Minireview

Tetraodon genome confirms *Takifugu* findings: most fish are ancient polyploids

Yves Van de Peer

Address: Bioinformatics and Evolutionary Genomics, Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. E-mail: yves.vandepeer@psb.ugent.be

Published: 25 November 2004

Genome ***Biology*** 2004, **5:**250

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2004/5/12/250

Abstract

An evolutionary hypothesis suggested by studies of the genome of the tiger pufferfish *Takifugu rubripes* has now been confirmed by comparison with the genome of a close relative, the spotted green pufferfish *Tetraodon nigroviridis*. Ray-finned fish underwent a whole-genome duplication some 350 million years ago that might explain their evolutionary success.

In 1993, Sydney Brenner and colleagues [1] proposed sequencing the pufferfish genome as a cost-effective way to identify and characterize human genes. The genome of the pufferfish is only about one-eighth of the size of that of human but was expected to contain a similar gene repertoire. Ten years later, not only has a draft genome sequence been released for *Takifugu rubripes* (Fugu, also known as the Japanese or tiger pufferfish) [2], but also for *Tetraodon nigroviridis* (green spotted pufferfish) [3], a close relative that diverged from *Takifugu* 18-30 million years ago (Mya). By comparing the two pufferfish genomes with that of human, several hundred novel human genes have already been uncovered, as was predicted by Brenner and colleagues [1]. But the pufferfish genome sequencing projects have also yielded a surprising finding: ray-finned fish (Actinopterygii), such as pufferfish might have more genes than lobe-finned fish (coelacanths and lungfish) and land vertebrates, because of additional gene-duplication events [4]. The recent release of the *Tetraodon* genome sequence [3] provides overwhelming evidence that a genome-duplication event did indeed occur early in the evolution of ray-finned fish.

A fish-specific genome duplication

Some of the first data pointing to a possible genome duplication in fish came from *Hox* genes and *Hox* gene clusters. *Hox* genes encode DNA-binding proteins that specify cell fate along the anterior-posterior axis of bilaterian animal embryos and occur in one or more clusters of up to 13 genes. Whereas lobe-finned fish, amphibians, reptiles, birds and mammals have four clusters, extra *Hox* gene clusters have been discovered in zebrafish, *Medaka*, Nile tilapia and pufferfish [4]. The observation that such distantly related species [5] all have seven or eight *Hox* gene clusters suggested the occurrence of an additional genome-duplication event in the ray-finned fish lineage before the divergence of most teleost (bony fish) species. More recent comparative genomic studies have turned up many more genes and gene clusters for which there are two copies in fish but one in other vertebrates [6]. The findings that different paralogous pairs seem to have originated at about the same time, that different fish species seem to share ancient gene duplications, and that different paralogs are found on different linkage groups in the same order as other duplicated genes, all support the hypothesis that these genes arose through a large-scale gene-duplication event. It is worth noting, however, that some authors have argued that an ancestral whole-genome-duplication event was not responsible for the abundance of duplicated fish genes [7].

Additional evidence for a genome duplication in ray-finned fish was provided by analyzing the complete *Takifugu* genome, a draft sequence of which was published in 2002 [2]. Two recent studies identified duplicated genes in this genome and used phylogenetic trees to estimate the ages of

these duplicates [8,9]. Vandepoele *et al.* [8] constructed phylogenetic trees for all gene families containing between two and ten duplicated *Takifugu* genes, which amounts to a total of 3,077 families. For each gene family, the relative date of duplication events was determined to test whether gene duplications occurred before or after the split between fish and land vertebrates. To this end, neighbor-joining trees were created for each of the *Takifugu* gene families with homologous sequences from mouse and human. Absolute dating of duplication events was achieved through inference from linearized trees [10]. In such trees - where branch length is directly proportional to time - the split between ray-finned fish and land vertebrates, dated at 450 Mya, was used as a calibration point for the dating of gene-duplication events. A major fraction (about one-third) of the duplicated genes in *Takifugu* could be ascribed to a large-scale gene-duplication event specific to the fish lineage, which was estimated to have occurred about 320 Mya (Figure 1).

