doi:10.1111/j.1420-9101.2010.01960.x

Genetics of male nuptial colour divergence between sympatric sister species of a Lake Victoria cichlid fish

I. S. MAGALHAES*†[1] & O. SEEHAUSEN*†

*Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Balzerstrasse, Bern, Switzerland
†Fish Ecology and Evolution, Eawag Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biogeochemistry, Seestrasse, Kastaninenbaum, Switzerland

Keywords:

Castle-Wright estimator;
cichlids;
male colour;
sympatric speciation.

Abstract

The hypothesis of sympatric speciation by sexual selection has been contentious. Several recent theoretical models of sympatric speciation by disruptive sexual selection were tailored to apply to African cichlids. Most of this 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. We investigated the genetic architecture controlling male nuptial colouration in a sympatric sibling species pair of cichlid fish from Lake Victoria, which 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 two to four genes with significant epistasis and dominance effects. Yellow colouration appears to be controlled by one gene with complete dominance. The two colours appear to be epistatically linked. Knowledge on how male colouration segregates in hybrid generations and on the number of genes controlling differences between species can help us assess whether assumptions made in simulation models of sympatric speciation by sexual selection are realistic. In the particular case of the two sister species that we studied a small number of genes causing major differences in male colouration may have facilitated the divergence in male colouration associated with speciation.

Introduction

Animal colouration has long been recognized as an important trait in intraspecific and interspecific signalling (Endler, 1992). Colour polymorphisms are known to affect mate choice in Guppies from Trinidad (Endler & Houde, 1995), Australian Bowerbirds (Endler & Mielke, 2005), 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 disruptive sexual selection acting on secondary sexual characters, such as colouration, alone may drive population divergence and speciation in sympatry, even in the absence of ecological adaptation (Turner & Burrows, 1995; Payne & Krakauer, 1997; Higashi *et al.*, 1999; van Doorn *et al.*, 1998, 2004). The likelihood of this outcome, however, has been questioned as 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 simple additive genetics with few loci controlling traits involved in sexual isolation (reviewed by Gourbiere, 2004). Most models suggest that the likelihood of speciation decreases as the number of loci controlling sexual isolation increases.

Most of the models on speciation by disruptive sexual selection 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).

Correspondence: Isabel S. Magalhaes, Department of Biological Sciences, University of Hull, Cottingham Road, Hull HU6 7RX, UK.
Tel.: +44 1482 466034; fax: +44 1482 465458;
e-mail: I.Santos-Magalhaes@hull.ac.uk
[1]Present address: Department of Biological Sciences, University of Hull, Cottingham Road, Hull HU6 7RX, UK.

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; Maeda *et al.*, 2009). 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, 1999; Seehausen & Schluter, 2004). Well-studied examples 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 *P. nyererei* males have red dorsum and dorsal fin and yellow flanks, whereas *P. 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 or greyish. At places where the water transparency is high, the two phenotypes are ecologically and genetically distinct sister species with only the two extreme male colour phenotypes present, but at places with low water transparency intermediate male colour phenotypes are present and even dominate the population (Seehausen *et al.*, 2008). Also, females of both species display assortative mating preferences when they can see male nuptial colouration but not when the latter are masked by monochromatic light (Seehausen & van Alphen, 1998). 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). Both male colouration and female preference for it are main contributors to reproductive isolation of these sister species and are therefore considered speciation traits. Female mating preference differences between the species were 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.*, 2008), but the number of genes controlling male colour differences had so far not been estimated. 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 at an incipient stage of speciation. If differences in male colour are heritable and controlled by several genes with additive effects, then 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 in nature appears to correlate with the presence of red colouration, and tested for the hypothesis that the presence of yellow colour is controlled by one gene with complete dominance. Knowledge on the heritability of and genetic architecture underlying the difference in speciation traits is crucial to understand possible pathways of speciation and the relevance of assumptions of theoretical studies of speciation (Turelli *et al.*, 2001).

