JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 306B:35–44 (2006)

A BAC Library of the East African Haplochromine Cichlid Fish *Astatotilapia burtoni*

MICHAEL LANG[1†], TSUTOMU MIYAKE[2], INGO BRAASCH[1], DEBORAH TINNEMORE[2], NICOL SIEGEL[1], WALTER SALZBURGER[1], CHRIS T. AMEMIYA[2*], AND AXEL MEYER[1*]
[1]*Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany*
[2]*Benaroya Research Institute at Virginia Mason, Seattle, Washington 98101*

ABSTRACT A BAC library was constructed from *Astatotilapia burtoni*, a haplochromine cichlid that is found in Lake Tanganyika, East Africa, and its surrounding rivers. The library was generated from genomic DNA of blood cells and comprises 96,768 individual clones. Its median insert size is 150 kb and the coverage is expected to represent about 14 genome equivalents. The coverage evaluation was based on genome size estimates that were obtained by flow cytometry. In addition, hybridization screens with five probes largely corroborate the above coverage estimate, although the number of clones ranged from 5 to 22 authenticated clones per single copy probe. The BAC library described here is expected to be useful to the scientific community interested in cichlid genomics as an important resource to gain new insights into the rapid evolution of the great species diversity of haplochromine cichlid fishes. *J. Exp. Zool. (Mol. Dev. Evol.) 306B:35–44, 2006.*

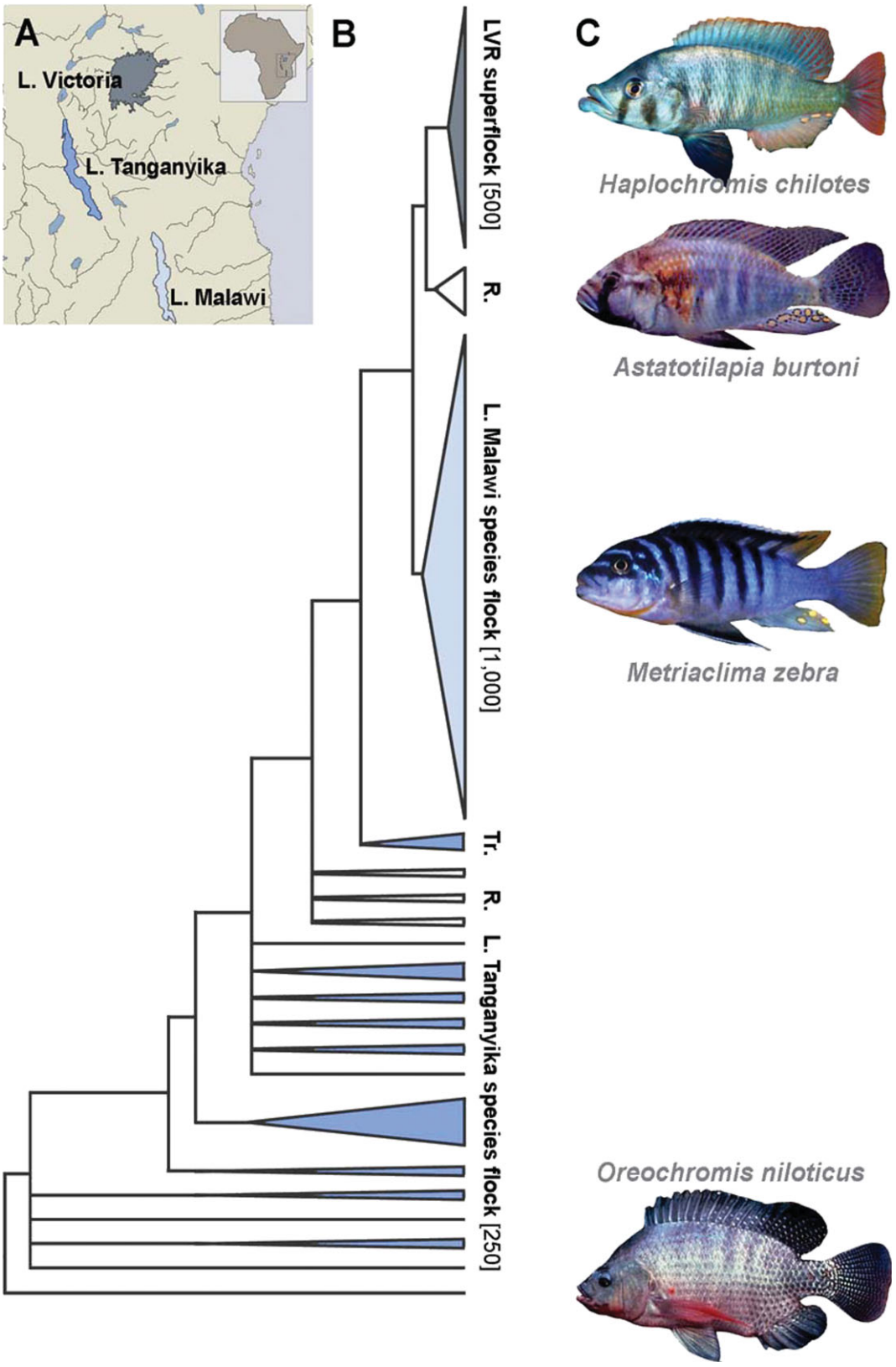
The species flocks of cichlid fishes of the East African Great Lakes Victoria, Tanganyika and Malawi are extraordinary examples for explosive speciation and adaptive radiation (Fryer and Iles, '72; Meyer et al., '90; Meyer, '93; Stiassny and Meyer, '99; Danley and Kocher, 2001; Turner et al., 2001; Kocher, 2004; Salzburger and Meyer, 2004). Each of the three lakes harbors several hundreds of endemic cichlid species that have evolved via intralacustrine speciation (speciation within the same lake). With an estimated number of about 1,700–2,000 species, the tribe Haplochromini represents, by far, the most species-rich assemblage of cichlid fishes. The entire species flocks of Lake Malawi and the Lake Victoria region superflock exclusively comprise haplochromines (Meyer, '93; Turner et al., 2001; Salzburger et al., 2002a, 2005; Verheyen et al., 2003; Salzburger and Meyer, 2004). The ages of these species flocks are dated at about 100,000 years for the Lake Victoria region superflock, and around 700,000 years for Lake Malawi (Meyer et al., '90; Kocher et al., '93; Danley and Kocher, 2001; Verheyen et al., 2003).

