A Microsatellite-Based Genetic Linkage Map of the Cichlid Fish, *Astatotilapia burtoni* and a Comparison of Genetic Architectures Among Rapidly Speciating Cichlids

Matthias Sanetra and Axel Meyer

Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology,

University of Konstanz, 78457 Konstanz, Germany

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. EU564211–EU564333
Running title: Genetic Map of the cichlid fish *Astatotilapia burtoni*

*Keywords: Astatotilapia burtoni, Hox cluster, linkage map, microsatellites, ESTs*

1*Corresponding author: Axel Meyer, Lehrstuhl für Zoologie und Evolutionsbiologie, Fachbereich Biologie, Universität Konstanz, Fach M617, Universitätsstrasse 10, 78457 Konstanz, Germany*

Email: axel.meyer@uni-konstanz.de

Tel: +49 7531 88 4163

Fax: +49 7531 88 3018
ABSTRACT

Cichlid fishes are extremely species rich and formed species flocks within record-breaking short time spans. We constructed a medium density genetic linkage map of *Astatotilapia burtoni* from the Lake Tanganyika region based on 208 microsatellite markers, 148 of which were newly developed. Sequences originated largely from a partial genomic library, but some EST and BAC clones were also used. The mapping cross was derived from two inbred laboratory lines to obtain F2 progeny by intercrossing. In correspondence with its karyotype, the map revealed 20 linkage groups spanning 2,265 cM of the genome (size ~950 Mb) with an average marker spacing of 10.7 cM. With adjacent microsatellites the seven *Hox* clusters, *ParaHox* *C1* and two paralogs of *pdgfrβ* were mapped each to different linkage groups, thus supporting the fish-specific genome duplication hypothesis. Genes for *ovarian cytochrome P450 aromatase* (*CYP19A1*), *long wavelength-sensitive opsin* (*LWS*) and *microfibril-associated glycoprotein 4* (*magp4*) were mapped and determined to be interesting candidate genes for cichlid speciation. The *A. burtoni* linkage map was compared to the other two available maps for cichlids, *Oreochromis* spp. and *M. zebra/L. fuelleborni*, using shared markers. This linkage map will facilitate efficient genome scans (QTL) and future comparative genomic analyses of cichlids.
The cichlid fishes of East Africa provide well-known examples for rapid
diversification and explosive speciation owing to their phylogenetically young age
and therefore comparatively extremely short evolutionary time span for the origin of
more than 2,000 species (MEYER et al. 1990; MEYER 1993; KORNFIELD and SMITH
2000; VERHEYEN et al. 2003; KOCHER 2004; SALZBURGER and MEYER 2004;
SALZBURGER et al. 2005). Astonishingly large numbers of species make up the three
species flocks, each composed of hundreds of species, in the Lakes Victoria, Malawi,
and Tanganyika (FRYER and ILES 1972). Despite this huge phenotypic diversity
displayed by each of the species flocks, molecular phylogenetic studies on this
problem revealed that many of the species evolved similar morphologies convergently
in each of these three adaptive radiations (Meyer et al. 1990; KOCHER et al. 1993;
MEYER 1993; STIASSNY and MEYER 1999). These striking phenotypic similarities
among cichlid fishes from different species flocks that evolved in parallel make the
study of the underlying genetic architecture of cichlids particularly interesting.

The observed redundant patterns in the evolutionary diversification of cichlid
fishes support the view that the three large East African lakes are a natural experiment
of evolution, in which this repeated evolution might ultimately help to better
understand the processes that led to the repeatedly evolved patterns of diversification.
Particularly the species of cichlids of the Lakes Victoria and Malawi adaptive
radiations are very young and genetically extremely similar. Comparative studies on
the genomic organization of these closely related yet morphologically diverse fishes
will help to unravel the genetics of speciation (Kocher 2004; Albertson and Kocher 2006; Hoegg et al. 2007). Investigation of the molecular basis of those different phenotypes, i.e., the genetic and transcriptional changes that underlie differences among organisms, can be achieved through detailed comparisons of genome and transcriptome scans also including candidate gene approaches (Streelman and Kocher 2000; Braasch et al. 2006; Salzburger et al. 2007; Gerrard and Meyer 2007; Salzburger et al. 2008). For instance, the gene for long wavelength-sensitive opsin (LWS) has been reported to be involved in ecological adaptation and mate choice of cichlids (Carleton et al. 2005; Maan et al. 2006; Terai et al. 2006), while a microfibril-associated glycoprotein (magp4) is a good candidate for examining species differences with regard to jaw development (Kobayashi et al. 2006).

Species-specific linkage maps have recently become established as important genetic tools in an effort to aid in the more detailed knowledge of genotype-phenotype relationships (Albertson et al. 2003; Erickson et al. 2004; Albertson and Kocher 2006). The latter approach is known as quantitative trait locus or QTL scan, which makes use of the linkage disequilibrium created through experimental crosses between different species or laboratory strains (Falconer and Mackay 1996; Lynch and Walsh 1998). The ability to produce fertile interspecific crosses among some of these species of cichlids (Crapon de Caprona 1984) and the general popularity in captive breeding, support the establishment of cichlid fishes as a model system in comparative evolutionary genomic research.
The cichlid species *Astatotilapia burtoni* occurs in Lake Tanganyika and the surrounding river systems exhibit a rather generalist life style and is likely to represent a relatively ancestral type of cichlid (SALZBURGER *et al.* 2005). Its phylogenetic placement “between” the species flock of +/- 500 endemic species of Lake Victoria and up to 1000 endemic species of Lake Malawi makes it a very interesting species to study in this regard (MEYER *et al.* 1991). Since *A. burtoni* occupies a crucial phylogenetic position at the base of the extremely species-rich tribe of cichlids, the Haplochromini (SALZBURGER *et al.* 2002), which make up the large radiations of lakes Victoria and Malawi, its genome can serve as a sort-of base line from which comparisons to the endemic cichids of these lakes will be exceptionally insightful. Given these close genetic affinities, most of the genomic resources developed for *A. burtoni* will also be applicable to the large haplochromine cichlid species flocks from lakes Victoria and Malawi.

