First publ. in: Trends in Genetics 17 (2001), 6, pp. 299-301

Genome duplication, divergent resolution and speciation

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

What are the evolutionary consequences of gene duplication? One answer is speciation, according to a model initially called Reciprocal Silencing and recently expanded and renamed Divergent Resolution. This model shows how the loss of different copies of a duplicated gene in allopatric populations (divergent resolution) can promote speciation by genetically isolating these populations should they become reunited. Genome duplication events produce thousands of duplicated genes. Therefore, lineages with a history of genome duplication might have been especially prone to speciation via divergent resolution.

Genome duplications have been credited, albeit controversially, with providing the genetic raw material necessary for major transitions in evolution, including the evolution of multicellularity, bilateral symmetry and the evolution of vertebrates[1–5]. The controversy stems largely from the results of phylogenetic studies that raise doubts about the timing of certain genome duplication events relative to major transitions and show that many gene trees do not have the topology predicted by genome duplication hypotheses[6–10]. Despite doubts over the origins of duplicated genes, the mechanistic link between gene duplication and evolutionary transitions has been less contentious. Although there are only a few examples of duplicated genes evolving completely new functions[11–13], Ohno's proposition that redundant genes produced during large-scale gene duplication events can evolve previously nonexistent functions important for the evolution of phenotypic 'complexity' has become well entrenched in the literature.

Recently, several studies have reported evidence for a complete genome duplication in the ancestor of actinopterygian fishes (ray-finned fishes)[14–22]. Two distantly related actinopterygian species, zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*), possess seven unlinked Hox gene clusters[14,18]. In Sarcopterygii (the sister group to Actinopterygia that includes lobe-finned fish, amphibians, reptiles and mammals) there appear to be only four unlinked Hox gene clusters. Thus, a genome duplication event in the ancestor of ray-finned fishes

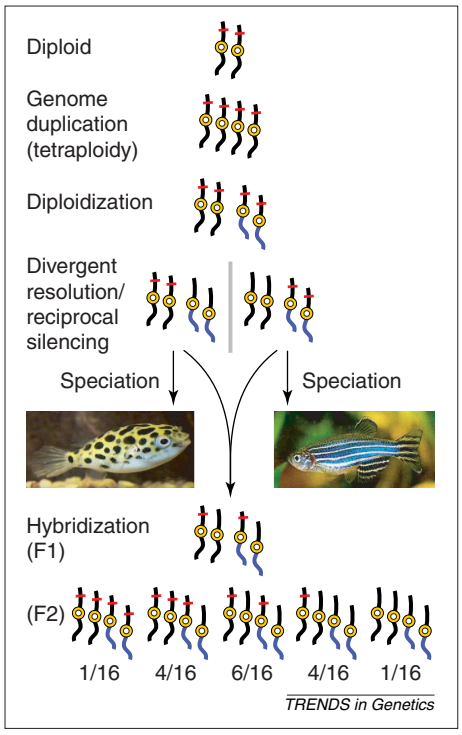

Fig. 1. Speciation through genome duplication and divergent resolution. The hypothesis that natural selection will favour speciation over hybridization in populations fixed for different copies of a duplicated locus. Red bars represent a locus duplicated (along with all other loci) during a tetraploidy event. In this hypothetical example, diploidization is driven by a reciprocal translocation depicted by a change in chromatid shape and colour. A pufferfish (left) and a zebrafish (right) are shown as examples of the descendants of the two populations. It will be possible to test the hypothesis that divergent resolution has occurred in ray-finned fish when the genome sequencing projects for these two species are completed. Hybridization of populations fixed for different copies of a duplicated locus produces heterozygous individuals possessing a functional allele and a pseudogene at each locus of the duplicated gene. Crosses between these F1 individuals produce F2 individuals with between zero and four alleles in the proportions shown.

