

For Zebrafish, revised 9/27/06

Construction of bacterial artificial chromosome libraries for two model fish species: the Lake Malawi cichlid (*Metriaclima zebra*), and the blind cavefish (*Astyanax mexicanus*).

Federica DiPalma¹, Celeste Kidd¹, Richard Borowsky² and Thomas D. Kocher¹

¹Hubbard Center for Genome Studies, University of New Hampshire

²Department of Biology, New York University

Corresponding Author:

Thomas D. Kocher, Hubbard Center for Genome Studies, University of New Hampshire,
35 Colovos Road, Gregg Hall, 4th Floor, Durham, NH 03824.

Tel: (603) 862-2115, Fax: (603) 862-2940

Thomas.d.kocher@unh.edu

Abstract

Teleost fishes have become important models for studying the evolution of the genetic mechanisms of development. A key resource for comparative genomics and positional cloning are large-insert libraries constructed in bacterial artificial chromosomes. We have constructed BAC libraries for two species of teleost fish that are important models for the study of developmental evolution. *Metriaclima zebra* is one of several hundred closely-related, morphologically diverse haplochromine cichlids which have evolved over the last one million years in Lake Malawi, East Africa. The Mexican tetra, *Astyanax mexicanus* is well known for adaptations related to the recent evolution of blind cave-dwelling forms. Clones and high-density filters for each library are available to the scientific community through the Hubbard Center for Genome Studies.

Introduction

Because of their extraordinary morphological diversity, teleost fishes have become widely used model organisms for the study of developmental evolution. Studies of danios closely related to the laboratory zebrafish have provided important insights into the evolution of pigment pattern diversity (Quigley *et al.*, 2005). Studies of the three-spine stickleback have uncovered the genetic basis for parallel evolution of body armor and spines (Colosimo *et al.*, 2005). Studies of gene expression have begun to identify the genetic mechanism of pelvic fin loss in pufferfish (Tanaka *et al.*, 2005). The ease with which genomic resources can be developed now allows the selection of model species to be based primarily on the particular phenotypes of interest to the investigator, rather than on the availability of particular genomic resources.

Haplochromine cichlids have undergone a spectacular radiation in the lakes of East Africa. More than 2,000 species of cichlids have evolved in these lakes over the past 5 million years. These species can be viewed as a collection of natural mutants, screened by natural selection of a diversity of adult phenotypes. The radiation of cichlid fishes in East Africa is an ideal model system for studying the genetic basis of speciation and the mechanisms of evolutionary change in morphology, pigmentation and behavior (Kocher 2004).

The Mexican tetra is best known for its eyeless forms (the blind cavefish) which have evolved in several cave systems in east-central Mexico (Mitchell 1977; Dowling *et al.* 2002). This species is used as a model for studying the evolution of developmental

pathways associated with adaptation to subterranean habitats (Borowsky & Wilkens 2002) including changes in pigmentation (McCauley *et al.*, 2004) and eye development (Jeffery 2005).

BAC libraries are a key resource needed to support physical mapping, identification, and functional characterization of the genes underlying the evolutionary transformation of developmental pathways (Miyake & Amemiya 2004). Here we describe the construction of bacterial artificial chromosome libraries for a haplochromine cichlid (*Metriaclima zebra*) and the Mexican tetra (*Astyanax mexicanus*).

Results

Library construction

The BAC libraries were constructed using the *Hind*III cloning site in commercially prepared pCC1 BAC vector from Epicentre Laboratories. Briefly, high molecular weight (HMW) DNA embedded in agarose was partially digested using *Hind*III, and size selected by pulsed-field gel electrophoresis (PFGE) as described by (Asakawa *et al.*, 1997). DNA fractions of different sizes were excised from the gel, electroeluted as described by (Strong *et al.*, 1997), and ligated to the *Hind*III-digested and dephosphorylated pCC1 BAC vector (Epicentre). Size fractions and ligation products consistently yielding BAC clones with inserts between 80 and 150 kb were chosen for library construction. Both BAC libraries are the products of several size fractionations, of

different ligations and many transformations, each resulting in varying degrees of cloning efficiency and quality of the clones generated.

