Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochimica et Biophysica Acta 1727 (2005) 65–74
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
http://www.elsevier.com/locate/bba

Fish possess multiple copies of *fgfrl1*, the gene for a novel FGF receptor

Beat Trueb[a,*], Stephan C.F. Neuhauss[b], Stefan Baertschi[a], Thorsten Rieckmann[a], Christof Schild[a], Sara Taeschler[a]

[a]*ITI Research Institute, University of Bern, Murtenstrasse 35, CH-3010 Bern, Switzerland*
[b]*Swiss Federal Institute of Technology, Department of Biology, and Brain Research Institute of the University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*

Received 30 August 2004; received in revised form 17 November 2004; accepted 6 December 2004
Available online 23 December 2004

Abstract

FGFRL1 is a novel FGF receptor that lacks the intracellular tyrosine kinase domain. While mammals, including man and mouse, possess a single copy of the FGFRL1 gene, fish have at least two copies, *fgfrl1a* and *fgfrl1b*. In zebrafish, both genes are located on chromosome 14, separated by about 10 cM. The two genes show a similar expression pattern in several zebrafish tissues, although the expression of *fgfrl1b* appears to be weaker than that of *fgfrl1a*. A clear difference is observed in the ovary of *Fugu rubripes*, which expresses *fgfrl1a* but not *fgfrl1b*. It is therefore possible that subfunctionalization has played a role in maintaining the two *fgfrl1* genes during the evolution of fish. In human beings, the FGFRL1 gene is located on chromosome 4, adjacent to the SPON2, CTBP1 and MEAEA genes. These genes are also found adjacent to the *fgfrl1a* gene of *Fugu*, suggesting that FGFRL1, SPON2, CTBP1 and MEAEA were preserved as a coherent block during the evolution of *Fugu* and man.

Keywords: *Danio rerio*; Fibroblast growth factor receptor, FGFR; *Fugu rubripes*; Whole-genome duplication; Zebrafish

1. Introduction

The FGF signaling system controls the proliferation, differentiation and migration of many different cell types [1–3]. FGF signaling is therefore involved in a multitude of biological processes, including development, organogenesis, angiogenesis, wound healing and oncogenesis [4]. So far, 24 different FGF ligands have been described that bind to a total of 4 different FGF receptors termed FGFR1–FGFR4.

We have recently discovered a fifth member of the FGFR family that we have named FGF receptor-like 1 (FGFRL1) [5–7]. The cDNA for this protein was isolated from a subtracted, cartilage-specific cDNA library. Independently, the novel receptor has also been discovered by two other research groups and termed FGFR5 [8,9]. Similar to all FGFRs, the novel receptor contains a signal peptide, three extracellular Ig-like domains, the first being separated from the second by a hydrophilic box (acidic box), and a transmembrane domain. In contrast to the other FGFRs, it is lacking the intracellular tyrosine kinase domain, but instead contains an unrelated short tail at the C-terminus that is rich in histidine and threonine residues. The gene for FGFRL1 is located on human chromosome 4 in region 4p16, in close proximity to the gene for FGFR3 [5]. It comprises 6 exons, and each of these exons codes for an entire, functional domain (signal peptide, three Ig-like domains, acidic box, transmembrane domain). Northern blotting and in situ hybridization experiments demonstrated that the novel receptor is preferentially expressed in musculoskeletal tissues, but low expression is also found in many other tissues [5–9]. Experiments with green fluorescent protein revealed that FGFRL1 is located at the plasma membrane, where it could interact with FGF ligands [7]. In vitro, recombinant FGFRL1 binds FGF2 and interacts with heparin.

Abbreviations: EST, expressed sequence tag; FGF, fibroblast growth factor; FGFRL, fibroblast growth factor receptor-like; LG, linkage group; PCR, polymerase chain reaction; UTR, untranslated region

* Corresponding author. Tel.: +41 31 632 87 26; fax +41 31 632 49 63.
E-mail address: beat.trueb@iti.unibe.ch (B. Trueb).

0167-4781/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbaexp.2004.12.001

Based on the absence of the intracellular tyrosine kinase domain, we have speculated that the novel receptor does not signal by itself but rather modulates the function of other FGFRs similar to a decoy receptor, which binds ligands but does not pass on the signal. Consistent with this idea, we found that FGFRL1 exerts an inhibitory effect on cell proliferation when overexpressed in MG63 osteosarcoma cells [7].

To learn more about the putative function of the novel receptor, we analyzed what animals might possess a homologous gene. The gene for FGFRL1was identified in man, mouse and chicken, but no similar gene was found in the roundworm *C. elegans* or the fruit fly *Drosophila* [7]. It is therefore possible that the FGFRL1 gene evolved together with the vertebrates. In this study, we investigated whether the most divergent class of the vertebrates compared to humans, the fish, would also contain a related gene.

2. Materials and methods

2.1. EST clones

EST clones were obtained from the German Resource Center for Genome Research RZPD in Berlin (https://www.rzpd.de/), from the MRC Geneservice in Cambridge (http://www.hgmp.mrc.ac.uk/geneservice/) and from Dr. J. Peng, Institute of Molecular and Cell Biology, Laboratory of Functional Genomics, 30 Medical Drive, 117609 Singapore [10]. The clones were grown in the original bacteria using LB media containing the required antibiotics. Plasmids were isolated from overnight cultures with the help of QIAprep Spin columns (Qiagen, Hilden, Germany). The nucleotide sequences of all plasmids were determined by the dideoxy chain termination method using a cycle sequencing machine.

2.2. Sequence comparisons

All sequences were analysed with the computer software package of the Genetics Computer Group in Madison, WI (GCG Version 10.2). Database comparisons were performed via the Internet against the GenEMBL Databank (http://www.ncbi.nlm.nih.gov/Genbank/), the blast server of the Sanger Center (http://www.sanger.ac.uk/cgi-bin/blast/) and the IMCB *Fugu* Genome blast server (http://www.fugu-sg.org/).