Molecular phylogenies have corroborated the derived status of haplochromines among the East African cichlid lineages, suggesting that the cichlid faunas of the Lake Victoria region and of Lake Malawi were seeded by generalists that inhabited rivers, swamp and marsh areas (Meyer et al., '91; Salzburger et al., 2002a, 2005; Verheyen et al., 2003).

The riverine haplochromine *Astatotilapia burtoni* (Günther, 1894) is thought to be a fitting representative of one of these potential founding lineages (Meyer et al., '91; Kocher et al., '93; Meyer, '93; Salzburger et al., 2002a; Verheyen et al., 2003). *A. burtoni* was shown to be a sister group to both the Lake Victoria region superflock and the Lake Malawi species flock, while its generalist life-style and body plan place it close to the proposed ancestral lineage (Meyer et al., '91) (Fig. 1).

Grant sponsors: Universität Konstanz and Deutache Forschungsgemeinschaft to Am; National Institute of Health and National Science Foundation to CTA; European Union (Marie Curie Individual Fellowship) and Landesstiftung Baden-Württemberg gGmbH to WS.

†Michael Lang's, current address is: Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, Spain.
*Correspondence to: Axel Meyer, Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany. E-mail: axel.meyer@uni-konstanz.de
*Correspondence to: Chris T. Amemiya, Benaroya Research Institute at Virginia Mason, Seattle, Washington 98101.
E-mail: camemiya@benaroyaresearch.org
Received 13 February 2005; Accepted 20 June 2005
Published online 27 October 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21068.


A
L. Victoria
L. Tanganyika
L. Malawi
B
LVR superflock [500]
R.
L. Malawi species flock [1,000]
Tr.
R.
L. Tanganyika species flock [250]
C
Haplochromis chilotes
Astatotilapia burtoni
Metriaclima zebra
Oreochromis niloticus

Besides being a popular model species in evolutionary biology, *A. burtoni* is also used in neuroethological studies (Wickler, '62; Francis et al., '93; Hofmann et al., '99). Territorialism and breeding behavior of males is strongly dependent on their current social status, which often changes within weeks. This is accompanied by alteration of coloration, growth and size of corresponding neural cells (Hofmann and Fernald, 2000).

Currently, several projects seek to gain new insights into the genomic and molecular biological bases for the adaptive evolution of haplochromine cichlids. EST projects (http://www.tigr.org/tdb/tgi/) (Watanabe et al., 2004) and microarray studies promise to detect differentiation in gene expression patterns (Renn et al., 2004). QTL mapping has already identified genomic regions important for jaw morphologies and coloration phenotypes of the Lake Malawi cichlid *Metriaclima zebra* (Albertson et al., 2003; Streelman et al., 2003). Recently, a genome sequencing consortium has been established in an effort to propose sequencing of the Nile tilapia *Oreochromis niloticus* genome (Kocher et al., 2004).

BAC libraries (Shizuya et al., '92; Miyake and Amemiya, 2004) are essential resources for the analyses of extended genomic regions or entire genomes of higher organisms. Once prepared from a species of interest, genomic regions of about 150 kb length are stably preserved and accessible by conventional library screening methods.

So far, BAC libraries of the Nile tilapia *O. niloticus* (Katagiri et al., 2001) and of the Lake Victoria cichlid *Haplochromis chilotes* (Watanabe et al., 2003) have been constructed. Moreover, a BAC library of the Malawi cichlid *M. zebra* is available at the Hubbard Center for Genome Studies (http://hcgs.unh.edu/BAC/Metriaclima.html). The latter two species provide genome resources from haplochromines from the Lake Victoria region superflock and the Lake Malawi species flock, respectively. Comparative studies that include these highly specialized haplochromines and the generalist *A. burtoni* promise to be informative with regard to the underlying genetic mechanisms that led to the enormous degree of diversification observed in cichlids within such a short time. It has been shown in previous work (see, e.g., Zardoya et al., '96; Salzburger et al., 2002b) that it is possible to obtain DNA sequence data of a large phylogenetic spectrum in cichlids. This could be achieved by application of standard PCR protocols with primers designed in non-coding nuclear DNA regions. Hence, DNA sequence data of many cichlid species should be easily obtained, once having detected the corresponding DNA region by BAC library screening of only one species (e.g., *A. burtoni*).

Here we describe the construction of a BAC library for *A. burtoni*. It is an arrayed BAC library that comprises approximately 14 × coverage of the genome and a median insert size of 150 kb.

