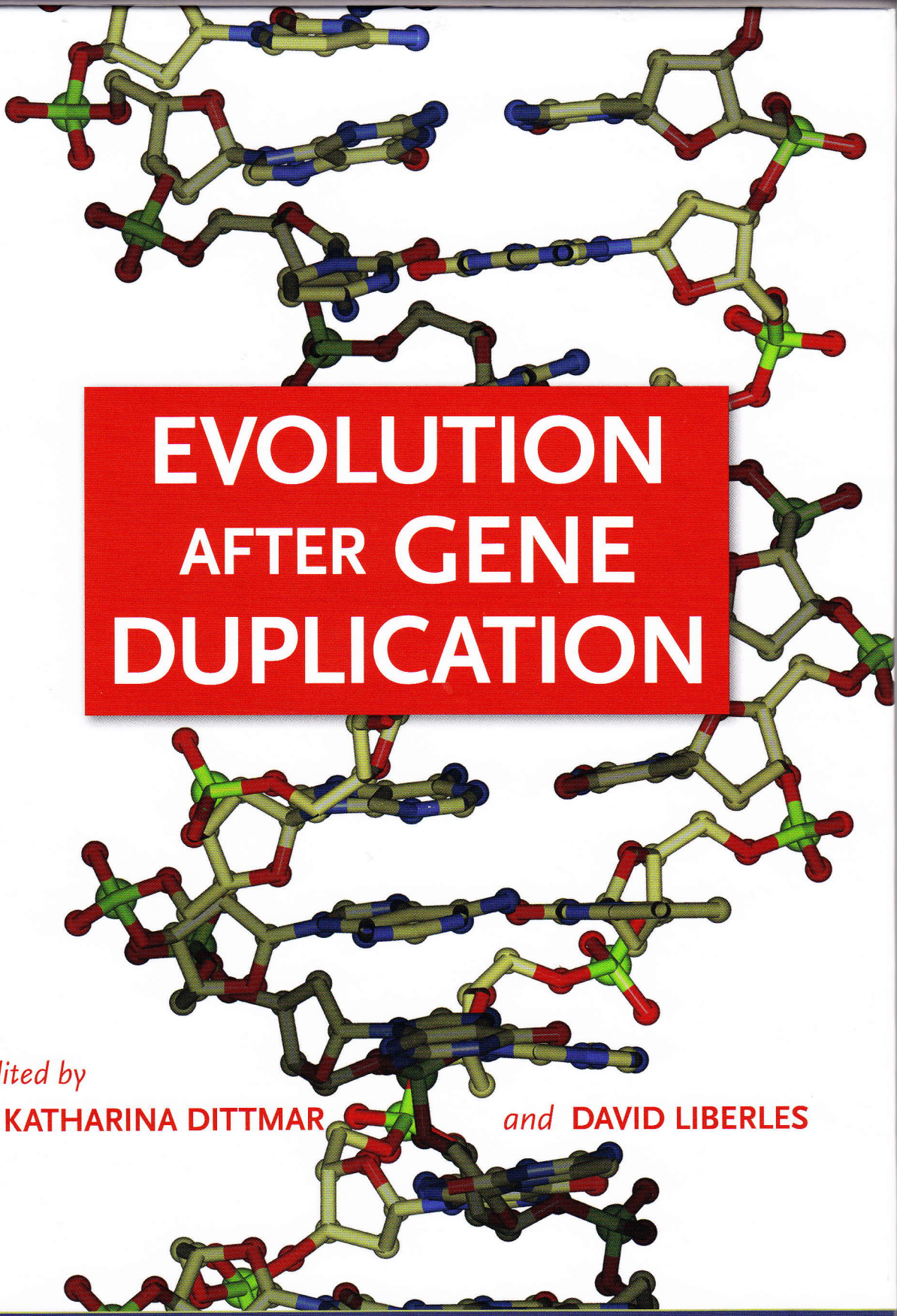

EVOLUTION AFTER GENE DUPLICATION

Edited by
KATHARINA DITTMAR *and* DAVID LIBERLES

WILEY-BLACKWELL

16 Whole Genome Duplications and the Radiation of Vertebrates

SHIGEHIRO KURAKU

Evolutionary Biology and Zoology, Department of Biology, University of Konstanz, Konstanz, Germany

AXEL MEYER

Evolutionary Biology and Zoology, Department of Biology, University of Konstanz, Konstanz, Germany; Center for Advanced Study, Berlin, Germany

1 INTRODUCTION

Almost 40 years ago, Susumo Ohno (1970) now famously said: "Duplication created where selection merely modified." Ohno made this statement, which is still considered by most researchers to be rather heretical, virtually in the absence of empirical data, at least by the standards and knowledge of the age of genomics that we are in (see Meyer and Van de Peer, 2003). However, during the last four decades, particularly the last 10 years, both the profusion and importance of all kinds of duplications in the genome have become more widely recognized. Duplications as construction principles of evolution have increased in acceptance even outside the field of genomics. It is seen increasingly by researchers in the field of "evo-devo" as an important mechanism by which organisms are permitted to experiment with the evolution of novel gene function, without having to do this slowly or having to lose the original function of a gene (copy) altogether.

