2 October 2008 | www.nature.com/nature | £10

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature

HIV/AIDS
The long road to vaccines

SPINTRONICS
Diamonds make sense

STEM CELLS
Let's get clinical

SPECIATION IN COLOUR

A textbook example of evolution in action

NATUREJOBS
Biotech & Pharma

9 770028 083095 40

nature International weekly journal of science

About the cover

Speciation in colour: a textbook example of evolution in action The cichlid fish of African lakes are textbook examples of rapid speciation but the mechanisms involved remain elusive. Observations of the cichlids in Lake Victoria now demonstrate the ecological and molecular basis of divergent evolution of the visual system (seen as divergence of vision genes, male coloration and female preferences) leading to speciation by sensory drive through interacting natural and sexual selection. The sensory drive hypothesis predicts that divergent adaptation in sensory and signalling systems to different environments can cause premating isolation between populations. As well as providing clear evidence that speciation can occur through sensory drive without geographical isolation, this work provides a mechanistic explanation for the collapse of cichlid fish species diversity during the anthropogenic eutrophication of Lake Victoria. On the cover, the colourful cichlids *Pundamilia nyererei* and *Pundamilia pundamilia*, as well as a wild intermediate (top) with the light spectra from different water depths in the background. [Article p. 620; News & Views p. 601; www.nature.com/podcast] Cover images: Ole Seehausen & Inke van der Sluijs.

ARTICLES

Speciation through sensory drive in cichlid fish

Ole Seehausen[1,2], Yohey Terai[3], Isabel S. Magalhaes[1,2], Karen L. Carleton[4], Hillary D. J. Mrosso[5], Ryutaro Miyagi[3], Inke van der Sluijs[6]†, Maria V. Schneider[2]†, Martine E. Maan[6]†, Hidenori Tachida[7], Hiroo Imai[8] & Norihiro Okada[3]

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: speciation 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.

The sensory drive hypothesis for speciation[1,2] 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[3,4] and speciation[5], 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[6]. Here we provide ecological, population genetic and molecular evidence for each of the predictions of sensory drive speciation[2] in sympatric cichlid fish inhabiting light gradients in Lake Victoria (East Africa).

Lake Victoria is spatially highly heterogeneous in water clarity and ambient light[7,8], 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[9]. Vertebrate visual pigments consist of a light-absorbing component, the chromophore, and a protein moiety, the opsin[10]. 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[11]. Eight different visual pigments have been found in all haplochromine cichlids[12–14], but only a subset of these is expressed in any individual species[12,14,15]. 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[16]. Of these, the *LWS* opsin gene is by far the most variable among Lake Victoria cichlids[13,17], with sequence variation being five times greater than in Lake Malawi cichlids despite a tenfold greater age of the latter species flock[18].

Female Lake Victoria cichlids have mating preferences for conspicuously coloured males[19]. Perception of conspicuousness is influenced by ambient and background light, signal transmission, receiver sensitivity and higher level processing[2,20]. 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[8]. 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[16,21], with a lower detection threshold for red but a higher one for blue light[22,23]. 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[1,2,16,23].

Examining the role of environmental gradients in speciation requires tests at replicate gradients, as is recognized both in evolutionary ecology[24–26] and in population genomics[27]. A recent model of clinal speciation through sensory drive[5], as well as other models of clinal speciation[28–30], 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[31] 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[28–30]. 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

[1]Institute of Zoology, University of Bern, Baltzerstr. 6, CH-3012 Bern, Switzerland. [2]Eawag, Swiss Federal Institute for Aquatic Science and Technology, Centre of Ecology, Evolution & Biogeochemistry, Department of Fish Ecology & Evolution, 6047 Kastanienbaum, Switzerland. [3]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. [4]Department of Biology, University of Maryland, College Park, Maryland 20742, USA. [5]Tanzania Fisheries Research Institute, Mwanza Centre, PO Box 475 Mwanza, Tanzania. [6]Department of Animal Ecology, Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands. [7]Department of Biology, Faculty of Sciences, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan. [8]Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, 484-8506 Japan. †Present addresses: Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montréal, Québec H3A 1B1, Canada (I.v.d.S.); The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK (M.V.S.); University of Texas at Austin, Integrative Biology, 1 University Station C0930, Austin, Texas 78712, USA (M.E.M.).

between islands[9]. *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[9].

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 nyererei*[32] (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[33,34]. The red *P. nyererei* occurs at greater mean water depths, in more red-shifted ambient light than the blue *P. pundamilia*[23]. *P. nyererei* have a lower threshold for the detection of red light, whereas *P. pundamilia* possess a lower threshold for detection of blue light[23]. Earlier we found that red and blue fish tended to possess different alleles at the *LWS* opsin gene locus[16]. Here we fully develop this system to test predictions of sensory drive speciation.