For *A. burtoni*, a BAC library (LANG *et al.* 2006) as well as expressed sequence tags (ESTs) have been generated (SALZBURGER *et al.* 2008), and cDNA microarrays are available as well (RENN *et al.* 2004; W. SALZBURGER, H. A. HOFMANN and A. MEYER unpublished results). In addition, there is detailed knowledge on *Hox* genes (HOEGG and MEYER 2005; HOEGG *et al.* 2007), *Para-Hox* genes (SIEGEL *et al.* 2007) and several other genes related to coloration (BRAASCH *et al.* 2006; SALZBURGER *et al.* 2007) and fertilization (GERRARD and MEYER 2007) for this key species. Genomic resources available from other cichlids include the Tilapia, *Oreochromis niloticus* (BAC library: KATAGIRI *et al.* 2001; Genetic maps: KOCHER *et al.* 1998; LEE *et al.*
Genetic Map of *Astatotilapia burtoni* M. Sanetra and A. Meyer

2005; Katagiri *et al.* 2005), *Haplochromis chilotes* (BAC library: Watanabe *et al.* 2003; ESTs: Watanabe *et al.* 2004; cDNA microarrays: Kobayashi *et al.* 2006), and *Metriaclima zebra* (BAC library: available at http://hcgs.unh.edu/BAC/Metriaclima.html; ESTs: Salzburger *et al.* 2008; Genetic map: Albertson *et al.* 2003). Recently, the National Institute of Health (NIH) has committed to sequencing four cichlid genomes. A detailed description of genomic resources developed for cichlid fishes can be found at http://hcgs.unh.edu/cichlid/.

AFLPs and microsatellite loci (also termed SSR) are the most common markers used in the development of linkage maps and QTL studies. Microsatellites are preferable because of their codominant nature and extremely high degrees of intraspecific allele polymorphism, which makes them most effective. On the other hand, their generation requires high costs and is time consuming (reviewed in Erickson *et al.* 2004). Most linkage maps based on microsatellites have been constructed for economically important fish species, such as the Atlantic salmon (GilbeY *et al.* 2004), rainbow trout (Sakamoto *et al.* 2000), European sea bass (Chistiakov *et al.* 2005), and Nile tilapia (Kocher *et al.* 1998; Lee *et al.* 2005) in order to search for loci that affect commercially important traits. Research in the fields of ecology and evolution has recently become interested in identifying the genetic basis of adaptive trait evolution especially in natural populations of non-model organisms. The past decade has seen a proliferation of studies that employ linkage maps together with QTL approaches to shed light on evolutionary processes, for instance the parallel evolution of benthic and limnetic forms in threespine...
sticklebacks (Peichel et al. 2001; Colosimo et al. 2004; Miller et al. 2007) and reduction of eyes and pigmentation in the Mexican cavefish, Astyanax mexicanus (Protas et al. 2007).

Here we report on the construction of a microsatellite linkage map of the cichlid fish A. burtoni based on an F2 intercross derived from two inbred laboratory strains. We identified 20 linkage groups. The map also incorporates some EST based markers and nuclear genes from sequenced BAC clones, e.g., the seven reported Hox genes and the two paralogs of pdgfrβ, a gene involved in coloration (Braasch et al. 2006). This linkage map will thus provide a useful future tool in studying the genetic basis of adaptive traits that played a major role in the rapid diversification of cichlid fishes.

MATERIALS AND METHODS

Experimental crosses: We crossed an A. burtoni female derived from our University of Konstanz stock with a male stemming from a laboratory stock that originated in the laboratory of Russell D. Fernald at Stanford University and is now also held at the University of Texas at Austin (Hans A. Hofmann). The stocks are originally from the Tanzania and Zambia regions of Lake Tanganyika respectively. Both stocks had been kept in the laboratory for about 10 years and therefore were significantly inbred. The resulting F1 generation was raised to sexual maturity and groups of several females with one or two males were established for the F1 intercross. Young fry of the F2 generation were taken from the mouths’ of F1 females usually consisting of 10-50 individuals. Genotyping at 10 microsatellite markers revealed
family relationships within each group. The final mapping population included 168 F2 offspring derived by intercrossing one male with five different females, thus constituting a half-sib family. However, we first established linkage groups by genotyping a subset of the first 90 F2 individuals that were born and then added the remaining 78 individuals for those markers with low LOD scores and/or those with some missing data.

**Microsatellite markers:** A microsatellite enriched library was prepared from *A. burtoni* DNA using a magnetic bead enrichment protocol and (CA)$_{15}$ and (CT)$_{15}$ probes (for a detailed description see SANETRA and MEYER 2005). A total of 1156 clones were sequenced, and 683 clones (enrichment rate ~60%) contained repeat motifs (including ~10% duplicate clones). Primer sets were designed for 278 putative loci using the Primer3 software (ROZEN and SKALETSKY 2000). Markers were considered informative when at least one parent was heterozygous, which was the case for 148 of these loci.