A very similar approach was followed for the analysis of the *Takifugu* genome by Christoffels *et al.* [9], who obtained essentially the same result: by constructing linearized trees, the whole-genome-duplication event was estimated to have occurred approximately 350 Mya. To test whether the sudden increase in the number of duplicated genes in the *Takifugu* genome was the result of an entire-genome duplication rather than an increased rate of independent tandem-duplication events, both Vandepoele *et al.* [8] and Christoffels *et al.* [9] investigated the appearance of duplicated genes in duplicated blocks on chromosomes. Statistically significant regions of micro-colinearity were identified within the complete *Takifugu* genome, showing the same gene content and gene order. Indeed, both studies reported a large number of duplicated genes in so-called paralogons - homologous genomic segments that can be proved to have been created by duplication [11] - and concluded that most findings were congruent with a large-scale, probably whole-genome duplication event in a ray-finned ancestor that gave rise to the *Takifugu* and other fish lineages.

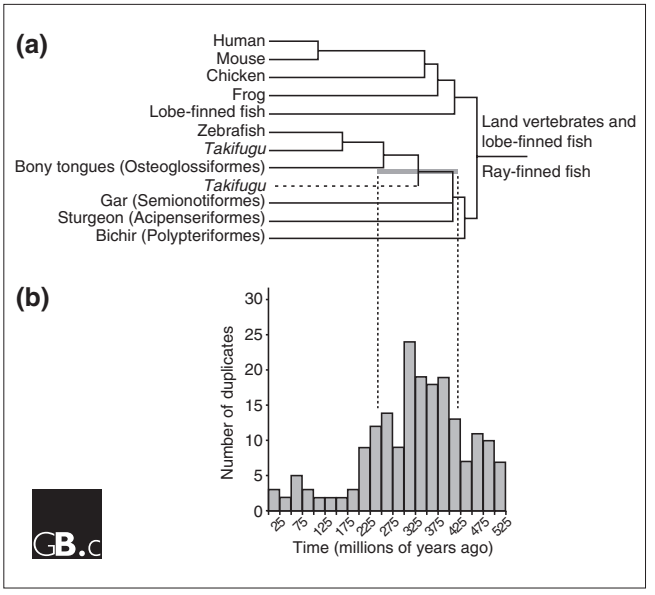

Figure 1
A phylogenetic tree showing the vertebrate phylogenetic relationships and superimposed pufferfish gene-duplication events. **(a)** A generally accepted tree illustrating the relationships between several vertebrate species. The gray horizontal bar denotes the fish-specific genome-duplication event inferred from absolute dating of *Takifugu* paralogs. The broken line indicates the position of the duplicated copy of the *Takifugu* genome that originated between the divergence of gar and the bony tongues. **(b)** The bar chart shows the number of paralogous genes that could be dated through the construction of linearized trees. Modified from [8,9].

Comparing genomes

Because of the highly fragmented nature of the initial *Takifugu* genome assembly, it was difficult to prove that the large-scale gene-duplication event had indeed affected the whole genome. The recent release of the well-assembled *Tetraodon* genome [3] seems to have settled this issue in two ways. First, Jaillon *et al.* [3] analyzed the chromosomal distribution of ancient duplicates and observed that genes on one chromosome have a strong tendency to have duplicate copies on a single other chromosome. As would be expected from a whole-genome-duplication event, all chromosomes are involved. Second, by using a comparative approach in which they compared the *Tetraodon* genome with that of human, which has not undergone the genome-duplication event (Figure 1), Jaillon *et al.* [3] showed that almost every region in the human genome clearly corresponds to two regions in the *Tetraodon* genome. This type of comparative analysis (Figure 2) has proved very powerful for unveiling genome-duplication events. Recently, such an approach provided overwhelming evidence for the long-suggested [12], but contested (see, for example, [13]) ancient whole-genome duplication in the yeast *Saccharomyces cerevisiae*, by comparing its genome with that of different relatives that diverged prior to the duplication events [14,15].

