Research article

Many genes in fish have species-specific asymmetric rates of molecular evolution

Dirk Steinke[1], Walter Salzburger[1,2], Ingo Braasch[1,3] and Axel Meyer*[1]

Address: [1]Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany, [2]Center of Junior Research Fellows, University of Konstanz, D-78457 Konstanz, Germany and [3]Department of Physiological Chemistry I, Biozentrum, University of Wuerzburg, Germany

Email: Dirk Steinke - Dirk.Steinke@uni-konstanz.de; Walter Salzburger - Walter.Salzburger@uni-kostanz.de; Ingo Braasch - ingo.braasch@biozenturm.uni-wuerzburg.de; Axel Meyer* - Axel.Meyer@uni-konstanz.de

* Corresponding author

Abstract

Background: Gene and genome duplication events increase the amount of genetic material that might then contribute to an increase in the genomic and phenotypic complexity of organisms during evolution. Thus, it has been argued that there is a relationship between gene copy number and morphological complexity and/or species diversity. This hypothesis implies that duplicated genes have subdivided or evolved novel functions compared to their pre-duplication proto-orthologs. Such a functional divergence might be caused by an increase in evolutionary rates in one ortholog, by changes in expression, regulatory evolution, insertion of repetitive elements, or due to positive Darwinian selection in one copy. We studied a set of 2466 genes that were present in *Danio rerio*, *Takifugu rubripes*, *Tetraodon nigroviridis* and *Oryzias latipes* to test (i) for forces of positive Darwinian selection; (ii) how frequently duplicated genes are retained, and (iii) whether novel gene functions might have evolved.

Results: 25% (610) of all investigated genes show significantly smaller or higher genetic distances in the genomes of particular fish species compared to their human ortholog than their orthologs in other fish according to relative rate tests. We identified 49 new paralogous pairs of duplicated genes in fish, in which one of the paralogs is under positive Darwinian selection and shows a significantly higher rate of molecular evolution in one of the four fish species, whereas the other copy apparently did not undergo adaptive changes since it retained the original rate of evolution. Among the genes under positive Darwinian selection, we found a surprisingly high number of ATP binding proteins and transcription factors.

Conclusion: The significant rate difference suggests that the function of these rate-changed genes might be essential for the respective fish species. We demonstrate that the measurement of positive selection is a powerful tool to identify divergence rates of duplicated genes and that this method has the capacity to identify potentially interesting candidates for adaptive gene evolution.

Background

Biology is a discipline rooted in comparisons. Comparative studies have led to the assembly of a detailed catalogue of biological similarities and, also, of differences between species, yielding insights into the mechanisms by which organisms and their genomes adapt to a wide range of ecological niches. Genomics is the most recent biological discipline to employ comparison-based approaches. During the last five years, whole genome sequences have become available for several vertebrates: e.g., human, mouse, rat, chicken, zebrafish, and two pufferfish species (*Takifugu rubripes* and *Tetraodon nigroviridis*) [1-6]. The increasing wealth of sequence data allows whole genome comparisons for the study of the evolutionary forces that shape genomes [7]. Comparative strategies have identified chromosomal blocks of DNA sequences that are conserved over long evolutionary time spans. Such a degree of evolutionary conservation has, for example, been a powerful guide in sorting functional from non-functional DNA [8-11] and to assign putative gene function.

Ray-finned fishes, which comprise ~25,000 extant species [12], are the most species-rich group of vertebrates. They show enormous differences in their morphology and adaptations to divergent environments. Their sister group are the lobe-finned fishes, which include the other half of all bony vertebrates, such as coelacanths, lungfishes and the tetrapods (amphibian, reptiles, birds and mammals). The ray-finned fishes and the lobe-finned fishes diverged between 400–450 million years ago [13]. Although this large evolutionary distance would imply that only a rather small fraction of the functional portions of their genomes are shared, comparative studies revealed that most human coding sequences (~91%) are homologous to genes in fish [5]. Natural selection is known to leave its footprint on protein-coding sequences in a genome by affecting rates of silent and replacement rates differentially. In sequences that have evolved under positive selection, the number of retained mutations is closer to those that arose by mutation than under purifying selection where amino acid replacement mutations are selected against.

It has been suggested that the large number of fish species and their tremendous morphological diversity might be causally related to a genome duplication event that is specific to the teleost lineage [14-21]. Since gene and genome duplication events are likely to increase the genetic raw-material, it has been speculated that there is a relationship between gene copy number and morphological complexity and, by extension, also species diversity [22,23]. This would imply that one copy of a duplicated gene has diverged from the roles of the pre-duplication ortholog. Such a divergence could be demonstrated by an increase in evolutionary rate, expression differences, regulatory evolution, and/or by evidence for positive Darwinian

selection. Duplicated genes may be redundant after the duplication event, which means that inactivation of one of the two duplicates might have little or no effect on the phenotype [24-26]. Therefore, since at least one of the copies is free from any functional constraint, mutations in this gene-copy might be selectively neutral, having the potential to turn one copy into a non-functional pseudogene. Alternatively, one of the duplicates might adopt a new function through neofunctionalization [22,27,28], or the ancestral function might get divided between the paralogs (subfunctionalization) [29,30]. Recent studies revealed that subfunctionalization can occur rapidly and is often accompanied by prolonged and substantial rates of neofunctionalization in a large proportion of duplicated genes. Thus, a new model, termed sub-neofunctionalization (SNF) has been proposed [31,32]. Other authors argue, that the evolution of new functions may start with the duplication of an existing gene in a sense a preadaptation for that function, followed by a period of evolution among the gene copies, resulting in the preservation of the most effective variant and the 'pseudogenization' and eventual loss of the remaining copies [33].

Post-duplication secondary gene loss is relatively frequent. However, it has been estimated that, nonetheless, ~20%–50% of paralogous genes are retained for longer evolutionary time spans after a genome duplication event [34,35]. A selective advantage due to a new and possibly unique function seems to be sufficient to retain a copy and to prevent degenerative substitutions that would ultimately drive the other copy to become a pseudogene. Among other factors positive Darwinian selection can be responsible for functional divergence between two duplicates (e.g., [36-38]). When a gene with multiple functions is duplicated, the duplicates are redundant only for as long as each copy retains the ability to perform all ancestral roles [29,30]. According to the duplication-degeneration-complementation (DCC) model [30], degenerative mutations preserve rather than disrupt duplicated genes, but also change their functions or at least restrict their original functions which later might become more specialized.

In the current study, simultaneous sequence comparisons of the entire protein-coding portions of the genomes of four fish model species (*Danio rerio*, *Takifugu rubripes*, *Tetraodon nigroviridis* and *Oryzias latipes*) with the human outgroup genome were conducted in order to study the evolutionary extent of sequence conservation and divergence in duplicated fish genes. To facilitate gene identification for functional genomic studies, each data set has been annotated using the structured vocabulary provided by the Gene Ontology Consortium (2001), based on molecular studies of the gene function in *Homo sapiens*. To detect lineage-specific evolutionary processes, we
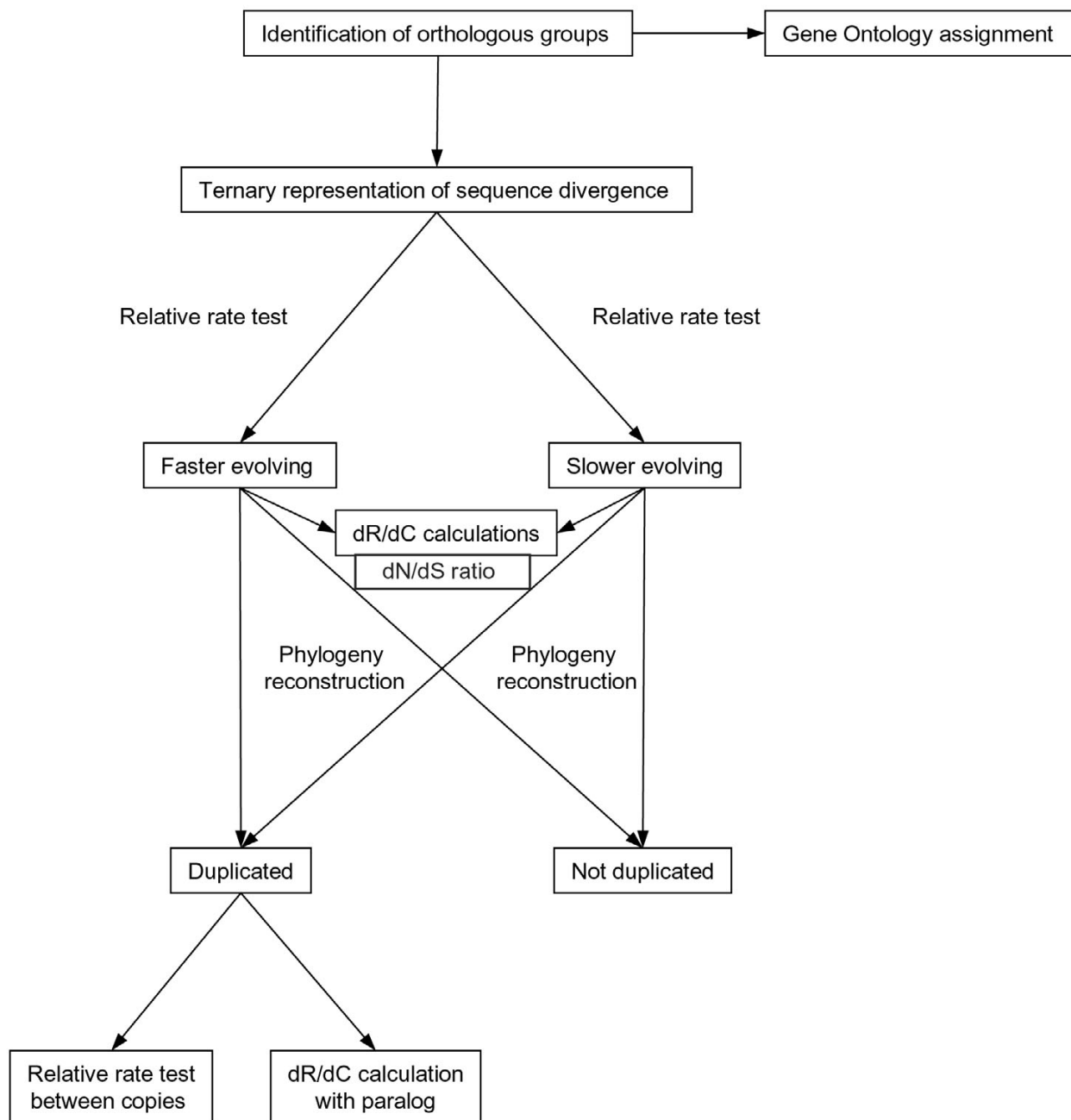
Figure 1
Flowchart of the analysis routine used in this study.