Gene and genome duplications might increase the potential of evolutionary lines to produced diverse phenotypes of organisms (Ohno, 1970). One could classify genomic duplications based on their size or the mechanism that produced them. The first category would then be duplications of individual nucleotides followed by small numbers of base pairs such as dinucleotide motifs of microsatellites (e.g., CA repeats). Other potential categories could be the duplications of small sets of functional contiguous nucleotides such as enhancers and promoters. This category might be followed by exon duplications and entire gene duplication that might occur through tandem duplications or retropositions. Duplication of chromosomal regions that are larger than a single gene might include chromosome arms or even entire chromosomes. The largest, and presumably rarest form of duplication would be duplication of the whole genome. Whole genome duplications (WGDs) can apparently be recognized in genomes of

Evolution After Gene Duplication, Edited by Katharina Dittmar and David Liberles

"higher" organisms that have been sequenced completely. Usually, up to three genome (but interestingly, never more than that) WGDs have been inferred so far in the genomes of higher animals, fungi, and plants. WGD might be a major force that not only changes the genome content dramatically, but potentially creates a surplus of newly duplicated genes, which might also result in new genetic networks and, possibly, evolutionary phenotypic novelties [reviewed by Semon and Wolfe (2007a)].

Before whole genome sequences became available, whole genome duplications had been analyzed primarily for early vertebrate evolution, based on evidence obtained through molecular phylogenetic analyses of biologically important gene families (e.g., *Hox* genes, genes involved in the adaptive immune system). In such an analysis, one can commonly recognize an increase in numbers of genes among various gene families (Holland et al., 1994; Kasahara et al., 1996; Wittbrodt et al., 1998). More recently, by analyzing a larger set of gene sequences, whether they were complete or incomplete genome sequences, it was shown that similarly arranged sets of genes (synteny) are located on different chromosomes within a single genome (e.g., Pebusque et al., 1998) and that many chromosomal segments are similar within a genome (reviewed in Kasahara, 2007). Intragenome redundancy was later revealed for teleost fishes as well (reviewed in Meyer and Van de Peer, 2005).

In this chapter we briefly describe current knowledge regarding large-scale gene duplications that occurred at the basal lineages of vertebrates. We focus on the teleost-specific genome duplication (TSGD). This teleost-specific genome duplication is also called the third round (3R) genome duplication because the basal lineage of vertebrates experienced two previous rounds (1R and 2R) of whole genome duplications. Here we give specific attention to the 1R, 2R, and 3R WGD in the basal vertebrate lineages.

2 TELEOST-SPECIFIC GENOME DUPLICATION

2.1 Background

The idea of the teleost-specific whole genome duplication event in the actinopterygian lineage was proposed much later than that for 1R/2R WGDs. This is probably because studies at the molecular level for animals leading to human (mouse, chicken, and *Xenopus*) go back further than those on actinopterygians. Identification of higher numbers of genes for some gene families in teleost fishes than in tetrapods was the first DNA-based evidence on this issue (Wittbrodt et al., 1998). Some fish (e.g., salmonids and cyprinids) and amphibian lineages, including *Xenopus laevis*, have experienced additional independent whole genome duplication(s) more recently (see Gregory, 2005; Semon and Wolfe, 2008). The clumping of duplicated genes in specific genomic segments of modern fishes suggested a genome doubling that is not shared by tetrapods (e.g., Amores et al., 1998; reviewed by Meyer and Van de Peer, 2005). More recently, this fish-specific genome duplication has been shown with certainty through analyses involving large-scale sequence data (Taylor et al., 2001a, 2003), including those of draft genome sequences of teleost fish models, pufferfish, and medaka (Aparicio et al., 2002; Jaillon et al., 2004; Kasahara et al., 2007).

The question of how far back in the evolution of fishes the 3R duplication occurred remained open because the initial comparative genomic analyses were based on the genomes of rather modern "model" teleost fishes. Studies of more basal lineages of

