First publ. in: Journal of Molecular Evolution 53 (2001), pp. 434-444

The Ghost of Selection Past: Rates of Evolution and Functional Divergence of Anciently Duplicated Genes

Yves Van de Peer,[1] John S. Taylor,[1] Ingo Braasch,[1] Axel Meyer[1,2]

[1] Department of Biology, University of Konstanz, 78457 Konstanz, Germany
[2] DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Received: 4 January 2001 / Accepted: 29 March 2001

Abstract. The duplication of genes and even complete genomes may be a prerequisite for major evolutionary transitions and the origin of evolutionary novelties. However, the evolutionary mechanisms of gene evolution and the origin of novel gene functions after gene duplication have been a subject of many debates. Recently, we compiled 26 groups of orthologous genes, which included one gene from human, mouse, and chicken, one or two genes from the tetraploid *Xenopus* and two genes from zebrafish. Comparative analysis and mapping data showed that these pairs of zebrafish genes were probably produced during a fish-specific genome duplication that occurred between 300 and 450 Mya, before the teleost radiation (Taylor et al. 2001). As discussed here, many of these retained duplicated genes code for DNA binding proteins. Different models have been developed to explain the retention of duplicated genes and in particular the subfunctionalization model of Force et al. (1999) could explain why so many developmental control genes have been retained. Other models are harder to reconcile with this particular set of duplicated genes. Most genes seem to have been subjected to strong purifying selection, keeping properties such as charge and polarity the same in both duplicates, although some evidence was found for positive Darwinian selection, in particular for *Hox* genes. However, since only the cumulative pattern of nucleotide substitutions can be studied, clear indications of positive Darwinian selection or neutrality may be hard to find for such anciently duplicated genes. Nevertheless, an increase in evolutionary rate in about half of the duplicated genes seems to suggest that either positive Darwinian selection has occurred or that functional constraints have been relaxed at one point in time during functional divergence.

Key words: Genome duplication — Functional divergence — Positive Darwinian selection — Relative rate test

Introduction

Fish are, at least by the criterion of number of species, the most successful group of vertebrates. Recently, it has been suggested that the large number of fish species (about 25,000, Nelson 1994) and their tremendous morphological diversity might be due to a genome duplication event specific to the teleost lineage (Amores et al. 1998; Wittbrodt et al. 1998; Meyer and Schartl 1999). Since gene and genome duplication events increase the amount of genetic material that may be necessary for increasing the genomic and phenotypic complexity of organisms (Ohno 1970; Sidow 1996; Holland 1998, 1999; Force et al. 1999; Lundin 1999; Naruse et al. 2000), it is tempting, albeit controversial (e.g. Aparicio 2000; Kappen 2000) to speculate on a cause-effect relationship between gene copy number and morphological complexity and/or species diversity. Wittbrodt et al. (1998) and Amores et al. (1998) have suggested that the

Correspondence to: Yves Van de Peer; *present address:* Department of Plant Genetics, University of Ghent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium; *email:* yvdp@gengenp.rug.ac.be

potentially more complex genomic architecture of fish might have permitted them to adapt and speciate quickly in response to changing environments. Many studies have indeed shown that speciation can occur very rapidly in fish, the most well-known case undoubtedly being the speciation of African cichlids (Meyer 1993; Stiassny and Meyer 1999). The hypothesis that genome duplication is responsible for increased phenotypic complexity predicts that duplicated genes have diverged from the roles played by their pre-duplication homologs. This divergence could be demonstrated by an increase in evolutionary rate or by evidence for positive Darwinian selection.

Indications for a fish-specific genome duplication first came from studies based on *Hox* genes and *Hox* clusters. *Hox* genes encode DNA-binding proteins that specify cell fate along the anterior-posterior axis of bilaterian animal embryos and they occur in one or more clusters of up to 13 genes per cluster (Gehring 1998). It is thought that the ancestral *Hox* gene cluster arose from a single gene by a number of tandem duplications. The observation that protostome invertebrates and the deuterostome cephalochordate *Amphioxus* possess a single *Hox* cluster while Sarcopterygia, a monophyletic group including lobe-finned fish such as the coelacanth and lungfishes, amphibians, reptiles, birds, and mammals, have four clusters, (Holland and Garcia-Fernandez 1996; Holland 1997) supports the hypothesis of two rounds of entire genome duplications early in vertebrate evolution. Recently, "extra" *Hox* gene clusters have been discovered in fish. Amores et al. (1998) described the existence of seven *Hox* clusters in zebrafish (*Danio rerio*). Similar results have also been described for medaka (*Oryzias latipes*), which has 22 *Hox* genes mapped on seven different linkage groups (Naruse et al. 2000) and for the African cichlid fish *Oreochromis niloticus,* which has at least six *Hox* gene clusters (Málaga-Trillo and Meyer 2001). In pufferfish (*Fugu rubripes*), five *Hox* clusters have been found (Amores et al. 2001) and there appear to be two A clusters (Aparicio et al. 1997). These data strongly point to a *Hox* cluster duplication in Actinopterygii (ray-finned fishes) that occurred before the divergence of zebrafish, medaka, and pufferfish, at least 100 Mya (Nelson 1994; Santini and Tyler 1999). However, criticisms of the teleost genome duplication hypothesis have focused on the fact that *Hox* genes reveal the history of only a small portion of the entire genome. Although many other multigene families have been described that have more genes in fish than in other vertebrates (Wittbrodt et al. 1998; see also Postlethwait et al. 2000), many (if not most) fish have much smaller genomes than humans (Ohno 1970). For example, Morizot et al. (1991) have estimated that the genome of the platyfish (genus *Xiphophorus*) is five times smaller than the human genome. Elgar et al. (1999) estimated that the pufferfish genome is eight times smaller than the human genome and suggested that the duplication of *Hox* clusters by regional duplication is easier to reconcile with fish genome size data than a complete genome duplication. However, the small genome of *Fugu* may be due to the reduction of intergenic and noncoding regions (Venkatesh et al. 2000). Furthermore, mapping data suggest that duplications are not limited to *Hox* clusters. Large chromosome segments or entire chromosomes appear to be duplicated (Amores et al. 1998; Force et al. 1999; Woods et al. 2000; Postlethwait et al. 2000).

The fish-specific genome duplication hypothesis predicts that fish have more genes than other vertebrates that do not share this genome duplication. Recently, we surveyed more than 220 different genes and found 26 cases where one human gene has two orthologs in zebrafish (Taylor et al. 2001). Furthermore, the zebrafish paralogs showed sister relationships in phylogenetic trees (Fig. 1), and seemed to have been formed at the same time, somewhere between 300 and 450 Mya. Combined with the observation that the zebrafish paralogs were found on many different linkage groups and showed conserved synteny with other genes (Barbazuk et al. 2000; Postlethwait et al. 2000), this provides strong support for a fish-specific genome duplication rather than many independent tandem duplications (see also Amores et al. 1998; Meyer and Schartl 1999; Gates et al. 1999; Taylor et al. 2001).