Characterization of the BAC libraries

The *Metriaclima zebra* BAC library consists of a total of 56,832 clones providing 5.72 haploid genome equivalents assuming a genome size of 1.06 Gb (Table 1). The library is composed of two segments. The first 19,968 clones were generated using HMW DNA isolated from pooled brain tissues. The remaining 36,864 clones were generated using HMW DNA isolated from pooled brains and red blood cells. The size distribution of 90 clones randomly picked from the two library segments is shown in Figure 1a. The insert size ranges between 45 and 235 Kb. *NotI* digestion of the 90 randomly picked BAC clones followed by PFG electrophoresis indicates an average insert size of 110 Kb (Fig. 2a), and that 3% of the library comprises clones with no inserts or with vector artifacts (e.g. deleted vector bands).

The *Astyanax mexicanus* BAC library consists of a single library segment of 58,752 BAC clones providing 5 haploid genomic equivalents assuming a genome size of 1.2Gb (Table 1). The library was generated using HMW DNA isolated from pooled muscle tissues. The insert size distribution is shown in Figure 1b with inserts ranging from 45 to 195 Kb. Analysis of 90 randomly selected clones indicates an average insert size of 105 kb (Fig. 2b) and that clones without inserts comprise 2% of the library.

Gene content

To confirm the estimated coverage of each we screened the *Metriaclima* library filters with several single-copy probes. One filter of the library, corresponding to 1.8X genome equivalents was screened with each probe. Probes for the green (RH2B), red (LWS) and UV (SWS1) opsin genes each resulted in a positive hit. Each of these positives hybridization results were confirmed by PCR and sequencing to demonstrate the expected gene content. The single strong positive for each probe is only slightly lower than the expected value of 1.8 genome equivalents on the filter. Similar success has been obtained from the *Astyanax* library. A probe for sonic hedgehog (*shh*) recovered 3 clones of *shh* and 1 clone of the related protein, *twhh*, from filters with an expected coverage of 3.6x genome equivalents (M. Protas, pers comm.).

Discussion

Critical factors involved in the construction of our BAC libraries

Preparation of the BAC vector for cloning is a labor-intensive process. Each step must be optimized by titration of restriction enzyme and alkaline phosphatase treatments in order to achieve low non-recombinant background levels and maximum efficiency. To render our BAC ligation process more efficient, we purchased the pre-restricted pCC1 BAC vector from Epicentre.

There are three cloning sites (*HindIII*, *BamHI* and *EcoRI*) within the *Lac Z* gene of the pCC1 BAC vector. Initially, we chose to construct our libraries in the *EcoRI* site to take advantage of a method for controlled partial digestion using a competition between *EcoRI* endonuclease and *EcoRI* methylase (Larin *et al.* 1991; Bonnema *et al.* 1996). However, we were unable to successfully clone *EcoRI* partially digested HMW DNA into the *EcoRI*-digested pCC1 BAC vector. Analysis of putative transformants revealed a high percentage (50-60%) of deleted vector bands with no inserts (data not shown). We attribute this to a high level of *EcoRI* enzyme star activity in the vector preparation we purchased from Epicentre. During BAC vector preparation, if the restriction enzyme is not appropriately titrated, over digestion results in cutting at a secondary star site. If dephosphorylation is then incomplete, the vector can religate between the restriction site and the star site resulting in small deletions in the vector, which affect the expression of the selection gene (Osoegawa *et al.*, 1999).

For this reason we decided to construct our libraries using the *HindIII* cloning site in the pCC1 BAC vector. Optimal partial digestion conditions using the *HindIII* endonuclease were determined by limiting the enzyme concentration and by keeping the reaction time constant in order to generate the maximum concentration of DNA within the desired size range (100 to 250 Kb). The optimal amount of *HindIII* for each HMW DNA preparation was determined empirically. The unit concentration that yielded the maximum amount of DNA in the desired range was used for library construction. This procedure is known to be highly dependent on both DNA concentration and the source of tissue, and is very difficult to reproduce consistently. Diffusion of molecular reagents into the agarose is a critical limiting factor; enzymes and buffers have easier access to the DNA near the surface of the plug than in the middle of the plug, and as a result enzyme concentration and incubation time are difficult to control. In order to improve access of the restriction enzymes to the HMW DNA, we cut each agarose plug into three smaller pieces prior to restriction digestion.