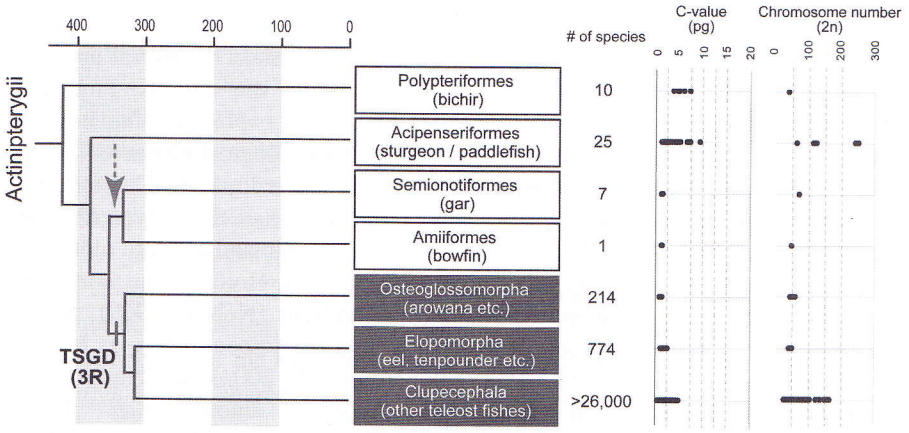

Figure 1 Phylogenetic and genomic properties of key lineages representing pre- and post-WGD conditions for the teleost-specific whole genome duplication. Phylogenetic relationships are based on Kikugawa et al. (2004) and Inoue et al. (2003). Divergence times are based on Azuma et al. (2008). Animal groups with a pre-WGD state are shown in white boxes, while those with a post-WGD state are shown in black boxes. Information regarding *C*-values and chromosome numbers was retrieved from the Animal Genome Size Database (www.genomesize.com; Gregory et al., 2007). Note that multiple entries for the same species in this genome size database are included in the graphs.

Actinopterygii revealed that their divergences from the stem lineage of fishes preceded this genome duplication event (Crow et al., 2006; Hoegg et al., 2004). Now it is thought that TSGD occurred in the lineage leading to all extant teleost fishes after the separation of more basal actinopterygian lineages: the Polypteriformes (bichir), Acipenseriformes (sturgeons and paddlefish), Amiiformes (bowfin), and Semionotiformes (gars) (Hoegg et al., 2004) (Figure 1). The TSGD was formerly called the fish-specific genome duplication (FSGD), but is now based on the more precise knowledge of the phylogenetic timing of the event. Currently, it is now more correctly called the teleost-specific genome duplication (TSGD; Kuraku and Meyer, 2009). The latter term is also more accurate because the word *fish* is applied to a paraphyletic assemblage that includes cyclostomes, chondrichthyes, lungfishes, and coelacanths, which did not experience this genomic event.

The phylogenetic timing of the TSGD has been revealed using two different approaches. First, based on a gene family tree approach, where timings of gene duplications are estimated directly, an analysis using duplicated sets of paralogs of the *Fugu* genome derived from the TSGD suggested that it occurred 320 ± 67 million years ago (Mya) (Vandepoele et al., 2004). Second, the relative timing of the TSGD was estimated based on the absolute timings of the split of nonteleost actinopterygians, and was inferred to have occurred approximately 380 to 300 Mya. The latter approach was taken, based on whole mitochondria DNA (mtDNA) genome sequences (Azuma et al., 2008), as well as a combination of both mtDNA and nuclear protein-coding genes. This study came up with a more recent estimate (316 to 226 Mya; Hurley et al., 2007).

2.2 Evolution Before the TSGD

As already mentioned, the TSGD occurred after the separation of several ancient lineages from the actinopterygian stem lineage that led to the teleosts. Phylogenetic relationships among these pre-TSGD lineages have been explored using both mitochondrial genes and nuclear genes as well as using morphological characters (Inoue et al., 2003; Kikugawa et al., 2004; reviewed by Meyer and Zardoya, 2003). Both of these two types of genes produced consistent results and identified the Polypteriformes as the most basal actinopterygian lineage and found that bowfin and gars are more closely related to each other than to any other groups. Previously, Holostei was considered a monophyletic group based on morphological observations (Nelson, 1969). However, the position of the amia-gar group in relation to that of Acipenseriformes is still not resolved consistently (Inoue et al., 2003; Kikugawa et al., 2004) (see Figure 1).

According to information from the Animal Genome Size Database (www.genomesize.com; Gregory et al., 2007), some members of Polypteriformes have much larger genome sizes than those of most other actinopterygian fishes, except for some species of the Acipenseriformes. Many sturgeons have hugely increased their genome sizes and their number of chromosomes as a result of repeated lineage-specific polyploidization events (Gregory, 2005; Peng et al., 2007). It is interesting to note that all of these pre-TSGD actinopterygian lineages show a low level of species diversity, and only about 40 extant species belong to the four most basal actinopterygian lineages (Figure 1).

The finding that the ancestors of some basal lineages of actinopterygian fishes did not experience the TSGD is interesting, as it allows us to study their genomes in an effort to examine a pre-TSGD (i.e., a 2R) genomic condition. Although the amount of data is still limited, the pre-TSGD condition of gene repertoires or selected genomic segments haboring them has been confirmed for some more cases (e.g., Chiu et al., 2004; Hoegg and Meyer, 2007). For example, for the *ParaHox* gene family, *Amia calva*, an extant member of one of the lineages that diverged immediately before the TSGD, seems to possess a similar gene organization (*Gsx, Xlox*, and *Cdx*) in the same transcription orientation to that of human and amphioxus (Mulley et al., 2006). The *ParaHox* gene repertoire of teleost fishes whose ancestor experienced the third vertebrate genome duplication is different. Their ancestral set of *ParaHox* gene clusters dispersed across different genomic regions as a result of successive subsequent gene losses after the TSGD (Siegel et al., 2007).

