

# Genome mapping of the orange blotch colour pattern in cichlid fishes

J. TODD STREELMAN, R. CRAIG ALBERTSON\* and THOMAS D. KOCHER

Hubbard Center For Genome Studies, 4th Floor, Environmental Technology Building, University of New Hampshire, 35 Colovos Road, Durham, NH 03824, USA

## Abstract

The dramatic variation of cichlid fish colour pattern is thought to function in mate choice, evolve by sexual selection, and contribute to explosive speciation. Here, we combine linkage mapping and population genetic analyses to identify a single region of the cichlid genome responsible for the orange blotch (OB) colour phenotype. In each analysis, OB is tightly linked to the *c-ski1* gene. Additionally, we use comparative mapping information from the *Takifugu rubripes* and human genomes to suggest positional candidate loci for OB. Our work should engender a more comprehensive understanding of the molecular ecology of OB and its role in cichlid speciation. Moreover, we have assembled the components of a method to focus upon the genetic basis of evolutionarily and ecologically significant phenotypes.

**Keywords:** cichlid, colour pattern, orange blotch, QTL, sexual selection, speciation

Received 14 February 2003; revision received 3 June 2003; accepted 3 June 2003

## Introduction

Colour pattern is a central feature of cichlid fish behaviour and evolution. Hues from red to yellow to blue, in conjunction with stripes and bars, decorate cichlid bodies in a bewildering array of species-specific combinations. The brightest and gaudiest individuals are usually male. Females are less colourful, consistent with their greater role in parental care. In this context, cichlid colour is believed to confer mate recognition signals and may evolve by sexual selection via female choice (Dominey 1984; Seehausen & van Alphen 1998). Sexual selection is probably a major evolutionary force driving the diversification of cichlid species flocks (Seehausen 2000; Danley & Kocher 2001).

A few cichlid colour phenotypes are expressed mainly in females. One is called orange blotch (hereafter OB), with individuals resembling calico cats (Fig. 1, right). The apparent sex linkage of OB may be explained by proximity to a hypothetical female determining locus (Lande *et al.* 2001). OB males are rare in nature (less than 1%) and are thought to be recombinants between sex and colour (Seehausen *et al.* 1999). Early research suggested that the

OB / BB polymorphism was oligogenic and that OB was dominant (Holzberg 1978). The OB pattern has been observed in numerous species from Lakes Victoria and Malawi (Table 1 in Lande *et al.* 2001) but not in Lake Tanganyika or elsewhere. OB is almost always at lower frequency than the blue with black bar pattern, which is found in both males and females (BB, Fig. 1, left). In Lake Victoria, the frequency of OB is higher in areas of relatively clear water (Seehausen *et al.* 1999).

The mechanisms maintaining this apparently bright colour pattern in females are not known. Against some backgrounds in shallow water, OB may be cryptic to kingfishers (Stauffer, pers. comm.). However, in deeper water the pattern is almost certainly conspicuous to piscine predators. There is some evidence that males spawned by OB mothers will choose OB mates (Seehausen *et al.* 1999). Recently, Lande *et al.* (2001) have modelled how blotched phenotypes,



**Fig. 1** Representative BB (male) and OB (female) individuals of *Metriaclima zebra*.

Correspondence: J. Todd Streeleman. Fax: (603) 862 2940; E-mail: jts3@hopper.unh.edu

\*Present address: Department of Cytokine Biology, The Forsyth Institute, 140 The Fenway, Boston MA 02115, USA

in complex association with reversal of sex determining mechanisms, might contribute to cichlid sympatric speciation by sexual selection alone.

Here, we use two different approaches to identify the region of the cichlid genome responsible for OB. First, we take advantage of linkage disequilibrium, generated from an interspecific cross, to localize OB to a particular chromosomal region. We then exploit greater levels of recombination in a natural population to identify a single marker on this linkage group which is associated closely with colour. The combination of these complementary strategies, quantitative trait locus (QTL) mapping and association scans provides an efficient means to discover the linkage between phenotype and genotype (Streelman & Kocher 2000). Finally, using comparative information from the pufferfish (*Takifugu rubripes*) and human genomes, we identify positional candidate genes which might be responsible for this pigmentation phenotype.

