Correspondence

Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis

John S. Taylor*, Yves Van de Peer† and Axel Meyer*

The discovery of seven unlinked *Hox* gene clusters in zebrafish (*Danio rerio*) and two *HoxA* gene clusters in the pufferfish (*Takifugu rubripes*) led to the hypothesis that a genome duplication event occurred early during the evolution of ray-finned fishes (Actinopterygii) [1]. Shortly thereafter, Wittbrodt and co-workers [2] uncovered many additional examples of genes that appeared to have been duplicated in fish. The fish-specific genome duplication hypothesis was further supported by the discovery of synteny within the zebrafish genome (*i.e.* the observation that many duplicated zebrafish genes map to the same pairs of chromosomes) [3,4] and by the discovery of seven unlinked *Hox* gene clusters in the Japanese medaka (*Oryzias latipes*) [5].

However, in several recent papers Robinson-Rechavi *et al.* [6–8] have argued that an ancestral whole-genome duplication event might not be responsible for the abundance of duplicated fish genes. These authors counted orthologous genes in fish and mouse and, where extra genes were found in fish, compared the number of gene duplications occurring in a single fish lineage to the number of gene duplications shared by more than one lineage [6,7]. They found that most mouse genes surveyed occurred only once in fish. Duplicated fish genes were detected, but most were the products of lineage-specific duplication events and not an ancient duplication event.

Here we discuss three major problems with the approach used by Robinson-Rechavi *et al.* First, to test the ancient fish-specific genome duplication event, only genes that were available in the Hovergen database [9] from mouse and at least three major fish lineages (*i.e.* orders) were employed. Few genes met this three-lineage criterion: 33 in one study [6] and a mostly overlapping set of 37 genes in another [7]. The authors argue that their analyses of this short list of genes provide very little evidence for the ancient fish-specific genome duplication hypothesis: in these phylogenetic analyses only 7 gene families out of 37 (19%) follow a pattern consistent with an ancestral whole genome duplication origin. On the other hand, for 11 (30%), all detected duplications happened after the divergence of fish lineages. Finally for 19 gene families (51%) no duplication was observed among fish [7].

However, the failure to find many examples of gene duplicates shared by at least two lineages (*i.e.* 'lineage-shared' gene duplicates) in a very limited dataset should not be considered evidence against an ancient genome duplication event. As Robinson-Rechavi *et al.* [6] point out, the amount of sequencing done in mouse is about 25 times higher than in zebrafish. Furthermore, most duplicated genes are likely to be lost, a prediction based upon theoretical and empirical data [10,11]. Considering these two factors, which severely limited the number of duplicates that one might expect to find, the discovery that lineage-shared duplicates occur in 7 out of 37 genes might be considered evidence in favour of an ancient whole-genome duplication rather than evidence against it.