attempted to identify genes that seem to have evolved with significantly divergent (slower or faster) rates of amino acid substitution in one particular species. To this end, we applied a non-parametric relative rate test [39]. Duplicated genes identified with this approach were further studied to test for evidence of positive Darwinian selection and whether the hypothesis for the retention of duplicated genes and the evolution of novel gene functions in one copy is supported. To test whether sequences have been subjected to positive Darwinian selection, the ratio of the proportion of radical nonsynonymous difference (dR) per radical nonsynonymous site and the pro-
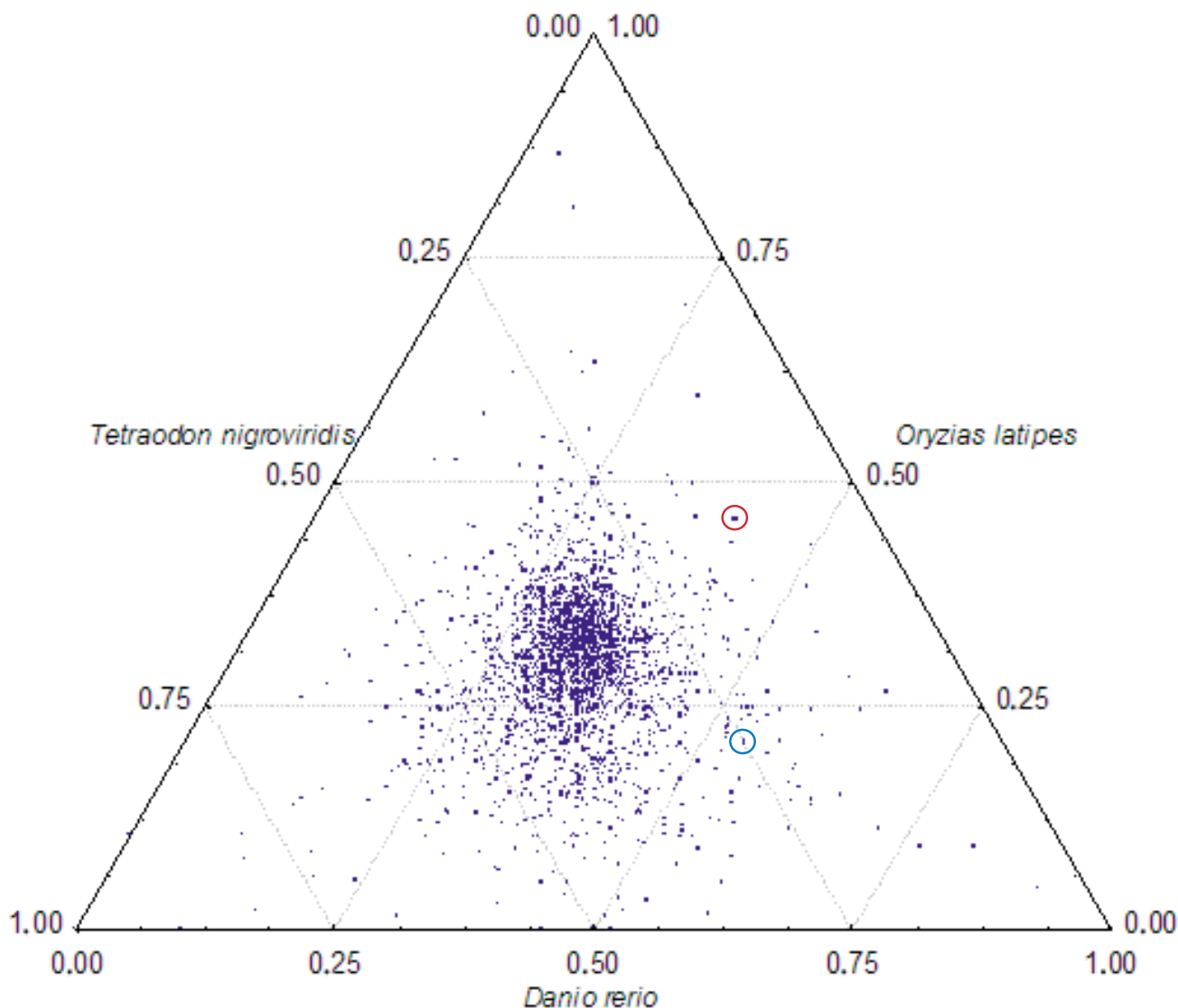
Figure 2
Example of a ternary representation of distances of fish species to human orthologs (*Danio rerio, Oryzias latipes, Tetraodon nigro-viridis*). The red circle represents Lysyl-oxidase-like-1 (LOXL-1), a gene with significantly lower distance from *Tetraodon nigro-viridis* to the human ortholog than to other fish species. The blue circle represents the ankyrin repeat domain 10, a gene with significantly higher distance from *Danio rerio* to the human ortholog than to other fish species.

portion of conservative nonsynonymous site (dC) was calculated using the an approach of Hughes *et al.* [40]. Although different methods have been developed to detect positive selection based on the rate of nonsynonymous (dN) and synonymous substitutions (dS), we note that this ratio can be possibly used to detect positive selection for recently diverged genes only (30–50 MYA) as demonstrated in previous studies [41,42]. It has also been argued that positive selection is of an episodic nature. Meaning that, after a period of positive selection, purify-

ing selection usually blurs the substitution pattern that is indicative of positive selection [38,43]. Since the method by Hughes *et al.* [40] compares nonsynonymous sites and the resulting amino acid changes only, positive selection would need to be active for a much longer period. It should be noted though, that this method may be less sensitive than methods based on the dN/dS ratio [43,44]. Furthermore, a recent study revealed that the dR/dC measure is influenced by the transition/transversion ratio and amino acid composition of the investigated sequences

Table 1: Abundance and proportion of protein genes of a given fish species with significantly lower or higher distance to the human ortholog than other species. The total number of comparisons was N = 2466. [a] denotes genes with significantly lower or higher distance for both pufferfish species

Tetraodon nigroviridis		*Takifugu rubripes*		*Danio rerio*		*Oryzias latipes*		Pufferfishes[a]	
lower	higher	lower	higher	lower	higher	lower	higher	lower	higher
25	35	58	81	80	64	208	26	19	14
$\sum$ = 60		$\sum$ = 139		$\sum$ = 144		$\sum$ = 234		$\sum$ = 33	
1.01%	1.42%	2.35%	3.28%	3.24%	2.60%	8.43%	1.05%	0.77%	0.57%
$\sum$ = 2.43%		$\sum$ = 5.64%		$\sum$ = 5.84%		$\sum$ = 9.49%		$\sum$ = 1.34%	

[45]. Therefore, inferences about positive selection based on the dR/dC method should be treated with some caution.

Results

BLAST similarity searches of data sets of protein sequences of *Tetraodon nigroviridis* revealed a total of 12422 significant hits (e-value: 10^{-50}) when compared to the human genome (outgroup). 12176 hits were found when comparing *Takifugu rubripes* to the human genome, 9619 hits were found for *Danio rerio* and 4681 hits were obtained for *Oryzias latipes*. For our comparative approach (Figure 1), 2466 orthology groups consisting of genes found in all four fish species could be assigned. Figure 2 shows an example of a ternary diagram of p-distances of *Tetraodon nigroviridis*, *Danio rerio* and *Oryzias latipes* amino acid sequences always with respect to the *Homo sapiens* genes. A total of 390 genes were found that have a smaller distance to the human ortholog than their orthologs in other

fish species according to relative rate tests, whereas 220 genes show higher distances in one of the tested fish species' genomes (Table 1). In *Oryzias latipes* a higher percentage (8.43%) of genes show a smaller genetic distance to their human orthologs than in the other fish genomes. The number of genes with higher or lower distance for particular fish species is depicted in Figure 3. The majority of genes detected in *Oryzias latipes* and *Danio rerio* show smaller distances, whereas a higher amount of genes in both pufferfish species seem to have evolved faster (59%). All genes under positive Darwinian selection according to Hughes *et al.* [40] with asymmetric sequence divergence to the human ortholog are given for *Tetraodon nigroviridis* (Table 2), *Takifugu rubripes* (Table 3), *Danio rerio* (Table 4) and *Oryzias latipes* (Table 5). A complete list of the genes with divergent evolutionary rates from all fish species is provided in the supplementary material [see Supplementary File 1]. In all dN/dS calculations (data not shown) dS

Table 2: *Tetraodon nigroviridis* protein genes under positive Darwinian selection with significantly lower ([a]) or higher ([b]) distance to the human ortholog than other fish species. Duplicated genes are given with rate differences between the two copies according to the relative rate test of Tajima (1993) at the 5% level with one asterisk and at the 1% level with two asterisks.