## Materials and methods

### QTL scan for the OB/BB polymorphism

We crossed a BB male *Labeotropheus fuelleborni* (LF) with an OB female *Metriaclima zebra* (MZ); these animals were one generation removed from wild-caught stock. The  $F_1$  hybrids were viable, fertile and segregated 50:50 for OB.  $F_1$  were intercrossed to produce 152  $F_2$ , which were scored as either OB or BB. The  $F_2$  were typed for 136 codominant microsatellite markers or restriction fragment length polymorphisms (RFLPs) in known genes. Roughly half these markers were chosen to maximize genomic coverage, from the more extensive linkage map of the tilapia, a closely related cichlid (Lee *et al.* in prep; <http://hogs.unh.edu/comp>). We constructed a genetic map using JoinMap 3.0 (Van Ooijen & Voorrips 2001), assigning 127 markers to 24 linkage groups with a LOD score threshold of 4.0. Further details of map construction have been published elsewhere (Albertson *et al.* 2003). The cichlid proto-oncogene *c-ski1* (Huang *et al.* 1999) was mapped using primers (F: gtcagtcacattcctggctg, R: ttctatgctctgcgggtttt) that amplify a CT microsatellite in the 5' UTR. An additional set of markers was mapped in the LF  $\times$  MZ cross after initial comparison of the OB linkage group to its homologue in tilapia. This included a tandem array of cone opsin genes (*LWS*, *SWS2a*, *SWS2b*; Carleton and Kocher 2001), mapped by polymerase chain reaction (PCR)-RFLP using primers F0a and R0 (Carleton & Kocher 2001) and the restriction enzyme *Taq* 1, as well as the microsatellite marker GM 264 (GenBank Accession no. BV005384).

We used the nonparametric mapping function of MapQTL 4.0 (Van Ooijen *et al.* 2002) to estimate the genomic position of the binary OB/BB colour polymorphism. This approach employs the Kruskal–Wallis rank sum test, which orders individuals according to phenotype, and then sorts them, one locus at a time, by marker genotype. Under the null

hypothesis of no segregating QTL, the Kruskal–Wallis test statistic ( $K$ ) approximates a  $\chi^2$  distribution, with degrees of freedom equal to the number of genotypic classes minus one. Since the test is performed on many linked and unlinked loci, we used a Bonferroni correction to derive an appropriate significance threshold ( $\alpha = 0.05$ ,  $P < 0.0004$ ).

We employed recombination breakpoint analysis to define molecular markers that flank the OB/BB polymorphism. The rationale behind this approach is that  $F_2$  individuals segregate as independent combinations of parental genomes. For markers on the OB linkage group, we identified the allele (OB is dominant), inherited from the MZ dam, sufficient to confer the affected colour pattern. We then reconstructed OB chromosome breakpoints in  $F_2$  fishes.

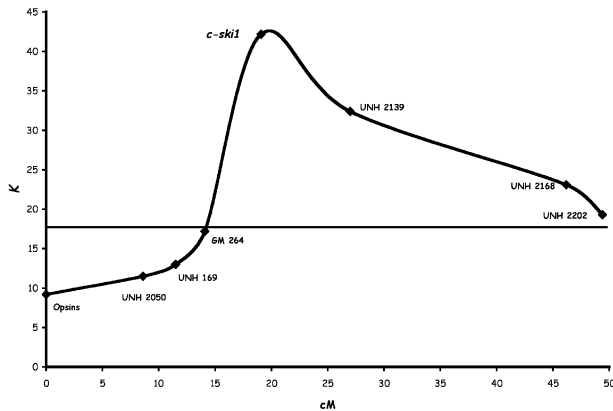
### Association scan for OB in nature

We wanted to test the hypothesis that molecular markers, linked to OB in our laboratory cross, segregate with the OB/BB colour polymorphism in nature. Natural populations harbour a greater number of recombinant individuals than laboratory crosses and can serve as independent or complementary mapping 'panels' (Lander & Schork 1994). The ability to detect an association between allele frequency and phenotype in assigned experimental vs. control groups is determined by the demographic history of alleles in populations and is affected by local rates of recombination, selection and allelic heterogeneity (Kruglyak 1999). We reasoned that significant linkage disequilibrium might exist, in nature, between the causative OB mutation and molecular markers from the OB chromosome. If this is so, we expect significant differences in allele frequency at linked markers for OB vs. BB individuals.

