.423	0.096
	PC3	2.500	1	2.500	7.370	**0.011**	7.370	0.747
Error	colour	27.942	29	0.964				
	PC1	12.546	29	0.433				
	PC2	10.651	29	0.367				
	PC3	9.836	29	0.339				
Total	colour	171.000	35					
	PC1	40.688	35					
	PC2	58.065	35					
	PC3	15.544	35					

The colouration of males bred from “blue” parents ranged from zero to three on the colour scale, with most individuals of types zero, one or two. Males bred from “red” parents ranged from one to four on the colour scale, with most individuals of type four (Fig.7a). There was considerable overlap between lab-bred families of the two incipient species on morphology PC axis 1. However, males bred from “red” parents often had lower scores than males bred from “blue” parents (Fig. 7b), resembling the situation in wild fish. There was complete overlap along the PC2 and the PC3 axes.

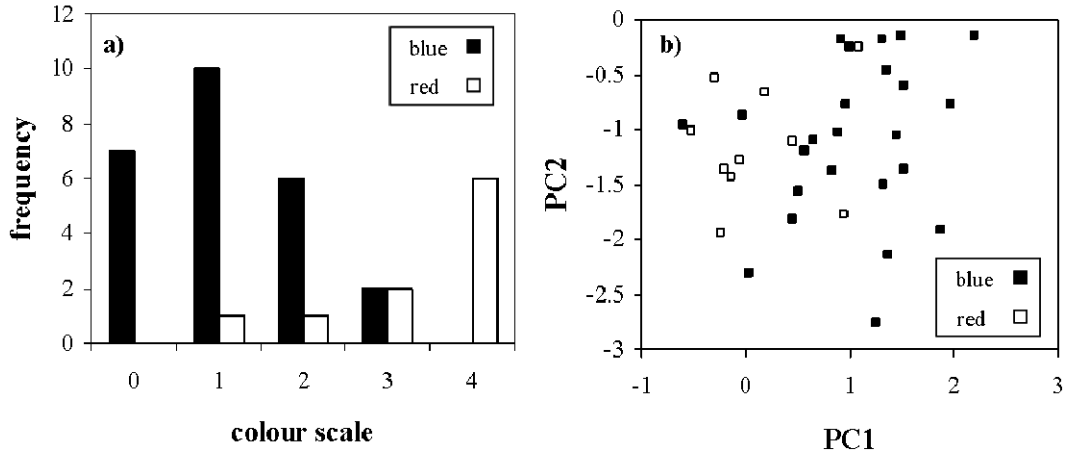

Figure 7. Frequency distributions of a) colour scores and b) Principal Components1 and 2 for lab bred individuals.

Divergence between fish bred from "blue" versus "red" wild phenotypes in the common garden experiment was highest for colouration (P_{ST} = 0.525), followed by PC1 (P_{ST} = 0.323), PC3 (P_{ST} = 0.049) and PC2 (P_{ST} = 0.0005) (Table 4). Hence, all P_{ST} estimates among populations grown in a common environment were smaller than the P_{ST} among wild populations, but those for PC2 and PC3 were much smaller. Due to the small sample sizes, confidence intervals were large, and no P_{ST} estimate was significantly different from the neutral expectation except that for colour. If we assume that variation among the parental fish that we used for breeding was representative for the variation in the wild populations, we can apply the confidence intervals estimated from wild fish. Even with this approach it was clear that P_{ST} values from morphological PC2 and PC3 were close to the expectations from drift alone. In contrast, the PC1 and colour-derived P_{ST} values were significantly outside the neutral expectation also in the fish raised under identical conditions (Fig. 5c).

We took 50 random subsamples of 6 individuals from each of the two wild caught incipient species groups defined by male nuptial colouration (”blue” and “red”). We then made 625 pairwise comparisons and generated a distribution of P_{ST} values for each Principal Component. The P_{ST} values for PC2 and PC3 of our common garden populations laid below and very close to the lower confidence intervals of the P_{ST} distributions generated from the resampling of the wild populations, respectively (Table 6). The P_{ST} value of PC1 however, was similar to the P_{ST} value generated from the resampling of wild fish, although its confidence interval was larger. Intrapopulation variances were not significantly different between wild and lab bred populations for PC1, PC2 and PC3 of “blue” phenotypes (Levene’s test, PC1: P = 0.263, PC2: P = 0.856, PC3: P = 0.226) and “red” (Levene’s test, PC1: P = 0.309, PC2: P = 0.156, PC3: P = 0.132), but they were significantly larger for colouration of lab bred populations (Levene’s test, “blue”: P = 0.002*;* “red”: P = 0.000).

Table 6. Means and confidence intervals of P_{ST} distributions generated from 625 pairwise comparisons for PC1, PC2 and PC3.

	P_{ST} (PC1)	P_{ST} (PC2)	P_{ST} (PC3)
mean	0.386	0.186	0.132
97.50%	0.811	0.661	0.628
2.50%	0.133	0.023	0.017

Discussion

Our study found that male nuptial colouration is correlated with water depth and has a bimodal distribution and that divergence at neutral loci is extremely low but significant, which is consistent with previous findings (Seehausen et al., 2008). We have also found that differentiation in morphological traits in wild fish is correlated with water depth and is generally much larger than the degree of differentiation for neutral genetic makers, although never as large as differentiation in colours. However, several of the morphological differences were lost when fish were raised in a common laboratory environment. In the following section we will discuss each of these findings and their general implications (see Fig. 8 for summary diagram of methods, results and conclusion).

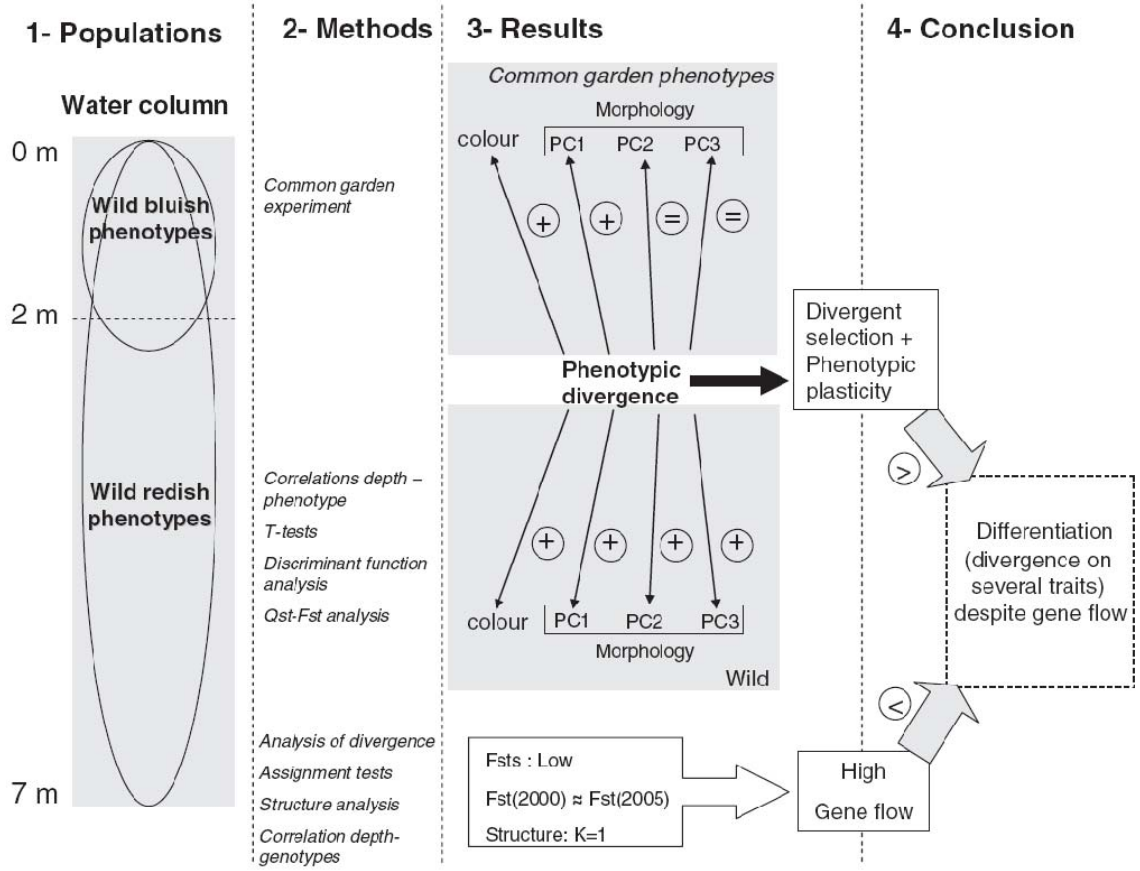

Figure 8. Summary diagram showing in panel 1) the distribution of colour phenotypes across water depth; 2) main methods used to analyse phenotypic and genetic divergence and heritability of the phenotypic traits; 3) main results showing the level of divergence in quantitative traits for wild and lab bred phenotypes and divergence at neutral markers. Symbols + and = and are used if the trait divergence is equal or higher than neutral expectations respectively; 4) conclusion showing how the different processes inferred from the results affect the measured characters and lead to their differentiation (>) or homogenization (<).

Genetic structure and differentiation

Our results suggest that the phenotypically variable populations of *Pundamilia* around Kissenda Island are significantly structured at unlinked putatively neutral marker loci, but differentiation is weak and gene flow widespread. Water depth appears to be a structuring force, with significant isolation by depth in an individual-based analysis, even though the effect is small. Groups of individuals made based on catching depth were not significantly differentiated, but groups made based on male nuptial colour were significantly differentiated. Male nuptial colour in turn is significantly associated with water depth. The biological meaningfulness of small but significant F_{ST} values has rightly been questioned (Waples, 1998). However, in our case, biological meaningfulness of these low but significant measures of differentiation is supported by our result that similarly weak but also similarly significant differentiation was found in two independent samples taken with an interval of five years, whereas year had no significant effect on molecular variance. Hence, the differentiation between phenotype groups is unlikely just a spurious result of random from year to year variation. From this, we conclude that populations with blue or red male nuptial colouration are indeed incipient species at this location, even though there clearly is evidence for introgressive hybridization between them.

The more frequent occurrence of linkage disequilibrium in "blue", when compared to "red", suggests that "blue" phenotypes are more strongly introgressed. "Blue" phenotypes occupy predominantly shallower water depth, where they coexist syntopically with "red" phenotypes. The latter, on the other hand, are distributed throughout the whole depth gradient, with maximum abundance at greater depth. They overlap with "blue" only in part of their depth range, the shallow water. The considerable overlap of spawning habitat and the fact that genetic differentiation between the colour phenotypes was larger than that between the depth ranges, suggest that factors other than spawning habitat segregation must contribute to the partial isolation between these incipient species. The assignment tests confirmed the F_{ST} values; there is significantly non-random distribution of genotypes between the colour phenotype groups, suggesting either assortative mating within types or very strong selection against intermediates.

Phenotypic differentiation vs. neutral genetic divergence

Estimation of Q_{ST} or its analogous P_{ST} and F_{ST} values is challenging and subjected to a number of potential pitfalls (Hendry, 2002; O'Hara & Merilä 2005; Whitlock, 2008). An error of most concern in the calculations of Q_{ST} which could also apply to P_{ST} values is the imprecision of the estimations when using low sample sizes, which tend to bias Q_{ST} downwards (O'Hara & Merilä 2005). F_{ST} calculations and interpretations in high gene flow species can also be problematic since sampling error across loci may account for 1/2S of the obtained F_{ST} value (Waples, 1998). The observed F_{ST} value might hence be an overestimate. Hence, maximum caution is warranted when interpreting Q_{ST}/P_{ST}-F_{ST} comparisons. In our case, the above mentioned potential sources of error seem conservative as we are searching for the signal of divergent selection.

Hendry (2002) also raised two important issues on the relationship between Q_{ST} and F_{ST}: (1) an assumption of this approach is that mutation rates are considerably lower than migration rates, which might be violated when markers with high mutation rates are used or populations with low gene flow analysed; (2) as divergence time increases and

migration and mutation rates decrease demonstrating that Q_{ST} is higher than F_{ST} becomes increasingly harder, even if divergence of quantitative traits was driven by selection. Although we have used markers known to have high mutation rates, the F_{ST} values indicate extremely high gene flow between populations; therefore migration rates are most likely considerably higher than mutation rates.

Whitlock (2008) recently suggested that comparing the Q_{ST} (or its equivalent value P_{ST}) for one trait with the mean multilocus F_{ST} was incorrect, and that it should instead be compared with the distribution of F_{ST} values. Our results are robust to this concern. We only considered evidence for selection the non-overlap of the 95% confidence intervals, which are good representations of single locus F_{ST} distributions. Our P_{ST} estimates were in every significant case larger than the largest of our 10 single locus F_{ST} values.

All the P_{ST} values from wild fish were higher than the F_{ST} values. Confidence intervals were generally large, which is expected when using a small number of populations (O'Hara & Merilä, 2005). However, aside from one exception (PC3 when populations were divided by depth), they did not overlap with the confidence intervals of the F_{ST} values and had their lower end above the highest single locus F_{ST} values. These results therefore are consistent with the hypothesis of divergent selection on male nuptial colour and head morphology (PC1 and PC2).