Duplicated genes may be redundant, which means that inactivation of one of the two duplicates should have little or no effect on the phenotype (Nowak et al. 1997; Gibson and Spring 1998; Lynch and Conery 2000). Therefore, since one of the copies is freed from functional constraint, mutations in this gene will be selectively neutral and will eventually turn the gene into a non-functional pseudogene. On the other hand, by chance, a series of non-deleterious mutations might turn the duplicate gene into a gene with a new function (Ohno 1973). Ohno's model, which Hughes (1994) first called the "mutation during non-functionality" (MDN) model and later the "mutation during redundancy" (MDR) model (Hughes 1999), has been widely adopted as an explanation for the evolution of functionally novel genes. However, many have criticized the MDR model and numerous other models have been put forward to explain the retention and functional divergence of genes (Hughes 1994, 1999; Walsh 1995; Nowak et al. 1997; Gibson and Spring 1998; Wagner 1998; Force et al. 1999).

Gibson and Spring (1998) argued that selection can prevent the loss of redundant genes (i.e., duplicates) if those genes code for components of multidomain proteins because mutant alleles disrupt such proteins. Hughes (1994) and Force et al. (1999) argued that when a gene with multiple function is duplicated, the duplicates are redundant only for as long as each retains the ability to perform all ancestral roles. When one duplicate experiences a mutation that prevents it from carrying out

one of its ancestral roles, the other duplicate is no longer redundant. This is consistent with Sidow's (1996) proposition that a single unique function in an ocean of redundancy is enough to keep the gene afloat and prevent degenerative substitutions. According to Force et al.'s (1999) duplication-degeneration-complementation (DDC) model, degenerative mutations preserve rather than destroy duplicated genes but also change their functions or at least restrict their functions to become more specialized.

Also, positive Darwinian selection can be responsible for functional divergence between the duplicates (e.g. Zhang et al. 1998; Duda and Palumbi 1999; Hughes et al. 2000). Most studies that look for evidence of positive Darwinian selection compare the ratio of nonsynonymous (p_N) and synonymous (p_S) substitutions (Hughes 1999; Nei and Kumar 2000). Since most amino acid changes are disadvantageous, synonymous substitutions occur at a higher rate than nonsynonymous ones in most genes, due to purifying selection. Under neutral evolution, the rates of synonymous and nonsynonymous substitutions are expected to be equal (Kimura 1983). However, under positive Darwinian selection, natural selection favors amino acid replacements. As a result, nonsynonymous mutations get fixed at a faster rate than synonymous mutations (ratio p_N:p_S > 1), as has been shown for genes such as primate lysozyme genes (Messier and Stewart 1997), pregnancy-associated glycoprotein genes (Hughes et al. 2000), primate ribonuclease genes (Zhang and Nei 2000), conotoxin genes (Duda and Palumbi, 1999), and many more (Endo et al. 1996 and references therein). Unfortunately, the ratio of nonsynonymous over synonymous mutations can only be demonstrated to be higher than one for recently duplicated genes (Hughes 1999; Kumar and Nei 2000). Once the gene has adapted to its specific function, purifying selection is expected to predominate, allowing the number of synonymous substitutions per site to catch up and eventually exceed the number of nonsynonymous substitutions per site. Therefore, positive Darwinian selection cannot be detected 30–50 million years after gene duplication (Hughes 1999; Hughes et al. 2000). Since the zebrafish paralogs are between 300 and 450 million years old (Taylor et al. 2001), positive selection cannot be detected by considering the p_N:p_S ratio. However, Hughes et al. (1990) developed an alternative method for testing whether sequences have been subjected to positive Darwinian selection by evaluating whether nonsynonymous mutations occur in such a way as to change protein charge or polarity to a greater extent than is expected under random substitution. This method involves the computation of the proportion of radical nonsynonymous difference (p_{NR}) per radical nonsynonymous site versus the proportion of conservative nonsynonymous difference per conservative nonsynonymous site (p_{NC}). When $p_{NR} > p_{NC}$, nonsynonymous differences occur in such a way as to change the property of interest to a greater extent than expected at random. Since this method looks at nonsynonymous sites only and the resulting amino acid changes, the occurrence of positive Darwinian selection should be evident for a much longer period. It should be noted though that this method might be less sensitive to detect positive selection than looking at the ratio p_N:p_S (Vacquier et al. 1996; Hughes 1999), but as mentioned before, it is the only option available for anciently diverged genes.

In the current study, pairs of anciently duplicated zebrafish genes were studied to see whether evidence could be found for positive Darwinian selection and whether the retention of genes and the evolution of novel gene functions is supported by some of the previous models proposed. A relative rate test was also applied to see whether one of the duplicates evolved at a faster rate after duplication, which could point to either positive Darwinian selection or relaxed functional constraint on one of the duplicates.

Materials and Methods

Sequence data. Detailed information on the identification and retrieval of the duplicated zebrafish sequences can be found in Taylor et al. (2001). Two different alignments were created for the genes listed in Table 1. The first alignment contained the two zebrafish paralogs plus their human ortholog, while the second alignment contained only the two zebrafish paralogs. Sequences were first aligned at the amino acid level using CLUSTAL_X (Thompson et al. 1997). Then, the corresponding nucleic acid sequences were collected and aligned using the amino acid alignments as guides. Editing of the alignments was done with the Bioedit sequence editor (http://www.mbio.ncsu.edu/RNaseP/info/programs/BIOEDIT/bioedit.html). Only regions that could be unambiguously aligned were used for further analyses and whenever there was doubt about amino acids being homologous, they were removed from the alignment. Accession numbers of the sequences used can be found in Taylor et al. (2001). Sequence alignments will be made available on our website (http://www.evolutionsbiologie.uni-konstanz.de/). The nucleotide sequence for one of the *Gdf6* zebrafish paralogs is not available.

Positive Darwinian selection. We looked for evidence for positive Darwinian selection using the program SCR3 (Hughes et al. 1990). This program tests whether nonsynonymous mutations occur in such a way as to change some amino acid property of interest to a greater extent than is expected under random substitution (neutral evolution). Each nonsynonymous site is defined as either conservative or radical. Conservative substitutions will lead to an amino acid replacement that is conservative with respect to charge or polarity, while a substitution at a radical site will lead to an amino acid replacement that changes charge or polarity (Hughes et al. 1990). Although the estimation of nucleotide substitutions can be sensitive to the relative rates of occurrence of transitions and transversions, we have used the default transition:transversion ratio of 0.5 for all comparisons. The alternative of using a transition:transversion ratio estimated from fourfold degenerate positions (Hughes et al. 2000; Lynch and Conery 2000), is unlikely to give a better estimate since these positions are saturated when the two zebrafish paralogs are compared.