We genotyped individuals of MZ from Chiofu Bay, Lake Malawi (segregating for OB/BB, collected in July 2001), for the following microsatellite markers from linkage group 12 (linkage group numbers from Albertson *et al.* 2003): UNH 2050, UNH 169, GM 264, *c-ski1*, UNH 2139 and UNH 2168. We used GENEPOP (Raymond & Rousset 1995) to conduct (1) probability tests of Hardy–Weinberg equilibrium at each locus, (2) exact tests of allelic association among loci (linkage disequilibrium) and (3) per locus exact tests of genic differentiation (allele frequency differences) between BB and OB individuals. All significance values were Bonferroni-corrected for multiple tests.

### Comparative mapping of OB

Markers flanking OB were used in blastN searches of the *Takifugu rubripes* genome (Aparicio *et al.* 2002; <http://genome.jgi-psf.org/fugu6/fugu6.home.html>). Resulting genomic scaffolds were browsed for open reading frames to identify genes, which are putatively located in the interval containing OB.



**Fig. 2** A single region of the genome is associated with the OB/BB colour polymorphism in the  $F_2$  of LF and MZ. Markers *c-ski1* ( $K = 42$ ;  $P < 0.0001$ ), UNH 2139 ( $K = 31$ ;  $P < 0.0001$ ), UNH 2168 ( $K = 23$ ;  $P < 0.0001$ ) and UNH 2202 ( $K = 19$ ;  $P < 0.0001$ ) from linkage group 12 show significant linkage to colour using the Kruskal–Wallis nonparametric mapping function of MapQTL 4.0. Shown is the value of the test statistic,  $K$ , for molecular markers along LG 12. The bar indicates the Bonferroni-corrected significance level ( $\alpha = 0.05$ ,  $P < 0.0004$ ).

## Results

### QTL scan for the OB/BB polymorphism

The OB/BB colour polymorphism is associated significantly with four molecular markers located on linkage group 12 (critical value  $\alpha = 0.05$ ,  $P < 0.0004$ ; Fig. 2): *c-ski1* ( $K = 42$ ;  $P < 0.0001$ ), UNH 2139 ( $K = 31$ ;  $P < 0.0001$ ), UNH 2168 ( $K = 23$ ;  $P < 0.0001$ ) and UNH 2202 ( $K = 19$ ;  $P < 0.0001$ ). The  $K$  score for marker GM 264 ( $K = 17$ ) is high but not statistically significant ( $P < 0.0005$ ) after Bonferroni correction. The OB chromosome is dominant; we identified a single allele, sufficient to confer the OB pattern, at each of the loci linked to colour. Both the Kruskal–Wallis rank sum test and interval mapping (data not shown) indicate that a region of the linkage group near the *c-ski1* locus is most strongly associated with OB.

We used recombinant breakpoint analysis to infer the molecular markers flanking the OB/BB polymorphism. Following the inheritance of the OB chromosome, from the MZ dam through individual  $F_2$  fishes, allows us to pinpoint chromosome breakpoints. For instance, an OB  $F_2$  individual carried the 'OB' allele inherited from the MZ dam at markers *c-ski1* and UNH 2139 but not GM 264 and UNH 169. A second  $F_2$  animal (scored BB) carried the 'OB' allele from the MZ dam at markers UNH 169 and GM 264 but not *c-ski1* or UNH 2139. Taken together, these breakpoints suggest that the position of OB lies between markers GM 264 and *c-ski1*. Recombination fractions at GM 264 (0.03) and *c-ski1* (0.02) imply that OB is roughly 2 centimorgans (cM) from *c-ski1*.