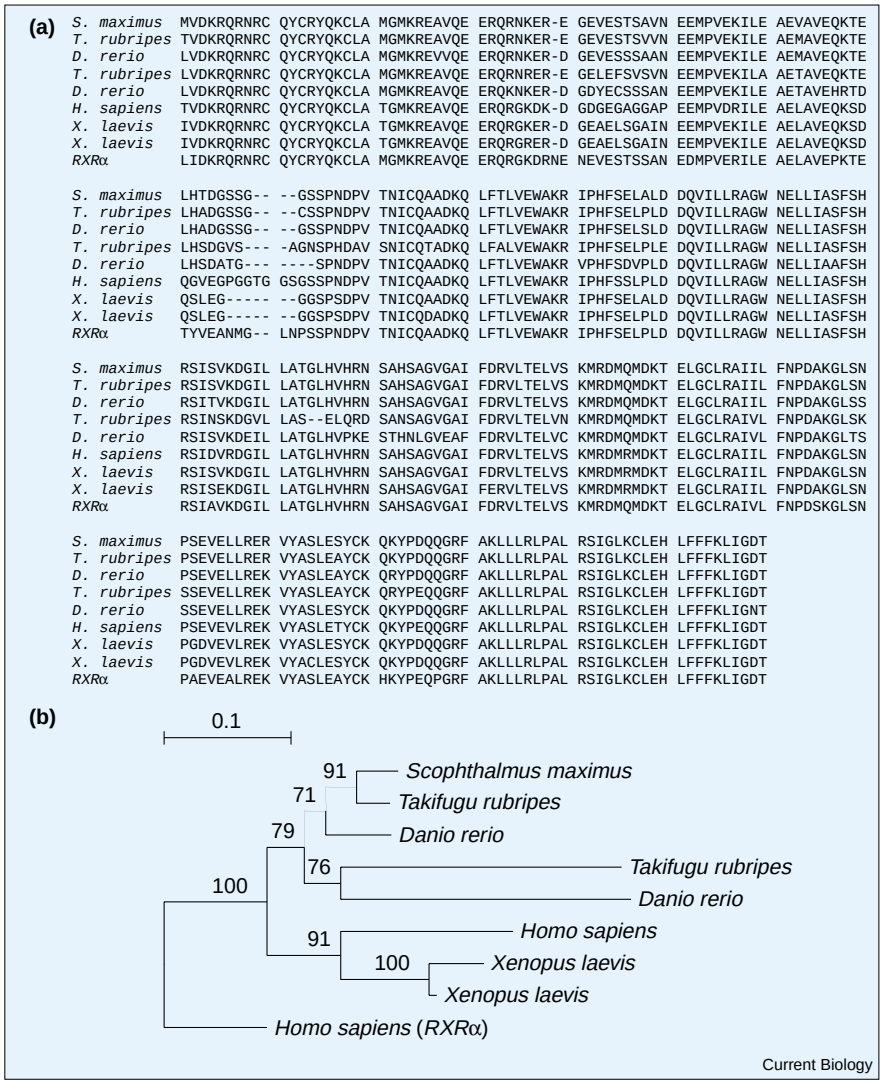
Our second criticism involves the bias introduced by tetraploid species. In one study [6], Robinson-Rechavi *et al.* uncovered 26 lineage-specific gene duplication events and only 13 gene duplication events shared by at least two lineages. Because most of the duplicated fish genes arose more recently than the divergence of major fish groups, Robinson-Rechavi *et al.* argue that it does not appear possible to support the view that the abundance of duplicated genes in fish arose mainly through a unique genome duplication event [7]. However, in Actinopterygii there are many clear cases of lineage-specific tetraploidy [12–17]. It is not possible to determine precisely the extent to which relatively recent tetraploidy events have contributed to the high number of lineage-specific gene duplication events observed by Robinson-Rechavi *et al.* because unfortunately the names of the species they surveyed were not provided. But the inclusion of 9 lineage-specific gene duplication events from Salmoniformes — all species tetraploid — is one obvious source of bias.

The order Cypriniformes, with 10 lineage-specific gene duplication events, also includes tetraploid species that have received a high degree of sequencing effort such as goldfish, *Carassius auratus*, and carp, *Cyprinus carpio*. Thus, much lineage-specific gene duplication is expected in ray-finned fishes, especially in two of the orders surveyed by Robinson-Rechavi *et al.*, and these data should also not be interpreted as evidence against a genome duplication in the ancestor of ray-finned fishes.

Our third criticism deals with the phylogenetic analyses discussed in reference [6]. In addition to surveying the Hovergen database for orthologous sequences in mouse and fish, Robinson-Rechavi *et al.* [6] sequenced nuclear hormone receptors in 7 fish species. In at least one case these new data were not analysed rigorously and this error also led to erroneous conclusions with respect to gene duplication in fish.