Additional 191 microsatellite primer sequences were collected from available genomic resources for other cichlids and their usefulness for mapping the *A. burtoni* genome was tested. Finally, we were able to employ a total of 60 informative markers derived from the Tilapia *Oreochromis niloticus* (**KOCHER** et al. 1998; UNH106, UNH152, UNH130, UNH192), *Copadichromis cyclicos* (**KELLOG** et al. 1995; UNH002), *Astatoreochromis allaudi* (**WU** et al. 1999; OSU9D, OSU13D, OSU19T, OSU20D), *Tropheus moorii* (**ZARDOYA** et al. 1996; TmoM5, TmoM7, TmoM27), *Pundamilia pundamilia* (**TAYLOR** 2002; Ppun1, Ppun5-7, Ppun12, Ppun16, Ppun18-20,
Ppun24, Ppun34-35, Ppun41), and *Metriaclima zebra* (ALBERTSON et al. 2003; UNH2004, 2005, 2008, 2032, 2037, 2044, 2046, 2056, 2058, 2059, 2069, 2071, 2075, 2080, 2084, 2094, 2100, 2104, 2112, 2116, 2117, 2125, 2134, 2139, 2141, 2149, 2150, 2153, 2163, 2166, 2169, 2181, 2185, 2191, 2204). A comprehensive list of the markers used is provided in Table S1.

We also searched for microsatellite repeat motifs in 9,375 non-redundant cDNA clones derived from a library using *A. burtoni* brains and mixed tissue including both sexes and all stages of development (SALZBURGER et al. 2008). For 21 microsatellite containing cDNA clones we developed PCR primers, 13 of which were polymorphic and gave reproducible results. Genbank accession numbers for these markers are as follows: Abur221 -CN470695; Abur223 - CN469772; Abur228 – CN470764; Abur224 – DY626128; Abur225 – DY630453; Abur226 – DY630491; Abur227 – DY626763; Abur230 – DY626468; Abur233 – DY629660; Abur234 – DY630828; Abur235 – DY627273; Abur239 – DY629088; Abur240 – DY630681. We compared these sequences to the cDNA and peptide database of medaka (*Oryzias latipes*), as the most closely related model organism to cichlids, using Blastview at www.ensembl.org/Multi/blastview.

**Genotyping procedures:** Microsatellites were amplified in 10-µl PCR reactions containing 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM dNTPs, 4 pmol of each locus specific primer, 0.8 units *Taq* polymerase (Genaxxon), and 10-30 ng genomic DNA. Forward primers were labeled with a fluorescent dye (6-FAM, HEX, or NED). In addition, for 212 of the initially designed primer pairs, the M13 method
for fluorescent labelling of PCR products (Schülke 2000) was used for economic reasons (see Table S1). With the latter method, 1 pmol of forward primer, 4 pmol reverse primer, and 4 pmol M13 6-FAM or HEX-labelled primer were used. For the second round of genotyping, we used a PCR multiplexing kit (Qiagen) to amplify 3-5 loci in a single PCR reaction in a 12 µl volume containing 6.25 µl of 2x Qiagen Multiplex PCR Master Mix and 1.25 µl of a mix of primers. Final concentration of primers was 0.2 µM. All PCRs were run on a Perkin Elmer (Norwalk, CT) GeneAmp PCR 9700.

Three basic temperature protocols depending on labelling method and multiplex scheme were used. **Forward primer labelling:** one cycle of 3 min at 94°C, 35 cycles at 94°C for 30 sec, 48-58°C for 30 sec, 72°C for 90 sec, and a final extension step at 72°C for 10 min. **M13 primer labelling:** one cycle of 3 min at 96°C, 5 cycles at 96°C for 30 sec, 62-56°C for 30 sec, 72°C for 30 sec, 35 cycles at 96°C for 30 sec, 58-53°C for 30 sec, 72°C for 30 sec and a final extension step at 72°C for 10 min. **Multiplex PCR:** 95°C for 15 min, then 94°C for 30 sec, 50-60°C for 90 sec, 72°C for 60 sec. 35 cycles of PCR were performed with a final extension period of 30 min 72°C (for details see Table S1). PCR products for 4-6 loci were combined with a mixture of ABI Genescan-500 ROX size standard and analyzed with an ABI 3100 Automated Sequencer (Applied Biosystems). More details on fluorescent dye labelling, multiplexing schemes, and annealing temperatures ($T_a$) for each locus are given in supplemental Appendix S1 at http://www.genetics.org/supplemental/. Allele sizes were scored with the Genotyper.
3.7 (Applied Biosystems) software package and transferred to an electronic spreadsheet.

**Type I markers:** Several clones containing interesting candidate genes had been sequenced from the BAC library of *A. burtoni* (*Lang et al. 2006*). These clones incorporated the homologous sequences of *pdgfrβ* (*Braasch et al. 2006*), and all *Hox* genes reported in *A. burtoni* (*Hoeegg et al. 2007*). Putative microsatellites were derived from the clone sequences with the Tandem Repeat Finder v. 3.2.1 software (*Benson 1999*), so that two informative markers were obtained from BAC clone 26M7 containing *pdgfrβ*a (Abur209, 212) and one for 20D21 containing *pdgfrβ*b (Abur218). The corresponding BAC clones (clone number, accession number) for *HoxAa* (116M8, EF594313), *HoxAb* (150O18, EF594311), *HoxBa* (170E12, EF594310), *HoxBb* (34I18, EF594314), *HoxCa* (103K21, EF594312), *HoxDa* (32B18, EF594315), and *HoxDb* (19E16, EF594316) each yielded one to three polymorphic microsatellite markers. A tetranucleotide ATCT repeat was used for marker development of clone 99M12 (*Siegel et al. 2007*), which includes *ParaHoxC1* representing a dispersed *Hox*-like gene cluster.

A homolog of the human *microfibril-associated glycoprotein 4 (magp4)* has been reported from cichlid EST clones derived from *Haplochromis chilotes* (*Kobayashi et al. 2006*). We used three of these clones (Acc. No. BJ679835, BJ676254, BJ680594) to form 1,214 bp of continuous sequence (Fig. 1). We initially designed two pairs of primers to amplify from genomic DNA of *A. burtoni*, Magp4_1F (5’-TCAGACCTCCACCAAACAGTC) and Magp4_1R (5’-
The long wavelength-sensitive opsin (LWS) has been characterized from a BAC library of *H. chilotes*. Numerous primers to amplify upstream and downstream regions of the gene are available (Terai et al. 2006). We used the following primers to amplify this region from genomic DNA in *A. burtoni* all giving ~1 kb fragments: LWSB_LF and LWSB_R1, LWSB_F2 and LWSB_R3, LWSB_F3 and LWSB_R4, LWSB_F4 and LWSB_R5, LWSB_F8 and LWSB_R9, LWSB_F9 and LWSB_R10.