Both the F_{ST} value and the P_{ST} values were higher when populations were divided based on colour than when populations were divided based on depth range. This difference can be explained by the fact that, even though there is a correlation between colour and water depth, there is considerable overlap between colour-phenotypes in shallow water. Colour may be a better estimate of a male's long term average water depth than the depth at which we caught the fish. However this also implies that factors other than spawning site isolation by water depth contribute to partial isolation between phenotypes at Kissenda Island.

Divergent selection and phenotypic plasticity in morphology

Morphology was more divergent than expected by drift between fishes from different water depth and between fishes of different breeding colouration. The individual-based analysis also revealed a small but highly significant effect of depth distances between individuals on their morphology. However, in contrast to breeding colouration, variance in morphology was unimodally distributed with large overlap between the depth zones and colour phenotypes. None the less, 91.5% of the individuals were correctly assigned in our discriminant functions analysis when groups were based on male colour, and all *t*-tests were significant, indicating clear differences in morphology between the incipient species.

We compared the phenotypic differences between wild caught fish with those between animals raised in a "common garden" lab environment. This allows us to estimate the relative contribution of phenotypic plasticity to differences found in the wild.

Colouration and Principal Component 1 of morphology were largely maintained when populations were raised in a common garden environment, whereas the differences in PC2 and PC3 dropped to the neutral F_{ST} line in lab bred fish, suggesting that the differences in several morphological traits observed in wild fish, are due to phenotypic plasticity. The P_{ST} values for lab bred PC2 and PC3 were also significantly lower than

expected by chance through random sampling of wild fish for breeding. Therefore it is unlikely that the loss of P_{ST} in PC2 and PC3 was due to a sampling bias. Hence, we refer to these axes as "genetic" PC axis 1 and "environmental" PC axes 2 and 3. Head length, lower jaw length and width, snout length and width, cheek depth and inter-orbital depth had more than 80% of their relative variance loading to PC1 and are therefore most likely traits with a high heritable component. These traits, all related to head morphology, were also the most important measures contributing to differences between groups in the discriminant functions analysis. The causes for divergence in these traits most likely arise from differences in feeding resources available at different water depths. Studies on the feeding ecology of cichlids from Lake Victoria have suggested that at islands where depth segregation between the sister species *P. pundamilia* and *P. nyererei* is strong, they have different diets (Bouton et al., 1997; Mrosso et al., unp.), with individuals of *P. pundamilia* feeding more on benthic food while *P. nyererei* feed more on plankton. These differences in diet are perhaps a consequence of the different microhabitats, shallow water and crevices in the blue *P. pundamilia* and greater average depth in the red *P. nyererei* (Seehausen, 1997; Seehausen et al., 2008) and require phenotypic adaptation, some of which, we found to be achieved through heritable divergence.

On the other hand, divergence in the wild populations in the traits body depth, pre-orbital depth, eye length, eye depth and to a smaller extent head width was largely or partly due to phenotypic plasticity. With the exception of eye length and eye depth, these traits had positive loadings on genetic PC axis 1, as well as on environmental PC axes 2 and 3. From this we infer that environmentally induced divergence and genetic adaptation operate in the same direction, which is an additional argument for the action of divergent selection between the incipient species.

In conclusion, we found that the *Pundamilia* population at Kissenda Island in Lake Victoria is phenotypically and genetically structured along a gradient of water depth. Two incipient species with different male breeding colour occupy different but overlapping depth ranges and are weakly but consistently differentiated at neutral gene loci. This confirms previous findings of a correlation between ambient light colour along the water depth gradient, male nuptial colour and speciation. We show that male nuptial colour is heritable. Additionally divergent selection acts on ecologically relevant morphological traits along the same gradient. Response to this selection is partly accommodated by heritable adaptation and partly by phenotypic plasticity, where the latter appears to accentuate differences between the incipient species. Hence, divergent selection acts on several traits, contributing to the maintenance of differentiation despite high levels of gene flow.

Acknowledgements

We would like to thank Bob O'Hara for writing the codes for R and Gaku Takimoto for implementing them. We also thank Denis Roy for the very useful discussions, Pascal Vonlanthen for comments on the manuscript, and the members of the Fish Ecology & Evolution lab for helpful feedback and discussion. We thank the Tanzanian Fisheries Research Institute for hospitality and access to facilities. Mhoja Kayeba and Mohammed Haluna provided crucial assistance in the field. This research was supported by Swiss National Science Foundation Project 3100A0-106573, to O.S. .

References

Barel, C. D. N., Vanoijen, M. J. P., Witte, F. & Wittemaas, E. L. M. 1977. Introduction to Taxonomy and Morphology of Haplochromine Cichlidae from Lake Victoria - Manual to Greenwoods Revision Papers. *Neth. J. Zool.* **27**: 333-338.

Bolnick, D. I. 2004. Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. *Evolution* **58**: 608-618.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411**: 944-948.

Bouton, N., Seehausen, O. & van Alphen, J. J. M. 1997. Resource partitioning among rock-dwelling haplochromines (Pisces : Cichlidae) from Lake Victoria. *Ecol. Freshw. Fish* **6**: 225-240.

Brewer, M. J. 2003. Discretisation for inference on normal mixture models. *Stat. Comput.* **13**: 209-219.

Burnham, K. P., Anderson, D. R. 2002. *Model selection and Multimodel inference, a practical information theoretic approach*. New York Springer.

Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O. 2005. Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Mol. Ecol.* **14**: 4341-4353.

Dieckmann, U. & Doebeli, M. 1999. On the origin of species by sympatric speciation. *Nature* **400**: 354-357.

Dijkstra, P. D., Seehausen, O., Gricar, B. L. A., Maan, M. E. & Groothuis, T. G. G. 2006. Can male-male competition stabilize speciation? A test in Lake Victoria haplochromine cichlid fish. *Behav. Ecol. and Sociobiol.* **59**: 704-713.

Dobzhansky, T. 1970. *Genetics and the evolutionary process*. Columbia University Press, New York and London.

Drossel, B. & McKane, A. 2000. Competitive speciation in quantitative genetic models. *J. of Theor. Biol.* **204**: 467-478.

Emelianov, I., Marec, F. & Mallet, J. 2004. Genomic evidence for divergence with gene flow in host races of the larch budmoth. *P. Roy. Soc. B-Biol. Sci.* 271: 97-105

Excoffier, L., Estoup, A. & Cornuet, J. M. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727-1738.

Fleming, I. A. & Gross, M. R. 1994. Breeding Competition in a Pacific Salmon (Coho, Oncorhynchus-Kisutch) - Measures of Natural and Sexual Selection. *Evolution* **48**: 637-657.

Gavrilets, S. 2004. *Fitness lanscapes and the origin of species*. Princeton University press.

Glaubitz, J. C. 2004. CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol. Ecol. Notes* **4**: 309-310.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J. Hered.* **86**: 485-486.

Goudet, J. & Buchi, L. 2006. The effects of dominance, regular inbreeding and sampling design on Q(ST), an estimator of population differentiation for quantitative traits. *Genetics* **172**: 1337-1347.

Grant, P.R. & Grant, B. R. 2008. Fission and Fusion of Darwin's finches populations. *Philos. T. Roy. Soc. B* **363**: 2821-2829.

Hardy, O. J. & Vekemans, X. 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 4: 618-620.

Hendry, A. 2002. QST > = ≠ < FST. *Trends Ecol. Evol.* 17:502.

Hendry, A. P., Grant, P. R., Grant, B. R., Ford, H. A., Brewer, M. J. & Podos, J. 2006. Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. *P. Roy. Soc. B-Biol. Sci.***273**: 1887-1894.

Higashi, M., Takimoto, G. & Yamamura, N. 1999. Sympatric speciation by sexual selection. *Nature* **402**: 523-526.

Kondrashov, A. S. & Kondrashov, F. A. 1999. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351-354.

Lande, R. 1992. Neutral Theory of Quantitative Genetic Variance in an Island Model with Local Extinction and Colonization. *Evolution* **46**: 381-389.

Leinonen, T., Cano, J. M., Makinen, H. & Merila, J. 2006. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J. Evol. Biol.* **19**: 1803-1812.

Leinonen, T., O'Hara, R. B., Cano, J. M. & Merilä, J. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J. Evol. Biol.* **21**: 1-17.

Lu, G. Q. & Bernatchez, L. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (Coregonus clupeaformis): Support for the ecological speciation hypothesis. *Evolution* **53**: 1491-1505.

Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. 2006. Sensory drive in cichlid speciation. *Am Nat.* **167**: 947-954.

McKay, J. K. & Latta, R. G. 2002. Adaptive population divergence: markers, QTL and traits. *Trends Ecol. Evol.* **17**: 285-291.

Merilä, J. 1997. Quantitative trait and allozyme divergence in the greenfinch (Carduelis chloris, Aves: Fringillidae). *Biol. J. Linn. Soc.* **61**: 243-266.

Merilä, J. & Crnokrak, P. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.***14**: 892-903.

Naisbit, R. E., Jiggins, C. D. & Mallet, J. 2001. Disruptive sexual selection against hybrids contributes to speciation between *Heliconius cydno* and *Heliconius melpomene*. *P. Roy. Soc. B-Biol. Sci.* **268**: 1849-1854.

Nosil, P & Sandoval C. P. 2008. Ecological niche dimensionality and the evolutionary diversification of stick insects. *PLoS ONE* 3: 1-11.

Nosil, P. & Harmon, L. J. . Niche dimensionality and ecological speciation. In: *Ecology and Evolution*, (R. Butlin, D. Schluter and J. R. Bridle, eds.). Cambridge University press, in press.

Nosil, P., Harmon, L. J. & Seehausen O. 2008. Ecological explanations for incomplete speciation. *Trends Ecol. Evol.* , in press.

O'Hara, R. B. & Merilä, J. 2005. Bias and precision in Q(ST) estimates: Problems and some solutions. *Genetics* **171**: 1331-1339.

Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. 1995. Microsatellite Analysis of Population-Structure in Canadian Polar Bears. *Mol. Ecol.* **4**: 347-354.

Pritchard, J. K., Stephens, M. &Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Rice, W. R. 1989. Analyzing Tables of Statistical Tests. *Evolution* **43**: 223-225.

Rogers, S. M. & Bernatchez, L. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonas sp Salmonidae) species pairs. *Mol. Biol. and Evol.* **24**: 1423-1438.

Rousset, F. 2000. Genetic differentiation between individuals. *J. Evol. Biol.* 13: 58-62.

Rundle, H. D., Nagel, L., Boughman, J. W. & Schluter, D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**: 306-308.

Schliewen, U., Rassmann, K., Markmann, M., Markert, J., Kocher, T. & Tautz, D. 2001. Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Mol. Ecol.* **10**: 1471-1488.

Schluter, D. 1996a. Ecological causes of adaptive radiation. *Am. Nat.* **148**: S40-S64.

Schluter, D. 1996b. Ecological speciation in postglacial fishes. *Philos. T. Roy. Soc. B* **351**: 807-814.

Seehausen, O. 1997. Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid (Haplochromis nyererei). *Ecol. Freshw. Fish* **6**: 57-60.

Seehausen, O., van Alphen, J. J. M. & Witte, F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808-1811.

Seehausen, O. & van Alphen, J.J.M. 1999. Can sympatric speciation by disruptive sexual selection explain rapid evolution of cichlid diversity in Lake Victoria? *Ecol. Lett.* 2: 262-271.

Seehausen, O., Witte, F., van Alphen, J.J.M. & Bouton, N. 1998. Direct mate choice maintains diversity among sympatric cichlids in Lake Victoria. *J. of Fish Biol*. 53 (Suppl. A): 37-55.

Seehausen, O. 2006. Conservation: Losing biodiversity by reverse speciation. *Curr. Biol.* 16: 334-337.

Seehausen, O.. Progressive levels of trait divergence along a "speciation transect" in the Lake Victoria cichlid fish *Pundamilia*. In: *Ecology and Evolution*, (R. Butlin, D. Schluter and J. R. Bridle, eds.). Cambridge University press, in press.

Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., vander Sluijs, I., Schneider, M. V., Maan M. E., Tachida, H., Imai, H. & Okada N. 2008. Speciation through sensory drive in cichlid fish. *Nature***, 455**: 620-626.

Slatkin, M. 1987. Gene Flow and the Geographic Structure of Natural-Populations. *Science* **236**: 787-792.

Spitze, K. 1993. Population-Structure in *Daphnia obtusa* - Quantitative Genetic and Allozymic Variation. *Genetics* **135**: 367-374.

Taylor, M. I., Meardon, F., Turner, G., Seehausen, O., Mrosso, H. D. J. & Rico, C. 2002. Characterization of tetranucleotide microsatellite loci in a Lake Victorian, haplochromine cichlid fish: a *Pundamilia pundamilia* x *Pundamilia nyererei* hybrid. *Mol. Ecol. Notes* **2**: 443-445.

van der Sluijs, I., van Alphen, J. J. M. & Seehausen, O. 2008. Preference polymorphism for coloration but no speciation in a population of Lake Victoria cichlids. *Behav. Ecol.* **19**: 177-183.

