First publ. in: Journal of experimental zoology / Part B, Molecular and developmental evolution, 306B (2006), pp. 567–574

A BAC Library for the Goldfish *Carassius auratus auratus* (Cyprinidae, Cypriniformes)

JING LUO[†]*, MICHAEL LANG[‡], WALTER SALZBURGER, NICOL SIEGEL, KAI N. STÖLTING, AND AXEL MEYER
Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany

ABSTRACT A goldfish (*Carassius auratus auratus*) bacterial artificial chromosome genomic library (BAC library) was constructed from one aquarium-bred male specimen (tetraploid, $4n = 100$, genome size = 3.52 pg/cell). The library consists of 128,352 positive clones with an average insert size of 150.4 kb, covering the genome 11-fold. All clones were spotted onto nylon filters and thus are available for screening of genomic regions of interest, such as candidate genes, gene families, or large-sized syntenic DNA regions of cyprinid species. Preliminary screens with two genes were conducted with hybridizing probes to the genes RAG1 and lgi1. RAG1 is a single-copy gene in zebrafish and is duplicated in *C. a. auratus*. We found a very close correlation between the number of positive BAC clones and the expected library coverage. Two copies of lgi1 were found in zebrafish. We have detected four different copies in *C. a. auratus*, not in the expected abundance, which indicates some variation in the coverage of the BAC library. The preliminary screens indicate that many duplicated genes that resulted from the ancient fish-specific genome duplication persist in the tetraploid goldfish genome. Hence, the BAC library will provide a useful resource for the future work on comparative genomics, polyploidy, diploidization, and evolutionary genomics in fishes.

Recent studies in comparative evolutionary genomics suggest that most likely two rounds of genome duplication occurred during vertebrate evolution (Ohno, '70; Amores et al., '98; Skrabanek and Wolfe, '98; Meyer and Schartl, '99; Wolfe, 2001). One round occurred on the shared lineage leading to the vertebrates and cephalochordates and a second round of duplication later in vertebrate evolution, somewhere around the divergence of fish (or amphibians). Another (third) round of genome duplication occurred in the teleost lineage, the most species-rich group of vertebrates. This event is thought to have taken place during the Devonian (Ohno and Atkin, '66; Ohno, '70; Amores et al., '98; Taylor et al., 2003; Vandepoele et al., 2004; Crow et al., 2005). It has been suggested that the fish-specific genome duplication is at least partially responsible for the species diversity of teleosts (Vogel, '98; Gregory and Hebert, '99; Meyer and Van de Peer, 2005; Yan et al., 2005). Thus, present vertebrate genomes can be regarded as degenerate polyploidy since genome duplication(s) have happened historically (Ohno, '70; Wolfe, 2001; Taylor et al., 2003; Vandepoele et al., 2004).

It must be better understood how duplicated genomes diploidize over evolutionary time and how long this takes (Ferris and White, '77; Allendorf, '78; Wolfe, 2001). More concrete, it must be elucidated how the organism can cope with the duplicated genome in the short run; how the dosage balance of genes is obtained in a genome that contains two or more copies of a gene; or what the mechanisms are that lead to the retention or eradication of duplicated genes in a genome. There has been discussion about to what extend duplicated genes take over novel gene functions and by which means this might occur

Grant sponsor: Max-Planck Gesellschaft; Grant sponsor: Universität Konstanz ; Grant sponsor: Deutsche Forschungsgemeinschaft.
*Correspondence to: J. Luo and A. Meyer, Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany.
E-mail: axel.meyer@uni-konstanz.de and jl.jingluo@gmail.com

[†]Current address: Institute of Hydrobiology, Chinese Academy of Sciences, 430072 Wuhan, Hubei, P.R. China.

[‡]Current address: Departament de Genètica, Universitat de Barcelona, 08028 Barcelona, España.
Received 27 September 2005; Accepted 30 March 2006
Published online 17 May 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21112.

(Force et al., '99; Lynch et al., 2001). Some of the above questions have been explored in botanic taxa. Yet, the theories derived from work on plants need to be tested in vertebrates (Furlong and Holland, 2004; Le comber and Smith, 2004).

Polyploidy occurs in several diverse groups of fish and might have played an important role in regulatory evolution (Le Comber and Smith, 2004). The importance of genome duplication in shaping the evolution of genomes can be best examined by comparative genomic analyses of DNA sequences from several closely related organisms that vary in their relative number of genome duplications, i.e., closely related diploid and tetraploid organisms.