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. *Pundamilia 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 (three families) and vice versa (four 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. To acquire additional information on the heritability of colour, backcrosses were bred from a blue phenotype F1 hybrid male (*P. pundamilia* male × *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 6 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 (Hoya Corporation, Tokyo, Japan). 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 (Olympus Imaging Corp., Tokyo, Japan). These males were brought into standardized dominant territorial motivational state, assessed by the 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 × *P. nyererei* female and one family from *P. nyererei* male × *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 backcrosses and 26 F2s) on at least two different days.

Colour analysis

The salient differences in nuptial colouration 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., San Jose, CA, USA), 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, IL, USA). Using the function 'Define by colour', the delimiting criteria to select the area covered in red were defined by a combination of the parameters hue (colour) and saturation (richness of colour) (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 nonwhite 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; Fig. 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 performed without the knowledge of the cross type and on photoshopped images of the fish body without fins.

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_B^2 is the variance between individuals, and σ_W^2 is the variance within individuals (Becker, 1992). Repeatability is 1 when all variance is between individuals, whereas 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

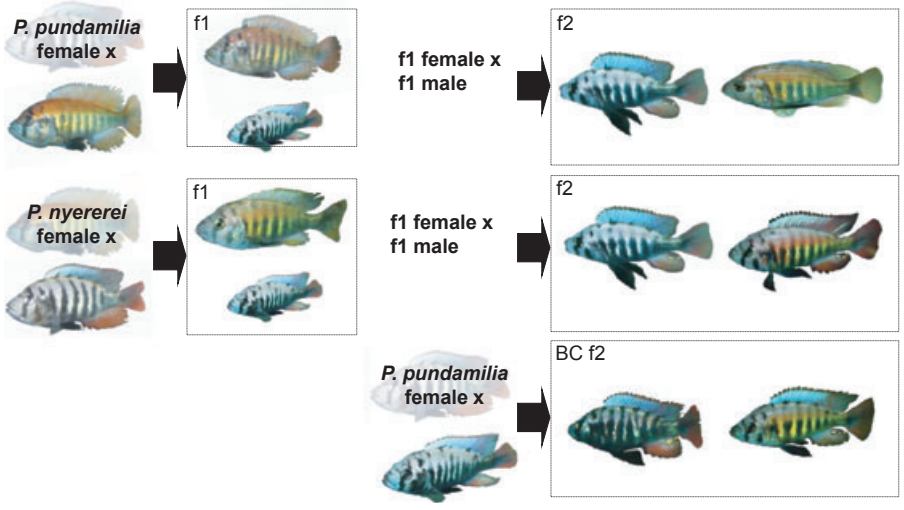

Fig. 1 Crossing scheme and colour patterns of parental and hybrid males. Paled-out images symbolize male colour of species of dam. Small images in the F1 boxes indicate the occurrence of a few entirely blue males in these otherwise rather homogeneous families of intermediate male colour.

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., Chicago, IL, USA). 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 (http://www.r-project.org), to test whether 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 ln transformed. As we only had backcrosses in one direction (to *P. pundamilia*), these were not included in the tests of additivity 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),

$$z_i = \mu_0 + M_{i2}\alpha + M_{i3}\delta + \in_i \qquad (1)$$

where z_i is the trait mean in the ith 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 ith 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 be discounted, including epistasis in the model requires more than six lines. We only used four 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:

$$\hat{a} = (M^T V^{-1} M)^{-1} M^T V^{-1} z \qquad (2)$$

and the predicted line means by each of the models are estimated by the equation:

$$\hat{z} = M\hat{a} \qquad (3)$$

where M is the matrix of coefficients of additive effects M_{i2} in the additive model:

$$M_{i2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \qquad (4)$$

and the matrix M_{i3} in the additive–dominance model:

$$M_{i3} = \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad (5)$$

V is the diagonal weighting matrix of sampling variances (squared standard errors) of the observed line means:

$$V = \begin{pmatrix} [SE(\mu_{P1})]^2 & 0 & 0 & 0 \\ 0 & [SE(\mu_{P2})]^2 & 0 & 0 \\ 0 & 0 & [SE(\mu_{F1})]^2 & 0 \\ 0 & 0 & 0 & [SE(\mu_{F2})]^2 \end{pmatrix} \qquad (6)$$