MATERIALS AND METHODS

High molecular weight DNA extraction

Specimens of *A. burtoni* were reared under laboratory conditions in aquaria (12 hr light; 12 hr dark). Genomic DNA was obtained from the arterial blood of one male *A. burtoni* (approximately 8 cm head to tail length) from an inbred strain, following Amemiya et al. ('96). The specimens were sacrificed after anesthetization with MS222, cooled in ice-water and the blood was collected from the dorsal aorta posterior of the insertion of the dorsal fin. A 1 ml syringe (needle gauge 27) was used and the blood was collected 2 min after injection of 30 μl heparin, 100 mg/ml. The blood obtained was further mixed with one-fourth volume of heparin, 100 mg/ml, and the blood cell concentration was then estimated with a hemacytometer (Neubauer Improved) from 1:100 dilutions in 0.85 × PBS. Aliquots of cells were diluted in 0.85 × PBS, mixed with *Incert Agarose* (Cambrex Bio Science, Baltimore, MD) and poured into 80 μl plug molds (BioRad Laboratories, Hercules, CA), giving a final agarose concentration of 1%. Plugs were treated overnight at room

Fig. 1. BAC libraries of East African cichlid fishes. (**A**) Map of East Africa showing the three Great Lakes Victoria, Tanganyika and Malawi. (**B**) Mitochondrial phylogeny of the East African cichlids modified from current phylogenies (Klett and Meyer, 2002; Salzburger et al., 2002a; Verheyen et al., 2003; Salzburger and Meyer, 2004). The cichlid species flock from Lake Tanganyika is the oldest and genetically most diverse. Some riverine cichlid lineages (R.) and the Tanganyikan Tropheini (Tr.) are sister to the Lake Malawi cichlid species flock, another riverine clade (R.) including *Astatotilapia burtoni*, and the Lake Victoria region (LVR) superflock. The sizes of the clades represent the respective species numbers. Species number for the three large assemblages are depicted in square brackets. (**C**) BAC libraries are currently available for four cichlid species: *H. chilotes* from Lake Victoria (Watanabe et al., 2003), *Metriaclima zebra* from Lake Malawi (http://hcgs.unh.edu/BAC/Metriaclima.html), *Astatotilapia burtoni* (present study) and the Nile tilapia *Oreochromis niloticus* (Katagiri et al., 2001).

temperature in cell lysis solution (1% LDS, 10 mM Tris pH 8.0, 100 mM EDTA pH 8.0). The solution was substituted several times. Finally, plugs were stored in 20% NDS (0.2% *N*-laurylsarcosine, 2 mM Tris pH 9.0, 0.14 M EDTA pH 9.0). Before further treatment, pulsed-field-gel-electrophoresis (BioRad, XA Mapper) was used to discard low molecular weight DNA from the plugs.

BAC library construction

For size fragmentation, genomic DNA of each plug was partially digested for 2.5 hr in 500 μl reaction volumes containing the restriction enzyme *Eco*RI (New England Biolabs, Ipswich, MA) and *Eco*RI methylase (New England Biolabs). The optimal quantities of enzyme activity had to be estimated previously by titration experiments employing 1/6 plug per sample. Other components were 2.6 mM spermidine, 0.5 mg/ml BSA (New England Biolabs) and reaction buffer (0.08 mM *S*-adenosylmethionine, 2 mM $MgCl_2$, 100 mM NaCl, 50 mM Tris-Cl, 1 mM DTT). The reaction was stopped with 150 μl of a solution containing 300 μg proteinase K, 2.9% *N*-laurylsarcosine and 0.29 M EDTA, followed by an incubation at 37°C for 1 hr. Proteinase K activity was inhibited by addition of 12 μl PMSF, 100 mM. DNA fragments were separated on a 1% agarose gel using pulsed-field-gel-electrophoresis (BioRad CHEF XA Mapper), essentially as described by Osoegawa et al. ('98). Gel slices were prepared that contained size fragments of high molecular weight and the DNA fragments were electroeluted and dialyzed according to Strong et al. ('97) and Danke et al. (2004). A small fraction of the high molecular weight DNA was analyzed on a 1% agarose gel and the DNA concentration was estimated by comparison with defined amounts of the DNA standard size ladder λ-*Hind*III. The Copy Control™ Kit (Epicentre, Madison, WI) was used for ligation reactions. Approximately 50 ng of partially digested DNA and 25 ng of the BAC vector CopyControl™ pCC1BAC™ (*Eco*RI) were ligated in 50 μl reaction volumes that were carried out overnight at 15°C.

After ligation, samples were inactivated by a brief proteinase K treatment, desalted by drop dialysis against ddH_2O for 2 hr using 0.025 μm nitrocellulose filters (Millipore, Billerica, MA) and the sample volume was then reduced three-fold by drop dialysis against PEG8000 (30% in 1/2 × TE). For test transformations, approximately 15–20 ng DNA of desalted ligation mix was added to 20 μl electrocompetent cells DHB10T1 (Invitrogen, Carlslbad, CA). Electroporation was carried out with the Cell-Porator *Escherichia coli* Pulser (Gibco-BRL, Invitrogen) with subsequent incubation of cells in 500 μl SOC for 1 hr at 37°C at 250 rpm. For large-scale transformations, approximately 75–100 ng DNA was added to one vial (~100 μl) of electrocompetent cells. Electroporation was performed with the Electro Cell Manipulator 600 (BTX, Harvard Apparatus, Inc., Holliston, MA) and 2 mm cuvettes (MβP, San Diego, CA, USA) applying 2.5 kV pulses. After expressing in 5.0 ml SOC for 1 hr at 37°C, glycerol was added to the transformation mix to a final concentration of 10%. Small aliquots were removed (for titering the number of colonies in the transformed material) and the rest snap-frozen with liquid nitrogen and stored at −80°C.