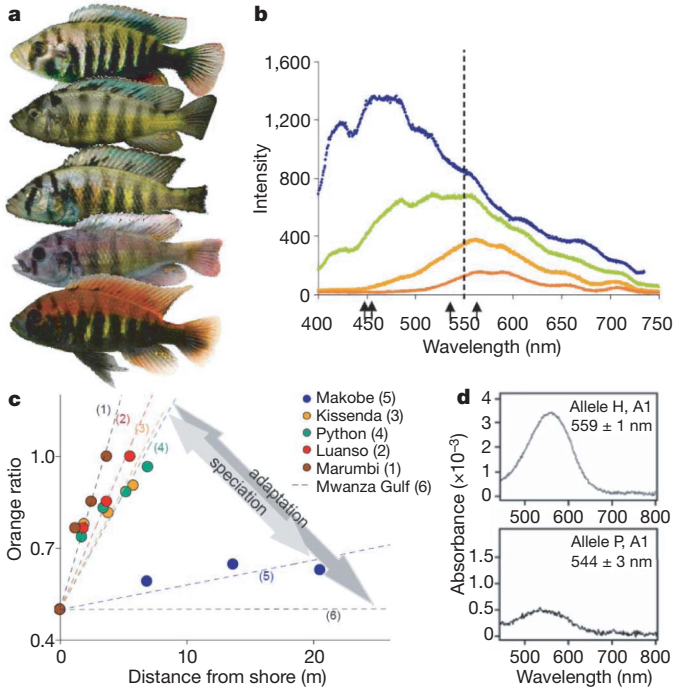

Figure 1 | Male phenotypes, light gradients and *LWS* opsin absorbance. **a**, Variation in male nuptial colouration. Five phenotype classes from 0 ('blue', typical *P. pundamilia*; top) to 4 ('red', typical *P. nyererei*; bottom). **b**, An example of a moderately steep light gradient (Python island): 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. The line through 550 nm indicates the 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[16]. **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[9]. 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.

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[9] (see Supplementary Information).

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[35], causing successive shifts of ambient light towards longer wavelengths with increasing water depth (this study), and also with increasing turbidity (earlier study)[7,8]. The rate at which ambient light changes with increasing depth is positively correlated with turbidity[8] (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 islands (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 our earlier work[9]. 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).

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.

Table 1 | The five environmental gradients of this study

Island	Water clarity (cm Secchi) (mean ± s.d.)	Shoreline slope (mean ± s.d.)	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

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[36] 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/

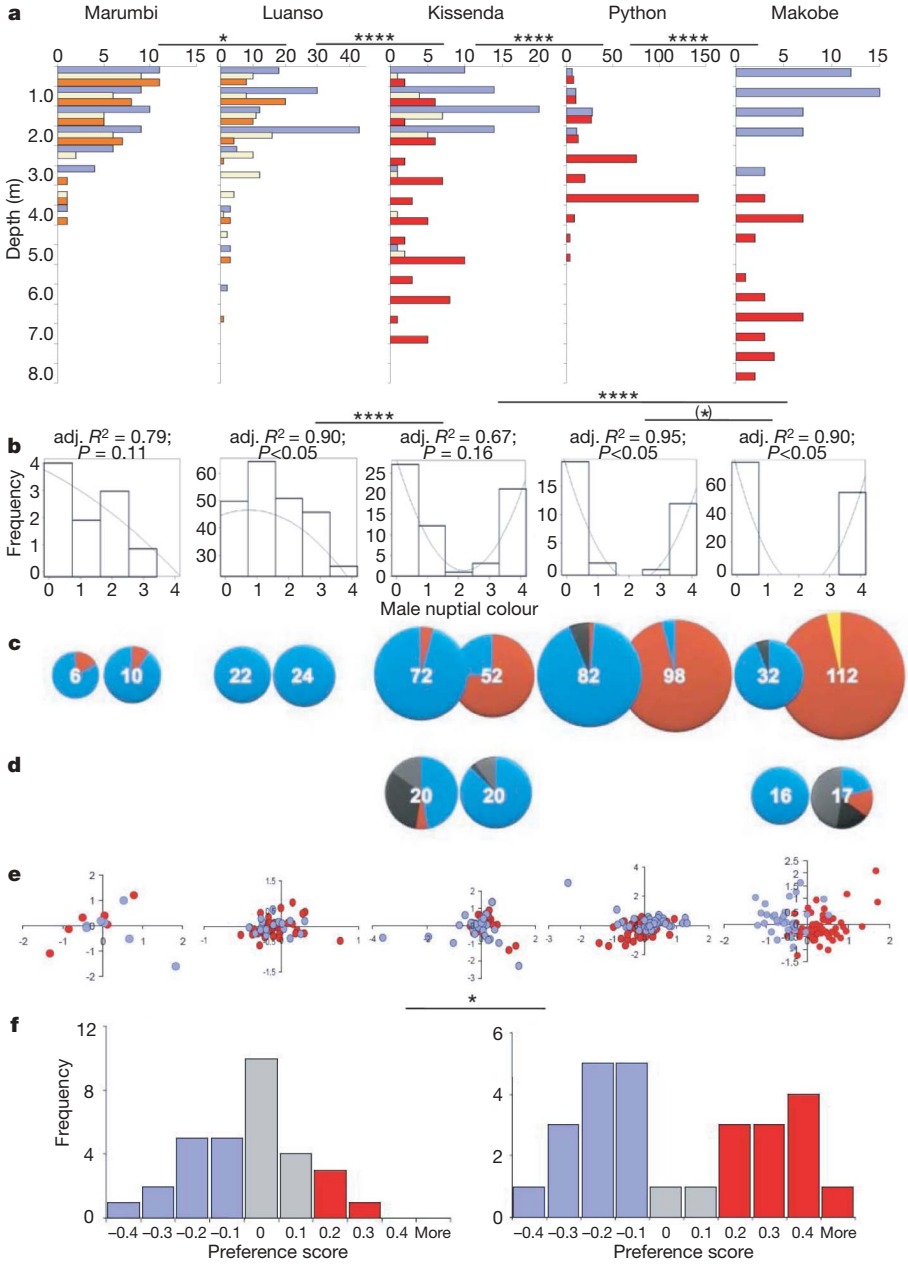

Figure 2 | Ecological, phenotypic, genetic and behavioural differentiation between blue and red *Pundamilia* nuptial phenotypes at five islands. All data for the same island are presented in the same column. Significant differences between islands indicated by asterisks (all tests two-tailed): $*P < 0.05$, $****P < 0.0001$, $(*)P < 0.1$. **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 are included that could be assigned to 'blueish' and 'reddish' phenotypes (altogether 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 the *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[41] (left) and Python island[40] (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* value of a *G*-test.