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:

$$\chi^2 = \sum_{i=1}^{k} \frac{(z_i - \hat{z}_i)^2}{SE(z_i)^2} \qquad (7)$$

where k is the number of observed lines, 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 performed by sequential model fitting (Lynch & Walsh, 1998). If the additive model is rejected, then dominance or epistasis effects also contribute to the difference between the lines. If the additive plus dominance model is also rejected, then 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 because of the segregation of the parental-line 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}, σ^2_{F1} and σ^2_{F2} in V by the new values obtained from eqn 10. This is performed until the values stabilize. The result of eqn 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 $(M^T V^{-1} M)^{-1}$.

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 outbred populations and their crosses. Cockerham (1986) added a correction for sampling error. We use the equation of Cockerham (1986) as:

$$n_E = \frac{(\mu_{P1} - \mu_{P2})^2 - \sigma^2_{P1} - \sigma^2_{P2}}{8\sigma^2_S} \quad (13)$$

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 earlier.

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 eqn 12. C_α is not known so we used a range of corrections to represent the effect of different underlying distributions of allelic effects: $C_\alpha = 0$ which considers all allelic effects equal; $C_\alpha = 0.25$ which considers a normal distribution of allelic effects; $C_\alpha = 1$ which assumes a negative exponential distribution, and $C_\alpha = 4$ 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.*, 2003).

The variance of the estimate was calculated by using the equation (Zeng, 1992):

$$\mathrm{var}(n_{e(zeng)}) = \frac{4\bar{c}^2 + (1 + C_\alpha)^2 \mathrm{var}(\hat{n}_e)}{[1 - \hat{n}_e(1 - 2\bar{c})]^4} \qquad (17)$$

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 back crosses (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 the 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.

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.

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).

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.

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 ($\Lambda = 4.752$, d.f. = 1, $P = 0.029$).

Estimation of minimum number of genes controlling red body colour

After ten iterations, the values for σ_S^2 and $\mathrm{var}(\sigma_S^2)$ stabilized at 0.77 and 0.17, respectively. These values were then inserted in eqns 13 and 14 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 [$n_E = 1.61$, $\mathrm{var}(\hat{n}_E) = 3.12$]. 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 [$C_\alpha = 0$: $n_E = 1.65$, $\mathrm{var}(\hat{n}_E) = 3.75$]. The minimum number of loci remained similar when a normal distribution [$C_\alpha = 0.25$: $n_E = 1.82$, $\mathrm{var}(\hat{n}_E) = 2.36$] or a negative exponential distribution [$C_\alpha = 1$: $n_E = 2.30$, $\mathrm{var}(\hat{n}_E) = 14.99$] of allelic effects was 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 [$C_\alpha = 4$: $n_E = 4.26$, $\mathrm{var}(\hat{n}_E) = 93.6761$].

