Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 337 (2005) 216–223
BBRC
www.elsevier.com/locate/ybbrc

Duplicated cytoglobin genes in teleost fishes ☆

Christine Fuchs [a], Anne Luckhardt [a], Frank Gerlach [a], Thorsten Burmester [b], Thomas Hankeln [a,*]

[a] Institute of Molecular Genetics, Johannes Gutenberg-University Mainz, D-55099 Mainz, Germany
[b] Institute of Zoology, Johannes Gutenberg-University Mainz, D-55099 Mainz, Germany

Received 18 August 2005
Available online 15 September 2005

Abstract

Cytoglobin is a recently discovered myoglobin-related O_2-binding protein of vertebrates with uncertain function. It occurs as single-copy gene in mammals. Here, we demonstrate the presence of two paralogous cytoglobin genes (*Cygb-1* and *Cygb-2*) in the teleost fishes *Danio rerio*, *Oryzias latipes*, *Tetraodon nigroviridis*, and *Takifugu rubripes*. The globin-typical introns at positions B12.2 and G7.0 are conserved in both genes, whereas the C-terminal exon found in mammalian cytoglobin is absent in the fish genes. Phylogenetic analyses show that the two cytoglobin genes diverged early in teleost evolution. This is confirmed by gene synteny analyses, which suggest a large-scale duplication event. Although both cytoglobin genes are highly conserved and have evolved under purifying selection, substitution rates are significantly higher in *Cygb-1* than in *Cygb-2*. Similar to their mammalian ortholog, both fish cytoglobins are expressed in a broad range of tissues. However, *Cygb-2* is more than 250-fold stronger expressed in neuronal tissues, suggesting a subfunctionalization of the two cytoglobin paralogs after gene duplication.

Keywords: Cytoglobin; Globin; Genome duplication; Evolution; Synteny

Cytoglobin (Cygb) is a heme-containing respiratory protein that has only recently been added to the protein family of vertebrate globins [1–3]. Cygb shares a common phylogenetic ancestry with vertebrate myoglobin (Mb), from which it diverged by duplication about 450 million years ago (MYA) before radiation of the jawed vertebrates [3]. Although sharing only about 30% of the amino acids with vertebrate Mb, Cygb proteins exhibit the classical globin fold [4,5]. Due to N- and C-terminal extensions, mammalian Cygb proteins are about 40 amino acids longer than typical globins and measure 190 amino acids. Nevertheless, recombinant mouse and human Cygbs bind O_2 with an Mb-like affinity in the range of $P_{50} = 1$ Torr [2,6]. In contrast to the penta-coordination of the heme-iron in deoxygenated vertebrate hemoglobin (Hb) and Mb, Cygb displays a hexa-coordinate binding scheme in which the distal histidine is bound to the Fe^{2+} ion.

The physiological function of Cygb is not well understood. While initially proposed to be universally expressed (thus its name) [3], recent studies in mammals have confined Cygb expression to the fibroblast cell lineage (i.e., fibroblasts, chondroblasts, osteoblasts, and hepatic stellate cells), and to distinct neuronal cells in the central and peripheral nervous systems, including the retina [7–10]. These data suggested that Cygb might in fact fulfill different, tissue-specific functional roles, e.g., by being involved in the detoxification of reactive oxygen or nitrogen species or by supplying O_2 to specific enzymatic reactions, such as the hydroxylation of collagen or the synthesis of nitric oxide [11].

Cygb sequences have been identified by now in various mammals, chicken (*Gallus gallus*), frog (*Xenopus tropicalis* and *Xenopus laevis*), and zebrafish (*Danio rerio*) [1–3,12–14]. So far, only a single *Cygb* gene has been identified in all of these vertebrate taxa. Here, we report that actually two distinct, paralogous *Cygb* gene lineages exist in the

☆ *Abbreviations:* Cygb, cytoglobin; Hb, hemoglobin; Mb, myoglobin; Ngb, neuroglobin; GbX, globin X; MYA, million years ago; EST, expressed sequence tag; CDS, coding sequence.

* Corresponding author. Fax: +49 6131 39 24585.
E-mail address: hankeln@uni-mainz.de (T. Hankeln).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.08.271

teleost fishes, which is interesting in the light of the proposed whole-genome duplication in early fish evolution [15,16]. We study the genomic organization, molecular evolution, and expression patterns of the Cygb paralogs in the zebrafish *D. rerio*, the medaka *Oryzias latipes*, and in the pufferfishes *Tetraodon nigroviridis* and *Takifugu rubripes*.

Materials and methods

Animals. Specimens of the zebrafish *D. rerio* and the spotted green pufferfish *T. nigroviridis* were purchased from a local pet shop and kept in culture at 26 °C at the Institute of Zoology, University of Mainz. Specimens of the medaka *O. latipes* were kindly provided by Professor Dr. M. Schartl, University of Würzburg.

Database analyses, sequence alignment, and phylogenetic inference. The BLAST algorithm [17] was applied to search the genomic DNA sequence databases of *T. nigroviridis* (http://www.genoscope.fr/) and *T. rubripes* (http://genome.jgi.psf.org/fugu6/fugu6.home.html), respectively. The NCBI database (http://www.ncbi.nlm.nih.gov/BLAST/) was used to search expressed sequence tags (ESTs) of both pufferfish species as well as of *D. rerio* and *O. latipes*. The nucleotide sequences were extracted from the databases, and assembled and translated using the DNAstar 5.08 (Lasergene) and Genedoc 2.6 [18] software. Large-scale genomic analysis was performed based on the data shown by the genomic browsers of Genoscope (*T. nigroviridis*) and the NCBI Map Viewer database (*Homo sapiens*; http://www.ncbi.nlm.nih.gov/mapview/).

