Rearrangement Rate following the Whole-Genome Duplication in Teleosts

Marie Sémon and Kenneth H. Wolfe

Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

It is now clear that a whole-genome duplication (WGD) occurred at the base of the teleost fish lineage. Like the other anciently polyploid genomes investigated so far, teleost genomes now behave like diploids with chromosomes forming pairs at meiosis. The diploidization process is currently poorly understood. It is associated with many gene deletions, such that one of the duplicates is lost at most loci and has also been proposed to coincide with an increase in genomic instability. Here we ask whether WGD is a determinant of the genomic rearrangement rate in teleosts. We study variability of the rates of rearrangement along a vertebrate phylogenetic tree, composed of 3 tetrapods (human, chicken, and mouse) and 3 teleost fishes (zebrafish, *Tetraodon*, and *Takifugu*), whose complete genome sequences are available. We devise a simple parsimony method for counting rearrangements, which takes into account various methodological complications caused by the WGD and the subsequent gene losses. We show that there does appear to be an increase in rearrangement rate after WGD, but that there is also a great deal of additional variability in rearrangement rates across species.

Introduction

With almost 24,000 living species (Nelson 1994), teleosts are the most diversified group of vertebrates. They constitute more than 96% of the ray-finned fishes (Actinopterygii), which diverged from the lobe-finned fishes lineage (Sarcopterygii) in the early Devonian (Nelson 1994; Basden et al. 2000). Most of these fishes are diploid organisms, that is, their chromosomes are arranged in bivalents (pairs) at meiosis. However, it has long been suspected that a whole-genome duplication (WGD) occurred in the ray-finned fish lineage because many gene families contain more genes in teleost fishes than in mammals (Amores et al. 1998; Meyer and Schartl 1999; Robinson-Rechavi et al. 2001). This was finally verified by the comparative analysis of the *Tetraodon* (*Tetraodon nigroviridis*) and human genomes (Jaillon et al. 2004): many human genome segments map to 2 different *Tetraodon* chromosomes, a characteristic signature of a polyploidization event. Recent studies, dating the divergence between gene duplicates using a molecular clock (Christoffels et al. 2004; Vandepoele et al. 2004) or localizing the duplication in phylogenetic trees (Hoegg et al. 2004; Crow et al. 2006), have shown that the WGD occurred 320–404 MYA.

Immediately after this WGD, each gene was present in 2 copies and chromosomes were arranged in tetravalents (containing 2 homeologous pairs of chromosomes) at meiosis. In the time since the WGD, the genomes have largely returned to a diploid form, and modern teleost genomes are described as paleopolyploid (i.e., genetically diploid but structurally polyploid) (Jaillon et al. 2004). The process of diploidization is not yet well understood (Comai 2005; Ma and Gustafson 2005; Chen and Ni 2006). For example, it is not known whether in most paleopolyploids the switch from multivalent to bivalent meiosis occurred through changes in a small number of master chromosome–pairing genes, such as the wheat *Ph1* locus (Griffiths et al. 2006), or through a more gradual loss of pairing between homeologs as a result of the loss of one copy of most genes.

Key words: whole-genome duplication, rearrangement rates, teleost, vertebrate.

E-mail: khwolfe@tcd.ie.

Mol. Biol. Evol. 24(3):860–867. 2007
doi:10.1093/molbev/msm003
Advance Access publication January 11, 2007

© The Author 2007. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

It has also been hypothesized that the rate of genomic rearrangement increases after WGD (Comai 2005). There are at least 2 reasons why this might be the case. First, WGD creates many homologous regions, which can provide a substrate for recombination and therefore increase the neutral rate of rearrangements. For example, WGD doubles the sizes of multigene families, increasing the likelihood of ectopic exchanges. Similarly, it doubles the content of other homologous sequences such as repeated elements, and these are known to promote rearrangements: imperfect exchange promoted by Alu elements during meiosis has been associated with human diseases, for instance (Deininger et al. 2003). Alternatively, an increase in rearrangement rates may be selectively favored if it leads to structural divergence of chromosomes and thus facilitate the separation of homeologous chromosomes at meiosis. The tetravalents created by WGD may be deleterious because they could increase the likelihood of chromosomal nondisjunction (Comai 2005). Under both of these scenarios, we expect the rearrangement rate to increase directly after WGD and then to decrease again as diploidization progresses.

There is currently no clear evidence that genomic rearrangements increase after WGD, although there are suggestive results from synthetic plant polyploids, in which an increase in rearrangement rates has been observed (Song et al. 1995; Pontes et al. 2004; Chen and Ni 2006). To perform a general and comprehensive test in vertebrates of whether the genomic rearrangement rate increases directly after WGD and subsequently decreases, we must calculate genome-wide rearrangement rates for multiple post-WGD species (teleosts) and nonduplicated outgroups (tetrapods).