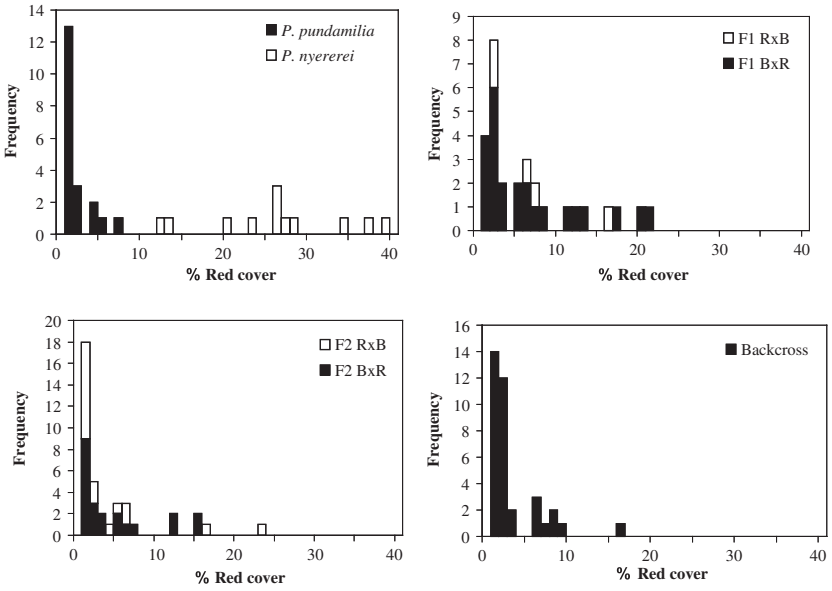

Fig. 2 Distribution of male colour phenotypes (by % red cover of body) in (from top left to bottom right) parental lineages, F_1 hybrids, F_2 hybrids and backcrosses of blue F1 hybrid male to blue females.

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						
									95% confidence interval of the difference	
Comparison	n1,n2	F	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std. error 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 R × B-B × R	6,23	0.377	0.544	0.240	27	0.812	0.669	2.790	−5.057	6.394
F2 R × B-B × R	19,22	0.081	0.778	0.146	37	0.884	0.255	1.739	−3.269	3.779

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 ($\chi^2 = 0.53$, d.f. = 1, $P = 0.47$). All F1 individuals with no yellow present in their body were from *P. pundamilia* male × *P. nyererei* female families. In the F2 generation, we observed 80% ($\chi^2 = 0.09$, d.f. = 1, $P = 0.76$); and in the BC generation, 53% of males with yellow bodies ($\chi^2 = 0$, d.f. = 1, $P = 1$), compared to the expected 75% and 50%, respectively.

Discussion

The Castle-Wright method of estimating gene number gave a narrow range of estimates for male red colouration that suggest that differences between *P. pundamilia* and *P. nyererei* are determined by the action of 2–4 loci. The differences in yellow colouration are most likely determined by one single locus with dominance. We will first

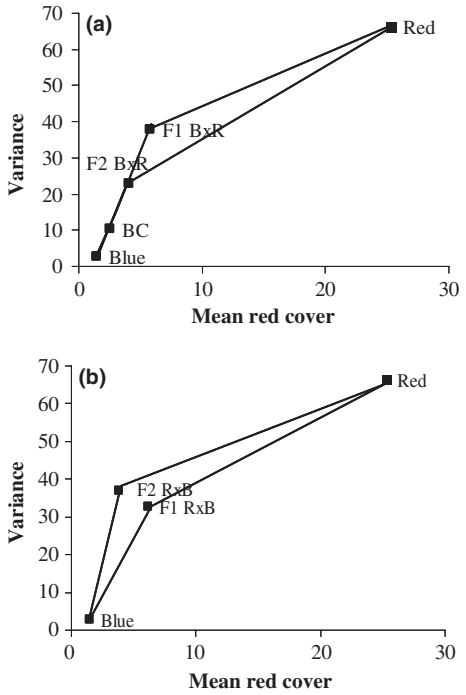

Fig. 3 Plot of means of red cover against its variance for the five lines (a) *Pundamilia nyererei* parental (Red), *Pundamilia pundamilia* parental (Blue), F_1 (B × R), F_2 (B × R) and backcross of blue phenotype F1 hybrid males to *P. pundamilia* females (BC); and (b) Red, Blue, F_1 (R × B), F_2 (R × B).

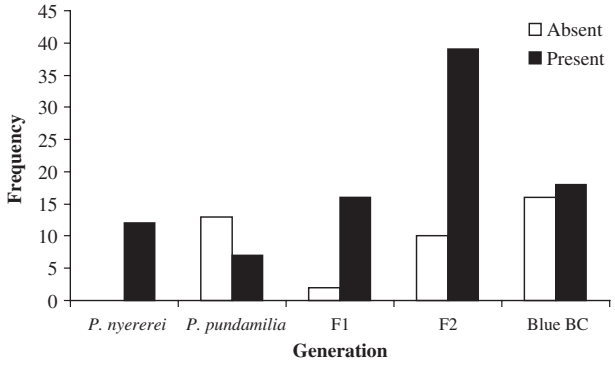

Fig. 4 Frequency of individuals with yellow colouration present or absent in each generation.

discuss the potential weaknesses and strengths in our data and then discuss the implications of our findings, in the light of speciation models.

