

From 2R to 3R: evidence for a fish-specific genome duplication (FSGD)

Axel Meyer^{1*} and Yves Van de Peer²

Summary

An important mechanism for the evolution of phenotypic complexity, diversity and innovation, and the origin of novel gene functions is the duplication of genes and entire genomes. Recent phylogenomic studies suggest that, during the evolution of vertebrates, the entire genome was duplicated in two rounds (2R) of duplication. Later, ~350 mya, in the stem lineage of ray-finned (actinopterygian) fishes, but not in that of the land vertebrates, a third genome duplication occurred—the fish-specific genome duplication (FSGD or 3R), leading, at least initially, to up to eight copies of the ancestral deuterostome genome. Therefore, the sarcopterygian (lobe-finned fishes and tetrapods) genome possessed originally only half as many genes compared to the derived fishes, just like the most-basal and species-poor lineages of extant fishes that diverged from the fish stem lineage before the 3R duplication. Most duplicated genes were secondarily lost, yet some evolved new functions. The genomic complexity of the teleosts might be the reason for their evolutionary success and astounding biological diversity. *BioEssays* 27:937–945, 2005.

© 2005 Wiley Periodicals, Inc.

Introduction

That there might be a causal link between increasing genomic and organismal complexity is an intriguing hypothesis that had been already suggested by J.B.S. Haldane in his landmark *The causes of evolution* and later by Ed Lewis and has received renewed attention since complete vertebrate genome sequences have become available recently. More than 30 years

ago, Susumo Ohno⁽¹⁾ convincingly put forward the hypotheses that duplications of genes and even entire genomes could be one of the major mechanisms responsible for increasing complexity during evolution. Ohno went so far as to suggest that gene duplication might be a more important mechanism in evolution than natural selection and now famously proclaimed: “*Natural selection merely modified, while redundancy created*”. Ohno’s suggestion was that, after duplications, one copy of a gene would retain the original function and that the additional copy would be free to evolve a new one. This “Ohno-mechanism” was confirmed by some recent functional genomic analyses of yeast genomes.

Several processes, such as tandem duplications, segmental duplications or even entire genome duplications, can lead to an increased number of genes in genomes. Larger genomes might tend to facilitate the functional diversification of genes, lead to larger gene families and thereby permit more complex interactions and gene networks to evolve. Ultimately, increased phenotypic complexity during evolution might thus be the result of more complex genomes.^(2,3) Genomes of protostomes such as *Drosophila* and *Caenorhabditis*, and ancestral deuterostome lineages such as *Branchiostoma* (Amphioxus) tend to have smaller gene families, or only even single copies of genes, while the genomes of mammals typically have more genes, often three or four per gene family.^(4,5) Yet the number of genes per se cannot be the whole explanation for increasing developmental and morphological complexity during evolution. That “regulatory evolution”, (i.e. increasing complexity in the regulatory networks that control the expression of genes) also played a decisive role during evolution is undisputed.⁽⁶⁾ However, the relative importance of both gene duplication and divergent transcriptional regulation in explaining the evolution of biological complexity remains a subject of vigorous debate.

2R: two rounds of genome duplications during the evolution of deuterostomes

The “one-two-four” (or 1-2-4) rule is the currently prevalent model to explain the evolution of gene families and of vertebrate genomes more generally. Based on this model, two rounds of genome duplication (2R) occurred early in the evolution of deuterostomes, which duplicated the ancestral

¹Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Germany.

²Bioinformatics and Evolutionary Genomics, Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), Ghent University, Belgium

*Correspondence to: Axel Meyer, Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany. E-mail: axel.meyer@uni-konstanz.de

Received 19 October 2004; Accepted 10 May 2005

DOI 10.1002/bies.20293

Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: mya, million years ago.

possibly cephalochordate-like genome to two (1R), after the first duplication, and then to four genomes after the second (2R) genome duplication.^(7,8) Several kinds of evidence based on analyses of gene content and gene family size support the 2R hypothesis. For example, some of the first empirical support for the 2R hypothesis came from studies of the evolution of *Hox* gene clusters in deuterostomes, where a general pattern from a single protostome or early deuterostome *Hox* cluster to four *Hox* clusters in tetrapods was observed (Fig. 1). Also, it is commonly found that clusters of other genes remain linked, often even in the same gene order, on different chromosomes [e.g. Refs. 5,8–14]. This synteny of gene clusters is often retained across large evolutionary distances among vertebrates. Some of the observed syn-

tenies could have arisen not by duplications of the entire genome, but only of portions (e.g. arms or entire chromosomes) of it. But, it could not easily be explained by numerous individual gene duplication events, which clearly are the most common type of duplication mechanism,⁽¹¹⁾ since those events would not maintain syntenic relationships across distantly related genomes.⁽¹⁴⁾

The 2R model predicts that vertebrate genomes contain four paralogs for each proto-ortholog of the cephalochordate ancestral genome.^(4,5) Based on quadruplicate paralogy between gene or genomic segments, some have indeed argued strongly for two rounds of genome duplication.^(3,8,12,13) Others, often analyzing the same data but with different techniques, found clear evidence only for one genome-doubling