A phylogenetic analysis of the sequences aligned by PileUp was performed with the Phylogeny Interface Environment (PIE) provided by the Bioinformatics Division of the MRC Rosalind Franklin Centre for Genomics Research (http://www.hgmp.mrc.ac.uk). The pairwise number of synonymous (Ks) and non-synonymous (Ka) substitutions was calculated with the GCG program *Diverge* for the aligned nucleotide sequences. Only the sequences of the extracellular domain (nucleotides 101–1219 of the human sequence) were included since the signal peptides and the intracellular domains did not align properly and even showed a frameshift in the case of rodent FGFRL1 [7]. The distances between the sequences were determined with the Kimura 2-parameter of the GCG program *Distance* using the same nucleotide sequence alignment.

2.3. RNA isolation and Northern blotting

Total RNA was isolated from various fish tissues by the guanidinium isothiocyanate method [11] using the RNeasy Kit of Qiagen GmbH (Hilden, Germany). The RNA (10 μg/lane) was separated on a 1% agarose gel in the presence of 1 M formaldehyde and transferred by vacuum blotting to a Nylon membrane (GeneScreen, NEN, Boston, MA). The blot was hybridized at 42 °C with the cDNA probes described below in a buffer containing 50% formamide [12]. After 24 h, the blot was washed at regular stringency and exposed to BioMax MS film (Eastman Kodak, Rochester, NY). RNA from the ovary of *Fugu rubripes* was obtained from the MRC Geneservice, Babraham, Cambridge.

The probe for *Danio fgfrl1a* corresponded to the *Eco*RI/*Xba*I insert of EST clone 6794698, the probe for *Danio fgfrl1b* to the *Eco*RI/*Eco*RI insert of EST clone 052-H06-2. The probes for *Fugu fgfrl1a* and *fgfrl1b* were prepared by PCR using genomic DNA from *F. rubripes* (MRC Geneservice, Babraham, Cambridge) and the following primers: *Fugu fgfrl1a* upper primer AGCAG-CCCCCAGGCAACATTATC, *Fugu fgfrl1a* lower primer TCAGTCGGGCTGATGC-CTGAGAC, *Fugu fgfrl1b* upper primer CCATATTCTCGCCAGCTTACAAC, *Fugu fgfrl1b* lower primer CGGAGTACCATCTTCATCACCAG. Amplification was performed for 40 cycles with an annealing temperature of 60°C (*fgfrl1a*) and 58°C (*fgfrl1b*), respectively. The PCR products were purified on a 1% agarose gel and subcloned into pBlueSK$^+$ utilizing the internal *Sac*II/*Pst*I sites (*fgfrl1a*) or *Sac*II/*Xho*I sites (*fgfrl1b*). The authenticity of all probes was verified by DNA sequencing. The cDNA inserts were excised from the cloning vectors and labeled by the random primed oligolabeling method with [α^{32}P] dCTP [13].

2.4. Quantitative PCR

Total RNA from different zebrafish tissues was denatured at 65 °C and cooled to room temperature. The RNA (1 μg) was transcribed into first strand cDNA by reverse transcriptase from the Moloney Murine Leukemia Virus (1.5 u/μl), as suggested by the supplier of the enzyme (Stratagene). A mixture of oligo dT$_{(25)}$dG, oligo dT$_{(25)}$dA and oligo dT$_{(25)}$dC served as primer. Following transcription, the enzyme was inactivated by heating to 95 °C. The cDNA was quantified by real time PCR with the ABI 7700 Sequence Detection System utilizing the SYBR Green PCR master mix (Applied Biosystems). The following primer pairs were used: *fgfrl1a* CCAGTCCAGAAAGCACTGCTTCCTCAG/TGTTTGGCCGACTGCAGAAAAGTGTG (annealing temperature 60 °C), *fgfrl1b* TAGTCGAAGACGTCAGAGCGCTTTAC/TCCTGCAGTCCTACAGTACGGTTGTG (annealing temperature 58 °C). All experiments were performed in duplicate, and the cDNA sequence of ribosomal protein S9 was amplified in parallel as a control.

2.5. Radiation hybrid mapping

Radiation mapping was performed with the mouse/zebrafish radiation hybrid panel LN54 [14]. The following primer pairs were used: *fgfrl1a* upper primer CCAGTCCAG-AAAGCACTGCTTCCTCAG, *fgfrl1a* lower primer TGTTT-GGCCGACTGCAGAAAAGTGTG, *fgfrl1b* upper primer TAGTCGAAGACGTCAGAGCGCTTTAC, *fgfrl1b* lower primer TCCTGCAGTCCTACAGTACGGTTGTG. PCR was performed