It was known from cytological work that certain groups of fishes such as salmonids (e.g., salmon, trout, and their relatives) as well as cyprinids (e.g., carp, goldfish, and zebrafish) have the tendency to duplicate their genomes. Many species of cyprinids are polyploid, e.g., tetraploid ($N = 4$; goldfish and common carp; $2n = 100$), hexaploid ($N = 6$; *Schizothorax prenanti* Tchang; $2n = 148$), octaploid ($N = 8$; *Carassius auratus gebelio* and *C. a. langsdorfi*; $2n = 200\pm$), or rarely, triploid ($N = 3$; *Phoxinus 2 eos-neogaeus*) (Dawley and Goddard, '88; Yu et al., '89; Murakami et al., 2001; He et al., personal communication). Young polyploid cyprinid species can be found that underwent genome duplication events relatively recently, less than 10 million years ago (mya) (Lynch and Conery, 2000). Those species can be used to test the effects of the drastic genomic events, such as dosage balance or fast stabilization of duplicated genomes via retention/exclusion of redundant genomic DNA regions.

Molecular phylogenies suggest that an ancestral polyploidization event has taken place in cyprinids after the divergence of Danioninae and Cyprininae (Howes, '91; He et al., unpublished data). Moreover, genome duplication events have occurred recurrently in these lineages during the divergence of Cyprininae, Schizothoracinae, and Barbinae (Li et al., personal communication). Thus, this group of fish provides us an excellent model to test hypotheses relevant to the vertebrate genome duplication and diversification.

Bacterial artificial chromosome genomic libraries (BAC libraries) are essential resources for the analysis of genomic regions or entire genomes of higher organisms (Shizuya et al., '92). Once prepared from a species of interest, genomic DNA regions of about 150 kB in length are preserved in BAC clones and are accessible through library screening methods. Large genomic regions can be further analyzed by functional- or bioinformatic studies. BAC libraries are also an important genomic resource for genome sequencing projects (Katagiri et al., 2001; Osoegawa et al., 2001; Amores et al., 2004; Miyake and Amemiya, 2004; Donnison et al., 2005; Thorsen et al., 2005).

Here, we report the construction of a BAC library for the goldfish (*C. a. auratus*). The entire genus was thought to be tetraploid (15–20 mya according to Zan et al., '86; Risinger and Larhammar, '93; Yang and Gui, 2004). Our recent work (Luo et al., unpublished data) has also shown that within the last two million years, hexaploid populations have recurrently appeared in the different lineages of tetraploid populations. The goldfish is the artificial breed of the wild goldfish *Carassius auratus auratus* from China (Wang, 2000). It shows diverse morphologies in body shape and coloration. In addition, goldfish are widely used as model organisms in physiological research. The BAC library presented might aid investigations that concern the genomic consequences of genome duplication and tetraploidization in particular.

MATERIALS AND METHODS

DNA extraction and BAC library construction

DNA extraction and library construction was performed as described by Amemiya et al. ('96); Osoegawa et al. ('98); Danke et al. (2004), and as recently reported in Lang et al. (2006). In total, 300 μl of blood with heparin was obtained from the specimen. The concentration of blood cells was quantified to be approximately 10^9 cells/ml and quantities corresponding to 5.5×10^6 cells were embedded in 80 μl of 1% InCert agarose (FMC) for DNA extraction (Lang et al., 2006). Titration experiments were carried out with the restriction enzyme *Eco*RI (New England Biolabs (NEB), Ipswich, MA, USA) and *Eco*RI methylase (NEB) in order to optimize the conditions for size fragmentation of chromosomal DNA (Lang et al., 2006). Finally, genomic DNA from four agarose plugs was partially digested for 2.5 hr with 7 u restriction enzyme *Eco*RI and 42 u *Eco*RI methylase.

DNA fragments were separated on a 1% agarose gel using pulse field gel electrophoresis (BioRad CHEF DR III, Hercules, CA) as described by Osoegawa et al. ('98). Field inversion of the electric current is required for adequate separation of high-molecular-weight (HMW) DNA and