Insert size estimation

Test amounts of transformation mix were grown on LB agar plates (12.5 μg/ml chloramphenicol, 0.4 mM IPTG and 40 μg/ml XGAL). Clones were isolated randomly and prepared manually using a modification of a standard plasmid miniprep protocol (Sambrook and Russell, 2001). A tenth of each prepared BAC clone DNA was digested with the restriction enzyme *Not*I (Invitrogen) and the samples were then analyzed on a pulsed-field gel. The size of each clone was estimated manually using low-range PFG markers (New England Biolabs) as references.

Genome size estimation

Arterial blood was obtained as described above. A volume of 50 μl blood was obtained from one male specimen and mixed with 450 μl Liquemin N25000 (F. Hoffmann-LaRoche Ltd., Basel, Switzerland) and stored at room temperature. Cells were washed and prepared as described in Danke et al. (2004); chicken red blood cell nuclei were used as internal standards. Genome sizes were estimated through flow cytometry (FACSCalibur, Becton-Dickinson, Franklin Lakes, NJ).

Library arrays

To array the library, the transformed cells were grown on LB agar plates (12.5 μg/ml chloramphenicol, 0.4 mM IPTG and 40 μg/ml XGAL) and single white colonies were transferred into 384-well microtiter plates (Genetix M7001) using a colony-picking robot (Norgren Systems, Littleton, NH). A total array system (Biorobotics, Cambridge, MA) was used for library replication and

to prepare nylon filter sets (22 × 22 cm), containing BAC colony DNAs.

Library screening

One nylon filter set was hybridized overnight in 0.6 M NaCl, 0.02 M EDTA, 0.2 M Tris pH 8.0, 0.5% SDS and 0.05% sodium pyrophosphate at 65°C with gene-specific DNA probes that ranged in size from 277 to 725 bp. The probes for five genes were amplified with gene-specific primers (Table 1) and labeled with α^{32}P-dCTP by random priming (Feinberg and Vogelstein, '83). The filters were washed in wash buffer (1 × SSC, 0.1% SDS, 0.05% sodium pyrophosphate) for 30 min at room temperature and then for 40 min at 37°C. DNAs from resultant positive clones were isolated and prepared according to the FastPlasmid Mini Kit (Eppendorf, Hamburg, Germany) or manually as explained above. The clones were further analyzed by PCR and direct sequencing on an ABI 3100 automated capillary DNA sequencer using the Big Dye Termination Reaction chemistry and gene-specific primers. In some cases, positive hybridizing clones were re-spotted on nylon filters and screened once more with gene-specific probes. Resultant positive clones were prepared manually and analyzed by PCR for validation. Restriction digestions of positive clones were prepared with the restriction endonucleases *Not*I and *Eco*RI, followed by pulsed-field gel electrophoresis or standard gel electrophoresis, respectively.

Phylogenetic reconstruction

A neighbor-joining tree of Sox10 protein sequences from several vertebrate taxa was reconstructed applying the gamma model in MEGA2 (Kumar et al., 2001). The obtained phylogeny was assessed by bootstrapping (5,000 replicas). The tree was rooted with human SOX9. GenBank accession numbers for the sox10 sequences are given in Fig. 3A and Table 1.

RESULTS AND DISCUSSION

We extracted genomic DNA from the arterial blood of one male *A. burtoni* from an inbred strain. A total of 600 μl of blood with heparin was obtained from the specimen. The blood cell concentration was quantified to be approximately 10^9 cells/ml and quantities corresponding to 1×10^7 cells were poured into 80 μl agarose plugs. The obtained amount of DNA in the plugs was approximately 20 μg. Titration experiments were carried out to optimize size fragmentation of genomic DNA with the restriction endonuclease *Eco*RI and *Eco*RI methylase. We used 1/6 plug and variable amounts of *Eco*RI restriction enzyme (1–6 units) and *Eco*RI methylase (10–60 units) for each sample in a 500 μl reaction volume. The goal was to obtain a tight, yet homogeneous distribution of partially digested DNA upon analytical pulsed-field gel electrophoresis. Based on this titration, five agarose plugs containing genomic DNA (approx. 150 μg) were partially digested for 2.5 hr with 4 units *Eco*RI and 50 units *Eco*RI methylase. DNA fragments were separated by preparative pulsed-field gel electrophoresis and several gel slices were prepared that contained partially digested DNA fragments, ranging from 50 to 250 kb. Test ligations and transformations with the various fractions revealed that a particular

TABLE 1. List of genes used for screening the BAC library, including gene-specific primers, the location of the primers on the gene, the length of the probe, GenBank accession numbers for gene sequences in Astatotilapia burtoni, and number of positive clones after PCR verification

Gene	Primer sequence	Location	Probe length	No. of positive BAC clones	Accession number
tfap2a	F: 5′-CAA AGA GTT CAC RGA CCT GCT G-3′	Exon 7	277 bp	6	AY970938
	R: 5′-CTT TCT GTG CTT CTC RTC TTT GTC-3′	Exon 7			
mc1r	F: 5′-SGC GGA YCG YTA YAT CAC CA-3′	Exon 1	486 bp	21	AY970939
	R: 5′-TCC TGR CTS CGG TAV GCG TA-3′	Exon 1			
ednrb1	F: 5′-AAR GAY TGG TGG CTK TTC AG-3′	Exon 4	725 bp	18	AY970945
	R: 5′-GAK GCC ATG TTG ATS CCA AT-3′	Exon 6			
aim1	F: 5′-CTG GGM GGA GCW TGT GGT TAC-3′	Exon 3	322 bp	5	AY970940
	R: 5′-TGG CTT CWC CSA GRG CWG AGA A-3′	Exon 3			
sox10b	F: 5′-CCW ACC ACY CCC AAG ACS GAA C-3′	Exon 3	659 bp	22 (sox10b)	AY970944
	R: 5′-CCC ATR TAR GAG AAG GCR GAG T-3′	Exon 3		15 (sox10a)	AY970942
	F: 5′-GAG GAG CAS AGC WTK TCG GAG-3′	Exon 1	PCR/seq. only		AY970941 (*sox10a*)
	R: 5′-ASA GCT TCC CCA GYG TYT TGC-3′	Exon 1			AY970943 (*sox10b*)

