*First publ. in: BioEssays 20 (1998), pp. 511–515*

# More genes in fish?

J. Wittbrodt,[1] A. Meyer,[2] and M. Schartl[3]*

Summary

Certain species of fish have recently become important model systems in comparative genomics and in developmental biology, in certain instances because of their small genome sizes (e.g., in the pufferfish) and, in other cases, because of the opportunity they provide to combine an easily accessible and experimentally manipulable embryology with the power of genetic approaches (e.g., in the zebrafish). The resulting accumulation of genomic information indicates that, surprisingly, many gene families of fish consist of more members than in mammals. Most modern fish, including the zebrafish and medakka, are diploid organisms; however, the greater number of genes in fish was possibly caused by additional ancient genome duplications which happened in the lineage leading to modern ray-finned fishes but not along the lineage leading to tetrapods. Since these two lineages shared their last common ancestor (in the Devonian about 360 million years ago) individual duplicated members of gene families were later lost in fish. Interestingly, comparative data indicate that, in some cases, genes in mammals even serve somewhat different functions than their homologues in fish, highlighting that the degree of evolutionary relatedness of genes is not always a reliable predictor of their evolutionary conservation and their similarity of function. Since fish are phenotypically probably not more complex than mammals, it is possible that evolution took alternative paths to the ''economics of genomics'' through alternative solutions to gene regulation. It is suggested that the more complex genomic architecture of fish permitted them to adapt and speciate quickly in response to changing selective regimes. *BioEssays* **20**:511–515, 1998. 

## Introduction

In recent years, fish have become mainstream models for biological research.[1–3] Among the many beneficial characteristics that have favored this development is the suitability of small ''laboratory fish'' species for the analysis of vertebrate development by a saturation mutagenesis approach, in particular, zebrafish. Another attractive feature and potential advantage that has attracted considerable interest is the small genome size of several fish species compared with humans. Considering that most important human genes are expected to be present and conserved in fish, it is felt that their compact genomes provide versatile and heuristic tools for genome projects. For instance, the Huntington chorea disease gene from the *Fugu* fish is 7.4 times smaller than its human counterpart,[4] and the Medakafish p450 aromatase gene spans only 2.5 kb, compared with 70 kb in humans,[5] solely due to smaller introns (despite a perfect conservation in exon sizes and arrangements). Unexpectedly, the steadily increasing database on characterized fish genes strongly suggests that many multigene families studied so far in fish contain more functional members than in mammals. The exception rather than the rule are examples like the odorant receptor gene family which is smaller in fish.[6,7] This means

[1]Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany; E-mail: jwittbr@gwdg.de
[2]Department of Biology, University of Konstanz, Konstanz, Germany; E-mail: Axel.Meyer@uni-konstanz.de
[3]Physiologische Chemie I, Biozentrum der Julius-Maximilians Universität, Würzburg, Germany; E-mail: phchl@biozentrum.uni-wuerzburg.de
*Correspondence to: M. Schartl, Physiologische Chemie I, Biozentrum der Julius-Maximilians Universität, 97074 Würzburg, Germany; E-mail: phchl@biozentrum.uni-wuerzburg.de

that fish have additional genes that, due to conserved structural features and phylogenetic analyses, can be unequivocally ascribed to a certain gene family (thus being true paralogues), although the same type of analysis fails to discover the direct mammalian counterparts (the orthologues).