for exclusion of small DNA molecules during gel electrophoresis (Osoegawa et al., '98). This was achieved by the construction and addition of a 24-channel relay apparatus that was connected in-between the pulse field power module and the gel tank electrodes. Details about this apparatus are available upon request. Gel slices were prepared that contained size fragments of HMW DNA and the DNA fragments were electro-eluted and dialysed (Strong et al., '97; Danke et al., 2004). Then, 20 μl of HMW DNA fragment solution (approximately 50 ng) was ligated into the vector CopyControl™ pCC1BAC™ (*Eco*RI) (25 ng/μl) in 50 μl reaction volumes at 15°C for 12 hr. For transformation, the ligation mix was added to electro-competent cells DHB10T1 (Invitrogen, Carlsbad, CA) and electroporation was performed with the GenePulser (BioRad). Test ligations and transformations were performed. About 15–20 ng desalted DNA obtained from each ligation reaction were used in a 1 mm gap cuvette (MβP) with 20–25 μl electro-competent cells DHB10T1 at 1.25 kV pulses. Large-scale electroporation was carried out with approximately 75–100 ng DNA and one vial (100–130 μl) of electro-competent cells in a 2 mm gap cuvette (MβP, San Diego, CA, USA) and 2.5 kV pulses. Transformation products were supplemented with glycerol to a final concentration of 10% and 75 μl of each transformation product was grown on selective LB plates for determining the number and size of recombinant BAC clones. The remaining of the large-scale products were snap-frozen with liquid nitrogen and stored at −80°C.

Insert size estimation

To estimate the insert sizes, test amounts of transformation mix were grown on LB agar plates with chloramphenicol (12.5 μg/ml). BAC clones were isolated randomly and BAC clone DNA was prepared manually from 2 ml overnight cultures (Sambrook and Russel, 2001). Approximately one tenth of each BAC clone DNA was digested with *Not*I (Invitrogen) and the samples were run on a pulsed-field gel. The size of each clone was estimated manually with Low-Range PFG Markers (NEB).

Genome size estimation

Arterial blood was obtained as stated in Lang et al. (2006). A volume of 200 μl blood was mixed with 50 μl heparin (100 μg/μl, Roche) and stored at 4°C. Cells were washed twice with 0.85 × PBS and centrifuged for 5 min at 400*g* after each wash. Samples were stained for 15–30 min with PI buffer (0.85 × PBS, containing 50 μg/ml propidium iodide, 50 μg/ml RNase, 1% Triton X-100, 0.1% sodium citrate). Genome sizes were estimated by flow cytometry (Beckman, Fullerton, CA, USA) by comparison to the known genome size of *Astatotilapia burtoni* (Perciformes; Cichlidae) and *Gallus domesticus* (Gregory, 2005; Lang et al., 2006).

Arraying of high-density bacterial colony filters

BAC clones were picked with the robot Q-PIX (Genetix Ltd., New Milton, UK) into 384-well plates, filled with LB broth, 10% glycerol. The position P23 in each plate was left empty and the position P24 was used for a positive control. In total, 336 plates containing 128,352 (336 plate × 382 wells/plate) BAC clones were picked that theoretically cover the goldfish genome 11-fold. Two replicas were made from each plate. One replica was used to spot BAC arrays on nylon filters for future screening. Seven filters were spotted to cover 336 × 384-well microtiter plates leading to a total of 129,024 clones including negative and positive clones.

Library screening

The nylon filter set (seven filters) was hybridized with gene-specific biotinylated DNA probes, using the NEBlot Phototope Kit (NEB). The probes of RAG1 and lgi1 were synthesized by random priming from 100 ng of PCR product of 1,500 and 400 bp in length, respectively, amplified from *C. a. auratus* genomic DNA. RAG1 was amplified with degenerate primers, designed by Lópes et al. (2004) RAG1F1 5′-GTG AGC TGC AGT CAG TAC CAT AAG ATG T-3′ and RAG1R1 5′-CTG AGT CCT TGT GAG CTT CCA TRA AYT T-3′. This primer pair targets the region spanning between nucleotide positions 2,215 and 3,772 of the RAG1 gene of *Oncorhynchus mykiss* (accession no. NC001717) (Lópes et al., 2004). For lgi1, the primers lgi1_For1087 5′-YAC MGS GAC ACB GAT GTG G-3′ and lgi1_Rev1517 5′-GWG AAD GMH CTN GGY GCC TG-3′ were used, designed on zebrafish lgi1 and other available sequences (Gu et al., 2005).