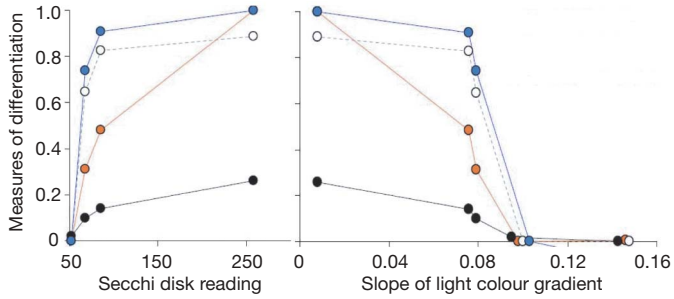

Figure 3 | Measures of differentiation between sympatric *Pundamilia* phenotypes plotted against water transparency (left) and light slope (right). Blue symbols and line: Spearman rank correlations between colour and *LWS* genotype (best fit to water clarity $R^2 = 0.79$, $P = 0.045$, d$f = 4$; best 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$f = 4$).

green opsin[37]. Focusing on these three positions, we divided alleles into three groups described previously[9]: 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[9] (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 measured previously by microspectrophotometry, reporting too that P pigments were blue-shifted relative to H pigments[16]. 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 ($\chi^2 > 66$, d$f = 1$, $P < 0.0001$; 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). 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 microsatellite loci carrying the signature of differentiation increases steadily from Marumbi and Luanso (0 out of 11) to Makobe island (7 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

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	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.

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.

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[39] (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[19,33,34]. 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[40,41]. Combining published data[40,41] 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 ($\chi^2 = 22$, $df = 10$, $P = 0.03$, 10,000 randomizations). However, this relationship was not significant when restricted to F_2 hybrid females ($\chi^2 = 10.2$, $df = 6$, $P = 0.13$, 10,000 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[40]. 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[16]. 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 λ_{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 panmictic 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[9]—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[5,28–30,42]. In contrast, when migration exceeds selection, divergence cannot occur either[43,44]. This explains the absence of speciation on the steepest of our gradients.

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[8]. 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.

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[16]	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"[16]	Zue island	Red chest	0	0	100	0	6
Total							548

* Hybridizing populations (neither *P. pundamilia* nor *P. nyererei*, but the hybridizing blue (*P. pundamilia*-like) and red (*P. nyererei*-like) populations shown in Fig. 2 (this study)).
† *n* represents *n* haplotypes sequenced.

METHODS SUMMARY

Ambient, absorbance and transmittance light spectra were measured with an Ocean Optics PS 1000 spectrophotometer and a 100 μm optical fibre, in the shade between 8:50 and 9:00 in the morning. We calculated the 'transmittance orange ratio' as the ratio of transmittance in the 550–700 nm range over the total visible range. The 'light slope' was obtained by regressing the transmittance orange ratio against distance (m) from the shore along the lake floor. Male fish in breeding colouration were collected by angling and netting; 480 males were photographed immediately in a photo cuvette. Water depth was measured and recorded to the nearest 0.5 m for each of 960 males. To determine the functional relevance of the observed amino acid substitutions in the *LWS* genes, the sequence of the P allele was reconstructed from the H allele by *in vitro* mutagenesis. The pigments were then expressed, reconstituted and purified as described elsewhere[9]. Absorption spectra of reconstituted pigments were measured before and after irradiation with light (>490 nm). DNA was extracted from fin tissue of 305 individuals and amplified using 11 microsatellite primers. The fragments were analysed on a Beckman Coulter CEQ 8000 Genetic Analysis System. Determination of the opsin genes was as described previously[13]. We sequenced exons 2–5 of *LWS* (872 bp), which encode the trans-membrane region, from 263 individuals (526 haplotypes). We sequenced exons 1–5 (including introns) of the *SWS2A* gene from males of Makobe island and Kissenda island. For detection of selection, the *LWS* gene and its 5-kb upstream and 3.5-kb downstream flanking sequences (total 10.5 kb) were amplified by long PCR[9] and sequenced from 10 red and 9 blue males from Python island. To determine female mating preferences, we conducted laboratory two-way mate choice assays with females from Luanso island and Python island and laboratory-bred F_2 hybrids from Python island[40].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 April; accepted 25 July 2008.