2.3 Evolution After the TSGD

It has been suggested that the TSGD somehow permitted the remarkable diversification of teleost species we see in extant teleost fishes. It will be interesting to explore the relevance of this genome doubling for morphological diversification by identifying and studying fish lineages that diverged from the stem lineage immediately after this event. The most interesting lineage in this regard will be the Osteoglossomorpha (e.g., arowana, arapaima), the earliest post-TSGD lineage (Hoegg et al., 2004; Azuma et al., 2008). The time that elapsed between the TSGD and the split of these fishes from the stem lineage is thought to be less than 10 million years, a short amount of time compared to the following history of the post-TSGD fish lineages (−300 Mya). This estimate was based on evolutionary rates of *Hox* genes (Crow et al., 2006).

The Osteoglossomorpha (<200 species), as well as the lineage of the Elopomorpha (e.g., eels, tenpounders), the sister group to the Euteleostei, are rather low in terms of species diversity (<750 species), compared to the diversity of all other extant teleost fishes (<26, 000 species) (Figure 1).

The Osteoglossomorpha and Elopomorpha seem to have a rather limited range of genome sizes and chromosome numbers compared to the pre-TSGD fish lineages (except those in Polypteriformes and Acipenseriformes) and other teleosts (except those whose genomes polyploidized lineage, specifically after the TSGD: e.g., goldfish, salmons) (Figure 1). As Ohno proposed, many of the teleost fishes have karyotypes with 48 chromosomes for diploid genomes (Ohno, 1970). Attempts to reconstruct an ancestral karyotype of the hypothetical teleost ancestor produced karyotypes similar to those of this widely shared format (Jaillon et al., 2007; Kasahara et al., 2007). The constancy of karyotype organization across diverse lineages of actinopterygian fishes (with some exceptions) may be the reason there was no pioneering hypothesis of TSGD based solely on classical cytogenetic studies.

In fact, the oldest teleost crown groups are, based on palaeontological data, believed to have diversified around at 150 Mya (Patterson, 1993; Benton, 1997; Figure 2). Therefore, based on palaeontological data, a time interval of about 150 to 170 million years followed after the TSGD before the phenotypic diversification of teleost fishes really began. In addition, other now extinct lineages of fish initially diversified but subsequently lost that diversity again. These observations have been used as counterevidence against the argument of a direct causal relationship between the TSGD and the huge diversification of fishes. As has often been observed before, palaeontological dates are always minimal age estimates of lineages, and molecular dates often lead to much older divergence estimates. Recent molecular datings based on *Hox* genes (Crow et al., 2006) and mtDNA (Azuma et al., 2008) have also suggested that the Osteoglossomorpha and Elopomorpha diverged immediately after the TSGD, questioning palaeontological data which suggested that a long time interval before the phenotypic diversification followed the TSGD.

To date, analyses of the genomic consequences of the TSGD have concentrated on large-scale sequence data of model systems. A molecular phylogenetic analysis, including 10 fish model species, has shown that the most ancient divergence between those species is the split between the third most early-branching lineage, Otocephala, including zebrafish, and a group containing other model fishes (pufferfishes, medaka, stickleback). These two lineages diverged at approximately 290 to 270 Mya (Steinke et al., 2006b). This estimate is roughly consistent with that obtained by whole mitochondrial genomes (310 to 270 Mya; Azuma et al., 2008). It should be noted that comparisons among only model fishes cannot cover the entire teleost fish diversity following the TSGD. The future inclusion of the Osteoglossomorpha and the Elopomorpha would provide important additional information, as some of the species for these groups are model species. The inclusion of these species would also aid identification of previously overlooked secondary lineage-specific modifications of genomic features (Taylor et al., 2003; Steinke et al., 2006a; Loh et al., 2008; Postlethwait, 2007; Semon and Wolfe, 2007b,c).