Primers AroI_Ex7F (5’-GGTGATCGCAGCTCCGGACACTCTCTCC), AroI_Ex8R (5’-CCTGTGTTCAAGATGATGTTGTGC), and AroI_1600R (5’-GTACAGCTAAAGGTTCGGGTC) to amplify 600–1,200 bp of ovarian cytochrome *P450 aromatase (CYP19A1)* were designed using a genomic sequence from tilapia deposited in GenBank (AF472620) and sequences from Lake Malawi cichlids at http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml (D. T. Gerrard and A. Meyer unpublished results). Following SNP identification, nested primers, AroI_Fw (5’-ATGGCTGCATTCCACCAC) and AroI_Rv (5’-TTCTTCATGCTTCTGCTCCTC),
were designed for *A. burtoni* producing 447 bp of intron sequence located between exon7 and 8.

The primary sequences of the three gene regions, *magp4*, *LWS*, and *CYP19A1*, were amplified using annealing temperatures of 55°C, 58°C, 54°C, respectively, and the following PCR conditions: one cycle of 3 min at 94°C, 35 cycles at 94°C for 30 sec, 55-58°C for 30 sec, 72°C for 90 sec, and a final extension step at 72°C for 10 min. Resulting sequences of *A. burtoni* were screened for SNPs in the parents of the mapping cross, and identified SNPs were then sequenced and scored using nested primers (with annealing of 55°C) in 168 F2 individuals.

**Linkage mapping:** Linkage distances and maker orderings were determined with the LocusMap 1.1 software (GARBE and DA 2003). The Kosambi mapping function, which accounts for double recombinations, was used to convert recombination frequencies to centimorgans (cM) for all analyses. Non-inheritance errors provided by Locusmap were checked by re-evaluating the original chromatograms and either corrected or omitted from the dataset. Most of these errors were due to rounding errors of a 1-base difference in allele size, others were classified as possible allele-drop-outs or allele mutations. The assignment of markers to linkage groups was carried out by using a sex-averaged LOD-threshold of ≥4.0. This value was increased from the commonly used LOD score of 3.0 to minimize the risk of false linkage due to the large number of two-way tests being performed (OTT 1991). From the likelihood scores of marker orders, the most likely order of three or more markers
within a linkage group could be inferred. Graphics of the linkage groups were produced with the MapChart 2.1 software (VORRIPS 2002).

RESULTS

**Polymorphic microsatellites:** We have characterized 278 new microsatellite sequences from genomic DNA of *A. burtoni*, 225 of which could be used to amplify PCR products of the expected size. One hundred forty-eight markers were informative (at least one parent heterozygous), while 49 were not variable in this mapping cross. The remaining 28 loci showed banding patterns that were difficult to interpret and thus could not be reliably scored. An additional 60 microsatellite markers were derived from prior studies which employed microsatellite markers in cichlids (KOCHE ET AL. 1998; KELLOG ET AL. 1995; ZARDOYA ET AL. 1996; TAYLOR 2002; ALBERTSON ET AL. 2003). Most of the microsatellites used consist of pure and compound dinucleotide tandem repeats, mainly comprising CA and to a lesser extent CT repeat motifs. In addition, there were 13 tetranucleotide loci developed from the genus *Pundamilia* (TAYLOR 2002) and one from a BAC clone containing *ParaHoxC1*. An investigation of the relationship between repeat length and rate of polymorphism showed no significant correlation, however no repeats shorter than 8 times were used initially.

**A. burtoni linkage map:** Significant linkages were identified for 202 genetic markers, including 189 microsatellite loci and 13 type I (gene) markers (Fig. 2). Only nine markers could not be linked to any other marker, which gives a proportion of
linked markers of 95%. A total of 188 markers were genetically distinct, whereas 14 markers could not be distinguished on the basis of recombination and thus were considered as pairs of two markers each on the map. We found 20 linkage groups (LG) with the number of markers per group ranging from three (LG10) to 32 (LG3). The number of linkage groups found does not differ from the 20 chromosomes expected from the karyotype (THOMPSON 1981) while incomplete but large genetic maps usually have more linkage groups than chromosomes (e.g., LEE et al. 2005). The largest linkage groups were LG3, LG4, and LG7 with 373, 349, and 236 cM, respectively. Accordingly, the *A. burtoni* karyotype shows at least two large subtelocentric-telocentric chromosomes, although karyotypic size variation among chromosomes was not as pronounced when compared to *Tilapia* and *Sarotherodon* (THOMPSON 1981).

The total sex-averaged length of the map was 2,265 cM. Marker spacing was on average 10.7 cM, however the range was quite variable from 0.4 to 154 cM. Thirty-five percent of marker intervals were <5 cM long, 27% ranged between 5 and 10 cM, 19% varied between 10 and 20 cM, and 19% were >20 cM. Since the genome size of this species has been reported to be 0.97 pg (LANG et al. 2006), which approximately equals 950 Mb, we estimated the physical-to-map distance being 420 kb/cM. Although we used fewer markers, this value is comparable to the second generation map of *Tilapia* due to the larger genome of that latter species (LEE et al. 2005).