**Table 1** Summary statistics for microsatellite markers employed in a population survey of *Metriaclima zebra* from Chiofu Bay, Lake Malawi. Asterisks (\*) indicate markers which depart from Hardy–Weinberg equilibrium (critical value  $\alpha = 0.05$ ,  $P < 0.0083$ ). Exact tests evaluate allele frequency differences between OB and BB individuals (critical value  $\alpha = 0.05$ ,  $P < 0.0083$ )

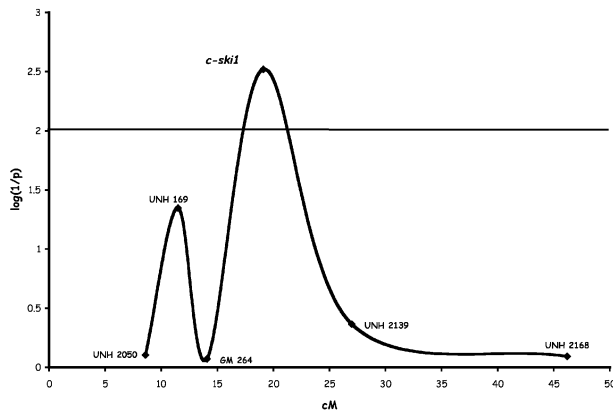
Locus	$n$ (individuals)	no. alleles	$H_O$	$H_E$	Exact test ( $P$ )
UNH 2050	67	52	0.96	0.98	NS
UNH 169	62	3	0.39	0.42	NS
GM 264*	43	23	0.79	0.94	NS
<i>c-ski1</i>	70	14	0.86	0.89	0.003
UNH 2139*	70	31	0.95	0.95	NS
UNH 2168	65	18	0.80	0.88	NS

### Association scan for OB in nature

We wanted to test the hypothesis that markers from the OB linkage group are associated with colour in natural populations. To accomplish this, we tested allele frequency differences between OB and BB individuals from the wild for markers linked to colour in our laboratory cross. We scored six microsatellite markers from linkage group 12 (UNH 2050, UNH 169, GM 264, *c-ski1*, UNH 2139 and UNH 2168) for polymorphism in the Chiofu Bay population of MZ, which segregates for OB/BB (Table 1). Markers UNH 2139 and GM 264 departed from Hardy–Weinberg equilibrium (critical value  $\alpha = 0.05$ ,  $P < 0.0083$ ). We found no evidence of linkage disequilibrium between alleles at the six loci (all  $P > 0.1$ ). Only *c-ski1* showed significant differences in allele frequency between OB and BB individuals ( $P = 0.003$ , critical value  $\alpha = 0.05$ ,  $P < 0.0083$ ; Table 1; Fig. 3). These data from a population sample support the inference derived from our laboratory cross that the OB gene is linked to *c-ski1*.

### Comparative mapping of OB

Flanking markers *c-ski1* and GM 264 show significant similarity to regions of *Takifugu* genome scaffolds. Base pairs (bp) 1–1406 of *c-ski1* (Huang *et al.* 1999) match ( $E$ -value = 0.00) one end of scaffold 61 (~400 Kb) and positions 1541–3116 match ( $1 \times 10^{-180}$ ) another end of scaffold 825 (~100 Kb). Sequence from GM 264 shows similarity ( $3 \times 10^{-21}$ ) to scaffold 2321 (26 296 bp). Open reading frames (ORFs, Table 2) on these scaffolds include a predicted protein with similarity to medaka *B* (scaffold 61), *Pax7* (scaffold 61), multiple predicted proteins with forkhead domains (scaffolds 61 and 2321), the glutamate receptor AMPA 2.2 (scaffold 825) and *Blu* (scaffold 825).



**Fig. 3** *c-ski1* is associated with colour in a population of MZ from Chiofu Bay, Lake Malawi. For each marker UNH 2050, UNH 169, GM 264, *c-ski1*, UNH 2139 and UNH 2168, we report the  $\log(1/p)$  from exact tests of allele frequency differences, implemented in GENEPOP. Only *c-ski1* shows significant genic differentiation ( $P = 0.003$ ) between OB and BB individuals. The bar indicates the Bonferroni-corrected significance level ( $\alpha = 0.05$ ,  $P < 0.0083$ ).

## Discussion

Variation in colour pattern is a hallmark of the rapid evolution of cichlid fishes in the great lakes of East Africa (Seehausen 2000). Differences among species or populations are common in nature (Arnegard *et al.* 1999; Smith & Kornfield 2002) and occupy a central place in theoretical models of cichlid speciation, whether as the phenotype which drives diversification by sexual selection (Turner & Burrows 1995; Higashi *et al.* 1999; Lande *et al.* 2001) or as a marker of fitness-

bearing traits in ecological selection scenarios (Dieckmann & Doebeli 1999; Kondrashov & Kondrashov 1999).