A comparative analysis between the human and *Tetraodon* genomes has also allowed inference of the basic structure of the ancestral bony vertebrate genome, and the reconstruction of much of the evolutionary history of ancient and recent chromosomal rearrangements leading to the modern human karyotype. By matching up the genes on the *Tetraodon* chromosomes with homologs on human chromosomes, Jaillon *et al.* [3] inferred that the ancestor of both fish and land vertebrates had no more than 12 chromosomes, a number that has been previously suggested on the basis of linkage relationships between zebrafish, *Medaka*, and human [16]. Comparison of the genomes of *Tetraodon* and human also showed that chromosome evolution in both

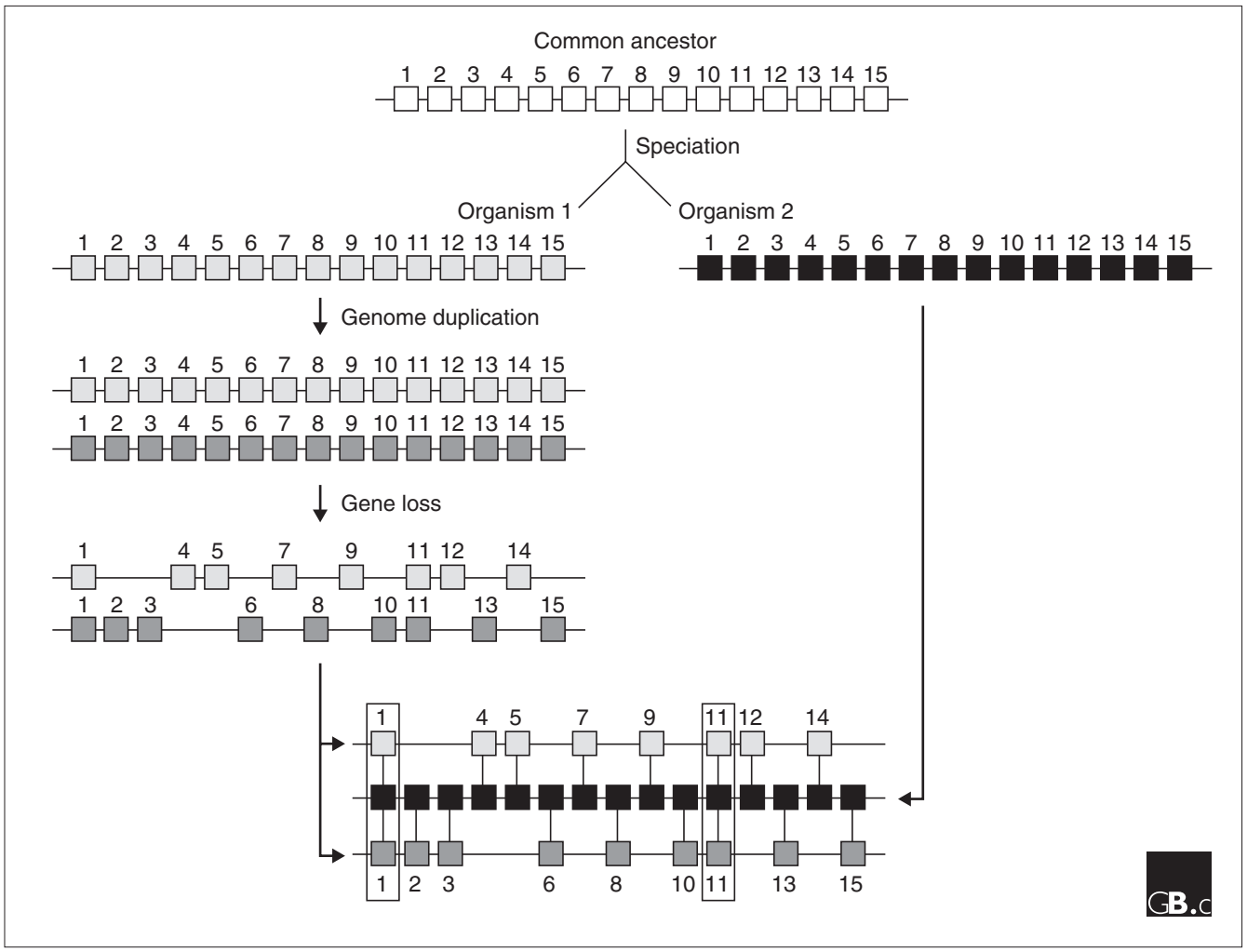

Figure 2
Uncovering genome duplications through comparative analysis with related sequences. The hypothetical genomes of two related organisms are shown, each containing the same set of genes. Both genomes are initially identical, but the genome of Organism 1 is duplicated, resulting in a second identical set of chromosomes and genes. After some time, homologous chromosomes lose a different set of genes, keeping two copies for only a minority of the duplicated genes. For the sake of simplicity, the genome of Organism 2 is assumed to remain unchanged. Within Organism 1, the only evidence for a duplication event comes from the conserved order of the anchor points formed by genes 1 and 11 (indicated by boxed regions). Comparison with the genome of Organism 2, however, shows a pattern of so-called 'double conserved synteny' where the duplicated nature of Organism 1 is revealed.

lineages differed considerably. Whereas all but one of the ancestral *Tetraodon* chromosomes had not undergone interchromosomal exchange for 450 Mya, only one human chromosome was similarly undisturbed. A possible explanation for the difference in genome evolution might be the massive integration of transposable elements in the human genome, with an increased overall frequency of chromosome breaks as a result [3].

Evolutionary implications