1. Schluter, D. & Price, T. Honesty, perception and population divergence in sexually selected traits. *Proc. R. Soc. Lond. B* **253,** 117–122 (1993).
2. Boughman, J. W. How sensory drive can promote speciation. *Trends Ecol. Evol.* **17,** 571–577 (2002).
3. Gray, S. M. & McKinnon, J. S. Linking color polymorphism maintenance and speciation. *Trends Ecol. Evol.* **22,** 71–79 (2007).
4. Chunco, A. J., McKinnon, J. S. & Servedio, M. R. Microhabitat variation and sexual selection can maintain male color polymorphisms. *Evolution* **61,** 2504–2515 (2007).
5. Kawata, M., Shoji, A., Kawamura, S. & Seehausen, O. A genetically explicit model of speciation by sensory drive within a continuous population in aquatic environments. *BMC Evol. Biol.* **7,** 99 (2007).
6. Boughman, J. W. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature* **411,** 944–948 (2001).
7. Levring, T. & Fish, G. R. The penetration of light in some tropical East African waters. *Oikos* **7,** 98–109 (1956).
8. Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277,** 1808–1811 (1997).
9. Terai, Y. *et al.* Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS Biol.* **4,** 2244–2251 (2006).
10. Shichida, Y. *The Retinal Basis of Vision: Visual pigment: photochemistry and molecular evolution* (ed. Toyoda, J.-I.) 23–35 (Elsevier Science, 1999).
11. Yokoyama, S., Blow, N. S. & Radlwimmer, F. B. Molecular evolution of color vision of zebra finch. *Gene* **259,** 17–24 (2000).
12. Carleton, K. L. & Kocher, T. D. Cone opsin genes of African cichlid fishes: Tuning spectral sensitivity by differential gene expression. *Mol. Biol. Evol.* **18,** 1540–1550 (2001).
13. Terai, Y., Mayer, W. E., Klein, J., Tichy, H. & Okada, N. The effect of selection on a long wavelength-sensitive (*LWS*) opsin gene of Lake Victoria cichlid fishes. *Proc. Natl Acad. Sci. USA* **99,** 15501–15506 (2002).
14. Parry, J. W. L. *et al.* Mix and match color vision: Tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Curr. Biol.* **15,** 1734–1739 (2005).
15. Carleton, K. *et al.* Visual sensitivities tuned by heterochronic shifts in opsin gene expression. *BMC Biol.* **6,** 22 (2008).
16. Carleton, K. L., Parry, J. W. L., Bowmaker, J. K., Hunt, D. M. & Seehausen, O. Colour vision and speciation in Lake Victoria cichlids of the genus *Pundamilia*. *Mol. Ecol.* **14,** 4341–4353 (2005).
17. Spady, T. C. *et al.* Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol. Biol. Evol.* **22,** 1412–1422 (2005).
18. Genner, M. J. *et al.* Age of cichlids: New dates for ancient lake fish radiations. *Mol. Biol. Evol.* **24,** 1269–1282 (2007).
19. Maan, M. E. *et al.* Intraspecific sexual selection on a speciation trait, male coloration, in the Lake Victoria cichlid *Pundamilia nyererei*. *Proc. R. Soc. Lond. B* **271,** 2445–2452 (2004).
20. Endler, J. A. Some general comments on the evolution and design of animal communication systems. *Phil. Trans. R. Soc. Lond. B* **340,** 215–225 (1993).
21. Vandermeer, H. J., Anker, G. C. & Barel, C. D. N. Ecomorphology of retinal structures in zooplanktivorous haplochromine cichlids (Pisces) from Lake Victoria. *Environ. Biol. Fishes* **44,** 115–132 (1995).
22. Smit, S. A. & Anker, G. C. Photopic sensitivity to red and blue light related to retinal differences in two zooplanktivorous haplochromine species (Teleostei, Cichlidae). *Neth. J. Zool.* **47,** 9–20 (1997).
23. Maan, M. E., Hofker, K. D., van Alphen, J. J. M. & Seehausen, O. Sensory drive in cichlid speciation. *Am. Nat.* **167,** 947–954 (2006).
24. Endler, J. A. Gene flow and population differentiation. *Science* **179,** 243–250 (1973).
25. Schluter, D. & Nagel, L. M. Parallel speciation by natural selection. *Am. Nat.* **146,** 292–301 (1995).
26. Nosil, P., Egan, S. R. & Funk, D. J. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62,** 316–336 (2008).
27. Stinchcombe, J. T. & Hoektsra, H. E. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100,** 158–170 (2008).
28. Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421,** 259–264 (2003).
29. Gavrilets, S. *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press (2004).
30. Leimar, O., Doebeli, M. & Dieckmann, U. Evolution of phenotypic clusters through competition and local adaptation along an environmental gradient. *Evolution* **62,** 807–822 (2008).
31. Nosil, P., Vines, T. H. & Funk, D. J. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59,** 705–719 (2005).
32. Seehausen, O. *Lake Victoria Rock Cichlids. Taxonomy, Ecology and Distribution.* (Verduijn Cichlids, 1996).
33. Seehausen, O. & van Alphen, J. J. M. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behav. Ecol. Sociobiol.* **42,** 1–8 (1998).
34. Stelkens, R. B., Pierotti, M. E. R., Joyce, D. A., Smith, A. M., van der Sluijs, I. & Seehausen, O. Female mating preferences facilitate disruptive sexual selection on male nuptial colouration in hybrid cichlid fish. *Phil. Trans. R. Soc. B* 363 2861–2870 (2008).
35. Okullo, W. *et al.* Parameterization of the inherent optical properties of Murchison Bay, Lake Victoria. *Appl. Opt.* **46,** 8553–8561 (2007).
36. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289,** 739–745 (2000).
37. Asenjo, A. B., Rim, J. & Oprian, D. D. Molecular determinants of human red/green color discrimination. *Neuron* **12,** 1131–1138 (1994).