Previous studies have compared gene order between human and fish or rearrangement rates between multiple species that have not undergone genome duplication. However, the heterogeneity of methods used makes it difficult to compare the results. Several pairwise comparative analyses of gene order between teleost fish and humans have been carried out, beginning well before the availability of whole-genome sequences (Amores et al. 1998; McLysaght et al. 2000; Naruse et al. 2000; Postlethwait et al. 2000; Woods et al. 2000; Bernot and Weissenbach 2004). Subsequent analyses with the complete genomes of *Tetraodon*, zebrafish, and *Takifugu* (*Takifugu rubripes*) have largely confirmed that synteny (gene composition of the chromosomes) tends to be conserved between human and fish, but that the precise gene order along the syntenic chromosomes is often

different in the 2 species (Jaillon et al. 2004; Naruse et al. 2004; Woods et al. 2005). Furthermore, interchromosomal rearrangements are much more frequent in the human lineage than in the teleost lineages (Jaillon et al. 2004; Naruse et al. 2004; Kai et al. 2005; Woods et al. 2005).

Methods have also been developed to compare whole-genome organization across multiple species (Burt et al. 1999; Bourque et al. 2004, 2005; Hillier et al. 2004; Murphy et al. 2005), sometimes with the reconstruction of ancestral chromosomes. However, none of these methods have incorporated data from paleopolyploid species (except yeast; Fischer et al. 2006), possibly because WGD complicates comparative mapping across genomes.

The calculation of relative genome-wide rearrangement rates across a data set that includes both nonduplicated and duplicated genomes is made nontrivial by the initial doubling of chromosome number and subsequent massive gene loss. Here we propose a new method to allow direct comparisons of rearrangement rates in WGD and non-WGD species. Using this method we measure the level of gene-order conservation for each branch of a tree containing several teleost and tetrapod species whose complete genomes are available. Our data are compatible with an increase of rearrangement rate after WGD, but the rearrangement rate also varies considerably across species.

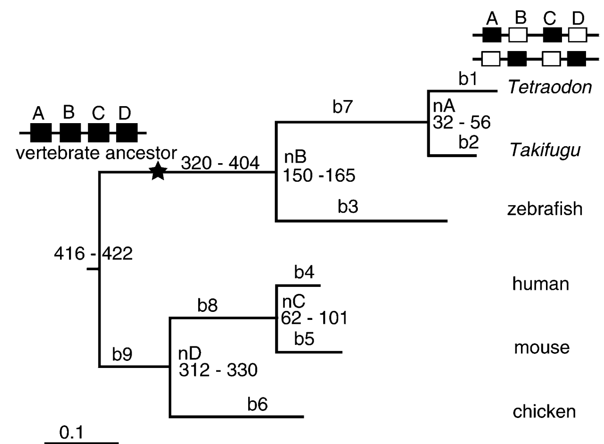

FIG. 1.—Phylogenetic relationships among the 6 vertebrate species under study. We have included 3 teleost species (*Tetraodon*, *Takifugu*, and zebrafish) and 3 nonteleost species (chicken, human, and mouse) for which whole-genome sequences are available. The star indicates the place where the WGD event occurred in the tree. The branch and nodes are labeled as used in the text. b9 is the branch between nB and nD. The branch lengths represent amino acid sequence divergence (scale bar is represented under the tree), and the divergence time estimated using fossil data are shown on each node of the tree.

Methods

Data Set Construction

We selected sequences from Homolens (Penel S, Duret L, personal communication, http://pbil.univ-lyon1.fr/databases/homolens.html), a database relying on the Ensembl release 36 annotations (Hubbard et al. 2005, release November 2005). We built our data set of orthologs in the 6 genomes under study (*Homo sapiens*, *Mus musculus*, *Gallus gallus*, *T. nigroviridis*, *T. rubripes*, and *Danio rerio*) using a phylogenetic approach (Dufayard et al. 2005). We discarded all the trees that were not compatible with the species tree or that contained duplicates. In paleopolyploid teleost genomes, most genes are single copy. In most cases all the species have the same orthologous copy, but sometimes paralogous copies are retained in different species (Sémon and Wolfe 2007). Because the tree has the same topology in both cases, it was unfortunately not possible to discriminate between orthologs and paralogs between zebrafish and *Tetraodon* (or between zebrafish and *Takifugu*). Inclusion of these genes artificially inflates the estimated rate of interchromosomal rearrangement. We obtained 1,150 sets of mapped orthologs, from which we identified 1,128 pairs of proximate genes in the human genome, 1,131 in mouse, 1,112 in chicken, 1,125 in *Tetraodon*, but only 930 pairs in zebrafish and 877 pairs in *Takifugu*. The number of pairs P for each species is given by $P = N - C$, where N is the number of orthologs (1,150) and C is the number of chromosomes or contigs onto which genes are mapped in that species. We ignore all genes other than our set of 1,150 orthologs.

Branch Length Computation

The topology of the phylogenetic tree for the 6 species considered here is widely accepted (Nelson 1994; Wittbrodt et al. 2002; fig. 1). There appears to be a considerable variation in divergence time estimates, especially depending on the type of data used (molecular or fossils; Yamanoue et al. 2006). We have estimated the branch lengths for this tree in 2 ways. First, we concatenated the protein alignments corresponding to each of the ortholog sets. We then estimated amino acid divergence using Tree-Puzzle (Schmidt et al. 2002), forcing the topology to follow the species tree. The resulting branch lengths are shown in figure 1.