**Hox gene clusters:** In *A. burtoni*, seven *Hox* gene clusters have been detected by means of BAC library screening and sequencing of positive clones (HOEGG et al. 2005).
The mapping results using microsatellite flanking regions revealed a distinctive distribution of *Hox* clusters throughout the genome. *HoxAa*, *HoxAb*, *HoxBa*, and *HoxBb* were mapped to LG2, LG18, LG7, and LG20, respectively. While *HoxCa* was assigned to LG5, *HoxDa* and *HoxDb* mapped to LG9 and LG16, respectively. Overall, this pattern supports the hypothesis of a fish-specific genome duplication, because all *Hox* clusters are found on different chromosomes (HOEGG and MEYER 2005; MEYER and VAN DE PEER 2005). *ParaHoxC1* is on LG3 with no association to one of the other conventional *Hox* clusters corroborating the idea that these *Hox*-like genes constitute dispersed homeobox genes forming novel clusters somewhere else in the genome (Siegel *et al.* 2007). There is also close physical association of *ParaHoxC1* to the genes *pdgfrα* and *kita* being involved in teleost coloration (PARICHY *et al.* 1999; SALZBURGER *et al.* 2007).

**Platelet-derived growth factor receptor β (pdgfrβ):** Two paralogons of this gene are known to be present in cichlids due to the fish-specific genome duplication (BRAASCH *et al.* 2006). The mapping of adjacent microsatellite markers of type (GT/CA)$_{12-15}$ from corresponding BAC clones showed the location of *pdgfrβa* on LG4, while *pdgfrβb* was mapped to LG14. In the same BAC clone, the two markers Abur209 and Abur212 are flanking the *pdgfrβa–csf1r* tandem, spanning approximately 50kb between them. It is thus not surprising that they are regarded identical on the map. The observation of *pdgfrα* (see the above part on *ParaHoxC1*) and the A- and B-copies of *pdgfrβ* occurring on different linkage groups, as in *Hox*...
genes, favors the origin of this gene family by whole genome duplication and not due to tandem duplication.

**Long wavelength-sensitive opsin (LWS):** In the sequence upstream of the LWS-gene using the primers LWSB_F2 and LWSB_R3, we found a SNP at position 231 (using nested primers) in the parents of the mapping cross. The male parent was CT heterozygous and the female parent was TT homozygous, thereby permitting linkage analyses through genotyping of F₂ individuals. The analysis revealed this gene important for color vision and probably speciation by female mate choice (TERAI et al. 2006) to be located on LG13 in *A. burtoni*. LWS does therefore not appear to be linked to any other candidate gene employed in this study.

**Microfibril-associated glycoprotein 4 (magp4):** Amplification from genomic DNA with the two primer pairs directly derived from EST clones of *H. chilotes* led to the discovery of an intronic region in sequence BJ679835 (Fig. 1). Since the expected size of the product from cDNA was 544 bp and the observed size was 2.2 kb, the size of the intron could be estimated to be ~ 1.65 kb in length. Within this intron we identified a SNP at position 392 in the sequence given by primer Magp4_2R when aligned to clone BJ679835. At this site the male parent was TT homozygote and the female parent was CC homozygote, while all F₁ individuals were CT heterozygotes, as expected. We also compared the amplified coding regions between *H. chilotes* and *A. burtoni*, which displayed 15 substitutions over 195 bp. For SNP genotyping of the F₂ we used nested primers in the intron region yielding a shorter fragment of 339 bp, subsequent linkage analyses of which showed the *magp4* locus to be positioned on the
small LG20 (five markers, 22cM). The latter also comprises the HoxBb cluster and the marker UNH2069, which shows an associated QTL for jaw morphology in the Malawi cichlid *Metriaclima* (ALBERTSON et al. 2003).

**Ovarian cytochrome P450 aromatase (CYP19A1):** We were able to identify two SNPs (AF472620:g.4167A>C; AF472620:g.4440G>T) in the parental DNA of the *A. burtoni* mapping cross, one in the intron bridging exons 7 and 8 and the second one in exon 8. The male parent was g.4167CC and g.4440GG homozygote, while the female parent was g.4167AA and g.4440TT homozygote. Linkage analysis with microsatellite markers using an F2 intercross revealed the map location of *CYP19A1* on LG11 in *A. burtoni*. In the Nile tilapia, *CYP19A1* was found in the vicinity of the presumed SEX locus on LG1 (LEE AND KOCHER 2007). This linkage group corresponds to the original LG6 reported by ALBERTSON et al. (2003) for *Metriaclima*, but this has now been renamed in accordance to the *Tilapia* map as LG1 (ALBERTSON et al. 2005).

**EST-linked microsatellites:** Thirteen informative microsatellite markers (Abur 221, 223-228, 230, 233-235, 239-240; Table S1) were produced from EST clones of *A. burtoni*. These markers occurred widely distributed on several different linkage groups, such as LG3-5, LG7, LG9, LG11, LG13-14, and LG19, which would make them suitable for genome scans in order to discover functional polymorphisms. Comparisons of *A. burtoni* EST clones with known sequences of medaka revealed a few homologies with protein coding genes. The sequence adjoining marker Abur 224 on LG7 was indicated as part of the transcript of *deoxyhypusine synthase (DHS)*,
which occurs in a single copy in medaka and is essential for cell viability. Despite the short overlap of 42bp and therefore high e-value (2.2e-11, PID 79%), alignment structure and orientation to the poly-A tail with the microsatellite in the 3’UTR strongly support this assumption. Abur226 on LG14 (6.4e-42, PID 78%) corresponds to protein phosphatase 1 regulatory subunit 14A (CPI-17), a cytosolic inhibitory protein of PP1 with a molecular weight of 17kDa. Abur233 on LG9 (5.6e-50, PID 77%) was found adjacent to a member of the protein family keratin, type I (cytoskeletal cytokeratin) accounting for the keratin filaments in epithelia. Teleost fish show an excess of keratin type I over type II genes, thus as many as 17 type I gene members are present in medaka. For more annotations of EST sequences from this library see SALZBURGER et al. (2008).