The orange blotch colour phenotype is unusual because it is expressed almost exclusively in females and is not dependent upon sexual maturity. The ontogeny of OB seems to involve the altered deployment of melanophores at metamorphosis. In LF, MZ and the hybrids we have examined, BB and OB individuals are indistinguishable as larvae and after release from maternal oral incubation (21 days post-fertilization, dpf). At metamorphosis (80–100 dpf), melanophores of BB individuals form vertical bars on a bluish-brown background. By contrast, the melanophores of presumptive OB individuals do not form distinct bars, but rather mass in random and variably shaped blotches, superimposed on an orange body colour.

Lande *et al.* (2001) have taken advantage of the sex linkage of blotched colour patterns to offer a model of cichlid sympatric speciation via reversal of sex determining mechanisms and sexual selection. Under restrictive conditions, the model predicts speciation in a few dozen generations. In general, efforts to grasp the function of colour pattern in speciation would be facilitated greatly by a molecular understanding of the phenotype in question. To this end, we have used genomic and population approaches to characterize the genetic basis of OB.

### *OB is linked to c-ski1 in the laboratory and in nature*

We used a QTL scan of polymorphism in a laboratory cross to identify a single region of the cichlid genome associated

**Table 2** Summary of selected genes found on *Takifugu rubripes* genomic scaffolds. Scaffolds 61 and 825 are positive for *c-ski1*; scaffold 2321 is positive for GM 264. The third column indicates the position in bp, along the scaffold, which marks the beginning of the corresponding ORF. Approximate scaffold sizes are: 61 = 400 Kb, 825 = 100 Kb, 2321 = 26 000 bp. The fourth column lists the position in the human genome for each ORF, respectively

Scaffold	Gene	Start	Human homologue
61	<i>c-ski1</i>	13328	1p36
61	HKR3	93074	1p36
61	Forkhead domain	126137	<i>Foxd2</i> (1p34) <i>Foxd3</i> (1p32) <i>Foxe3</i> (1p32)
61	similar to membrane-associated transporter (B)	241897	???
61	T1R1 sweet taste receptor	268996	1p36
61	<i>Pax7</i>	314620	1p36
825	Hyaluronoglucosaminidase 1	16905	3p21
825	<i>Blu</i>	35471	3p21
825	<i>A6</i>	52476	3p21
825	Receptor for activated protein kinase C	57212	3p21
825	AMPA 2.2 (glutamate receptor)	80486	4q32
825	Phospholipase C – L2	84645	3p22–21
825	<i>c-ski1</i>	94984	1p36
2321	Forkhead domain	17228	<i>Foxd2</i> (1p34) <i>Foxd3</i> (1p32) <i>Foxe3</i> (1p32)

with the OB colour pattern. The OB allele inherited from the MZ dam is dominant in our cross, supporting conclusions from earlier research (Holzberg 1978). The region of the OB linkage group associated with colour is large, spanning approximately 30 cM (Fig. 2). We have used two independent approaches to refine the position of OB. First, we used breakpoint analysis to localize OB between markers GM 264 and *c-ski1*. The recombination fractions at GM 264 (0.03) and *c-ski1* (0.02) suggest that OB is roughly 2 cM from *c-ski1*. Using a general approximation for vertebrate genomes (to convert genetic to physical distance, 1 cM = 0.5–1 megabase [Mb]; Lynch & Walsh 1998), we conjecture that OB is within 2 Mb of *c-ski1*.

Second, we used an association scan to test markers, linked to colour in the laboratory, for allele frequency differences between OB and BB individuals from nature. Only *c-ski1* showed significant genic differentiation with respect to colour pattern (Table 1; Fig. 3). Taken together, the results of linkage and population genetic analyses suggest that OB lies in close proximity to *c-ski1*.