Relative rate tests. To investigate whether one of the two zebrafish paralogs evolved at a faster rate since their duplication, a relative rate

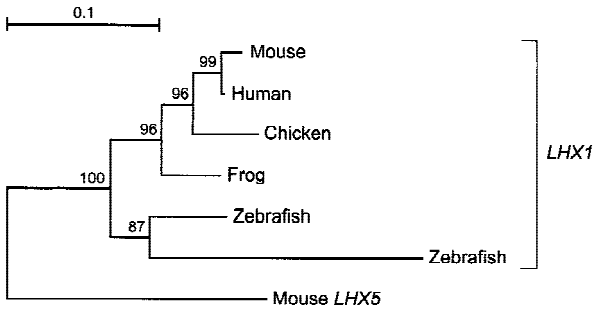

Fig. 1. Neighbor-joining (Saitou and Nei 1987) tree of the *Lhx1* genes, showing a sister-group relationship for the zebrafish genes. This tree is part of a larger tree including many other members of the *Lhx* gene family (Taylor et al. 2001). Evolutionary distances were computed according to Kimura (1983). An identical tree topology is obtained with maximum likelihood and maximum parsimony. Bootstrap values (Felsenstein 1985) above 50%, out of 500 replications, are indicated.

test was applied to each of the genes. We applied the nonparametric rate test (Nei and Kumar 2000) developed by Tajima (1993) and implemented in MEGA2, and compared the zebrafish paralogs with their human ortholog. The advantage of using a nonparametric test is that the results are not influenced by the choice of a, possibly wrong, substitution model (Nei and Kumar 2000). The rate test of Tajima compares two sequences with an outgroup sequence and counts the number of unique substitutions in both lineages. When both genes evolve under the molecular clock hypothesis, both genes are expected to have accumulated a similar number of 'unique' substitutions (Tajima 1993; Nei and Kumar 2000). On the other hand, when one of the duplicates has accumulated a significantly larger number of substitutions, the molecular clock does not apply and one of the paralogs is inferred to have experienced an increased evolutionary rate. The relative rate test applied in this study was based both on the amino acid sequences and on the corresponding nucleic acid sequences using first and second codon positions. Third codon positions are saturated when zebrafish and human sequences are compared (Taylor et al. 2001, see also Fig. 3).

Results and Discussion

Positive Darwinian Selection

Initially, we looked for positive selection using the complete gene sequence. However, since substitutions in least variable regions are bound to be more conservative and would tend to 'dilute' the signal of selection (Endo et al. 1996; Hughes et al. 2000), we also looked for positive Darwinian selection using only the variable regions of the gene. Variable regions were selected on the basis of amino acid differences between the two zebrafish paralogs (Fig. 2). It should be noted though that for many genes it was difficult to discriminate between conserved and variable regions since variable sites are scattered over the whole sequence. In such cases, the analysis involved the complete gene.

As can be seen in Table 1, no genes show evidence for positive Darwinian selection ($p_{NR} > p_{NC}$) regarding charge. For change in polarity, all three *Hox* genes show signs of positive Darwinian selection, although the value for *HoxC6* is close to 1, which would point to neutrality. A p_{NR}:p_{NC} value of approximately 1 is also true for *Lhx1* and *Otx1*. Neutrality would imply relaxation of functional constraint on these genes, or at least of part of the genes. When analyzing the more variable parts of these genes, additional evidence for positive Darwinian selection regarding polarity was found for the *Pax2* gene ($p_{NC} = 0.275$, $p_{NR} = 0.284$, p_{NR}:$p_{NC} = 1.033$). For all other genes, the p_{NR}:p_{NC} ratio is much lower than 1, which provides strong evidence for purifying selection, that is keeping properties such as charge and polarity the same in both duplicates. Although we do not find evidence for positive Darwinian selection for most genes, this does not rule out the possibility of positive Darwinian selection shortly after the genome duplication. Although often p_N:p_S and p_{NR}:p_{PC} ratios are correlated (e.g. Hughes et al. 1990; Hughes et al. 2000), Vacquier et al. (1997) found strong evidence for positive Darwinian selection in two homologous fertilization proteins if they looked at p_N:p_S ratios, but not if they looked at p_{NR}:p_{NC} ratios. They concluded that, although positive Darwinian selection had taken place, structural constraints were responsible for conservation in both charge and polarity. The high ratios of p_{NC}:p_{NR} we observe for the zebrafish duplicates (Table 1) seem to suggest that structural constraints keep them from changing their charge or polarity.

Accelerated Rates of Evolution

To examine whether one of the duplicates evolved at a faster rate after the duplication, a relative-rate test was performed (see Materials and Methods) comparing each zebrafish duplicate to its human orthologue (Table 2). Thirteen duplicates do not show a statistically significant increase in rate of evolution. However, thirteen genes do show an increased rate in one of the zebrafish paralogs. This is most obvious for *Fkd5, Hoxb5, Hoxb6, Lhx1* (see also Fig. 1 and Fig. 3 in Taylor et al. 2001), *Ntn1, Otx1,* and *Rxr*β, where an increased rate is detected both on the amino acid level and on the basis of first and second codon positions. For the *Hox* genes, and maybe also for the other homeobox containing genes *Lhx1* and *Otx1,* the increase in rate may be the result of positive Darwinian selection, as shown in Table 1. For *Brn1, En1,* and *Ephb4,* an increase in rate is statistically significant when the amino acid sequences are compared, while for *Dll1* and *Isl2,* an increase in evolutionary rate is only detected on the basis of first and second codon positions. The latter finding may be surprising, but as can be seen in Table 2, the difference in unique amino acid substitutions for *Dll1* is close to significance ($p = 0.065$). For *Isl2* this value is not statistically significant due to the small number of unique substitutions in both zebrafish paralogs.