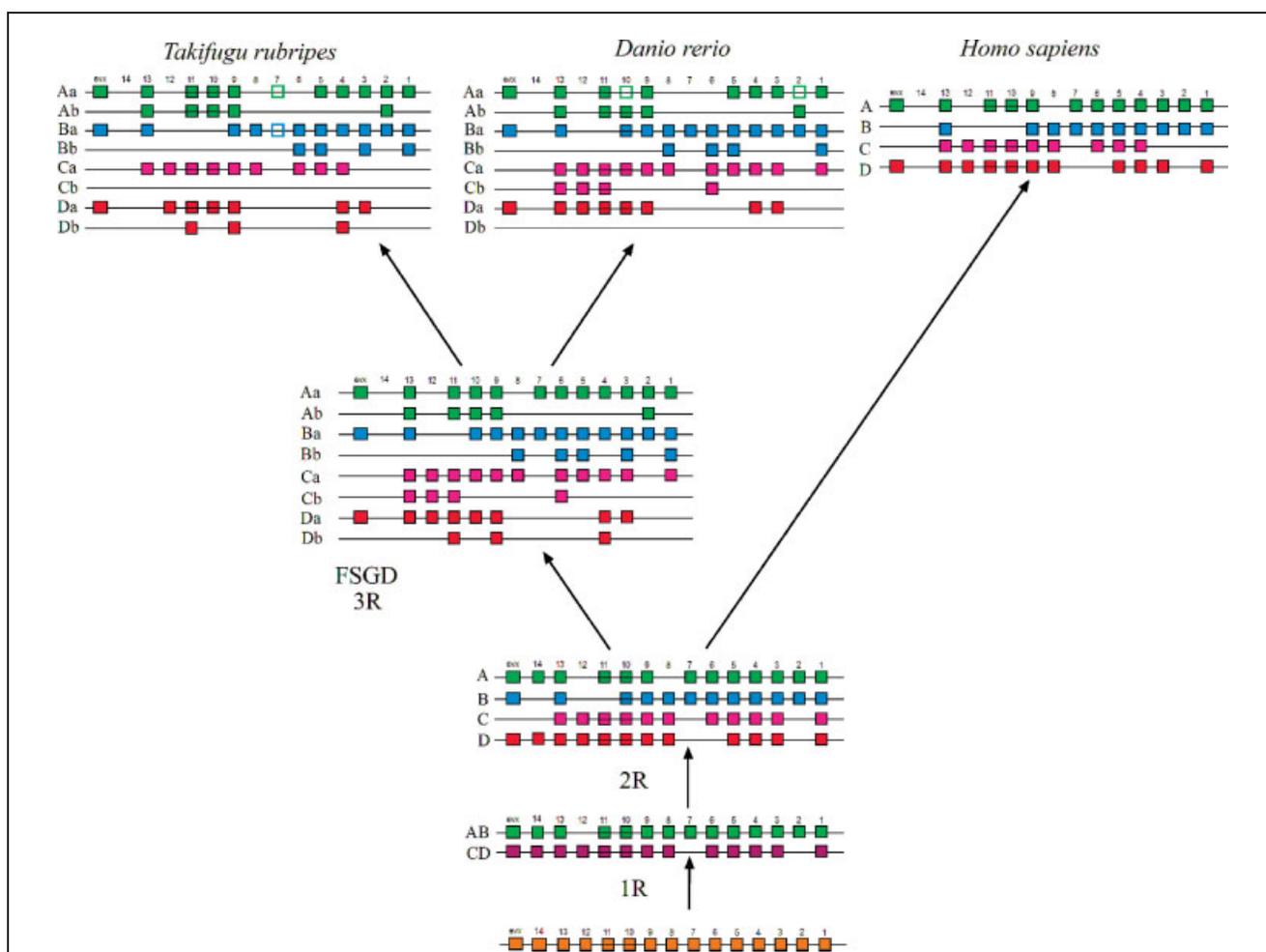


Figure 1. Evolution of Hox clusters and the inferred genomic events. Genomic architecture of the fugu, zebrafish, human and cephalochordate (in orange) Hox clusters are known. The intermediate architectures of Hox clusters are inferred based on the 3R model and are hypothetical. During the evolution of ray-finned fishes, the FSGD occurred and also the inferred fish-ancestral Hox cluster architecture is shown. 1R is believed to have occurred following the branching off the cephalochordates from the chordate stem lineage and before the cyclostomes and chondrichthyans branched off. 2R is thought to have occurred quickly thereafter. The FSGD (3R) occurred after the basal lineages of fishes branched off from the actinopterygian stem lineage and before the origin of the derived (teleost) ray-finned fishes.

event.^(15–17) While, until recently, yet others have taken a neutral position, arguing that current data neither support nor reject the 2R hypothesis.^(5,18,19) Other recent evidence for and against the 2R hypothesis has been put forward and several modifications have been proposed, assuming a diversity of small- and large-scale gene duplication events.⁽²⁰⁾

If quadruplicate paralogy regions are historical remnants of two whole genome duplications, this should be reflected in the topologies of phylogenetic trees reconstructed from their constituent genes.^(5,14,18) Therefore, this pattern of two rounds of genome duplication would be expected to bring about a symmetrical nested set of phylogenetic relationships among the four genes of gene families that reflects their evolutionary history ((AB)(CD)) with the age of the AB split the same as the age of the CD split.⁽¹⁹⁾ The fact that many, or even most, tree topologies based on duplicated vertebrate genes do not show a 2 + 2 topology has, by some, therefore been considered to be evidence against the 2R hypothesis.⁽²⁰⁾

Furlong and Holland,⁽²¹⁾ however, have argued that incongruent tree topologies are not necessarily in disagreement with two rounds of sequential genome duplication, but are even to be expected when two auto-tetraploidy events have taken place in close succession. Also Gibson and Spring⁽²²⁾ have observed that the period between both genome duplication events during early vertebrate evolution could have been as short as 10 million years or maybe even shorter. Gene trees would thus simply reflect the random order of diploidization of chromosomes, rather than the order of gene duplication and tree topologies would, in general, turn out to be asymmetrical. In such cases particularly for sequences that have been duplicated more than 600 million years ago,⁽²³⁾ gene quartets would not contain sufficient phylogenetic signal to resolve internal branches and inferred tree topologies would essentially be random. In addition, due to directional selection and possibly the acquisition of novel gene functions, many duplicate genes exhibit unequal rates of evolution after duplication,⁽²⁴⁾ which poses a difficulty for most phylogenetic algorithms and often results in tree reconstruction artifacts.