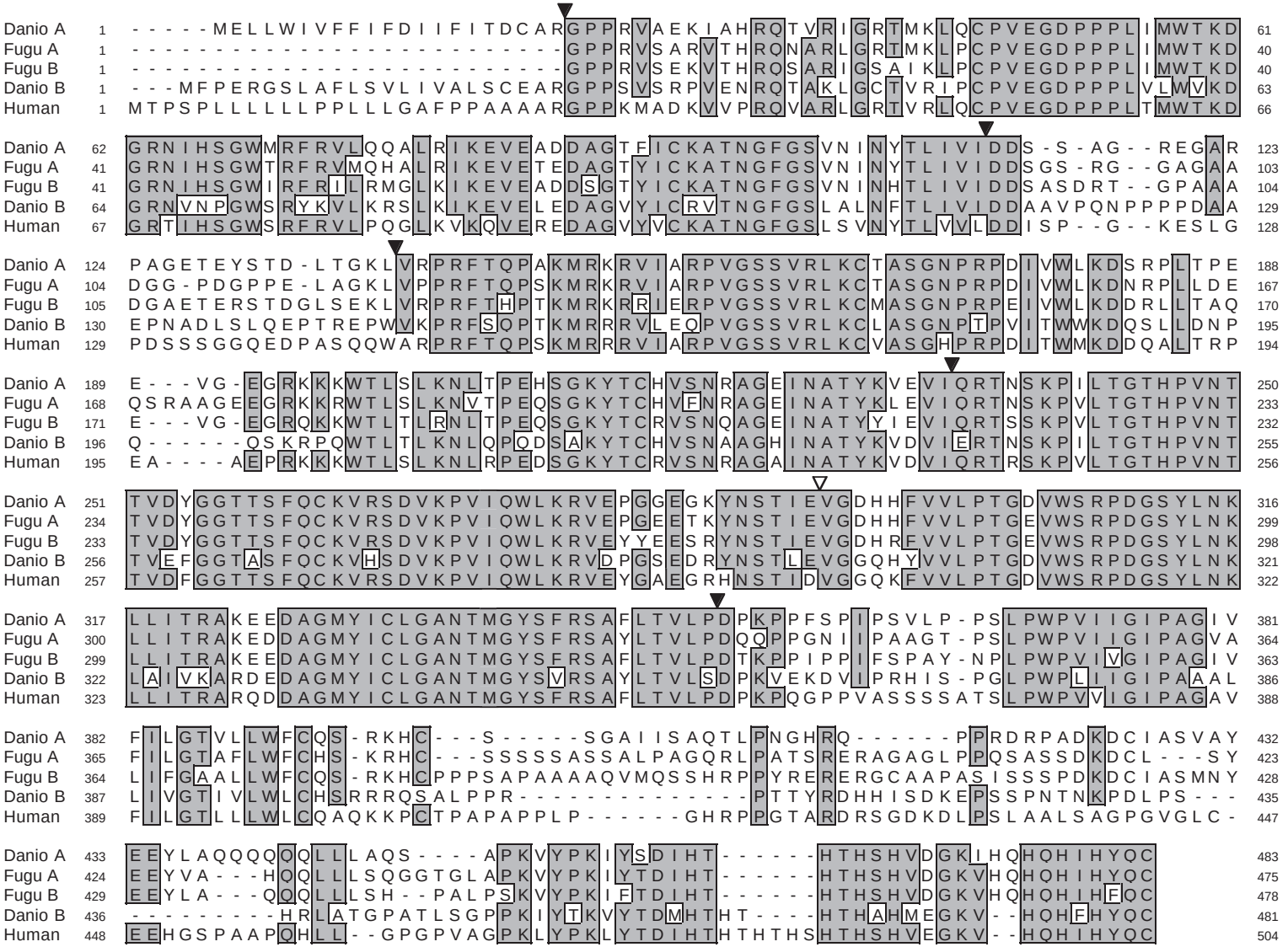

Fig. 1. Alignment of the *fgfrl1* amino acid sequences from zebrafish, pufferfish and man. Identical residues are boxed. The positions of introns in the corresponding genes are marked by black triangles. An open triangle shows the position of intron 5a that occurs only in the *fgfrl1a* gene from *Fugu* (cf. also Table 2). The accession numbers are: human AJ277437, *DanioA* AJ574916, *DanioB* AJ781308, *FuguA* BN000669, *FuguB* BN000670.

for 35 cycles with an annealing temperature of 60 °C (*fgfrl1a*) or 58 °C (*fgfrl1b*). The PCR products were separated on an agarose gel and the data were analyzed with the radiation hybrid mapper computation engine provided on the Internet (http://mgchd1.nichd.nih.gov:8000/zfrh/beta.cgi).

3. Results

3.1. More than one copy of fgfrl1 in Danio rerio

A detailed search of all entries in Genbank was performed with the amino acid sequence of human FGFRL1 as query and the program TBlastN. This search identified 7 EST clones from the zebrafish *D. rerio* that showed 55–82% sequence identity with the human sequence. Five of these clones were ordered from nearby Resource Centers and the complete nucleotide sequences of the clones were determined. An alignment of these sequences demonstrated that the five EST clones formed two separate contigues, representing two different cDNAs. One cDNA was termed *fgfrl1a* (clones 6794698, 4729713), the other *fgfrl1b* (clones 7054853, 052-H06-2, 069-B08-2). The first contigue spanned a total of 2749 nucleotides and contained a typical poly(A) tail at the 3' end (accession number AJ574916). This sequence harbored an open reading frame that could be translated into a polypeptide of 483 amino acid residues (Mr 54 kDa, Fig. 1). The second set covered a total of 3407 nucleotides and contained also a poly(A) tail (accession number AJ781308). This sequence could be translated into a polypeptide of 481 residues (Mr 53 kDa, Fig. 1). At the amino acid level, the two zebrafish sequences shared 57% sequence identity or 68% sequence similarity if conservative replacements were included. Compared to the human FGFRL1 sequence, the two zebrafish sequences showed 68% and 61% identity, respectively. The partial nucleotide sequences of the other EST clones for zebrafish *fgfrl1* that were not characterized in detail in our study could all be assigned to either the *fgfrl1a* or the *fgfrl1b* contigue.

Both zebrafish sequences displayed the typical domain structure of the FGF receptor-like protein with a signal peptide, three Ig-like domains, a hydrophilic linker (acidic box) separating the first and the second Ig-like domains, a transmembrane domain and a relatively short intracellular domain with a peculiar stretch of alternating histidine and threonine residues (Fig. 1). We concluded that zebrafish contain at least two genes for the novel FGF receptor-like protein. Minor sequence variations were noticed between the individual clones of each set, but these variations barely affected the deduced amino acid sequence (see below).

3.2. Expression of fgfrl1a and fgfrl1b in zebrafish tissues

To examine whether the two *fgfrl1* genes were indeed expressed, we utilized the clones for Northern blotting experiments with RNA from zebrafish tissues. Due to the small size of the fish, RNA was prepared from entire parts of the body (head, trunk) and from large organs (ovary, spine, gut) (Fig. 2). A probe specific for *fgfrl1a* hybridized to an mRNA that migrated as a closely spaced doublet in a region corresponding to 3.0–3.2 kb. Assuming a poly(A) tail of 300 nucleotides, this size is in accordance with the length of the cDNA sequence (2749 nt). The mRNA was expressed in the spine, the head and the trunk of adult fish, but barely in the gut or the ovary. A signal was also detected with an RNA preparation from fish larvae of 5 days post fertilization. In general, the signals obtained were very weak such that the blot had to be exposed for several days.

A probe specific for *fgfrl1b* hybridized to a single mRNA that migrated with a mobility corresponding to 3.6 kb (Fig. 2), consistent with the length of the *fgfrl1b* cDNA sequence (3407 nt). The signal was detected in the RNA preparation from fish larvae, but barely in any preparation from adult tissues.

Since the expression levels were very low, our results were verified by quantitative PCR. For this purpose, our RNA preparations were transcribed into first strand cDNA. A primer pair for *fgfrl1a* detected fair amounts of the *fgfrl1a* transcript in the preparations from larvae and from adult head, but low levels in all the other preparations from adult animals (Fig. 3). A primer pair for *fgfrl1b* yielded a similar distribution. The *fgfrl1a* gene was expressed in the zebrafish larvae and in the head, but barely in any other tissue (Fig. 3).