The deduced Cygb amino acid sequences were manually added to an alignment of selected globin sequences, using the known Cygb crystal structure [19] and previously published alignments [3,12,20] as templates. The sequences used for the alignment are: Neuroglobin (Ngb) from *H. sapiens* (HsaNGB, Accession No. AJ245946), the pufferfish *T. nigroviridis* (TniNgb, AJ315609), the zebrafish *D. rerio* (DreNgb, AJ315610), the rainbow trout *Oncorhynchus mykiss* (OmyNgb, AJ547800), the chicken *G. gallus* (GgaNgb, AJ635192), and the Western clawed frog *X. tropicalis* (XtrNgb, AJ634914), globin X from *X. tropicalis* (XtrGbX, AJ634915) and *D. rerio* (DreGbX, AJ635194), Mb from the zebrafish (DreMb, AAR00323), human (HsaMB, M14603), and chicken (GgaMb, XM_416292), Hb α and β from human (HsaHB α, J00153, HsaHB β, M36640), and the tetrapod Cygbs from human (HsaCYGB, AJ315162), mouse (MmuCygb, AJ315163), chicken (GgaCygb, AJ812227), and *X. tropicalis* (XtrCygb, BC076983), the teleost Cygbs from *D. rerio* (DreCygb-1, AJ320232, DreCygb-2, AJ635229), *O. latipes* (OlaCygb-1, AJ635227, OlaCygb-2, AJ635228), *T. nigroviridis* (TniCygb-1, AJ635230, TniCygb-2, AJ635231), and *T. rubripes* (TruCygb-1, genomic Scaffold_2147, TruCygb-2, Scaffold_6463 (exons 1 and 2) and Scaffold_5409 (exon 3) at IMCB (http://www.fugu-sg.org/)) as well as the trout (OmyCygb-1, AJ547802). Final alignments are available from the authors upon request.

Bayesian phylogenetic reconstructions were performed using MrBayes 3.0 β 4 [21] applying the PAM model of amino acid evolution [22] with 500,000 generations. Trees were sampled every 10th generation and posterior probabilities were estimated on the final 25,000 trees (burn-in = 25,000). MEGA3.0 software [23] was used for phylogenetic tree reconstructions using distance data. Amino acid distance matrices were calculated using the PAM models with uniform rates among sites. Neighbor-joining trees were compiled with MEGA3.0. The reliability of the branching pattern was tested by bootstrap analyses with 1000 replications. The calculation of the nonsynonymous (dn) and synonymous (ds) nucleotide substitution rates was accomplished using the tool SNAP (http://www.hiv.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html), which implements the method of Nei and Gojobori [24]. Amino acid and nucleotide substitution rates were estimated from the PAM distance matrix, assuming that the divergence of Actinopterygii and Sarcopterygii took place about 420 MYA [28].

Cloning and sequencing of teleost Cygb-1 and Cygb-2 cDNAs. Selected tissues of freshly killed specimens of *D. rerio* and *T. nigroviridis* were dissected and stored at −80 °C, while specimens of *O. latipes* were prepared as whole animals. Total RNA was extracted using the RNeasy Kit (Qiagen) according to the manufacturer's instructions, including a DNase I digestion step to remove genomic DNA. One microgram of the zebrafish tissues, 0.5 μg from each green spotted pufferfish tissue, and 0.5 μg of the whole animal RNA of medaka were reverse-transcribed using the SUPERSCRIPT II reverse transcriptase (Invitrogen) with an oligo(dT) primer. The integrity of the cDNA was tested by PCR using β-actin-specific primers. The complete coding sequences (CDS) of the bioinformatically identified teleost Cygb genes were amplified using 1/10 of the cDNA and 2 U Taq polymerase (Sigma) in a standard PCR protocol with 30 s annealing time at 55–58 °C. The following RT-PCR primers were employed: *D. rerio*: 5′-ATGGATGAGGAAATCGCTGC-3′ and 5′-CTG TGTCATCTTCTCTCTGTTGGC-3′ (*β-actin*), 5′-ATGGAGAAAGAG AGGGAGA-3′ and 5′-TCAGACTGCTGAGCTGGATAG-3′ (*Cygb-2*); *O. latipes*: 5′-ATGGATGATGACATTGCCGC-3′ and 5′-GTCATCTTC TCCCTGTTGGCTTTG-3′ (*β-actin*), 5′-ATGGAGAGCAGGCACCCT C-3′ and 5′-TCAGCTGGTGAGTTGGAGAG-3′ (*Cygb-1*), and 5′-ATG TCTTGCAGGGAGTCTCCACCGCC-3′ in combination with 5′-TCAG ACTGCTGAGCTGGATAG-3′ (*Cygb-2*); *T. nigroviridis*: 5′-ATGGATG AGGAAATCGCCGCC-3′ and 5′-GTCATCTTTTCCCTGTTGGCCTT G-3′ (*β-actin*), 5′-ATGACCATATGGAAAGGATGCAGAGGG-3′ and 5′-TCACTCCACTGAGGACGACA-3′ (*Cygb-1*), and 5′-ATGACCATA TGGTTYTTTGTYAATTTCCC-3′ combined with 5′-ATCGCGGATC CTCACACGGCTGAGCTGGAGAC-3′ (*Cygb-2*). PCR products were cloned into the pGEM T-Easy vector system (Promega) and both strands were sequenced by a commercial sequencing service (GENterprise).

Quantitative real-time RT-PCR (qRT-PCR). Equal amounts of total RNA of *D. rerio* were reverse-transcribed as described above. As a first step of sample normalization, the concentration of the DNase I treated RNA was measured by optical density (Eppendorf Biophotometer). For normalization on the cDNA level, cDNAs were quantified relatively to each other using the PicoGreen reagent (Molecular Probes), which specifically detects RNA-cDNA heteroduplexes. Traces of dsDNA had previously been removed by DNase I digestion. The fluorescence was detected using the TBS-380 Mini fluorometer (Turner BioSystems).