Besides positive Darwinian selection, increases in the

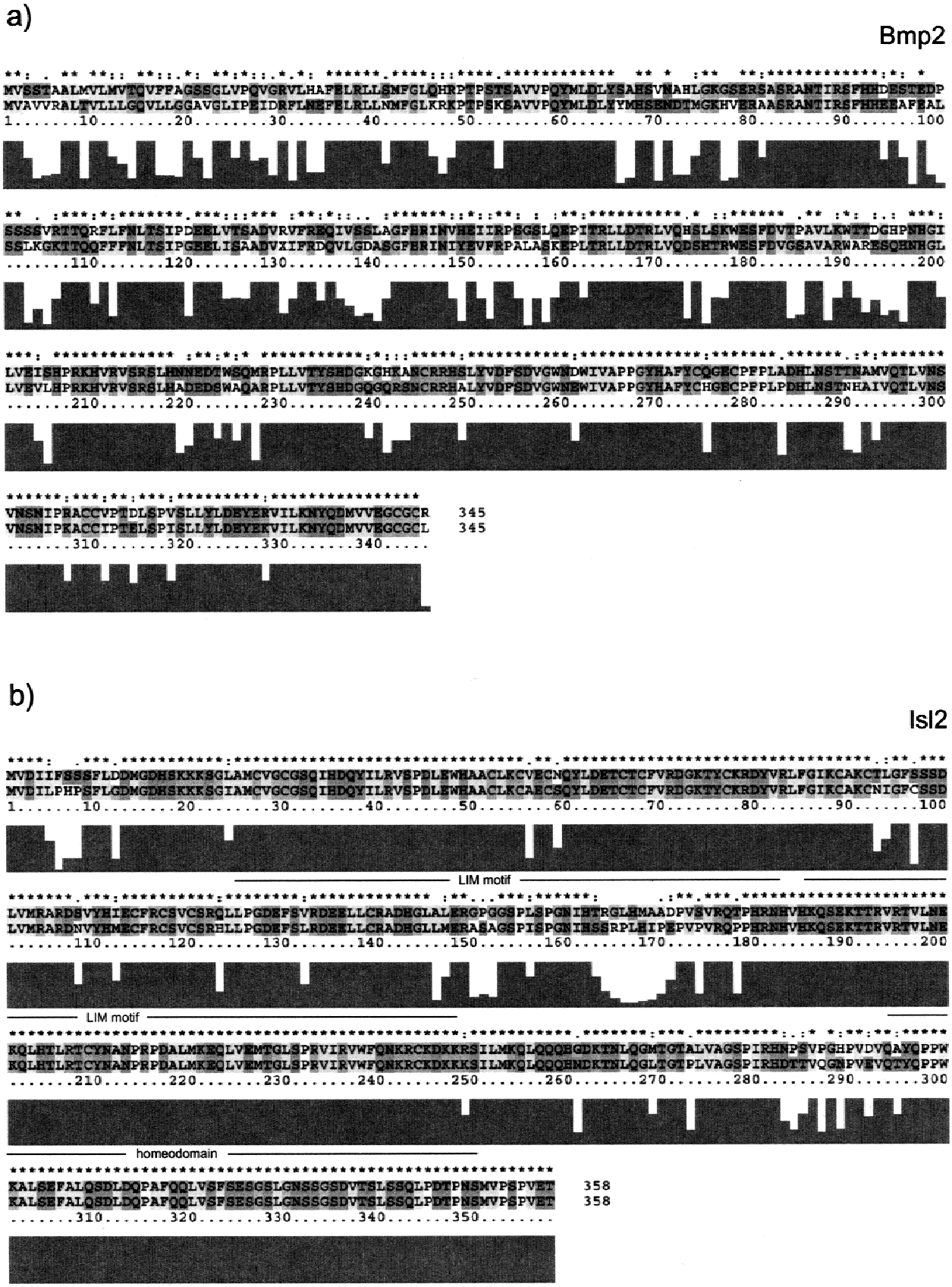

Fig. 2. **(A)** Amino acid variability map for *Bmp2* zebrafish paralogs. **(B)** Amino acid variability map for *Isl2* zebrafish paralogs. The highly conserved homeodomain and the 2 LIM domains (Freyd et al. 1990; Karlsson et al. 1990; Toyama and Dawid 1997) are indicated. Asterisks above the alignment indicate positions that have a single, conserved residue. A double point indicates that one of the following "strong" groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, and FYW. A single point indicates that one of the following "weaker" groups is conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY (Thompson et al. 1997). Variability maps for the other zebrafish paralogs can be found at *http://www.evolutionsbiologie.uni-konstanz.de/.*

Table 1. Ratio of radical and conservative amino acid changes in zebrafish paralogs. P_S is the number of synonymous mutations per synonymous site; p_N is the number of nonsynonymous substitutions per nonsynonymous site; p_{NC} is the proportion of conservative nonsynonymous difference per conservative nonsynonymous site; p_{NR} is the proportion of radical (non-conservative) nonsynonymous difference per radical nonsynonymous site. Genes that show a ratio p_{NR}:p_{NC} > 1 are indicated in bold

			Charge		Polarity		p_{NR}/p_{NC}	
Gene[a]	p_S[b]	p_N	p_{NC}	p_{NR}	p_{NC}	p_{NR}	Charge	Polarity
Bmp2	0.679	0.198	0.235	0.145	0.221	0.159	0.616	0.718
Brn1	0.749	0.058	0.074	0.037	0.073	0.033	0.493	0.457
Dll1	0.772	0.125	0.149	0.089	0.144	0.090	0.598	0.619
Dlx2	0.760	0.154	0.188	0.104	0.174	0.120	0.553	0.692
En1	0.713	0.083	0.121	0.040	0.100	0.053	0.335	0.535
En2	0.766	0.106	0.147	0.061	0.111	0.098	0.414	0.876
Eph-b4	0.640	0.175	0.213	0.116	0.193	0.147	0.546	0.761
Fkd5	0.690	0.093	0.104	0.076	0.110	0.065	0.727	0.589
Flot1	0.632	0.056	0.075	0.033	0.067	0.036	0.442	0.542
Hoxb5	0.597	0.086	0.141	0.010	0.077	0.105	0.072	**1.375**
Hoxb6	0.674	0.196	0.245	0.129	0.183	0.221	0.528	**1.210**
Hoxc6	0.685	0.143	0.194	0.074	0.141	0.149	0.380	**1.063**
Isl2	0.774	0.074	0.104	0.034	0.090	0.045	0.322	0.506
Jak2	0.693	0.161	0.177	0.141	0.165	0.153	0.798	0.924
Lhx1	0.783	0.117	0.166	0.052	0.116	0.119	0.311	**1.032**
Msx3	0.854	0.107	0.125	0.079	0.140	0.054	0.636	0.384
Ntn1	0.653	0.079	0.104	0.047	0.098	0.044	0.449	0.452
Otx1	0.707	0.152	0.202	0.072	0.147	0.162	0.354	**1.106**
Pax2	0.668	0.040	0.060	0.011	0.040	0.040	0.174	0.985
*Rar*α	0.704	0.042	0.053	0.027	0.049	0.030	0.500	0.609
*Rxr*β	0.677	0.089	0.125	0.041	0.105	0.061	0.328	0.579
Shh	0.798	0.162	0.195	0.116	0.204	0.089	0.597	0.436
Sna(*il*)	0.509	0.076	0.101	0.041	0.085	0.062	0.408	0.723
Snap25	0.611	0.054	0.067	0.038	0.054	0.054	0.576	1.000
Sox11	0.573	0.076	0.111	0.030	0.092	0.046	0.270	0.504

[a] The name of the gene used is the name of the human gene. This may differ from the name given to the zebrafish gene (Taylor et al. 2001).
[b] The high number of synonymous mutations per synonymous site points to saturation of synonymous substitutions when both zebrafish paralogs are compared (see also Fig. 3).