| GenBank Acc#	annotation according to human (UniGene)	duplicate rate difference
CAG12160.1	lysyl oxidase-like 1 [a]	0.247**
CAF89330.1	nuclear LIM interactor-interacting factor 2 [a]	0.120**
CAG00079.1	chloride intracellular channel 5 [a]	0.053*
CAG07128.1	twisted gastrulation [a]	-
CAF96302.1	TAFA2 [a]	0.014
CAF96493.1	serine/threonine protein kinase 6; aurora-A; IPL1-related kinase [a]	0.067**
CAG00266.1	hypothetical protein DKFZp564D0478 [a]	0.147*
CAF94681.1	eukaryotic translation elongation factor 1 gamma [a]	-
CAF96262.1	TTK protein kinase [a]	-
CAF91195.1	zinc finger protein 207 [a]	-
CAF91040.1	exosome component 7[a]	-
CAF87120.1	hypothetical protein FLJ10996 [a]	-
CAF92823.1	ubiquitin domain containing 1 [a]	0.128
CAF91601.1	peroxisomal lon protease [a]	-
CAF95463.1	hypothetical protein FLJ21156 [a]	-
CAF91120.1	syntaxin binding protein 1 [b]	0.350**
CAG08971.1	short-chain dehydrogenase/reductase [b]	-
CAG13862.1	exostosin 1 [b]	0.074
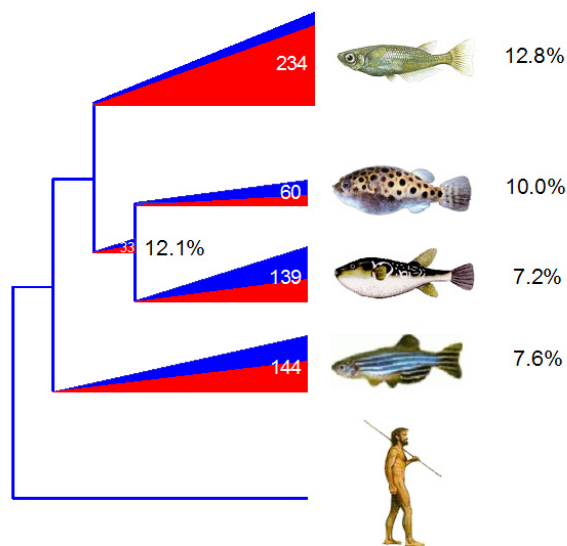
Figure 3
Phylogeny of the studied fish species. The proportion of genes with significantly lower (red) or higher (blue) distance to the human ortholog than to other fish species, are mapped onto the phylogeny as proportional triangles. Numbers within the triangle represent the total abundance of those genes. The percentages represent the corresponding proportion of transcription factors.

outperforms dN, which is expected given the old divergence times between the investigated species.

All the genes with divergent evolutionary rates have been tentatively annotated according to the best hit match according to the Homo sapiens UniGene data base. For these duplicates, the rate difference and the results of the relative rate tests between the two paralogs are given in the corresponding table (Table 2, 3, 4, 5). The annotation of the zebrafish best hits was additionally validated by comparisons to the ZFIN database (Table 4). Table 6 shows the number of fish specific paralogous pairs of genes discovered in this study with different evolutionary rates as compared to their human ortholog. Furthermore, the number of pairs where one paralog appears to have evolved under positive Darwinian selection is listed in Table 6. Accordingly, a total of 49 fish specific paralogous genes could be identified. Of these, 24 show a statistically significant accelerated rate of evolution in one of the two copies and 14 show positive Darwinian selection in one paralogous copy.

We then compared the relative frequency of genes of different gene functions according to the Gene Ontology (GO) classification. Figure 4 shows the number of genes

with different functions for the complete dataset as well as for those genes where one of the fish species shows higher or smaller genetic distances. Some functions are likely to be overrepresented due to small sample sizes, like copper ion binding proteins, where 60% of the total numbers of proteins of this gene function show higher or lower distances in one fish species. This applies also to peptidases (40%), sugar binding proteins (60%) and members of the Wnt signaling pathway (100%). On the other hand, only one group (ubiquitin-protein ligases) shows a lower percentage (7.5%) than within the total number of proteins of this function in all comparisons (Figure 4).

Overrepresentation of ATP-binding proteins and transcription factors
The relative frequency of genes of a certain functions of all fish genes under positive Darwinian selection (Figure 5) shows that more ATP binding proteins and transcription factors were found than could be expected based on the number of ATP binding proteins and transcription factors in general. A χ^2 (df = 2) test (depicted in Figure 4) confirmed that ATP binding proteins, hydrolases, oxidoreductases, transferases, and transcriptions factors occur significantly more often than could be expected by chance (highlighted by asterisks in Figure 4). The proportion of transcription factors detected in each species is on average similar (~10%) to other species (Figure 3). Overall, the majority of genes with divergent evolutionary rates among species are DNA-, ATP- and protein-binding proteins and enzymes (Figure 5). A comparison of the rates of Gene Ontology (GO) groups among the investigated fish genes is provided in Figure 6. Some GO groups include genes with divergent evolutionary rates only in zebrafish and medaka (calmodulin binding proteins, iron binding proteins, kinases, structural components, sugar binding proteins and members of the Wnt signaling pathway). In other GO groups the frequency of genes with divergent evolutionary rates from both pufferfish species is higher than 50% (e.g., calcium and copper ion binding proteins, methyltransferases, oxidoreductases and peptidases).

Discussion
Our final data-set of genes that were recovered from all four species included a total of 2466 orthologous duplicates. This number was mainly limited by the small number of available *Oryzias latipes* sequences (4681). To examine whether these genes evolved at faster or slower evolutionary rates within one of the four fish species, relative rate tests were performed comparing each fish gene with the orthologs in other fish genomes. The human ortholog was used as outgroup. About two thirds of the 610 genes that showed significantly different evolutionary rates in one or more fish species with respect to the others showed a lower rate of molecular evolution for only one fish species (especially in *Oryzias latipes*). This implies that

Table 3: *Takifugu rubripes* protein genes under positive Darwinian selection with significantly lower ([a]) or higher ([b]) distance to the human ortholog than other fish species. Duplicated genes are given with rate differences between the two copies according to the relative rate test of Tajima (1993) at the 5% level with one asterisk and at the 1% level with two asterisks.

JGI Acc#	annotation according to human (UniGene)	duplicate rate difference
FRUP00000147506	syntaxin 3A [a]	0.178**
FRUP00000132184	lysyl oxidase-like 1 [a]	0.207**
FRUP00000139282	chloride intracellular channel 5 [a]	0.074
FRUP00000156028	twisted gastrulation [a]	-
FRUP00000136434	mitochondrial elongation factor G1 [a]	0.170*
FRUP00000139600	TAFA2 [a]	-
FRUP00000129212	serine/threonine protein kinase 6 [a]	-
FRUP00000153498	exostosin 1 [a]	0.086
FRUP00000133663	zinc finger protein 106 homolog [b]	0.201**
FRUP00000162759	hydroxysteroid (17-beta) dehydrogenase 4 [a]	-
FRUP00000133083	dUTP pyrophosphatase [a]	-
FRUP00000133686	excision repair cross-complementing 1 isoform 2 [a]	-
FRUP00000132409	cytochrome c oxidase subunit Va [a]	-
FRUP00000129950	glycine cleavage system protein H [a]	-
FRUP00000140923	mitogen-activated protein kinase kinase kinase kinase 2 [a]	-
FRUP00000157532	tryptophan rich basic protein [a]	-
FRUP00000133839	stress-induced-phosphoprotein 1 [a]	-
FRUP00000136741	dual specificity phosphatase 14 [a]	-
FRUP00000148346	transitional epithelia response protein [a]	-
FRUP00000149078	drebrin-like [a]	-
FRUP00000151439	jumonji domain containing 2C [a]	0.319**
FRUP00000134740	tumor necrosis factor type 1 receptor associated protein [a]	0.099**
FRUP00000133087	leprecan-like 1 [a]	0.317**
FRUP00000144902	calcium binding protein 5 [a]	0.012
FRUP00000162608	neurogenic differentiation 4 [a]	-
FRUP00000130779	FLJ21963 protein [a]	-
FRUP00000143840	thioredoxin domain containing [a]	-
FRUP00000154905	terminal deoxynucleotidyltransferase interacting factor 1 [a]	-
FRUP00000128255	SMILE protein [a]	-

these genes are conserved in that particular lineage, most likely as a result of purifying selection (Figure 3). This high fraction of genes with a smaller genetic distance, i.e. slower evolutionary rate, might be due to the fact that comparisons based on BLAST searches are biased, so that genes with a higher rate are not recovered with the stringent e-value threshold we have applied.

The split between Sarcopterygii and Actinopterygii occurred about 400–450 million years ago [46]. Therefore it is possible that genes accumulated so many mutations and back mutations so that these genes could not be homologized applying very stringent BLAST conditions necessary for reliable annotations. However, we were able to identify 220 genes that show a statistically significant increased rate of molecular evolution. These genes presumably have been subjected to relaxed functional constraints. Genes, which did not accumulate more mutations than the average, are likely to have been subjected to purifying selection and thus, were not free to evolve. [43,47].

Table 6: All fish specific paralogous pairs of proteins discovered in this study with significantly lower or higher distance to the human ortholog than other fish species. "dR/dC>1" represents the number of pairs where at least one paralog is under positive Darwinian selection according to the method of Hughes *et al.* (1990). "rel. rate sign." represents the number of pairs where at least one paralog shows a significantly higher rate of evolution according to the relative rate test of Tajima (1993).

		dR/dC>1		rel. rate sign.		total number of pairs
lower distance	13	44.83%		16	55.17%	29
higher distance	1	5.00%		8	40.00%	20
Σ	14	28.57%		24	48.98%	49

Table 4: *Danio rerio* genes under positive Darwinian selection with significantly lower ([a]) or higher ([b]) distance to the human ortholog than other fish species. Duplicated genes are given with rate differences between the two copies according to the relative rate test of Tajima (1993) at the 5% level with one asterisk and at the 1% level with two asterisks. Known gene symbols according to ZFIN are also provided.