Real-time PCR experiments were carried out on an ABI Prism 7000 SDS (Applied Biosystems). For a 20 μl PCR, 2 μl cDNA and 18 μl of a mastermix ("Absolute QPCR SYBR Green ROX Mix" (ABgene) containing dNTPs, a "hot start" polymerase (Thermo Start DNA Polymerase, ABgene House), buffer, SYBR Green, and the primer oligonucleotides at 10 pmol/μl), were used. Primer sequences were 5′-GCTGGAGTGGCTGTGCTG-3′ and 5′-GTTGAGAGCGTTCAAGACTCGCTGTCCG-3′ (Dre *Cygb-1*), and 5′-ACATGACCCAGAGAAAGTATCATCTG-3′ in combination with 5′-CTCTCCAAACTCCTCTGCAAG-3′ (Dre *Cygb-2*); the final concentration of the oligonucleotides was 190 nM. The cycling protocol included an initial activation step at 95 °C for 15 min followed by 40 cycles of a four-step protocol (94 °C for 15 s, 56 °C for 30 s, 72 °C for 30 s, and a detection step at 76.8 °C for 30 s). Dissociation curves were obtained from 60 to 92 °C to analyze the specificity of the amplification reactions. The qRT-PCR results were evaluated by means of the standard curve method. For standard curve generation, known amounts of plasmids containing the appropriate gene amplicon as insert were used at dilutions ranging from 10^7 to 10^3 copies.

Results and discussion

Two Cygb paralogs in teleosts

Analyses of the complete genome sequences from human and mouse, and the nearly complete draft versions of chicken and frog revealed that *Cygb* is a single-copy gene in these taxa [3,13,14]. Initially, we detected only a single cognate Cygb gene in the then-fragmentary genome sequence data of *D. rerio* [3], which is now named *Cygb-1*. Here, we have identified orthologs of *Cygb-1* via database

analyses in the medaka *O. latipes* as well as in the green spotted pufferfish *T. nigroviridis* and the torafugu *T. rubripes*. Refined database searches in the current genomic and EST sequence data revealed the presence of a second, related *Cygb* gene (*Cygb-2*) in all of these teleost model species.

Based on the in silico predictions, the coding regions of *Cygb-1* were amplified via RT-PCR, cloned, and sequenced from *D. rerio*, *T. nigroviridis*, and *O. latipes*. These coding sequences range from 525 bp (zebrafish) to 540 bp (pufferfishes) (Table 1A), corresponding to proteins from 174 to 179 amino acids. The observed length heterogeneities are confined to the N- and C-termini, with only a single indel in the globin fold (Fig. 1). The average sequence identities/similarities between the teleost Cygbs-1, and human, chicken, and *X. tropicalis* Cygb sequences amount to 52/78% (Table 1C). All teleost Cygb-1 genes have the standard exon–intron structure of vertebrate *Hb* and *Mb* genes [25,26], comprising three exons interrupted at the conserved intron positions B12.2 (i.e., between codon positions 2 and 3 of the 12th amino acid of helix B) and G7.0 (Fig. 1). The C-terminal intron at HC11.2 and a small fourth exon, hitherto found only in mammalian *Cygb* genes, are absent in all fish Cygb-1 genes. As noted earlier, this fourth C-terminal exon probably represents an evolutionary innovation of mammals [3].

The *Cygb-2* coding sequences of zebrafish, pufferfish, and medaka were isolated as described above. The coding regions comprise between 540 and 591 bp, and the deduced proteins cover 179–196 amino acids (Table 1A). On average, the Cygb-2 proteins display about 53/81% amino acid identity/similarity to their Cygb-1 paralogs in the respective species (Table 1C). Relative to human Cygb, teleost Cygb-2 sequences are 66% identical and 85% similar, showing that fish Cygb-2 is somewhat more closely related to mammalian Cygb than fish Cygb-1. Like the Cygb-1 sequences, all teleost Cygb-2 proteins display the conserved sequence hallmarks of functional globins, including the distal and proximal histidines at helix positions E7 and F8, and the phenylalanine residue at position CD1 (Fig. 1). Unlike most other globins, mammalian Cygbs contain conserved cysteines, positioned at B2 and E9, which may create a disulfide bridge in vitro [6]. Formation of this disulfide bridge under oxidizing conditions increases the O_2 affinity of mammalian Cygb. However, these cysteines are only conserved in TniCygb-2 and TruCygb-2, whereas in the other teleost Cygbs at least one cysteine is replaced by another amino acid (Fig. 1). Thus, it is rather unlikely that teleost Cygbs display a similar dependence of the O_2 affinity on the redox state like mammalian Cygbs.

Notably, the Cygb-2 sequences of medaka and pufferfish are substantially longer than Cygb-1. The increased length of Cygb-2 versus Cygb-1 results exclusively from N-terminal protein extensions, which are rich in prolines. Like Cygb-1, all Cygb-2 proteins lack the mammalian-type C-terminal exon extension and feature the typical three exon–two intron structure with the B12.2 and G7.0 introns. We note that the N-terminal extension is absent in zebrafish Cygb-2, resulting in a protein of only 179 amino acids. Interestingly, the nucleotide sequence upstream of the Cygb-2 start codon in the *D. rerio* genome can be conceptionally translated into a proline-rich amino acid sequence, and a CTG motif is found in place of the usual ATG start codon in the Cygb-2 genes of the other species. We therefore speculate that Cygb-2 was shortened at its N-terminus by a mutation in the *D. rerio* lineage. The functional implications of the shortened Cygb-2 version in zebrafish remain to be investigated. It is also conceivable, however, that such terminal length differences are selectively neutral, since a recombinant mouse Cygb without extensions displayed essentially equivalent ligand binding characteristics like the longer wild-type version [6].