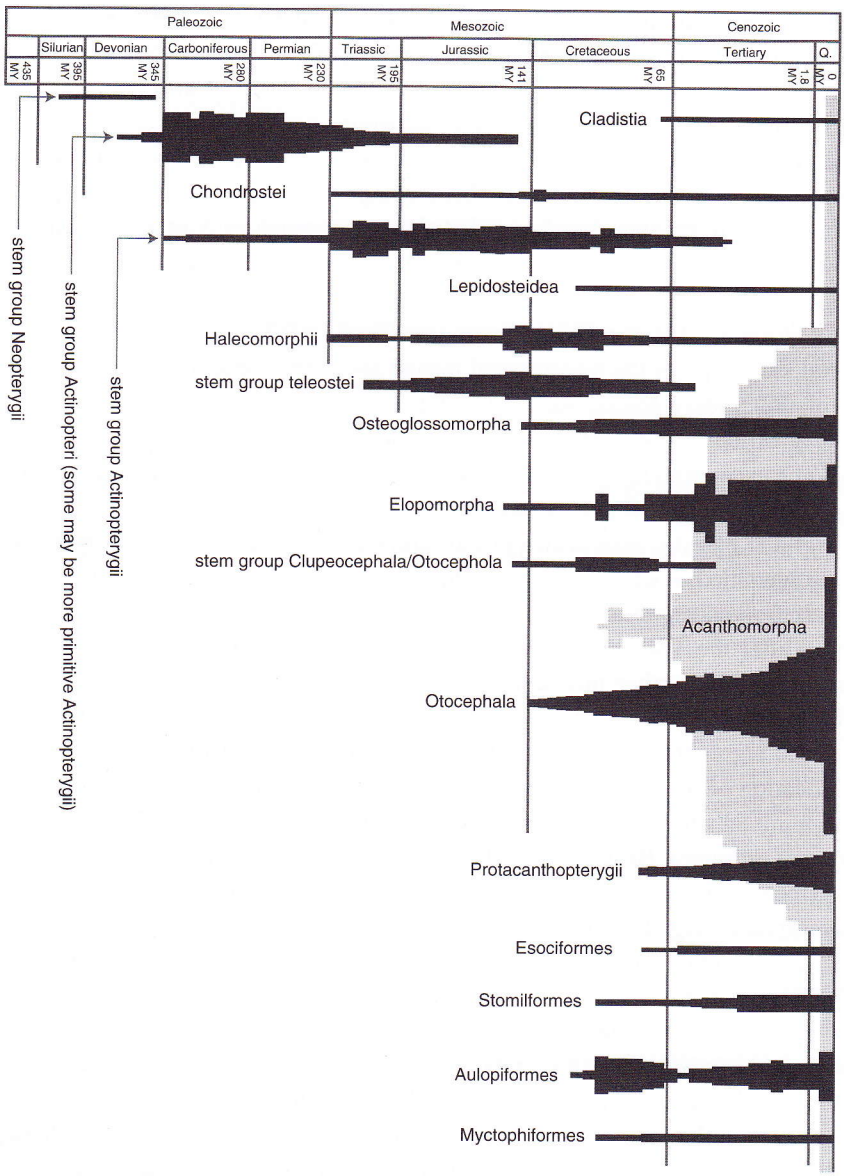

Figure 2 Patterns and timing of the diversification of fishes. Diversity (x-axis) and timing (y-axis) of the diversification of the major actinopterygian fish lineages are shown based on fossil evidence. Acanthomorpha, which has extremely large number of species, is shown in gray in the background. MY, million years; Q, quaternary. (Adapted from Patterson, 1993.)

2.4 Discussion

Reliable divergence time estimates of different teleost fish lineages are a key if one wants to test for a causal relationship between the TSGD and the phenotypic diversification and speciation of teleost fishes. If there was a considerable temporal interval following the TSGD before many species-rich lineages originated, this would tend to

weaken the argument of a causal link between any genomic and large-scale phenotypic events. Divergence times obtained by whole mitochondria genome–based phylogenetic analyses await further confirmation with nuclear protein-coding genes. Since the early-branching lineages after the TSGD (Osteoglossomorpha and Elopomorpha) show relatively low species diversity, there may already be a case against a link between the TSGD and a massive species diversification. Indeed, there are many biological strategies, which in concert with ecological opportunities generate diverse phenotypes without apparent genomic events.

3 1R/2R GENOME DUPLICATION

3.1 Background

Ohno (1970) was the first to propose that two-round whole genome duplications occurred during the evolution of chordates, based on genome sizes and karyotypes (Figure 3). This hypothesis was later supported through the identification of multiple copies of genes in model vertebrates (of course, jawed vertebrates such as human, mouse, and chicken) that were orthologous to a single ancestral invertebrate gene (Lundin, 1993; Holland et al., 1994; Sidow, 1996). More recently, large-scale sequencing studies of particular genomic regions provided evidence that similar arrays of genes are found in a single genome (mostly mammals), suggesting that this redundancy is derived from a duplication event that probably involved the entire genome [Holland et al., 1994; Kasahara et al., 1996; Katsanis et al., 1996; Lundin, 1993; Pebusque et al., 1998; Sidow, 1996); reviewed by Kasahara (2007)]. Based on genomewide sequence analyses, the hypothesis of two rounds of whole genome duplications (termed 1R and 2R) has been confirmed with certainty (McLysaght et al., 2002; Dehal and Boore, 2005; Hufton et al., 2008; Putnam et al., 2008).

At a similar time as the two WGD events, jawless vertebrates, the Cyclostomata, diverged from the vertebrate stem lineage leading to the Gnathostomata (jawed vertebrates). The Cyclostomata is the only extant lineage of the Agnatha, the jawless vertebrates, which now consists of the Myxiniformes (hagfishes) and the Petromyzoniformes (lampreys) (Kuraku et al., 2009b). It was long thought that lampreys are more closely related to jawed vertebrates than to hagfishes (Janvier, 1996; see also Janvier, 2007), but recently most molecular phylogenetic studies have supported the monophyly of cyclostomes [summarized by Kuraku and Kuratani (2006)]. The origin of the agnathans has been debated vigorously in relation to the phylogenetic timing of the 1R and 2R WGDs that occurred during the evolutionary history of vertebrates (Escriva et al., 2002; Putnam et al., 2008). A recent study that involved both large-scale phylogenetic analyses and intensive taxon sampling as well as cDNA sequencing suggested that the cyclostomes diverged after the two rounds of WGD; this hypothesis was termed the *pan-vertebrate tetraploidization* (PV4) *hypothesis* (Kuraku, 2008). This name was used to indicate that all vertebrates are characterized by post-2R genomes, and that subsequent genomic events such as the TSGD led to further modifications of the genomes of some vertebrate lineages, such as the teleosts.