Ensembl Acc#	annotation according to human (UniGene)	duplicate rate difference	ZFIN
ENSDARP00000003385	ankyrin repeat domain 28 [a]	0.285**	ankrd
ENSDARP00000024204	vesicle-associated membrane protein A isoform 2 [a]	-	vapa
ENSDARP00000006783	cofilin 2 [a]	0.078*	cfl2
ENSDARP00000026380	KIAA0073 protein [a]	-	-
ENSDARP00000039499	haloacid dehalogenase-like hydrolase domain containing 2 [a]	-	-
ENSDARP00000005886	arsenate resistance protein ARS2 isoform b [a]	-	ars2
ENSDARP00000049098	thyroid hormone receptor interactor 13 [a]	-	trip11
ENSDARP00000010419	DnaJ subfamily A member 2 [a]	0.135	dnaja2
ENSDARP00000048880	gamma-glutamyl Carboxylase [a]	-	ggcx
ENSDARP00000025803	SEC14 (S. cerevisiae)-like 1 [a]	0.114	sec14ll
ENSDARP00000001039	hypothetical protein MGC10882 [a]	-	-
ENSDARP00000038616	glutaryl-Coenzyme A dehydrogenase isoform a precursor [a]	0.083	gcdh
ENSDARP00000024124	activin A type II receptor precursor [a]	0.057**	acvr2a
ENSDARP00000025487	DKFZP434B168 [a]	-	dkfzp434b168
ENSDARP00000024082	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP [a]	-	atic
ENSDARP00000038219	roundabout 1 isoform a [a]	0.276*	robo1
ENSDARP00000008117	HIRA interacting protein 5 [a]	-	-
ENSDARP00000026984	beta catenin-like 1 [a]	-	-
ENSDARP00000046992	carbohydrate (chondroitin) synthase 1 [a]	-	chys1
ENSDARP00000016016	phosphatidylserine decarboxylase [a]	-	-
ENSDARP00000023471	ADP-ribosylation factor-like 6 interacting protein [a]	-	arl6ip
ENSDARP00000002978	aspartate aminotransferase 1 [a]	0.015**	-
ENSDARP00000016111	ADP-ribosylation factor 4-like [a]	0.032**	arl4
ENSDARP00000018086	androgen-induced 1 [a]	-	aig1
ENSDARP00000015026	signal recognition particle 72 kDa [a]	-	wu:fi03d11
ENSDARP00000038278	cytidine deaminase [a]	-	aicda
ENSDARP00000006977	ankyrin repeat domain 10 [b]	0.211	ankrd10
ENSDARP00000009789	chromobox homolog 3 HP1 gamma [a]	0.009**	cbx
ENSDARP00000016540	RAB25 [a]	0.170*	rab

We found considerable numbers of genes under positive Darwinian selection according to Hughes *et al.* [40] with relaxed substitution rates at the amino acid level. It seems that in such cases selection acted more strongly to conserve the amino acid sequence of a gene in a particular lineage, possibly to maintain their ancestral function. In many cases, this function still seems to be shared with tetrapods and their common ancestors. Remarkably, 66% of the genes (Figure 5) that appear to have evolved under positive Darwinian selection turn out to be binding proteins, especially transcription factors and ATP binding proteins. Transcription factors represent on average ~10% of genes with significantly higher or lower evolutionary rates in each species, which is higher than the average proportion of transcription factors in vertebrate proteomes (~3%) (according to the EBI Eukaryotic Genome database). When identified as deviating gene, then it was in almost all cases because of a smaller rate of substitution, suggesting that these genes are highly conserved in at least one of the studied fish species. Mutations in a binding domain, as well as in regulatory regions [48,49], for example in a transcription factor, could negatively effect the

expression of genes and therefore such mutations might be selected against.

None of the detected genes show evidence of positive selection using a dN/dS calculation following the method of Yang [50]. Due to the age of the investigated species, saturation might have blurred the signal of positive selection. However, a recent study [45] showed that the dR/dC ration is influenced by the transition/transversion ratio and amino acid composition of the investigated sequences. Therefore, inferences on positive selection based on this method should be treated cautiously.

A considerable high number of genes that have a higher or lower evolutionary rates are somehow related to neural development or functions of the brain. This might reflect differences in behavior and cognitive abilities between the investigated fish species.

We identified 49 new duplicated genes in this study, which had not been identified as such before, and which are likely to be the result of the fish-specific genome dupli-

Table 5: *Oryzias latipes* **genes under positive Darwinian selection with significantly lower distance to the human ortholog than other fish species. Duplicated genes are given with rate differences between the two copies according to the relative rate test of Tajima (1993) at the 5% level with one asterisk and at the 1% level with two asterisks.**

GenBank Acc#	annotation according to human (UniGene)	duplicate rate difference
AJ457222	DNA topoisomerase I	-
AU167343	CLIP-associating protein I	-
AU167618	Ras-related associated with diabetes	-
AU167923	atrophin-1 interacting protein 1	-
AU176665	TAF6-like RNA polymerase II	-
AU177030	protein inhibitor of activated STAT X isoform beta	-
AU177176	metastasis suppressor 1	-
AU177627	Utrophin	-
AU180234	SNF2 histone linker PHD RING helicase	-
AV668786	AMPK-related protein kinase 5	-
AV670534	hippocampus abundant transcript 1	0.080*
BJ000375	ephrin receptor EphA2	0.149*
BJ003764	mitogen-activated protein kinase kinase kinase kinase 3	-
BJ004731	NICE-4 protein	-
BJ005724	tsJ homolog 1 isoform a	-
BJ007326	transcription factor T	-
BJ007726	G protein-coupled receptor 155	-
BJ008665	transducin-like enhancer protein 4	-
BJ008817	protein kinase C and casein kinase substrate in neurons 2	-
BJ021405	nucleoporin 62 kDa	-
BJ488812	fibulin 6	-
BJ490986	TRAF2 and NCK interacting kinase	-
BJ493563	PDZ domain containing ring finger 1	-
BJ495426	endothelin converting enzyme 1	0.084
BJ501370	DIP2-like protein isoform a	-
BJ517858	microfibrillar-associated protein 1	-
BJ519269	BCL2/adenovirus E1B 19kD interacting protein 2	-
BJ524171	component of oligomeric golgi complex 4	-
BJ527988	UDP-glucose ceramide glucosyltransferase-like 1	-
BJ539899	nuclear phosphoprotein PWP1	-
BJ543391	ribosomal protein L27a	-
BJ704447	pre-B-cell colony enhancing factor 1 isoform a	-
BJ704659	TCDD-inducible poly(ADP-ribose) polymerase	-
BJ706497	ubiquitin specific protease 34	-
BJ713400	neuro-oncological ventral antigen 1 isoform 2	-
BJ717464	thimet oligopeptidase 1	-
BJ724303	dihydrolipoamide S-succinyltransferase	-
BJ728080	myeloid/lymphoid or mixed-lineage leukemia 3	-
BJ728082	chromodomain helicase DNA binding protein 4	-
BJ728281	engulfment and cell motility 1 isoform 1	-
BJ728299	transcription factor AP-2 alpha	-
BJ728451	spermine synthase	-
BJ728665	Rb1-inducible coiled coil protein 1	-
BJ729946	intersectin 1 isoform ITSN-l	-
BJ730543	B-cell lymphoma 6 protein;	-
BJ730756	RE1-silencing transcription factor	-
BJ731205	MLL septin-like fusion	-
BJ733043	T-cell lymphoma invasion and metastasis 1	-
BJ733490	RAD21 homolog	-
BJ733549	SH3-domain binding protein 4	-
BJ734934	serine/threonine kinase 2	-
BJ735209	methylene tetrahydrofolate dehydrogenase 2 precursor	-
BJ735433	Huntingtin	-
BJ736888	bromodomain containing protein 2	-
BJ737071	fibrinogen C domain containing 1	0.061
BJ742416	JM1 protein	-
BJ743130	Moesin	-
BJ746379	microtubule associated serine/threonine kinase-like	-

cation [17,18,20,51]. Twenty-one of those show an increased rate in only one of the fish paralogs in a particular fish lineage. In 14 of these cases the increase in rate is likely to be the result of positive Darwinian selection (Table 6). Increases in the evolutionary rate in one copy could also be explained by the classical Ohno model of gene evolution [52], which predicts that one copy will evolve more rapidly at nonsynonymous sites compared to the other, as an effect of their redundancy. Nevertheless, it is difficult to demonstrate clear signs of positive Darwinian selection, when the duplication event is ancient [41]. For genes that show a faster evolutionary rate in one species and, in addition, evidence for positive Darwinian selection, one might expect concomitant divergence in function. On the other hand, for paralogs where positive Darwinian selection could not be demonstrated and where the evolutionary rates have not increased, one

might assume that these genes have been under purifying selection or that these genes are about to lose their function according to the duplication-degeneration-complementation (DCC) model [30] due to the accumulation of degenerative mutations. Therefore, such genes may have a similar function or even be completely redundant functionally with respect to their ortholog. Although it might seem unlikely that two duplicates of one ancestral genes perform exactly the same function(s) after more than 300 million years of evolution [51,53], redundancy has been shown to be widespread in genomes of higher organisms [23-25].