**Comparison between the A. burtoni and M. zebra/L. fuelleborni map:** A genetic linkage map is available from a hybrid cross of two closely related Lake Malawi cichlids, *Labeotropheus fuelleborni* and *Metriaclima zebra* (ALBERTSON et al. 2003). It contains 127 microsatellite markers, of which we were able to use 35 as informative markers in the linkage map of *A. burtoni*. A comparison of the two maps revealed good concordance in some parts, in which all markers located in a single linkage group in Lake Malawi cichlids, e.g. LG2 with UNH2037, 2059, and 2080, were found in a single LG12 in *A. burtoni* as well. In addition, numerous markers appeared jointly albeit on differently named linkage groups in the two maps, thus enabling patterns of correspondence of those linkage groups. It appears that *A. burtoni* linkage groups LG1, LG4, LG7, LG9, LG11, and LG12 correspond to LG3, LG4,
LG16, LG5, LG6, and LG2 in the Lake Malawi cichlid map, respectively, using the criterion of at least two shared markers. The observed correspondences of microsatellite markers are reported in Table 1. The largest difference between the location of markers shared between the two maps occurred in LG3 in *A. burtoni*, which comprised markers from both LG1 and LG10 in Lake Malawi cichlids also in different orders. This result, however, is not surprising since markers in the Lake Malawi map were assigned to 24 linkage groups, while the map of *A. burtoni* comprises only 20 linkage groups.

**Comparison between the *A. burtoni* and the Tilapia (*Oreochromis* spp.) map:**

We initially screened 51 microsatellite markers from the available linkage maps of the tilapia (Kocher *et al.* 1998; Lee *et al.* 2005) for their use in linkage mapping of *A. burtoni*, and found rather low levels of polymorphism in our test cross, although amplification success was around 50%. Thus, only a small number of markers are shared between the two maps, namely UNH106, UNH130, UNH192, UNH2150, UNH2166, and UNH2191, which could be used for comparison. In the tilapia map, containing 24 linkage groups, markers on LG3, LG9, LG11, LG15, and LG23 indicated correspondence of these linkage groups to LG17, LG8, LG18, LG7, and LG3, respectively, in the *A. burtoni* map.

SNP mapping of *CYP19A1* in *A. burtoni* revealed this gene on LG11, which therefore has a closer relationship to LG1 in tilapia containing *SEX* and *CYP19A1* (Lee *et al.* 2006). In accordance to that, Albertson *et al.* (2005) have renamed LG6 from an earlier paper on Metriaclima (Albertson *et al.* 2003) to be now LG1 as in
tilapia. Hence, similarities among all three East African cichlid maps could be used to find out probable synonymies of linkage groups. An attempt to reconcile the nomenclature of present maps is depicted in Table S2. Also, taking the tilapia map as a standard some of these relationships were included in Figure 2.

**DISCUSSION**

This is the third genetic map of a cichlid fish. *Astatotilapia burtoni* is from Lake Tanganyika and its surrounding rivers. It significantly adds to the knowledge of previous linkage maps for the generalist and geographically widespread tilapia, *Oreochromis* spp. (KOCHER et al. 1998; LEE et al. 2005), and the specialist Lake Malawi endemic *M. zebra/L. fueilleborni* (ALBERTSON et al. 2003). In the *A. burtoni* map we identified linkages among 202 genetic markers, mainly microsatellites and 13 type I (gene) markers, which were assigned to 4 larger and 16 smaller linkage groups. This observation corresponds quite well to the 20 chromosomes and their morphology reported from investigations of the karyotype (THOMPSON 1981). While most chromosomes are relatively small and of metacentric-submetacentric or metacentric type, there are also four large subtelocentric-telocentric chromosomes displayed by the karyotype. Lake Malawi cichlids, on the other hand, have a slightly higher chromosome number with N = 23 (THOMPSON 1981), which is also in agreement with the 24 linkage groups found by ALBERTSON et al. (2003) for an interspecific cross between *M. zebra* and *L. fueilleborni*. The current map of the tilapia, though one of the
most detailed fish linkage maps, still shows 2 linkage groups more than is expected from the 22 chromosomes of the karyotype (LEE et al. 2005).

We were able to establish several correspondences of the *A. burtoni* and the Lake Malawi cichlid map, suggesting that the increase in chromosome number might have been partly due to a fission of large LG3 in *A. burtoni* having generated linkage groups 1 and 10 in *M. zebra*. Concerning most other linkage groups that could be compared between the two species (Table 1), syntenies of microsatellite markers were common with only a few unexpected syntenies among shared markers, such as UNH2084 on LG4 instead of LG1 in *A. burtoni*. The correspondence of markers used for the map of tilapia could for a large part not been directly explored because of the evolutionary distance between the species. Many of these markers amplified did not show polymorphism in the *A. burtoni* mapping cross so that recent views on the wide application range of the tilapia map for the > 2000 species of cichlid fishes in the East African lakes (LEE et al. 2005) might have been too optimistic. Nevertheless, using the combined information from the three maps of East African cichlids many linkage groups could be brought in line with the tilapia map on the basis of shared markers (see Table S2, Figure2).

Comparison of the map location of the gene for *ovarian cytochrome P450 aromatase (CYP19A1)* on LG1 in tilapia (LEE and KOCHER 2007) together with overlapping markers used in *M. zebra* and *A. burtoni* (Table S2), renders the synonymy of LG6 in *M. zebra* and LG11 in *A. burtoni* to that tilapia LG1 highly likely (see also ALBERTSON et al. 2005). *CYP19A1* is involved in sex differentiation
of mammals, and could also be important in determining sex in vertebrate species that lack sex chromosomes. However, in the Nile tilapia, this gene was found 27 cM away from the presumed sex-determining locus on LG1, rendering its function as a master control gene for sex determination unlikely. In general, the sex locus in tilapia (*O. niloticus*) behaves like an XY male heterogametic system (Lee et al. 2003; Lee and Kocher 2007). On the other hand, Lee et al. 2004 found microsatellite markers consistent with a WZ (female heterogametic) system on LG3 in *O. aureus*. Thus, the mechanism of sex-determination appears highly variable among species of African cichlids as is the case in fishes more generally, where the differentiation is influenced by both genetic and environmental factors (Volff et al. 2007). The occurrence of *CYP19A1* on LG11 in *A. burtoni* therefore raises the possibility that a sex determining factor is located on this chromosome as well, in the event that Nile tilapia and haplochromine cichlids have retained similar mechanisms of sex determination.