### Observations suggesting divergent function despite highest sequence similarity

So far the identification of genes in fish has been done predominantly through the so called ''homology approach.'' Either molecular probes, derived for example from *Drosophila* or mouse genes, were used to clone sequences that cross-hybridized, or fish genes were found by other methods such as positional cloning or degenerate primer polymerase chain reaction (PCR). These fish genes were identified by virtue of their sequence similarity to known, related genes from other organisms. Sometimes, however, there are complications with this otherwise straightforward approach. In several cases, fish genes that appeared orthologous to certain genes from mammals, by the criterion of sequence similarity, did not have the same function in fish as that of, for example, mice or humans. The *echidna hedgehog* gene in zebrafish, originally postulated to represent an entirely new class of hedgehog genes,[8] appeared to be the orthologue of the indian hedgehog class of genes based on phylogenetic analysis.[9,10] The *echidna hedgehog* gene in zebrafish, however, and its orthologue, the indian *hedgehog* gene in mice, for example, appear to be quite different with respect to expression pattern and function.[8] This example illustrates that the function of putatively orthologous genes can, nonetheless, be different. For the lactate dehydrogenase (LDH) C isozymes in fish and mammals, it has been shown by phylogenetic analysis that they are not orthologous, but rather are derived from independent duplication events from the LDH A and B genes, respectively.[11] Genes with homeoboxes related to that of the *Drosophila muscle segment homeobox* (*msh*) gene, termed *msx* genes, have been isolated from mammals, birds, amphibians, and zebrafish. A detailed phylogenetic analysis of the sequences and studies on the temporal and spatial expression pattern of the individual gene family members led to the convincing conclusion that both structure and function of *msx* genes in fish and tetrapods diverged independently in the lineages giving rise to these organisms.[12] Although a given zebrafish *msx* gene may be most similar in sequence to a particular one of the mammalian genes, it will not necessarily be the orthologue and applying the concept of homology in order to interpret structural similarity or to extrapolate a known function from one taxonomic group to another is precluded in this case. A similar situation is found for *otx* genes in zebrafish and medaka. While the fish *otx2* is highly conserved in comparison to its mammalian counterpart, neither *otx1* nor *otx3* in fish can be unequivocally identified by sequence to be the orthologue of *otx1* in higher vertebrates[13,14] (F. Loosli and J. Wittbrodt, unpublished results). Comparison of the expression patterns of the fish and mouse genes strongly suggests that structure and function of the *otx* genes in fish and tetrapods also diverged independently, similar to the above-mentioned situation with *msx* genes.

### Evidence for larger gene families in teleosts

An even more striking complication when looking at fish genomes is the widespread observation that many multigene families potentially comprise more members in fish than in mammals (Table 1), where they were originally detected. For instance, the *msx* and *otx* gene families not only display functional differences between putative fish and mammalian orthologues but also have more paralogues in fishes than in mammals.[12,13] Another example of this kind is the activin group of transforming growth factor-β (TGF-β) like molecules. In fish, two different copies of the TGF-βA and -βB genes are found (R. Köster and J. Wittbrodt, unpublished observations), but only one copy of each gene exists in higher vertebrates. Allozyme studies had indicated previously that a number of metabolic enzymes may be encoded by more gene family members in fish than in mammals.[15] The number and function of homeobox (Hox) genes, their arrangement in clusters, and the precise phylogenetic timing of the duplications of the Hox genes have received considerable attention. From the single ancestral Hox cluster,[16] at least four clusters evolved and are known to exist in mammals. The phylogenetic question of when the duplications from the ancestral Hox cluster to four occurred during the evolutionary history of vertebrates has only recently become the focus of attention.[17–19] Importantly, recent data suggest that, once again, fish actually appear to have more genes than mammals. Prince et al.[20] discovered that zebrafish have at least two additional Hox gene clusters (termed X and Y) giving a total of at least six compared to the four Hox clusters (A–D) that seem to be prevalent in mammals. So far, four clusters have been found in *Fugu* and were structurally characterized in detail. While the A, B, and C clusters match nicely with their mammalian counterpart in respect to genomic organization and sequence similarity, the fourth cluster is enigmatic. It has fewer Hox genes than occur in the mammalian D cluster. It is possible that the bona fide Hox D of *Fugu* remains to be found and it has been suggested that the unusual cluster is an additional one.[21]

It is certainly too early, and our database on fish genes is still too small, to conclude that this is a general phenomenon. Studies on more ''basal'' fish (e.g., the coelacanth, the lungfishes, or the sturgeons) and ''comparative genomics'' of different teleosts is required to answer these questions. It

TABLE 1. Examples of Larger Multigene Families in Teleost Fish Compared With Higher Vertebrates

| Multigene family | No. of paralogue seqs. in higher vertebrates | Additional paralogue(s) in fish without higher vertebrate orthologue(s) | Species | Ref. |
|---|---|---|---|---|
| Actins | 6 | +3 | *Fugu* | 28 |
| 5-Hydroxytryptamine type 1A receptor | 1 | +1 | *Fugu* | 29 |
| G-protein-α | 16 | +1 (Gαp1) | *Fugu* | 30 |
| D1 dopamine receptors | 3 | +1 | European eel | 31 |
| Thyroid hormone receptors | 2 (α, β) | +2 (THRα2), (THRβ2) | Japanese flounder | 32 |
| Neurotrophin receptors (TRK) | 3 | +2 | Zebrafish | 33 |
| Receptor tyrosine kinases of subclass I (EGFR family) | 4 | +1 (Xmrk) | Platyfish, Medaka, Rainbow trout | 34<br>A. Gomez, C. Winkler, and M. Schartl (unpublished observations) |
| Notch | 3 | +3 (notch 1b, 4, 5) | Zebrafish | 35,36 |
| Gonadotropin-releasing hormone | 1 | +1 (sbGnRH) | Several species | 37 |
| Neurotrophins (NGF family) | 4 | +1 (NT-6) | Platyfish, Medaka | 38<br>R. Götz (personal communication) |
| Insulin-like growth factor-1 | 1 | +2 | Rainbow trout | T.T. Chen (personal communication) |
| Insulin-like growth factor-1 receptor | 1 | +2 | Salmon | 39 |
| Activin βA | 1 | +1 (βA2) | Medaka, Goldfish, Zebrafish | R. Köster, F. Rosa, and J. Wittbrodt (unpublished observations) |
| Activin βB | 1 | +1 (βB2) | | |
| Hedgehogs | 3 | +2 | Zebrafish | 40 |
| msh-class (*msx*) homeoproteins | 3 | +2 | Zebrafish | 12,41 |
| Orthodenticle-related genes (otx) | 2 | +1 | Zebrafish, Medaka | 13<br>J. Wittbrodt and F. Lossli (unpublished observations) |
| Engrailed | 2 | +1 | Zebrafish | 42 |
| Distalless-like genes | 6 | +2 | Zebrafish | 43,44 |