As mentioned above, on the basis of previous analyses of the *Takifugu* genome, the whole-genome-duplication event in fish is thought to have occurred somewhere between 300 and 350 Mya [8,9]. An interesting question is whether this date correlates with a decisive period in the evolution of the fish. For instance, if the genome duplication had been responsible for the biological diversification and large number of ray-finned fish, as suggested previously [4,17], it must have occurred prior to the radiation of most fish lineages. The class Actinopterygii includes more than 23,500 species [18], of which the vast majority are teleosts or ray-finned fish. Interestingly, all older, more basal groups of ray-finned fish, namely Polypteriformes (bichirs), Acipenseriformes (sturgeons and paddlefish), Semionotiformes (gars), and Amiformes (bowfin), have only a few extant species (Figure 1). Most members of these basal actinopterygian lineages are considered to be 'living

fossils', because their morphology has remained unchanged over very long evolutionary time periods.

In a recent study, Hoegg *et al*. [19] have tried to determine the timing of the duplication event in relation to the origin of lineages of teleost and 'nonteleost' fish by sequencing three nuclear genes - *fzd8*, *sox11* and *tyrosinase* - from sturgeons, gars, bony tongues, and a tenpounder. For these three genes, two copies have been described previously in derived teleost model species, such as zebrafish and pufferfish, but only one orthologous copy has been found in tetrapods. The specific clustering of the genes in individual gene trees for these three genes and a dataset of concatenated genes support the hypothesis that the fish-specific genome-duplication event took place after the split of the Acipenseriformes and the Semionotiformes from the lineage leading to teleost fish, but before the divergence of Osteoglossiformes (bony tongues) and the other more derived groups of fish (Figure 1). This is in good agreement with the recent analyses of the *Takifugu* genome, as fossil data age the Semionotiformes at between 245 and 286 million years, whereas molecular estimates for the Amiiformes, which are of approximately the same age as the Semionotiformes, hint at a separation from the Teleostei stem lineage about 367-404 Mya. Likewise, molecular data suggest an age of 335 million years for the Osteoglossiformes [19]. The inferred relative and absolute dates for the fish-specific genome duplication event seem to separate the species-poor branching lineages from the species-rich teleost lineages, providing evidence that the fish-specific genome duplication might be related causally to an increase in species and morphological diversity.