Van Oppen, M. J. H., Turner, G. F., Rico, C., Deutsch, J. C., Ibrahim, K. M., Robinson, R. L. & Hewitt, G. M. 1997. Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *P. Roy. Soc. B-Biol. Sci.* **264**: 1803-1812.

Waples, R. S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *J. Hered.* 89: 438-450.

Whitlock, M. C. 2008. Evolutionary inference from Q(ST). *Mol. Ecol.* 8: 1885-1896

Wright, S. 1951. The Genetical Structure of Populations. *Annals of Eugenics* **15**: 323-354.

Wu, L., Kaufman, L. & Fuerst, P. A. 1999. Isolation of microsatellite markers in *Astatoreochromis alluaudi* and their cross-species amplifications in other African cichlids. *Mol. Ecol.* **8**: 895-897.

Zardoya, R., Vollmer, D. M., Craddock, C., Streelman, J. T., Karl, S. & Meyer, A. 1996. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *P. Roy. Soc. B-Biol. Sci.* **263**: 1589-1598.

Appendix 1. Allelic variability at 10 microsatellite loci in *Pundamilia* from Kissenda island. Number of samples analysed in each population (N), number of alleles at each locus (A), allelic richness (AR), range of allele size (AS), observed heterozygosity (HO) and expected heterozygosity (HE) at each locus.

		shallow	deep	blue (2005)	red (2005)	blue (2000)	red (2000)
	N	52	55	60	46	33	26
Ppun5	A	21	21	22	19	20	19
	A_R	20.75	20.51	20.43	19.00	17.99	17.68
	A_S	186-278	190-274	186-278	190-270	182-270	194-278
	H_O	0.96	0.86	0.89	0.93	0.86	0.92
	H_E	0.93	0.93	0.91	0.94	0.95	0.94
Ppun7	A	24	24	25	21	18	19
	A_R	23.68	22.96	23.74	20.64	16.64	18.09
	A_S	186-314	166-294	186-290	166-314	194-282	166-274
	H_O	0.98	0.91	0.97	0.91	0.87	0.92
	H_E	0.95	0.95	0.95	0.95	0.94	0.95
Ppun17	A	17	17	16	19	14	15
	A_R	16.43	16.51	14.37	18.53	13.17	14.22
	A_S	86-170	94-162	94-170	86-162	98-162	106-166
	H_O	0.88	0.83	0.83	0.89	0.85	0.79
	H_E	0.86	0.91	0.85	0.92	0.89	0.90
Ppun32	A	5	4	4	5	7	4
	A_R	4.92	4.00	4.00	4.91	6.01	3.88
	A_S	153-163	153-159	153-159	153-163	153-165	153-161
	H_O	0.70	0.67	0.71	0.65	0.67	0.58
	H_E	0.71	0.67	0.70	0.67	0.72	0.65
Pzeb3	A	4	3	4	4	3	3
	A_R	4.00	2.89	3.92	3.93	3.00	3.00
	A_S	310-316	312-316	310-316	310-316	310-314	310-314
	H_O	0.50	0.48	0.42	0.58	0.63	0.52
	H_E	0.45	0.43	0.39	0.49	0.53	0.54
Pzeb5	A	4	4	4	4	2	3
	A_R	4.00	3.85	3.98	3.87	2.00	2.81
	A_S	118-130	118-130	118-130	118-130	118-126	118-128
	H_O	0.67	0.57	0.68	0.53	0.45	0.35
	H_E	0.56	0.50	0.55	0.50	0.51	0.46

TMO5	A	28	27	28	25	21	22
	A_R	27.17	26.13	25.17	24.78	18.14	19.96
	A_S	317-383	319-377	317-383	321-375	323-379	299-375
	H_O	0.94	0.96	0.93	0.98	0.82	0.85
	H_E	0.95	0.96	0.95	0.96	0.94	0.95
osu16d	A	27	24	28	23	22	18
	A_R	26.36	23.23	25.17	22.70	18.29	16.70
	A_S	82-154	84-148	82-154	84-136	80-156	86-138
	H_O	0.98	0.96	0.97	0.98	0.94	0.92
	H_E	0.95	0.95	0.95	0.95	0.94	0.94
osu19t	A	28	29	25	25	18	21
	A_R	27.32	27.99	22.52	24.86	15.74	19.60
	A_S	107-173	107-175	107-175	107-175	109-171	109-187
	H_O	0.92	0.93	0.93	0.91	0.85	0.81
	H_E	0.95	0.96	0.94	0.96	0.92	0.96
osu20d	A	41	41	44	33	29	23
	A_R	39.28	38.88	37.62	32.11	21.82	20.64
	A_S	153-287	139-273	139-287	155-273	155-273	157-281
	H_O	0.90	0.94	0.93	0.91	0.88	0.81
	H_E	0.97	0.97	0.97	0.96	0.94	0.93
Average	H_O	0.84	0.81	0.83	0.83	0.77	0.72
	H_E	0.82	0.82	0.81	0.82	0.81	0.80

Chapter 6

Genetics of male colour in a Lake Victoria cichlid fish species pair and its role in sympatric speciation

Genetics of male colour in a Lake Victoria cichlid species pair and its role in sympatric speciation

Isabel S. Magalhaes, Ole Seehausen

Abstract

The hypothesis of sympatric speciation by sexual selection has been contentious. African cichlid fish are a speciation model system in which this mechanism has been proposed repeatedly. Most recent theoretical models of sympatric speciation by disruptive sexual selection were tailored to apply to African cichlids. Major theoretical constrains lie in the demographic instability of the process and in the difficulty of evolving linkage disequilibrium between female preference and male trait. Most theoretical work concludes that the genetic architecture of female preference and male trait is a key determinant of the likelihood of disruptive sexual selection to result in speciation. Here we investigate the genetic architecture controlling male nuptial colouration in a sympatric sibling species pair of cichlid fish from Lake Victoria. These species differ conspicuously in male colouration and female mating preferences for these. We estimated that the difference between the species in male nuptial red colouration is controlled by a minimum number of to two to four genes, which do not only have additive effects. Yellow colouration appears to be controlled by one gene with complete dominance. The two colours appear to be epistatically linked. Knowledge on the number of genes controlling male colouration can help us further assess whether assumptions made in simulation models of sympatric speciation by sexual selection are realistic.

Introduction

Animal colouration has for long been recognized as an important trait in intra- and interspecific signalling (Endler, 1992). Colour polymorphisms are know to be important as mate choice cues in Guppies from Trinidad (Endler & Houde, 1995), Australian Bowerbirds (Endler & Mielke, 2005), Papua new Guinea coral reef fish (Messmer et al., 2005), sticklebacks (Boughman,2001) and African cichlids (Seehausen et al., 1999; Turner et al., 2001; Kocher, 2004). Theoretical work suggested that sexual selection acting on colouration by and on its own may drive population divergence and speciation (Turner & Burrows, 1995; Payne & Krakauer, 1997; Higashi et al., 1999; van Doorn et al., 1998; van Doorn et al., 2004). These models suggest that disruptive sexual selection on secondary sexual characters, such as colouration, alone can be sufficient for sympatric speciation in the absence of ecological adaptation. The likelihood of this outcome, however, has been questioned since conditions required are stringent (Arnegard & Kondrashov, 2004) and it is debatable whether assumptions made in these models are realistic. Models of sympatric speciation tend to assume that simple additive genetics with few loci control traits involved in sexual isolation (reviewed by Gourbiere, 2004). Most models suggest that the likelihood of speciation decreases as the number of loci controlling colouration increases.

Most of the models have been inspired by empirical evidence coming from research on African cichlid fish (Turner & Burrows, 1995; van Doorn et al., 1998; Higashi et al., 1999; Lande et al., 2001; van Doorn et al., 2004). African cichlids form the Earth's most species rich vertebrate assemblages within geographically narrowly confined areas: individual lakes. They have incredibly high speciation rates. In the particular case of Lake Victoria more than 500 endemic species have probably evolved in just 15 000 years (Johnson et al., 1996; Stager & Johnson, 2008). Well differentiated species coexist with incipient species and colour morphs, providing a unique natural laboratory for evolutionary biologists studying speciation. Male nuptial colouration is one of the most compelling features of these fish and is known to play a central role in the evolution and maintenance of species richness (Seehausen et al, 1997; Seehausen et al., 1999; Seehausen & Schluter, 2004). A well studied example of this are the sister species *Pundamilia pundamilia* and *Pundamilia nyererei*, which are widely and sympatrically distributed at rocky islands in Lake Victoria. The species differ primarily in male breeding colouration. Males of both species have blackish underparts and blackish vertical bars on the flanks but *Pundamilia nyererei* males have red dorsum and dorsal fin and yellow flanks whereas *Pundamilia pundamilia* males are blue-grey dorsally and have a metallic blue dorsal fin and blue-grey flanks (Seehausen, 1997). Females of both species are cryptically coloured and yellowish (many but not all populations of *P. nyererei*) or greyish (most *P. pundamilia* populations). A recent study found that in populations living on moderately shallow to moderately steep light gradients there were strong correlations between ambient light colour, male nuptial colour, visual pigments and female mating preference for male nuptial colour, supporting a scenario of speciation through sensory drive (Seehausen et al., 2008). Female mating preference was estimated to be controlled by between one and four loci (Haesler & Seehausen, 2005). Male colouration is known to be heritable (Seehausen et al., 1997; van der Sluijs et al., in prep), but the number of genes controlling male colour differences had not been estimated yet.

In this study, the amount of red male nuptial colouration in weakly differentiated populations of the two species was analysed to confirm that red colouration is heritable and to estimate the minimum number of genes controlling differences in red colouration between species. If differences in male colour are heritable, and controlled by several genes with additive effects, expectations are that the first hybrid generation will express relatively uniform colouration, whereas colour should segregate again among second hybrid generation males. Additionally we also analysed the presence of yellow body colouration, which appears to correlate with the presence of red colouration, and tested for the hypothesis that its presence is controlled by one gene with complete dominance.

Materials and Methods

Collection of individuals and breeding

Individuals from the parental generation used in this study were collected at Python Island, in the Mwanza gulf of Lake Victoria. *P. nyererei* were collected in 1991 and *P. pundamilia* in 1992. Both species were maintained through seven generations in separate breeding populations. In 1999 F1 hybrids were produced. F1 hybrids were bred from *P. nyererei* males and *P. pundamilia* females (3 families) and vice-versa (4 families). F1 hybrid males were variable in colour, most individuals in all families were partially red, but in several families some were entirely blue. F2 hybrids were bred between 2000 and 2001 from randomly chosen F1 hybrids. In order to acquire additional information on the heritability of colour, backcrosses were bred from a blue phenotype F1 hybrid male (*P. pundamilia* male x *P nyererei* female) with several *P. pundamilia* females. Backcrosses from red phenotype F1 hybrid males, and backcrosses to *P. nyererei* were, however, not available for analysis. All individuals were raised to maturity in family groups. Females from this experiment were used previously to estimate the number of genes determining female mating preferences (Haesler & Seehausen, 2005).

Photography

Every male was photographed in breeding condition when fully mature and at least six month old. Males of several F1 families were photographed in a photo cuvette with standardized background using slide films exposed at 100 ISO with a Pentax Super A Reflex Camera. These pictures were then digitized using a slide scanner. Males of some F1 and all F2 and Backcross families were photographed in aquaria using an Olympus Camedia Master digital camera. These males were brought into standardized dominant territorial motivational state, assessed by expression of dark vertical bars. This motivational state was achieved by giving males visual access through a transparent partition to other males in two adjacent compartments.

One hundred and forty seven males were photographed and included in the analysis: 12 *P. nyererei*, 20 *P. pundamilia*, 29 F1s (two families from *P. pundamilia* male x *P. nyererei* female and one family from *P. nyererei* male x *P. pundamilia* female), 50 F2s (from 15 families) and 36 backcrosses (from six families) (Fig. 1). To estimate the repeatability of our measurements, we photographed 42 individuals (16 backrosses and 26 F2s) on at least two different days.

Colour analysis

The salient differences in nuptial coloration between male *P. pundamilia* and male *P. nyererei* lie in the distribution of hues on the flanks and the dorsal section of the body (Fig. 1). Males of *P. pundamilia* have blue-grey flanks and a metallic blue spinuous dorsal fin, whereas males of *P. nyererei* have a bright red spinuous dorsal fin, bright red dorsum and upper flanks above the lateral line, and yellow lower flanks below the lateral line. We used digital analyses of images to investigate the genetics underlying the extension of orange and red on the body. Using Photoshop 6.0 (Adobe Systems Inc.) we cropped pictures to remove background, eyes and fins and kept only the body. The body area covered by red colouration was quantified in Sigma Scan 4.0 (SPSS Inc., Chicago Illinois, USA). The delimiting criteria were defined by a combination of hue and saturation (hue: 0-35, saturation: 25-100% plus hue: 225-255, saturation: 30-100%). The total body area of the fish was captured by setting the brightness criterion such that all non-white pixels were selected. The percentage of body covered by red was calculated by dividing the number of pixels that matched the red colour criteria by the total number of pixels occupied by the body of the fish.