should be noted, however, that the examples that we have to date come from a wide variety of gene classes, and the fish that have contributed these examples represent the major branches of the teleost fish phylogenetic tree.

## Possible explanations and evolutionary implications

Why might there be more genes in fish than in mammals? Total genome duplications (increases in ploidy) and individual gene duplications obviously played an important evolutionary role in shaping the vertebrate genome[22] and probably initiate a process that leads to expansion of a gene family. It may be that, in teleosts, the balance between the rate of gene duplication formation and the rate of loss of duplicates is slightly shifted toward the former, with new duplicates having little or no selective cost. Whole genome duplications in ancestral vertebrates generally appear to have occurred before the separation of fish and higher vertebrates.[23] Others may have been lineage-specific duplications. On the single gene level, duplications are an important driving force of molecular evolution, where one of the two copies of a gene, now freed from performing an indispensable task, may undergo stochastic variations of its coding sequence and potentially acquire a new or related function through selection or genetic drift. Similarly, the evolution of the regulatory elements and their nexus of interactions will shape and control novel functions of already existing structural genes.[24] If the two copies of a gene took divergent evolutionary routes in different lineages toward new functions, this would explain

why homologues are not necessarily identical or even similar in terms of function and why function cannot be part of the definition of homology. Thus, the level of relatedness among genes can only be decided on the basis of phylogenetic analysis.

Since most of the multigene families have been under intense scrutiny in mammals for many years, it is unlikely that there are mammalian orthologues of the larger fish gene families that have simply remained undetected. Additional fish paralogues also could be simply the result of recent whole genome tetraploidization, as has been documented for some teleost lineages, such as the salmonid fish. Indeed, in such species many genes have two nonallelic copies that are functional, but so far all evidence points toward those two copies being fully redundant, and they cannot be differentiated as individual members of the gene family. However, most examples of additional fish paralogues compiled in Table 1 come from species that are not members of such tetraploid lineages. *Fugu* and zebrafish, in particular, the most widely studied teleosts contributing the most examples for larger gene families in fish, are not polyploid species (e.g., cf. ref. 25). Thus, the more ancient whole genome duplication events provided the basis for the development of multigene families. Hence, the other possibility is that ''higher vertebrates'' lost more of these duplicated genes while they were maintained in fish.

Are fish generally more complex than mammals, requiring more genes to structure and maintain their organismic organization? The obvious answer is ''no.'' A theoretical explanation can be taken from the economy of gene regulation. Differential gene expression is a way by which one and the same coding sequence can exert several functions in different cell types and at different developmental stages. These functions are regulated by specific regulatory elements that are compacted in a single genetic locus. This has been shown for many mammalian genes and seems to be the prevalent situation. Alternatively, separate gene copies with individual regulatory elements for each different function can exist. It will be interesting to see how the mode of gene regulation and differential gene expression in fish compares with that of mammals.

Or do fish have more genes than they actually need, which are maintained at apparently no or low enough evolutionary cost not to be selected against? Such ''dispensable'' gene copies could then even serve as a backup for the indispensable ones, a situation known as redundancy. As organisms that typically produce an overabundance of embryos and larvae, they would not specifically benefit from a high degree of gene redundancy to compensate for mutations and disturbances of normal development.