Tetraodon nigroviridis
Of the 93 genes with faster or slower evolutionary rates specific to *Tetraodon nigroviridis*, 40 were duplicated (43%). Unique increases in the rate of evolution for one
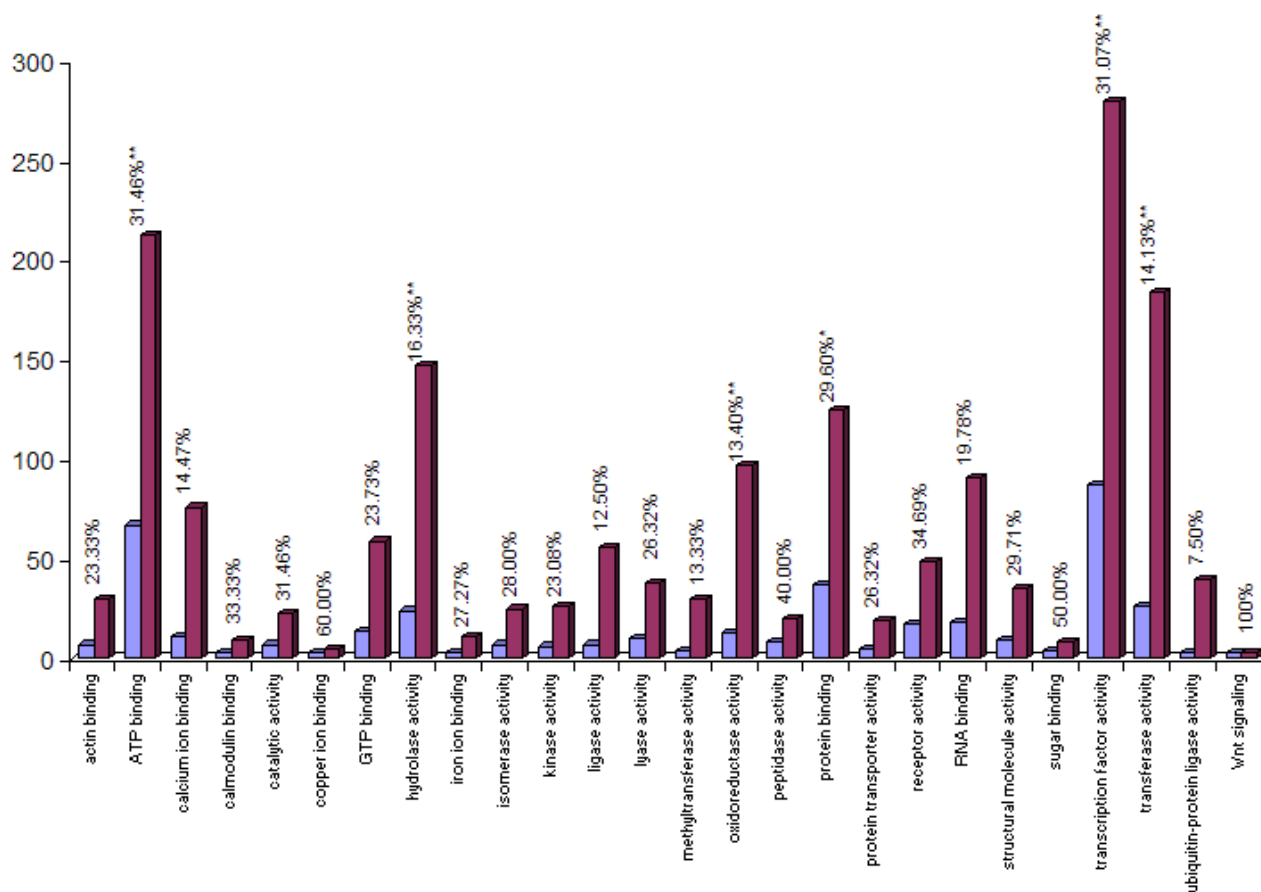

Figure 4
Abundances of gene functions (according to GO) of all fish protein genes with significantly lower or higher distance to the human ortholog than to other fish species (blue) compared to total abundances of gene functions (violet). The percentages represent the rates of the chosen genes compared to the total number of proteins of a given gene function. Significant differences between total and divergent gene abundances according to a χ^2-Test are given at the 5% level with one asterisk and at the 1% level with two asterisks.
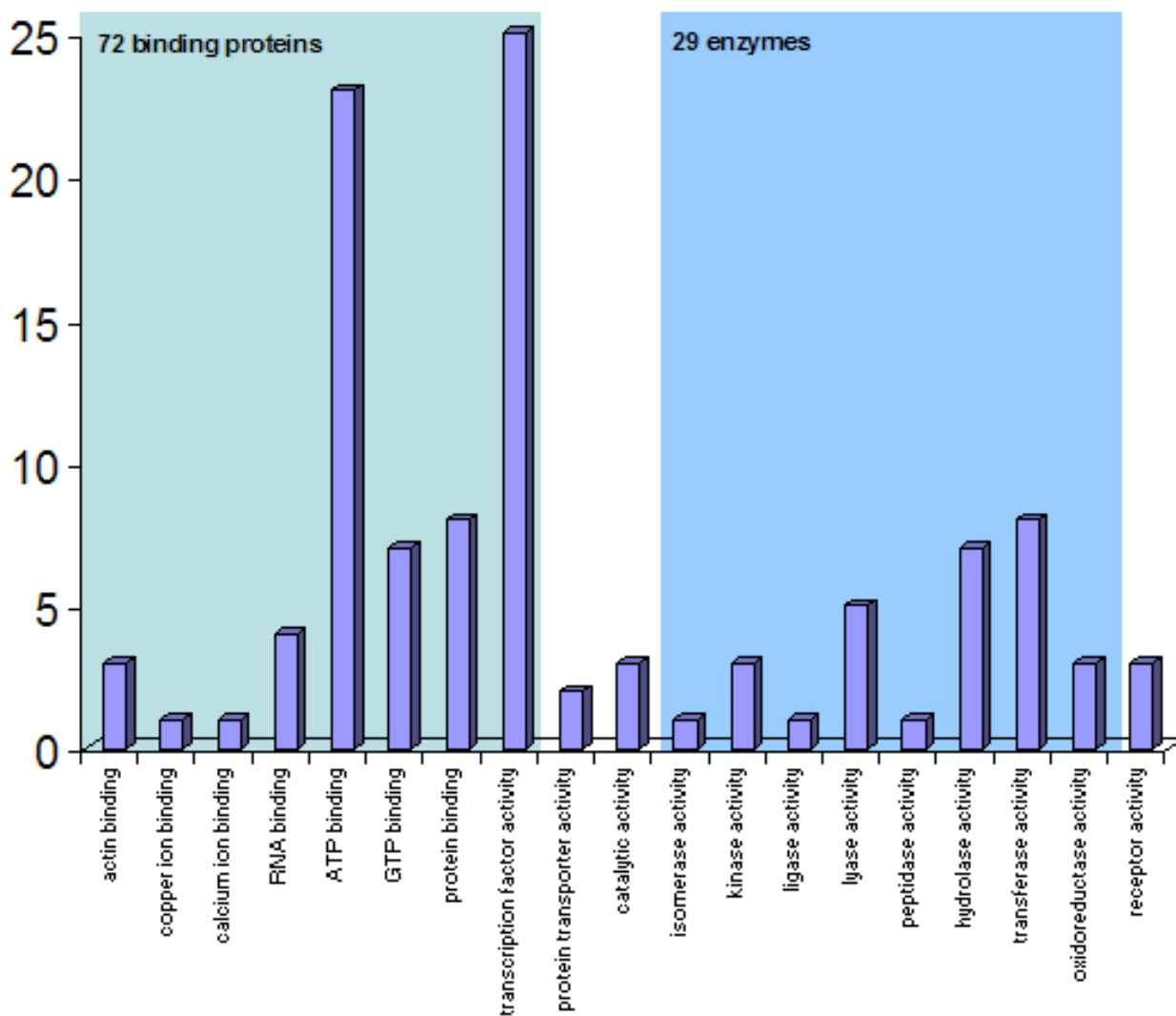
Figure 5
Abundances of gene functions (according to GO) of all fish protein genes under positive Darwinian selection with significantly lower or higher distance to the human ortholog than to other fish species.

copy of the duplicated genes, as a possible result of positive Darwinian selection, were most obvious in three cases (Table 2): The nuclear LIM interactor-interacting factor is an evolutionarily conserved transcriptional regulator that acts globally to silence neuronal genes [54]. The syntaxin binding protein 1 is a neural-specific, syntaxin-binding protein that may participate in the regulation of synaptic vesicle docking and fusion [55]. The hypothetical protein DKFZp564D0478 is known to be expressed in the hippocampus of vertebrates, although nothing is known about the function of this protein. Another positively selected duplicated gene is the ubiquitin domain containing 1

(UBTD1). There is no significant increase in the rate of evolution and it is only known that this protein is also expressed in the hippocampus [56]. Other genes are single-copy genes mainly coding for enzymes. However, the eukaryotic translation elongation factor I gamma gene encodes for a subunit of the elongation factor-1 complex, which is responsible for the delivery of aminoacyl tRNAs to the ribosome.

Takifugu rubripes
A total of 172 genes could be identified, which show a faster or slower rate of molecular evolution in *Takifugu*
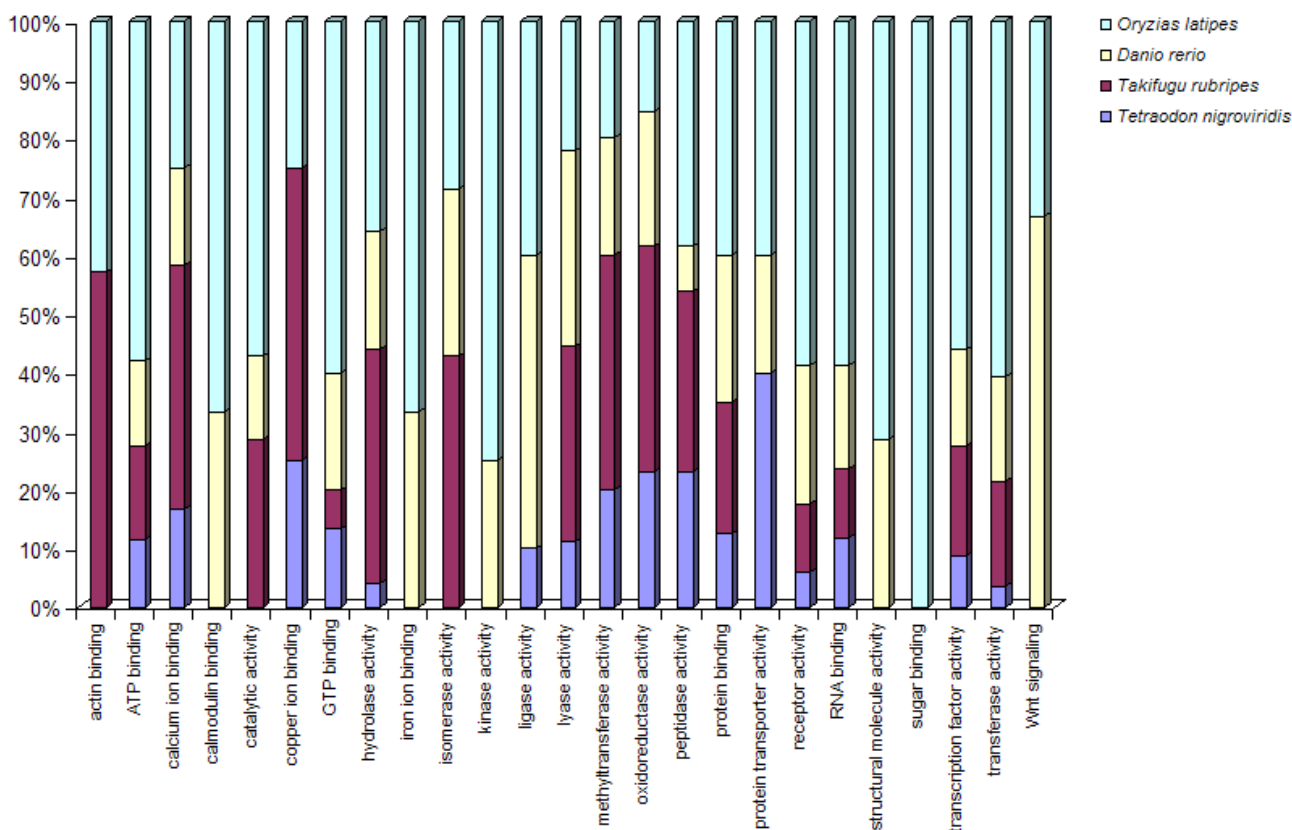
Figure 6
Comparison of rates of gene functions (according to GO) of all fish protein genes with significantly lower distance to the human ortholog than other fish species.