Only the red cover of the body was accessible to analysis with image analysis software. However, expression of red colouration on upper flanks and dorsum in *Pundamilia* appears conditional to expression of yellow on the lower flanks (Seehausen et al., 2008; Figure 1). Every fish with any red colour on the upper flanks or dorsum in our data set had yellow lower flanks. Additionally, a variable proportion of the males without red had yellow lower flanks. We scored and analysed yellow as a simple presence/absence trait. Scoring was done without knowledge of the cross type and on "dissected" images of the fish body without fins.

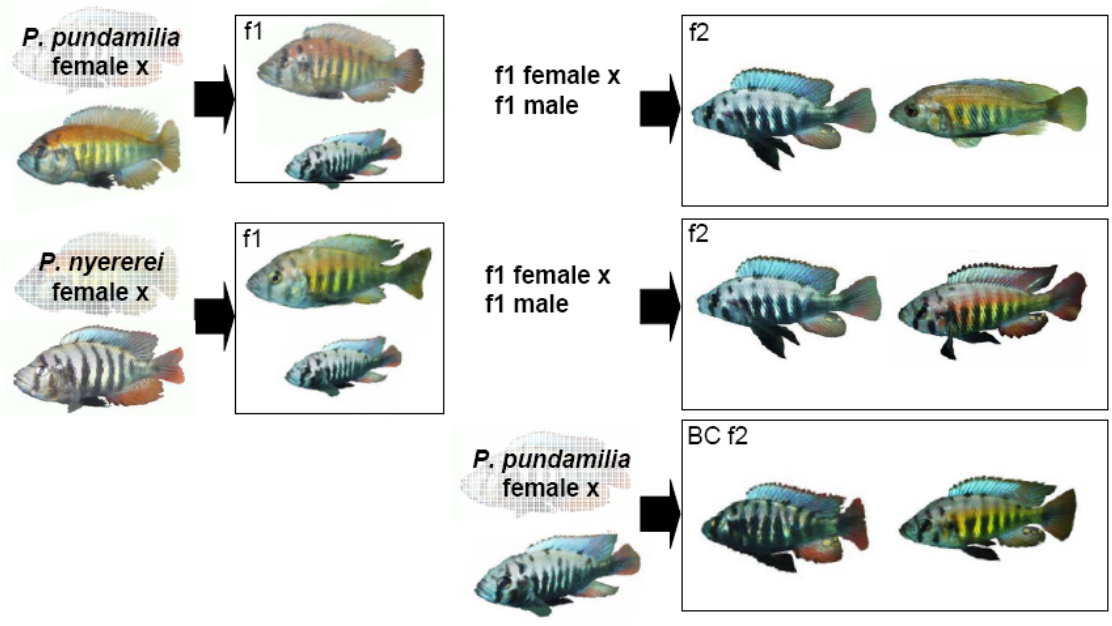

Figure 1. Crossing scheme and colour patterns of parental and hybrid males. Paled-out images symbolize male colour of species of dam. Small images in the F1 boxes indicate the occurrence of a few entirely blue males in these otherwise rather homogeneous families of intermediate male colour.

Data analysis

Repeatability of relative size of body area covered by red colour was calculated using 16 backcross and 26 F2 males as

$$R = \frac{\sigma_B^2}{\sigma_w^2 + \sigma_B^2}$$

where σ^2_B is the variance between individuals and σ^2_W is the variance within individuals. Repeatability is 1 when all variance is between individuals, while it is zero when all variance is within individuals.

As visual representations we plotted frequency histograms of the percentage of red cover in the parental and hybrid generations. We performed independent samples t-tests to compare means between *P. pundamilia, P. nyererei*, F1, F2, and backcrosses, and tested for homogeneity of variances using SPSS 14.0 (SPSS Inc.). We also performed t-tests between the two directions of the crosses within the F1 and F2 generations separately to test for sex-linkage of the red cover.

For yellow colouration we plotted frequency histograms of the individuals in each generation with yellow colouration absent or present. We performed binomial tests for equality of proportions, using R (www.r-project.org), to test if there were significant differences in the proportions of individuals with yellow colouration between generations. We also performed binomial tests for equality of proportions between the two directions of the crosses within the F1 and F2 generations separately to test for sex-linkage of the yellow body colour.

Tests of epistasis, additivity and dominance effects

Methods to estimate the number of genes, to test for additivity and dominance and to estimate variance components assume normally distributed values within each line. Percentage of red cover was not normally distributed in any line, except in *P. nyererei*, and was therefore log transformed. As we only had backcrosses in one direction (to *P. pundamilia*) these were not included in the tests of additvity and dominance, nor in the estimation of number of genes.

We tested for the contribution of additivity and dominance to the difference in red cover between the two species and their crosses by using the joint-scaling method (Cavalli, 1952; Hayman, 1960a; Mather & Jinks, 1971). If a simple additive model fits red cover, each cross line's mean phenotype should be the average of the mean phenotypes of its parental lines. Dominance will cause hybrids to resemble one parental line more than the other. The joint -scaling method is summarized in Lynch & Walsh (1998).

This method fits a multiple regression model to the observed line cross means (Hayman, 1960a),

$$z_i = \mu_0 + M_{i2}\alpha + M_{i3}\delta + \in_i \quad (1)$$

where z_i is the trait mean in the i*th* line, μ_0 is the mean of all line means, α is the additive genetic effect, δ is the dominance effect and ε_i is the sampling error associated with the i*th* line. M_{i2} and M_{i3} are matrices of coefficients of additive and dominance effects respectively. When epistasis is not included in the model, it is implicitly included in the error term along with the sampling error. However, unless the effects of dominance can de discounted, including epistasis in the model requires more than six lines. We only used

4 lines. Therefore, the effect of epistasis was not estimated separately and is contained in the error term.

The parameters μ_0, α and δ are estimated by using the equation:

$$\hat{a} = (M^T V^{-1} M)^{-1} M^T V^{-1} z \quad (2)$$

and the predicted line means by each of the models are estimated by the equation:

$$\hat{z} = M\hat{a} \quad (3)$$

where M is the matrix of coefficients of additive effects M_{i2} in the additive model:

$$M_{i2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \quad (4)$$

and the matrix M_{i3} in the additive-dominance model:

$$M_{i3} = \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (5)$$

V is the diagonal weighting matrix of sampling variances (squared standard errors) of the observed line means:

$$V = \begin{pmatrix} [SE(\mu_{P1})]^2 & 0 & 0 & 0 \\ 0 & [SE(\mu_{P2})]^2 & 0 & 0 \\ 0 & 0 & [SE(\mu_{F1})]^2 & 0 \\ 0 & 0 & 0 & [SE(\mu_{F2})]^2 \end{pmatrix} \quad (6)$$

and z is the vector of observed line means. The standard errors of the predicted line means are also obtained by square rooting the diagonal values of the matrix $M(M^T V^{-1} M)^{-1} M^T$.

The observed means of each line were then compared with the predicted means from each model using the equation:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\bar{z}_i - \hat{z}_i)^2}{SE(z_i)^2} \quad (7)$$

where k is the number of observed lines, $\bar{Z}_i$ and $\hat{Z}_i$ are the line means observed and expected under the model respectively, and $SE(z_i)$ is the standard error of the observed line means. This test statistic is chi-square distributed, under the null assumption that genes have purely additive action, with the degrees of freedom being the number of lines minus the number of estimated parameters. The fitting of the models was done by sequential model fitting (Lynch & Walsh, 1998). If the additive model is rejected, dominance or epistasis effects also contribute to the difference between the lines. If the additive plus dominance model is also rejected, epistasis and / or linkage are contributing to the difference between the lines. The fitting of the additive model was tested against the fitting of the additive-dominance model by the difference (Lynch & Walsh, 1998):

$$\Lambda = \chi^2_A - \chi^2_{AD} \quad (8)$$

where χ^2_A and χ^2_{AD} are the chi-square values of the additive and additive-dominance models respectively. This difference is equivalent to a likelihood ratio test-statistic, with a chi-square distribution and degrees of freedom equivalent to the number of parameters estimated by one model minus the numbers of parameters estimated by the other.

Estimation of the minimum number of loci controlling differences in red body colour

Before estimating the number of genes we needed to estimate the segregational variance (σ^2_S) and its variance (var(σ^2_S)). The segregational variance describes the excess variance that appears in the F2 generation due to the segregation of the parental-lines genes. We estimated these values by applying a joint-scaling test (Hayman, 1960b) similar to the one used to estimate additivity and dominance, except that it is applied to the lines variances instead of their means.

To estimate σ^2_S and var(σ^2_S) we used the equations (Hayman, 1960b):

$$\hat{a} = (M^T V^{-1} M)^{-1} M^T V^{-1} v \quad (9)$$

and

$$\hat{v} = M\hat{a} \quad (10)$$

M is the matrix of coefficients from equations predicting line variances from parental and segregation variance (Hayman, 1960b):

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 1 \end{pmatrix} \quad (11)$$

V is the sampling covariance matrix (Hayman, 1960b):

$$V = \begin{pmatrix} 2(\sigma^2_{P1})^2/n+2 & 0 & 0 & 0 \\ 0 & 2(\sigma^2_{P2})^2/n+2 & 0 & 0 \\ 0 & 0 & 2(\sigma^2_{F1})^2/n+2 & 0 \\ 0 & 0 & 0 & 2(\sigma^2_{F2})^2/n+2 \end{pmatrix} \quad (12)$$

where σ^2_{P1}, σ^2_{P2}, $\sigma^2_{F1,}$and σ^2_{F2} are the observed variances for each line and n is the sample size of each line; v is the vector of observed line variances. Equations (9) and (10) are iterated by replacing σ^2_{P1}, σ^2_{P2}, $\sigma^2_{F1,}$ and σ^2_{F2} in V by the new values obtained from equation (10). This is done until the values stabilize. The result of equation (9), â, is a vector of the variance components for the two parental generations,σ^2_{P1} and σ^2_{P2}, and the segregational variance σ^2_S; var(σ^2_{P1}), var(σ^2_{P2}) and var(σ^2_S) are the diagonal elements of the final estimate of $(M^T V^{-1} M)^{-1}$.

In order to estimate the minimum number of genes contributing to the species differences in male colouration we used the Castle-Wright estimator. The original Castle-Wright equation uses the means of inbred lines and the variances of their F1, F2 and backcrosses (Castle, 1921; Wright, 1968). The method was extended by Lande (1981) to accommodate outbreeding populations and their crosses. Cockerham (1986) added a correction for sampling error. We use the equation of Cockerham (1986) as:

$$n_E = \frac{(\mu_{P_1} - \mu_{P_2})^2 - \sigma^2_{P_1} - \sigma^2_{P_2}}{8\,\sigma^2_S} \quad (13)$$

where n_E is the estimate for the minimum number of genes, μ_{P1}, μ_{P2}, σ^2_{P1} and σ^2_{P2} are the means and the variances of the two parental lines respectively. σ^2_S is the segregational variance estimated as mentioned above.
The variance of the estimate was calculated by using the equation (Cockerham, 1986):

$$\mathrm{var}(\hat{n}_E) \approx \hat{n}^2_E \left[\frac{4(\sigma^2_{P_1} + \sigma^2_{P_2})}{(\mu_{P1} - \mu_{P2})} + \frac{\mathrm{var}(\sigma^2_S)}{(\sigma^2_S)^2} \right] \quad (14)$$

Given that the Castle - Wright estimator assumes additive genetic variation, which has not been tested for male colouration in *Pundamilia*, we added a correction to the estimator that takes into account linkage and loci of unequal effect (Zeng, 1992):

$$n_{e(zeng)} = \frac{2\bar{c}n_e + C_\alpha(n_e - 1)}{1 - n_e(1 - 2\bar{c})} \quad (15)$$

where $\bar{c}$ is the recombination index and C_α is the squared coefficient of variation of effects and n_e is estimated by using equation 12. C_α is not known so we used a range of corrections to represent the effect of different underlying distributions of allelic effects: C_α = 0 which considers all allelic effects equal; C_α = 0.25 which considers a normal distribution of allelic effects; C_α = 1 which assumes a negative exponential distribution, and C_α = 4 which assumes a leptokurtic (L-shaped) distribution of allelic effects. The recombination index was estimated from the haploid number of chromosomes (M) (Lynch, Walsh, 1998), using the equation:

$$\bar{c} = \frac{M - 1}{2M} \quad (16)$$

The haploid number of chromosomes was assumed to be 22 (Albertson et al., 2003b).

The variance of the estimate was calculated by using the equation (Zeng, 1992):

$$\mathrm{var}(n_{e(zeng)}) = \frac{4\bar{c}^2 + (1 + C_\alpha)^2\,\mathrm{var}(\hat{n}_e)}{[1 - \hat{n}_e(1 - 2\bar{c})]^4} \quad (17)$$

Test for a model of one locus with complete dominance controlling yellow body colour

If one locus with complete dominance controls yellow colouration predictions are that 100% F1 males, 75% F2 and 50% blue BC have yellow body colouration. Using the total number of individuals in each generation we estimated the expected number of individuals with and without yellow colouration and compared them against the observed number of individuals using chi-square tests.

Results

Repeatability of colour measurements

Repeatability of our measurements of red colouration was 0.87 in F2 generation and 0.88 in backcrosses. Pooled over both crosses a repeatability of 88% was obtained, suggesting a maximum for possible heritabilities around 88%.