To further improve the consistency of our partial digestions, we used only HMW DNA isolated from tissues which yielded the highest quality DNA. We embedded intact cells in agarose plugs and lysed them by placing the samples in detergent containing a high concentration of EDTA to inhibit nuclease activity. We included proteinase K at a concentration of 2mg/ml in these solutions to ensure all cellular proteins were digested. We then removed the cellular components released by this digestion via diffusion during repeated washings of the agarose plugs. Prior to partial digestion, we tested each HMW DNA preparation for nuclease activity by analysing uncut and completely digested DNA

by PFGE. The best results were obtained with red blood cells and cells from tissues such as brain or gonads. Lower quality HMW DNA was isolated using cells from muscle and liver tissues.

Applications

These libraries are being used successfully in a number of positional cloning projects at HCGS and elsewhere. The arrayed libraries, as well as high-density replica filters, are available for distribution through the Hubbard Center for Genome Studies (<http://hcgs.unh.edu/BAC/>).

The *Metriaclima* BAC library complements existing genomic resources for cichlid fishes, including a genetic linkage map for Lake Malawi haplochromines (Albertson *et al.*, 2003), and genetic and physical maps of the tilapia genome (Lee *et al.*, 2005; Katagiri *et al.*, 2005). BAC libraries are now available for several species of East African cichlids, including tilapia (Katagiri *et al.*, 2001) and haplochromine cichlids from lakes Victoria (Watanabe *et al.*, 2003) and Tanganyika (Lang *et al.*, 2005). The library for *Metriaclima* is being used in the positional cloning of genes controlling differences in jaw and tooth morphology, pigmentation and sex determination in Lake Malawi cichlids (Albertson and Kocher 2006; Streelman *et al.* 2003).

Astyanax has become an important model for studying the genetic basis of adaptive evolution. Among the adaptations to subterranean life are reductions of eyes (Yamamoto and Jeffery 2000), pigmentation (Protas *et al.*, 2006), and teeth (Stock *et al.*, 2006) and

expansion of alternative sensory modalities (Franz-Ondendaal and Hall, 2006). The *Astyanax* library will be an important resource for positional cloning the genes underlying these traits.

Material and Methods

Sources of DNA

The *Metriaclima zebra* used were lab-reared F₂₋₄ animals bred from wild-caught parents collected from Mazinzi Reef, southeast arm of Lake Malawi, in 1996.

The *Astyanax mexicanus* specimens were lab-reared F₁ full-sibs of wild-caught parents collected in January 2002 from Arroyo Sarco (Rio Sabinas drainage), near the village of Encino, Tamaulipas, Mexico (Darnell 1962). They displayed the typical phenotypes of the surface forms of this species.

BAC Vector

For construction of our BAC libraries we used the CopyControl pCC1BAC *Hind*III Cloning-Ready Vector from Epicentre Technologies.

Isolation of High Molecular Weight (HMW) DNA

HMW DNA was isolated from brain and muscle by disrupting the tissues with a mortar and pestle. Cells were then washed twice in PBS, re-suspended in PBS at a concentration of 4×10^7 cells/ml, and equilibrated to 50°C. When whole blood was used as the source tissue, cells were washed in PBS three to four times, re-suspended in PBS at a concentration of 4×10^7 cells/ml, and equilibrated to 50°C. An equal volume of 1% low melting point agarose (InCert, BioWhittaker Molecular Applications, BMA) in PBS was added to the cell suspension. The mixture was then transferred into ice-cold disposable plug moulds (Bio-Rad). Plugs were incubated for 72 hours at 50°C in 0.1 M EDTA (pH

8.0), 0.01M Tris-HCl (pH 7.6), 0.02 M NaCl, 1% (w/v) Sarcosyl (Sigma) and 2mg/ml proteinase K (Sigma) with gentle shaking. Plugs were then washed three times in 0.02 M Tris-HCl (pH8.0) and 0.05 M EDTA for 1 hour at RT. PMSF (1mM final concentration; Sigma) was added to the last two washes to inactivate proteinase K.