evolutionary rate in one copy could be explained by the classical (Ohno 1997) or MDR model of gene evolution, which predicts that one copy will evolve more rapidly at nonsynonymous sites than the other, due to redundancy. The widely accepted classical model (Ohno 1973) predicts that neutral mutations can turn a duplicated gene into a pseudogene, or alternatively, by chance a series of mutations, can alter a gene sufficiently to take on a new function. Therefore, it seems plausible to assume that genes with a statistically significant increase in substitution rates have been subjected to relaxed functional constraints, while genes where none of the duplicates seems to have undergone accelerated substitution rates (see Table 2) have been subjected to purifying selection and thus, were not free to accumulate substitutions at random (Hughes 1999). However, in this case it remains difficult to explain why relaxation or increase in evolutionary rate in one of the duplicates has occurred in zebrafish, but not in *Xenopus*. Comparison of DNA sequences of duplicate genes of the tetraploid frog *Xenopus laevis* showed no evidence for positive Darwinian selection or increase in evolutionary rate (Hughes and Hughes 1993; Hughes, personal communication). One might argue that rate differences in *Xenopus* are not detected because the time of divergence was too recent (about 30 million years) by comparison with the duplications in zebrafish, but this is unlikely because gene silencing due to relaxation is expected to happen within a few million years (Nowak et al. 1997; Lynch and Conery 2000). Furthermore, if functional divergence of genes occurred according to the MDR model, one would expect to see acceleration throughout the complete gene, since mutation is random, for third codon positions, as shown in Figs. 3a and 3b. At the amino acid level, this is clearly not the case (Fig. 2B). The only other explanation for the increased rate in one of the duplicates is positive Darwinian selection. If there was positive Darwinian selection for a functional change, accumulation of substitutions would be predicted to occur only in the domains relevant to that function. Nevertheless, as discussed in the previous section, it might be very difficult to find clear traces of positive Darwinian selection, when the duplication is ancient.

Redundancy or Functional Divergence?

For genes that show a faster evolutionary rate and/or evidence for positive Darwinian selection one might ex-

Table 2. Results of the nonparametric relative rate test of Tajima (1993) comparing the two zebrafish paralogs with their human ortholog. Genes that show a statistically significant increase in rate of evolution in one of the duplicates are indicated in **bold.** Substitutions were computed for first and second codon positions and for amino acids

	Nucleotide sequence (1st and 2nd)					Amino acid sequence				
Gene	Sites	m1[a]	m2[b]	χ^2	Significant[c]	Sites	m1	m2	χ^2	Significant
Bmp2	691	56	60	0.14	No ($p = 0.710$)	344	39	29	1.47	No ($p = 0.225$)
Brn1	826	25	26	0.02	No ($p = 0.889$)	**412**	**23**	**9**	**6.13**	**Yes ($p = 0.013$)**
Dll1	**1220**	**82**	**42**	**12.90**	**Yes ($p = 0.000$)**	610	51	34	3.40	No ($p = 0.065$)
Dlx2	368	17	24	1.20	No ($p = 0.274$)	184	10	16	1.38	No ($p = 0.239$)
En1	234	12	7	1.32	No ($p = 0.251$)	**117**	**11**	**1**	**8.33**	**Yes ($p = 0.004$)**
En2	362	19	21	0.10	No ($p = 0.752$)	180	13	9	0.73	No ($p = 0.394$)
Eph-b4	1616	89	110	2.22	No ($p = 0.137$)	**808**	**32**	**54**	**5.63**	**Yes ($p = 0.018$)**
Flot1	428	8	16	2.67	No ($p = 0.102$)	214	7	12	1.32	No ($p = 0.251$)
Fkd5	**552**	**8**	**41**	**22.22**	**Yes ($p = 0.000$)**	**276**	**8**	**30**	**12.74**	**Yes ($p = 0.000$)**
Gdf6[d]						**303**	**12**	**27**	**5.77**	**Yes ($p = 0.016$)**
Hoxb5	**488**	**9**	**31**	**12.1**	**Yes ($p = 0.001$)**	**243**	**6**	**23**	**9.97**	**Yes ($p = 0.002$)**
Hoxb6	**430**	**18**	**54**	**18.00**	**Yes ($p = 0.000$)**	**215**	**7**	**29**	**13.44**	**Yes ($p = 0.000$)**
Hoxc6	394	24	26	0.08	No ($p = 0.777$)	197	18	16	0.12	No ($p = 0.732$)
Isl2	**718**	**35**	**20**	**4.09**	**Yes ($p = 0.043$)**	358	17	9	2.46	No ($p = 0.117$)
Jak2	942	76	57	2.71	No ($p = 0.099$)	471	42	31	1.66	No ($p = 0.198$)
Lhx	**782**	**23**	**67**	**21.5**	**Yes ($p = 0.000$)**	**391**	**10**	**54**	**30.25**	**Yes ($p = 0.000$)**
Msx3	262	13	16	0.31	No ($p = 0.577$)	130	4	8	1.33	No ($p = 0.248$)
Ntn1	**1188**	**25**	**56**	**11.86**	**Yes ($p = 0.001$)**	**593**	**13**	**44**	**16.86**	**Yes ($p = 0.000$)**
Otx1	**358**	**13**	**47**	**19.27**	**Yes ($p = 0.000$)**	**179**	**6**	**30**	**16.00**	**Yes ($p = 0.000$)**
Pax2	772	26	16	2.38	No ($p = 0.123$)	385	12	8	0.80	No ($p = 0.371$)
*Rar*α	772	15	23	1.68	No ($p = 0.194$)	385	5	12	2.88	No ($p = 0.090$)
*Rxr*β	**780**	**21**	**41**	**6.45**	**Yes ($p = 0.011$)**	**390**	**8**	**31**	**13.56**	**Yes ($p = 0.000$)**
Shh	784	49	61	1.31	No ($p = 0.253$)	391	18	29	2.57	No ($p = 0.109$)
Sna(*il*)	318	10	7	0.53	No ($p = 0.467$)	159	6	4	0.40	No ($p = 0.527$)
Snap25	241	4	4	0.00	No ($p = 1.000$)	120	2	8	3.60	No ($p = 0.058$)
Sox11	524	20	17	0.24	No ($p = 0.622$)	261	13	15	0.14	No ($p = 0.705$)

[a] m1 is the number of unique substitutions in zebrafish paralog 1.
[b] m2 is the number of unique substitutions in zebrafish paralog 2.
[c] Significant at the 95% confidence level ($p < 0.05$).
[d] The nucleotide sequence for one of the zebrafish genes is not available.

pect divergence in function. Contrarily, 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 the same purifying selection and, therefore, may have a similar function or even be completely redundant. Although it might seem unlikely that genes perform completely redundant functions after at least 300 million years of evolution, redundancy has been shown to be widespread in genomes of higher organisms (Nowak et al. 1997 and references therein; Gibson and Spring 1998). For example, *En1* and *En2* are both homologs of the *Drosophila* segmentation gene *engrailed.* Knock-out experiments in mice have shown that *En2* is functionally redundant since mice where the *En2* homeodomain was deleted showed no obvious defects in embryonic development (Joyner et al. 1991).