Still, the remaining opponents⁽²⁵⁾ of the 2R hypothesis argue that sets of tandem gene duplications are alone sufficient to explain the greater genomic complexity of vertebrates compared to that of their deuterostome ancestors and feel that phylogenetic trees based on vertebrate genes are mostly incongruent with two rounds of genome duplication. But, such a 'clean' pattern should not be expected for every gene family,^(5,21,26) for other reasons as well, including the fact that about 600 million years of independent evolution in all remaining deuterostome lineages makes it difficult to find appropriate genomes for phylogenomic comparisons since also lineage-specific genomic events further add to the difficulty of reconstructing an ancestral genome of such antiquity.^(5,26) Duplicated genes seem to be lost relatively easily and quickly after the initial duplication event, reducing the size of gene

families secondarily, while gains of function are less likely to occur.^(11,24) Moreover, lineage-specific (tandem) duplications could both increase the size of gene families and, depending on when they occurred, obscure the topology of a gene tree. Such independent losses and duplications of individual genes or sections of the genome have occurred frequently during the evolution of vertebrates from their deuterostome ancestor.⁽¹⁰⁾ In summary, the evidence for 2R seems, in our opinion, and by that of most other workers in the field^(4,10,13,21,23) to be sufficiently strong as to provisionally accept this model for the evolution of the vertebrate genome.

3R: the fish-specific genome duplication (FSGD)

During the last six years, evidence has accumulated to suggest that many fish might have even more genes than humans^(14,27,28) and recent data^(23,29–33) suggest that an additional whole genome duplication occurred in the fish lineage, extending the 1-2-4 to an 1-2-4-8 rule.⁽¹⁴⁾ Whether most of these additional genes originated through a complete fish-specific genome duplication (FSGD or 3R) or through many lineage-specific tandem gene or smaller block duplications was initially debated just as for the 2R hypothesis. However, a consensus has emerged in support of the 3R hypotheses.^(14,23,29–38) Here, we review recent evidence for the 3R hypothesis.

The first indications for an actinopterygian-specific genome duplication again came from studies based on Hox genes and Hox clusters, in particular those of the zebrafish and fugu. Extra Hox gene clusters were discovered in the zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), the African cichlid (*Oreochromis niloticus*), the pufferfish (*Takifugu rubripes*) this suggested that all teleost fishes experienced an additional genome duplication in ray-finned fishes (Actinopterygii) before the divergence of most teleost species [reviewed in Refs. 30,34–36,38]. Comparative genomic studies have also revealed many more genes and gene clusters for which two copies exist in teleost fishes but only one cognate copy in other vertebrates. The observations that different paralogous pairs seem to have originated at about the same time,⁽³¹⁾ that different ray-finned fish species seem to share ancient gene duplications⁽³²⁾ and that different paralogs are found on different linkage groups and show synteny with other duplicated chromosomal regions⁽³²⁾ support the hypothesis that these genes arose through a complete genome duplication event during the evolution of the actinopterygian lineage (see other recent reviews,^(35–38)).

The most conclusive evidence for a complete genome duplication in ray-finned fishes came from the comparative analyses of the nearly complete *Takifugu* genome and in particular the recently determined *Tetraodon* genome with the human genome.^(28,29) The approach taken by Jaillon et al.⁽²⁹⁾ was to compare the chromosomal distribution of genes of the

Tetraodon genome with that in the human genome. It was observed that many incidents of human synteny segments were found in duplicate on two different *Tetraodon* chromosomes. Both Vandepoele et al.⁽²³⁾ and Christoffels et al.⁽³³⁾ identified duplicated genes in *Takifugu* and used phylogenetic analyses to estimate the ages of their duplication events. Vandepoele et al.⁽²³⁾ constructed phylogenetic trees for all gene families containing 2 to 10 duplicated sets of *Takifugu* genes. For each gene family, relative dating of duplication events was performed to test whether gene duplications occurred before or after the divergence of the lineages that led to ray-finned fishes and land vertebrates, subsequently referred to as 1R/2R and 3R, respectively (Fig. 2). To this end, neighbor-joining trees were created for each of the *Takifugu* gene families with homologous sequences from mice and humans. *Ciona* (ascidian) and *Drosophila* sequences were used as outgroups. About 750 gene families could reliably be used for relative dating purposes. Absolute dating of duplication events was performed by inference from linearized trees.⁽³⁹⁾ In such trees, where branch length is directly proportional to time, the split between ray-finned fishes (actinopterygians) and land vertebrates, dated at 450 million years ago was used as a calibration point for the dating of gene duplication events. Combining the results of relative and absolute dating, 565 duplication events could be inferred of which 1/3rd was ascribed to the 3R (FSGD) and 2/3rd to the 2R

duplication events (Fig. 2). This age distribution of duplicated genes is what would be expected if three whole genome duplication doublings have occurred in early vertebrates and ray-finned fishes and assuming equal amounts of gene loss following independent genome doublings.⁽²³⁾

To test whether the sudden increase in the number of duplicated genes in the *Takifugu* genome is the result of an entire genome duplication rather than an increased rate of local gene duplications events, Vandepoele et al.⁽²³⁾ also investigated whether duplicated genes appear in duplicated blocks on chromosomes. The identification of duplicated blocks is conventionally considered strong evidence for large-scale duplication events. By applying a newly developed algorithm to detect co-linearity⁽⁴⁰⁾ to scaffolds of the pufferfish genome sequence, about 160 duplicated blocks were identified, containing over 540 paralogous gene pairs. To date the origin of these duplicated chromosomal blocks, linearized trees were again inferred, but this time only for the set of anchor points, (i.e. the paralogous genes within duplicated blocks.) Dating of duplicates in duplicated segments clearly showed that the different blocks of duplicated segments all arose at approximately the same time about 320 million years ago (Fig. 2).⁽²³⁾ A similar estimate for the FSGD was obtained by Christoffels et al.⁽³³⁾ who followed a similar approach when analyzing the whole *Takifugu* genome. Based on the construction of 88 linearized trees, these authors estimated that