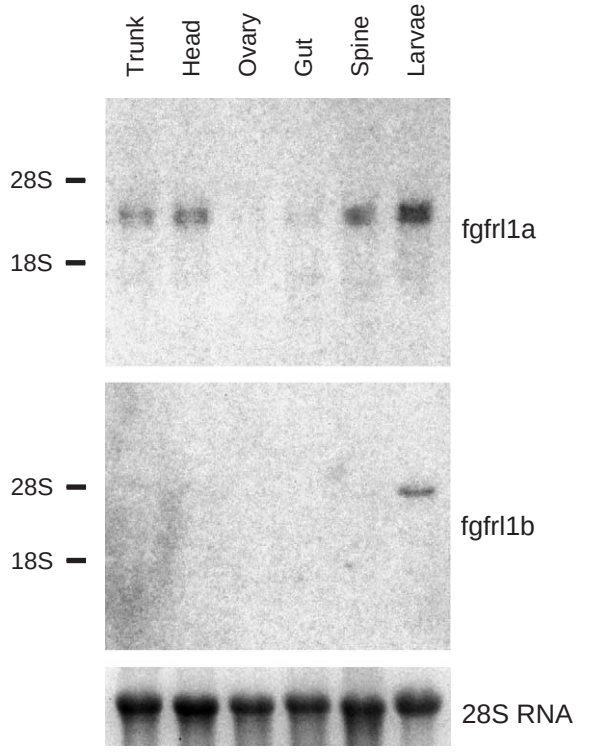

Fig. 2. Expression of the *fgfrl1* genes in zebrafish. Probes specific for *fgfrl1a* and *fgfrl1b* were hybridized to a Northern blot containing RNA from different tissues of adult zebrafish, as well as RNA from zebrafish larvae (5 days post fertilization). The 28S ribosomal RNA stained with ethidium bromide is included as loading control.

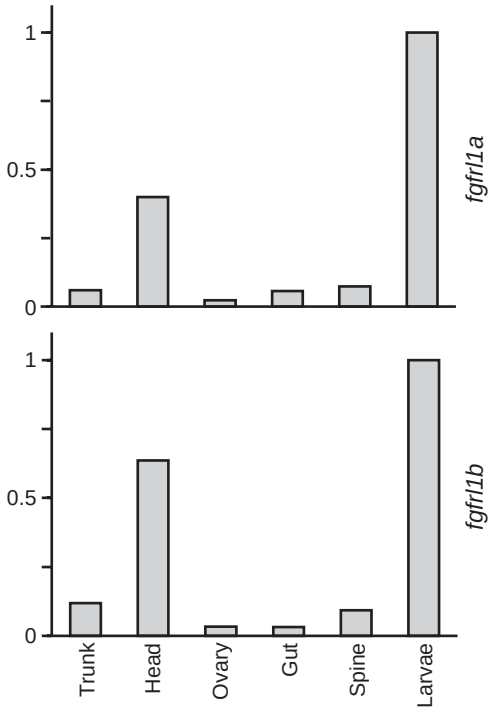

Fig. 3. Quantitative PCR. RNA from different zebrafish tissues and from zebrafish larvae was transcribed into first strand cDNA. The expression of *fgfrl1a* and *fgfrl1b* was quantified by real time PCR using two specific primer pairs. The results are given in relation to the expression level of the *fgfrl1* gene in fish larvae (=1).

These data are in keeping with the results obtained by Northern blotting, except for the expression of *fgfrl1b* in the adult head. Obviously, the mRNA levels for both *fgfrl1* genes are close to the detection limit of Northern blotting in the adult fish. While the expression of *fgfrl1a* can still be demonstrated by this method in some adult tissues, the expression of *fgfrl1b* is below detection level and requires PCR for analysis.

3.3. The structure of the fgfrl1 genes

Since a great deal of the genome from *D. rerio* has been sequenced, we compared the cDNA sequences with the draft of the genomic sequence. Exons corresponding to the 5′ end of *fgfrl1a* (nucleotides 1–432) were found in genomic clones CH211-288O15, CH211-13A5 and CH211-159I8. Clone CH211-159I8 has been sequenced completely and annotated as belonging to linkage group 14. It contained the first two exons of the zebrafish *fgfrl1a* cDNA. Interestingly, the splice sites of these exons matched the splice sites of the exons from the human FGFRL1 gene.

Exons corresponding to the complete *fgfrl1b* sequence were found in genomic clones DKEY-149K8 and RP71-24B12. Curiously enough, the two clones have been annotated as belonging to two different linkage groups, LG 15 and LG 14. One possibility to explain this apparent discrepancy is the existence of two nearly identical genes, *fgfrl1bα* and *fgfrl1bβ*, that would have been generated by a recent gene duplication event. Since there is no evidence for such a recent duplication, we favor another possibility, the existence of two different haplotypes in the original DNA source used for sequencing (see Discussion). To further investigate this possibility, we deduced the exact exon–intron structure of the two genes (Table 1).

fgfrl1bα was found to span ~84 000 nucleotides from the ATG start codon to the polyadenylation signal AATAAA. Exon 1 contained the sequence for the signal peptide, exons 2, 4 and 5 corresponded to the sequences for the three Ig-like domains, exon 3 coded for the hydrophilic linker (acidic box), and exon 6 corresponded to the sequence for the transmembrane segment, the intracellular domain and the 3'UTR (Table 1). Particularly large introns were found between exons 1 and 2 (48 324 bp) and between exons 2 and 3 (26 137 bp). All introns interrupted the codons for the amino acids after the first nucleotide (splice phase 1). The position of the splice sites corresponded exactly to those of the human FGFRL1 gene.

fgfrl1bβ spanned ~78 000 nucleotides. It was also divided into six exons by 5 introns of splice phase 1, and the position of all the introns corresponded to those of the *fgfrl1bα* gene (Table 1). Particularly, large introns were observed between exons 1 and 2 (42 383 bp) and exons 2 and 3 (26 033 bp).