Figure 1

Amino acid alignment (a) and phylogeny (b) of *RXRβ* genes; note, zebrafish *RXRβ* genes are named *RXRε* and *RXRδ*. Gene identification numbers from top to bottom: turbot, *S. maximus*, gi14994052; pufferfish, *T. rubripes*, JGI4030; zebrafish, *D. rerio* (*RXRε*), gi1046299; pufferfish, *T. rubripes*, JGI191; zebrafish, *D. rerio* (*RXRδ*), gi1046297; human, *H. sapiens*, gi1350911; frog, *X. laevis*, duplicates, gi840922 & gi1085220; human RXRα, outgroup, gi4506755. Phylogeny based upon Poisson-corrected genetic distances and Neighbour-joining clustering as implemented in TREECON [21]. Bootstrap support [22] for each node based upon 500 reiterations. Distance scale: 0.1 substitutions per site. We note that the pufferfish sequences were not available when Robinson-Rechavi *et al.* published their work but that the relationship between turbot RXR and zebrafish RXRε does not depend upon their inclusion in the analysis. The turbot *RXR* gene is also more similar to zebrafish *RXRε* than it is to *RXRδ* at the nucleotide level (all positions and when only 3rd codon positions are considered). The duplication event results in the formation of two paralogous clades of *RXRβ* genes in fish. Branch lengths suggest an increase in evolutionary rate of one of these two clades.



Zebrafish has two *RXRβ* genes [18]. Robinson-Rechavi *et al.* sequenced a single *RXRβ* gene from turbot (*Scophthalmus maximus*) and came to the conclusion that the duplication of *RXRβ* occurred only in the zebrafish lineage. They state that the information about gene duplication obtained in one fish order cannot be extended to another; There may be two copies in the zebrafish, yet only one in the turbot, as for *RXRβ* for example [6]. However, we found that this new turbot sequence is more closely related to one of the two zebrafish *RXRβ* genes, *RXRε*, than it is to the other, *RXRδ* (Figure 1). The topology of the phylogeny shown in Figure 1 supports the hypothesis that the duplication of *RXRβ* genes occurred before the divergence of the zebrafish and turbot ancestors and it is therefore consistent with the ancient fish-specific genome duplication hypothesis. The second turbot *RXR* gene, the orthologue of zebrafish *RXRδ*, appears to have been lost or is waiting to be discovered. The *RXRβ* phylogeny makes an important point: to show that two species experienced the same gene or genome duplication event, it is not necessary to find two genes in both species.

In summary, Robinson-Rechavi and co-workers restricted their comparisons to sets of orthologous genes with representatives from at least three different fish orders and this seriously limited their ability to test the fish-specific genome duplication hypothesis. Using a phylogenetic approach, we have uncovered a large number of anciently duplicated genes in the zebrafish [19,20] and very few of these genes have been characterised in species from three different actinopterygian orders. The comparisons of the numbers of lineage-specific and lineage-shared

gene duplicates presented by Robinson-Rechavi *et al*. are biased by the inclusion of Salmoniformes sequences, and perhaps sequences from other tetraploid species. More importantly, the number of recent gene duplicates has no bearing on the ancient fish-specific genome duplication hypothesis. And a lack of rigor in their phylogenetic analysis led them to date inaccurately at least one gene duplication event. We conclude that the frequently published view [6–8] that extra genes in fish were not produced during an ancient fish-specific genome duplication event is not supported by Robinson-Rechavi *et al*.'s data.

Acknowledgements

We thank Henner Brinkmann for comments on the manuscript and for help with phylogenetic analyses. *Takifugu rubripes* sequences were provided freely by the Fugu Genome Consortium for use in this correspondence only. J.S.T. is supported by a postdoctoral fellowship from NSERC (Canada). We also thank the Deutsche Forschungsgemeinschaft for grants to Y.VdP. (842/2-1) and A.M. (1725/2-1, 1725/3-1, 1725/4-1, 1725/5-1).