In parallel, we used dates from the fossil record to calibrate the tree. A recent paper provides geological evidence for the age of splits between all of the animal genome models. The authors cite the appropriate paleontology literature and propose precise minimum and looser maximum age constraints for 30 divergences among key genome model organisms (Benton and Donoghue 2006). They proposed 61.5–100.5 Myr for the human–mouse divergence, 312.3–330.4 Myr for mammals–birds, and 416.1–421.8 Myr for actinopterygian–sarcopterygians. According to the date of radiation of tetraodontids estimated using the fossil record, *Tetraodon* and *Takifugu* diverged 32.3–56.0 MYA. The divergence of zebrafish and *Tetraodon* occurred 149.9–165.2 MYA.

Rearrangement Rate Estimation

We measured the conservation of proximity between 2 genes in our set as an approximation of the rearrangement rate (McLysaght et al. 2000). Here, we define 2 genes to be “proximate” in a species if no other gene in our set of 1,150 orthologs lies between them. We calculated the proportion of pairs of genes that are proximate at a particular node on the tree (start point) and that are still closer than a specific distance at another node (end point). This presupposes that break points are not reused because if they are, our method will only take the biggest rearrangement into account. For clarity, we use the notation for numbering of nodes and

internal branches as shown in figure 1. We defined as "intrachromosomal" rearrangement a case where the genes are proximate at the start point and are still located on the same chromosome, but are separated by at least one intervening gene (in our set of 1,150 orthologs) at the end point. We defined as "interchromosomal" rearrangement a case where the genes are proximate at the start point and are not located on the same chromosome in the end point.

We used a simple parsimony method to estimate the conservation of proximity in comparisons internal to teleost fishes or to tetrapods. For instance, to estimate the conservation of proximity on the terminal branch b4 (fig. 1), we evaluated the proportion of gene pairs that were proximate at the time of divergence between human and mouse (node nC) and that are not in proximity in human anymore. To find a set of genes that were proximate at the node nC, we took the pairs of genes that are proximate in chicken and in mouse today, which can be inferred by parsimony to have been proximate at nC. Of course, this is not the complete set of genes that were proximate at nC, but it should be unbiased with regard to the fate of the pair in the human lineage. We then split this data set according to their proximity in human, to obtain the fractional conservation of proximity on branch b4. A similar approach was used to estimate the rearrangement rates in the mouse, *Tetraodon*, and *Takifugu* lineages (branches b5, b1, and b2; fig. 1).

To estimate the rearrangement rate in the chicken lineage (branch b6, fig. 1), we needed to start with a data set of gene pairs that were in proximity in the ancestor of birds and mammals (nD). By parsimony, any gene pair that is proximate in at least 1 fish genome and 1 mammalian genome should have been proximate at node nD. Because of the WGD, it was not trivial to obtain this subset of pairs. Take for instance 4 genes that were ordered ABCD in the ancestral vertebrate genome (fig. 1). After the WGD and the episode of massive gene losses that occurred in the fish lineage, we expect to observe 2 segments in teleosts, for example, AC and BD, mapping on paralogous chromosomes. Therefore, a pair of proximate genes BC in human can lose proximity in teleost fishes, without any rearrangements other than gene deletion after the fish WGD (fig. 1). We used information contained in both fish segments and counted no rearrangement between B and C after the WGD if both AC and BD are proximate in fish. A similar process was used to evaluate the rearrangement rate between node nD and human (branches b8 + b4) and node nD and mouse (branches b8 + b5), which then allows us to estimate b8 by subtraction of the previously calculated values of b4 or b5.

We used a similar method to estimate the rearrangement rate on branch b3, the zebrafish lineage after the split between zebrafish and *Tetraodon* (nB). By parsimony, a pair of genes that is linked in *Tetraodon* (or in *Takifugu*) and in chicken (or human or mouse) must have been linked at nB. Because of the WGD and the subsequent gene losses, genes that are proximate in fish are not necessarily proximate in human even in the absence of rearrangements (fig. 1). Therefore, we pick pairs of genes that are proximate in fish and separated by at most 2 genes in chicken and human (or chicken and mouse), and then we test whether they are proximate in zebrafish to obtain the fraction of rearrangements on branch b3. We performed this analysis on a subset of the complete data set containing only 858 ortholog families, where the zebrafish gene has been mapped on to a chromosome.