DNA fraction with an average size of 150 kb yielded ~2400 recombinant BAC clones per 1 µl of DNA solution. Given the total volume of 150 µl of the partial digest DNA solution, this fraction could have potentially yielded about 360,000 recombinant clones and thus was used for library construction.

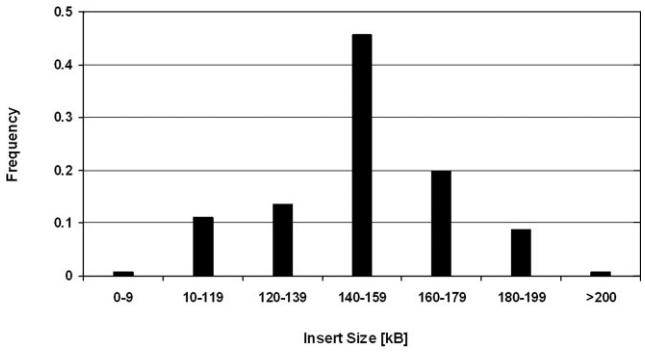


Fig. 2. Size distribution of BAC clone inserts. A total of 162 randomly picked clones were prepared and digested with the restriction enzyme *Not*I. Samples were run on a CHEF gel and insert sizes were calculated manually by comparison with standard size markers.

A total of 162 clones were isolated randomly from the transformed bacterial material in order to assess the distribution of clone insert lengths of the library. DNAs from respective clones were digested with the restriction endonuclease *Not*I and analyzed on a pulsed-field gel. The size of each clone was estimated manually using a standard pulsed-field DNA size marker (New England Biolabs, Low Range Marker) as reference.

The clone insert size had a mean of 145 kb and a median of 150 kb (Fig. 2). The insert size distribution obtained fits well with the DNA insert sizes that were used in the cloning procedure. The portion of clones that were smaller than 100 kb was estimated to be about 10%; one clone (out of 162) did not contain any insert.

The genome size of *A. burtoni* was estimated in order to obtain a reference value for assessing the completeness and the genomic coverage of the BAC library. Several studies with closely related species suggested the genome size to be 1.0–1.2 pg per haploid genome (Hinegardner and Rosen, '68, '72; see also Ohno and Atkin, '66; Majumdar and McAndrew, '86; Vinogradov, '98; Gregory, 2001). However, to obtain a confident estimate, we