References

1. Amores A, Force A, Yan Y-L, Jolly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L *et al.*: Zebrafish ***hox*** clusters and vertebrate genome evolution. *Science* 1998, 282:1711-1714.
2. Wittbrodt J, Meyer A, Schartl M: More genes in fish? *BioEssays* 1998, 20:511-515.
3. Gates MA, Kim L, Egan ES, Cardozo T, Sirotkin HI, Dougan ST, Lashkari D, Abagyan R, Schier AF, Talbot W: A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res* 1999, 9:334-347.
4. Postlethwait JH, Yan Y-L, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z *et al.*: Vertebrate genome evolution and the zebrafish gene map. *Nat Genet* 2000, 18:345-349.
5. Naruse K, Fukamachi S, Hirochi M, Kondo M, Matsuoko T, Hanamura N, Morita Y, Hasegawa K, Hishigaki R *et al.*: A detailed linkage map of Medaka, ***Oryzias latipes***: Comparative genomics and genome evolution. *Genetics* 2000, 154:1773-1784.
6. Robinson-Rechavi M, Marchand O, Escriva H., Bardet P-L, Zelus D, Hughes S, Laudet V: Euteleost fish genomes are characterized by expansion of gene families. *Genome Res* 2001, 11:781-788.
7. Robinson-Rechavi M, Marchand O, Escriva H, Laudet V: An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Curr Biol* 2001, 11:R458-R459.
8. Robinson-Rechavi M, Laudet V: Evolutionary rates of duplicated genes in fish and mammals. *Mol Biol Evol* 2001, 18:681-683.
9. Duret L, Mouchiroud D, Gouy M: HOVERGEN: a database of homologous vertebrate genes. *Nucleic Acids Res* 1994, 22:2360-2365.
10. Li W-H: Rate of gene silencing at duplicated loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* 1980, 95:237-258.
11. Lynch M, Conery JS: The evolutionary fate and consequences of duplicate genes. *Science* 2000, 260:1151-1155.
12. Uyeno T, Smith GR: Tetraploid origin of the karyotype of catostomid fishes. *Science* 1972, 175:644-646.
13. Allendorf FW: Protein polymorphism and the rate of loss of duplicate gene expression. *Nature* 1978, 272:76-78.
14. Ferris SD, Whitt GS: Loss of duplicate gene expression after polyploidization. *Nature* 1977, 265:258-260.
15. Dingerkus G, Howell WM: Karyotypic analysis and evidence of tetraploidy in the North American paddlefish, ***Polydon spathula***. *Science* 1976, 194:842-844.
16. Risinger C, Larhammar D: Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. *Proc Natl Acad Sci USA* 1993, 90:10598-10602.
17. Schmidtke J, Schulte B, Kuhl P, Engel W: Gene action in fish of tetraploid origin. V. Cellular RNA and protein content and enzyme activities in cyprinid, clupeoid, and salmonid species. *Biochem Genet* 14:975-980.
18. Jones BB, Ohno CK, Allenby G, Boffa MB, Levin AA, Grippo JF, Petkovich M: New retinoid X receptor subtypes in zebrafish (***Danio rerio***) differentially modulate transcription and do not bind 9-cis retinoic acid. *Mol Cell Biol* 1995, 15:5226-5234.
19. Taylor JS, Van de Peer Y, Braasch I, Meyer A: Comparative genomics provides evidence for an ancient genome duplication event in fish. *Phil Trans R Soc Lond B* 2001, 356:1661-1679.
20. Van de Peer Y, Taylor JS, Braasch I, Meyer A: The ghosts of selection past: rates of evolution and functional divergence in anciently duplicated genes. *J Mol Evol* 2001, 53:434-444.
21. Van de Peer Y, De Wachter Y: TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Applic Biosci* 1994, 10:569-570.
22. Felsenstein J: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985, 39:783-791.

Addresses: *Department of Biology, University of Konstanz, D-78457 Konstanz, Germany. †Department of Genetics, Vlaams Instituut voor Biotechnologie (VIB), University of Gent, B-9000, Gent, Belgium.
Correspondence: John Taylor
E-mail: john.taylor@uni-konstanz.de