Remarkably, many of the retained duplicated genes turn out to be transcription factors (Table 3). In this respect, it is important to note that in a study listing a large number of genes for which multiple copies were found in vertebrates for a single *Drosophila* gene, many of them were transcription factors as well (Spring 1997). Is it a coincidence that mainly duplicated transcription factor genes have been retained? Our survey of genes may have been biased. Since the zebrafish is a model organism for developmental biologists, developmental control genes such as transcription factors may be the genes most studied and sequenced. On the other hand, the retention may be due to the fact that, once duplicated, these genes are harder to get rid of. It is not hard to imagine that mutations in a DNA-binding domain of a transcription factor could negatively effect the expression of genes, even when the original copy of the transcription factor is still present.

Gibson and Spring (1998) have suggested that alteration of a single domain in a multidomain protein might lead to nonfunctional complexes that exhibit a so-called dominant-negative phenotype. Their model is based on the observation that, for several genes, point mutations lead to a much more severe phenotype than when the (duplicated) gene is simply knocked out. In this case, one would expect selection against deleterious point mutations resulting in the retention of the genes. As a matter of fact, the gene is not only retained, it is also kept redundant. Although in the strict sense transcription fac-

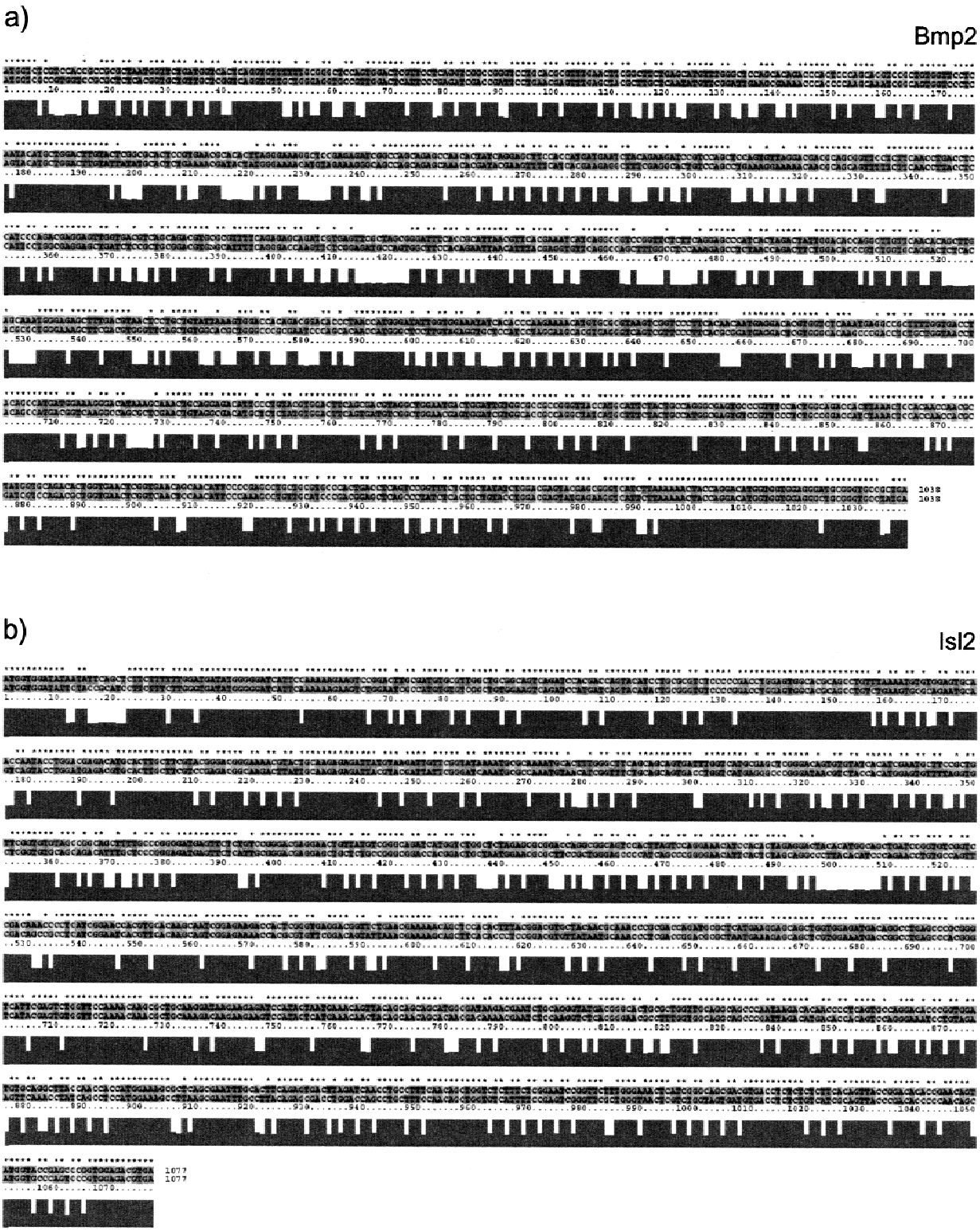

Fig. 3. **(A)** Nucleotide sequence variability map for *Bmp2* zebrafish paralogs. **(B)** Nucleotide sequence variability map for *Isl2* zebrafish paralogs.

tors are not multidomain proteins (Gibson and Spring 1998), many of them can bind to DNA as homodimers or heterodimers. If one gene copy receives a mutation that makes it nonfunctional, for example by affecting its DNA-binding capacity, it will still be able to dimerize, but transcriptional regulation will no longer be possible. This has been demonstrated for the *Drosophila* paired gene, which is homologous to the mammalian *Pax* gene. Point mutations that disrupt DNA binding have a dominant-negative effect on the normal gene regulation (Miskiewics et al. 1996). However, apart from *Pax,* not many dominant-negative mutations have been described

Table 3. Anciently duplicated genes in zebrafish and their gene products

Gene	Gene product
Bmp2	Secreted signaling protein
Brn1	DNA-binding protein
Dll1	DNA-binding protein
Dlx2	DNA-binding protein
En1	DNA-binding protein
En2	DNA-binding protein
Eph-b4	Receptor tyrosine kinase
Fkd5	DNA-binding protein
Flot1	Membrane-associated protein
Gdf6	Secreted signalling protein
Hoxb5	DNA-binding protein
Hoxb6	DNA-binding protein
Hoxc6	DNA-binding protein
Isl2	DNA-binding protein
Jak2	Non receptor tyrosine kinase
Lhx	DNA-binding protein
Msx3	DNA-binding protein
Ntn1	Secreted protein
Otx1	DNA-binding protein
Pax2	DNA-binding protein
*Rar*α	DNA-binding protein
*Rxr*β	DNA-binding protein
Shh	Secreted signalling protein
Sna(*il*)	DNA-binding protein
Snap25	Vesicular protein
Sox11	DNA-binding protein

for transcription factors and it is, therefore, questionable whether the model of Gibson and Spring can explain the retention of most of the zebrafish duplicates. Furthermore, most of the zebrafish transcription factor duplicates are probably not redundant in the strict sense (e.g. Ekker et al. 1997; Force et al. 1999).