Partial digestion of HMW DNA and size selection

Before digestion, plugs were washed extensively in TE for 24 hours, then cut into three smaller blocks of equal size. Each block was placed in a microfuge tube containing 500ul of 1x *Hind*III restriction buffer (NEB) at 4^oC for 1 hour. Fresh 1x restriction buffer was then added with 15 units of *Hind*III (NEB) and plugs were incubated on ice for another 2 hours to allow diffusion of the enzyme. The partial digestion reaction was performed by incubating the plugs at 37^oC for 35 minutes and stopped by EDTA to a final concentration of 20 mM.

Size selection was carried out as described in (Asakawa *et al.*, 1997), with minor modifications. In brief, partially digested DNA was separated by PFGE on a CHEF-DRII apparatus (Bio-Rad) in 0.5x TBE at 14^oC, 5 V/cm for 6 hours, with a 5s-20s pulse time. At the end of this first electrophoresis step, the gel portion containing DNA of 50 Kb or less in size was removed and discarded. Also the portion of the gel containing the original plugs was removed and discarded. Fresh 1% agarose was added to the remaining gel and a second electrophoresis step was performed using the same conditions for 19 hours. Gel slices containing size fractionated DNA were obtained by cutting horizontally at 0.5cm intervals in the range of 100-250 Kb. Each excised gel slice was subsequently inverted and buried in 1% low-melting-point agarose gel. A third electrophoresis step using the same conditions

for 19 hours was carried out to concentrate the widely spread DNA fragments in each gel slice into a sharp single band. The band of size selected genomic DNA was then excised and dialyzed in TAE at 4°C overnight.

Ligation and Electroporation

Size fractionated DNA was recovered from each gel band by electroelution in dialysis bags (Spectra/Por 7, Spectrum Laboratories Inc.) as described by (Strong *et al.*, 1997).

Partially digested HMW DNA was then ligated to 25 ng of dephosphorylated, *HindIII* digested pCC1BAC (Epicentre Technologies) at a 1:10 molar ratio of insert to vector with 400 units of T4 ligase (NEB cohesive end ligation units) in 50 µl reaction at 16°C overnight. The ligation mixture was dialyzed by spotting onto a 0.025µm microdialysis filter (Millipore) floating in 0.2X TE buffer for 1 hour at room temperature. The dialyzed ligation mixture was then concentrated by spotting onto a microdialysis filter floating on a solution containing 30% (w/v) PEG in TE buffer (Osegawa *et al.*, 1999).

1-2 µl of dialyzed concentrated ligation was used to transform 20 µl of ElectroMAX DH10B competent cells (Invitrogen). Electroporation was carried out using a Bio-Rad Gene Pulser at 200Ω, 2.5KV and 25µF. Cells were incubated in 1 ml SOC medium for 45 minutes at 37°C while shaking. Cells were then spread on LB plates (Bio-assay plates, Genetix) containing 12.5 µg/ml Chloramphenicol, 40 µg/ml X-gal and 100 µg/ml IPTG and incubated at 37°C for 24 hours to allow the blue color to develop sufficiently.

Storage of the BAC library

White recombinant colonies were picked to 384-well plates (Genetix) containing 70µl of freezing buffer (Sambrook *et al.*, 1989) using a Q-bot workstation (Genetix). Plates were incubated overnight at 37°C, replicated and stored at -80°C.

Analysis of transformant colonies

90 BAC clones were randomly picked from the transformant plates and grown overnight in 1.5 ml LB containing 12.5 µg/ml chloramphenicol. DNA was isolated by simple miniprep following the alkaline lysis method in (Sambrook *et al.*, 1989). The insert was released from the vector by restriction digestion with *NotI* (New England Biolabs). Samples were then run on a CHEF DRII apparatus (Bio-Rad) at 6V/cm, 0.5x TBE, 14°C for 10 hours with a pulse time of 5s-15s. The presence of vector and insert was determined by UV visualization after ethidium bromide staining.