Molecular evolution of the teleost Cygb paralogs

The Cygb-1 and -2 amino acid sequences of *D. rerio*, *O. latipes*, *T. nigroviridis*, and *T. rubripes* were included in an alignment of selected vertebrate globins. Phylogenetic trees were constructed employing the Bayesian approach (Fig. 2) and the neighbor-joining distance method, which yielded equivalent results (not shown). A simplified tree, designed for better overview, was prepared by averaging over external branch lengths (Fig. 2A). As previously proposed [20], it shows the presence of three globin lineages in vertebrates, one comprising the neuroglobins (Ngb) and globin X (GbX), one comprising Hbs and Mbs, and another clade comprising all Cygb sequences. The monophyletic grouping of Mb and Cygb (as proposed by [3]) is not reflected in Fig. 2A due to a lack of resolution at the base of the tree. The common evolutionary ancestry of Mb and Cygb has, however, been established by independent genomic data pointing at a gene duplication event [3].

The detailed phylogenetic tree comprising all available Cygb sequences, rooted by Hb and Mb, shows that tetrapod and teleost Cygbs are monophyletic, as expected from the species relationships (Fig. 2B). Paralogy of teleost Cygb-1 and -2 is well supported by the tree. Branch lengths for Cygb-1 tend to be longer than for Cygb-2, in line with a faster rate of sequence evolution of the Cygb-1 paralogs (see below). At the species level, the two related pufferfishes group with the medaka at the exclusion of the zebrafish, in agreement with phylogenomic data [27].

Assuming that Actinopterygii and Sarcopterygii diverged about 420 MYA [28], mean substitution rates of 0.84×10^{-9} and 0.61×10^{-9} amino acid replacements per site per year were calculated for teleost Cygb-1 and Cygb-2, respectively (Table 1B). These evolutionary rates are higher than those observed for chicken (0.43×10^{-9}) and for mammalian Cygbs (0.3×10^{-9}) [13,29], but nevertheless indicate a conservative evolution of the teleost Cygb paralogs. Accordingly, the calculated ratios of nonsynonymous (dn) to synonymous (ds) nucleotide substitutions

Table 1
Sequence features, distance comparisons, and molecular evolutionary rates of teleost Cygbs

	bp	aa	kDa
(A)			
Dre Cygb-1	525	174	19.9
Ola Cygb-1	534	176	20.0
Tni Cygb-1	540	179	20.1
Tru Cygb-1	540	179	19.9
Dre Cygb-2	540	179	20.6
Ola Cygb-2	585	194	22.0
Tni Cygb-2	573	190	21.8
Tru Cygb-2	591	196	22.3

	Amino acid substitution rates ($\times 10^{-9}$)	Nucleotide substitutions		
		dn ($\times^{-9}$)	ds ($\times 10^{3}$)	dn/ds
(B)				
Teleost Cygb-1: Hsa Cygb	0.84	0.48	1.31	0.37
Teleost Cygb-1: Gga Cygb	0.87	0.49	1.66	0.30
Teleost Cygb-1: Xtr Cygb	0.88	0.49	2.18	0.23
Teleost Cygb-2: Hsa Cygb	0.61	0.36	1.65	0.22
Teleost Cygb-2: Gga Cygb	0.69	0.37	1.32	0.28
Teleost Cygb-2: Xtr Cygb	0.69	0.39	2.71	0.14

	Dre Cygb-1	Ola Cygb-1	Tni Cygb-1	Tru Cygb-1	Dre Cygb-2	Ola Cygb-2	Tni Cygb-2	Tru Cygb-2	Hsa Cygb	Gga Cygb	Xtr Cygb
(C)											
Dre Cygb-1		52/83	51/79	50/77	52/83	47/81	46/81	46/80	49/78	45/79	48/78
Ola Cygb-1	62		71/93	70/90	60/84	56/83	55/81	56/81	55/79	55/82	51/81
Tni Cygb-1	63	76		84/96	55/82	51/80	53/79	49/78	53/78	50/82	50/82
Tru Cygb-1	62	75	91		56/80	51/79	53/78	51/78	51/76	50/80	48/80
Dre Cygb-2	63	66	66	64		74/88	72/88	72/88	67/84	64/86	64/86
Ola Cygb-2	62	68	67	66	71		80/93	81/91	66/86	64/82	59/82
Tni Cygb-2	60	69	67	66	71	82		87/98	66/86	58/82	57/82
Tru Cygb-2	61	70	65	66	72	83	88		65/85	60/81	56/80
Hsa Cygb	62	64	64	63	67	72	72	72		73/91	69/87
Gga Cygb	61	63	63	63	66	68	67	68	77		73/91
Xtr Cygb	59	61	63	61	66	66	63	62	71	74	

(A) Gives the lengths of the CDS from teleost *Cygb-1* and *Cygb-2*, and the deduced amino acids as well as the computed protein molecular masses. The amino acid evolution distance data, shown in (B), were estimated according to the PAM model [22], assuming that the lineages leading to Actinoterygii and Sarcopterygii separated 420 MYA [28]. Corrected nonsynonymous (dn) and synonymous (ds) nucleotide substitutions per site were calculated by the method of Nei and Gojobori [24]. All rates are given as changes per site per year. Percent values of the nucleotide identities (lower left) as well as the amino acid identities and similarities (upper right) of teleost Cygb-1 and -2 genes and Cygb of human (Hsa), chicken (Gga), and *X. tropicalis* (Xtr) are given in (C).