could explain the observation that zebrafish and medaka have almost twice as many Hox clusters as, for example, humans. Additional evidence favouring the fish-specific genome duplication hypothesis comes from comparative mapping studies, which have identified a large number of mammalian genes with two zebrafish 'co-orthologs'[19–21]. Furthermore, most of these zebrafish co-orthologs are related in the way predicted by the fish-specific genome duplication hypothesis[22]. Amores *et al.*[14] proposed that this fish-specific genome duplication event provided gene copies that helped spur the teleost radiation, and Meyer and Schartl[17] suggested that there might be a cause–effect relationship between gene copy number and species diversity. This hypothesis, that new genes produced during a genome duplication facilitated the teleost radiation, mirrors Ohno's proposal that 'big leaps' in evolution required the creation of new gene loci with previously nonexistent functions[1]. However, instead of promoting speciation, the fate awaiting most duplicated genes appears to be loss or silencing due to degenerative mutations in coding and/or regulatory regions[23–26].

Lynch and Conery[26] and Lynch and Force[27] have recently presented a model that suggests that this loss or silencing of duplicated genes might be more important to the evolution of species diversity than the evolution of new functions in duplicated genes. Lynch and Conery[26] described how the loss of different copies of a duplicated gene in geographically separated populations could genetically isolate these populations, should they become reunited (Fig. 1). Their model is as follows. Consider individuals from populations that are fixed for different unlinked copies of a duplicated gene. If these individuals mate, their 'hybrid' progeny would be heterozygous, possessing a functional allele and a pseudogene at each locus of the duplicated gene. Crosses between the F1 individuals produce some F2 individuals with only pseudogenes at both loci in question (~6%). The remaining F2 progeny would receive between one and four functional alleles at these loci. The loss of different duplicates in different populations is called 'divergent resolution', and Lynch and Conery stated that with 'tens to hundreds' of gene duplicates present in most eukaryotic genomes, divergent resolution would result in the passive build-up of reproductive isolation.

Lynch and Force[27] expanded this model of speciation by divergent resolution by pointing out that for 'haploinsufficient' loci, the consequence of inheriting one functional allele can be as severe as inheriting only pseudogenes (haploinsufficiency occurs when the gene product from one functional allele is inadequate to support normal function). Lynch and Force also brought together 'subfunctionalization' and divergent resolution. Subfunctionalization occurs when duplicated genes divide the functions of their single ancestral gene[28]. Unless the subdivision of expression by duplicate genes follows an identical pattern in different populations, hybrids will lack one or more subfunctions even in the absence of gene loss[27]. Lastly, Lynch and Force make the connection between divergent resolution and genome duplication events. Divergent resolution of the thousands to tens of thousands of genes and regulatory regions produced by a genome duplication event could be a much more effective isolating mechanism than the one involving independently duplicated genes originally described by Lynch and Conery.

'Lynch and Conery[26] and Lynch and Force[27]…[suggest] that this loss or silencing of duplicated genes might be more important to the evolution of species diversity than the evolution of new functions in duplicated genes.'

Data from the family Salmonidae are consistent with this hypothesis that reproductive isolation via divergent resolution following genome duplication can lead to speciation. Salmonid fishes (e.g. salmon, trout, whitefish and grayling) evolved from a tetraploid ancestor and are in various stages of diploidization[29,30]. This is important because divergent resolution following genome duplication genetically isolates populations only when the genes involved occur on chromosomes that have re-established disomic inheritance. As predicted by the hypothesis that tetraploidy promotes speciation, the Salmonidae includes many more species (*c.* 70) than its diploid sister taxon, Osmeridae (smelt and their relatives), which contains only about ten species[31]. The hypothesis that divergent resolution has played a role in salmonid speciation is consistent with the results of isozyme studies. These demonstrate that in this family the number of duplicated loci retained varies among populations and among species[32,33].

Ferris *et al.*[34] provided different but equally intriguing data consistent with the hypothesis that divergent resolution leads to speciation in the fish family Catostomidae. Like salmonids, catostomids (suckers) evolved from a tetraploid ancestor[34]. Within this family the loss of expression of duplicated genes appears to be correlated with speciation – the clades within Catostomidae containing the most species are those clades with the most gene silencing[34]. Ferris *et al.* suggested that gene loss might have been enhanced by reductions in population size accompanying speciation events; that is, the speciation process might have caused duplicate gene silencing. However, an alternative interpretation of the data is that gene loss (i.e. divergent resolution) was at least partially responsible for speciation in these fishes.