What does this combined approach tell us about the genetic basis of OB? Should we expect to find differences in allele frequencies at *c-ski1* in other populations of MZ, or other species segregating for OB? Strictly speaking, the inference space of our QTL experiment is limited to our mapping panel, while the inference space of our association scan at Chiofu is limited to this population (Streelman & Kocher 2000). It is notable, perhaps, that we detect an association between *c-ski1* and colour over an estimated 2 Mb, given that the average extent of linkage disequilibrium (LD) in well-studied genomes (i.e. human) is thought to be much smaller (3 kilobases [Kb] & Kruglyak 1999). Several lines of evidence suggest that extended linkage disequilibrium may be a reasonable expectation for cichlid genomes in general and for the OB phenotype in particular. First, the average value of LD in humans obscures a growing appreciation that genomes are mosaics of high and low LD blocks (Daly *et al.* 2001; Jeffreys *et al.* 2001). Furthermore, populations with differing demographic histories will probably vary in the extent of LD over the same genomic interval (Stumpf & Goldstein 2003). For instance, isolated populations sometimes exhibit greater LD (up to 11.5 Mb) than others with contrasting evolutionary trajectories (Angius *et al.* 2002; Jorgensen *et al.* 2002).

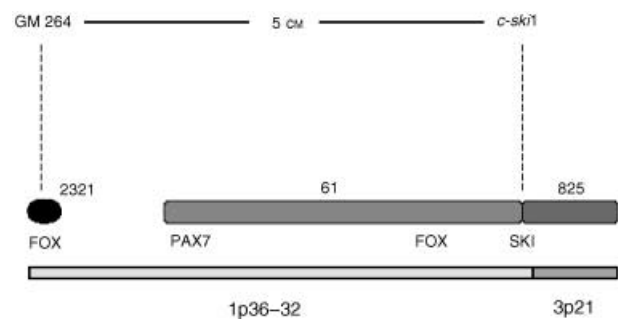
Cichlid populations are highly structured, isolated by elements of the physical environment (e.g. sandy substrates, deep water) and, on average, are young (Kornfield & Smith 2000). This means that novel phenotypes may become established in populations with small effective sizes (10–100). OB is a sex-linked phenotype in nature and is found usually in low frequency (8–40% of females, Lande *et al.* 2001). Each of these conditions (variant founded by a small number of individuals, variant in low frequency) can lead to greater LD (Kruglyak 1999). In general, we know too

little about the extent of LD in cichlid genomes and the variance in LD across populations. The linkage of *c-ski1* to the OB phenotype in the Chiofu population, as well as the observation that OB is found in isolated populations of MZ and other species from two African rift lakes (Malawi and Victoria) would seem to present a unique opportunity to gather relevant data.

#### Candidate genes for OB

We used comparative genomic analysis to infer the identity of genes located in the OB region. Sequences from flanking markers GM 264 and *c-ski1* were employed to search the *T. rubripes* genome. Each marker had strong matches to *Takifugu* genome scaffolds (Table 2). In all, these comprise about 500 Kb of sequence. Given the compression of the *Takifugu* genome with respect to other vertebrates (2–6×), this region approximates the size of the OB interval. Assuming conserved synteny among bony fish genomes, on the scale of a few Mb the ORFs on *Takifugu* scaffolds 61, 825 and 2321 should correspond roughly to the set of the genes around cichlid *c-ski1*.

We interpret the matches of different regions of *c-ski1* to ends of scaffolds 61 and 825 to indicate that these segments are neighbours in the genomic sequence. Unfortunately, because scaffolds 61 and 825 (*c-ski1*-positive) do not join scaffold 2321 (GM 264-positive) we cannot order, with certainty, the position of OB with respect to the *Takifugu* genome. However, as scaffold 2321 contains an ORF with similarity to genes on human chromosome 1p36–32, we can hypothesize about the orientation of the markers (Fig. 4).



**Fig. 4** Comparative mapping of the OB interval. OB is flanked by markers GM 264 and *c-ski1* and is probably located within 2 cM of *c-ski1* (see text). Markers GM 264 and *c-ski1* have strong hits to *Takifugu* genome scaffolds 2321 (26 296 bp) 61 (~400 Kb) plus 825 (~100 Kb), respectively (scaffolds drawn to approximate scale). Open reading frames located on scaffold 61 include PAX7, a forkhead domain (FOX), a protein with similarity to medaka B and other markers found on human 1p36 (Table 2). A single ORF containing a forkhead domain is present on scaffold 2321. Scaffold 825 contains ORFs found mostly on human chromosome 3p21 (Table 2). Although we suggest this orientation of markers with respect to the *Takifugu* genome, we cannot formally exclude the possibility that GM 264 is located on the other side of *c-ski1*.