Differences in male colouration

We found that the percentage of red cover was normally distributed in the *P. nyererei* parental line, but in *P. pundamilia* and all cross lines, it was skewed to the left, showing a large amount of individuals having no red and then successively smaller frequencies of individuals having some red cover (Fig. 2). Therefore all values were ln transformed for further analysis. After the transformation, all lines except the F_2 line (Shapiro-Wilk test, $P = 0.025$) conformed to normality.

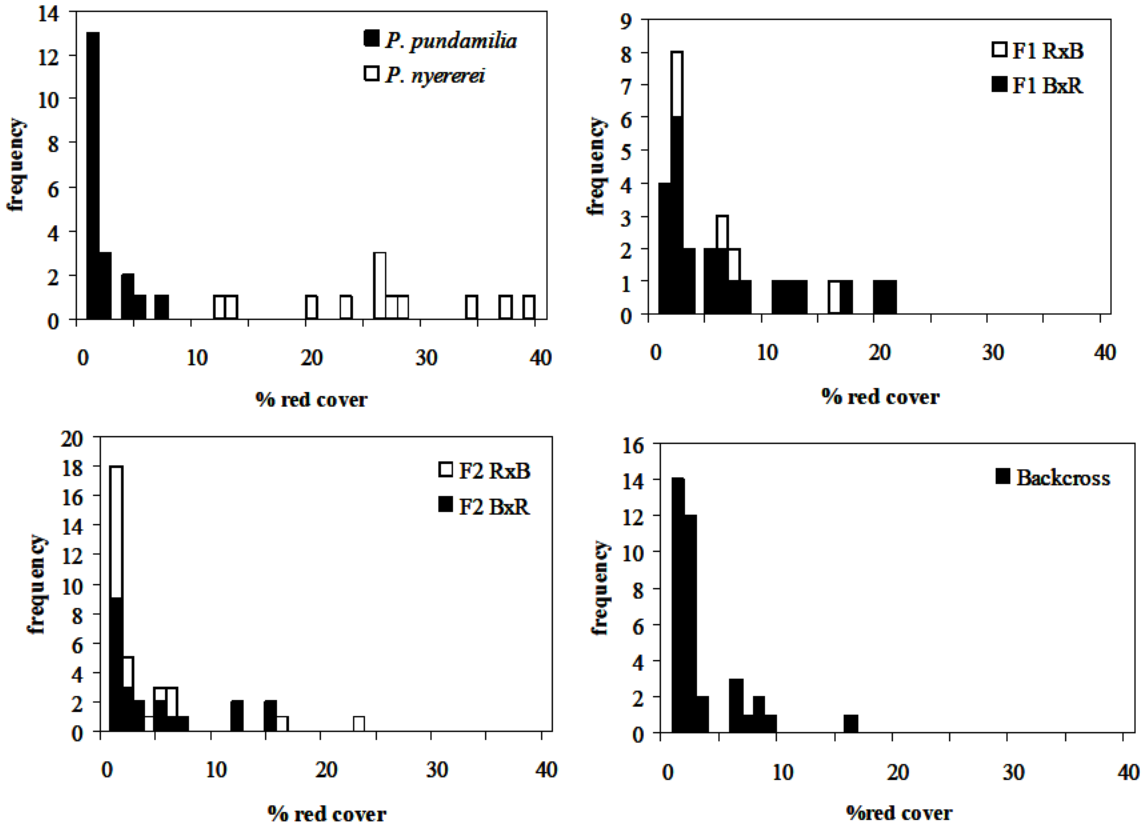

Figure 2. Distribution of male colour phenotypes (by % red cover of body) in (from top left to bottom right) parental lineages, F_1 hybrids, F_2 hybrids and backcrosses of blue F1 hybrid male to blue females.

The maximum red cover was 40% and was found in the *P. nyererei* parental class. The full range of phenotypes on the red side was not recovered in any of the line crosses. The F_1 and F_2 generations had a maximum red cover of 22% and 24% respectively.

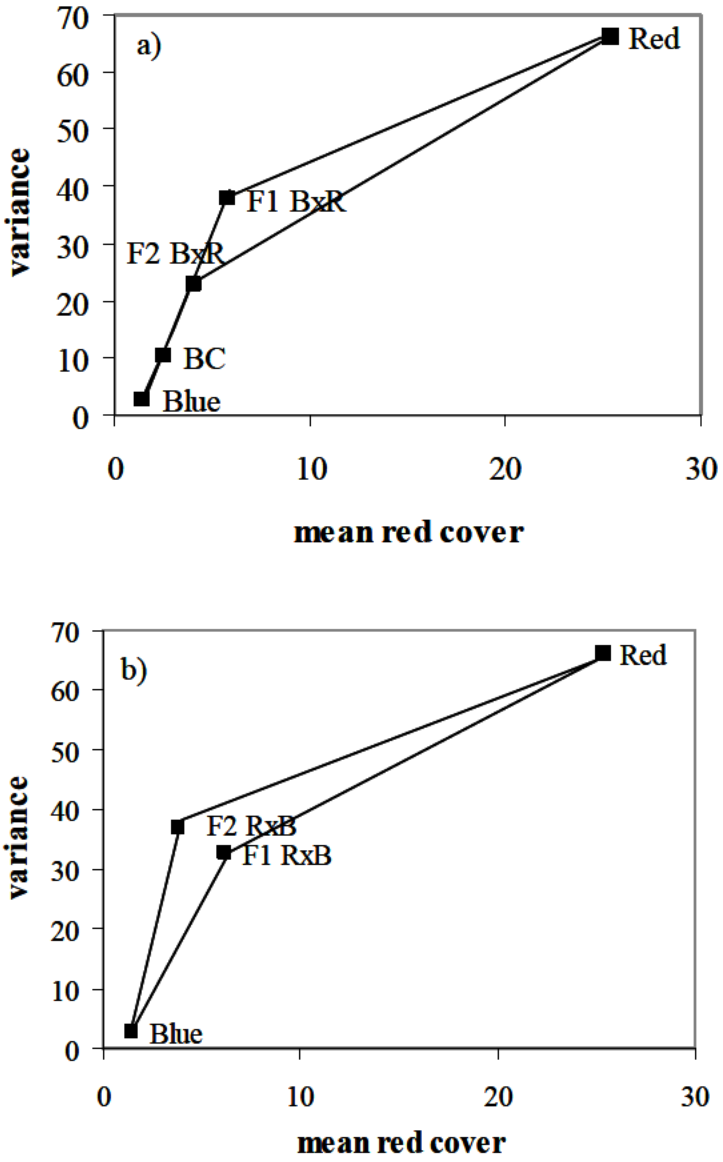

Figure 3. Plot of means of red cover against its variance for the 5 lines a) *P. nyererei* parental (Red), *P. pundamilia* parental (Blue) F_1 (BxR), F_2 (BxR) and backcross of blue phenotype F1 hybrid males to *P. pundamilia* females (BC); and b) Red, Blue, F_1 (RxB), F_2 (RxB).

The variance in red cover was significantly higher in the *P. nyererei* parental line than in *P.pundamilia* parental and backcross lines and the variance of the F1 was significantly higher than in *P.pundamilia* parental and backcross lines (Table 1, Fig 3a).

The parental *P. nyererei* generation had a significantly higher amount of red than all other lines (Table 1). The F1 and F2 generations had means of red cover intermediate to that of the two parental lines, but skewed towards *P. pundamilia*. The mean red cover of F1 males was significantly lower than that of *P. nyererei* males but significantly higher than that in the *P.pundamilia* and BC males. In the F2 line the mean red cover was significantly higher than that of *P. pundamilia* males but significantly lower than those of *P. nyererei* males. The direction of the parental cross had no effect on mean or variance of red cover of either F1 or F2 hybrids (Table 1; Fig. 3).

Table 1. Results of Levene's tests for equality of variances and of independent samples t-tests for equality of means. Significant values ($p<0.05$) are shown in bold face.

	Levene's Test for Equality of Variances			t-test for Equality of Means						
comparison	n1,n2	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Red vs Blue	12,20	13.177	**0.001**	10.048	11.552	**0.000**	23.903	2.379	18.698	29.109
Red vs F1	12,29	0.546	0.464	8.520	39	**0.000**	19.484	2.287	14.858	24.110
Red vs F2	12,50	3.548	0.064	12.128	60	**0.000**	21.789	1.797	18.196	25.383
Red vs blue BC	12,36	9.848	**0.003**	9.445	12.173	**0.000**	22.769	2.411	17.525	28.014
Blue vs F1	20,29	17.341	**0.000**	-3.772	33.973	**0.001**	-4.419	1.171	-6.800	-2.038
Blue vs F2	20,50	8.272	**0.005**	-2.717	67.105	**0.008**	-2.114	0.778	-3.667	-0.561
Blue vs blue BC	20,36	4.018	0.050	-1.463	54	0.149	-1.134	0.775	-2.688	0.420
F1 vs F2	29,50	2.422	0.124	1.870	77	0.065	2.305	1.233	-0.150	4.760
F1 vs blue BC	29,36	10.774	**0.002**	2.661	40.869	**0.011**	3.285	1.235	0.792	5.779
F2 vs blue BC	50,36	3.144	0.080	1.058	84	0.293	0.980	0.927	-0.863	2.823
F1 RxB-BxR	6,23	0.377	0.544	0.240	27	0.812	0.669	2.790	-5.057	6.394
F2 RxB-BxR	19,22	0.081	0.778	0.146	37	0.884	0.255	1.739	-3.269	3.779

In the *P. nyererei* parental generation every fish had yellow flanks, but in all other generations there were at least some individuals without yellow flanks (Fig. 4).

The *P. nyererei* parental generation had a significantly higher proportion of individuals with yellow flanks than the *P. pundamilia* parental and the blue backcross generations (Table 2). The proportions of individuals with yellow flanks in F1, F2 and backcross hybrids were intermediate to those of the parental lines. Among the hybrid generations, the largest proportion of individuals with yellow occurred in the F1 generation, followed by the F2 and the BC generations. Both F1 and F2 generations had significantly higher proportions of males with yellow than the BC and the *P. pundamilia* parental generations. The direction of the parental cross had no effect on proportion of individuals with yellow in either F1 or F2 hybrids.

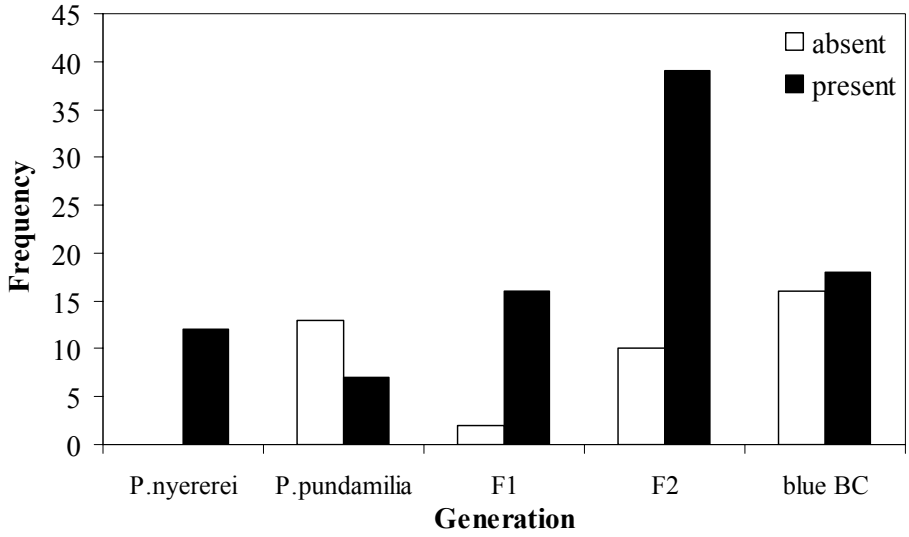

Figure 4. Frequency of individuals with yellow colouration present or absent in each generation.

Table 2. Results of binomial tests for equality of proportions of individuals with yellow colouration. Significant values ($p<0.05$) are shown in bold face.

comparison	n1,n2	X-squared	Sig.	df	95% Confidence Interval of the Difference	
Red vs Blue	12,20	10.580	**0.001**	1	0.374	0.926
Red vs F1	12,18	0.201	0.654	1	-0.104	0.326
Red vs F2	12,49	1.629	0.202	1	0.039	0.369
Red vs blue BC	12,34	6.709	**0.010**	1	0.246	0.695
Blue vs F1	20,18	9.370	**0.002**	1	-0.846	-0.232
Blue vs F2	20,49	10.781	**0.001**	1	-0.719	-0.173
Blue vs blue BC	20,34	0.989	0.320	1	-0.487	0.128
F1 vs F2	18,49	0.271	0.603	1	-0.129	0.315
F1 vs blue BC	18,34	5.225	**0.022**	1	0.095	0.624
F2 vs blue BC	49,34	5.446	**0.020**	1	0.039	0.494
F1 RxB-BxR	5,13	0.009	0.926	1	-0.181	0.488
F2 RxB-BxR	17,21	1.279	0.258	1	0.647	0.857

Additivity and Dominance

Joint-scaling showed that the red cover was neither adequately explained by a simple additive model nor by an additive plus dominance model (Table 3). The predicted means of the additive models were significantly different to the observed means ($\chi 2$ =17.805, d.f. =2, P = 0.0001) (Table 3a). When dominance was added to the model the fit was still poor with significant differences between the observed and predicted means ($\chi 2$ = 13.051, d.f. =1, P=0.0003) (Table 3b). However adding dominance significantly improved the fit of the model (Λ= 4.752, d.f. =1, P = 0.029).