Heritability and variation in colouration

We crossed laboratory lines of *P. pundamilia* and *P. nyererei* that exhibited significantly different male colouration with nonoverlapping variance. Our parental lines were maintained through several generations before F1 hybrids were produced. Nonetheless, large variation found in the proportion of red colouration in our *P. nyererei* males indicates that the extent of red colouration in *P. nyererei* may have a nonheritable plastic component (Seehausen *et al.*, 1997), but possibly also that the population may not be fixed for red alleles at all loci. Both sources of variation could have affected downwards our estimate of the number of genes. One of the assumptions of the Castle-Wright estimator is additive positive effects fixed in one parental line and additive negative effects in the other, but as the Lande (1981) extended equation accommodates outbred populations with parental lines not fixed only the effects of

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 R × B-B × R	5,13	0.009	0.926	1	−0.181	0.488
F2 R × B-B × R	17,21	1.279	0.258	1	0.647	0.857

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	*Pundamilia nyererei*	*Pundamilia 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: $\mu_o = 1.171$, $\alpha_c = 1.915$				
(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: $\mu_o = 1.080$, $\alpha_c = 1.788$, $\delta_c = -0.275$				

nonheritable components are not accounted for in our estimations. 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. The existence of *P. pundamilia* parental line males with yellow flanks and the existence of F1 hybrids without yellow on their flanks indicate that some *P. pundamilia* males and *P. nyererei* females of the parental lines were heterozygous 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).

Estimation of number of genes

Results from the Castle-Wright estimator suggest that the 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. The Castle-Wright estimator uses several assumptions, namely additivity of genes, normality of data in each line and statistically independent data. To meet some of the assumptions, which were initially not met by our data, we normalized the data and used the correction to the estimator that takes into account linkage and loci of unequal effect (eqn 15, Zeng, 1992). For reliable estimations, the Castle-Wright estimator also requires a large sample size from each line and generation: at least 20–30 of the parental and F1 populations and around 100 or more of the F2 and backcross generations (Lande, 1981). Although our sample size for the F1 and parental generations is close to the required, the F2 and backcross generations are not. Small sample sizes will contribute to increase the variances and may therefore lead to a downwards bias in the estimated number of genes. Therefore, we interpret the estimates obtained as a minimum number of genes. We further wish to stress that the objective of our analysis is not to provide the exact number of genes that control differences in the trait, but to estimate it to the nearest order of magnitude.

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 largely explain the presence/absence of yellow on the lower flanks.

Evidence for nonlinearity of gene effects

The variances in red cover did not conform to a simple additive model. There are three lines of evidence for the nonlinearity of gene effects: (i) several entirely blue phenotypes occur in most F_1 crosses; (ii) the phenotypes of F_1, F_2 and backcross males deviate from normality and are skewed towards blue; (iii) the here reported backcrosses were derived from an entirely blue F1 hybrid male and several different true-breeding blue *P. 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 nonadditive genetic architecture of colouration are 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).

Fig. 5 Male phenotypes in one of the BC families. The father was an entirely blue F1 hybrid male (top left), the mother a *Pundamilia 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.

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 colouration between the species. Additional loci may be involved in explaining the variation in red cover within *P. nyererei*.

Implications for speciation models

Models of sympatric speciation by sexual selection vary in the number of genes for the male sexual character at which speciation is most strongly facilitated. However, this number is often below five in models with nonecological sexual selection (Turner & Burrows, 1995; Arnegard & Kondrashov, 2004; Gourbiere, 2004). On the other hand, in a model of sensory drive speciation in which natural selection on the visual system and sexual selection on male colour interact (Kawata *et al.*, 2007), speciation was most probable when male colour was either polygenic or determined by one major locus with epistasis plus few additional loci with small effect. There is some evidence that the *Pundamilia* species pair that we studied here has indeed speciated by the interaction of natural selection on the visual system and sexual selection on male colouration (Seehausen *et al.*, 2008). Our best estimate of the genetic architecture of male breeding colouration in this species pair, i.e. relatively few genes with epistasis is therefore not inconsistent with assumed genetic architectures that are permissive of sympatric speciation in models.