The mRNAs transcribed from the two *fgfrl1b* genes showed 7 substitutions within the open reading frame. All substitutions except one will remain silent since they occur at the third position of the codons (wobble positions). Only one substitution would change asparagine 429 present in the intracellular domain into isoleucine. At least 17 variations including substitutions, deletions and insertions were noted in the 3′ UTR. Furthermore, considerable differences were found within the introns. According to detailed sequence comparisons, EST clones 7054853, 7036193, 069-B08-2 and 7146898 appeared to correspond to the *fgfrl1bα* gene variant, whereas EST clones 052-H06-2 and 7154242 better matched the *fgfrl1bβ* gene variant.

The chromosomal localization of the zebrafish genes was investigated by radiation hybrid mapping using the mouse/zebrafish panel LN54. A primer pair was designed that should specifically amplify a 330 bp fragment from the *fgfrl1a* gene. A second pair was selected in a way that it would specifically amplify a 334 bp fragment from both, the *fgfrl1bα* and the *fgfrl1bβ* gene variants. Using the first primer pair, *fgfrl1a* was mapped to linkage group LG 14, 16.96 cR from marker z6847 (between fe50g09 and fe37e12, LOD score 13.8). With the second primer pair, *fgfrl1b* was mapped to linkage group 14, 51.25 cR from marker z9017 (between z17288 and z21080, LOD score 15.6). Thus, *fgfrl1a* and *fgfrl1b* reside on the same chromosome, separated by about 10 cM. No other linkage was found for *fgfrl1b* (the second best linkage group was LG 17 with a LOD score of 3.6). It is therefore likely that the two gene variants *fgfrl1bα* and *fgfrl1bβ* do in fact

Table 1
Structure of the *fgfrl1b* Genes from *D. rerio*

Exon	Acceptor	Donor	Splice phase	Position on DKEY-149K8
Danio fgfrl1bα				
1		ATGTTT...GCGAGAG gtgagt		50758–50827
		MetPhe...AlaArgG	1	
2	ctccag	GTCCCC...GTAATTG gtaaag		99150–99422
		lyProP...ValIleA	1	
3	tttcag	ATGATG...CCATGGG gtaaat		125558–125650
		spAspA...ProTrpV	1	
4	ctgcag	TGAAGC...GTCATTG gtgagt		127860–128138
		alLysP...ValIleG	1	
5	ccacag	AGCGCA...CTCTCTG gtcagt		130562–130915
		luArgT...LeuSerA	1	
6	ttgcag	ATCCCA...tataaga aataaa		132696–134988
		spProL		

Exon	Acceptor	Donor	Splice phase	Position on RP71-24B12
Danio fgfrl1bβ				
1		ATGTTT...GCGAGAG gtgagt		118609–118540
		MetPhe...AlaArgG	1	
2	ctccag	GTCCCC...GTAATTG gtaaag		76158–75886
		lyProP...ValIleA	1	
3	tttcag	ATGATG...CCGTGGG gtaaat		49854–49762
		spAspA...ProTrpV	1	
4	ctgcag	TGAAGC...GTCATTG gtgagt		47363–47085
		alLysP...ValIleG	1	
5	ccacag	AGCGCA...CTCTCTG gtcagt		44936–44583
		luArgT...LeuSerA	1	
6	ttgcag	ATCCCA...tataaga aataaa		42905–40638
		spProL		

represent two different haplotypes of a highly variable region in the zebrafish genome.

3.4. Two different fgfrl1 genes in F. rubripes

The cDNA sequences for *fgfrl1a* and *fgfrl1b* from *D. rerio* were also compared to all EST clones from the pufferfish *F. rubripes*. Although the database contains nearly 25 000 ESTs from 15 different adult and juvenile *Fugu* tissues [15], no matches were obtained. We therefore compared the cDNA sequences to the draft of the *Fugu* genome sequence. In this way, two scaffolds were identified that appeared to contain the *Fugu fgfrl1a* and *fgfrl1b* genes. No evidence was obtained for the existence of a third homologous gene.

The *Fugu fgfrl1b* gene was found on the minus strand of scaffold M188. The first exon, which would code for the signal peptide, could not unequivocally be identified because signal peptides are not conserved among different species. However, all the other exons could clearly be identified (Table 2). Exons 2, 4 and 5 were found to code for the three Ig-like domains as described above, exon 3 corresponded to the hydrophilic box, and exon 6 contained the information of the transmembrane domain, the intracellular domain and the 3′ UTR. All exons were separated by introns of splice phase 1 as found in the human and the zebrafish genes.

The *Fugu fgfrl1a* gene was identified on the plus strand of scaffold M70. It was split by introns into functional domains as described above (Table 2). However, it contained an additional intron of 99 bp that divided exon 5 into two parts, exons 5a and 5b. The additional intron was of splice phase 0, while all the other introns were of splice phase 1. It is therefore likely that the *fgfrl1a* gene has acquired an additional intron after duplication of the ancestral *fgfrl1* gene.

An open reading frame of 1425 nucleotides corresponding to 475 amino acids could be assembled from the exons of the *Fugu fgfrl1a* gene (accession number BN000669). An open reading frame of 1434 nucleotides corresponding to 478 amino acids was assembled from the exons of the *fgfrl1b* gene (accession number BN000670). The two *Fugu* proteins deduced from these reading frames showed 76% sequence identity among each other or 81% and 60% identity, respectively, to *fgfrl1a* and *fgfrl1b* from zebrafish (cf. Fig. 1).