Another possible evolutionary advantage of multiple copies of genes is that there is a higher number of genes ready to acquire new or different functions permitting faster adaptation and evolution. Moreover, with more genes the complexity of the architecture of interactions among these genes is expected to increase. One could speculate that this genomic complexity of fish might permit rapid responses (in terms of changing morphology phenotypically or in terms of speciation) to changing environmental challenges and/or persistence in the face of varying and adverse selective regimes. In this regard, it might be important to point out that in African cichlid fishes speciation can happen very rapidly indeed[26,27] and that presently approximately 25,000 species of teleost fish are known, compared with roughly only 4,000 mammalian species. This abundance of fish species has been explained by the fact that teleost fish can adapt to the most divergent ecological conditions. A more ''flexible genome'' of fish might preadapt them for rapid adaptation or could be the consequence of this process. Surely, other hypotheses can be proposed as well. The observation of more genes in fish compared with mammals remains an interesting phenomenon that calls for further evolutionary and developmental studies. In practical terms, this phenomenon must be taken into consideration as a possible complication in genetic analysis based on homology cloning and in fish genome projects.

## Acknowledgments

We thank Daniel Chourrout, Ana Gómez, Felix Loosli, Kimberley Sweeney, and Ingo Schlupp for constructive comments on drafts of this manuscript.

## References

1 Powers DA (1989) Fish as model systems. *Science* 246:352–358.
2 Nuesslein-Volhard C (1994) Of flies and fishes. *Science* 266:572–574.
3 Driever W, Stemple D, Schier A, Solnica-Krezel L (1994) Zebrafish: Genetic tools for studying vertebrate development. *Trends Genet* 10:152–159.
4 Baxendale S, Abdulla S, Elgar G, Buck D, Berks M, Micklem G, Durbin R, Bates G, Brenner S, Beck S, Lehrach H (1995) Comparative sequence analysis of the human and pufferfish Huntington's disease genes. *Nature Genet* 10:67–76.
5 Tanaka M, Fukada S, Matsuyama M, Nagahama Y (1995) Structure and promoter analysis of the cytochrome P-450 aromatase gene of the teleost fish, medaka (*Oryzias latipes*). *J Biochem* 117:719–725.
6 Barth AL, Justice NJ, Ngai J (1996) Asynchronous onset of odorant receptor expression in the developing zebrafish olfactory system. *Neuron* 16:23–34.
7 Weth F, Nadler W, Korsching S (1996) Nested expression domains for odorant receptors in zebrafish olfactory epithelium. *Proc Natl Acad Sci USA* 93:13321–13326.
8 Currie PD, Ingham PW (1996) Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature* 382:452–455.
9 Zardoya R, Abouheif E, Meyer A (1996) Evolutionary analyses of hedgehog and hoxd-10 genes in fish species closely related to the zebrafish. *Proc Natl Acad Sci USA* 93:13036–13041.
10 Zardoya R, Abouheif E, Meyer A (1996) Evolution and orthology of hedgehog genes. *Trends Genet* 12:496–497.
11 Quattro JM, Woods HA, Powers DA (1993) Sequence analysis of teleost retina-specific lactate dehydrogenase C evolutionary implications for the vertebrate lactate dehydrogenase gene family. *Proc Natl Acad Sci USA* 90:242–246.
12 Ekker M, Akimenko M-A, Allende JK, 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* 14:1008–1020.
13 Mori H, Miyazaki Y, Morita T, Nitta H, Mishina M (1994) Different spatio-temporal expression of three otx homeoprotein transcripts during zebrafish embryogenesis. *Mol Brain Res* 27:221–231.
14 Li Y, Allende ML, Finkelstein R, Weinberg ES (1994) Expression of two zebrafish orthodenticle-related genes in the embryonic brain. *Mech Dev* 48:229–244.