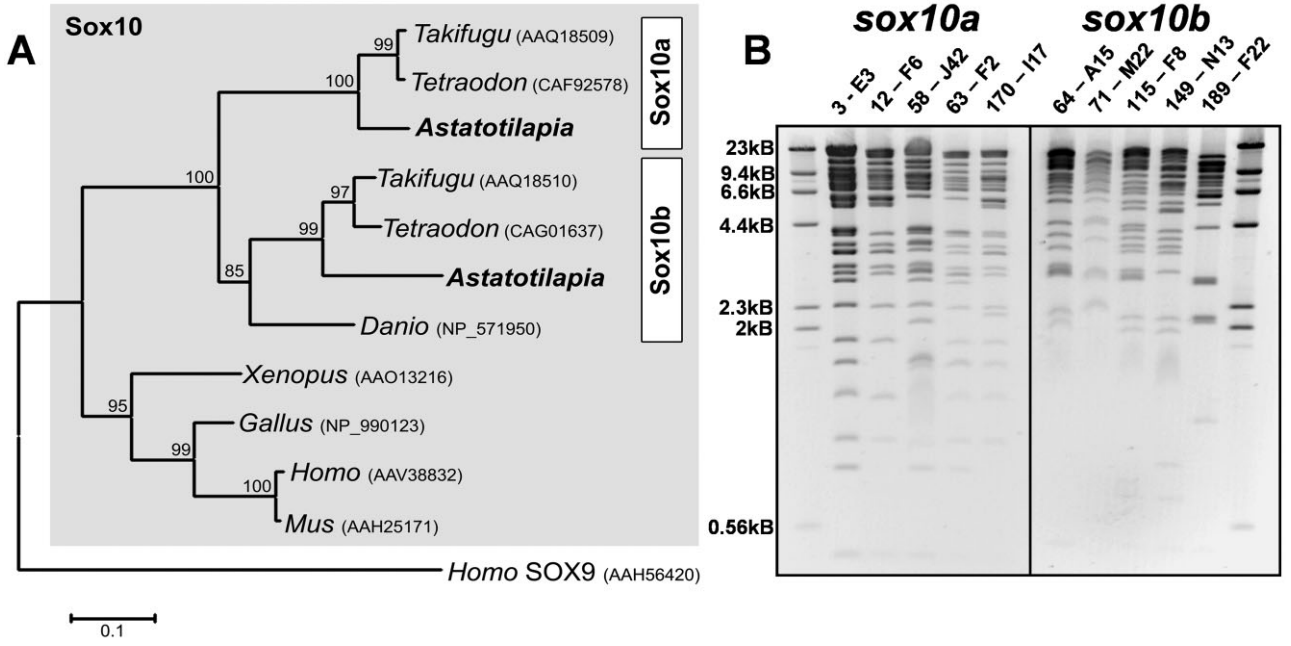


Fig. 3. Characterization of sox10a or sox10b containing BAC clones. (**A**) Phylogeny of vertebrate *Sox10*. A neighbor-joining tree of *Sox10* protein sequences from several vertebrate taxa. Numbers above nodes indicate bootstrap values (5,000 replicas). The tree was rooted with human *SOX9*. The cichlid *Sox10* paralogs (corresponding to the concatenated sequences of exons 1 and 3) are depicted in bold. *Sox10* terminology is adapted from Koopman et al. (2004). The zebrafish *Sox10* paralog with a function in pigment cell development (Dutton et al., 2001) belongs to the *Sox10b* branch. (**B**) Agarose gel (1%) with *Eco*RI restriction digestions of ten BAC clones containing *sox10a* or *sox10b*, and the DNA size standard Lambda-*Hind*III. The restriction patterns of *sox10a* and *sox10b* clones are clearly distinct, confirming that the *sox10* paralogs are located in different genomic regions.

measured the genome size of *A. burtoni* red blood cells of one male specimen, using flow cytometry. In agreement with the previous studies, a value of 0.97 pg per haploid genome was estimated from measurement of several thousand cells in two separate flow cytometer experiments (data not shown). This value is only slightly below the range of genome sizes of other cichlid fish species (see above). Actually, a wide variance is also found by comparing different genome size estimates for a single species, the Nile tilapia *O. niloticus*, ranging from 0.95 to 1.2 pg per haploid genome (Gregory, 2001). Thus, we cannot see any discrepancy of the genome size estimate of *A. burtoni* to the estimates of closely related species.

A total of 96,768 clones were isolated into 384-well microtiter dishes, comprising an estimated 14 × coverage of the *A. burtoni* genome. Two replicas of the library were prepared, and an approximate 13 × fraction of the library was spotted onto 22 × 22 cm nylon hybridization membranes.

The depth of the library was independently corroborated by screening the BAC library filter set with gene-specific probes for five genes. The complete nylon filter set was screened with two or three different probes per screen that were specific to the genes *aim1*, *tfap2*, *mc1r*, *ednrb1* and *sox10b*. Further verification of hybridizing clones was obtained by PCR amplification with gene-specific DNA sequences and direct sequencing of the obtained PCR products. The screens for *aim1* and *tfap2* were additionally verified by re-spotting positive clones on nylon membranes for a second hybridization with the corresponding probes. Strong hybridizing clones were prepared and re-analyzed by PCR. In this way, a more thorough detection of true positives was obtained.

The genes that were analyzed in our study play important roles at different developmental levels of teleost color patterning: *tfap2a* is required for early steps in the development of the neural crest, from which pigment cells are derived (Knight et al., 2003). *sox10b* further specifies the non-ectomesenchymal neural crest derivatives including pigment cells (Dutton et al., 2001; Koopman et al., 2004). *ednrb1* is not only expressed by different pigment cell types but also has a patterning function since loss-of-function mutations lead, for example, to disrupted adult stripes in zebrafish (Parichy et al., 2000). *mc1r* and *aim1* are important for the differentiation of the dark pigment cells, the melanin-producing melanophores. While *mc1r* has a role in melanin regulation (Barsh, 2003), *aim1* is thought to be involved in melanin synthesis (Fukamachi et al., 2001). Since the diversification of color patterns has played an important role in the adaptive radiations of East African lake cichlids (Meyer, '93; Danley and Kocher, 2001; Allender et al., 2003; Terai et al., 2003), these genes represent candidate ''speciation genes''.

The screens for *mc1r* and *ednrb1* revealed 21 and 18 positive clones, respectively. In comparison with the estimated filter coverage of 13 times the *A. burtoni* genome, these results suggest an over-representation of the corresponding genomic regions since PCR amplification and direct sequencing of the positive clones revealed identical DNA sequences (i.e., belonged to the same gene).

Recently, it was shown that *sox10* is duplicated in the pufferfish *Takifugu rubripes* and the two copies were named *sox10a* and *sox10b* (Koopman et al., 2004). This finding could be confirmed for *A. burtoni*. DNA sequence comparisons discriminated between 15 *sox10a*- and 22 *sox10b*-specific clones out of the total of 37 positives. The probe for the screen was designed from a region of exon 3 of *sox10b* and the deduced amino acid sequences of both obtained gene fragments revealed uninterrupted open reading frames.

Phylogenetic reconstruction of the *sox10* gene fragments indicates that they are *sox10a* and *sox10b*, as defined by Koopman et al. (2004) (Fig. 3A). In addition to the obtained 659 bp *sox10* exon 3 DNA fragments, a 424 bp section of s*ox10* exon 1 coding sequence was amplified and sequenced directly from each clone with gene-specific primers (Table 1). The concatenated DNA sequences, including published data, were used for phylogeny reconstruction (Fig. 3A) and the obtained gene tree shows that the duplication of the *sox10* gene is older than the split of the lineages of Ostariophysi (represented by zebrafish) and Neoteleostei (represented by fugu and cichlid). In contrast to tetrapods, ray-finned (actinopterygian) fishes have undergone a whole genome duplication early in teleost evolution (see, e.g., Taylor et al., 2003). It is, therefore, most likely that *sox10* was duplicated during this genome duplication event. The *Eco*RI fingerprint patterns of five positive clones each for *sox10a* and *sox10b* are shown in Fig. 3B. The restriction fragment patterns are clearly distinct between the two duplicates, indicating that the genes are located in different genomic regions.