Table 3. Joint-scaling test for the percentage of red cover (ln transformed). Sample means and means predicted by an additive and an additive + dominance model are shown with their standard error (SE). The estimated parameters for the modelled mean (μ_o), modelled additive component (α_c) and modelled dominance component (δ_c) are also shown.

line	*P.nyererei*	*P.pundamilia*	F1	F2
Sample mean	3.18	-0.19	1.18	0.50
SE	0.11	0.25	0.22	0.19
a) additive model				
Mean (additive model)	3.09	-0.74	1.17	1.17
SE	0.10	0.19	0.10	0.10
Parameter estimates: μ_o = 1.171, α_c = 1.915				
b) additive + dominance model				
Mean (additive + dominance model)	3.14	-0.43	0.80	1.08
SE	0.11	0.24	0.19	0.11
Parameter estimates: μ_o = 1.080, α_c = 1.788, δ_c = -0.275				

Estimation of minimum number of genes controlling red body colour

After ten iterations the values for σ^2_S and var(σ^2_S) stabilized at 0.77 and 0.17 respectively. These values were then inserted in equations 12 and 13 to obtain an estimate of the minimum number of loci and its variance. The Castle-Wright estimates indicated that a minimum number of two loci control differences in red cover ($n_E = 1.61$, var($\hat{n}_E$) = 3.12). Using Zeng's (1992) formula to take into account linkage and loci of unequal effect gave very similar values for when equal allelic effects were assumed ($C_\alpha = 0$: $n_E = 1.65$, var($\hat{n}_E$) = 3.75). The minimum number of loci remained similar when a normal distribution ($C_\alpha = 0.25$: $n_E = 1.82$, var($\hat{n}_E$) = 2.36) or a negative exponential distribution ($C_\alpha = 1$: $n_E = 2.30$, var($\hat{n}_E$) = 14.99) of allelic effects were assumed, but the variance increased. Only when a leptokurtic distribution of allelic effects was assumed was the minimum number of genes controlling red colouration higher than two ($C_\alpha = 4$: $n_E = 4.26$, var($\hat{n}_E$) = 93.6761).

Test for a model of one locus with complete dominance controlling yellow colouration

The observed proportions of males with yellow body did not differ significantly from the expected proportions under a model of one locus with complete dominance. In the F1 generation we observed 89% of individuals with yellow body compared to the 100% that were expected ($\chi2$ =0.53, d.f. =1, P = 0.47). In the F2 generation we observed 80% ($\chi2$ =0.09, d.f. =1, P = 0.76), and in the BC generation 53% of males with yellow bodies ($\chi2$ =0, d.f. =1, P = 1), compared to the expected 75% and 50% respectively.

Discussion

Results from the Wright-Castle estimator suggest that difference in male red colouration between the closely related and occasionally hybridizing cichlid species *P. pundamilia* and *P.nyererei* is heritable and determined by a minimum number of between two and four genetic loci. As a model of simple additivity did not fit our data well, two loci with additive effects appear not to be a likely explanation for the differences in mean red cover between the species. Considering a normal distribution or a negative exponential distribution of allelic effects could be more realistic. In these two cases, however, the variance increased considerably suggesting two is the minimum number of loci, but that in fact several more loci could be involved in the control of the differences in mean red cover. There may also be additional loci controlling the presence of red on the spinuous dorsal fin. However, the spinuous dorsal fin is generally blue unless the dorsal body surface is red, in which case the spinuous dorsal fin can be entirely red too. Additionally, the model of one locus with dominance appears to explain the presence/absence of yellow on the lower flanks.

We hybridized laboratory lines of these species that exhibited significantly different colouration with non-overlapping variance. Large variation found in the proportion of red colouration in our *P. nyererei* males indicates that the population may not be fixed for red alleles at all loci. Albeit much smaller, the variation observed in our *P. pundamilia* males was likely measurement error. It is clear from visual inspection of the photos that none of these males had any red colour on its flanks. Also, the existence

of *P. pundamilia* parental males with yellow flanks and the existence of F1 hybrids with no yellow on their flanks indicate that in both parental lines there were heterozygotes at the locus controlling yellow.

The mean red cover of males in the first and second hybrid generations and in the backcross generation was intermediate to those of males from the two parental generations. However the hybrids had a relatively low average percentage of red cover and therefore were closer to the mean of *P. pundamilia* than to that of *P. nyererei* in their red cover. On the other hand, except for a few entirely blue individuals, most F1 hybrid males were yellow on the lower flanks, resembling *P. nyererei*. Across all cross lines, the presence of red on flanks and dorsum was coupled to the presence of yellow flanks. Some individuals were yellow but had no red on the upper flanks, but none were red without having yellow on the lower flanks. This closely matched observations on wild populations with variable male colouration (Van der Sluijs et al., 2008).

The variances in red cover did not conform to a simple additive model. There are three lines of evidence for the non-linearity of gene effects: (1) several entirely blue phenotypes occur in most F_1 crosses; (2) the phenotypes of F_1, F_2 and back-cross males deviate from normality and are skewed towards blue; (3) the here reported back-crosses were derived from an entirely blue F1 hybrid male and different true-breeding blue *Pundamilia pundamilia* females. Yet they consistently produced some yellow and reddish phenotypes alongside blue phenotypes (Fig. 5). The mean trait value of these backcrosses was only slightly lower than that in the F_2 hybrids. These results are supported by the fact that the additive model made a very poor fit to the data. The indications of non-additive genetic architecture of colouration is consistent with results from other studies on the genetics of animal colour traits involved in species-specific sexual signalling such as male nuptial blue and yellow colouration in Lake Malawi cichlids (Barson et al., 2007) and plumage colour in birds (Mundy et al., 2004).

Adding dominance significantly increased the fit of the model, suggesting genes with dominant effects may contribute to variation in red colouration. However, observed and expected means were still significantly different, indicating that epistasis and /or linkage may also play a role in the genetics of red cover. Additionally, the observation that all fish with red colour on dorsal flanks have yellow lower flanks, whereas yellow lower flanks occur with or without red on dorsal flanks, suggests epistasis between the gene for yellow flank and the genes for red flank. Thus in total, we infer a minimum number of between 2 and 4 loci to explain the differences in male nuptial coloration between the species. Additional loci may be involved in explaining the variation in red cover within *P. nyererei*.

Figure 5. Male phenotypes in one of the BC families. The father was an entirely blue F1 hybrid male (top left), the mother a *P. pundamilia* female of the true-breeding line represented by the male top right. Yet, reddish males occurred in the family, clearly showing evidence for epistasis.

Models of sympatric speciation driven by sexual selection vary in the number of genes for the male sexual character that most strongly facilitate speciation. However, this number is often below five (Turner & Burrows, 1995; Arnegard, & Kondrashov, 2004; Gourbiere, 2004). The values obtained in our investigation varied around this number. Our results are therefore not inconsistent with values that might be permissive of sympatric speciation under the conditions of these models. However, the assumptions on the genetics of male colouration being controlled by genes with additive effects, most commonly applied in models of sympatric speciation by disruptive sexual selection are clearly not met by the observed genetics of this trait in a pair of fully sympatric sister species of Lake Victoria cichlids.

References

Albertson, R. C., Streelman, J. T., Kocher, T. D. 2003. Genetic basis of adaptive shape differences in the cichlid head. *J. Hered*. **94**: 291-301.

Albertson, R. C., Streelman, J. T. & Kocher, T. D. 2003b. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Nat. Acad. Sci. USA* **100**: 5252-5257.

Arnegard, M. E. & Kondrashov, A.S. 2004. Sympatric speciation by sexual selection alone is unlikely. *Evolution* **58**: 222-237.

Barson, N.J., Knight, M.E. & Turner, G. F. 2007. The genetic architecture of male colour differences between a sympatric Lake Malawi cichlid species pair. *J. Evol. Biol*. **20**: 45-53.

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* 411: 944-948.

Bush, G. L. 1994. Sympatric Speciation in Animals - New Wine in Old Bottles. *Trends Ecol. Evol*. **9**: 285-288.

Castle, W.E. 1921. Estimating the number of genetic factors concerned in blending inheritance. Am. Nat. **55**: 564-566.

Cavalli, L. L. 1952. *An analysis of linkage in quantitative inheritance. Quantitative inheritance*. 135-144. H. M. S. O., London.

Cockerham, C. C. 1986. Modifications in Estimating the Number of Genes for a Quantitative Character. *Genetics* **114**: 659-664.

Endler, J. A. & Mielke, P. W. 2005. Comparing entire colour patterns as birds see them. *Biol. J. Linn. Soc*. **86**: 405-431.

Endler, J. A. & Houde, A. E. 1995. geographic variation in female preferences for male traits in *Peocilia reticulata*. *Evolution* 49: 456-468.

Endler, J.A. 1992. Signals, signal conditions and the direction of evolution. *Am. Nat.* **139**: 125-153.

Gourbiere, S. 2004. How do natural and sexual selection contribute to sympatric speciation? *J. Evol. Biol.* **17**: 1297-1309.

Haesler, M.P. & Seehausen, O. 2005. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *P. Roy. Soc B-Biol. Sci*. **272**: 237-245.

Hayman, B. I., 1960a. The separation of epistatic from additive dominance variation in generation means. *Genetica* **31**: 371-390.

Hayman, B. I., 1960b. Maximum likelihood estimation of genetic components of variation. *Biometrics* **16**: 369-381.

Higashi, M., Takimoto, G. & Yamamura, N. 1999. Sympatric speciation by sexual selection. *Nature* **402**: 523-526.

Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J. W. 1996. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* 273: 1091-1093.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Rev. Gen*. **5**: 288-298.

Lande, R., Seehausen, O. & van Alphen, J. J. M. 2001. Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica* **112**: 435-443.

Lande, R. 1982. Rapid Origin of Sexual Isolation and Character Divergence in a Cline. *Evolution* **36**: 213-223.

Lande, R. 1981. The Minimum Number of Genes Contributing to Quantitative Variation between and within Populations. *Genetics* **99**: 541-553.

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusets.

Mather, K. & Jinks, J. L., 1982. *Biometrical genetics*. Chapman and Hall, N.Y.

Mayr, E. 1942. *Systematics and the origin of species.* Columbia University Press.

Messmer, V., Jones, G. P., van Herwerden, L. & Munday, P. L. 2005. Genetic and ecological characterisation of colour dimorphism in a coral reef fish. *Env. Biol. Fishes* **74**: 175-183.

Mundy, N. I., Badcock, N. S., Hart, T., Scribner, K., Janssen, K. & Nadeau, N. J 2004. Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science* **303**: 1870-1873.

Payne, R. J. H. & Krakauer, D. C. 1997. Sexual selection, space, and speciation. *Evolution* **51**: 1-9.

Schluter, D. 2000. Ecological character displacement in adaptive radiation. *Am. Nat.* **156**: 4-16.

Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., van der Sluijs, I., Schneider, M. V., Maan, M. E., Tachida, H., Imai, H. & Okada, N. 2008. Speciation through sensory drive in cichlid fish. *Nature* **455**: 620-626.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. Nat. Acad. Sci. USA* **271**: 1345-1353.

Seehausen, O., van Alphen, J. J. M. & Lande, R. 1999. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett*. **2**: 367-378.

Seehausen, O., van Alphen, J. J. M. & Witte, F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**: 1808-1811.

Seehausen, O. 1997. Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid. *Ecol. Freshw. Fish* **6**: 59-66.

Shaw, K. L. & Parsons, Y. M. 2002. Divergence of mate recognition behavior and its consequences for genetic architectures of speciation. Am. Nat. 159: 61-75.

Stager, J. C. & Johnson, T. C. 2008. The late Pleistocene desiccation of Lake Victoria and the origin of its endemic biota. *Hydrobiologia* **596**: 5-16.

Turner, G. F.,Seehausen, O.,Knight, M. E.,Allender, C. J. & Robinson, R. L. 2001. How many species of cichlid fishes are there in African lakes? Mol. Ecol. 10: 793-806.

Turner, G.F. & Burrows, M.T. 1995. A Model of Sympatric Speciation by Sexual Selection. *P. Roy. Soc B-Biol. Sci.* **260**: 287-292.

Van der Sluijs, I., Seehausen, O., Van Dooren, T. J. M. & Van Alphen, J. J. M. . No evidence for a genetic eassociation between female mating preference and male secondary sexual trait in a Lake Victoria cichlid fish. In prep.

van der Sluijs, I., van Alphen, J. J. M. & Seehausen, O. 2008. Preference polymorphism for coloration but no speciation in a population of Lake Victoria cichlids. *Behav. Ecol.* **19**: 177-183.

van Doorn, G.S., Dieckmann, U. & Weissing, F. J. 2004. Sympatric speciation by sexual selection: A critical reevaluation. *Am. Nat*. *163*: 709-725.

van Doorn, G.S., Noest, A.J. & Hogeweg, P. 1998. Sympatric speciation and extinction driven by environment dependent sexual selection. *P. Roy. Soc B-Biol. Sci*. **26**: 1915-1919.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol*. **16**: 381-390.