Pre-hybridization and hybridization was performed in hybridization bottles (70 × 300 mm^2) (Thermo, Waltham, MA, USA). First, filters were set to incubate overnight at 68°C in 50 ml pre-hybridization solution (6 × SSC, 0.5% SDS,

100 μg/ml denaturated salmon sperm DNA, 5 × Denhardt's solution). Denatured, biotinylated DNA probe was added to each bottle and hybridization was carried out overnight at 68°C. Filters were washed twice at RT in 2 × SSC, 0.5% SDS and twice in 0.1 × SSC, 0.1% SDS at 68°C. Chemiluminescent detection of labeled probe was performed using the Phototope-Star Detection kit (NEB) in Bio-Assay dishes [24.5 × 24.5 cm^2] (Nunc, Rochester, NY, USA) followed by 4–7 min exposure to films (Hyperfilm™, Amersham, Little Chalfont, UK) in autoradiography cassettes. BAC clone DNA of positive clones was prepared as described above. The positive clones were further analyzed by PCR with above-stated primers and restriction digestions, followed by gel electrophoresis.

Phylogenetic analysis and divergence time estimation

RAG1 DNA sequences were aligned using the DNASTAR software package 5.0 (DNASTAR) and confirmed by eye. Afterwards, pairwise sequence alignments combined with the four downloaded sequences, *Pimephales promelas* (AY430210), *Hesperoleucus symmetricus* (AY059468), *Lavinia exilicauda* (AY059469), and *Danio rerio* (U71093) were conducted with BioEdit (Hall, '99). We performed phylogenetic analysis using PAUP 4.0b10 (Swofford, 2002). Here, we used maximum likelihood to reconstruct phylogenetic relationships of paralogs, applying the SYM+I model of molecular evolution that was chosen as the best fitting model by means of a maximum-likelihood ratio test with MODELTEST 3.06 (Posada and Crandall, '98). In addition, node support was assessed by bootstrapping with 1,000 replicates.

Time since divergence (T) between the two RAG1 paralogs was calculated using the formula $T = D_{ML}/2r$, where D_{ML} is the maximum likelihood distance between the homologous sequences and r is the rate of substitution.

RESULTS AND DISCUSSION

BAC library construction

Genomic DNA was obtained from arterial blood of one sacrificed male goldfish and partially digested with the restriction enzymes *Eco*RI and *Eco*RI methylase. HMW genomic DNA fragments were prepared and further ligated into pCC1BACTM (Epicentre, Madison, WI, USA) (see *Materials and Methods*). HMW DNA fragments of an average size of 150 kb were used for library construction as cloning of those DNA fragments yielded approximately 2,490 recombinant BAC clones per microliter of DNA solution. Overall, this fraction could have possibly yielded around 448,200 recombinant clones.

A total of 141 clones were isolated randomly to assess the distribution of clone insert lengths of the library from large-scale transformations. *Not*I digestions were prepared from each clone (see *Materials and Methods*) and BAC clone insert sizes were estimated by comparisons with HMW DNA molecular weight standards (NEB Low-Range Marker). The distribution of clone insert sizes had a mean of 150 kb and a median of 145 kb (Fig. 1). The proportion of clones in the library that were smaller than 100 kb was estimated to be 6.4%; seven clones did not contain any insert.

The genome size of *C. a. auratus* was an important reference value for assessing the completeness and the genomic coverage of the BAC library. We obtained this value experimentally by fluorescence-activated flow cytometry. Literature data are available but those estimates vary considerably among each other. Our estimate of the gemome size was 3.52 pg/cell, which is very close to the average of estimates from the literature data (3.59 pg/cell, including our data) (Ohno et al., '67; Hinegardner, '68; Wolf et al., '69; Beamish et al., '71; Hinegardner and Rosen, '72; Hafez et al., '78; Mauro and Micheli, '79; Zan et al., '86; Ojima and Yamamoto, '90; Vinogradov, '98; Collares-Pereira and Moreira da Costa, '99; Ciudad et al., 2002; Hardie and Hebert, 2004;

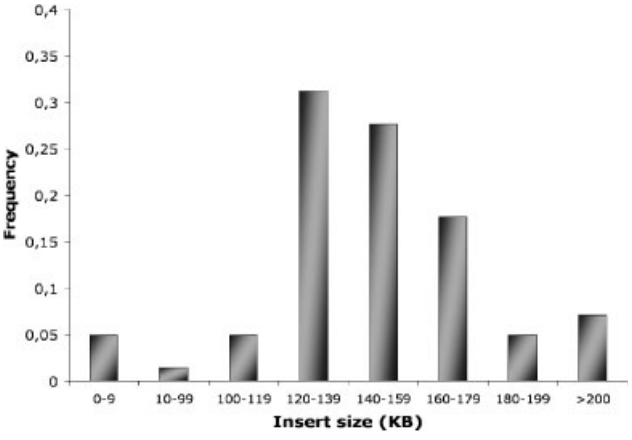

Fig. 1. Size distribution of BAC clone inserts. Diagram was prepared from 141 samples. BAC clones were randomly picked, digested with the restriction enzyme *Not*I and run on a CHEF gel. Insert sizes were estimated manually by comparison with standard size markers (PFG Low-Range Marker, NEB).