38. McDonald, J. H. Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* **15,** 377–384 (1998).
39. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116,** 153–159 (1987).
40. Haesler, M. P. & Seehausen, O. Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proc. R. Soc. B* **272,** 237–245 (2005).
41. van der Sluijs, I., van Alphen, J. J. M. & Seehausen, O. Preference polymorphism for coloration but no speciation in a population of Lake Victoria cichlids. *Behav. Ecol.* **19,** 177–183 (2008).
42. Nosil, P., Crespi, B. J. & Sandoval, C. P. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proc. R. Soc. Lond. B* **270,** 1911–1918 (2008).
43. Nosil, P. & Crespi, B. J. 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 (2004).
44. Rasanen, K. & Hendry, A. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol. Lett.* **11,** 624–636 (2008).

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

Acknowledgements We acknowledge the Tanzania Commission for Science & Technology for research permissions, the Tanzania Fisheries Research Institute, and its Muranza Centre director E. F. B. Katunzi, for 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, D. Schluter, 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, and 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.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.S. (ole.seehausen@aqua.unibe.ch) or N.O. (nokada@bio.titech.ac.jp).

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[41] (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[45]. 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.

***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[9] 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[46] 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[47], 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 (ref. 48). F_{ST} estimates and their significance were calculated over 100 permutations, as implemented in Arlequin[46].

Population genetics of opsin genes. Determination of the *LWS* gene was as described previously[13]. 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 (ref. 49). 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[9] 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[9] 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[38] 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[40]. 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.

45. Venables, W. N. & Ripley, B. D. Modern applied statistics with S. (Springer, 2002).
46. Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1,** 47–50 (2005).
47. Rice, W. R. Analyzing tables of statistical tests. *Evolution* **43,** 223–225 (1989).
48. Belkhir K., Borsa P. & Chikhi L.. Raufaste, N. & Bonhomme, F. Genetix Version 4.05 for Windows Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France) (1996–2004); http://www.genetix.univ-montp2.fr/genetix/genetix.htm.
49. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., Rozas, R. & Dna, S. P. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496–2497 (2003).

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'ATGAAGGGTAAACGTGATATGGA3'), SWS2A_F2 (5'CACCACAAACAACAAATACAACAA3'), SWS2A_R1 (5'AGGCCCGACTTTGGAAACTTC3'), SWS2A_R2 (5'AAAAGATAATCGTGGTCAAAGGAA3').

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[4]

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	
allele 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
Ma91HA1.txt	C...	H
Ma91HA2.txt	C...	H
Ma92HA1.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
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

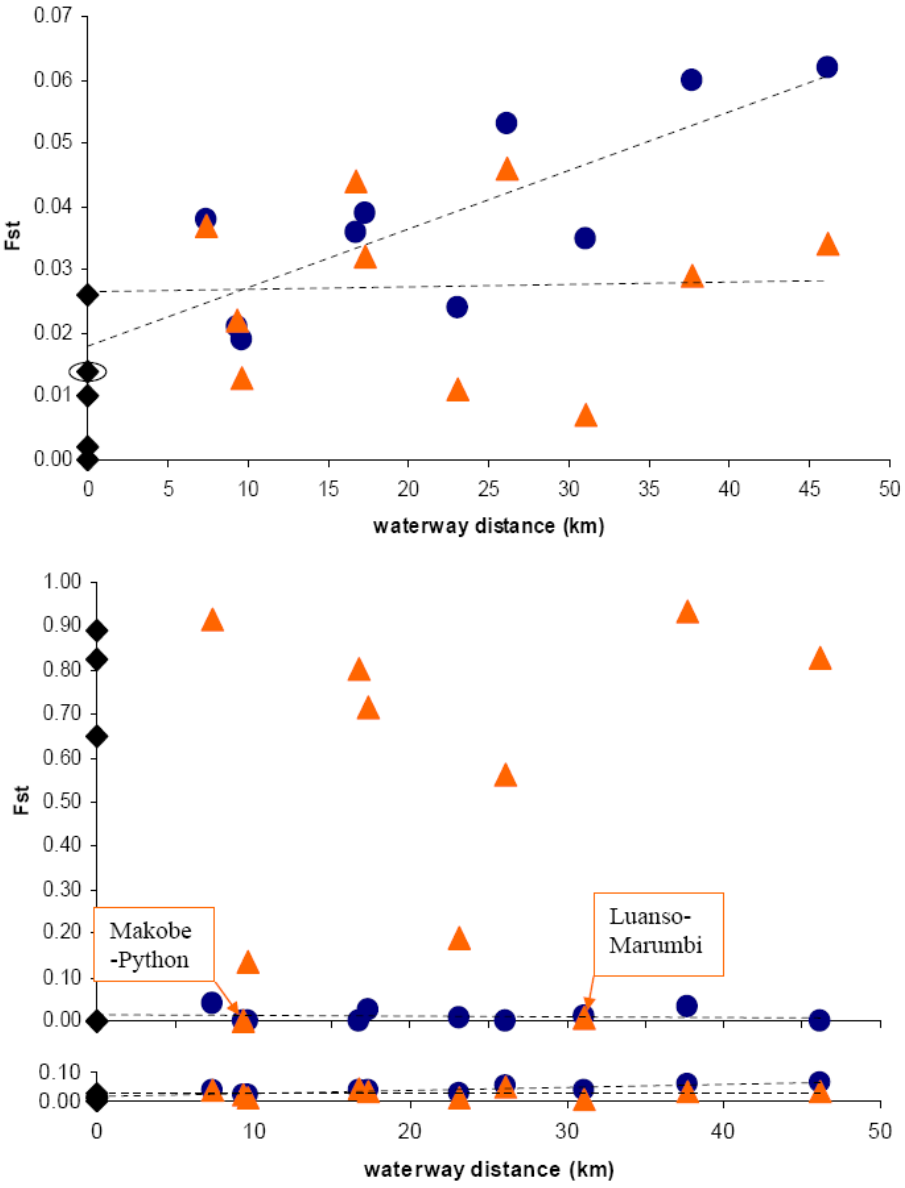

Figure S1. Contrasting patterns of population genetic divergence at neutral and adaptive loci.