Wright, S. 1968. Evolution and the genetics of populations. University of Chicago Press.

Zeng, Z. B. 1992. Correcting the Bias of Wright Estimates of the Number of Genes Affecting a Quantitative Character - a Further Improved Method. *Genetics* 131: 987-1001.

Chapter 7

Summary and Conclusion

Summary and conclusion

In this thesis I have investigated mechanisms of speciation in the presence of gene flow using Lake Victoria haplochomine cichlids. Divergent sexual and ecological selection, both mediated by spatial and environmental heterogeneity, appear to be central to the origin and maintenance of diversity in this very young species flock. I aimed for a better understanding of how they work and on their outcomes.

Summary

The role of divergent and disruptive selection

While many articles focus on the likelihood of divergent sexual selection or divergent ecological selection each leading to sympatric speciation separately, few have studied both mechanisms in the same model to analyse the relative contribution of each of them, and the scope of their interaction (but see Lande & Kirkpatrick, 1988; van Doorn et al. 1998; Kirkpatrick & Ravigne, 2002; Gourbiere 2004). Nonetheless empirical studies on organisms such as land snails (Chiba 1999), sticklebacks (McKinnon & Rundle, 2002) are examples of species pairs that diverged in secondary sexual characters and show ecological character displacement at the same time, thus providing evidence for the role of the two types of selection. The most species rich vertebrate adaptive radiations are also characterized by species that diverged both in traits subject to ecological selection and traits subject to sexual selection (Streelman & Danley, 2003). However, a problem widespread in speciation research is that mechanisms of speciation are inferred from species differences that may have arisen during or after speciation. Such inference is problematic even when very young species are compared because ecological and sexual character displacement can both start already before speciation is completed. To identify the mechanisms that initiate speciation one would ideally study the same speciation event repeatedly through time. This will usually be impossible, but a related approach is to study replicate divergence events involving very similar phenotypes, at different stages of completion (Seehausen, in press). This thesis applies the latter approach to three different phenotypic polymorphisms that can characterize closely related and sympatric species of haplochromine cichlids, and can also occur as segregating polymorphisms within populations.

The female colour morphs of *N. omnicaeruleus* from Makobe Island are completely sympatric and syntopic, and appear to be stably coexisting over ecological time although ecological and eco-morphological differences are nearly absent. However, nearly no neutral genetic differentiation between colour morphs was found (chapter 2). This result repeated itself in a population from a closely related species, *N. greenwoodi* segregating a very similar polymorphism. The proposed mechanism for the stabilization of these polymorphisms entails that colouration is submitted not only to selection exerted by female and male mating preference (Seehausen et al., 1999), but also to intra-specific negative frequency-dependent sexual selection exerted by biased aggression of females towards their own morph (Dijkstra et al., 2008). That significant differentiation at neutral loci between the sympatric morphs has not evolved may be because the very low numbers of blotched males hinders the spread and fixation of female mating preferences for these novel male types (Seehausen et al., 1999), or because effects of disruptive sexual selection cannot overcome the homogenizing effects of gene flow in complete sympatry. There is however an indication of non-random mating between sympatric

colour morphs, manifested in significantly non-random genetic assignment and an excess of linkage disequilibrium in one of the morphs. This is the first confirmation with neutral markers that nonrandom mating in this system has an effect on gene flow, suggesting sexual selection may be important in maintaining this colour polymorphism. This same study also indicates potential for ecological character displacement among the colour morphs of *N. omnicaeruleus*, which constitutes evidence for the action of divergent ecological selection following or parallel to sexual selection.

The existence of trophic morphs within species of *Neochromis* corroborates the idea that ecological selection too maintains phenotypic diversity (chapter 3). At Makobe island, two eco-morphologically highly distinct morphs, *N.* "unicuspid scraper" and *N. omnicaeruleus* appear to be eco-morphs within a single species, despite evidence for non-random mating coming from non-random genetic assignment. At Igombe island, *N.* "unicuspid scraper" and *N. greenwoodi* diverge in some of the same eco-morphological and dental traits, but in this case the morphs are significantly differentiated at neutral genetic loci. While in *N.* "Bihiru scraper", at Bihiru island, phenotypic response to disruptive selection is only observed on one trait, tooth shape, and there is no evidence for neutral loci differentiation These examples provide evidence for disruptive ecological selection maintaining eco-morph differentiation in dentition, and divergent selection maintaining differentiation in head and jaw shape despite high gene flow. The results from these three replicate cases, presenting phenotypic divergence associated with variable extents of neutral genetic differentiation, are similar to patterns expected in studies of parallel speciation with replicates at different stages of speciation.

The populations of the sister species *Pundamilia pundamilia* and *Pundamilia nyererei* from Kissenda island provide evidence for the role of divergent and disruptive selection acting on several traits too. The two species diverge in male nuptial colouration. Divergence in male colouration is driven by disruptive sexual selection exerted by female mate choice, which is a by-product of adaptation to different light environments (chapter 4). Additionally, they also appear to diverge in several eco-morphological traits. Morphological divergence is partly accommodated by heritable adaptation and partly by phenotypic plasticity, where the latter appears to be adaptive by accentuating the differences between the two incipient species (chapter 5). Although the divergence at neutral loci indicates that there are restrictions to gene flow between the two incipient species at three of the five replicate cases investigated, they are at two of these extremely weak. Also in this system, the five replicate cases present different levels of divergence in several traits associated with variable genetic differentiation, representing different stages of completion of speciation.

Altogether these studies provide evidence for divergent and disruptive selection on several unrelated traits maintaining phenotypic divergence. The balance between gene flow and selection appears to dictate the extent of phenotypic divergence, with different strengths of selection with gene flow resulting in different levels of completion of speciation. However, in all cases, gene flow between morphs or incipient species appears to be too high for selection to lead to the completion of speciation.

The role of environmental heterogeneity

About one third of the shores of the mainland and islands in the southern part of Lake Victoria are made of rocks (Seehausen, 1996). In these areas the substrate can be

made of gravel, small rock gardens, boulder fields or steep rock walls, depending on the steepness of the slope of drowned shore (Seehausen, 1996). Important physical and biological microhabitat characteristics depend on the shore slope and boulder size. Water clarity also varies considerably between different sites in the lake. Inside the Mwanza Gulf secchi depth can be as low as 50 cm, while around offshore islands the water is relatively clear and secchi depth can exceed 300 cm. The most diverse communities of cichlids have been found in locations with clear water, in habitats with medium sized rocks and moderately steep slopes (Seehausen, 1996). Together, variation in spatial structure and water clarity generates important variation in selection pressures affecting the evolution and maintenance of haplochromine diversity. In two chapters of this thesis replicate cases of speciation in places that vary in environmental heterogeneity are directly compared.

The water clarity around Igombe, Makobe and Bihiru Islands is relatively similar. However, the steepness of shoreline slope and spatial linearity of the environmental gradient vary considerably. These factors appear to correlate well with the extent of phenotypic differentiation of eco-morphs observed at each island. Populations of the different species of the genus *Neochromis* inhabiting the shores of these islands appear to experience selection pressures of different strength on trophic morphology along environmental gradients, although genetic differentiation was only significant in one case (chapter 3). At Makobe, an island with a gently sloping and spatially linear resource gradient (the conditions most conducive of clinal speciation), there is evidence for correlations between multiple morphological variables and water depth at individual level. Although I found no evidence for depth segregation between colour morphs of *N. omnicaeruleus* (chapter 2), I found significant evidence for depth segregation between the trophic morphs of this species (chapter 3). The fact that trophic morphs diverge in dentition and in 8 morphological distances mostly related to resource use, and that variation in these traits is correlated with water depth suggest divergent ecological selection due to resource use may be acting along the environmental gradient. The lack of neutral genetic differentiation of the morphs suggests that phenotypic response to this selection was possible despite gene flow, while selection was apparently insufficient to cause speciation. At Igombe, an island with a slightly steeper environmental gradient than Makobe, there is evidence for the differentiation of eco-morphs in dentition and 4 morphological traits, but the correlation between trophic morphology and water depth is weaker. At this island there is evidence for weak but significant differentiation at neutral loci between the morphs, suggesting divergent ecological selection can cause incipient speciation. Finally, at the island with the steepest environmental gradient and a patchy mosaic-like habitat, Bihiru, I found phenotypic response to disruptive selection only on one trait, tooth shape, and no evidence for neutral loci differentiation.

Besides variation in shore line slope, boulder size and water clarity, the high organic content of Lake Victoria waters creates habitat heterogeneity in light regime with short wave lengths (blue light) being selectively absorbed and scattered in the upper part of the water column, resulting in a more red-shifted spectrum towards deeper water (Maan et al., 2006). As a result the visual system of haplochomine cichlids has undergone diversifying selection in a way that spectral sensitivities tend to match the ambient spectrum of the environment (Carleton & Kocher, 2001; Carleton et al, 2005). Therefore, perception of colours and their conspicuousness against the environmental background

changes throughout the water column. As a result female choice for conspicuous males may select for different male colours at different depths, leading to divergent selection on male nuptial colouration (Seehausen et al., 1997; Carleton et al, 2005; Maan et al., 2006).

The sister species pair *P. pundamilia* and *P. nyererei* are geographically fully sympatric. They inhabit locations that differ in light gradients primarily mediated by water depth. The populations from five of these locations were investigated (chapter 4). At the island with the shallowest light gradient analysed, Makobe, correspondence has been established between differences in depth distribution, wavelength sensitivity, male nuptial colouration, female mating preferences and ambient light spectrum between *P. pundamilia* and *P. nyererei*. There were strong bimodalities in male colouration and female preference, strong divergence in the gene controlling wave length sensitivity and significant differentiation between microsatellite loci. At the two islands with intermediate light gradients, Python and Kissenda, these associations become progressively weaker but are still significant. Finally at Marumbi and Luanso islands, which have very steep light gradients, the two species are replaced by single panmictic populations with very little variation in genes controlling visual sensitivity and no differentiation at neutral loci, although there was some variation in colour and mating preference.

Both studies on the genus *Neochromis* and on *Pundamilia* suggest that variation in environmental heterogeneity explain variation in the extent of phenotypic and genetic differentiation between incipient species. These studies consistently show that at locations with steep environmental gradients there is less diversity than at locations with shallow gradients. Therefore, steepening of environmental gradients, predicts loss of ecological and species diversity. From a biodiversity conservation point of view this is particularly relevant since eutrophication of Lake Victoria, partly due to human activities, is deteriorating water clarity and leading to an increase in the steepness of slopes of environmental gradients.

Conclusions

The main conclusion from this thesis is that divergent selection, (sexual or ecological) acting on several types of traits can maintain high levels of phenotypic diversity and differentiation despite high gene flow among Lake Victoria cichlids. Variation in the steepness of environmental gradients can explain much of the variation in the extent of phenotypic differentiation: better differentiation is associated with shallower gradients (chapter 3 and 5). Nonetheless, phenotypic differentiation was also observed in the absence of spatial differentiation (chapter 2). Neutral genetic differentiation, on the other hand appears to require the existence of environmental gradients. When there was no spatial differentiation, incipient species or morphs failed to show significant genetic differentiation (chapters 2, 3 and 4). Light and spatial gradients that create shallow and intermediate environmental gradients are associated with the largest extent of phenotypic and genetic differentiation indicating these are factors conducive to speciation.

Attempts to identify mechanisms that initiate species formation require information about the past that most of the times does not exist. We cannot completely rule out that phenotypic differentiation was initiated in an allopatric phase, however this scenario seems extremely unlikely since the incipient species and morphs studied in this

thesis diverge in ecologically relevant traits, which strongly indicates that phenotypic divergence was a direct result of ecological interaction. Weak or inexistent neutral genetic differentiation between incipient species also points to phenotypic divergence as a result of selection in the presence of gene flow being the cause behind the patterns observed presently. However, independently of how phenotypic differentiation was initiated, from an evolution and conservation point of view it is impossible to ignore the importance of ecological and divergent selection and environmental heterogeneity in maintaining diversity at present, particularly in young species that have only very recently diverged such as Lake Victoria cichlids. I hope my thesis provided evidence for this and that it has also provided a better understanding of these mechanisms and how their interactions contribute to the maintenance of biodiversity.

Future research

Each of the species complexes and polymorphisms analysed in this thesis deserves further investigation.

Future research on the *Neochromis* sex-linked polymorphism focusing on the identification of the number of genes controlling the sex-linked colouration and on their interactions could help further understanding the role sexual selection as the main mechanism maintaining phenotypic diversity and also why selection on colour has resulted in discrete polymorphisms with partial assortative mating but has at least so far not caused complete speciation. Also, further investigations of the indications of eco-morphological divergence of the WB morph of *N. omnicaeruleus* would be interesting in order to explore the hypothesis of ecological character displacement.