Gregory, 2005). We relied on our estimate as a reference value for the assessment of the genomic coverage of the BAC library since it did not differ significantly from the calculated average goldfish genome size.

Arraying filters and screening

It was calculated that a total of 119,348 clones would comprise a 10-fold coverage of the goldfish genome. In total, 128,352 clones were isolated into 336 microtiter plates (384-well) leading to an 11-fold coverage of the *C. a. auratus* genome, with an average insert size of 150.4 kB. One well of each plate was left empty while another well contained a characterized BAC clone from a previously prepared genomic library of *A. burtoni* (Lang et al., 2006). Three replicas of the whole library were prepared. Additionally, the entire library was used for preparation of a nylon filter set for later screening purposes. Two sets of nylon filters were spotted and each set was composed of seven filters.

The utility of the library was examined by screening the BAC library nylon filter set with gene-specific probes for lgi1 and RAG1 fragments. RAG1 is a single-copy gene in diploid cyprinids (Willett et al., '97) while lgi1 is known to be duplicated in zebrafish and pufferfish (Gu et al., 2005). The paralogous lgi1 genes are found in separate genomic regions as indicated by Gu et al. (2005) on two different chromosomes of zebrafish (chromosomes 12 and 13). The two lgi paralogs are expected to have originated approximately 250 mya from the teleost-specific genome duplication (Taylor et al., 2003; Vandepoele et al., 2004). The divergence time between Zebrafish and goldfish is around 60 mya according to Steinke et al. (2006). Thus, most likely the two paralogs of lgi1 were present in the ancestral species of the goldfish lineage. Hence, due to the recent genome duplication, there could be at most four paralogous lgi genes retained in the goldfish genome (Quiroz-Gutierrez and Ohno, '70; Risinger and Larhammar, '93; Yang and Gui, 2004).

We screened the whole set of seven nylon filters with the 11 × coverage of the BAC library. The screen for the gene lgi1 resulted in seven positive clones, which were confirmed by PCR and sequencing. In these seven positive clones, four different copies were distinguished, termed lgi1aa (2 clones), lgi1ab (1 clone), lgi1ba (1 clone), and lgi1bb (3 clones). There was 78.93% sequence identity between lgi1a and lgi1b (311 sites out of 394 nucleotides) and 84% identity at the amino acid level (111 residues out of 132 amino acids). Two variable sites were found comparing the nucleotide sequences of lgi1aa and lgi1ab, and between lgi1ba and lgi1bb only one variable site was found. Independent PCR- and sequencing reactions were conducted that confirmed these sites to be real mutations. Two paralogs (aa and bb) can be clearly distinguished by the fingerprint pattern, since the fingerprint patterns of lgi1aa/lgi1ab are more similar to each other than either of them to lgi1ba or to lgi1bb (Fig. 2A). All sequences have been deposited at GenBank (accession numbers DQ196521–DQ196524).

A second screen was carried out for the RAG1 gene. The probe for the RAG1 gene was ~1.5 kb in length. We detected 20 positive clones, which were confirmed by PCR amplification and sequence analysis of a 1.4 kb DNA fragment of corresponding BAC clone DNA. We obtained three different sequence variants. Two variants were almost identical to each other (99.8% identity) and were comprised of three and six clones, respectively. The third variant comprised of 11 positive clones, showed 92.8% DNA sequence identity to the other two variants. These sequences have been deposited in GenBank (accession numbers DQ196518–DQ196520). In addition, the *Eco*RI restriction pattern of BAC clones that contain the different RAG1 variants indicates two different genomic regions, while the two almost