Many examples of variation in locus copy number among tetraploid populations can also be found in the plant literature[35–37]. *Asplenium bradleyi* is a tetraploid fern formed by the hybridization of *A. platyneuron* and *A. montanum*, which are both diploid[36]. Comparisons of isozyme expression patterns in these species showed that one of the *A. bradleyi* populations had lost the isocitrate dehydrogenase (IDH) locus that came from *A. montanum*[36]. Bryan and Soltis[37] surveyed diploid, triploid and tetraploid populations of another fern species, *Polypodium virginianum*. Diploid populations all possessed a single leucine aminopeptidase (LAP) locus. Thirteen triploid or tetraploid populations possessed two LAP loci. However, in one tetraploid population, only the Lap-1 locus was observed and in a second allopatric tetraploid population, only Lap-2 was observed. From these isozyme studies in ferns, Werth and Windham[38] developed a model that is very similar to the divergent resolution model, in which 'reciprocal silencing' (the loss of different copies of duplicated genes in allopatric populations) would promote speciation.

Comparisons between tetraploid taxa and their diploid sister groups can determine if the association between genome duplication and speciation that we see in Paracanthopterygii (the group that includes salmonids and osmerids) is widespread. Some 50–70% of angiosperms appear to have experienced genome doubling[39], therefore it might be especially prudent to survey this group for a phylogenetic link between polyploidy and speciation. Genome sequencing projects will allow more extensive testing of the hypothesis that divergent resolution follows genome duplication. Ray-finned fish and flowering plants (Magnoliophyta) are both examples of rapid radiations, and polyploidy occurs at the base of both lineages[14–21,39–41]. To find examples of divergent resolution in these groups it will be necessary to identify paralogous regions within genomes and orthologous regions between genomes for distantly related species within each group. The pufferfish (*Fugu rubripes*), zebrafish (*Danio rerio*), *Arabidopsis* and rice (*Oryza sativa*) sequencing projects will facilitate this task.

Acknowledgements

J.S.T. is supported by a postdoctoral fellowship from NERC (Canada). The authors acknowledge support from the Deutsche Forschungsgemeinschaft and the University of Konstanz. A.M. thanks the Joint Genome Institute, Walnut Creek, USA, for their hospitality during his sabbatical stay there, while this manuscript was prepared. We thank Gerrit Begemann, Felix Breden and Henner Brinkmann; also Tancred Frickey, Joel Shore and Charles Werth for discussion and comments. This manuscript also benefited from the comments of three anonymous referees.