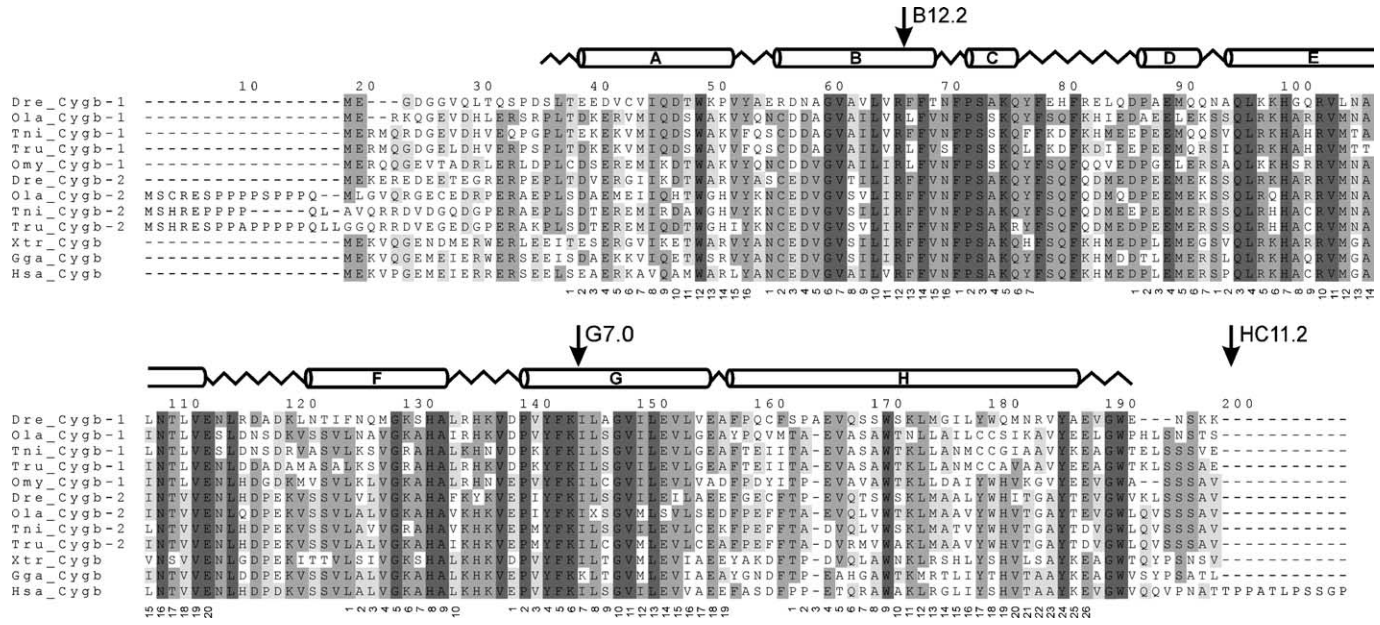

Fig. 1. Amino acid sequence alignment of the Cygb-1 and -2 paralogs from *D. rerio* (Dre), *O. latipes* (Ola), *T. nigroviridis* (Tni), and *T. rubripes* (Tru) with the orthologous single-copy Cygbs from *X. tropicalis* (Xtr), *G. gallus* (Gga), and *H. sapiens* (Hsa). The database accession numbers are given in Materials and methods. Conserved amino acids are shaded and intron positions interrupting the gene coding sequence are indicated by arrows. The typical globin α-helices (according to mammalian Mb as standard) are indicated graphically above the alignment and are designated A–H. The consensus numbering of positions according to the vertebrate globin fold is given below the alignment.

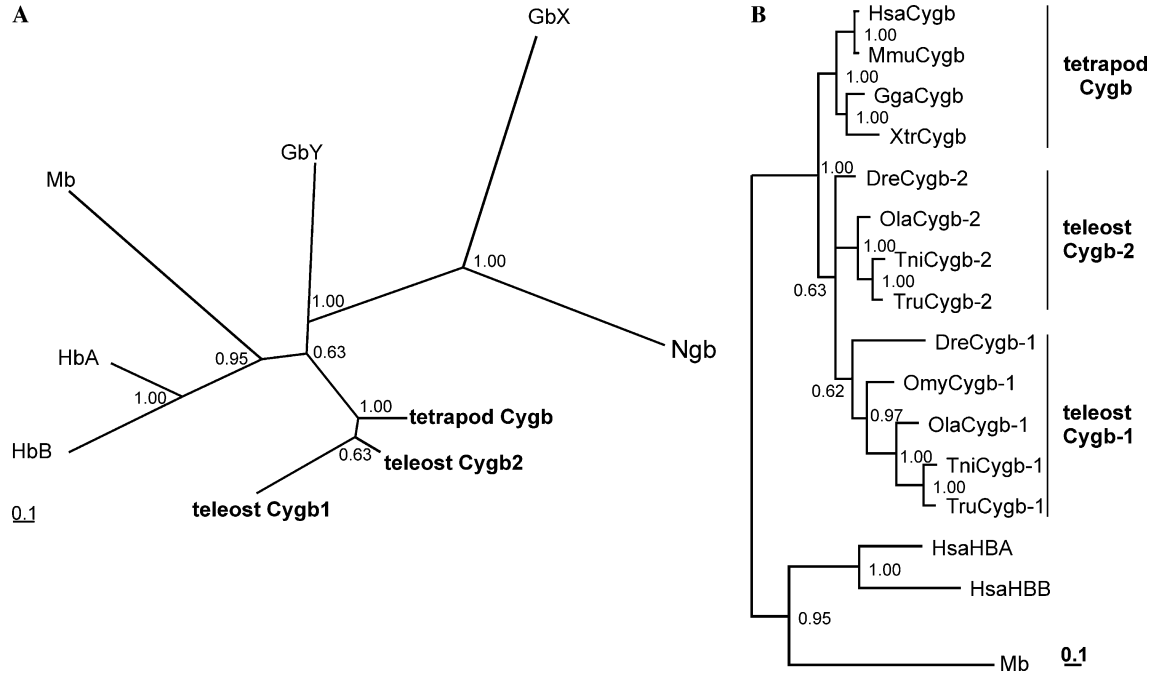

Fig. 2. Condensed phylogenetic tree clarifying the relationships of vertebrate Cygbs relative to other vertebrate globins (A) and a detailed tree, pinpointing the phylogenetic relationships of teleost Cygb-1 and -2 compared to tetrapod Cygbs (B). Trees have been constructed by MrBayes using the PAM model of amino acid evolution. The bar equals 0.1 PAM distance, and the numbers at the nodes represent posterior probabilities. See Materials and methods for protein names and their abbreviations.

indicate that both, Cygb-1 and Cygb-2, evolve under purifying selection in the teleosts (Table 1B). Analysis of a large *T. nigroviridis* protein dataset has revealed that the faster sequence evolution in pufferfish compared to mammalian proteins is a general feature, which may be explained by a higher rate of neutral mutations [16].