Our study adds to existing evidence from other empirical studies for nonlinearity of gene effects on traits that are directly or indirectly involved in interspecific mate choice (Mundy *et al.*, 2004; Barson *et al.*, 2007). Knowing how male colouration segregates in hybrids and its underlying genetic architecture can also help predict patterns of gene flow between species in the wild. To the extent that F1 hybrid males resemble *P. pundamilia* males better, they may have a higher probability of mating with *P. pundamilia* females, leading to asymmetrical introgression between the two species. There is indeed existing evidence from hybridising wild populations of these two species that blue phenotypes are more introgressed than red phenotypes (Magalhaes *et al.*, 2009).

Conclusion

Our data suggest that the differences in male colouration in a sister species pair of Lake Victoria cichlids are controlled by a relatively small number of genes, which is permissive of sympatric speciation by disruptive sexual selection on male colouration. Effects of epistasis and dominance may facilitate speciation as they are consistent with an increased likelihood of speciation in models of sensory drive speciation (Kawata *et al.*, 2007). Our empirical indications of epistasis and dominance may also help explain patterns of asymmetric introgression found in wild populations. Overall, we hope our data will be helpful for parameterizing increasingly realistic speciation models for explaining the unusual rates of nonallopatric speciation observed in African cichlid fish.

Acknowledgments

We thank Nicky Barson for help with the methods. We also thank Martine Maan and the members of the EAWAG 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 attributed to Ole Seehausen.

References

Albertson, R.C., Streelman, J.T. & Kocher, T.D. 2003b. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Natl. 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.

Becker, W.A. 1992. *Manual of Quantitative Genetics*. Academic Enterprises, Pullman, WA.

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

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*, pp. 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.

van Doorn, G.S., Noest, A.J. & Hogeweg, P. 1998. Sympatric speciation and extinction driven by environment dependent sexual selection. *Proc. R. Soc. Lond. B Biol. Sci.* **26**: 1915–1919.

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

Endler, J.A. 1992. Signals, signal conditions and the direction of evolution. *Am. Nat.* **139**: 125–153.

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. & Mielke, P.W. 2005. Comparing entire colour patterns as birds see them. *Biol. J. Linn. Soc.* **86**: 405–431.

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. *Proc. R. Soc. Lond. 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.

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 Evol. Biol.* **7**: 99.

Kocher, T.D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5**: 288–298.

Lande, R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**: 541–553.

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.

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusets.

Maeda, K., Takeda, M., Kamiya, K., Aibara, M., Mzighani, S.I., Nishida, M., Mizoiri, S., Sato, T., Terai, Y., Okada, N. & Tachida, H. 2009. Population structure of two closely related pelagic cichlids in Lake Victoria, *Haplochromis pyrrhocephalus* and *H. laparograma*. *Gene* **441**: 67–73.

Magalhaes, I.S., Mwaiko, S., Schneider, M.V. & Seehausen, O. 2009. Divergent selection and phenotypic plasticity during incipient speciation in Lake Victoria cichlid fish. *J. Evol. Biol.* **22**: 260–274.

Mather, K. & Jinks, J.L. 1971. *Biometrical Genetics*. Chapman and Hall, N.Y.

Messmer, V., Jones, G.P., van Herwerden, L. & Munday, P.L. 2005. Genetic and ecological characterisation of colour dimorphism in a coral reef fish. *Environ. 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.

Seehausen, O. 1997. Distribution of and reproductive isolation among color morphs of a rock-dwelling Lake Victoria cichlid. *Ecol. Freshw. Fish.* **6**: 59–66.

Seehausen, O. & Schluter, D. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. Natl. 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). *Beh. Ecol. Sociobiol.* **42**: 1–8.

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. & 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., 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.

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.

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.

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. *Proc. R. Soc. Lond. 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.

Wright, S. 1968. *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago.

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.

Received 18 June 2009; revised 5 January 2010; accepted 26 January 2010