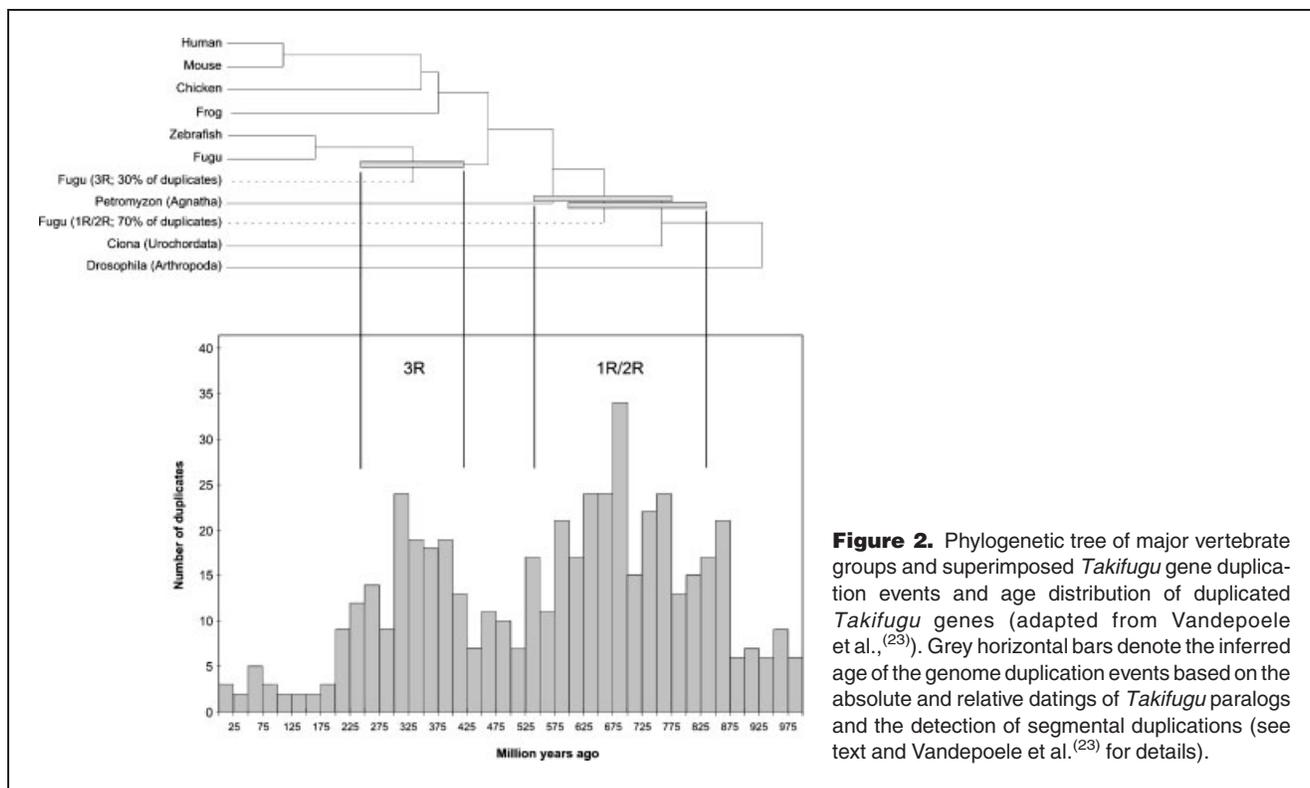


Figure 2. Phylogenetic tree of major vertebrate groups and superimposed *Takifugu* gene duplication events and age distribution of duplicated *Takifugu* genes (adapted from Vandepoele et al.,⁽²³⁾). Grey horizontal bars denote the inferred age of the genome duplication events based on the absolute and relative datings of *Takifugu* paralogs and the detection of segmental duplications (see text and Vandepoele et al.⁽²³⁾ for details).