Genomic synteny analysis

The genomic organization of the *Cygb* gene region from human chromosome 17q25.3 [3] and the two paralogous Cygb regions in the genome sequence from *T. nigroviridis* were compared in order to study the degree of syntenic conservation and possible large-scale relationships between the *Cygb-1* and *-2* gene regions. After initial identification of *Cygb-1* and *-2* by searching the *T. nigroviridis* genome browser, the anonymous predicted genes neighboring *Cygb-1* and *-2* were annotated by additional protein database searches and named according to their putative human orthologs (Fig. 3). This analysis was not equally feasible in the other teleosts due to the fragmentary status of their genome sequences.

The human–pufferfish genomic comparison revealed syntenic conservation, minimally extending over about 780 kb in the human sequence. The corresponding *T. nigroviridis* genomic regions were substantially shorter, as expected from the roughly 9-fold smaller genome size of pufferfishes compared to mammals [30]. Notably, both paralogous teleost *Cygb* gene regions revealed such syntenic conservation, with 6 (*Cygb-1*) and 8 (*Cygb-2*) genes matching the human genome sequence. This strongly suggests that the teleost *Cygb-1* and *-2* copies did not arise by a single-gene duplication event, but by a large-scale genomic duplication encompassing many neighboring genes. This genomic duplication was obviously followed by a differential loss of genes from the paralogous regions and rearrangements of gene order and transcriptional orientation (Fig. 3).

Tissue expression patterns of Cygb-1 and Cygb-2

To investigate a possible tissue-specific expression of the teleost *Cygb* gene paralogs, qRT-PCR analysis was performed in *D. rerio*. As expected from the mammalian *Cygb* expression data [2,3], teleost *Cygb* sequences displayed a broad expression profile with signals present in all tissues analyzed. *Cygb-1* mRNA was detected at highest levels in

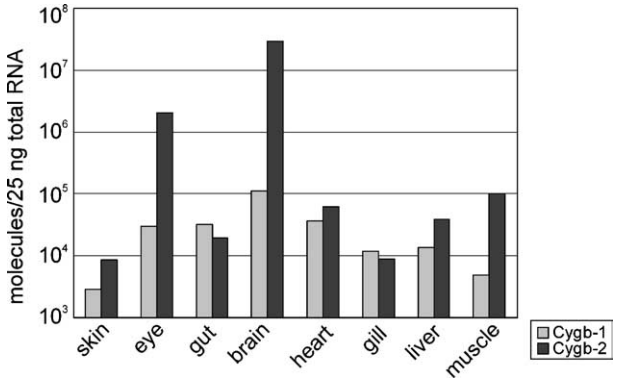

Fig. 4. *Cygb-1* and *Cygb-2* mRNA expression in *D. rerio* tissues, as measured by real-time quantitative RT-PCR. Both genes are expressed in all tissues analyzed, but *Cygb-2* reveals a pronounced expressional preference for brain and eye tissue.

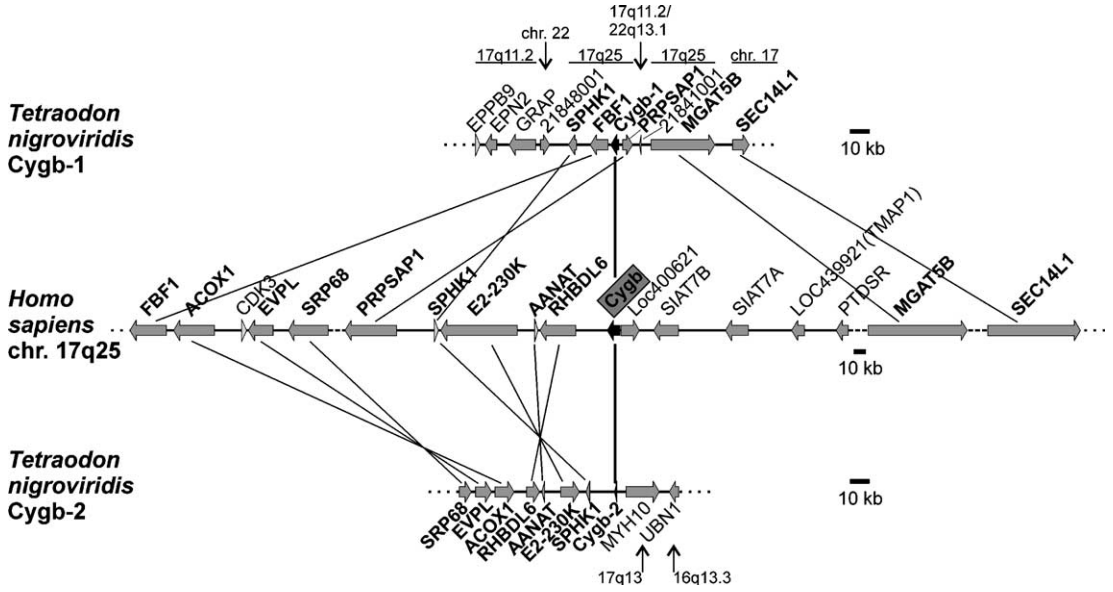

Fig. 3. Comparative genomic analysis of teleost *Cygb-1* and *Cygb-2* gene regions and the *Cygb* gene region in human on chromosome 17q25.3. The positions and transcriptional orientations of the genes are indicated by arrows. Horizontal solid lines indicate intergenic regions drawn to scale, while broken lines identify condensed intergenic regions. Dots at the very beginning and end of a genomic region denote that the database information is continuing beyond this point, but has not been studied further. Vertical solid lines indicate orthologous gene relationships predicted by database searching. Bold types indicate orthologous genes, while thin letters indicate genes present only in one species. All genes are termed according to their human orthologs. Complete gene names can be found in the NCBI Online Mendelian Inheritance in Man database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM). Predicted pufferfish genes, which did not yield a recognizable human ortholog, were labeled according to their Genoscope number.