There are a number of ORFs that fuel speculation about candidate loci for OB. Scaffolds 61 (*c-ski1*-positive) and 2321 (GM264-positive) contain ORFs corresponding to genes located on human chromosome 1p36–32 (Table 2; Fig. 4). *Pax7* is a member of the paired-box gene family encoding a transcription factor functioning in the neural crest, skeletal muscle and nervous system (Mansouri *et al.* 1996). *Pax7* is related closely to *Pax3* (Dahl *et al.* 1997), which is required for melanocyte development (Hornyak *et al.* 2001). Scaffolds 61 and 2321 each contain ORFs with a forkhead box. Forkhead genes comprise a large family of transcription factors named for the shape of their DNA-binding domain (Kaestner *et al.* 2000). Three forkhead genes (*Foxd2*, *Foxd3*, *Foxe3*) are present on human 1p36–32. *Foxd3* is capable of repressing melanogenesis (Kos *et al.* 2001). An ORF with similarity to medaka *B* is also located on scaffold 61. *B* is a novel transporter protein responsible for the orange-red medaka mutant (Fukamachi *et al.* 2001). The ability to exploit the *Takifugu* genome to identify candidate loci and comparative markers is a great advantage for this and other cichlid mapping projects.

#### Future directions

We have used both linkage mapping and population analyses to identify the genetic basis of an evolutionarily significant phenotype in vertebrates from nature. Each of these strategies is fundamental to the future of our research. Linkage mapping relies on disequilibrium among markers and provides a first pass to pinpoint large genomic regions associated with phenotype. To narrow these intervals, additional molecular markers must be tested in greater numbers of individuals. Comparative and physical genomic resources can contribute the necessary molecular loci, while further generations of this and other cross(es) will provide the recombinants necessary to refine the map position of OB.

We have demonstrated that individuals from a natural population segregating for OB show the signature of genic differentiation at *c-ski1*. Individuals from nature will provide an invaluable source of recombinants useful for association mapping. Finally, it has not escaped our attention that OB lies in close proximity to a quantitative trait locus for tooth shape (the interval UNH 2050 to *c-ski1*, Albertson *et al.* 2003) and a tandem array of three cone opsin genes (*LWS*, *SWS2a*, *SWS2b*, Fig. 2; Carleton & Kocher 2001). Given that cichlid speciation probably involves the combined forces of natural selection on ecological traits and sexual selection on colour patterns, we are very interested in these and other genomic linkages between colour, morphology and putative mate choice characters.

#### Acknowledgements

We thank members of the Kocher laboratory and two anonymous reviewers for comments on the manuscript; Aggrey Ambali, the

University of Malawi and the Malawi government for collecting permits. This work is supported by grants from the NSF to TDK (IBN 9905127); and the NSF/Alfred P. Sloan Foundation (DBI 9803946) and NIH (R03 DE 14446–01) to JTS.