rubripes as compared to the other four fish species. 74 (43%) of these genes are fish-specific duplicates. The total number is higher than in *Tetraodon*, which might reflect differences in sequence completeness. Positive Darwinian selection could be detected for 10% of the duplicated genes, most of which (70%) showed increased substitution rates in only one of the paralogs (Table 3): Syntaxin 3a is potentially involved in docking of synaptic vesicles at presynaptic active zones. In mammals, this gene occurs in different isoforms and is highly expressed in the larynx, which is evolutionarily and developmentally derived from branchial arches. The mitochondrial elongation factor G1 encodes one of the mitochondrial translation elongation factors [57]. The zinc finger protein 106 (ZFP106) is a conserved transcription factor of unknown function. However, its cDNA shares an extended region of identity with the *scr* homology domain 3 binding protein 3 (Sh3bp3) cDNA encoding a protein implicated in the

insulin-signaling pathway [58]. In situ hybridization of mouse embryos confirmed that ZFP106 is predominantly expressed in tissues with high developmental activity of either nuclear respiratory factor-1 (brown fat and developing brain) or myogenin (striated muscle). Jumonji domain containing 2C is also a transcription factor containing PHD finger motifs [59]. PHD finger motifs are zinc finger-like sequences found in nuclear proteins that participate in chromatin-mediated transcriptional regulation and are present in a number of proto-oncogenes. The TNF receptor-associated protein 1 (TRAP1) is a chaperone belonging to the HSP90 family that expresses an ATPase activity [60]. Remarkably, TRAP1 interacts with the C-terminal ends of the proteins encoded by both exostosin 1 (EXT1) and exostosin 2 (EXT2). EXT1 has apparently also evolved under positive selection and was duplicated in both examined pufferfish species. Leprecan-like 1 is a cartilage-associated protein precursor found in articular

chondrocytes and expressed in a variety of mammalian tissues [61].

Again the detected single copy genes are generally coding for a variety of enzymes. One exception is the protein encoded by *drebrin*-like. It is a cytoplasmic actin-binding protein thought to play a role in the process of neuronal growth and to be a member of the drebrin family of proteins that are developmentally regulated in the brain.

Pufferfishes
Out of a total of 265 genes, 33 appeared to show an accelerated rate of evolution for both species This relatively high number is in concordance with the findings that, in the *Tetraodon* genome neutral nucleotide sequence evolution per year is about twice as fast as in humans [6,62]. Thus, it might be possible that the genome of pufferfish species evolves faster in general compared to other vertebrates, and possibly even other fishes. This could be linked to the processes related to the extreme degree of genome compaction in pufferfishes. In ten of the above mentioned cases we were able to detect signals of positive Darwinian selection (Tables 2 and 3).

One of the these 10 genes, the extracellular copper enzyme Lysyl-oxidase-like-1 (LOXL-1) initiates the cross linking of collagens and elastin in the process of building and deposition of elastic fibers. LOXL-1 thus seems to have an essential role in elastogenesis and resilience. LOXL-1 Mutant mice have among other problems a defect in elastic fiber renewal in adult tissues including the lower dermis of the skin [63]. It is tempting to speculate that the lower divergence of LOXL-1 in pufferfishes might be explained by the importance of elasticity of tissues in these species due to their ability to inflate as defense mechanism [64-67]. In addition to cross-linking extracellular matrix proteins, the encoded protein may have a role in tumor suppression [68]. Chloride intracellular channels are involved in chloride ion transport within various subcellular compartments. The chloride intracellular channel 5 gene (CLIC5) specifically associates with the cytoskeleton of placenta microvilli [69]. As mentioned before, one copy of EXT1 shows a higher rate of evolution due to positive Darwinian selection. EXT1 is a transferase involved in the chain elongation step of heparan sulfate biosynthesis. It appears to be a tumor suppressor [70].

Examples of positively selected single copy genes in both pufferfish species were 'twisted gastrulation' and TAFA2. 'Twisted gastrulation' encodes a secreted BMP-binding protein that is a BMP signalling agonist in the dorsal-ventral patterning pathway [71]. The TAFA proteins are predominantly expressed in specific regions of the brain, and are postulated to function as brain-specific chemokines or neurokines, that act as a regulator of immune and nervous cells [72].

Danio rerio
Of the 144 genes with faster or slower evolutionary rates in *Danio rerio*, 38 belong to the group of duplicated genes. Among those genes with significantly higher rates in only one paralog (Table 4) are Cofilin 2, activin A type II receptor, Roundabout 1, GOT1, ARF4L, CBX3 and RAB25. Cofilin 2 controls reversibly actin polymerization and depolymerization in a ph-sensitive manner. It has the ability to bind g- and f-actin in a 1:1 ratio of cofilin to actin. It is the major component of intranuclear and cytoplasmic actin rods [73]. Activins are dimeric growth and differentiation factors, which belong to the transforming growth factor-beta (TGF-beta) superfamily of structurally related signaling proteins. Type II receptors are required for ligand-binding and for expression of type I receptors [74]. Roundabout 1 (ROBO1) encodes for an integral membrane protein that is both an axon guidance receptor and a cell adhesion receptor, and is involved in the decision by axons to cross the central nervous system midline [75]. Glutamic-oxaloacetic transaminase (GOT) is a pyridoxal phosphate-dependent enzyme, which exists in cytoplasmic and mitochondrial forms, GOT1 and GOT2. GOT plays a role in amino acid metabolism and the urea and tricarboxylic acid cycles. The two enzymes are homodimeric and show close homology [76]. The ADP-ribosylation factor 4-like is a member of the ADP-ribosylation factor family of GTP-binding proteins. ARF4L is closely similar to ARL4 and ARL7 and each has a nuclear localization signal and an unusually high guanine nucleotide exchange rate [77]. The protein encoded by the chromobox homolog 3 (CBX3) binds DNA and is a component of heterochromatin. This protein can also bind lamin B receptor, an integral membrane protein found in the inner nuclear membrane. The dual binding functions of the encoded protein may explain the association of heterochromatin with the inner nuclear membrane [78]. The gene encoding RAB25 may selectively regulate the apical recycling and/or transcytotic pathways [79]. Only little is known about the other genes detected in *Danio rerio*. Most of the single copy genes represent enzymes of varying function.

Oryzias latipes
A total of 234 genes could be identified of genes with divergent evolutionary rates specific to *Oryzias latipes* when compared to the other fish species. Of those, only 14 were found to be duplicated. Only four genes with positive selected paralogs could be detected in *Oryzias latipes* (Table 5). One of those, with an accelerated rate of evolution in one of the paralogs is the ephrin receptor EphA2, which belongs to the ephrin receptor subfamily of the protein-tyrosine kinase family. EPH and EPH-related

receptors have been implicated in mediating developmental events, particularly in the nervous system and limb development [80]. The ephrin receptors are divided into two groups based on the similarity of their extracellular domain sequences and their affinities for binding ephrin-A and ephrin-B ligands. This gene encodes a protein that binds ephrin-A ligands.

For two genes both paralogs show similar rates of evolution. The Endothelin-converting enzyme-1 is involved in the proteolytic processing of endothelin-1, -2, -3 to biologically active peptides [81]. Ficolin 1 encoded by FCN1 is predominantly expressed in the peripheral blood leukocytes, and has been postulated to function as a plasma protein with elastin-binding activity [82].

Other genes are single-copy genes mainly coding for enzymes with general functions. Remarkably, the number of positive selected genes in *Oryzias latipes* is higher than in the other investigated species and most of them have low substitution rates compared to other fish. Selection might thus act as maintenance for those genes and their function.

Conclusion
We identified genes under positive Darwinian selection using a combination of BLAST searches and phylogenetic methods. With these methods we could also demonstrate that the measurement of positive selection is a good method to identify divergent fates of duplicated genes. Most genes behave differently in particular species, which implies that their function is somehow essential, and possibly of adaptive value, for the investigated fish species. We identified 49 of previously unknown duplicated genes where one of the paralogs is under positive Darwinian selection and shows a significantly high rate of molecular evolution whereas the other copy did not undergo such dramatic changes, most likely due to purifying selection. The fact that these duplicated genes show lineage specific evolutionary rates in the investigated fish species suggests that even after such a long time since the duplication event these genes might still contribute to lineage specific features. One might assume, that these genes therefore play a role in the diversification of lineages. Models such as the DDC model [30] might explain retention and functional divergence of anciently duplicated genes. It is also possible that neofunctionalisation of one paralog occurred. However, when subfunctionalisation is responsible for the functional divergence of genes, this is probably limited to differences in timing and tissue specificity of expression. It has also been suggested that a proportion of duplicate genes undergo rapid subfunctionalisation, accompanied by prolonged and substantial neofunctionalisation [31]. So far, there is little evidence that the new paralogs described here have completely novel functions.

However, in several cases we could detect a significant increase in evolutionary rates in one of the duplicates. This is probably not due to relaxed functional constraints on the whole gene, but rather because duplicated genes experience a brief period of relaxed selection after duplication [26,41]. Duplicates that are being retained over longer evolutionary times are more likely to experience strong purifying selection.