brain, eye, gut, and heart (Fig. 4). *Cygb-2* was also widely expressed, but revealed a much more focused quantitative profile, with predominant expression in neural tissues like brain and eye. Comparison of transcript levels showed that *Cygb-2* was substantially stronger expressed than *Cygb-1* in nearly all tissues, except for gut and gills (Fig. 4). In brain and eye, *Cygb-2* was expressed about 300-fold and 250-fold stronger than *Cygb-1*. All other tissues showed less pronounced differences. Comparable expression patterns have been obtained by qualitative RT-PCR for tissues of a second teleost species, *T. nigroviridis* (data not shown).

Conclusions

Unlike mammals, birds, and frogs, each of the four teleost genomes studied harbors two paralogous gene copies of *Cygb*. These paralogs encode putatively functional respiratory proteins, which display substantial amino acid sequence differences and a quantitatively different tissue expression pattern. Genomic analysis shows that the teleost *Cygb* duplicates have arisen by a large-scale genomic duplication event. This finding is in agreement with a proposed whole-genome duplication, which is thought to have occurred about 320–350 MYA in the ancestor of ray-finned fishes [16,31,33]. Molecular clock calculations date the divergence of Cygb-1 and -2 approximately at that time range (data not shown).

While most gene duplicates are rendered nonfunctional by mutation or are even eliminated from the genome following such a whole-genome duplication, those paralogs that 'survive' can undergo different evolutionary fates. As proposed by Ohno [34], one of the two duplicates may be free from selective constraint to accumulate mutations and thus evolve novel functional properties. As an alternative to neo-functionalization, both paralogs can be stabilized by partitioning of subfunctions [32,33]. Based on the different tissue expression preferences, with *Cygb-2* clearly dominating in neural tissue, it is tempting to assume such a subfunctionalization scenario also for *Cygb-1* and *Cygb-2*. In mammals, the single-copy *Cygb* probably exerts different, yet undefined, cellular functions (i) in the cytoplasm and the nucleus of neural cell types and (ii) in the fibroblast cell lineage, where it occurs strictly cytoplasmatically [7,9–11]. It seems well conceivable that these two functions have partitioned in the teleost *Cygb* paralogs. Noteworthy, the 'neural' *Cygb-2* displays a markedly less amount of sequence change in evolution. One may speculate that this higher sequence conservation alludes to a function involving a more strictly controlled Cygb protein structure and possible interacting proteins. Fish *Cygb* paralogs will thus be valuable as a model by allowing for dissecting the function(s) of mammalian Cygb.

Acknowledgments

We thank Professor Dr. Manfred Schartl (Department of Physiological Chemistry I, University of Würzburg) for the generous gift of *O. latipes* animals, and we thank Valeska Heib for contributing the *O. mykiss* Cygb sequence. The dedicated efforts by the *Tetraodon* sequencing group at the French National Sequencing Center Genoscope (Evry cedex, France) and the *Takifugu* sequencing groups at the Institute of Molecular and Cell Biology (Singapore) and the Joint Genome Institute (Walnut Creek, California, USA) are gratefully acknowledged. This work has been supported by the Deutsche Forschungsgemeinschaft (Bu956/5 and Ha2103/3), the European Union (QLG3-CT-2002-01548), and the Stiftung für Innovation Rheinland-Pfalz.

References

[1] N. Kawada, D.B. Kristensen, K. Asahina, K. Nakatani, Y. Minamiyama, S. Seki, K. Yoshizato, Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells, J. Biol. Chem. 276 (2001) 25318–25323.

[2] J.T. Trent, M.S. Hargrove III, A ubiquitously expressed human hexacoordinate hemoglobin, J. Biol. Chem. 277 (2002) 19538–19545.

[3] T. Burmester, B. Ebner, B. Weich, T. Hankeln, Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues, Mol. Biol. Evol. 19 (2002) 416–421.

[4] D. de Sanctis, S. Dewilde, A. Pesce, P. Ascenzi, T. Burmester, T. Hankeln, L. Moens, M. Bolognesi, New insight into the haemoglobin superfamily: preliminary crystallo-graphic characterization of human cytoglobin, Acta Crystallogr. D Biol. Crystallogr. 59 (2003) 1285–1287.

[5] H. Sugimoto, M. Makino, H. Sawai, N. Kawada, K. Yoshizato, Y. Shiro, Structural basis of human cytoglobin for ligand binding, J. Mol. Biol. 339 (2004) 873–885.

[6] D. Hamdane, L. Kiger, S. Dewilde, B.N. Green, A. Pesce, J. Uzan, T. Burmester, T. Hankeln, M. Bolognesi, L. Moens, M.C. Marden, The redox state of the cell regulates the ligand binding affinity of human neuroglobin and cytoglobin, J. Biol. Chem. 278 (2003) 51713–51721.

[7] T. Hankeln, S. Wystub, T. Laufs, M. Schmidt, F. Gerlach, S. Saaler-Reinhardt, S. Reuss, T. Burmester, The cellular and subcellular localization of neuroglobin and cytoglobin—a clue to their function? IUBMB Life 56 (2004) 671–679.