The estimation of the rearrangement rate on branch b9 is also complicated by WGD. We illustrate our approach here for the example of the branch between the split *Tetraodon*/zebrafish (node nB) and the chicken genome (branch b9 + b6). We need first to build a data set X containing gene pairs that were linked at nB, which was done by taking pairs that are proximate in zebrafish and *Tetraodon* (or zebrafish and *Takifugu*). Then, we need to count the rearrangements that occurred between the teleost ancestor (nB) and the chicken genome. Take our 4 genes, ABCD, that are proximate in the vertebrate ancestor and that belong to 2 segments, AC and BD, in all teleost genomes after the diploidization. AC and BD will both belong to our data set X. If a rearrangement occurred between AB and CD somewhere on the chicken lineage after the vertebrate ancestor, we will count it twice, once when examining the pair AC and a second time when examining BD. To address this problem, we considered only once each segment of 4 genes that are proximate in the vertebrate ancestor. For this, we randomly picked a subset X′ of the data set X. More precisely, we randomly chose 1 pair of genes MN of the data set X and we kept this pair in the new data set X′. Then we picked all of the genes P that were proximate to M or N in the chicken genome and removed all of the pairs containing M, N, or 1 of the genes P from the data set X. We continued to pick pairs of genes in the data set X until no gene was left in the data set, and then we obtained a data set X′ for which we could safely measure the rate of proximity retention between nB and the chicken genome. We repeated the construction of X′ 100 times to stabilize the result. A similar process was used to compute the rearrangement rates on the branches b9 + b8 + b4 (from nB to the human genome) and b9 + b8 + b5 (from nB to the mouse genome).

We obtained the estimates of rearrangement rates on internal branches by subtraction. We used where possible the most reliable estimates of each rearrangement rate. For instance, we obtained rearrangement rate on the branch b7 (between the nodes nA and nB) using the rearrangement rates between the node nA and *Tetraodon*, and not *Takifugu*, because rearrangements on the *Takifugu* lineage are difficult to estimate due to the short scaffold length in this genome sequence (Aparicio et al. 2002). We calculated the rearrangement rates on the branch b8 using b8 + b4 and b9 using b9 + b6.

Results

We estimated the rates of rearrangement by parsimony along a vertebrate phylogenetic tree (fig. 1), composed of 3 tetrapods (human, chicken, and mouse), and 3 teleost fishes (zebrafish, *Tetraodon*, and *Takifugu*), whose complete sequences are available in Ensembl (Hubbard et al. 2005).

Conservation of Proximity between Species

We sampled 1,150 sets of 6 orthologous genes (one for each of the studied species) using the Homolens database

Table 1
Extent of Inter- and Intrachromosomal Rearrangements between Pairs of Species

(a) Same chromosome: number of gene pairs remaining on the same chromosome of end-point and start-point species

		End point						
		Takifugu	*Tetraodon*	Zebrafish	Zebrafish chromosome	Mouse	Human	Chicken
Start point	*Takifugu*	877*	871	358*	351	382	485	554
	Tetraodon	861*	1,125	412*	417	450	562	650
	Zebrafish	269*	498	930		271	319	371
	Zebrafish chromosome	237	453		858	244	288	335
	Mouse	290*	535	267*	263	1,131	1,034	877
	Human	303*	561	273*	270	1,013	1,128	932
	Chicken	309*	572	266*	268	812	927	1,112

(b) Proximity conserved: number of gene pairs remaining in proximity in the end-point and start-point species

		End point						
		Takifugu	*Tetraodon*	Zebrafish	Zebrafish chromosome	Mouse	Human	Chicken
Start point	*Takifugu*	877*	853	218*	194	222	232	234
	Tetraodon	854*	1,125	232*	209	245	256	251
	Zebrafish	219*	232	930		156	160	154
	Zebrafish chromosome	189*	201		858	134	138	135
	Mouse	223*	245	156*	137	1,131	930	619
	Human	233*	256	160*	143	930	1,128	668
	Chicken	233*	249	153*	137	619	667	1,112

(c) Proportion of rearrangements: proportion of gene pairs that are proximate in the start point but not syntenic in the end point

		End point						
		Takifugu	*Tetraodon*	Zebrafish	Zebrafish chromosome	Mouse	Human	Chicken
Start point	*Takifugu*		0.007	0.592*	0.600	0.564	0.447	0.368
	Tetraodon	0.235*		0.634*	0.629	0.600	0.500	0.422
	Zebrafish	0.711*	0.465			0.709	0.657	0.601
	Zebrafish chromosome	0.724*	0.472			0.716	0.664	0.610
	Mouse	0.744*	0.527	0.764*	0.767		0.086	0.225
	Human	0.731*	0.503	0.758*	0.761	0.102		0.174
	Chicken	0.722*	0.486	0.761*	0.759	0.270	0.166	

NOTE.—Each pair of species allows 2 different comparisons, depending on the species chosen as a start point. Figures in diagonals (underlined in a and b) indicate the number of pairs of proximate genes that were considered for each species. Stars indicate overestimates because the *Takifugu* and zebrafish genomes are not assembled into chromosomes. Zebrafish chromosome indicates comparisons involving only the gene families where the zebrafish gene is located on a chromosome and not on a scaffold. (a) Number of pairs where the genes are proximate in the start-point species and located on the same chromosome in the end-point species. (b) Number of pairs with conserved proximity: 2 genes are proximate in the start-point species and proximate (not separated by a gene in our data set) in the end-point species. (c) Proportion of rearrangements, considering interchromosomal rearrangements only. This is the ratio $(S - E)/S$, where S is the number of pairs studied in the start-point species and E is the number of pairs with conserved synteny in the end point (numbers from panel a).