In our BAC screens, we recovered two paralogous copies of *sox10* gene fragments while the

DNA probe was designed on the paralog *sox10b*. This is noteworthy, since the 659 bp DNA sequences that correspond to the probe differ by almost 30% between the two paralogous gene fragments. Given the hybridization conditions of our screens (see Materials and Methods), this shows that the screening of BAC clones proves to be a powerful tool to identify paralogous genes. In this case, the corresponding duplication event should have occurred more than 300 million years ago (Taylor et al., 2003).

The expression of *sox10a* and *sox10b* in *A. burtoni* remains to be analyzed with respect to tissue specificity and developmental stage. This could give important information about the probable function of the two genes, especially of *sox10a*, for which no such information is available.

The *Eco*RI fingerprint patterns of digested BAC clones (Fig. 3B) demonstrated that the clones do indeed result from partially digested genomic DNA and as a consequence are expected to contain essentially random areas of the genome and, hence, are overlapping. Therefore, this BAC library can be employed to obtain overlapping BAC clones for regional analysis of a given genomic region.

The screens for *aim1* and *tfap2* confirmed five and six positive clones, respectively. These comparatively low numbers of detected clones indicate that there is some variation in the coverage of the library. Noteworthy, the probes that were employed for screening for *aim1* and *tfap2* were, by far, the shortest ones with lengths of 322 bp (*aim*1) and 277 bp (*tfap*2a) (Table 1). Possibly, this might have influenced the efficiency of the screens since multiple probes were pooled in order to screen the library.

In summary, we were able to obtain a genomic BAC library of *A. burtoni* with a median BAC clone insert size of 150 kb and an approximate coverage of 14 genome equivalents. Screens of the library using nylon filter sets and radiolabeled gene-specific probes corroborate the coverage estimates, although the number of confirmed positives varies between genes. As a rough estimate, an average of ~14.5 positive clones per screen was obtained. In an attempt to screen for the gene *sox10b*, we additionally detected its fish-specific paralog *sox10a*. This shows that BAC libraries serve well to detect paralogous genes.

This BAC library is intended as a genomic resource for the ''cichlid community'' to facilitate the study of cichlid fishes for evolutionary and genomic purposes. The library is available to the scientific community via the laboratory of Axel Meyer, University of Konstanz, Konstanz, Germany. This library will aid in gaining new insights into the great diversification process in the particular species radiation of haplochromine cichlids.

ACKNOWLEDGMENTS

We thank Dr. Jing Luo for assisting in the Meyer laboratory, Joshua Danke in particular and the other members of the Amemiya laboratory for helping and assisting M.L. during his time in Seattle. M.L. is grateful to Elvira Juan for providing material and reagents that were required for work during revision of this article. The European Union (Marie Curie Individual Fellowship) and Landesstiftung Baden-Württemberg gGmbH grant was given to W.S., a grant from the National Institutes of Health and National Science Foundation to C.T.A., and grants from the Universität Konstanz and Deutsche Forschungsgemeinschaft to A.M.

LITERATURE CITED

Albertson RC, Streelman JT, Kocher TD. 2003. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc Natl Acad Sci USA 100:5252–5257.

Allender CJ, Seehausen O, Knight ME, Turner GF, Maclean N. 2003. Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations in nuptial coloration. Proc Natl Acad Sci 100: 14074–14079.

Amemiya CT, Ota T, Litman GW. 1996. Construction of P1 artificial chromosome (PAC) libraries from lower vertebrates. In: Birren B, Lai E. editors. Non-mammalian genomic analysis: a practical guide. Academic Press, New York. p 223–255.

Barsh GS. 2003. What controls variation in human skin color? PLOS Biol 1:19–22.

Danke J, Miyake T, Powers T, Schein J, Shin H, Bosdet I, Erdmann M, Caldwell R, Amemiya CT. 2004. Genome Resource for the Indonesian Coelacanth *Latimeria menadoensis*. J Exp Zool 301A:228–234.

Danley PD, Kocher TD. 2001. Speciation in rapidly diverging systems: lessons from Lake Malawi. Mol Ecol 10: 1075–1086.

Dutton KA, Pauliny A, Lopes SS, Elworthy S, Carney TJ, Rauch J, Geisler R, Haffter P, Kelsh RN. 2001. Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. Development 128:4113–4125.

Feinberg AP, Vogelstein B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13.

Francis RC, Soma K, Fernald RD. 1993. Social regulation of the brain-pituitary–gonadal axis. Proc Natl Acad Sci USA 90:7794–7798.

Fryer G, Iles TD. 1972. The cichlid fishes of the Great Lakes of Africa: their biology and evolution. Edinburgh: Oliver & Boyd. 641 p.

Fukamachi S, Shimada A, Shima A. 2001. Mutations in the gene encoding B, a novel transporter protein, reduce melanin content in medaka. Nat Genet 28:302.

Gregory TR. 2001. Animal Genomesize Database, URL: http://www.genomesize.com.

Hinegardner R, Rosen DE. 1968. Evolution of cellular DNA content in teleost fishes. Am Natural 102:517–523.

Hinegardner R, Rosen DE. 1972. Cellular DNA content and the evolution of teleostean fishes. Am Natural 106:621–644.