3.2 Evolution Before the 2R

The Vertebrata are a morphologically derived monophyletic group within the Phylum Chodata, which also includes the Cephalochordata (amphioxus) and the

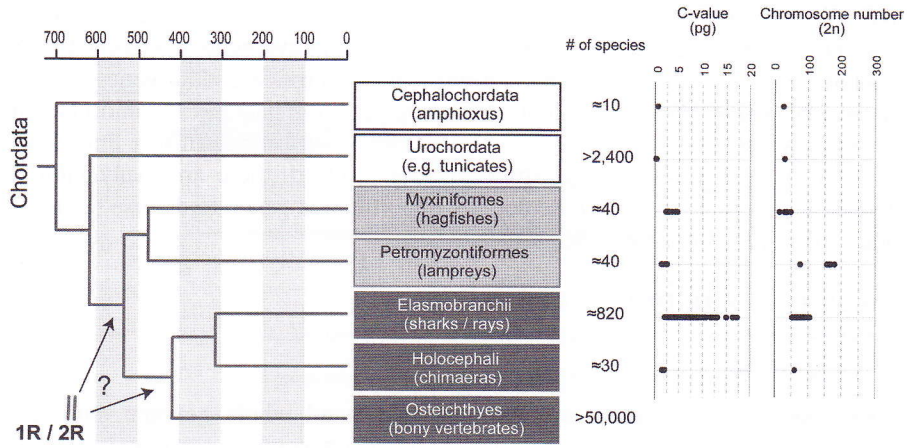

Figure 3 Phylogenetic and genomic properties of key lineages representing pre- and post-WGD conditions for the 1R/2R whole genome duplications. Phylogenetic relationships are based on Kikugawa et al. (2004), Kuraku and Kuratani (2006), and Putnam et al. (2008). Divergence times are based on related literature (Kuraku and Kuratani, 2006). Groups with a pre-WGD state are shown in white boxes, those with a post-WGD state in black boxes. It has not been shown clearly whether Myxiniformes and Petromyzoniformes diverged before or after the 1R/2R WGD. Information on *C*-values and chromosome numbers was retrieved from the Animal Genome Size Database. Note that multiple entries for the same species in this genome size database are included in the graphs.

Urochordata (tunicates) (Brusca and Brusca, 1990); Figure 3). Traditionally, cephalochordates were thought to be a sister group of vertebrates (e.g., Wada and Satoh, 1994), but "phylogenomic" analyses provided an updated phylogenetic hypothesis that identified urochordates as the sister group to vertebrates and placed cephalochordates as basal to chordates (Abascal et al., 2005; Bourlat et al., 2006; Delsuc et al., 2006; Dunn et al., 2008; Putnam et al., 2008) (Figure 3). In terms of numbers of species, the chordate subphyla do not show pronounced diversity (Figure 3).

As mentioned above, no consensus has been reached on the relative timing of the divergence between cyclostomes and gnathostomes relative to the timing of the 1R and 2R WGDs (Putnam et al., 2008; Kuraku et al., 2009a). In many cases, molecular phylogenetic trees containing cyclostome genes do not provide enough resolution or sufficiently strong statistical support on the relationships of multiple paralogs that are found in the genomes of jawed vertebrates (for more details, see Kuraku, 2008). A recently proposed view places cyclostomes after the 2R genome duplication (Kuraku et al., 2009a), while other studies suggest a pre-2R (Furlong et al., 2007) or a divergence time after the 1R but before the 2R duplication for cyclostomes (Escriva et al., 2002; Putnam et al., 2008).

3.3 Evolution After the 2R

Apart from cyclostomes, the earliest lineage that clearly diverged after the 2R WGD is the Chondrichthyes (cartilaginous fishes: namely, sharks, rays, skates, and chimaeras). Recently, the genome of the ghost shark *Callorhinchus milii* (elephant shark or elephantfish) has been sequenced with 1.4-fold coverage and has been deposited in the

NCBI Genome Survey Sequences (GSS) Database (Venkatesh et al., 2007; see also Venkatesh et al., 2005). Its genome contains convincing evidence that the lineage it belongs to experienced two rounds of genome duplication (Venkatesh et al., 2007). Other studies on this species provide further evidence that this lineage has highly conserved features in some genomic regions (Venkatesh et al., 2006; Yu et al., 2008). This conservation suggests that the common ancestor of chondrichthyans and osteichthyans apparently had a genome that was characterized by two consecutive rounds of genome duplications.