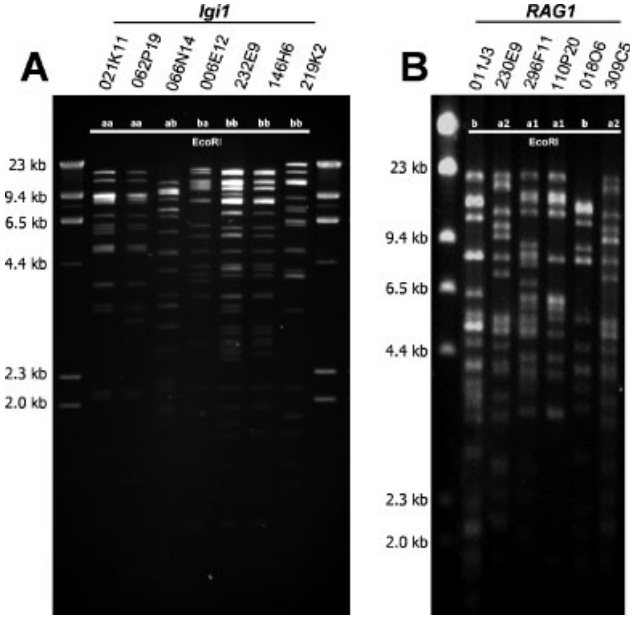

Fig. 2. Restriction digestions of BAC clones with the enzyme *Eco*RI; CHEF gels of restriction digestions with and *lgi 1* (**A**) *RAG 1* (**B**) containing BAC clones.

identical variants appear to be alleles of the same locus and are, hence, paralogous to the third variant (Fig. 2A). Fingerprint patterns of those BAC clones with the restriction enzymes *Bam*HI and *Hind*III showed similar results (data not shown). As discussed later on, the duplication of RAG1 in the lineage of *C. a. auratus* has most likely originated during the recent tetraploidization event.

The *Eco*RI fingerprint patterns of digested BAC clones (Fig. 2) show that the clones do indeed result from partially digested genomic DNA. As the BAC library covers the haploid genome approximately 11 times and as it is comprised of recombinant clones of essentially random areas of the goldfish genome, many BAC clone inserts should be overlapping. Hence, this BAC library might be employed for regional analysis of large genomic regions, assembled from BAC clone contigs.

In summary, the screening results for lgi1 did not accord well with the expected genomic coverage and more screens are necessary for a better evaluation of the depth of the library coverage. On the other hand, the number of positive clones obtained for RAG1 was very close to the expected 11-fold coverage of the *C. a. auratus* genome. To end with, positive clones were detected in both screens suggesting that the library represents a solid resource for future studies on genomic effects of genome duplications.

Phylogenetic relationships and divergence estimation

The family cyprinidae includes four subfamilies Cyprininae, Schizothoracinae, Barbinae, and Labeoninae. Species of three of those subfamilies, Cyprininae, Schizothoracinae, and Barbinae, are known to have undergone repeated genome duplication events (Fig. 3A; He et al., personal communication). The goldfish belongs to the subfamily Cyprininae (Howes, '91). Within this subfamily, ancestors of all species were thought to be tetraploids (Yu et al., '89). In addition, recent hexaploids were discovered in different lineages as well (Luo et al., unpublished data). Those genome duplication events have occurred within the last twenty million years (Risinger and Larhammar, '93; Yang and Gui, 2004).

In order to date the genome duplication event, RAG1 sequences were used to estimate the divergence time (T) between the two copies of RAG1 with the formula $T = D_{ML}/2r$, where D_{ML} is the maximum likelihood distance between duplicated sequences (Fig. 3B) and r is the rate of substitution. The rate r was estimated to be 2.53×10^{-9}/year, using the divergence time of 60 mya between the subfamilies Danionae and Cyprininae (Steinke et al., 2006). The divergence analysis with the RAG1 variants a and b dated the gene duplication to have occurred about 14.2–14.5 mya (Fig. 3B). As this estimation agrees with previous dating of goldfish-specific gene duplications (Risinger and Larhammar, '93; Yang and Gui, 2004), it is likely that the duplication of RAG1 resembles the genome duplication event of the goldfish. The lgi1 sequences were relatively short and conserved and therefore could not be used to estimate a divergence date of the duplicates.

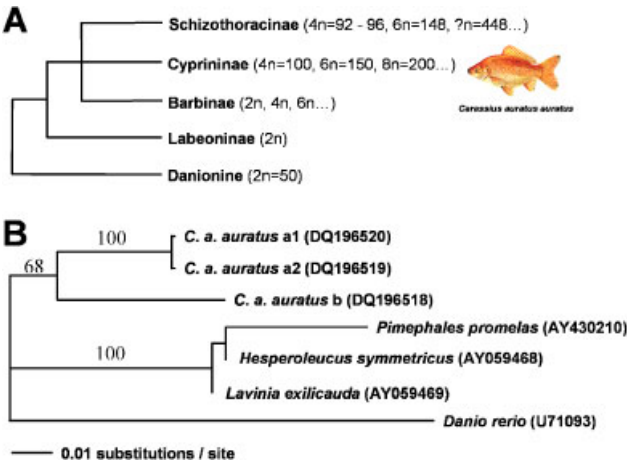

Fig. 3. Dating of the RAG1 gene duplication. (**A**) Phylogenetic relationships and documented polyploidy of Cyprinidae after Howes ('91). (**B**) Maximum Likelihood tree (PAUP software package, (Swofford, 2002) of three different variants of goldfish *RAG1* sequence fragments that were described in this study. ML distances were calculated with the SYM+I model. ML distance of *RAG1* variant *b* to *a*1 and *a*2 was 7.216% and 7.334%, respectively. The branch supports were estimated by bootstrapping with 1,000 replicas. Gene Bank accession numbers of each analyzed DNA sequence are indicated in brackets.