Methods
We used the database software EverEST 1.0 [83], which allows database searches on the basis of the BLAST algorithm and tests the association of results and phylogenetic analysis, making use of a relational database. Data sets of protein data from *Danio rerio* (Zebrafish Sequencing Group at the Sanger Institute, Zv4.0), EST data from *Oryzias latipes* (GenBank), protein data from *Takifugu rubripes* (JGI Fugu v3.0) and protein data from *Tetraodon nigroviridis* (GenBank) were screened against genome data from *Homo sapiens* (GenBank) using a translated BLAST routine (standard vertebrate code) with an expected value threshold of $< 1 \times 10^{-50}$. This relatively high value threshold was used in order to achieve high levels of confidence in the similarity searches, although such a stringent threshold might lead to the missing of some relevant orthologs. The query sequence and all best hits of every single search were aligned using the T-Coffee algorithm [84] implemented in EverEST. The automated analysis is depicted in Figure 1. Each sequence was tentatively assigned Gene Ontology (GO) classification based on annotation of the single 'best hit' match in BLASTX searches of *Homo sapiens* proteins (e$\leq 10^{-50}$). Annotations described here are at the "inferred from electronic annotation" (IEA) level of evidence (The Gene Ontology Consortium 2001). Following the alignment, sequence divergence for every possible human-fish pair was estimated in EverEST as the observed proportion of amino acid sites at which the two sequences under comparison were different (Poisson correction). All alignment positions with gaps were excluded previously (complete deletion). This option is generally desirable because different regions of DNA or amino acid sequences often evolve under different evolutionary forces.

The distances (relative p-distances) were used to construct three-coordinate ternary representations (implemented in EverEST) for cross-species comparisons to visualize species-specific genes as outliers. A relative rate test was applied to each of the orthologous groups. We applied the nonparametric rate test developed by [39] and implemented in MEGA 3.0 [85], and compared the genes with their human and their fish orthologs. For orthologous groups, where the p-distance between *Homo sapiens* and fish amino acid sequences was significantly ($p < 0.05$) higher or smaller compared to the other three fish species,

the ratio of the proportion of radical nonsynonymous difference (dR) per radical nonsynonymous site and the proportion of conservative nonsynonymous site (dC) was calculated using the program SCR3 [40]. This was done to evaluate the selective forces acting on those proteins. Conservative substitutions are amino acid replacements remain constant with respect to charge or polarity, while a substitution at a radical site involve amino acid replacements that changes charge or polarity [86]. In addition we conducted dN/dS calculations using the method by [50] implemented in PAML [87] for the 122 putative genes under positive selection according to dR/dC.

When BLAST identified one or more putative fish orthologs, protein sequences from all species were aligned using T-Coffee [84]. For each alignment, a preliminary tree was drawn. This tree facilitated the identification of identical sequences, sequences that varied only in length, and sequences within species that differed by few amino acids, all of which were removed from the alignment. Very similar sequences could be alleles at one locus or evidence of recent tandem duplications. In either case they were not likely to be important for our study of genome duplication. Phylogenies were reconstructed from the remaining sequences using Poisson-corrected genetic distances and the neighbor joining (NJ) algorithm [88] in MEGA 3.0 [85]. From these trees we identified sets of orthologous genes (i.e. genes which occurred only in monophyletic groups that matched the expected organismal topology). Regions where the alignment was unambiguous were retained and reanalyzed using NJ and maximum likelihood (ML) methods. For these last phylogenetic analyses the most closely related human paralogs (identified from the first NJ analyses) were used as outgroups. PHYML [89] was used to reconstruct ML. The best fitting models of sequence evolution for ML were obtained by ProtTest 1.2 [90].

To investigate whether or not one of the two fish paralogs evolved at a faster rate since their duplication, a relative rate test was applied to each of the genes. We applied the nonparametric rate test developed by [39] and implemented in MEGA 3.0 [85], and compared the paralogs with their human and another fish ortholog. The dR/dC ratio was calculated between the paralogs using the program SCR3 [40] to evaluate the selective forces acting on those.

Authors' contributions

DS conceived the study, carried out the comparative analyses, and drafted the manuscript. WS participated in the design of the study, and helped to draft the manuscript. IB participated in the comparative analyses, the design of the study, and helped to draft the manuscript. AM participated in the study design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

> ### Additional file 1
> *A complete list of genes with divergent evolutionary rates for all fish species of this study*
> Click here for file
> [http://www.biomedcentral.com/content/supplementary/1471-2164-7-20-S1.pdf]

References

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Glueckasmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X: **The sequence of the human genome.** *Science* 2001, **291:**1304-1351.
2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE,

Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.

3. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428**:493-521.

4. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobbson L, Kerje S, Andersson L, Crooijmans RP, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MD, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WR, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK, Wang J, Liu B, Yu J, Yang H, Nefedov M, Koriabine M, Dejong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yang S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte F, Taylor J, He J, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SM, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Betran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyras E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**:695-716.

5. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.** *Science* 2002, **297**:1301-1310.

6. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, de Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H: **Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate protokaryotype.** *Nature* 2004, **431**:946-957.

7. Liberles DA: **Datasets for evolutionary comparative genomics.** *Genome Biol* 2005, **6**:117.

8. Hardison RC: **Conserved noncoding sequences are reliable guides to regulatory elements.** *Trends Genet* 2000, **16**:369-372.

9. Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA: **Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons.** *Science* 2000, **288**:136-140.

10. Pennacchio LA, Rubin EM: **Genomic strategies to identify mammalian regulatory sequences.** *Nat Rev Genet* 2001, **2**:100-109.

11. Gottgens B, Barton LM, Chapman MA, Sinclair AM, Knudsen B, Grafham D, Gilbert JG, Rogers J, Bentley DR, Green AR: **Transcriptional regulation of the stem cell leukemia gene (SCL)--comparative analysis of five vertebrate SCL loci.** *Genome Res* 2002, **12**:749-759.

12. Nelson J: **Fishes of the world.** New York, Wiley; 1994.

13. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392**:917-920.

14. 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: **Zebrafish hox clusters and vertebrate genome evolution.** *Science* 1998, **282**:1711-1714.

15. Wittbrodt J, Meyer A, Schartl M: **More genes in fish?** *BioEssays* 1998, **20**:511-515.
16. Taylor JS, Van de Peer Y, Meyer A: **Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis.** *Curr Biol* 2001, **11**:R1005-8.
17. Taylor JS, Van de Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:1661-1679.
18. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y: **Genome duplication, a trait shared by 22000 species of ray-finned fish.** *Genome Res* 2003, **13**:382-390.
19. Chen WJ, Orti G, Meyer A: **Novel evolutionary relationship among four fish model systems.** *Trends in Genetics* 2004.
20. Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B: **Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes.** *Mol Biol Evol* 2004, **21**:1146-1151.
21. Meyer A, Van de Peer Y: **From 2R to 3R: evidence for the fish-specific genome duplication (FSGD).** *Bio Essays* 2005, **27**:1-9.
22. Ohno S: **Evolution by Gene Duplication.** New York, Springer-Verlag; 1970.
23. Dehal P, Boore JL: **Two rounds of whole genome duplication in the ancestral vertebrate.** *PLoS Biol* 2005, **3**:e314.
24. Nowak MA, Boerlijst MC, Cooke J, Smith JM: **Evolution of genetic redundancy.** *Nature* 1997, **388**:167-171.
25. Gibson TJ, Spring J: **Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins.** *Trends Genet* 1998, **14**:46-9; discussion 49-50.
26. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**:1151-1155.
27. Ohno S: **Ancient linkage groups and frozen accidents.** *Nature* 1973, **244**:259-262.
28. Golding GB, Dean AM: **The structural basis of molecular adaptation.** *Mol Biol Evol* 1998, **15**:355-369.
29. Hughes AL: **The evolution of functionally novel proteins after gene duplication.** *Proc Biol Sci* 1994, **256**:119-124.
30. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**:1531-1545.
31. He X, Zhang J: **Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution.** *Genetics* 2005, **169**:1157-1164.
32. Rastogi S, Liberles DA: **Subfunctionalization of duplicated genes as a transition state to neofunctionalization.** *BMC Evol Biol* 2005, **5**:28.
33. Francino P: **An adaptive radiation model for the origin of new gene functions.** *Nature Genetics* 2005, **37**:573-578.
34. Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS: **Zebrafish comparative genomics and the origins of vertebrate chromosomes.** *Genome Research* 2000, **10**:1890-1902.
35. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization.** *Genetics* 2000, **154**:459-473.
36. Zhang J, Rosenberg HF, Nei M: **Positive Darwinian selection after gene duplication in primate ribonuclease genes.** *Proc Natl Acad Sci U S A* 1998, **95**:3708-3713.
37. Duda TFJ, Palumbi SR: **Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod Conus.** *Proc Natl Acad Sci U S A* 1999, **96**:6820-6823.
38. Hughes AL, Green JA, Garbayo JM, Roberts RM: **Adaptive diversification within a large family of recently duplicated, placentally expressed genes.** *Proc Natl Acad Sci U S A* 2000, **97**:3319-3323.
39. Tajima F: **Simple methods for testing the molecular evolutionary clock hypothesis.** *Genetics* 1993, **135**:599-607.
40. Hughes AL, Ota T, Nei M: **Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules.** *Mol Biol Evol* 1990, **7**:515-524.
41. Van de Peer Y, Taylor JS, Braasch I, Meyer A: **The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes.** *J Mol Evol* 2001, **53**:436-446.
42. Raes J, Van de Peer Y: **Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silico.** *Appl Bioinformatics* 2003, **2**:91-101.