(Penel S, Duret L, personal communication). Our data set represents one gene out of 15 to 20 in each genome, which is comparable to recent studies comparing 2 fish genomes (Naruse et al. 2004; Woods et al. 2005). The limited resolution of this approach means that we will miss any local rearrangement that occurs between 2 consecutive genes in our data sets, but this should not influence our estimates of the relative rates of rearrangement in different branches of the tree.

In this data set, we first simply measure the conservation of proximity (McLysaght et al. 2000) between each pair of species to estimate the extent of rearrangement. We calculate the proportion of genes that are proximate (i.e., not separated by another gene in our set of 1,150 orthologs) in one species (the start point), whose orthologs are closer than a specific distance in the other species (the end point). We performed separate analyses of synteny conservation (measuring interchromosomal breakage, table 1a) and proximity conservation (measuring both intra- and interchromosomal breakage, table 1b). As expected, the levels of synteny and proximity conservation diminish as the phylogenetic distance increases (table 1). We note some exceptions to this rule; for instance, among the 1,128 pairs of proximate genes in human, roughly the same number are located on the same chromosome in mouse (1,013), as in chicken (932).

Proximity conservation between 2 species can be measured in 2 directions, depending on the species we define as the start point. Both definitions should give similar results, but this is not true if the karyotype of one species contains more chromosomes than the other or if one assembly is partial as in *Takifugu* (which contains 2,847 scaffolds with at least one gene). A pair of genes proximate in human could be artificially separated in *Takifugu* because they map on different scaffolds: this leads to overestimates of proximity losses. To address this problem, when possible we consider

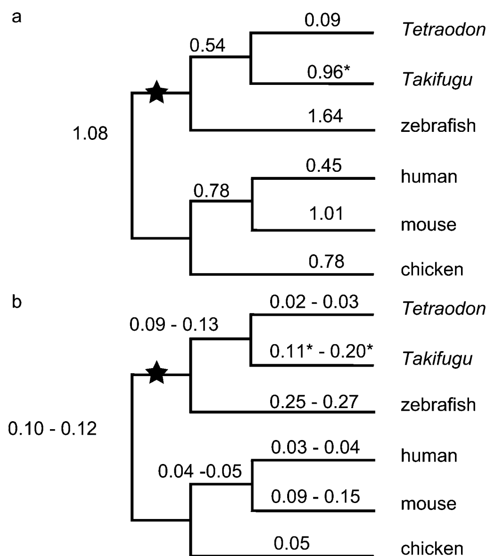

FIG. 2.—Rates of rearrangement in the 6 vertebrate species under study. Branch-specific rearrangement rates are indicated. These rates are obtained by normalizing our measure of proximity loss (sum of inter- and intrachromosome breakages) using (*a*) sequence divergence in the appropriate branch (units are proportional to proximity loss per amino acid substitution in our concatenated tree) or (*b*) estimates of the divergence time using the fossil record (units are proportional proximity loss per million years). The star indicates the place where the WGD event occurred in the tree. Asterisks indicate overestimates because the *Takifugu* genome is not assembled into chromosomes.

Takifugu as a start-point and not an end-point species. In zebrafish, we also analyze a second data set containing only 858 ortholog families where the zebrafish gene has been mapped on to a chromosome (table 1).

Parsimony Estimation of Rearrangement Rates on Each Branch of the Tree

We estimated the frequency of loss of proximity between pairs of genes on every branch of our vertebrate tree, taking into account complications caused by the WGD (see Methods). We then calibrated these estimates by using either sequence divergence as a proxy for time (fig. 2*a*) or dates from the fossil record (fig. 2*b*). Under the molecular clock hypothesis, the similarity between sequences from 2 species is directly proportional to the time that elapsed after their divergence. This is not always the case; the rate of substitution in both copies has been shown to accelerate after a gene duplication compared with the ancestral gene, for instance (Seoighe et al. 2003; Jordan et al. 2004). This may cancel with an acceleration of rearrangement rate after WGD, making our observations more conservative. The fossil record is not biased by this phenomenon and is therefore reliable for studies comparing diploid and paleopolyploid species. However, strong discrepancies exist between dates obtained using molecular and fossil data that have been attributed to incompleteness of fossil records (Yamanoue et al. 2006). The dates we used were published recently in a comprehensive study on the timing of divergence events between model species (Benton and Donoghue 2006).