On the basis of isozyme studies, Werth and Windham [20] developed a model in which the 'reciprocal silencing' of genes in geographically separated populations would promote speciation. A few years ago, this idea was revived in a model called 'divergent resolution', in which the loss or silencing of gene duplicates was postulated to be more important for the evolution of species diversity than the acquisition of new functions by duplicated genes. Divergent resolution occurs when different copies of a duplicated gene are lost on different chromosomes in different populations, thereby creating genetic barriers for reproduction between them [21,22]. Divergent resolution and lineage-specific subfunction partitioning [17] can promote incompatibility among populations within a species, and thus might facilitate evolutionary radiation. Gene duplications might, therefore, bring about rapid speciation in populations fixed for different copies of a duplicated locus. The fish-specific genome duplication has created many duplicates that could be divergently resolved. Potentially, such genes have played a prominent role in the radiation of the teleosts. Further studies of the genes encoded in these fish genomes may shed light on how important the fish-specific whole-genome duplication has been in the evolution of the ray-finned fish.

References

1. Brenner S, Elgar G, Sandford R, Macrae A, Venkatesh B, Aparicio S: **Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome.** *Nature* 1993, **366:**265-268.
2. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, *et al.*: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*.** *Science* 2002, **297:**1301-1310.
3. Jaillon O, Aury J-M, Brinet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozoef-Costaz C, Bernot A, *et al.*: **Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431:**946-957.
4. Meyer A, Schartl M: **Gene and genome duplications in verterbrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions.** *Curr Opin Cell Biol* 1999, **11:**699-704.
5. Chen W-J, Orti G, Meyer A: **Novel evolutionary relationships among four fish model systems.** *Trends Genet* 2004, **20:**424-431.
6. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer, Y: **Genome duplication, a trait shared by 22,000 species of ray-finned fish.** *Genome Res* 2003, **13:**382-390.
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. Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: paranome age and size differs considerably between ray-finned fishes and land vertebrates.** *Proc Natl Acad Sci USA* 2004, **101:**1638-1643.
9. Christoffels A, Koh EG, Brenner S, Aparicio S, Venkatesh B: ***Fugu* genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes.** *Mol Biol Evol* 2004, **21:**1146-1151.
10. Takezaki N, Rzhetsky A, Nei M: **Phylogenetic test of the molecular clock and linearized trees.** *Mol Biol Evol* 1995, **12:**823-833.
11. Van de Peer Y: **Computational approaches to unveiling ancient genome duplications.** *Nat Rev Genet* 2004, **5:**752-763.
12. Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome.** *Nature* 1997, **387:**708-713.
13. Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, *et al.*: **Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of redundancy compared to *Saccharomyces cerevisiae*.** *FEBS Lett* 2000, **487:**122-133.
14. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, *et al.*: **The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome.** *Science* 2004, **304:**304-307.
15. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.** *Nature* 2004, **428:**617-624.
16. Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H: **A Medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping.** *Genome Res* 2004, **14:**820-828.
17. Postlethwait P, Amores A, Cresco W, Singer A, Yan Y-L: **Subfunction partitioning, the teleost radiation and the annotation of the human genome.** *Trends Genet* 2004, **20:**481-490.
18. Nelson J: *Fishes of the World*. New York: Wiley; 1994.
19. Hoegg S, Brinkmann H, Taylor JS, Meyer A: **Phylogenetic timing of the fish-specific genome duplication correlates with phenotypic and taxonomic diversification in fishes.** *J Mol Evol* 2004, **59:**190-203.
20. Werth CR, Windham MD: **A model for divergent allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate gene expression.** *Am Nat* 1991, **137:**515-526.
21. Lynch M, Force A: **The origin of interspecific genomic incompatibility via gene duplication.** *Am Nat* 2000, **156:**590-605.
22. Taylor JS, Van de Peer Y, Meyer A: **Genome duplication, divergent resolution and speciation.** *Trends Genet* 2001, **17:**299-301.