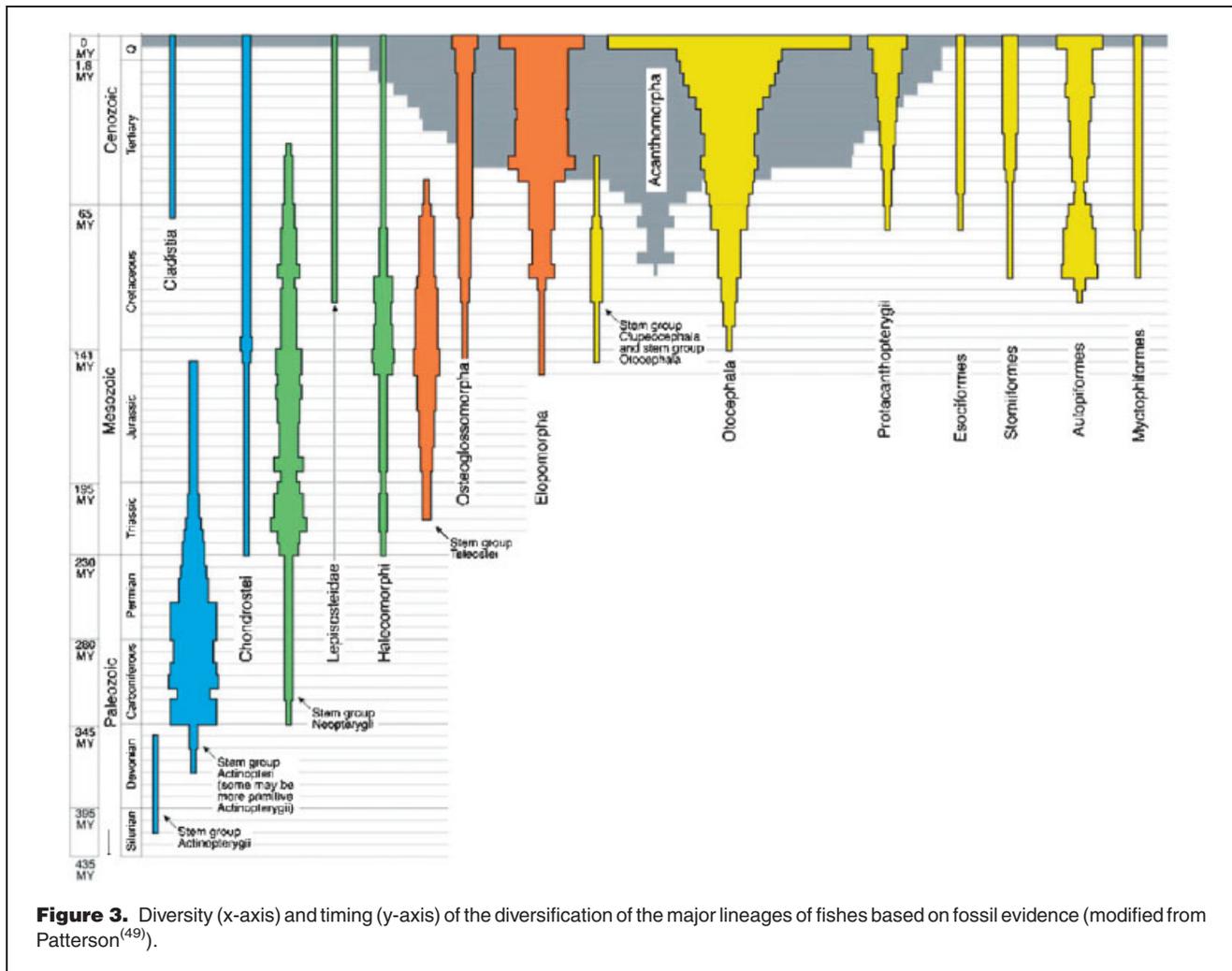


Figure 3. Diversity (x-axis) and timing (y-axis) of the diversification of the major lineages of fishes based on fossil evidence (modified from Patterson⁽⁴⁹⁾).

the FSGD occurred ~350 million years ago. These estimates would place the FSGD at the base of the radiation of all ray-finned fishes (Fig. 3).

Was the FSGD responsible for the dramatic rise of the teleosts?

If the FSGD (3R) was responsible for the biological diversification of ray-finned fishes, it must have occurred before their radiation began. Paleontological evidence suggests that most extant ray-finned fishes or “teleosts” first appeared only about 200 million years ago (Fig. 3). If the fossil record is reliable, there seems to be a major time gap between the FSGD, estimated to have occurred more than at least ~320 million years ago and the radiation of the teleosts, rendering it unlikely that the FSGD has been a major driving force behind the rapid radiation of teleosts, as has been suggested previously.^(10,14,23,27,35,36)

We suspect that, as is often found in molecular phylogenetic analyses, paleontological dates tend to underestimate

the real ages of lineages and palaeontological data can only provide minimum estimates for the age of particular lineages, and these are generally younger than estimates based on molecular data.⁽⁴¹⁾ Also, palaeontological and molecular estimates for fish divergences differ widely.⁽⁴¹⁾ The dating of fossils and their assignment to particular lineages are, by definition, always minimal ages of lineages, since the absence of fossil finds cannot rule out the existence of older fossils, which, if found, would lead to an increase in divergence age estimates.

In addition to absolute dating (through molecular clocks) genomic events can also be aged based on their phylogenetic distribution, (i.e. their distribution on known phylogenetic relationships—phylogenetic timing). This approach allows the determination of whether duplications have occurred prior to or after a lineage splitting event. The class Actinopterygii (or ray-finned fish) includes more than 23,500 species, of which the vast majority belong to the most derived division, the Teleostei (teleosts) (Figs. 3 and 4). Interestingly, all older, more