43. Hughes AL: **Adaptive evolution of genes and genomes.** New York, Oxford University Press; 1999.
44. Vacquier VD, Swanson WJ, Lee YH: **Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence?** *J Mol Evol* 1997, **44 Suppl 1**:S15-22.
45. Dagan T, Talmor Y, Graur D: **Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection.** *Molecular Biology and Evolution* 2002, **19**:1022-1025.
46. Benton MJ: **Phylogeny of the major tetrapod groups: morphological data and divergence dates.** *Journal of Molecular Evolution* 1990, **30**:409-424.
47. Zhang P, Gu Z, Li WH: **Different evolutionary patterns between young duplicate genes in the human genome.** *Genome Biology* 2003, **4**:R56.
48. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB: **Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila.** *Nature* 2005, **433**:481-487.
49. Gompel N, Carroll SB: **Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies.** *Nature* 2003, **424**:931-935.
50. Yang Z: **Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution.** *Mol Biol Evol* 1998, **15**:568-573.
51. Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: Paranome age and size differs considerably between ray-finned fishes and land vertebrates.** *Proc Natl Acad Sci USA* 2004, **101**:1638-1643.
52. Ohno S: **The reason for as well as the consequence of the Cambrian explosion in animal evolution.** *J Mol Evol* 1997, **44 Suppl 1**:S23-7.
53. Hoegg S, Brinkmann H, Taylor JS, Meyer A: **Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish.** *J Mol Evol* 2004, **59**:190-203.
54. Yeo M, Lee SK, Lee B, Ruiz EC, Pfaff SL, Gill GN: **Small CTD phosphatases function in silencing neuronal gene expression.** *Science* 2005, **307**:596-600.
55. Weimer RM, Richmond JE, Davis WS, Hadwiger G, Nonet ML, Jorgensen EM: **Defects in synaptic vesicle docking in unc-18 mutants.** *Nat Neurosci* 2003, **6**:1023-1030.
56. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, Hosoiri T, Kaku Y, Kodaira H, Kondo H, Sugawara M, Takahashi M, Kanda K, Yokoi T, Furuya T, Kikkawa E, Omura Y, Abe K, Kamihara K, Katsuta N, Sato K, Tanikawa M, Yamazaki M, Ninomiya K, Ishibashi T, Yamashita H, Murakawa K, Fujimori K, Tanai H, Kimata M, Watanabe M, Hiraoka S, Chiba Y, Ishida S, Ono Y, Takiguchi S, Watanabe S, Yosida M, Hotuta T, Kusano J, Kanehori K, Takahashi-Fujii A, Hara H, Tanase TO, Nomura Y, Togiya S, Komai F, Hara R, Takeuchi K, Arita M, Imose N, Musashino K, Yuuki H, Oshima A, Sasaki N, Aotsuka S, Yoshikawa Y, Matsunawa H, Ichihara T, Shiohata N, Sano S, Moriya S, Momiyama H, Satoh N, Takami S, Terashima Y, Suzuki O, Nakagawa S, Senoh A, Mizoguchi H, Goto Y, Shimizu F, Wakebe H, Hishigaki H, Watanabe T, Sugiyama A, Takemoto M, Kawakami B, Watanabe K, Kumagai A, Itakura S, Fukuzumi Y, Fujimori Y, Komiyama M, Tashiro H, Tanigami A, Fujiwara T, Ono T, Yamada K, Fujii Y, Ozaki K, Hirao M, Ohmori Y, Kawabata A, Hikiji T, Kobatake N, Inagaki H, Ikema Y, Okamoto S, Okitani R, Kawakami T, Noguchi S, Itoh T, Shigeta K, Senba T, Matsumura K, Nakajima Y, Mizuno T, Morinaga M, Sasaki M, Togashi T, Oyama M, Hata H, Komatsu T, Mizushima-Sugano J, Satoh T, Shirai Y, Takahashi Y, Nakagawa K, Okumura K, Nagase T, Nomura N, Kikuchi H, Masuho Y, Yamashita R, Nakai K, Yada T, Ohara O, Isogai T, Sugano S: **Complete sequencing and characterization of 21,243 full-length human cDNAs.** *Nat Genet* 2004, **36**:40-45.
57. Hammarsund M, Wilson W, Corcoran M, Merup M, Einhorn S, Grander D, Sangfelt O: **Identification and characterization of two novel human mitochondrial elongation factor genes, hEFG2 and hEFG1, phylogenetically conserved through evolution.** *Hum Genet* 2001, **109**:542-550.

58. Grasberger H, Ye H, Mashima H, Bell GI: **Dual promoter structure of ZFP106: regulation by myogenin and nuclear respiratory factor-1.** *Gene* 2005, **344:**143-159.

59. Yang ZQ, Imoto I, Fukuda Y, Pimkhaokham A, Shimada Y, Imamura M, Sugano S, Nakamura Y, Inazawa J: **Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines.** *Cancer Res* 2000, **60:**4735-4739.

60. Simmons AD, Musy MM, Lopes CS, Hwang LY, Yang YP, Lovett M: **A direct interaction between EXT proteins and glycosyltransferases is defective in hereditary multiple exostoses.** *Hum Mol Genet* 1999, **8:**2155-2164.

61. Jarnum S, Kjellman C, Darabi A, Nilsson I, Edvardsen K, Aman P: **LEPREL1, a novel ER and Golgi resident member of the Leprecan family.** *Biochem Biophys Res Commun* 2004, **317:**342-351.

62. Roest Crollius H, Weissenbach J: **Fish genomics and biology.** *Genome Research* 2005, **15:**1675-1682.

63. Liu X, Zhao Y, Gao J, Pawlyk B, Starcher B, Spencer JA, Yanagisawa H, Zuo J, Li T: **Elastic fiber homeostasis requires lysyl oxidase-like 1 protein.** *Nat Genet* 2004, **36:**178-182.

64. Wainwright PC, Turingan RG: **Evolution of pufferfish inflation behavior.** *Evolution* 1997, **51:**506-518.

65. Wainwright PC, Turingan RG, Brainerd EL: **Functional morphology of pufferfish inflation: mechanism of the buccal pump.** *Copeia* 1995:614-625.

66. Brainerd EL: **Pufferfish inflation: Functional morphology of postcranial structures in Diodon holocanthus (Tetraodontiformes).** *Journal of Morphology* 1994, **220:**243-261.

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

68. Goy A, Gilles F, Remache Y, Zelenetz AD: **Physical linkage of the lysyl oxidase-like (LOXL1) gene to the PML gene on human chromosome 15q22.** *Cytogenet Cell Genet* 2000, **88:**22-24.

69. Berryman M, Bretscher A: **Identification of a novel member of the chloride intracellular channel gene family (CLIC5) that associates with the actin cytoskeleton of placental microvilli.** *Mol Biol Cell* 2000, **11:**1509-1521.

70. McCormick C, Leduc Y, Martindale D, Mattison K, Esford LE, Dyer AP, Tufaro F: **The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate.** *Nat Genet* 1998, **19:**158-161.

71. Chang C, Holtzman DA, Chau S, Chickering T, Woolf EA, Holmgren LM, Bodorova J, Gearing DP, Holmes WE, Brivanlou AH: **Twisted gastrulation can function as a BMP antagonist.** *Nature* 2001, **410:**483-487.

72. Tom Tang Y, Emtage P, Funk WD, Hu T, Arterburn M, Park EE, Rupp F: **TAFA: a novel secreted family with conserved cysteine residues and restricted expression in the brain.** *Genomics* 2004, **83:**727-734.

73. Gillett GT, Fox MF, Rowe PS, Casimir CM, Povey S: **Mapping of human non-muscle type cofilin (CFL1) to chromosome 11q13 and muscle-type cofilin (CFL2) to chromosome 14.** *Ann Hum Genet* 1996, **60 (Pt 3):**201-211.

74. D'Abronzo FH, Swearingen B, Klibanski A, Alexander JM: **Mutational analysis of activin/transforming growth factor-beta type I and type II receptor kinases in human pituitary tumors.** *J Clin Endocrinol Metab* 1999, **84:**1716-1721.

75. Kidd T, Brose K, Mitchell KJ, Fetter RD, Tessier-Lavigne M, Goodman CS, Tear G: **Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors.** *Cell* 1998, **92:**205-215.

76. Wang CY, Huang YQ, Shi JD, Marron MP, Ruan QG, Hawkins-Lee B, Ochoa B, She JX: **Genetic homogeneity, high-resolution mapping, and mutation analysis of the urofacial (Ochoa) syndrome and exclusion of the glutamate oxaloacetate transaminase gene (GOT1) in the critical region as the disease gene.** *Am J Med Genet* 1999, **84:**454-459.

77. Smith SA, Holik PR, Stevens J, Melis R, White R, Albertsen H: **Isolation and mapping of a gene encoding a novel human ADP-ribosylation factor on chromosome 17q12-q21.** *Genomics* 1995, **28:**113-115.

78. Obuse C, Iwasaki O, Kiyomitsu T, Goshima G, Toyoda Y, Yanagida M: **A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1.** *Nat Cell Biol* 2004, **6:**1135-1141.

79. Prekeris R, Davies JM, Scheller RH: **Identification of a novel Rab11/25 binding domain present in Eferin and Rip proteins.** *J Biol Chem* 2001, **276:**38966-38970.

80. Lindberg RA, Hunter T: **cDNA cloning and characterization of eck, an epithelial cell receptor protein-tyrosine kinase in the eph/elk family of protein kinases.** *Mol Cell Biol* 1990, **10:**6316-6324.

81. Shimada K, Matsushita Y, Wakabayashi K, Takahashi M, Matsubara A, Iijima Y, Tanzawa K: **Cloning and functional expression of human endothelin-converting enzyme cDNA.** *Biochem Biophys Res Commun* 1995, **207:**807-812.

82. Harumiya S, Takeda K, Sugiura T, Fukumoto Y, Tachikawa H, Miyazono K, Fujimoto D, Ichijo H: **Characterization of ficolins as novel elastin-binding proteins and molecular cloning of human ficolin-1.** *J Biochem (Tokyo)* 1996, **120:**745-751.

83. Steinke D, Salzburger W, Meyer A: **EverEST - A phylogenomic EST database approach.** *Phyloinformatics* 2004, **6:**1-4.

84. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302:**205-217.

85. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5:**150-163.

86. Miyata T, Miyazawa S, Yasunaga T: **Two types of amino acid substitutions in protein evolution.** *J Mol Evol* 1979, **12:**219-236.

87. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13:**555-556.

88. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Molecular Biology and Evolution* 1987, **4:**406-425.

89. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic Biology* 2003, **52:**696-704.

90. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21:**2104-2105.