High-Density replica Filters

BAC clones from each library were spotted onto positively charged Nylon membranes (Performa, Genetix) in high density arrays using a Q-bot station (Genetix). On each 22x22cm membrane, 36,864 colonies were spotted in duplicate in a 4x4 array pattern representing 18,432 individual clones. After inoculation, membranes were placed onto bioassay dishes (Genetix) containing LB agar with 12.5 µg/ml chloramphenicol and incubated at 37°C for 12-16 hours. Each filter was processed by placing the filter, colony side up, on 3 mm Whatman paper saturated with the following solutions and for the specified time: 1) 10% (w/v) SDS, 5 minutes; 2) 0.5 N NaOH, 1.5 M NaCl, 5-10 minutes;

3) 0.5 M Tris-HCl (pH 7.5), 1.5 M NaCl, 1mM EDTA (pH 8.0), 10 minutes; 4) 2x SSC, 0.1% (w/v) SDS, 5 minutes; 5) 2x SSC, 5 minutes; 6) 0.4N NaOH, 20 minutes to fix DNA onto the filter. The filters were washed extensively with two times 1 litre washes of 5x SSC, 0.1% (w/v) SDS for 30 minutes each wash, and the cell debris removed by gently wiping the surface of the filters. Excessive background due to residual cellular debris in the filters was prevented by treating with 20 µg/ml of proteinase K (Sigma) in 0.01 M Tris-HCl pH 8.0, 0.005 M EDTA pH 8.0, 0.5% (w/v) SDS, for 1 hour at 37°C. Finally the filters were washed twice in one litre of 2x SSC for 5 minutes, air-dried and stored at 4°C.

Screening of BAC library filters

PCR products derived from green, red and UV genes were isolated from retinal cDNA and were used as probes. The specific genes were selected because they had been previously characterized (Parry *et al.*, 2005). Labelling of probes, hybridization and detection were carried out using the nonradioactive ECL Nucleic Acid Labelling and Detection Kit (Amersham Biosciences) according to the manufacturer's instructions.

Acknowledgements

This work was supported by a grant from the National Science Foundation program in Developmental Mechanisms (IOB #0208343).

References.

- Albertson RC, Streelman JT and Kocher TD. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc Natl Acad Sci U S A* 2003; 100(9): 5252-7.
- Albertson RC, Kocher TD. Genetic and developmental basis of adaptive variation in the cichlid feeding apparatus. *Heredity* 2006; 97: 211-221
- Asakawa S, Abe I, Kudoh Y, Kishi N, Wang YM, Kubota R, et al. Human BAC library: Construction and rapid screening. *Gene* 1997; 191(1): 69-79.
- Bonnema G, Hontelez J, Verkerk R, Zhang YQ, van Daelen R, van Kammen A, et al. An improved method of partially digesting plant megabase DNA suitable for YAC cloning: application to the construction of a 5.5 genome equivalent YAC library of tomato. *Plant J* 1996; 9(1): 125-33.
- Borowsky R and Wilkens H. Mapping a cave fish genome: polygenic systems and regressive evolution. *J Hered* 2002; 93(1): 19-21.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Jr., Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 2005; 307(5717): 1928-33.
- Darnell R. Fishes of the Rio Tamesi and related Coastal Lagoons in East-Central Mexico. Publications of the Institute of Marine Science, Port Aransas, Texas. 1962; 8.
- Dowling TE, Martasian DP and Jeffery WR. Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. *Mol Biol Evol* 2002; 19(4): 446-55.