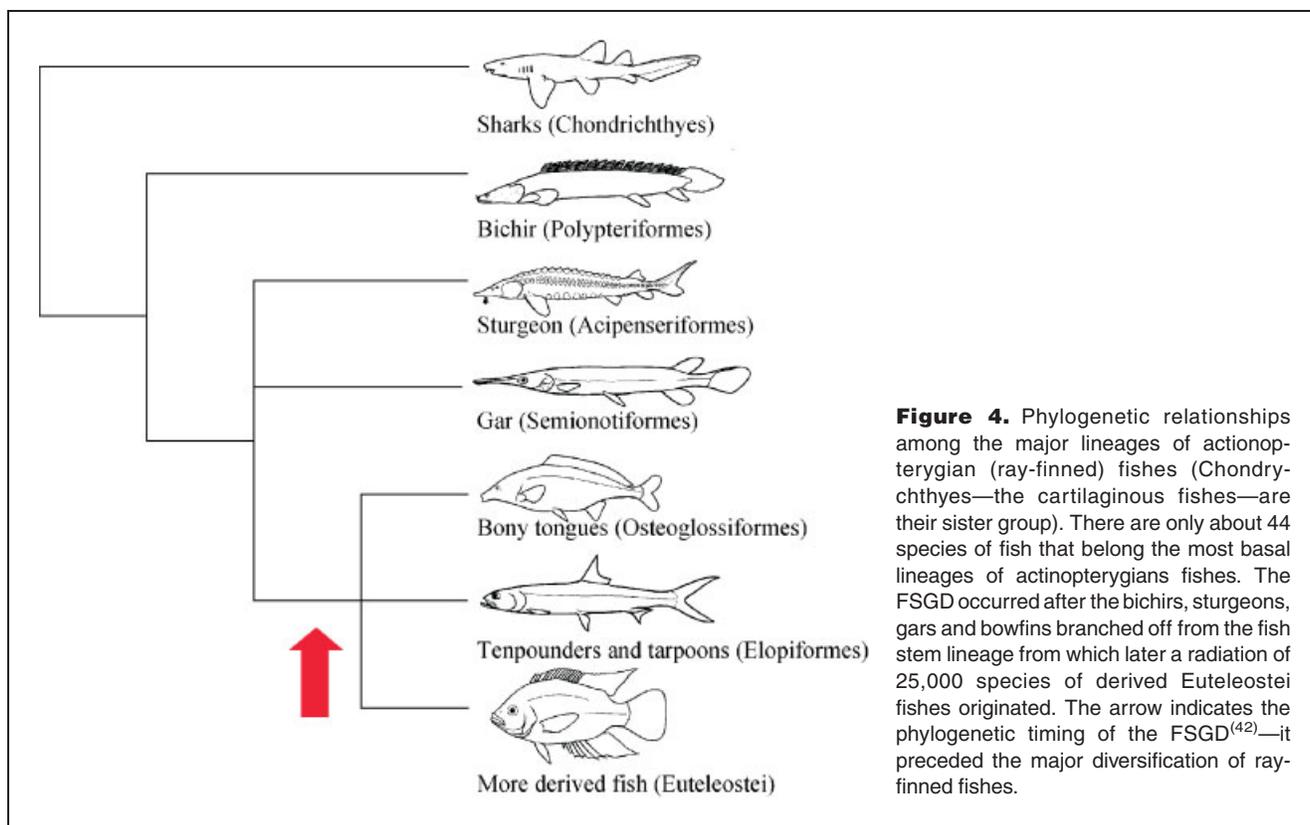


Figure 4. Phylogenetic relationships among the major lineages of actinopterygian (ray-finned) fishes (Chondrichthyes—the cartilaginous fishes—are their sister group). There are only about 44 species of fish that belong to the most basal lineages of actinopterygian fishes. The FSGD occurred after the bichirs, sturgeons, gars and bowfins branched off from the fish stem lineage from which later a radiation of 25,000 species of derived Euteleostei fishes originated. The arrow indicates the phylogenetic timing of the FSGD⁽⁴²⁾—it preceded the major diversification of ray-finned fishes.

basal groups of ray-finned fishes, namely Polypteriformes (bichirs), Acipenseriformes (sturgeons and paddlefish), Semionotiformes (gars) and Amiiformes (bowfin, probably the sister group to the gars) consist of only a few (~44 species) extant species (Figs. 3 and 4). Most members of these basal “species-poor” actinopterygian lineages are considered to be “living fossils”, because their morphology has remained unchanged over very long periods of time. Fossil data age the Semionotiformes between 245 and 286 mya (Permian), while molecular estimates for the Amiiformes, which are approximately of the same age as Semionotiformes, hint at a separation from the Teleostei stem lineage around 367–404 mya. Likewise, molecular data suggest an age of 335 mya for the Osteoglossomorpha (divergence dates reviewed in Ref. 42) (but see Fig. 3).

Hoegg et al.⁽⁴²⁾ attempted recently to determine the phylogenetic timing of the FSGD in relation to the origination of lineages of teleost and ‘non-teleost’ fishes. To this end, the three nuclear genes (*fzd8*, *sox11*, *tyrosinase*) were sequenced from sturgeons (Acipenseriformes), gars (Semionotiformes), bony tongues (Osteoglossomorpha) and a tenpounder (Elopomorpha) (Fig. 4). Two cognate paralogs have previously been described for these three genes, in derived teleosts such as zebrafish and pufferfish, yet only one orthologous copy for those duplicated genes is found in tetrapods. These three

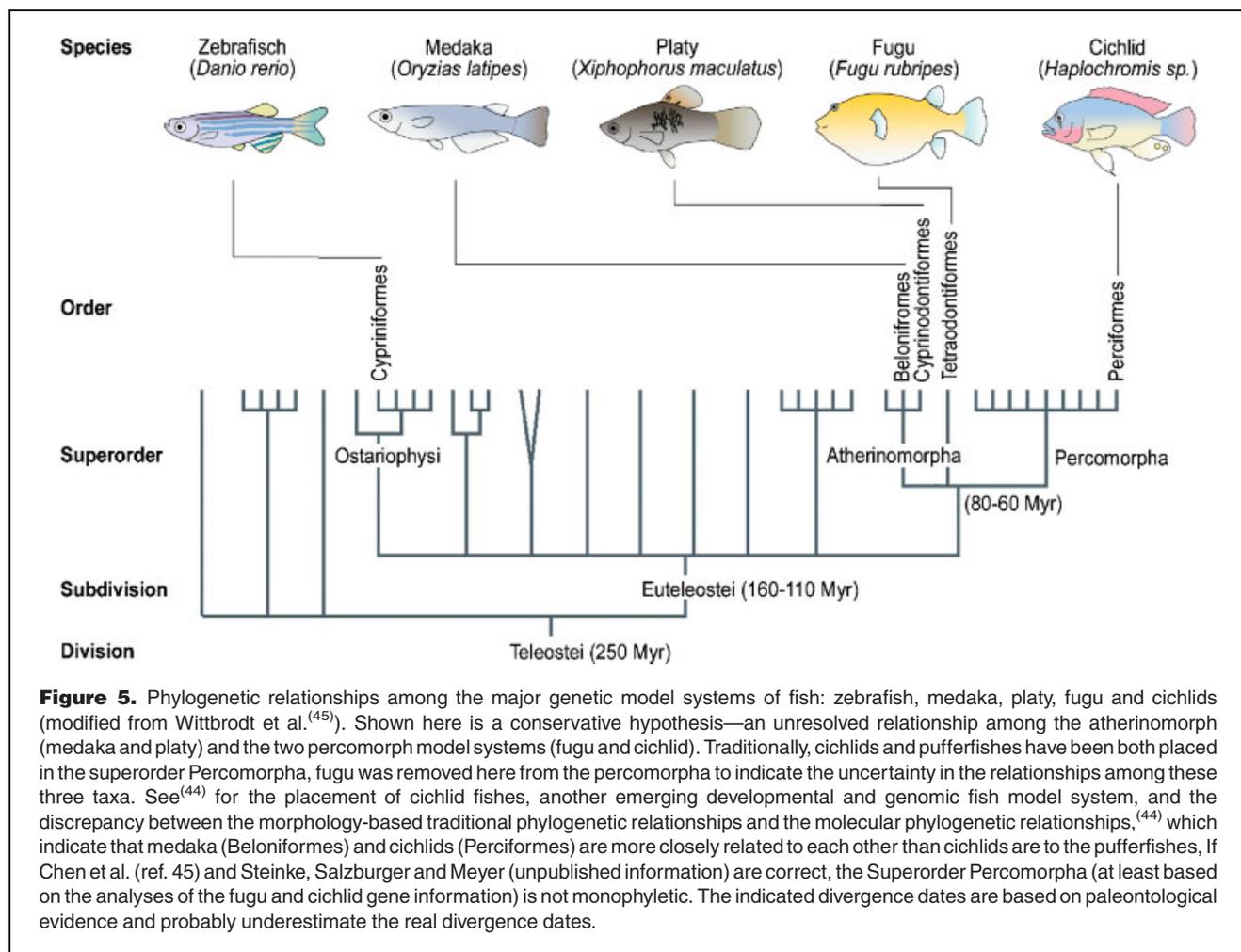
genes therefore suggested that, at least modern fishes (teleosts) are derived from a lineage in their common ancestry that experienced the 3R. But exactly how far back during the evolution of fishes the 3R duplication happened was asked in the study of Hoegg et al.⁽⁴²⁾ The specific clustering of the genes in individual gene trees and a concatenated dataset support the hypothesis that the fish-specific genome duplication event took place within the ray-finned (actinopterygian) fishes but after the divergence of the Chondrostei (Acipenseriformes and Polypteriformes) and after the Semionotiformes (including Amiiformes) separated from the actinopterygian stem lineage leading to teleosts, but before the divergence of Osteoglossiformes and other more derived groups of teleost fishes (Fig. 4). The FSGD therefore seems to be more precisely a teleost-specific genome duplication and can be dated to between 335 and 404 mya. This is in very good agreement with analyses of the complete pufferfish genome^(23,28) (see above), which showed an increased amount of duplicated genes that originated about 350 mya.^(23,33)