The surroundings of the *Fugu fgfrl1a* and *fgfrl1b* genes were inspected with the *Fugu*-Human Genome Synteny Viewer [16]. Interestingly, *Fugu* scaffold M70 revealed particularly high synteny to human chromosome 4 region 4p16.3, which also harbors the human FGFRL1 gene. This scaffold contained at least four genes that have orthologues in the human genome at region 4p16.3, namely FGFRL1, SPON2, CTBP1 and MEAEA (Fig. 4). Remarkably, the

Table 2
Structure of the *fgfrl1* genes from *F. rubripes*

Exon	Acceptor	Donor	Phase	Position on M70
Fugu fgfrl1a				
2	ttccag	GTCCGC...GTTATAG gtgcgt		12982–13254
		lyProP...ValIleA	1	
3	ccccag	ACGACT...AAACTGG gtaaca		31426–31503
		spAspS...LysLeuV	1	
4	ccgcag	TGCCTC...GTCATAC gtgagt		34893–35189
		alProP...ValIleG	1	
5a	cttcag	AGCGAA...CATCGAG gtgcac		36741–36913
		lnArgT...rIleGlu	0	
5b	ccaaag	GTCGGG...TTGCCAG gtaagg		37013–37193
		ValGly...LeuProA	1	
6	ttgcag	ATCAGC...ggcagga aataaa		37286–39023
		spGlnG		

Exon	Acceptor	Donor	Phase	Position on M188
Fugu fgfrl1b				
2	cctcag	GACCAC...GTCATTG gtgagt		72432–72160
		lyProP...ValIleA	1	
3	taacag	ATGACT...AAATTGG gtaagt		67638–67552
		spAspS...LysLeuV	1	
4	atcaag	TGCGCC...GTCATAC gtgagt		67349–67065
		alArgP...ValIleG	1	
5	ttgcag	AGAGGA...CTGCCAG gtgaga		66763–66410
		lnArgT...LeuProA	1	
6	ccgcag	ACACAA...ggtgtga aataaa		66315–64946
		spThrL		

order and the orientation of all these genes were identical in the human and in the *Fugu* genome, suggesting that they evolved as a coherent block in man as well as fish. In contrast, *Fugu* scaffold M188 showed the highest synteny to human chromosome 14. There was only one gene (ZFYVE28) that possessed an orthologue in the human genome at region 4p16.3 (Fig. 4). The adjacent genes had orthologues on human chromosomes 3p21 (KNSLP7), 4q28 (APG-1) and 4q27-28 (STK18).

3.5. Expression of the Fugu genes

Hybridization probes specific for the *fgfrl1a* and the *fgfrl1b* genes were prepared by PCR with genomic DNA from *F. rubripes*. The probes corresponded to the open reading frame of exon 6 and to part of the adjacent 3′ UTR. When hybridized to a Northern blot containing total RNA from the ovary of *F. rubripes*, the *fgfrl1a* specific probe yielded a broad band corresponding to a mRNA of 4.8 kb (Fig. 5). This result is in sharp contrast to the findings obtained above where a probe for *fgfrl1a* from zebrafish

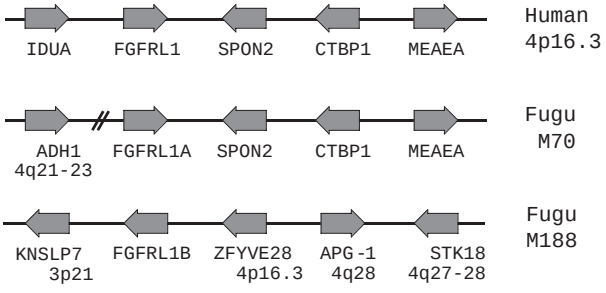

Fig. 4. Analysis of the neighborhood of the FGFRL1 genes in the human and the *Fugu* genome. The human genomic region of FGFRL1 at 4p16.3 was compared to the *Fugu* scaffolds M70 and M188 using the *Fugu*-Human Genome Synteny Viewer. Orthologous regions are marked with the corresponding human gene symbol and locus. ADH1 originates from scaffold M592 that overlaps with scaffold M70.

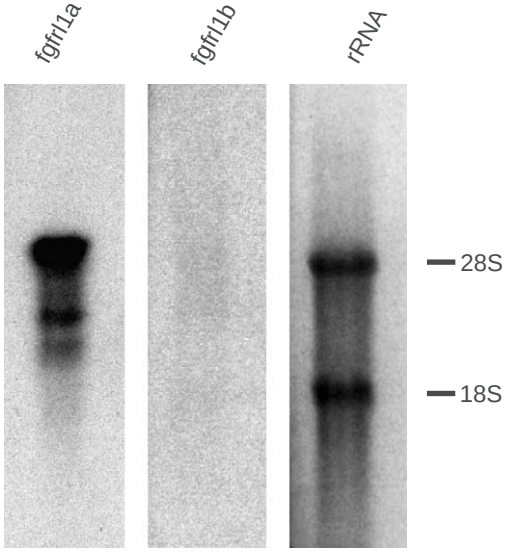

Fig. 5. Expression of the *fgfrl1* genes in *Fugu*. Probes specific for the *fgfrl1a* and *fgfrl1b* genes were hybridized to a Northern blot containing RNA from *Fugu* ovary. The ribosomal RNAs stained with ethidium bromide are included as size markers.

barely produced any signal with RNA from the ovary. Obviously, the expression patterns of the *fgfrl1* genes differ between the two fish species. No signal was obtained with the *fgfrl1b* specific probe from *F. rubripes* and RNA from *Fugu* ovary. Thus, the two *fgfrl1* genes show a clearly distinct spatial distribution in pufferfish. Since no RNA was available from any other *Fugu* source, our experiments could not be extended to other tissues or to other developmental states.

3.6. Phylogenetic analysis

A phylogenetic analysis performed with the two fish sequences confirmed our conclusions drawn above. The ratio of synonymous, silent (Ks) versus nonsynonymous (Ka) substitutions was 7.9 for *fgfrl1a* and *fgfrl1b* from zebrafish and 8.8 for the two *Fugu* genes. These values are similar, for example, to the ratio calculated for the human and chicken FGFRL1 gene (8.2). This result suggests that the two *fgfrl1* genes were subjected to selective pressure and did not evolve as pseudogenes. The distances between the nucleotide sequences of *fgfrl1a* and *fgfrl1b* were determined by the method of Kimura [17]. The Kimura 2-parameters were 0.49 for the two genes from *D. rerio* and 0.33 for those from *F. rubripes*. These values are comparable to the distances calculated between the FGFRL1 gene from man and the two genes from *Danio* (0.35 and 0.48, respectively) or the FGFRL1 gene from man and the two genes from *Fugu* (0.35 and 0.37, respectively). The duplication of the fish genes must therefore be very ancient and might have happened around the origin of teleost fish.

A phylogenetic tree was constructed with all the FGFRL1 sequences available. The tree was built by the neighbor-joining method, and the bootstrap values for distance, maximum parsimony and maximal likelihood were included as measures for reliability (Fig. 6). All mammalian sequences (human, mouse, rat) were found on one branch of this tree. A separate branch contained the avian and the amphibian sequences. The fish sequences were finally situated on a third, complex branch containing the two more closely related *fgfrl1a* sequences and the two more diverged *fgfrl1b* sequences.