3.4 Discussion

The expected pattern of two-round duplication steps (i.e., from one to two, and then from two to four; [[A, B], [C, D]]) has been found difficult to reconstruct for many gene families surveyed (reviewed by Kasahara, 2007; Kuraku, 2008). However, it is now almost accepted that the most parsimonious explanation for the large-scale redundancy within a genome is nothing but whole genome duplications. Thus, most inconsistencies with this hypothesis (generally, these are not statistically significant) are thought to be caused by the lack of phylogenetic signals preserved over long evolutionary times (Panopoulou and Poustka, 2005; Kuraku, 2008).

The recently proposed hypothesis on "post-2R cyclostome" is based partly on the identification of a larger number of paralogs for *RAR* and *Dlx* genes in cyclostomes (Kuraku et al., 2009a). Previously, fewer were reported (Neidert et al., 2001; Escriva et al., 2006). In the case of the TSGD, different lineages have experienced different modes of evolution (e.g., gene repertoires, evolutionary rate, and gene function). One misleading argument against the TSGD originated from a lack of understanding of these post-WGD variations (Robinson-Rechavi et al., 2001a,b see also Taylor et al., 2001b). The knowledge of these variations should be taken into account in studies focusing on the tempo and mode of the 2R WGDs that are still disputable.

4 CONCLUDING REMARKS

Whether or not this is coincident, lineages that diverged immediately before or after WGDs are usually those containing no model species whose genome sequences are deeply analyzed. Species in such groups tend not to be readily available, to be difficult to handle in experiments, or to be inconvenient for maintenance in laboratories compared with model organisms, whose phylogenetic positions are usually distant from WGD events.

In this chapter, we provide an overview of karyotypes and genome sizes for the major lineages of osteichthyans and chordate lineages (Figures 1 and 3) and demonstrate that these values can vary drastically even within lineages that have experienced an identical history of whole genome duplications. Since many of these lineages are hundreds of years old, this observation is not all that surprising, as these genomes have undergone lineage-specific evolution. However, it does seem somewhat surprising that these lineages, which have experienced such varied evolutionary histories, do not seem to differ in more pronounced genomic features, such as genome size and/or chromosome numbers. Why this is the case is still only poorly understood.

We now have a robustly confirmed phylogenetic tree for major vertebrate lineages. By using existing and emerging genomic sequence resources, an evolutionary history of their redundant genomes, ideally including those of nonmodel key animals, should be characterized more deeply.

Acknowledgments

We thank Masaki Miya for his kind help in preparing Figure 1, and the German Science Foundation and the Wissenschaftskolleg Berlin for support of A.M.

REFERENCES

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, et al. 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711–1714.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A. et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297:1301–1310.

Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M. 2008. Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. BMC Evol Biol 8:215.

Benton MJ. 1997. *Vertebrate Palaeontology*. New York: Chapman & Hall.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, et al. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444:85–88.

Brusca RC, Brusca GJ. 1990. *Invertebrates*. Sunderland, MA: Sinauer Associates.

Chiu CH, Dewar K, Wagner GP, Takahashi K, Ruddle F, Ledje C, Bartsch P, Scemama JL, Stellwag E, Fried C, et al. 2004. Bichir HoxA cluster sequence reveals surprising trends in ray-finned fish genomic evolution. Genome Res 14:11–17.

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.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3:e314.

Delsuc F, Brinkmann H, Chourrout D and Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965–968.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749.

Escriva H, Manzon L, Youson J, Laudet V. 2002. Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. Mol Biol Evol 19:1440–1450.

Escriva H, Bertrand S, Germain P, Robinson-Rechavi M, Umbhauer M, Cartry J, Duffraisse M, Holland L, Gronemeyer H, Laudet V. 2006. Neofunctionalization in vertebrates: the example of retinoic acid receptors. PLoS Genet 2:e102.

Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland PW. 2007. A degenerate ParaHox gene cluster in a degenerate vertebrate. Mol Biol Evol 24:2681–2686.

Gregory TR. 2005. *The Evolution of the Genome*. Burlington, MA: Elsevier Academic.

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD. 2007. Eukaryotic genome size databases. Nucleic Acids Res 35:D332–D338.

Hoegg S, Meyer A. 2007. Phylogenomic analyses of KCNA gene clusters in vertebrates: Why do gene clusters stay intact? BMC Evol Biol 7:139.

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.

Holland PW, Garcia-Fernàndez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. Dev. Suppl., pp. 125–133.

Hufton AL, Groth D, Vingron M, Lehrach H, AJ, P Panopoulou G. 2008. Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res 18:1582.

Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang Z, Thomas MG, Coates MI. 2007. A new time-scale for ray-finned fish evolution. Proc Biol Sci 274:489–498.

Inoue JG, Miya M, Tsukamoto K, Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish." Mol Phylogenet Evol 26:110–120.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431:946–957.

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467.

Janvier P. 1996. The dawn of the vertebrates: characters versus common ascent in the rise of current vertebrate phylogenies. Paleontology 39:259–287.

Janvier P. 2007. Evolutionary biology: born-again hagfishes. Nature 446:622–623.

Kasahara M. 2007. The 2R hypothesis: an update. Curr Opin Immunol 19:547–552.

Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T. 1996. Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. Proc Natl Acad Sci USA 93:9096–9101.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719.

Katsanis N, Fitzgibbon J, Fisher EM. 1996. Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 35:101–108.