Hofmann H, Fernald RD. 2000. Social status controls somatostatin neuron size and growth. J Neurosci 20:4740–4744.

Hofmann H, Benson ME, Fernald RD. 1999. Social status regulates growth rate: consequences for life-history strategies. Proc Natl Acad Sci USA 96:14171–14176.

Katagiri T, Asakawa S, Minagawa S, Shimizu N, Hirono I, Aoki T. 2001. Construction and characterisation of BAC libraries for three fish species: rainbow trout, carp and tilapia. Anim Genet 32:200–204.

Klett V, Meyer A. 2002. What, if anything, is a tilapia?—Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care systems in the African cichlid fishes. Mol Biol Evol 19:865–883.

Knight RD, Nair S, Nelson SS, Afshar A, Javidan Y, Geisler R, Rauch G-J, Schilling TF. 2003. Lockjaw encodes a zebrafish tfap2a required for early neural crest development. Development 130:5755–5768.

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

Kocher T, Conroy J, McKaye K, Stauffer J. 1993. Similar morphologies of cichlid fish in lakes Tanganyika and Malawi are due to convergence. Mol Phylogenet Evol 2:158–165.

Kocher T, Fernald R, Hofmann H, Meyer A, Okada N, Penman D. 2004. Genome sequence of a cichlid fish: the Tilapia (*Oreochromis niloticus*). Whitepaper.

Koopman P, Schepers G, Brenner S, Venkatesh B. 2004. Origin and diversity of the Sox transcription factor gene family: genome-wide analysis in Fugu rubripes*1. Gene 328:177–186.

Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245.

Majumdar KC, McAndrew BJ. 1986. Reative DNA content of somatic nuclei and chromosomal studies in three genera, tilapia, sarotherodon, and oreochromis of the tribe tilapiini (Pisces, Cichlidae). Genetica 68:175–188.

Meyer A. 1993. Phylogenetic relationships and evolutionary processes in East African cichlids. Trends Ecol Evol 8: 279–284.

Meyer A, Kocher TD, Basasibwaki P, Wilson AC. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550–553.

Meyer A, Kocher T, Wilson A. 1991. Scientific correspondence, African fishes. Nature 350:467–468.

Miyake T, Amemiya CT. 2004. BAC libraries and comparative genomics of aquatic chordate species. Comp Biochem Physiol C 138:233–244.

Ohno S, Atkin NB. 1966. Comparative DNA values and chromosome complements of eight species of fishes. Chromosoma 18:455–466.

Osoegawa K, Woon PY, Zhao B, Frengen E, Minako T, Catanese JJ, deJong PJ. 1998. An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52:1–8.

Parichy DM, Mellgren EM, Rawls JF, Lopes SS, Kelsh RN, Johnson SL. 2000. Mutational analysis of endothelin receptor b1 (rose) during neural crest and pigment pattern development in the zebrafish *Danio rerio*. Dev Biol 227: 294–306.

Renn SC, Aubin-Horth N, Hofmann H. 2004. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. BMC Genom 5:42.

Salzburger W, Meyer A. 2004. The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften 91: 277–290.

Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. 2002a. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. Syst Biol 51: 113–135.

Salzburger W, Baric S, Sturmbauer C. 2002b. Speciation via introgressive hybridization in East African cichlids? Mol Ecol 11:619–625.

Salzburger W, Mack T, Verheyen E, Meyer A. 2005. Out of Tanganyika: the repeated adaptive radiations of East African cichlid fishes. BMC Evol 21:17.

Sambrook J, Russell DW. 2001. Molecular cloning, a laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiri Y, Simon M. 1992. Cloning and stable maintenance of 300-kilobase fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89:8794–8797.

Stiassny MLJ, Meyer A. 1999. Cichlids of the Rift Lakes. Sci Am June:64–69.

Streelman JT, Albertson RC, Kocher TD. 2003. Genome mapping of the orange blotch colour pattern in cichlid fishes. Mol Ecol 12:2465–2471.

Strong SC, Ohta Y, Litman GW, Amemiya CT. 1997. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field-gel-separated partial digests of genomic DNA. Nucleic Acids Res 25:3959–3961.

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res 13:382–390.

Terai Y, Morikawa N, Kawakami K, Okada N. 2003. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. Proc Natl Acad Sci USA 100:12798–12803.

Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL. 2001. How many species of cichlid fishes are there in African lakes? Mol Ecol 10:793–806.

Verheyen E, Salzburger W, Snoeks J, Meyer A. 2003. Origin of the superflock of cichlid fishes from lake victoria, East Africa. Science 300:325–329.

Vinogradov AE. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry. Cytometry 31:100–109.

Watanabe M, Kobayashi N, Fujiyama A, Okada N. 2003. Construction of a BAC library for Haplochromis chilotes, a cichlid fish from Lake Victoria. Genes Genet Syst 78: 103–105.

Watanabe M, Kobayashi N, Shin-i T, Horiike T, Tateno Y, Kohara Y, Okada N. 2004. Extensive analysis of ORF

sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock: identity of EST sequences between Haplochromis chilotes and Haplochromis sp. ''Red-tailsheller''. Gene 343:263–269.

Wickler W. 1962. Zur Stammesgeschichte funktionell korrelierter Organ- und Verhaltensmerkmale: Ei-Attrappen und Maulbrüten bei afrikanischen Cichliden. Z Tierpsychol 19:129–164.

Zardoya R, Vollmer DM, Craddock C, Streelman JT, Karl S, Meyer A. 1996. Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). Proc R Soc Lond B 263:1589–1598.