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.

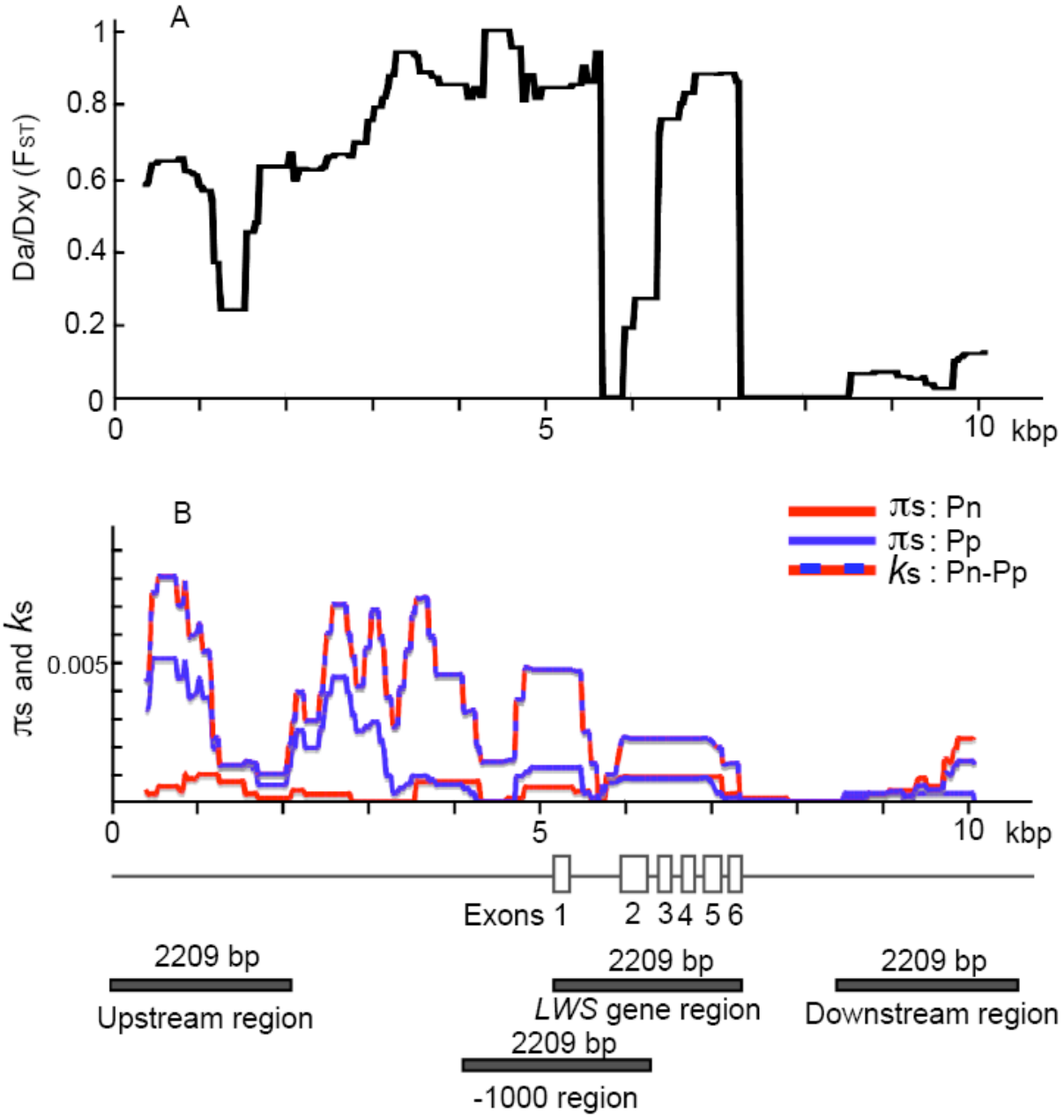

Figure S2. Detection of Selection Pressure on the *LWS* Gene. The genome structure of the *LWS* gene and its flanking region and (top) sliding-window analysis of Fst between blue and red male phenotypes from Python Island, and (bottom) sliding-window analysis of silent polymorphism (Π_s) in *P. pundamilia*-like (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.