Using microsatellites from BAC clone sequences, we were able to map all seven clusters of Hox genes (*Aa-Ca, Da, Db*) that have been reported from *A. burtoni* (Hoeegg et al. 2007) as well as *ParaHoxC1* (Siegel et al. 2007). The surprising variation of Hox clusters among vertebrates has been widely used to study the evolution of vertebrate genomic organization (Hoeegg and Meyer 2005). While the origin and timing of the four Hox clusters in Tetrapods is still much debated, there is mounting evidence that the eight clusters in ray-finned fish originated by whole genome duplication (3R hypothesis) (Malaga-Trillo and Meyer 2001; Hoeegg et al. 2004; Meyer and Van de Peer 2005). In accordance to that hypothesis, we found
all Hox clusters scattered throughout the genome and no linkages were established among them. Similarly, zebrafish have seven Hox clusters on seven different chromosomes (POSTLETHWAIT et al. 1998) but the Db cluster was lost instead of the Cb cluster in A. burtoni (HOEGG et al. 2007). The ParaHox C1 paralogon mapped to the center of the largest LG3, which did not carry any genes of the Hox complex.

Sequence comparisons showed that ParaHox C1 and its 3’ adjoining genes of Danio rerio is located on chromosome 20 (SIEGEL et al. 2007). In general, however, synteny of the duplicated genes in teleosts seem to be less conserved in the ParaHox genes compared to the Hox genes.

According to many studies, one important aspect for the vast amount of cichlid diversification is the significant role of color morphs in different populations and species with regard to mate choice and male-male competition (e.g., TURNER and BURROWS 1995; SEEHAUSEN and SCHLUTER 2004). Therefore, the mapping of candidate genes involved in teleost coloration can serve as a valuable tool to study speciation processes in cichlids. For instance, the orange blotch color pattern in M. zebra, which is expressed mainly in females, has been mapped to a particular chromosomal region and candidate genes for this pigmentation phenotype have been identified (STREELMAN et al. 2003). We mapped two paralogons of the tandems pdgfrβ-csf1r and one of pdgfrα-kit (being adjacent to ParaHoxC1), a family of receptor tyrosine kinase genes that have been shown to influence coloration in teleosts (Braasch et al. 2006; Salzburger et al. 2007). While the kit gene is essential for the development of neural crest-derived dark melanocytes in mammals and zebrafish
(PARICHY et al. 1999), csf1r promotes the development of yellow xanthophores in zebrafish (PARICHY and TURNER 2003) and there is some evidence for its developmental role in the generation of cichlid egg spots (SALZBURGER et al. 2007). The location of the pdgfrβ paralogons on two different linkage groups in the genetic map of *A. burtoni*, *pdgfrβ*a on LG4 and *pdgfrβ*b on LG14, lends further support to the occurrence of a fish-specific genome duplication. In general, the teleost A-paralogon has retained a longer stretch of synteny with the single copy of the tetrapod locus compared to the B-paralogon (see BRAASCH et al. 2006). It has therefore been suggested that the B-paralogon underwent some functional divergence (neofunctionalization) of the cell-surface receptors, rendering these duplicated receptor genes ideal targets for future QTL studies on cichlid coloration patterns.

Divergent evolution of the visual system is a likely mechanism to explain incipient speciation and diverse patterns in the males’ breeding coloration in cichlids (TERAI et al. 2006). Our study species *A. burtoni* is a close relative of the sibling species pairs in the genus *Pundamilia*, for which the sensory drive hypothesis (differences in male coloration evolving as a consequence of divergent visual sensitivities) has been proposed (MAAN et al. 2006). Specifically, the gene for *long wavelength-sensitive opsin* (*LWS*) shows the highest variability among cichlid opsins and appears to be under strong divergent selection at least in the Lake Victoria species flock (CARLETON et al. 2005). We were here able to locate the map position of this important gene in *A. burtoni* on LG13, which corresponds to LG5 in tilapia carrying the genes for the *Blue Opsin* and for *c-ski I* (LEE et al. 2005). Remarkably, the latter
marker has been found to be in close association to the orange blotch color polymorphism in *M. zebra* (Streelman et al. 2003). It will thus be interesting to examine through future advances in quantitative genetics whether other linked genes for coloration will be co-occurring on this particular chromosome.

Apart from color, evolutionary diversification in cichlids has also been driven by trophic specialization, which altered jaw morphologies and teeth shape in different species according to feeding type (e.g., Kocher 2004; Albertson and Kocher 2006; Streelman and Albertson 2006). A QTL study by Albertson et al. (2003) focusing on these feeding adaptations has shown that the oral jaws of Lake Malawi cichlids evolved in response to strong, divergent selection. In a closely related species pair of Lake Victoria cichlids, *Haplochromis chilotes* and sp. “rockkribensis”, showing divergent jaw types, a conspicuous difference in expression rates of the magp4 gene (encoding a microfibril-associated glycoprotein) was recently discovered (Kobayashi et al. 2006). This suggests that cichlid magp4 may be responsible for the disparity in jaw development between such morphologically different species. Magp4 is also known to be involved in a human heritable disease, the Smith-Magenis-Syndrome (SMS) resulting in a characteristic phenotype with flattened mid-face and striking jaw and forehead (Zhao et al. 1995). We mapped the ci-magp4 to the small LG20 in *A. burtoni*, which, in addition, comprises the *HoxBb* cluster and three anonymous microsatellite markers. The corresponding LG20 in *M. zebra* with marker UNH2069 carries a QTL for differences in lower jaw width (Albertson et al. 2003). Another candidate gene for craniofacial diversity, bone morphogenetic protein 4
Genetic Map of Astatotilapia burtoni

M. Sanetra and A. Meyer

(bmp4) occurs on LG19 in M. zebra (which corresponds to LG12 in A. burtoni), and explains more than 30% of the phenotypic variation in the opening and closing levers of the cichlid lower jaw (ALBERTSON et al. 2005).