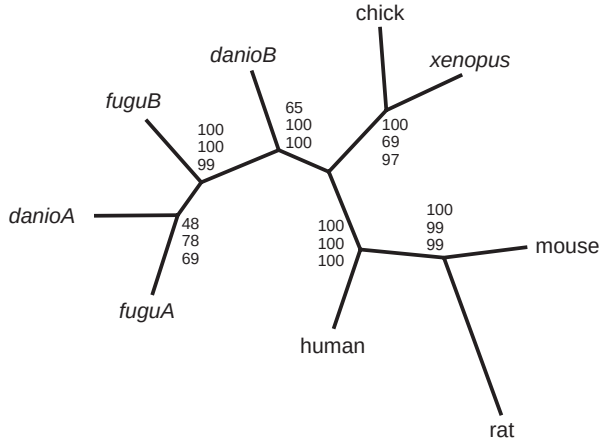

Fig. 6. Phylogenetic analysis of the vertebrate FGFRL1 proteins. An unrooted tree was built by the neighbor joining method, and the bootstrap values for neighbor-joining (top), maximum parsimony (middle) and maximum likelihood (bottom) are included at the nodes. The accession numbers are: human AJ277437, mouse AJ293947, rat AJ536020, *Xenopus* AJ616852, chick AJ535114, *DanioA* AJ574916, *DanioB* AJ781308, *FuguA* BN000669, *FuguB* BN000670.

4. Discussion

A complete genome duplication appears to be the most efficient way to increase the raw genetic material of an organism for further evolution. There is ample evidence that such whole-genome duplications have happened at the origin of yeast [18] and various plants [19]. It has also been noted some time ago that several genes unique in mammals have at least two copies in fish [20,21]. A systematic analysis of the draft genomic sequences from zebrafish and pufferfish provided further evidence for a whole-genome duplication during the evolution of ray-finned fish [22,23]. It was calculated that this genome duplication might have happened about 350 Myr ago, after the separation of ray-finned and lobe-finned fish, but probably before the origin of teleost fish [23]. While parts of the duplicated genes were lost again, another part was maintained during evolution. Preserved genes have adopted slightly different functions and this subfunctionalization might have further protected the genes from being lost [19,24,25].

A whole-genome duplication can explain our findings that *Fugu* and *Danio* have at least two copies of the gene for the novel receptor FGFRL1, whereas human beings and mice possess only one copy. It is possible that the duplicated fish genes were maintained due to subfunctionalization. The *fgfrl1a* and the *fgfrl1b* genes show a distinct expression pattern, at least in *F. rubripes*. In the ovary of this pufferfish, *fgfrl1a* was expressed at a fairly high level, while expression of *fgfrl1b* could not be detected.

The sequences of *fgfrl1a* and *fgfrl1b* from zebrafish have diverged to an extent (57% identity) that is greater than the divergence of human FGFRL1 and fish *fgfrl1a* (68% identity). It is therefore conceivable that the two fish receptors have adopted slightly different functional properties, such as different ligand specificity or different ligand affinity. Detailed comparisons and phylogenetic analyses demonstrate that fish *fgfrl1a* is more closely related to the mammalian gene than fish *fgfrl1b*. This observation is further substantiated by the fact that *Fugu fgfrl1a* is situated on a chromosomal segment (scaffold M70) that appears to be syntenic to human chromosomal region 4p16. Within this region, the order and the orientation of at least 4 genes, including FGFRL1, SPON2, CTBP1 and MEAEA, have been conserved during evolution. In contrast, *Fugu fgfrl1b* is located on a different chromosomal segment (scaffold M188) that does not show much synteny to human

chromosomal region 4p16. In this novel surrounding, *fgfrl1b* might have diverged more rapidly than *fgfrl1a*.

In addition to the existence of the two genes *fgfrl1a* and *fgfrl1b*, which can be explained by a whole-genome duplication, we identified two closely related versions of *fgfrl1b* in zebrafish, termed *fgfrl1bα* and *fgfrl1bβ*. The open reading frame of these gene variants differed only by 17 substitutions, but substantial differences were present in the 3'UTR as well as in the introns. In particular, the *fgfrl1bα* gene was about 6 kb longer than the *fgfrl1bβ* gene. There are two possibilities to explain the existence of these genes. Either the locus of *fgfrl1b* has undergone a recent duplication and the two versions have not yet diverged much except for the introns. So far, there is no evidence in support of a recent duplication of the *fgfrl1b* locus. *Fugu* does not appear to possess two copies of the *fgfrl1b* gene. Furthermore, radiation hybrid mapping in zebrafish yielded only one locus for *fgfrl1a* and one locus for *fgfrl1b*. Both loci are found on chromosome 14, and this assignment is consistent with the annotation of clone CH211-159I8 (harboring *fgfrl1a*) and clone RP71-24B12 (harboring *fgfrl1b*) on LG 14. Another possibility is more likely. The genomic DNA used for large scale sequencing might have contained material from at least two slightly different haplotypes. In fact, the DNA source originated from more than 1000 embryos that were pooled (www.sanger.ac.uk/Projects/D_rerio). Thus, a highly variable region that was sequenced from two different haplotypes could have been misassembled by automated computer alignment, and this false assembly could eventually lead to the assumption that a gene has recently been duplicated. The zebrafish sequencing group has realized this possibility and is currently addressing the issue by modifying the assembly code "Phusion".

We have come across the two paralogues *fgfrl1a* and *fgfrl1b* by sequencing of several EST clones from zebrafish. The completed cDNA sequences allowed us to predict the corresponding genes in the draft genomic sequence of *F. rubripes*. The *fgfrl1a* gene from *Fugu* has also been identified by the ENSEMBLE initiative during automatic annotation. However, the predicted sequence differs from our sequence in the regions of exon 3 and intron 5. Obviously, it was not possible to correctly identify all introns and exons based only on a mathematical algorithm. This fact emphasizes that it is inevitable to compare the genome of an organism to its "transcriptome" in order to unequivocally assign all transcribed regions. It is possible that the splice sites of exon 3 were misinterpreted by automatic annotation because this exon shows a particularly low degree of sequence conservation. Our prediction is supported by the fact that the amino acid sequence deduced from exon 3 contains the characteristic motif Asp-Asp-Ser at the 5′ end and the motif Gly/Glu-Lys-Leu at the 3′ end (cf. Fig. 1). These motifs are conserved among different species although the rest of exon 3 does not show much similarity.