The results in figure 2 (presented in more detail in table 2) show that rearrangement rates are variable in vertebrates. The WGD occurred on branch b9, between teleost species and tetrapods, and the rate of rearrangements on this branch is higher than most tetrapod branches (b8, b6, and b4). Because the split between human and chicken is much older than the divergence between zebrafish and *Tetraodon* (fig. 1), the large majority of branch b9 represents evolution along the teleost fish lineage rather than the tetrapod lineage (approximatively 285–295 Myr versus 110–128 Myr, respectively). More precisely, post-WGD evolution represents roughly half (180–274 Myr) of the total branch length (395–423 Myr). We can thus argue that this branch significantly reflects the effect of the WGD. After the WGD, the rearrangement rate tends to decrease ($b9 \geq b7 > b1$, fig. 2). We do not observe this tendency in *Takifugu* (b2), but the estimate on branch b2 is not reliable because of the fragmentary nature of this genome assembly. We do not observe a decrease in rearrangement rate on the zebrafish lineage either (b3); on the contrary, the genome of zebrafish is clearly the most rearranged of the 6 we have considered.

The ratio of inter- to intra-chromosomal rearrangement rates is variable among taxa and particularly high in the zebrafish and mouse lineage, where the majority (60%) of rearrangements occur between chromosomes. This ratio could be artificially increased by saturation. This happens when several rearrangements occur between the same pair of genes. In that situation, only one event of rearrangement will be detected by our method. If one breakage out of these multiple rearrangements is interchromosomal, all intrachromosomal rearrangements that precede or succeed will be ignored. If our estimations suffered from this saturation problem, we should observe a positive correlation between the total amount of rearrangements (intra + inter) in a given branch and the fraction of interchromosomal rearrangements (inter/total) in this branch. There is no relationship between these 2 variables in our data set (supplementary figure 1, Supplementary Material online); for instance, the total rearrangement rate is particularly high on the branch b6, but the relative proportion of interchromosome breakages is small.

Discussion

Relationship between WGD and Rearrangement Rate

We have shown that the rearrangement rate is highly variable across the vertebrate species we studied. We focused on the rearrangement rates in teleosts and tested whether we could link this variation to the WGD. We observed a small increase in rearrangement rate coincidental with the WGD (branch b9), followed by a decrease in the subsequent branches (b7 and b1) in the *Tetraodon* lineage (fig. 2). This is consistent with the hypothesis of an increase of rearrangement rate following polyploidization, followed by a decrease of rearrangement when the process of diploidization reaches completion. However, the trend we observe does not allow us to firmly conclude that WGD is responsible for the increase of rearrangement rate. This is mainly due to the fact that the reconstruction of rearrangements in the *Takifugu* lineage is not reliable and that the zebrafish genome seems to exhibit an unusually high rearrangement

Table 2
Rearrangement Rates for Each Branch of the Vertebrate Tree (b1–b9, labeled as in fig. 1)

	Branch	b1 *Tetraodon*	b2 *Takifugu*	b3 Zebrafish	b4 Human	b5 Mouse	b6 Chicken	b7	b8	b9
	Total Gene Pairs Studied	219	233	178	621	669	434	459	287	184
Extent of proximity loss	PL (total)	0.009	0.064	0.404	0.026	0.094	0.150	0.120	0.122	0.385
	PL (intra)	0.009	0.000	0.152	0.021	0.036	0.090	0.085	0.075	0.321
	PL (inter)	0.000	0.064	0.253	0.005	0.058	0.060	0.035	0.047	0.064
Branch length	Fossils (Myr)	32.25–56.0	32.25–56.0	149.85–165.2	61.5–100.5	61.5–100.5	312.3–330.4	93.85–132.95	211.8–268.9	337–382
	Molecular (a.a. substitution)	0.097	0.067	0.247	0.058	0.093	0.192	0.223	0.156	0.358
Rearrangement rates, molecular calibration (breaks in synteny per a.a. substitution)	RM (total)	0.093	0.955	1.636	0.448	1.011	0.781	0.538	0.782	1.075
	RM (intra)	0.093	0.000	0.615	0.362	0.387	0.469	0.381	0.481	0.897
	RM (inter)	0.000	0.955	1.024	0.086	0.624	0.312	0.157	0.301	0.179
Rearrangement rates, fossil calibration (breaks in synteny per Myr)	RF (total)	0.016–0.028	0.114–0.200	0.245–0.269	0.026–0.042	0.093–0.152	0.045–0.048	0.090–0.128	0.045–0.058	0.101–0.115
	RF (intra)	0.016–0.028	0.000	0.092–0.101	0.021–0.034	0.036–0.058	0.027–0.029	0.064–0.090	0.028–0.035	0.084–0.095
	RF (inter)	0.000	0.114–0.200	0.153–0.169	0.005–0.008	0.057–0.094	0.018–0.019	0.026–0.037	0.017–0.022	0.017–0.019

NOTE.—PL, proximity loss; RM, rearrangement rates, molecular calibration; RF, rearrangement rates, fossil calibration; a.a., amino acid. For each branch, the number of pairs of genes studied (total) is indicated. The frequencies of those pairs that have lost proximity are depicted in the rows PL (total). Analogous percentages were computed after separating interchromosomal and intrachromosomal rearrangements. The branches b7, b8, and b9 were estimated by subtraction (see Methods): b7 has been computed using b7 + b1, b9 using b9 + b6, and b8 using b8 + b4. The branch lengths are depicted in terms of amino acid sequence divergence (molecular) and fossil estimates (fossils). Both methods have been used to normalize the levels of proximity losses PL. The rearrangement rates we obtain are presented in the rows RM (normalization using molecular divergence) and RF (normalization using fossil record).

rate (see below). The addition of more genomes might help to answer our question.