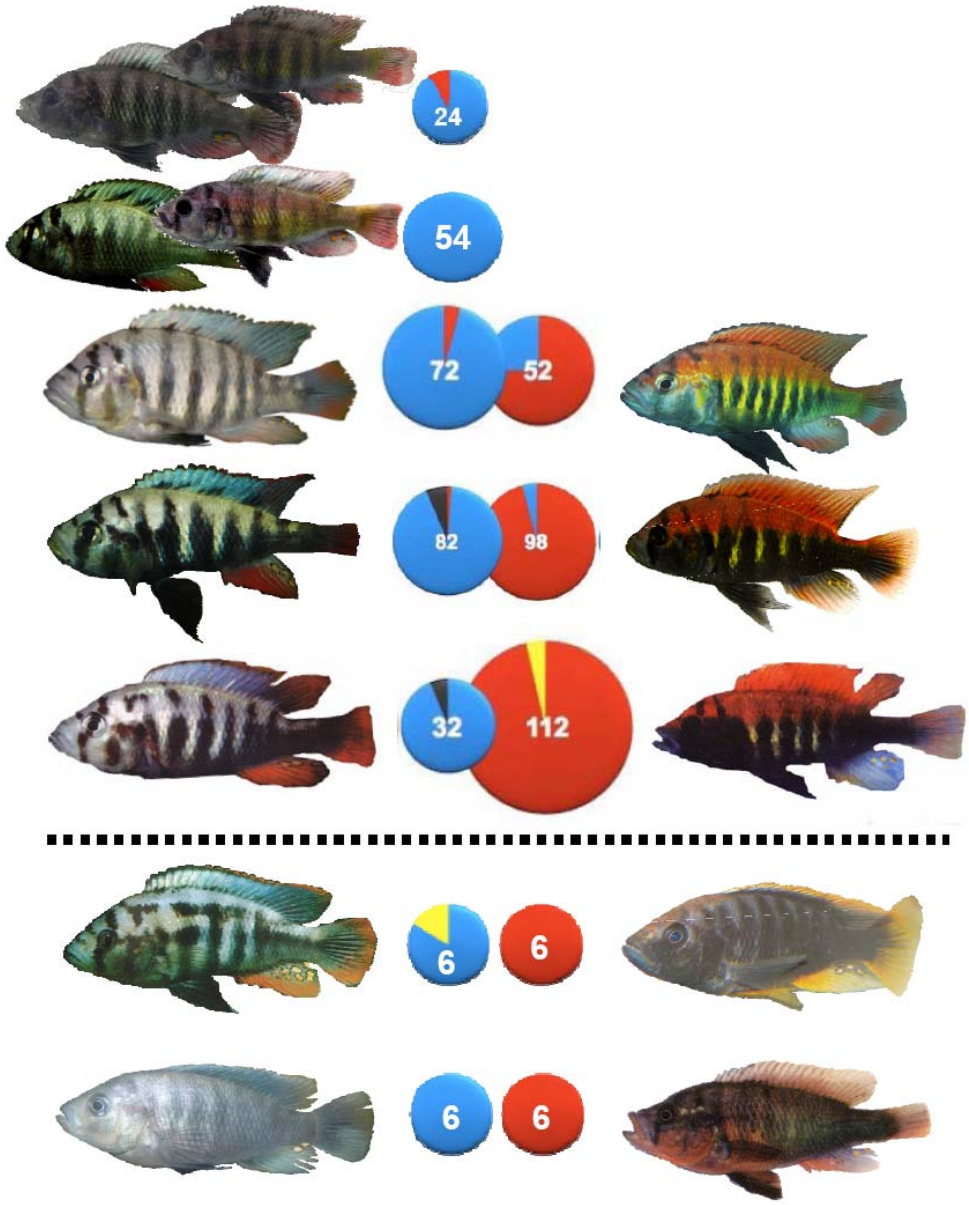

Figure S3. The male nuptial phenotypes and *LWS* allele group frequencies of all studied sympatric species and population pairs of *Pundamilia.* The hatched line separates the *P. pundamilia/P. nyererei* complex from other species pairs. Left column: blue types, right column: red dorsum and red chest 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.