#### References

- Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences USA*, **100**, 5252–5257.
- Angius A, Bebbere D, Petretto E *et al.* (2002) Not all isolates are equal: linkage disequilibrium analysis on Xq13.3 reveals different patterns in Sardinian sub-populations. *Human Genetics*, **111**, 9–15.
- Aparicio S, Chapman J, Stupka E, *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Takifugu rubripes*. *Science*, **297**, 1301–1310.
- Arnegard ME, Markert JA, Danley PD, Stauffer JR, Ambali AJ, Kocher TD (1999) Population structure and colour variation of the cichlid fish *Labeotropheus fuelleborni* along a recently formed archipelago of rocky habitat patches in southern Lake Malawi. *Proceedings of the Royal Society of London B*, **266**, 119–130.
- Carleton KL, Kocher TD (2001) Cone opsin genes of African cichlid fishes: tuning spectral sensitivity by differential gene expression. *Molecular Biology and Evolution*, **18**, 1540–1550.
- Dahl E, Koseki H, Balling R (1997) *Pax* genes and organogenesis. *Bioessays*, **19**, 755–765.
- Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. *Nature Genetics*, **29**, 229–232.
- Danley PD, Kocher TD (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology*, **10**, 1075–1086.
- Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354–357.
- Dominey WJ (1984) Effects of sexual selection and life history on speciation: species flocks in African cichlids and Hawaiian *Drosophila*. In: *Evolution of Fish Species Flocks* (eds Echelle AA, Kornfield I), pp. 231–249. University of Maine Press, Orono, ME.
- Fukamachi S, Shimada A, Shima A (2001) Mutations in the gene encoding *B*, a novel transporter protein, reduce melanin content in medaka. *Nature Genetics*, **28**, 381–385.
- Higashi MG, Takimoto G, Yamamura N (1999) Sympatric speciation by sexual selection. *Nature*, **402**, 523–526.
- Holzberg S (1978) A field and laboratory study of the behaviour and ecology of *Pseudotropheus zebra* (Boulenger), an endemic cichlid of Lake Malawi (Pisces: Cichlidae). *Zeitschrift für Zoological Syst Evological – Forschung*, **16**, 171–186.
- Hornyak TJ, Hayes DJ, Chiu LY, Ziff EB (2001) Transcription factors in melanocyte development: distinct roles for *Pax-3* and *Mitf*. *Mechanisms of Development*, **101**, 47–59.
- Huang CJ, Lin JY, Tsai HJ (1999) Two distinct *c-ski* cDNAs of fish, tilapia (*Oreochromis aurea*). *Molecular Reproduction and Development*, **54**, 223–231.
- Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics*, **29**, 217–222.
- Jorgensen TH, Degn B, Wang AG *et al.* (2002) Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands. *European Journal of Human Genetics*, **10**, 381–387.

- Kaestner KH, Knochel W, Martinez DE (2000) Unified nomenclature for the winged helix/forkhead transcription factors. *Genes and Development*, **14**, 142–146.
- Kondrashov AS, Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature*, **400**, 351–354.
- Kornfield I, Smith PF (2000) African cichlid fishes: model systems for evolutionary biology. *Annual Review of Ecology and Systematics*, **31**, 163–196.
- Kos R, Reedy MV, Johnson RL, Erickson CA (2001) The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development*, **128**, 1467–1479.
- Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, **22**, 139–144.
- Lande R, Seehausen O, van Alphen JJM (2001) Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica*, **112–113**, 435–443.
- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland MA.
- Mansouri A, Stoykova A, Torres M, Gruss P (1996) Dysgenesis of cephalic neural crest derivatives in *Pax7*<sup>-/-</sup> mutant mice. *Development*, **122**, 831–838.
- Raymond M, Rousset F (1995) GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Seehausen O (2000) Explosive speciation rates and unusual species richness in haplochromine cichlid fishes: effects of sexual selection. *Advances in Ecological Research*, **31**, 237–274.
- Seehausen O, van Alphen JJM (1998) The effect of male coloration on female mate choice in closely related Lake Victoria cichlids. *Behavioral Ecology and Sociobiology*, **42**, 1–8.
- Seehausen O, van Alphen JJM, Lande R (1999) Color polymorphism and sex ratio distortion in a cichlid fish as an incipient stage in sympatric speciation by sexual selection. *Ecology Letters*, **2**, 367–378.
- Smith PF, Kornfield I (2002) Phylogeography of Lake Malawi cichlids of the genus *Pseudotropheus*: significance of allopatric color variation. *Proceedings of the Royal Society of London B*, **269**, 2495–2502.
- Streelman JT, Kocher TD (2000) From phenotype to genotype. *Evolution and Development*, **2**, 166–173.
- Stumpf MP, Goldstein DB (2003) Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Current Biology*, **13**, 1–8.
- Turner GF, Burrows MT (1995) A model of sympatric speciation by sexual selection. *Proceedings of the Royal Society of London B*, **260**, 287–292.
- Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2002) *Qtlmap 4.0, Software for the Calculation of QTL Positions on Genetic Maps*. Plant Research International, Wageningen, the Netherlands.
- Van Ooijen JW, Voorrips RE (2001) *Joinmap 3.0, Software for the Calculation of Genetic Linkage Maps*. Plant Research International, Wageningen, the Netherlands.