Comparative genomic studies promise to be helpful for the understanding of the genetic events that follow polyploidization events. The BAC library presented here is intended as a genomic resource to facilitate the study of goldfish and carp genomes and to help gaining insights into the processes that accompany genome duplication processes. Copies of the library are located in the laboratory of Axel Meyer, at the University of Konstanz, Konstanz, Germany, and soon in the laboratory of Jing Luo, at the Hydrobiology

Institute of the Chinese Academy of Sciences, Wuhan, China. Each copy includes 336 384-well plates and is stored at −80°C. The library is available to the scientific community by the laboratory of Axel Meyer, University of Konstanz, Konstanz, Germany or by Jing Luo (Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China).

ACKNOWLEDGMENTS

We thank Ingo Braasch for helpful advice during the screening of the BAC library, the staff at GATC for clone picking and arraying, Claus Lang for technical support and Shuping He for discussions about cyprinid phylogeny. We also thank Dave Gerrard for comments on the manuscript, Yann Gibert for lgi sequences, and Elke Hespeler for technical assistance in the laboratory. This work was supported by the Max-Planck Gesellschaft to J.L. and A.M. and the Universität Konstanz, Germany and the Deutsche Forschungsgemeinschaft to A.M.

LITERATURE CITED

Allendorf FW. 1978. Protein polymorphism and the rate of loss of duplicate gene expression. Nature 272:76–78.

Amemiya CT, Ota T, Litman GW. 1996. Construction of P1 artificial chromosome (PAC) libraries from lower vertebrates. Nonmammalian Genom Anal: A Pract Guide 223–255.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH. 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711–1714.

Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C, Postlethwait JH. 2004. Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. Genome Res 14:1–10.

Beamish RJ, Merrilees MJ, Crossman EJ. 1971. Karyotypes and DNA values for members of the suborder Esocoidei (Osteichthyes: Salmoniformes). Chromosoma 34:436–447.

Ciudad J, Cid E, Velasco A, Lara JM, Aijón J, Orfao A. 2002. Flow cytometry measurement of the DNA contents of G0/G1 diploid cells from three different teleost fish species. Cytometry 48:20–25.

Collares-Pereira MJ, Moreira da Costa L. 1999. Intraspecific and interspecific genome size variation in Iberian Cyprinidae and the problem of diploidy and polyploidy, with a review of genome sizes within the family. Folia Zool 48: 61–76.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. 2005. The ''fish specific'' Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23: 121–136.

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.

Dawley RM, Goddard KA. 1988. Diploid-triploid mosaics among unisexual hybrids of the minnows *Phoxinus eos* and *Phoxinus neogaeus*. Evolution 42:649–659.

Donnison IS, O'Sullivan DM, Thomas A, Canter P, Moore B, Armstead I, Thomas H, Edwards KJ, King IP. 2005. Construction of a *Festuca pratensis* BAC library for map-based cloning in Festulolium substitution lines. Theor Appl Genet 110:846–851.

Ferris SD, White GS. 1977. Loss of duplicate gene-expression after polyploidization. Nature 265:258–260.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.

Furlong RF, Holland PWH. 2004. Polyploidy in vertebrate ancestry: Ohno and beyond. Biol J Linn Soc 82:425–430.

Gregory TR. 2005. Animal genome size database: http://www.genomesize.com

Gregory TR, Hebert PDN. 1999. The modulation of DNA content: proximate causes and ultimate consequences. Genome Res 9:317–324.

Gu W, Gibert Y, Wirth T, Elischer A, Bloch W, Meyer A, Steinlein OK, Begemann G. 2005. Using gene-history and expression analyses to assess the involvement of LGI genes in guman disorders. Mol Biol Evol 22:2209–2216.