Two types of regulatory mutations might change the function of transcription factors; those that alter the DNA-binding domain such that the set of genes with which it interacts is affected, and those that affect the expression of the gene, for example by a change in the DNA sequence of a *cis* regulatory element (Sidow 1996). For all the zebrafish genes described here that contain a DNA-binding domain (Table 3), this domain is extremely conserved (e.g. Fig. 2b). However, for some of these genes, such as *En1* (Force et al. 1999) and *Msx* (Ekker et al. 1997), the expression pattern or time of expression differs considerably. Therefore, if both paralogs have diverged in function, this is most likely due to mutations that affect the *cis* regulatory elements and not the structure of the gene product itself (Hughes 1994; Sidow 1996).

Force et al. (1999) and Lynch and Force (2000) have recently introduced another model, called the duplication-degeneration-complementation (DDC) model, to explain why duplicate genes might be retained. This model predicts that the likelihood of preservation is correlated with the number of 'subfunctions' that can be ascribed to a gene. The model starts from the assumption that a gene can perform several different functions, e.g. expression in different tissues and at different times during development, each of which may be controlled by different DNA regulatory elements. Several studies have shown that this is the case (Hughes 1994; Kirchhamer et al. 1996; Arnone and Davidson 1997). If duplicate genes lose different regulatory subfunctions, each affecting different spatial and/or temporal expression patterns, then they must complement each other by jointly retaining the full set of subfunctions present in the ancestral gene. Therefore, degenerative mutations facilitate the retention of duplicate functional genes, where both duplicates now perform different but necessary subfunctions. However, as predicted by the DDC model, the sum of the retained duplicates has to be equal to the total number of subfunctions performed by the ancestral gene. Before the DDC model, Hughes (1994) described a model that starts from the same assumptions, namely gene sharing, in which a single gene performs different functions. Gene duplication then allows each daughter gene to specialize for one of the functions of the ancestral genes. Force et al. (1999) showed that this might be the case for the *En1* genes in zebrafish. In mouse and chicken, *En1* is expressed in the developing pectoral appendage bud and in specific neurons of the hindbrain and spinal cord (Joyner and Martin 1987; Davis et al. 1991; Gardner and Barald 1992). In zebrafish, however, one of the paralogs is expressed in the pectoral appendage bud, while the second paralog is expressed in the hindbrain/spinal cord neurons (Force et al. 1999).

Possibly, retention of gene duplicates by subfunctionalization applies to many of the genes described in this study. Apart from *En1,* differences in the expression pattern of *Msx* zebrafish paralogs and homologous genes of other vertebrates also suggest subfunctionalization of the zebrafish genes after duplication (Ekker et al. 1997). Similar conclusions can be drawn for hedgehog genes (Laforest et al. 1998) and *Bmp2* (Martinez-Barbera et al. 1997). Overall, in order to determine whether one of the zebrafish paralogs is truly redundant, or whether subfunctionalization explains the retention and functional divergence of the duplicates, one has to identify mutants in differentially expressed duplicates.

Is the Classical Model Outdated?

Models such as the DDC model of Force et al. (1999) might explain retention and functional divergence of anciently duplicated genes. However, when subfunctionalization is responsible for the functional divergence of genes, this is probably limited to differences in timing and tissue specificity of expression. So far, there is little evidence that the paralogs described in this paper have completely different functions. In several cases we do have a statistical increase in evolutionary rate in one of the duplicates but this is probably not due to relaxed

functional constraints of the whole gene, as predicted by the MDR model. Although duplicated genes probably do experience a brief period of relaxed selection after duplication (Lynch and Conery 2000), the duplicates that are being retained are more likely to experience strong purifying selection (Table 1; see also Hughes and Hughes 1993; Hughes 1994; Lynch and Conery, 2000).

That duplicated genes can evolve previously nonexistent functions has been previously demonstrated. Expansion of repetitive regions in one copy of a duplicated pancreatic trypsinogen-like gene produced a gene for antifreeze glycoproteins in Antarctic fish (Cheng and Chen 1999) and mutations in duplicated opsin genes led to the evolution of trichromatic vision in new and old world primates (Dulai et al. 1999). However, it is questionable whether one would still be able to identify such duplicates after 300 million years of independent evolution. If the MDR model explained the functional divergence of duplicates these duplicates may be hard to find in database searches based on sequence similarity (e.g. BLAST). Although the MDR model does not seem to account for the majority of novel gene functions of duplicated genes (Hughes 1999), it is possible that a fraction of them have evolved beyond recognition. One possible solution to this problem is to consider synteny data. Studies investigating conserved synteny between different chromosomes might allow us to recognize regions of chromosomes that are paralogous and help identify duplicate genes that share little sequence similarity.

Acknowledgments. The authors acknowledge support from the German Science Foundation (DFG PE 842/2-1), the University of Konstanz, the Verband der Chemischen Industrie, and NSERC (Canada) for a Postdoctoral fellowship to JST. The authors also thank Austin Hughes for providing us with the SCR3 program and for useful discussions. Gerrit Begemann is acknowledged for helpful suggestions and comments on an earlier draft of the paper. YVdP is Research Fellow of the National Fund for Scientific Research–Flanders (Belgium).

References

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

Amores A, Amemiya CT, Postlethwait J (2001) Genome duplication and evolution of *Hox* clusters in teleosts. Am Zool (in press)

Aparicio S (2000) Vertebrate evolution: recent perspectives from fish. Trends Genet 16:54–56

Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S (1997) Organization of the *Fugu rubripes Hox* clusters: evidence for continuing evolution of vertebrate *Hox* complexes. Nat Genet 16:79–83

Arnone MI, Davidson EH (1997) The hardwiring of development: organization and function of genomic regulatory systems. Development 124:1851–1864

Barbazuk WB, Korf I, Kadavi C, Heyen J, Tata S, Wun E, Bedell JA, McPherson JD, Johnson SL (2000) The syntenic relationships of the zebrafish and human genomes. Genome Research 10:1351–1358

Cheng C-HC, Chen L (1999) Evolution of an antifreeze glycoprotein. Nature 401:443–444

Cyranoski D, Smaglik P (2000) Puffer fish joins genome stampede. Nature 408:6–7

Davis CA, Homyard DP, Millen KJ, Joyner AL (1991) Examining pattern formation in mouse, chicken and frog embryos with an *En*-specific antiserum. Development 2:287–298

Duda TF, Palumbi SR (1999) Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus.* Proc Natl Acad Sci USA 96:6820–6823