References

1 Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag
2 Lundin, L-G. (1999) Gene duplications in early metazoan evolution. *Cell Dev. Biol.* 10, 523–530
3 Sidow, A. (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* 6, 715–722
4 Holland, P.W.H. (1999) The effect of gene duplication on homology. In *Homology* (Bock, G.R. and Cardew, G., eds), pp. 226–242, Wiley
5 Patel, N.H. and Prince, V.E. (2000) Beyond the Hox complex. *Genome Biol.* 1, 1027.1–1027.4
6 Hughes, A.L. (1999) Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *J. Mol. Evol.* 48, 565–576
7 Wang, Y. and Gu, X. (2000) Evolution patterns of gene families generated in the early stages of vertebrates. *J. Mol. Evol.* 51, 88–96
8 Martin, A.P. (1999) Increasing genomic complexity by gene duplication and the origin of vertebrates. *Am. Nat.* 154, 111–128
9 Rivinsky, I. *et al.* (2001) Phylogenetic analysis of T-Box genes demonstrates the importance of *Amphioxus* for understanding evolution of the vertebrate genome. *Genetics* 156, 1249–1257
10 Friedman, R. and Hughes, A. (2001) Gene duplication and the structure of eukaryotic genomes. *Genome Res.* 11, 373–381
11 Cheng, C-H.C. and Chen, L. (1999) Evolution of an antifreeze glycoprotein. *Nature* 401, 443–444
12 Dulai, K.S. *et al.* (1999) The evolution of trichromatic colour vision by opsin gene duplication in New World and Old World primates. *Genome Res.* 9, 629–638
13 Manzanares, M. *et al.* (2000) Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature* 408, 854–857
14 Amores, A. *et al.* (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711–1714
15 Wittbrodt, J. *et al.* (1998) More genes in fish? *BioEssays* 20, 511–512
16 Meyer, A. and Málaga-Trillo, E. (1999) Vertebrate genomics: More fishy tales about *Hox* genes. *Curr. Biol.* 9, R210–R213
17 Meyer, A. and 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
18 Naruse, K. *et al.* (2000) A detailed linkage map of Medaka, *Oryzias latipes*: Comparative genomics and genome evolution. *Genetics* 154, 1773–1784
19 Gates, M.A. *et al.* (1999) A genetic linkage map for zebrafish: comparative analysis and localization of genes and expressed sequences. *Genome Res.* 9, 334–347
20 Postlethwait, J.H. *et al.* (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* 10, 1890–1902
21 Woods, I.G. *et al.* (2000) A comparative map of the zebrafish genome. *Genome Res.* 10, 1903–1914
22 Taylor, J.S. *et al.* Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos. Trans. R. Soc. London Ser. B* (in press).
23 Nei, M. (1969) Gene duplication and nucleotide substitution in evolution. *Nature* 221, 40–42
24 Allendorf, F.W. (1978) Protein polymorphism and the rate of loss of duplicate gene expression. *Nature* 272, 76–78
25 Li, W-H. (1980) Rate of gene silencing at duplicated loci: a theoretical study and interpretation of data from tetraploid fishes. *Genetics* 95, 237–258
26 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
27 Lynch, M. and Force, A.G. (2000) The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* 156, 590–605
28 Force, A. *et al.* (1999) The preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
29 Hartley, S.E. (1987) The chromosomes of salmonid fishes. *Biol. Rev.* 62, 197–214
30 Bailey, G.S. *et al.* (1978) Gene duplication in tetraploid fish: Model for gene silencing at unlinked duplicated loci. *Proc. Natl. Acad. Sci. U. S. A.* 75, 5575–5579
31 Nelson, J.S. (1984) *Fishes of the World* (2nd edn), Wiley
32 Allendorf, F.W. *et al.* (1984) Silencing of duplicate genes: A null allele polymorphism of lactate dehydrogenase in brown trout (*Salmo trutta*). *Mol. Biol. Evol.* 1, 238–248
33 May, B. *et al.* (1975) Biochemical genetic variation in pink and chum salmon. *J. Hered.* 66, 227–232
34 Ferris, S.D. *et al.* (1979) The roles of speciation and divergence time in the loss of duplicate gene expression. *Theor. Popul. Biol.* 15, 114–139
35 Wilson, H.D. *et al.* (1983) Loss of duplicate gene expression in tetraploid *Chenopodium*. *Biochem. Syst. Ecol.* 11, 7–13
36 Werth, C.R. *et al.* (1985) Recurring origins of allopolyploid species in *Asplenium*. *Science* 228, 731–733
37 Bryan, F.A. and Soltis, D.E. (1987) Electrophoretic evidence for allopolyploidy in the fern *Polypodium virginianum*. *Syst. Bot.* 12, 553–561
38 Werth, C.R. and Windham, M.D. (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* 137, 515–526
39 Wendel, J. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* 42, 225–249
40 Ku, H-M. *et al.* (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9121–9126
41 Poethig, R.S. (2001) Life with 25,000 genes. *Genome Res.* 11, 313–316

J.S. Taylor*
Y. Van de Peer
A. Meyer
Dept of Biology, University of Konstanz, 78457 Konstanz, Germany.
*e-mail: john.taylor@uni-konstanz.de