Kikugawa K, Katoh K, Kuraku S, Sakurai H, Ishida O, Iwabe N, Miyata T. 2004. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. BMC Biol 2:3.

Kuraku S. 2008. Insights into cyclostome phylogenomics: pre-2R or post-2R? Zool Sci 25:960–968.

Kuraku S Kuratani S. 2006. Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. Zool Sci 23:1053–1064.

Kuraku S, Meyer A. 2009. The evolution and maintenance of *Hox* gene clusters in vertebrates and the teleost-specific genome duplication. Int J Dev Biol. 53:765–773.

Kuraku S, Meyer A, Kuratani S. 2009a. Timing of whole genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before, or after? Mol Biol Evol 26:47–59.

Kuraku S, Ota S, Kuratani S. 2009b. Jawless fishes (Cyclostomata). In Hedges ESB, Kumar S (eds.), *Timetree of Life*. New York: Oxford University Press, pp. 315–319.

Loh YH, Brenner S, Venkatesh B. 2008. Investigation of loss and gain of introns in the compact genomes of pufferfishes (*Fugu* and *Tetraodon*). Mol Biol Evol 25:526–535.

Lundin LG. 1993. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16:1–19.

McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early chordate evolution. Nat Genet 31:200–204.

Meyer A, Van de Peer Y. 2003. "Natural selection merely modified while redundancy created"—Susumu Ohno's idea of the evolutionary importance of gene and genome duplications. J Struct Funct Genom 3:vii–ix.

Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27:937–945.

Meyer A, Zardoya R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. Annu Rev Ecol Evol Syst 34:311–338.

Mulley JF, Chiu CH, Holland PW. 2006. Breakup of a homeobox cluster after genome duplication in teleosts. Proc Natl Acad Sci USA 103:10369–10372.

Neidert AH, Virupannavar V, Hooker GW, Langeland JA. 2001. Lamprey *Dlx* genes and early vertebrate evolution. Proc Natl Acad Sci USA 98:1665–1670.

Nelson GJ. 1969. Gill arches and the phylogeny of fishes, with notes on the classification of vertebrates. Bull Am Mus Nat Hist 141:475–552.

Ohno S. 1970. *Evolution by Gene Duplication*. New York: Springer-Verlag.

Panopoulou G, Poustka AJ. 2005. Timing and mechanism of ancient vertebrate genome duplications: the adventure of a hypothesis. Trends Genet 21:559–567.

Patterson C. 1993. An overview of the early fossil record of acanthomorphs. Bull Mar Sci 52:29–59.

Pebusque MJ, Coulier F, Birnbaum D, Pontarotti P. 1998. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol Biol Evol 15:1145–1159.

Peng Z, Ludwig A, Wang D, Diogo R, Wei Q, He S. 2007. Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes. Mol Phylogenet Evol 42:854–862.

Postlethwait JH. 2007. The zebrafish genome in context: ohnologs gone missing. J Exp Zool B 308:563–577.

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071.

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

Robinson-Rechavi M, Marchand O, Escriva H, Laudet V. 2001b. An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr Biol 11:R458–R459.

Semon M, Wolfe KH. 2007a. Consequences of genome duplication. Curr Opin Genet Dev 17:505–512.

Semon M, Wolfe KH. 2007b. Rearrangement rate following the whole-genome duplication in teleosts. Mol Biol Evol 24:860–867.

Semon M, Wolfe KH. 2007c. Reciprocal gene loss between *Tetraodon* and zebrafish after whole genome duplication in their ancestor. Trends Genet 23:108–112.

Semon M, Wolfe KH. 2008. Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. Proc Natl Acad Sci USA 105:8333–8338.

Sidow A. 1996. Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6:715–722.

Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A. 2007. Comparative genomics of ParaHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. BMC Genom 8:312.

Steinke D, Salzburger W, Braasch I, Meyer A. 2006a. Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genom 7:20.

Steinke D, Salzburger W, Meyer A. 2006b. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. J Mol Evol 62:772–784.

Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001a. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B 356:1661–1679.

Taylor JS, Van de Peer Y, Meyer A. 2001b. Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis. Curr Biol 11:R1005–R1008.

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res 13:382–390.

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.

Venkatesh B, Tay A, Dandona N, Patil JG, Brenner S. 2005. A compact cartilaginous fish model genome. Curr Biol 15:R82–R83.

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, et al. 2006. Ancient noncoding elements conserved in the human genome. Science 314:1892.

Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, et al. 2007. Survey sequencing and comparative analysis of the elephant shark (Callorhinchus milii) Genome. PLoS Biol 5:e101.

Wada H, Satoh N. 1994. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. Proc Natl Acad Sci USA 91:1801–1804.

Wittbrodt J, Meyer A, Schartl M. 1998. More genes in fish? Bioessays 20:511–515.

Yu WP, Rajasegaran V, Yew K, Loh WL, Tay BH, Amemiya CT, Brenner S, Venkatesh B. 2008. Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genes: a comparative analysis of the protocadherin cluster. Proc Natl Acad Sci USA 105:3819–3824.