Dulai KS, von Dornum M, Mollon JD, Hunt DM (1999) The evolution of trichromatic colour vision by opsin gene duplication in new world and old world primates. Genome Res. 9:629–638

Ekker M, Akimenko MA, Allende ML, Smith R, Drouin G, Langille RM, Weinberg ES, Westerfield M (1997) Relationships among msx gene structure and function in zebrafish and other vertebrates. Mol Biol Evol 14:1008–1022

Elgar G, Clark MS, Meek S, Smith S, Warner S, Edwards YJK, Bouchireb N, Cottage A, Yeo GSH, Umrania Y, Williams G, Brenner S (1999) Generation and analysis of 25 Mb of genomic DNA from the pufferfish *Fugu rubripes* by sequence scanning. Genome Res 9:960–971

Endo T, Ikeo K, Gojobori T (1996) Large-scale search for genes on which positive selection may operate. Mol Biol Evol 13:685–690

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

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

Freyd G, Kim S, Horvitz RH (1990) Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gen *lin-11.* Nature 344:876–879

Gardner CA, Barald KF (1992) Expression patterns of engrailed-like proteins in the chick embryo. Dev Dyn 193:370–388

Gates MA, Kim L, Cardozo T, Sirotkin HI, Dougan ST, Lashkari D, Abagyan R, Schier AF, Talbot WS (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. Genome Research 9:334–347

Gehring WJ (1998) Master control genes in development and evolution: the homeobox story. Yale University Press, New Haven

Gibson TJ, Spring J (1998) Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. Trends Genet 14:46–49

Holland PW (1997) Vertebrate evolution: something fishy about *Hox* genes. Curr Biol 7:R570–R572

Holland PWH (1998) Major transitions in animal evolution: a developmental genetic perspective. Amer Zool 38:829–842

Holland PWH (1999) The effect of gene duplication on homology. In: Bock GR, Cardew G (eds) Homology. Wiley, Chichester, pp. 226–242

Holland PW, Garcia-Fernandez (1996) *Hox* genes and chordate evolution. Dev Biol 173:382–395

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B 256:119–124

Hughes AL (1999) Adaptive evolution of genes and genomes. Oxford University Press, New York

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol Biol Evol 7: 515–524

Hughes MK, Hughes AL (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis.* Mol Biol Evol 10:1360–1369

Hughes AL, Green JA, Garbayo JM, Roberts RM (2000) Adaptive diversification within a large family of recently duplicated, placentally expressed genes. Proc Natl Acad Sci USA 97:3319–3323

Joyner AL, Martin GR (1987) *En-1* and *En-2,* two mouse genes with

sequence homolog to the *Drosophila engrailed* gene: expression during embryogenesis. Genes Dev 1:29–38

Kappen C (2000) Analysis of a complete homeobox gene repertoire: implications for the evolution of diversity. Proc Natl Acad Sci USA 97:4481–4486

Karlsson O, Thor S, Norberg T, Ohlsson H, Edlund T (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. Nature 344:879–882

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kirchhamer CV, Yuh C-H, Davidson (1996) Modular *cis*-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples. Proc Natl Acad Sci USA 93:9322–9328

Laforest L, Brown CW, Poleo G, Geraudie J, Tada M, Ekker M, Akimenko M-A (1998) Involvement of the Sonic Hedgehog, patched 1 and *bmp2* genes in patterning of the zebrafish dermal fin rays. Development 125:4175–4184

Lundin L-G (1999) Gene duplications in early metazoan evolution. Cell Devel Biol 10:523–530

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

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473

Málaga-Trillo E, Meyer A (2001) Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. American Zool (in press)

Martinez-Barbera JP, Toresson H, Da Rocha S, Krauss S (1997) Cloning and expression of three members of the zebrafish *Bmp* family: *Bmp2a, Bmp2b* and *Bmp4.* Gene 198:53–59

Messier W, Stewart C-B (1997) Episodic adaptive evolution of primate lysosymes. Nature 385:151–154

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlids. Trends Ecol Evol 8:279–284

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

Miskiewicz P, Morissey D, Lan Y, Raj L, Kessler S, Fujioka M, Goto T, Weir M (1996) Both the paired domain and homeodomain are required for in vivo function of *Drosophila* paired. Development 122:2709–2718

Morizot DC, Slaugenhaupt SA, Kallman KD, Chakravarti A (1991) Genetic linkage map of fishes of the genus *Xiphophorous* (Teleostei: Poeciliidae). Genetics 127:399–410

Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, Shimada A, Wada H, Kusakabe T, Suzuki N, Kinoshita M, Kanamori A, Terado T, Kimura H, Nonaka M, Shima A (2000) A detailed linkage map of medaka, *Oryzias latipes:* comparative genomics and genome evolution. Genetics 154:1773–1784

Nelson JS (1994) Fishes of the world, 3rd edition. Wiley, New York

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Nowak MA, Boerlijst MC, Cooke J, Maynard Smith J (1997) Evolution of genetic redundancy. Nature 388:167–171

Ohno S (1970) Evolution by gene duplication. New York, Springer Verlag

Ohno S (1973) Ancient linkage groups and frozen accidents. Nature 244:259–262

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan Y-L, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res 10:1890–1902

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Santini F, Tyler JC (1999) A new phylogenetic hypothesis for the order Tetraodontiformes (Teleostei, Pisces), with placement of the most fossil basal lineages. Am Zool 39:10A

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Gen Devel 6:725–722

Spring J (1997) Vertebrate evolution by interspecific hybridisation—are we polyploid? FEBS Lett 400:2–8

Stiassny MLJ, Meyer A (1999) Cichlids of the African Rift Lakes. Scientific American February:64–69

Tajima F (1993) Simple methods for testing molecular clock hypothesis. Genetics 135:599–607

Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication in fish. Phil Trans Roy Soc (in press)

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Toyama R, Dawid IB (1997) *lim6,* A novel LIM homeobox gene in the zebrafish: comparison of its expression pattern with *lim1*. Dev Dyn 209:406–417

Vacquier VV, Swanson WJ, Lee Y-H (1997) Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? J Mol Evol 44:S15–S22

Venkatesh B, Gilligan P, Brenner S (2000) Fugu: a compact vertebrate reference genome. FEBS Lett 476:3–7

Wagner A (1998) The fate of duplicated genes: loss or new function? BioEssays 20:785–788

Walsh JB (1995) How often do duplicated genes evolve new functions? Genetics 139:421–428

Wittbrodt J, Meyer A, Schartl M (1998) More genes in fish? BioEssays 20:511–512

Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan Y-L, Huang H, Postlethwait JH, Talbot WS (2000) A comparative map of the zebrafish genome. Genome Res 10:1903–1914

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 95:3708–3713

Zhang J, Nei M (2000) Positive selection in the evolution of mammalian *interleukin-2* genes. Mol Biol Evol 17:1413–1416