



Are We Polyploids? A Brief History of One Hypothesis

Wojciech Makalowski

Genome Res. 2001 11: 667-670

Access the most recent version at doi:[10.1101/gr.188801](https://doi.org/10.1101/gr.188801)

References This article cites 19 articles, 10 of which can be accessed free at:
<http://genome.cshlp.org/content/11/5/667.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the words "LEARN MORE". On the right, there is a photograph of a woman wearing a red superhero mask and cape, and the Cellecta logo, which consists of a green molecular structure and the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Are We Polyploids? A Brief History of One Hypothesis

Wojciech Makalowski

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Formulation of the 2R Hypothesis

The late Susumu Ohno was one of the most prolific and influential modern biologists. "He has thought at least half of the thoughts that form the basis for the work being carried out all over the world in respect to genetic analysis." In 1970, Ohno published his seminal *Evolution by Gene Duplication* (Ohno 1970). The book brought (in addition to the now widely accepted title hypothesis) another intriguing thought, namely the hypothesis of two or more full genome duplications in early stages of vertebrate evolution. Interestingly, Ohno regarded tetraploidization as a more important evolutionary mechanism than tandem gene duplication: "... tandem duplication is meaningless unless supplemented periodically by simultaneous duplication of all gene loci by tetraploidy" (Ohno 1970). He based this suggestion solely on genome size differences in different chordates and evidence of recent tetraploidization in some fish lineages. In fact, he was not quite sure how often and when vertebrate genome duplications occurred. "I venture a guess that paedomorphosis from tunicate-like creatures to amphioxus-like creatures [...] was accompanied by a two to threefold increase in the genome size [...]. Whether this increase was accomplished exclusively by tandem duplication or by combination of tandem duplication and tetraploidy cannot be resolved at the moment" (Ohno 1970). Later, in the same chapter, he wrote: "It is our contention that the ancestors of reptiles, birds, and mammals have experienced at least one tetraploid evolution either at the stage of fish or at the stage of amphibians".

Three years later, the same author proposed "at least one round of tetraploid evolution" at the stage of fish or amphibian (Ohno 1973). This time the supportive evidence was the fact that, in mammals, many paralogous genes are unlinked. Such a situation would be expected in the case of whole chromosome duplication, as opposed to tandem duplications, which should result in closely linked paralogous genes. Of course, with time and chromosomal rearrangements one can expect deviations from this ideal picture, and in today's mammalian genomes some paralogs remain linked when others are unlinked.

In the last decade, the "two rounds of genome duplication in vertebrate evolution hypothesis" (2R hypothesis) resurfaced and gained in popularity, especially among developmental biologists. A series of articles have proposed different numbers and timing for the whole genome duplication (Fig. 1). The most popular version of the 2R hypothesis suggests one genome duplication at the root of the vertebrate lineage, followed by another around Agnatha and Gnathostomata. Some extreme proposals include the first genome duplication at the root of Chordata lineage and the last one at the amphibian lineage (Lundin 1993).

There is only one problem: The current mammalian gene number estimations based on both ESTs and draft sequence of the human genome reveal that our genome hosts much fewer protein coding genes than anticipated (Ewing and Green 2000; Roest Crolius et al. 2000; Lander et al. 2001; Venter et al. 2001). The 35,000 genes in the human genome means that, on average, for every invertebrate protein gene there are only two mammalian orthologs.

Of course one can argue that many redundant genes have been lost during vertebrate evolution. If this were true, we should be able to detect gene loss by simply plotting phylogenetic trees for different vertebrate gene families. Two rounds of the whole genome duplication would result in four vertebrate gene clusters of (AB) (CD) tree topology. This topology should be easily detectable even in incomplete data or after a gene loss in specific lineages. Figure 2 presents a hypothetical phylogenetic tree for a vertebrate gene family, assuming some gene loss or incomplete data and additional gene duplication in specific lineages. This simple test was recently applied extensively by A.L. Hughes (1999). Hughes constructed phylogenies for nine protein families involved in development. Only one of the trees, for dpp protein, exhibited the (AB) (CD) topology that would support the 2R hypothesis.

This issue of *Genome Research* brings more tests of the 2R hypothesis. First of all, Hughes and colleagues scrutinize the hypothesis with a series of the most rigorous tests to date. As his (Hughes 1999) previous approach has been subjected to some critique (or rather, doubt) regarding the final inference (Gibson and Spring 2000), Hughes has come out with even more sophisticated tests. This time Hughes and colleagues analyzed a set of

Arguments for the 2R Hypothesis

Gene numbers in different animal lineages were very often used in support of the 2R hypothesis. It seems that many invertebrates have gene number close to 15,000 (Table 1). With the initially accepted human gene number ~80,000, the fourfold ratio between mammalian and invertebrate gene numbers was in good agreement with the 2R hypothesis.

E-MAIL makalowski@ncbi.nlm.nih.gov;
FAX (301) 480-2918.