Comparison of Our Results with Other Studies

We can assess our confidence in our estimates of rearrangement rate in teleosts by comparing our results in tetrapods with those of previous studies. Multispecies comparisons have shown that the rate of interchromosomal rearrangement is low in the lineage leading to the earliest mammal and in the chicken lineage (Burt et al. 1999; Hillier et al. 2004; Jaillon et al. 2004). After the mammalian radiation, rates have varied radically in a lineage-specific way (Hillier et al. 2004; Bourque et al. 2005). Our observations are largely in agreement with all of these results. For instance, in accordance with Hillier et al. (2004), we observe that total rearrangement rate is 2–3 times higher in the mouse lineage than in the human lineage (compare b5 and b4, table 1 and fig. 2). We also confirmed that in mouse, interchromosomal rearrangements are more frequent than intrachromosomal rearrangements as they represent 60% of the total (Hillier et al. 2004). Comparisons between mouse populations (Britton-Davidian et al. 2000) and between rodent species (Zhao et al. 2004) have also shown that centromeric regions of mouse chromosomes are very active for interchromosomal rearrangements.

Species-Specific Determinants of Rearrangement Rate

The increase in rearrangement rate in teleosts may be due partly to the WGD, but there are alternative explanations. Changes in rearrangement rates have been attributed to species characteristics such as effective population size and generation time (Coghlan et al. 2005). Differential expansion of repetitive sequences may also have contributed to variation in rearrangement rates among species (Burt et al. 1999; Coghlan et al. 2005) because repeated elements are known to promote ectopic rearrangements (Deininger et al. 2003).

Jaillon et al. (2004) have suggested that the scarcity of transposable elements in the *Tetraodon* genome may be responsible for its low number of interchromosomal rearrangements compared with human genome. By contrast, the zebrafish genome is bigger (1,700 mb; Hubbard et al. 2005) than the *Tetraodon* genome (400 mb; Jaillon et al. 2004) and may therefore contain many more transposable elements. This led to the anticipation that rearrangement rate should be higher in zebrafish than in *Tetraodon* (Jaillon et al. 2004). However, this issue has been debated because many chromosomes exhibit a 1:1 correspondence between zebrafish and *Tetraodon*, which argues against an increase of rearrangement in the zebrafish lineage (Woods et al. 2005).

Here, we find that rearrangement rate is higher in zebrafish than in the *Tetraodon* lineage, which is in agreement with the expectations given the density of transposable elements. To investigate further the link between repeat density and rearrangement rates, we estimated the densities of repeated elements in each genome and compared them with the rearrangement rates at the terminal branches of the tree. In this data, there is no apparent relationship between the fraction of repeated elements and rearrangement rates

(supplementary table 1 and supplementary fig. 2, Supplementary Material online). For instance, the *Tetraodon* and human genomes have equally low rates of rearrangement, but they have, respectively, the smallest and largest fraction of repeated elements. The fact that we do not see a correlation between these measures of repeat density and rearrangement rate does not mean the 2 variables are not linked because our data set is small (only 6 points) and our measurements are done on different timescales. The fraction of repeats reflects both ancient "fossilized" repeated elements (that may be older than the external branch under study) and recent bursts of transposable element activity. For instance, the *Tetraodon* genome contains few transposable elements but they are mainly young and active (Jaillon et al. 2004), which contrasts with the chicken genome, which contains many more, but older, transposable elements (Hillier et al. 2004). We may expect that recent transposable elements, whose sequence is much more similar, impact more on rearrangements, even if they are fewer in number.

But in any case, it is difficult to determine the causality between repeated sequences and nonhomologous recombination. Both phenomena may merely reflect the intrinsic instability of some regions or some genomes (Coghlan et al. 2005). Add to this the fact that quiescent transposable elements have been shown to be activated in plants following a polyploidization event (Chen and Ni 2006), and it is clear that the relationship between WGD, repetitive element content, and rearrangement rate in vertebrate species is a complex one.

Supplementary Material

Supplementary table 1 and figures 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This study was supported by Irish Research Council for Science, Engineering and Technology and Science Foundation Ireland. The zebrafish genome is being sequenced and annotated by the Wellcome Trust Sanger Institute. The whole project is funded by the Wellcome Trust and all data can be accessed from the web page at http://www.sanger.ac.uk/Projects/D_rerio/. We thank Meg Woolfit and Brian Cusack for critical reading of the manuscript.

Literature Cited

Amores A, Force A, Yan YL, et al. (13 co-authors). 1998. Zebrafish hox clusters and vertebrate genome evolution. Science. 282:1711–1714.

Aparicio S, Chapman J, Stupka E, et al. (41 co-authors). 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science. 297:1301–1310.