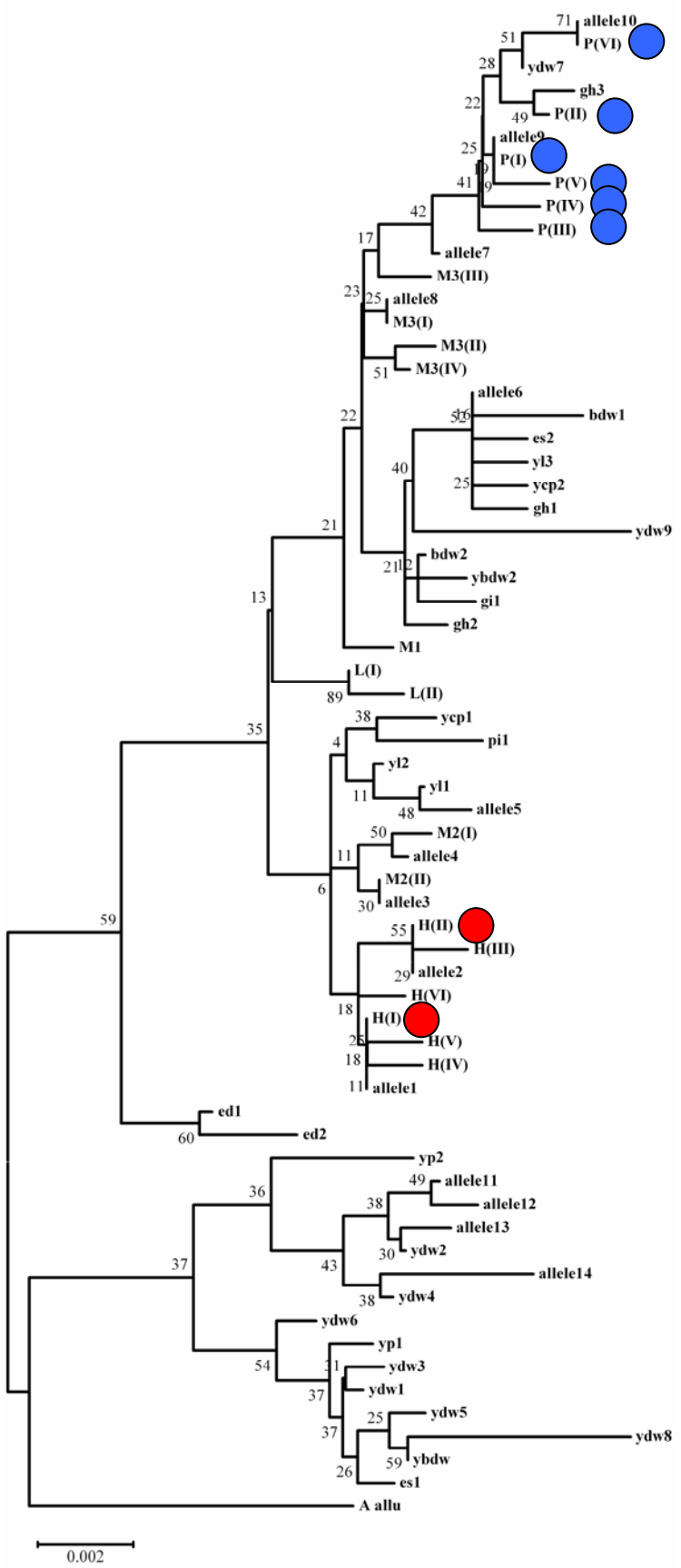

Figure S4a. Unrooted neighbourjoining genealogy of the *LWS* alleles of Lake Victoria cichlid fish built from 1098 sequences, including all currently known alleles. Numbers on branches are % bootstrap support. The *Pundamilia* H and P alleles are highlighted red and blue respectively.

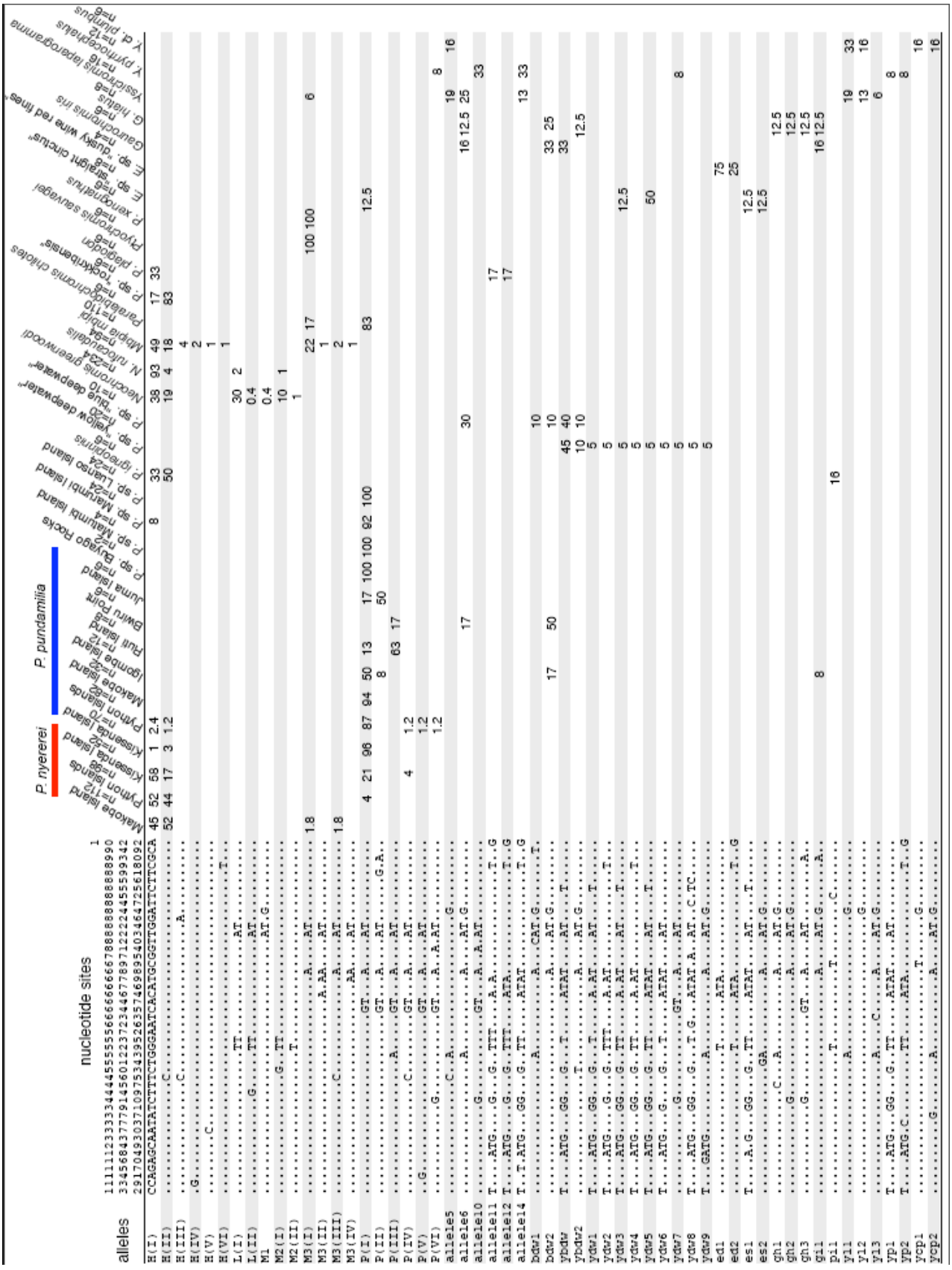

Figure S4b. An alignment of all informative sites of alleles and the frequencies (%) of each allele in *Pundamilia* and in various other species of Lake Victoria cichlids. The sample size is reported for each species under the species name and refers to the number of haplotypes (total n = 1098). Dots 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.

AdditionalReferences 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.