Some of the populations of males of the genus *Neochromis* that I analysed for the existence of trophic polymorphisms also appear to have subtle differences in colouration, mostly in the shades of blue in the body and in the amount red in the fins. Developing a method that could quantify these differences would allow exploring the possibility that not only divergent selection on eco-morphology but also divergent sexual selection on colouration are acting simultaneously in these populations. Further on, the understanding of this system of eco-morphological polymorphisms could benefit from more experimental quantitative genetics to characterize the genetic architecture of morphology and dentition.

The development of a linkage map and identification of the genes controlling colour, morphology and female mate preference, their location in the genome and the potential existence of linkage between them would greatly contribute to support the conclusions that so far have been made based on phenotypic analysis and quantitative genetics.

References

Butlin, R. K., Galindo, J. & Grahame, J. W. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos. T. Roy. Soc. B* **363**: 2997-3007.

Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O. 2005. Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Mol. Ecol.* **14**: 4341-4353.

Carleton, K. L. & Kocher, T. D. Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Mol. Biol. and Evol.* **18**: 1540-1550 (2001).

Chiba, S., 1999. Character displacement, frequency-dependent selection, and divergence of shell colour in land snails Mandarina (Pulmonata). *Biol. J. Linn. Soc.* **66**: 465-479.

Dijkstra, P. D., Seehausen, O. & Groothuis, T. G. G. 2008. Intrasexual competition among females and the stabilization of a conspicuous colour polymorphism in a Lake Victoria cichlid fish. *P. Roy. Soc. B-Biol. Sci*. **275**: 519-526.

Gourbiere, S. 2004. How do natural and sexual selection contribute to sympatric speciation? *J. Evol. Biol*. **17**: 1297-1309.

Lande, R. & Kirkpatrick M. 1988. Ecological speciation by sexual selection. J. Theor. Biol. **133**: 85-88.

Kirkpatrick, M. & Ravigne, V. 2002. Speciation by natural and sexual selection: Models and experiments. *Am. Nat.* **159**, S22-S35.

Kocher, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics* **5**: 288-298.

Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. 2006. Sensory drive in cichlid speciation. *Am. Nat.* **167**, 947-954.

McKinnon, J. S. & Rundle, H. D. 2002. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* **17**: 480-488.

Seehausen, O.. Progressive levels of trait divergence along a "speciation transect" in the Lake Victoria cichlid fish *Pundamilia*. In: *Ecology and Evolution*, (R. Butlin, D. Schluter and J. R. Bridle, eds.). Cambridge University press, in press.

Seehausen, O. 1996. *Lake Victoria Rock Cichlids. Taxonomy, Ecology and Distribution*: Zevenhuizen, NL: Cichlid Press.

Seehausen, O., van Alphen, J. J. M. & Lande, R. 1999b. Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecol. Lett.* **2**: 367-378.

Seehausen, O., vanAlphen, J. J. M. & Witte, F. 1997. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808-1811.

Streelman, J. T., & Danley P. D. 2003. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* **18**: 126-131.

Van Doorn, G. S., Noest, A. J. & Hogeweg, P. 1998. Sympatric speciation and extinction driven by environment dependent sexual selection. *P. Roy. Soc. B-Biol. Sci.* **265**: 1915-1919.

Affiliations of co-authors

Ole Seehausen
Aquatic Ecology and Macroevolution, Institute of Ecology and Evolution, University of Bern, Balzerstrasse 6, CH-3012 Bern, Switzerland; and Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland.

Salome Mwaiko
Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland; and Tanzanian Fisheries Research Institute, Mwanza Centre, PO Box 475 Mwanza, Tanzania.

Bänz Lundsgaard-Hansen
Aquatic Ecology and Macroevolution, Institute of Ecology and Evolution, University of Bern, Balzerstrasse 6, CH-3012 Bern, Switzerland; and Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland.

Maria Victoria Schneider
The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

Yohey Terai

Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan.

Karen L. Carleton

Department of Biology, University of Maryland, College Park, 20742 Maryland, USA.

Hillary D. J. Mrosso

Tanzania Fisheries Research Institute, Mwanza Centre, PO Box 475 Mwanza, Tanzania.

Ryutaro Miyagi

Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan.

Inke van der Sluijs

Department of Animal Ecology, Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands.

Current address: Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montréal, Québec, H3A 1B1, Canada

Martine E. Maan

Department of Animal Ecology, Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands; and University of Texas at Austin, Integrative Biology, 1 University Station C0930, Austin, Texas 78712, USA (M.E.M.).

Current address: EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland.

Hidenori Tachida

Department of Biology, Faculty of Sciences, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan.

Hiroo Imai

Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, 484-8506 Japan. 484-8506

Norihiro Okada

Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan.

Acknowledgements

A PhD project is definitely not an isolated work. I have been very fortunate for not only having worked on an extremely exciting project that I loved, but also for having met some great people while doing it. There are a lot of people that I would like to thank to for many different reasons.

First I would like to thank my PhD supervisor, Prof Dr. Ole Seehausen, for choosing me to do this great project, for very patiently teaching me almost everything I know until today about Evolutionary Biology and for showing me the wonderful world of cichlids.

I thank Prof. Dr. Michael Taborsky and Prof. Dr. Jerome Goudet for agreeing to act as my internal and external examiners, respectively.

I would like to thank all my colleagues (past and present) from the Fish Ecology and Evolution group: Alan Hudson, Pascal Vonlanthen, Salome Mwaiko, Denis Roy, Martine Maan, Marta Barluenga, Rudolf Muller, Severine Vuilleumier, John Hillary Mrosso, Vicky Schneider, Nellie Konijnendijk, Bänz Lundsgaard-Hansen, Pamela Woods, Kay Lucek, Erwin Schäfer, Nadja Pepe, Michelle, Andreas Taverna, Marcel Heasler, Etienne Besault, Rike Stelkens, Corinne Schmidt, Oliver Selz, Arjun Sivasundar for their help, discussions, comments, reviews on manuscripts, etc.

I would also like to thank some of my present and past colleagues from other groups of the EAWAG: Arne Hammrich, Scott Tiegs, Tonya del Sontro, Sara Schubert, Torsten Diem, Bine Sibler for their companionship, friendship and for the great parties at the Badehaus. Many thanks to everyone else at the EAWAG for the nice working environment.

I thank very much Kayeba Mhoja and Haluna Mohamed for the great work collecting all the fish, without which could not have done this work.

Many thanks as well to people from Hull University: Domino Joyce, Alan Smith, Paul Nichols, Pete Docker for always making me feel welcome there.

Many thanks to everyone from Leiden University for welcoming me there: Inke van der Sluijs, Machteld Verzijden, and Kees Ofker.

I thank all my friends from Portugal for making it so much fun to go back home and for helping me put my work as a scientist in perspective.

Finally, the most important people in my life and to whom I want to thank most is my family. I would like to thank my aunts Berta and Luisa for the weekly phone calls and all the advice and encouragement with my work. My two amazing sisters Sara and Diana were constant source of fun, entertainment, good advises and encouragement and for that I thank them both. Last but most important I thank my parents, the best parents in the world, who have always made me feel that wherever I am in the world I have a home to go back to; who gave me their unconditional support independently of my successes and failures throughout my PhD.

Erklärung

gemäss Art. 28 Abs. 2 RSL 05

Name/Vorname: .....

Matrikelnummer: .....

Studiengang: .....

Bachelor ☐ Master ☐ Dissertation ☐

Titel der Arbeit: .....

.....

.....

LeiterIn der Arbeit: .....

.....

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetztes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

.....

Ort/Datum

.....

Unterschrift

CURRICULUM VITAE

Name: Isabel Santos Magalhães
Date of birth: 17.02.1978
Place of Birth: Porto
Nationality: Portugal
Contact email: ismagalhaes@hotmail.com

Education:

2004-2009. PhD student at the Aquatic ecology and Macroevolution department at Bern University and at the Fishecology and Evolution department EAWAG, Switzerland: "The roles of divergent selection and environmental heterogeneity in speciation of Lake Victoria cichlid fish".

2002-2004. Masters in Environmental Sciences at Wageningen University, the Netherlands: "Development of defence mechanisms against predation in Daphnia in lake Berendonck, the Netherlands".

1996-2001. BSc in Biology at Porto University, Portugal, with specialisation in Applied Animal Biology.

Publications:

Magalhaes, I.S., Mwaiko S., Mrosso HDJ, Schneider M.V., Seehausen O. Disruptive selection and adaptive phenotypic plasticity during incipient speciation in Lake Victoria cichlid fish. *Jounal of Evolutionary Biology*.

Seehausen O., Terai Y., **Magalhaes I.S**., Carleton K.L., Mrosso H.D.J., Miyagi R., van der Sluijs I., Schneider M.V., Tachida H., Imai H., Okada N. Speciation through sensory drive in cichlid fish living on steep gradints of ambeint light. *Nature* 455: 620-625.

Magalhaes I. S., Lurling M., Roijackers R., Spaak, P.. Vertical distribution of Daphnia in Lake Berendonck (The Netherlands) during progressive hypolimnion oxygen depletion. Proceedings of the XXIX congress of the International Association of Theoretical and Applied Limnology (SIL).

Publications submited or in preparation:

Magalhaes I.S., Mwaiko S., Seehausen O. Sympatric colour polymorphisms associated with non-random gene flow in cichlid fish of Lake Victoria. Submited to *Journal of Evolutionary Biology*.

Magalhaes I.S., Lundsgaard-Hansen B., Mwaiko S., Seehausen O. **Eco-**morphological but not genetic differentiation within cichlid populations correlates with the slope of resource gradients. In prep.

Magalhaes I.S., Seehausen O. Genetics of male colour in a Lake Victoria cichlid species pair and its role in speciation. In prep.

Joyce, D.A., **Magalhaes I.S** & Genner, M.J. Sequential assessment on a bower building cichlid lek. In prep.

Conferences, workshops and courses attended:

Statistics course "Time series analysis" Dubendorf, Switzerland - 27.09.2007 and 04.10.2007

XII European Congress of Ichthyology, Cavtat (Dubrovnik), Croatia, 9-13.09.2007 **Oral presentation**

British Ecological Society Symposium "Speciation and Ecology", Sheffield, U.K. - 28-30.03.2007 **Poster**

Workshop "Adaptive vs. neutral genetic variability in conservation genetics" Tvärminne, Finland - 26.01.2007 and 29.01.2007 **Oral presentation**

Statistics course "Experimental design" Dubendorf, Switzerland - 18.01.2007 and 25.01.2007

Statistics course "Linear Regression" Dubendorf, Switzerland - 26.12.2006 and 30.12.2006

Statistics course "Introduction to R", Kastanienbaum, Switzerland - 06.07.2006 and 11.07.2006

EMBO course "Statistical methods for identification of genes under natural selection", Helsinki, Finland - 15.05 - 19.05.2006 **Poster**

Hybridization in Animals - extent, process and evolutionary impact (Symposium), Frankfurt, 12.-15.10.06. **Poster**

The Genetics of Speciation (Symposium), UBC Vancouver, 21.-24.07.2006. **Poster**

10th Congress of the European Society for Evolutionary Biology, Krakow, Poland, 15.-20.08.2005

XXIX SIL congress, 8-14 August 2004, Lahti, Finland. **Oral presentation**

Grants:

2008 SNSF grant for postdoctoral research at Hull University, UK, " duration: one year.

2001 Grant for internship at the Netherlands Institute For Sea Research (NIOZ) through the Europen programme "Leonardo da Vinci", duration: September 2001-June2002.

Awards

Third prize for best student oral presentation at XII EUROPEAN CONGRESS OF ICHTHYOLOGY, Cavtat (Dubrovnik), Croatia, 9-13.09.2007.

Language skills:

Portuguese (native speaker), English (written, read and spoken), French (read and spoken), Spanish (read and spoken), German and Dutch (basic knowledge).

Informatic skills:

Microsoft Office, SPSS, STATISTICA, Sigma Plot, Sigma scan, R.
Genetics software: Structure, Genetix, Beckman coulter software, Arlequin, Genepop, FSTAT, Convert, PAUP.

Field work:

2007 Behavioural studies and fish colection in Lake Malawi.
2005 Sample collection of cichlid fish in Lake Victoria, Tanzania.

2001 - 2002 Four scientific cruises, of one week each, in the North Sea for collection of invertebrates and box cores.
1998-2001 Participation in several bird ringing campaigns organised by the Institute for Nature Conservation) (ICN) in Portugal.

Teaching experience:
2006 - 2008 - teaching assistant for for several courses and practicals given by the department of Aquatic ecology & Macroevolution, at Bern University, Switzerland.
2007 - Supervision of a third year bachelor student during his Bachelor project at Bern University, Switzerland.
2003 –assistant student for the course "Principles of environmental sciences" given to first year MSc students in Wageningen University.

Other courses and activities:
2007 Nitrox certification.
2001 Certification for two star diver from CMAS.
2001 Participation in the course of "Basic life support" of the Portuguese Federation of Subaquatic Activities (FPAS).
2000 President of Porto's Youth Association for Science (AJC), Portugal.
1999 Participation in the course of monitors for environmental education in Portugal.
1999 Participation in the course of Scientific Ilustration in Portugal.
1998 Scuba diving certificate (CMAS).