Basden AM, Young GC, Coates MI, Ritchie A. 2000. The most primitive osteichthyan braincase? Nature. 403:185–188.

Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. Mol Biol Evol. 24:26–53.

Bernot A, Weissenbach J. 2004. Estimation of the extent of synteny between Tetraodon nigroviridis and Homo sapiens genomes. J Mol Evol. 59:556–569.

Bourque G, Pevzner PA, Tesler G. 2004. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. Genome Res. 14:507–516.

Bourque G, Zdobnov EM, Bork P, Pevzner PA, Tesler G. 2005. Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. Genome Res. 15:98–110.

Britton-Davidian J, Catalan J, da Graca Ramalhinho M, Ganem G, Auffray JC, Capela R, Biscoito M, Searle JB, da Luz Mathias M. 2000. Rapid chromosomal evolution in island mice. Nature. 403:158.

Burt DW, Bruley C, Dunn IC, et al. (13 co-authors). 1999. The dynamics of chromosome evolution in birds and mammals. Nature. 402:411–413.

Chen ZJ, Ni Z. 2006. Mechanisms of genomic and gene expression changes in plant polyploids. Bioessays. 28:240–252.

Christoffels A, Koh EG, Chia JM, 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.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet. 21:673–682.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6:836–846.

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. 2006. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol. 23:121–136.

Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. 2003. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev. 13:651–658.

Dufayard JF, Duret L, Penel S, Gouy M, Rechenmann F, Perriere G. 2005. Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. Bioinformatics. 21:2596–2603.

Fischer G, Rocha EP, Brunet F, Vergassola M, Dujon B. 2006. Highly variable rates of genome rearrangements between Hemiascomycetous yeast lineages. PLoS Genet. 2:e32.

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G. 2006. Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature. 439:749–752.

Hillier LW, Miller W, Birney E, et al. (174 co-authors). 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432:695–716.

Hoegg S, Brinkmann H, Taylor JS, Meyer A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol. 59:190–203.

Hubbard T, Andrews D, Caccamo M, et al. (52 co-authors). 2005. Ensembl 2005. Nucleic Acids Res. 33(Database issue):D447–D453.

Jaillon O, Aury JM, Brunet F, et al. (61 co-authors). 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431:946–957.

Jordan IK, Wolf YI, Koonin EV. 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Evol Biol. 4:22.

Kai W, Kikuchi K, Fujita M, Suetake H, Fujiwara A, Yoshiura Y, Ototake M, Venkatesh B, Miyaki K, Suzuki Y. 2005. A genetic linkage map for the tiger pufferfish, Takifugu rubripes. Genetics. 171:227–238.

Ma XF, Gustafson JP. 2005. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenet Genome Res. 109:236–249.

McLysaght A, Enright AJ, Skrabanek L, Wolfe KH. 2000. Estimation of synteny conservation and genome compaction between pufferfish (Fugu) and human. Yeast. 17:22–36.

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.

Murphy WJ, Larkin DM, Everts-van der Wind A, et al. (25 co-authors). 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science. 309:613–617.

Naruse K, Fukamachi S, Mitani H, et al. (20 co-authors). 2000. A detailed linkage map of medaka, Oryzias latipes: comparative genomics and genome evolution. Genetics. 154:1773–1784.

Naruse K, Tanaka M, Mita K, Shima A, Postlethwait J, Mitani H. 2004. A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revealed by comparative gene mapping. Genome Res. 14:820–828.

Nelson JS. 1994. Fishes of the world. New York: John Wiley and sons.

Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS. 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid Arabidopsis suecica genome. Proc Natl Acad Sci USA. 101:18240–18245.

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, 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.

Robinson-Rechavi M, Marchand O, Escriva H, Bardet PL, Zelus D, Hughes S, Laudet V. 2001. Euteleost fish genomes are characterized by expansion of gene families. Genome Res. 11:781–788.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18:502–504.

Sémon M, Wolfe KH. Forthcoming 2007. Reciprocal gene loss after whole genome duplication in teleost fishes. Trends Genet.

Seoighe C, Johnston CR, Shields DC. 2003. Significantly different patterns of amino acid replacement after gene duplication as compared to after speciation. Mol Biol Evol. 20:484–490.

Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc Natl Acad Sci USA. 92:7719–7723.

Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y. 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA. 101:1638–1643.

Wittbrodt J, Shima A, Schartl M. 2002. Medaka—a model organism from the far East. Nat Rev Genet. 3:53–64.

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

Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS. 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res. 15:1307–1314.

Yamanoue Y, Miya M, Inoue JG, Matsuura K, Nishida M. 2006. The mitochondrial genome of spotted green pufferfish Tetraodon nigroviridis (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. Genes Genet Syst. 81:29–39.

Zhao S, Shetty J, Hou L, Delcher A, Zhu B, Osoegawa K, de Jong P, Nierman WC, Strausberg RL, Fraser CM. 2004. Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. Genome Res. 14:1851–1860.

Richard Thomas, Associate Editor

Accepted January 3, 2007