[8] K. Nakatani, H. Okuyama, Y. Shimahara, S. Saeki, D.H. Kim, Y. Nakajima, S. Seki, N. Kawada, K. Yoshizato, Cytoglobin/STAP, its unique localization in splanchnic fibroblast-like cells and function in organ fibrogenesis, Lab. Invest. 84 (2004) 91–101.

[9] M. Schmidt, F. Gerlach, A. Avivi, L. Laufs, S. Wystub, J.C. Simpson, E. Nevo, S. Saaler-Reinhardt, S. Reuss, T. Hankeln, T. Burmester, Cytoglobin is a respiratory protein expressed in connective tissue and neurons that is up-regulated by hypoxia, J. Biol. Chem. 279 (2004) 8063–8069.

[10] M. Schmidt, T. Laufs, S. Reuss, T. Hankeln, T. Burmester, Divergent distribution of cytoglobin and neuroglobin in the murine eye, Neurosci. Lett. 374 (2005) 207–211.

[11] T. Hankeln, B. Ebner, C. Fuchs, F. Gerlach, M. Haberkamp, T.L. Laufs, A. Roesner, M. Schmidt, B. Weich, S. Wystub, S. Saaler-Reinhardt, S. Reuss, M. Bolognesi, D. de Sanctis, M.C. Marden, L. Kiger, L. Moens, S. Dewilde, E. Nevo, A. Avivi, R.E. Weber, A. Fago, T. Burmester, Neuroglobin and cytoglobin in search of their role in the vertebrate globin family, J. Inorg. Biochem. 99 (2005) 110–119.

[12] T. Burmester, M. Haberkamp, S. Mitz, A. Roesner, M. Schmidt, B. Ebner, F. Gerlach, C. Fuchs, T. Hankeln, Neuroglobin and cytoglobin: genes, proteins and evolution, IUBMB Life 56 (2004) 703–707.

[13] D. Kugelstadt, M. Haberkamp, T. Hankeln, T. Burmester, Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken, Biochem. Biophys. Res. Commun. 325 (2004) 719–725.

[14] C. Fuchs, T. Burmester, T. Hankeln, The Amphibian globin gene repertoire as revealed by the *Xenopus* genome, Cytogenet. Genome Res., in press.
[15] K. Vandepoele, W. De Vos, J.S. Taylor, A. Meyer, Y. Van de Peer, Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and vertebrates, Proc. Natl. Acad. Sci. USA 101 (2004) 1638–1643.
[16] O. Jaillon, J.M. Aury, F. Brunet, J.L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biemont, Z. Skalli, L. Cattolico, J. Poulain, V. De Berardinis, C. Cruaud, S. Duprat, P. Brottier, J.P. Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K.J. McKernan, P. McEwan, S. Bosak, M. Kellis, J.N. Volff, R. Guigo, M.C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quetier, W. Saurin, C. Scarpelli, P. Wincker, E.S. Lander, J. Weissenbach, H. Roest Crollius, Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype, Nature 431 (2004) 946–957.
[17] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[18] K.B. Nicholas, H.B. Nicholas Jr., GeneDoc: analysis and visualization of genetic variation, http://www.psc.edu/biomed/genedoc/ (1997).
[19] D. de Sanctis, S. Dewilde, A. Pesce, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, M. Bolognesi, Crystal structure of cytoglobin: the fourth globin type discovered in man displays heme hexacoordination, J. Mol. Biol. 336 (2004) 917–927.
[20] A. Roesner, C. Fuchs, T. Hankeln, T. Burmester, A globin gene of ancient evolutionary origin in lower vertebrates: Evidence for two distinct globin families in animals, Mol. Biol. Evol. 21 (2005) 12–20.
[21] J.P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics 17 (2001) 754–755.
[22] M.O. Dayhoff, B.C. Orcutt, Methods for identifying proteins by using partial sequences, Proc. Natl. Acad. Sci. USA 76 (1979) 2170–2174.
[23] S. Kumar, K. Tamura, M. Nei, MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment, Brief. Bioinform. 5 (2004) 150–163.
[24] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (1986) 418–426.
[25] R.E. Dickerson, I. Geis, Hemoglobin: Structure, Function, Evolution, and Pathology, Benjamin/Cummings Publ., California, 1983.
[26] R.C. Hardison, A brief history of hemoglobins: plant, animal, protist, and bacteria, Proc. Natl. Acad. Sci. USA 93 (1996) 5675–5679.
[27] W.J. Chen, G. Orti, A. Meyer, Novel evolutionary relationship among four fish model systems, Trends Genet. 9 (2004) 424–431.
[28] M.J. Benton, Vertebrate Paleontology, Unwin Hyman, London, 1990.
[29] S. Wystub, B. Ebner, C. Fuchs, B. Weich, T. Burmester, T. Hankeln, Interspecies comparison of neuroglobin, cytoglobin and myoglobin: sequence evolution and candidate regulatory elements, Cytogenet. Genome Res. 105 (2004) 65–78.
[30] S. Aparicio, J. Chapman, E. Stupka, N. Putnam, J.M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M.D. Gelpke, J. Roach, T. Oh, I.Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S.F. Smith, M.S. Clark, Y.J. Edwards, N. Doggett, A. Zharkikh, S.V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y.H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, S. Brenner, Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*, Science 297 (2002) 1301–1310.
[31] A. Christoffels, E.G. Koh, J.M. Chia, S. Brenner, S. Aparicio, B. Venkatesh, Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes, Mol. Biol. Evol. 6 (2004) 1146–1151.
[32] A. Force, M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, J. Postlethwait, Preservation of duplicate genes by complementary degenerative mutations, Genetics 151 (1999) 1531–1545.
[33] J. Postlethwait, A. Amores, W. Cresko, A. Singer, Y.L. Yan, Subfunction partitioning, the teleost radiation and the annotation of the human genome, Trends Genet. 20 (2004) 481–490.
[34] S. Ohno, Evolution by Gene Duplication, Springer, New York, 1970.