The inferred relative (phylogenetic) and absolute date for the FSGD separates the species-poor early branching ray-finned fish lineages from the extremely species-rich teleost lineage. The term fish-specific genome duplication, FSGD or 3R, seems still appropriate, although teleost-specific genome duplication would be most precise, since virtually all living

(~23,500 except for the ~44 most basal living ones) species that are commonly thought of as “fish” (a term that describes a paraphyletic group any way because it is applied to all, even sarcopterygian, fish such as lungfishes and coelacanths and cartilaginous fishes such as sharks, chimaeras and rays) experienced the FSGD in their evolutionary past. The absolute and phylogenetic timing of the 3R duplication provides suggestive evidence^(10,27,28,31,35–37) that the FSGD might be causally related to an increase in species as well as biological diversity. The phylogenetic timing of the FSGD following the divergence of non-teleostean, basal actinopterygian lineages (which consist of 44 species in 5 families) and the origin of the division Teleostei (which contains 23,637 species in 425 families⁽⁴³⁾) provides further correlational evidence to support a genomic link between 3R and the resulting organismal diversity. The subdivisions Osteoglossomorpha (217 species) and Elopomorpha (37 species) as the first lineage to diverge from the fish stem lineage after the presumptive genome duplication event show a somewhat elevated number of

species compared to more basal actinopterygians (44 species in five families).

Future comparative genomic studies on basal lineages of ray-finned fishes will likely shed additional light on this question. Most fish model systems currently used in genomic or developmental research are all relatively derived, except for the zebrafish, and mostly belong to the superorders Atherinomorpha and Percomorpha,^(10,44,45) highlighting the importance to also investigate species that belong to more ancient lineages such as the Acipenseriformes and the Polypteriformes (e.g. Refs. 46–48). Correct phylogenetic information is required to interpret comparative genomic and developmental data. Importantly, even the traditional morphology-based phylogenetic relationships among fish model systems might turn out not to be incorrect.⁽⁴⁴⁾ Chen et al.,⁽⁴⁴⁾ based on the phylogenetic analyses of several nuclear genes, found that cichlid fishes are probably more closely related to medaka (Beloniformes) than to pufferfishes as has been traditionally thought (Fig. 5). Also recent EST analyses of several



fish genomes tend to support this novel non-traditional grouping of Beloniformes and of perciform (at least the family Cichlidae or the suborder Labroidei) fishes (Steinke, Salzburger and Meyer, unpublished data). Figure 5, conservatively, depicts these relationships in an unresolved polytomy among the medaka, platy, fugu and cichlid model systems. Additional work on the higher order relationships among fishes will be required to test this hypothesis further.

Conclusions

The importance of the role of whole genome duplications in shaping the architectures of vertebrate genomes has been recognized and reaffirmed through the examination of complete genomic sequences of fishes and their comparative analyses.^(28,29) The strength of the evidence for two rounds (2R) of genome duplications early in the evolution of the vertebrate lineage and in particular for the (teleost) fish-specific genome duplication (3R or FSGD) has increased in recent years. The 3R genome duplication preceded the diversification of the major extant fish lineages. Teleost fishes with ~25,000 species are by far the most evolutionarily diverse group of vertebrates. The hypothesis of a cause–effect relationship between this teleost fish-specific genome duplication and the astonishing species diversity of this most-derived division of fishes is suggested by the result that species-poor extant fish lineages branched off the actinopterygian stem lineage before the 3R duplication led to an initial doubling of the gene number. Future comparative, phylogenomic and, importantly, functional genomic analyses of fish genomes are going to test this hypothesis further.

Acknowledgments

William Bemis kindly provided Fig. 3, Simone Hoegg Figures 1 and 4 and Jochen Wittbrodt and Walter Salzburger Fig. 5. The comments of two reviewers and Simone Hoegg greatly improved an earlier version of the manuscript. Financial support was provided by the University of Konstanz, and the Deutsche Forschungsgemeinschaft to AM.