- Franz-Odenaal TA, Hall BK. 2006. Modularity and sense organs in the blind cavefish, *Astyanax mexicanus*. *Evol Dev*. 8: 94-100.
- Jeffery WR. Adaptive evolution of eye degeneration in the Mexican blind cavefish. *J Hered* 2005; 96(3): 185-96.
- Katagiri T, Asakawa S, Minagawa S, Shimizu N, Hirono I and Aoki T. Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. *Anim Genet* 2001; 32(4): 200-4.
- Katagiri T, Kidd C, Tomasino E, Davis JT, Wishon C, Stern JE, et al. A BAC-based physical map of the Nile tilapia genome. *BMC Genomics* 2005; 6(1): 89.
- Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet* 2004; 5(4): 288-98.
- Lang M, Miyake T, Braasch I, Tinnemore D, Siegel N, Salzburger W, et al. A BAC library of the East African haplochromine cichlid fish *Astatotilapia burtoni*. *J Exp Zool B Mol Dev Evol* 2005.
- Larin Z, Monaco AP and Lehrach H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc Natl Acad Sci U S A* 1991; 88(10): 4123-7.
- Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, et al. A Second-Generation Genetic Linkage Map of Tilapia (*Oreochromis* spp.). *Genetics* 2005; 170(1): 237-44.
- McCauley DW, Hixon E and Jeffery WR. Evolution of pigment cell regression in the cavefish *Astyanax*: a late step in melanogenesis. *Evol Dev* 2004; 6(4): 209-18.

- Mitchell RW , Elliott WR. Mexican eyeless characin fishes, genus *Astyanax*: Environment, distribution and evolution. Special Publication of the Museum of Texas Technical University 1977; 12: 1-89.
- Miyake T and Amemiya CT. BAC libraries and comparative genomics of aquatic chordate species. *Comp Biochem Physiol C Toxicol Pharmacol* 2004; 138(3): 233-44.
- Parry JW, Carleton KL, Spady T, Carboo A, Hunt DM and Bowmaker JK. Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Curr Biol* 2005; 15(19): 1734-9.
- Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, et al. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet* 2006; 38(1): 107-11.
- Quigley IK, Manuel JL, Roberts RA, Nuckels RJ, Herrington ER, MacDonald EL, et al. Evolutionary diversification of pigment pattern in Danio fishes: differential fms dependence and stripe loss in *D. albolineatus*. *Development* 2005; 132(1): 89-104.
- Sambrook J, Fritsch E, F. and Maniatis T: *Molecular cloning: A laboratory manual*. Cold Spring Harbour Press, 1989.
- Streelman JT, Albertson RC and Kocher TD. Genome mapping of the orange blotch color pattern in cichlid fishes. *Molecular Ecology* 2003; 12: 2465-2471.
- Stock DW, Jackman WR, Trapani J. 2006. Developmental genetic mechanisms of evolutionary tooth loss in cypriniform fishes. *Development* 133: 3127-37.

- Strong SJ, Ohta Y, Litman GW and Amemiya CT. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. *Nucleic Acids Res* 1997; 25(19): 3959-61.
- Tanaka M, Hale LA, Amores A, Yan YL, Cresko WA, Suzuki T, et al. Developmental genetic basis for the evolution of pelvic fin loss in the pufferfish *Takifugu rubripes*. *Dev Biol* 2005; 281(2): 227-39.
- Watanabe M, Kobayashi N, Fujiyama A and Okada N. Construction of a BAC library for *Haplochromis chilotes*, a cichlid fish from Lake Victoria. *Genes Genet Syst* 2003; 78(1): 103-5.
- Yamamoto Y, Jeffery WR. 2000. Central role for the lens in cave fish eye degeneration. *Science* 289: 631-3.

Table 1. BAC library details

Library	Vector	Cloning site	No. Clones	Average Insert size	No. Plates	Genome coverage	Non-insert Clones
<i>M. zebra</i>	pCC1	HindIII	56,832	110 kb	148	5.72x	3% (3/90)
<i>A. mexicanus</i>	pCC1	HindIII	58,752	104 kb	153	5x	2% (2/90)

Figure 1. Size distribution of BAC clones from the *Metriaclima zebra* (a), and the *Astyanax mexicanus* (b) BAC libraries. For each library 90 randomly selected clones were analyzed for their sizes by PFGE following *NotI* digestion.

Figure 2. Insert size analysis of BAC clones. Pulsed-field gel electrophoresis (PFGE) of *NotI* digested BAC clones from the (a) *Metriaclima zebra* and the (b) *Astyanax mexicanus* BAC libraries