We have mapped 13 EST based microsatellite markers derived from brain and mixed tissue libraries of A. burtoni (SALZBURGER et al. 2008) showing a widespread distribution among chromosomal regions. Homology searches using the medaka database resulted in three reliable hits constituting protein coding genes for deoxyhypusine synthase (DHS), protein phosphatase 1 (CPI-17), and a member of the keratin, type I protein family. Despite the observation that the remaining EST-linked A. burtoni markers stem from anonymous clones, for which no homology with known sequences could be established, they should nevertheless provide a useful source for studying gene-associated polymorphisms in different chromosomal regions. It has been suggested that EST-linked microsatellite genome scans provide an efficient strategy for detecting signatures of divergent selection, especially in nonmodel organisms (VASEMÄGI et al. 2005). In particular, the broad interspecific application range of microsatellite markers in East African cichlids with their large diversity in morphology, behavior and life-history patterns provides a good opportunity to use such markers with known chromosomal location for comparative studies on polymorphism and to examine the footprints of selection.

Genetic linkage maps are valuable genomic resources that have been widely used in a number of different applications but particularly notable are recent advances of the QTL approach in evolutionary studies (ERICKSON et al. 2004; SLATE 2005). One
of the most debated questions in evolutionary biology is whether major genes play a key role in species differences or whether a large number of small changes does (ORR 2001). QTL studies have yielded significant contributions toward this issue (e.g., PEICHEL et al. 2001; HAWTHORNE and VIA 2001; ALBERTSON et al. 2003). In threespine sticklebacks, for instance, it was shown that a major QTL can cause large shifts in phenotype with regard to the morphological differences between the sympatric benthic and limnetic forms (COLOSIMO et al. 2004). Studies in East African cichlids discovered new relationships of the genomic regions involved in feeding adaptations and jaw types (ALBERTSON et al. 2003; ALBERTSON et al. 2005), including the discovery of a region explaining approximately 40% of phenotypic variance in cichlid tooth patterning (cusp number) (STREELMAN and ALBERTSON 2006). It is thus promising that whole-genome sequencing has been initiated for several species of cichlid fishes including draft sequencing (5x) of tilapia (O. niloticus) and 2x coverage each of three haplochromine species (http://hcgs.unh.edu/cichlid/).

Astatotilapia burtoni, H. chilotes and M. zebra will be among them, so that this medium density genetic map of A. burtoni presented here can be used in conjunction with QTL analyses to pinpoint evolutionarily important genes (those that are presumably selectively adaptive). Comparisons of the sequenced cichlid genomes will then give insights into the similarity or differences in the underlying molecular changes that caused their phenotypic divergence among closely related species or caused convergent similarities among more distantly related cichlid species.
This work was supported by a grant from the Deutsche Forschungsgemeinschaft to A.M. (ME 1725/11-2). We thank W. Salzburger for assistance with obtaining the mapping cross and for many fruitful discussions on the subject. The people who helped with the lab work in terms of genotyping and sequencing were: C. Chang-Rudolf, I. Eichstätter, U. Ernst, E. Hespeler, D. Leo, C. Michel, K. Nübling, M. Pehr, N. Siegl, and H. Singer. C. Stemshorn gave some advice concerning the use of the LocusMap software, and S. Kuraku helped with searching the medaka database.

LITERATURE CITED


KIJIMOTO, T., M. WATANABE, T. KIJIMOTO, K. FUJIMURA, M. NAKAZAWA et al., 2006

magp4 gene may contribute to the diversification of cichlid morphs and their speciation. Gene 373: 126-133.


A genetic linkage map of a cichlid fish, the Tilapia (Oreochromis niloticus). Genetics 148: 1225-1232.


Sinauer Associates, Sunderland, MA.


POSTLETHWAIT, J. H., Y.-L. YAN, M. A. GATES, S. HORNE, A. AMORES et al., 1998


VORRIPS, R. E., 2002 MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. **93**: 77-78.


TABLE 1

Oxford plot comparing the linkage maps of *Astatotilapia burtoni* and *Metriaclima zebra/Labeotropheus fuelleborni*

| A. burtoni | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1          |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 2          |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 3          | 6 |   | 3 |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 4          |   | 1 | 3 |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 6          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 7          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 8          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 9          |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 11         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 12         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 13         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 14         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 15         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 16         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 17         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 18         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 19         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 20         |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
FIGURE LEGENDS

FIGURE 1. Schematic drawing of the amplification of the magp4 gene from A. burtoni.
Position of primers, alignment of EST clones of H. chilotes, and the location of an intron in A. burtoni derived from genomic DNA are shown, the clones and primers are not drawn to scale.

FIGURE 2. Linkage map of Astatotilapia burtoni comprising 20 linkage groups.
Corresponding linkage groups for tilapia are in parentheses as inferred from shared markers among the three African cichlid maps. Distances in Kosambi centimorgans are given left to each linkage group.
Intron ~ 1,650 bp        Exon 244 bp

A. burtoni genomic sequence

H. chilotes EST clones
FIGURE 2
FIGURE 2. – Continued
FIGURE 2. – Continued
FIGURE 2. – Continued