REFERENCES

- Ohno S. 1970. Evolution by gene duplication. New York: Springer Verlag.
- Aburomia R, Khaner O, Sidow A. 2003. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. *J Struct Funct Genomics* 3:45–52.
- Holland PW. 2003. More genes in vertebrates? *J Struct Funct Genomics* 3:75–84.
- Spring J. 1997. Vertebrate evolution by interspecific hybridization—are we polyploid? *FEBS Lett* 400:2–8.
- Horton AC, Mahadevan NR, Ruvinsky I, Gibson-Brown JJ. 2003. Phylogenetic analyses alone are insufficient to determine whether genome duplication(s) occurred during early vertebrate evolution. *J Exp Zool* 299B:41–53.
- Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* 424:147–151.
- Sidow A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Gen Dev* 6:715–722.
- Sharman AC, Holland PWH. 1996. Conservation, duplication, and divergence of developmental genes during chordate evolution. *Neth J Zool* 46:47–67.
- Pebusque M-J, Couleir F, Birnbaum D, Pontarotti P. 1998. Ancient large-scale genome duplications: phylogenetic linkage analysis shed light on chordate genome evolution. *Mol Biol Evol* 15:1145–1159.
- Postlethwait P, Amores A, Cresco W, Singer A, Yan Y-L. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet* 20:481–490.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat Genet* 31:100–105.
- Lundin LG, Larhammer D, Hallböök F. 2003. Numerous groups of chromosomal regional paralogs strongly indicate two genome doublings at the root of the vertebrates. *J Struct Funct Genomics* 3:53–63.
- Meyer A, Scharl 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.
- McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. *Nat Genet* 31:200–204.
- Gu X, Wang Y, Gu J. 2002. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet* 31:205–209.
- Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, et al. 2003. New evidence for genome-wide duplications at the origin of vertebrates using an *Amphioxus* gene set and completed animal genomes. *Genome Res* 13:1056–1066.
- Makalowski W. 2001. Are we polyploids? A brief history of one hypothesis. *Genome Res* 11:667–670.
- Skrabaneck L, Wolfe KH. 1998. Eukaryote genome duplication—where's the evidence? *Curr Opin Gen Dev* 8:694–700.
- Hughes AL, Friedman R. 2003. Testing hypotheses of genome duplication in early vertebrates. *J Struct Funct Genomics* 3:85–93.
- Furlong RF, Holland PWH. 2002. Were vertebrates octoploid? *Phil Trans R Soc Lond B* 357:531–544.
- Gibson TJ, Spring J. 1998. Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *Trends Genet* 14:46–49.
- Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer A. 2004. 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* 101:1638–1643.
- Van de Peer Y, Taylor JS, Braasch I, Meyer A. 2001. Ghosts of selection past: rates of evolution of anciently duplicated genes. *J Mol Evol* 53:434–444.
- Friedman R, Hughes AL. 2001. Gene duplication and the structure of eukaryotic genomes. *Genome Res* 11:373–381.
- Gu X, Huang W. 2002. Testing the parsimony test of genome duplications: a counterexample. *Genome Res* 12:1–2.
- Wittbrodt J, Meyer A, Scharl M. 1998. More genes in fish? *BioEssays* 20:511–515.
- Aparicio S, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310.
- Jaillon O, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 31:946–957.
- Van de Peer Y, Taylor JS, Meyer A. 2003. Are all fishes ancient polyploids? *J Struct Funct Genomics* 3:65–73.
- 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* 356:1661–1679.
- Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res* 13:382–390.
- Christoffels A, Koh EG, Brenner S, Aparicio S, Venkatesh B. 2004. *Fugu* genome analysis provides evidence for a whole-genome duplication

- early during the evolution of ray-finned fishes. *Mol Biol Evol* 21:1146–1151.
34. Prohaska SJ, Stadler PF. 2004. The duplication of the Hox gene clusters in teleost fishes. *Theory in BioSciences* 23:89–110.
 35. Van de Peer Y. 2004. Tetraodon genome confirms Takifugu findings: most fish are ancient polyploids. *Genome Biology* 5:250.
 36. Volf J-N. 2005. Genome evolution and biodiversity in teleost fish. *Heredity* 94:280–294.
 37. Stellwag E. 2004. Are genome evolution, organism complexity and species diversity linked? *Integr Comp Biol* 44:358–365.
 38. Hoegg S, Meyer A. 2005. Hox clusters as models for vertebrate genomic evolution. *Trends Genet* 21:421–424.
 39. Takezaki N, Rzhetsky A, Nei M. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12:823–833.
 40. Simillion C, Vandepoelle K, Van de Peer Y. 2004. Recent developments in computational approaches for uncovering genomic homology. *BioEssays* 26:1225–1235.
 41. Meyer A, Zardoya R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. *Ann Review Ecol Syst* 34:311–338.
 42. Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with phenotypic and taxonomic diversification in fishes. *J Mol Evol* 59:190–203.
 43. Nelson J. 1994. *Fishes of the world*. New York: Wiley.
 44. Chen W-J, Guillermo Ortí G, Meyer A. 2004. Novel evolutionary relationship among four fish model systems. *Trends Genet* 20:424–431.
 45. Wittbrodt J, Shima A, Schartl M. 2002. Medaka—a model organism from the far east. *Nature Rev Genet* 3:53–64.
 46. Amores A, Suzuki T, yan Y-L, Pomery J, Singer A, et al. 2004. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-finned fish. *Genome Res* 14:1–10.
 47. Chui C, Dewar K, Wagner GP, Takahashi K, Ruddle F, et al. 2004. Bichir *HoxA* cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res* 14:11–17.
 48. Powers TP, Amemiya CT. 2004. Evolutionary plasticity of vertebrate *Hox* genes. *Curr Genomics* 5:459–472.
 49. Patterson C. 1993. An overview of the early fossil record of acanthomorphs. *Bull Mar Sci* 52:29–59.