Mammals possess 4 genes for FGF receptors (FGFR1–FGFR4) and one gene for the FGF receptor-like protein FGFRL1. Based on the hypothesis of a whole-genome duplication, one would expect to find up to 8 genes for FGF receptors and two genes for the receptor-like protein in fish. However, a preliminary analysis of the draft genomic sequence from zebrafish suggests that there are only 4–5 genes for FGF receptors. It is therefore likely that several duplicated receptor genes have been lost during evolution, whereas the additional copy of *fgfrl1* was maintained. The maintenance of two *fgfrl1* genes may point to an important function.

Acknowledgements

We thank Annette Bohnert for excellent technical assistance. This study was supported by the Swiss National Science Foundation (grant 31-102251), by Oncosuisse (grant OCS-01463-02-2004), by the Novartis Foundation (grant 04A02) and by the Foundation for Research on Muscular Diseases. SCFN is a Swiss Federal National Science Förderprofessor (NF PP00A-68868).

References

[1] G. Szebenyi, J.F. Fallon, Fibroblast growth factors as multifunctional signaling factors, Int. Rev. Cyt. 185 (1999) 45–106.
[2] C.J. Powers, S.W. McLeskey, A. Wellstein, Fibroblast growth factors, their receptors and signaling, Endocr.-Relat. Cancer 7 (2000) 165–197.
[3] D.M. Ornitz, N. Itoh, Fibroblast growth factors, Genome Biol. 2 (2001) 3005.1–3005.12.
[4] I. McIntosh, G.A. Bellus, E. Wang Jabs, The pleiotropic effects of fibroblast growth factor receptors in mammalian development, Cell Struct. Funct. 25 (2000) 85–96.
[5] M. Wiedemann, B. Trueb, Characterization of a novel protein (FGFRL1) from human cartilage related to FGF receptors, Genomics 69 (2000) 275–279.
[6] M. Wiedemann, B. Trueb, The mouse Fgfrl1 gene coding for a novel FGF receptor-like protein, Biochim. Biophys. Acta 1520 (2001) 247–250.
[7] B. Trueb, L. Zhuang, S. Taeschler, M. Wiedemann, Characterization of FGFRL1, a novel fibroblast growth factor (FGF) receptor preferentially expressed in skeletal tissues, J. Biol. Chem. 278 (2003) 33857–33865.
[8] I. Kim, S.-O. Moon, K.-H. Yu, U.-H. Kim, G.Y. Koh, A novel fibroblast growth factor receptor-5 preferentially expressed in the pancreas, Biochim. Biophys. Acta 1518 (2001) 152–156.
[9] M. Sleeman, J. Fraser, M. McDonald, S. Yuan, D. White, P. Grandison, K. Kumble, J.D. Watson, J.G. Murison, Identification of a new fibroblast growth factor receptor, FGFR5, Gene 271 (2001) 171–182.
[10] J. Lo, S. Lee, M. Xu, F. Liu, H. Ruan, A. Eun, Y. He, W. Ma, W. Wang, Z. Wen, J. Peng, 15'000 Unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis, Genome Res. 13 (2003) 455–466.
[11] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction, Anal. Biochem. 162 (1987) 156–159.
[12] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, Current Protocols in Molecular Biology, Greene Publishing Associates, New York, 1987.

[13] A.P. Feinberg, B. Vogelstein, A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 132 (1983) 6–13.
[14] N. Hukriede, D. Fisher, J. Epstein, L. Joly, P. Tellis, Y. Zhou, B. Barbazuk, K. Cox, L. Fenton-Noriega, C. Hersey, J. Miles, X. Sheng, A. Song, R. Waterman, S.L. Johnson, I.B. Dawid, M. Chevrette, L.I. Zon, J. McPherson, M. Ekker, The LN54 radiation hybrid map of zebrafish expressed sequences, Genome Res. 11 (2001) 2127–2132.
[15] M.S. Clark, Y.J.K. Edwards, D. Peterson, S.W. Clifton, A.J. Thompson, M. Sasaki, Y. Suzuki, K. Kikuchi, S. Watabe, K. Kawakami, S. Sugano, G. Elgar, S.L. Johnson, *Fugu* ESTs: new resources for transcription analysis and genome annotation, Genome Res. 13 (2003) 2747–2753.
[16] M. Halling-Brown, C. Sansom, D.S. Moss, G. Elgar, Y.J.K. Edwards, A *Fugu*-human genome synteny viewer: web software for graphical display and annotation reports of synteny between *Fugu* genomic sequence and human genes, Nucleic Acids Res. 32 (2004) 2618–2622.
[17] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, J. Mol. Evol. 16 (1980) 111–120.
[18] B. Dujon, et al., Genome evolution in yeasts, Nature 430 (2004) 35–44.
[19] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes, Science 290 (2000) 1151–1155.
[20] J. Spring, Vertebrate evolution by interspecific hybridisation—are we polyploid? FEBS Lett. 400 (1997) 2–8.
[21] A. Amores, A. Force, Y.L. Yan, L. Joly, C. Amemiya, A. Fritz, R.K. Ho, J. Langeland, V. Prince, Y.L. Wang, M. Westerfield, M. Ekker, J.H. Postlethwait, Zebrafish hox clusters and vertebrate genome evolution, Science 282 (1998) 1711–1714.
[22] J.S. Taylor, I. Braasch, T. Frickey, A. Meyer, Y. van de Peer, Genome duplication, a trait shared by 22,000 species of ray-finned fish, Genome Res. 13 (2003) 382–390.
[23] A. Christoffels, E.G.L. 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. 21 (2004) 1146–1151.
[24] C. Winkler, M. Schäfer, J. Duschl, M. Schartl, J.-N. Volff, Functional divergence of two zebrafish midkine growth factors following fish-specific gene duplication, Genome Res. 13 (2003) 1067–1081.
[25] 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.