Hafez R, Labat R, Quillier R. 1978. Teneurs nucléaires en A.D.N. et relations évolutives dans la famille des cyprinides (Teleostei). Bull Soc Hist Nat Toulouse 114:71–84.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis programm for Windows95/98/NT. Nucleic Acids Symp Ser 41:95–98.

Hardie DC, Hebert PDN. 2004. Genome-size evolution in fishes. Can J Fish Aquat Sci 61:1636–1646.

Hinegardner R. 1968. Evolution of cellular DNA content in teleost fishes. Am Nat 102:517–523.

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

Howes GJ. 1991. Systematics and biogeography: an overview. In: Winfield IJ, Nelson JS, editors. Cyprinid fishes, systematics, biology and exploitation. London: Chapman & Hall. p 1–33.

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.

Lang M, Miyake T, Braasch I, Tinnemore D, Siegel N, Salzburger W, Amemiya CT, Meyer A. 2006. A BAC Library of the East African Haplochromine Cichlid Fish *Astatotilapia burtoni*. J Exp Zool B: Mol Dev Evol 306:35–44.

Le Comber S, Smith C. 2004. Polyploidy in fishes. Biol J Linn Soc 82:432–442.

Lópes JA, Chen WJ, Orti G. 2004. Esociform phylogeny. Copeia 2004:449–464.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804.

Mauro ML, Micheli G. 1979. DNA reassociation kinetics in diploid and phylogenetically tetraploid cyprinidae. J Exp Zool 208:407–416.

Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11:699–704.

Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27: 937–945.

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

Murakami M, Matsuba C, Fujitani H. 2001. The maternal origins of the triploid ginbuna (*Carassius auratus langsdorfi*): phylogenetic relationships within the *C. auratus* taxa by partial mitochondrial D-loop sequencing. Genes Genet Syst 76:25–32.

Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.

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

Ohno S, Muramoto J, Christian L, Atkin NB. 1967. Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae. Chromosoma 23:1–9.

Ojima Y, Yamamoto K. 1990. Cellular DNA contents of fishes determined by flow cytometry. La Kromosomo II 57: 1871–1888.

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.

Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res 11:483–496.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Quiroz-Gutierrez A, Ohno S. 1970. The evidence of gene duplication for s-form NADP-linked isocitrate dehydrogenase in carp and goldfish. Biochem Genet 4:93–99.

Risinger C, Larhammar D. 1993. Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. Proc Natl Acad Sci USA 90:10598–10602.

Sambrook J, Russel DW. 2001. Molecular cloning, a laboratory manual, 3rd edition. 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-pair fragments of human DNA in Escherichia coli usig an F-factor-based vector. Proc Natl Acad Sci USA 89: 8794–8797.

Skrabanek L, Wolfe KH. 1998. Eukaryotic genome duplication—where's the evidence? Curr Opin Genet Dev 8: 694–700.

Steinke D, Salzburger W, Braasch I, Meyer A. 2006. Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genomics (online).

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.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10*. Sunderland, MA: Sinauer Associates.

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.

Thorsen J, Zhu B, Frengen E, Osoegawa K, de Jong PJ, Koop BF, Davidson WS, Hoyheim B. 2005. A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6:50.

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA 101:1638–1643.

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

Vogel G. 1998. Doubled genes may explain fish diversity. Science 281:1119–1121.

Wang CY. 2000. Goldfish in China. Beijing: Jindun Press.

Willett CE, Cherry JJ, Steiner LA. 1997. Characterization and expression of the recombination activating genes (RAG1 and rag2) of zebrafish. Immunogenetics 45:394–404.

Wolf U, Ritter H, Atkin NB, Ohno S. 1969. Polyploidization in the fish family Cyprinidae, Order Cypriniformes. I. DNA-content and chromosome sets in various species of Cyprinidae. Humangenetik 7:240–244.

Wolfe K. 2001. Yesterday's polyploidy and the mystery of diploidization. Nat Rev Genet 2:333–341.

Yan YL, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH. 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132:1069–1083.

Yang L, Gui JF. 2004. Positive selection on multiple antique allelic lineages of transferrin in the polyploid *Carassius auratus*. Mol Biol Evol 21:1264–1277.

Yu XJ, Zhou T, Li YC, L K, Zhou M. 1989. Chromosomes of Chinese freshwater fishes. Beijing: Science Press.

Zan R, Song Z, Liu W. 1986. Studies on karyotypes and nuclear DNA contents of some cyprinoid fishes, with notes on fish polyploids in China. In: Uyeno T, Arai R, Taniuchi T, Matsuura K, editors. Indo-Pacific fish biology, Tokyo: Ichthyological Society of Japan. p 877–885.