15 Morizot DC (1990) Use of fish gene maps to predict ancestral vertebrate genome organization. In Z-IM Ogita CL Markert (eds): *Isozymes: Structure, Function, and Use in Biology and Medicine*. New York: Wiley-Liss, pp 207–234.
16 Garcia-Fernandez J, Holland PWH (1994) Archetypal organization of the amphioxus Hox-gene cluster. *Nature* 370:563–365.
17 Meyer A (1996).The evolution of body plans: Hom–Hox cluster evolution, model sytems, and the importance of phylogeny. In Harvey, PH, Leigh Brown AJ, Maynard Smith J, Nee S (eds): *New Uses for New Phylogenies.* Oxford: Oxford University Press, pp 322–340.
18 Zhang J, Nei M (1996) Evolution of antennapedia-class homeobox genes. Genetics 142:295–303.
19 Bailey WJ, Kim J, Wagner GP, Ruddle FH (1997) Pylogenetic reconstruction of vertebrate Hox-cluster duplications. *Mol Biol Evol* 14: 843–853.
20 Prince VE, Joly L, Ekker M, Ho RK (1998) Zebrafish *hox* genes: Genomic organization and modified colinear expression patterns in the trunk. *Development* 125:407–420.
21 Holland PWH (1997) Vertebrate evolution: Something fishy about *Hox* genes. *Curr Biol* 7:570–572.
22 Ohno S (1985) Dispensable genes. *Trends Genet* 1:160–164.
23 Sparrow AH, Neumann AF (1976) Evolution of genome size by DNA doublings. 192:524–529.
24 Wilson AC, Carlson S, White TK (1977) Biochemical evolution. *Annu Rev Biochem* 46:573–639.
25 Ohno S, Muramoto J, Christian L (1967) Diploid–tetraploid relationship among Old-World members of the fish family Cyprinidae. *Chromosoma (Berl)* 23:1–9.
26 Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria Africa cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347:550–553.
27 Sturmbauer C, Meyer A (1992) Genetic divergence speciation and morphological stasis in a lineage of African cichlid fishes. *Nature* 358: 578–581.
28 Venkatesh B, Tay BH, Elgar G, Brenner S (1996) Isolation, characterization and evolution of nine pufferfish (*Fugu rubripes*) actin genes. *J Mol Biol* 259:655–665.
29 Yamaguchi F, Brenner S (1997) Molecular cloning of 5-hydroxytryptamine (5-HT) type 1 receptor genes from the Japanese puffer fish, *Fugu rubripes*. *Gene* 191:219–223.
30 Sarwal MM, Sontag J-M, Hoang L, Brenner S, Wilkie TM (1997) G protein alpha subunit multigene family in the Japanese puffer fish *Fugu rubripes:* PCR from a compact vertebrate genome. *Genome Res* 6:1207–1215.
31 Cardinaud B, Sugamori KS, Coudouel S, Vincent J-D, Niznik HB, Vernier P (1997) Early emergence of three dopamine D1 receptor subtypes in vertebrates. *J Biol Chem* 272:2778–2787.
32 Yamano K, Inui Y (1995). cDNA cloning of thyroid hormone receptor beta for the Japanese flounder. *Gen Comp Endocrinol* 99:197–203.
33 Martin SC, Marazzi G, Sandell JH, Heinrich G (1995) Five Trk receptors in the zebrafish. *Dev Biol* 169:745–758.
34 Wittbrodt J, Adam D, Malitschek B, Mäueler W, Raulf F, Telling A, Robertson S.M, and Schartl M (1989) Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. *Nature* 341:415–421.
35 Westin J, Lardelli M (1997) Three novel notch genes in zebrafish—Implications for vertebrate notch gene evolution and function. *Dev Gen Evol* 207:51–63.
36 Bierkamp C, Campos-Ortega JA (1993) A zebrafish homologue of the *Drosophila* neurogenic gene Notch and its pattern of transcription during early embryogenesis. *Mech Dev* 43:87–100.
37 Gothilf Y, Elizur A, Chow M, Chen TT, Zohar Y (1995) Molecular cloning and characterization of a novel gonadotropin-releasing hormone from the gilthead seabream (*Sparus aurata*). *Mol Mar Biol Biotech* 4:27–35.
38 Goetz R, Koester R, Winkler C, Raulf F, Lottspeich F, Schartl M, Thoenen H (1994) Neurotrophin-6 is a new member of the nerve growth factor family. *Nature* 372:266–269.
39 Chan SJ, Plitsetskaya EM, Urbinati E, Jin Y, Steiner DF (1997) Expression of multiple insulin and insulin-like growth factor receptor genes in salmon gill cartilage. *Proc Natl Acad Sci USA* 94:12446–12451.
40 Ekker SC, Ungar AR, Greenstein P, Kessler DPV, Porter JA, Moon RT, Beachy PA (1995) Patterning activities of vertebrate hedgehog proteins in the developing eye and brain. *Curr Biol* 5:944–955.
41 Akimenko MA, Johnson SL, Westerfield M, Ekker M (1995) Differential induction of four *msx* homeobox genes during fin development and regeneration in zebrafish. *Development* 121:347–357.
42 Ekker M, Wegner J, Akimenko MA, Westerfield M (1992) Expression of three zebrafish coordinate embryonic engrailed genes. *Development* 116:1001–1010.
43 Akimenko MA, Ekker M, Wegner J, Lin W, Westerfield M (1994) Combinatorial expression of three zebrafish genes related to distal-less: Part of a homeobox gene code for the head. *J Neurosci* 14:3475–3486.
44 Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FH, Weiss KM (1996) The evolution of the vertebrate Dlx gene family. *Proc Natl Acad Sci USA* 93:10858–10863.