Article and publication are at www.genome.org/cgi/doi/10.1101/gr.188801.

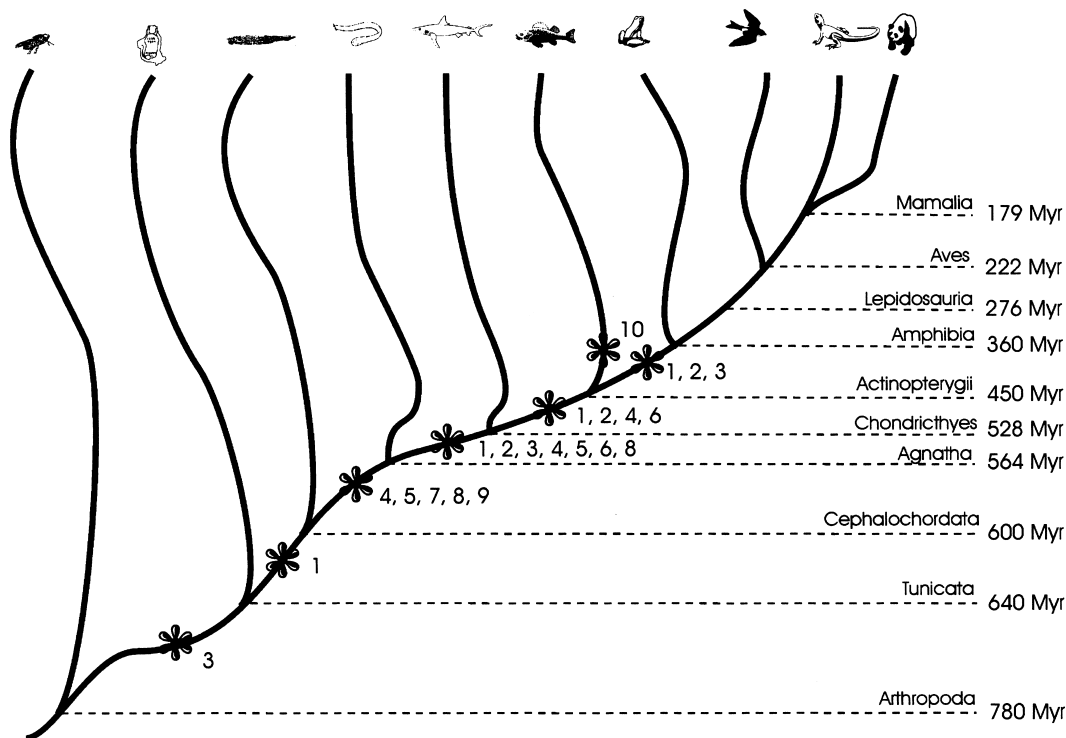


Figure 1 A schematic representation phylogeny of main vertebrate groups. Timescales are based on molecular clock (Kumar and Hedges 1998). Some of the proposed vertebrate whole genome duplications are marked by asterisks. The connected asterisks denote the range of the proposed genome duplication time. The references are as follows: 1, Ohno 1970; 2, Ohno 1973; 3, Lundin 1993; 4, Holland et al. 1994; 5, Sidow 1996; 6, Kasahara et al. 1996; 7, Spring 1997; 8, Ohno 1998; 9, Meyer and Schartl 1999; 10, Amores et al. 1998 and Mayer and Schartl 1999.

42 gene families that reside on human chromosome 2, 7, 12, and 17, the very same chromosomes that host the *Hox* clusters. The fact that two or more paralogs of several mammalian gene families are linked with *Hox* genes has been seen as a strong argument supporting the 2R hypothesis. It seems logical that those paralogs were created along with the *Hox* cluster during polyploidization; but very often what is obvious is not necessary true. To support the 2R hypothesis, clustered paralogous genes not only have to exhibit (AB) (CD) tree topology (Fig. 2), but additionally they ought to

diverge in concordance with the *Hox* family and by the proposed 2R hypothesis genome duplication time. Hughes et al. applied four different methods: construction of phylogenetic trees, estimation of the divergence time of paralogous genes, consistency of the phylogenetic trees, and, finally, the parsimony test (e.g., estimation of the minimum number of genetic events, like gene duplications, deletions, or translocations, required to explain the genes' current distribution on human chromosomes under two competing hypotheses [that is, independent duplication(s) of each

gene family and whole genome duplication(s)]. The results of their analysis did not support the 2R hypothesis. Thirty-five gene families provided evidence regarding the hypothesis, and in 29 cases the results were inconsistent with the hypothesis.

Nonetheless, the most intriguing results come from the parsimony analysis. Only 20 out of 42 gene families had an estimated genetic events number that differed for the two hypotheses. Although tandem duplication was the more parsimonious explanation in 14 cases (2 : 1 ratio against the 2R hypoth-

Table 1. Gene Number in Selected Animal Genomes

Species	Group	Gene number	Reference
<i>Drosophila melanogaster</i>	Arthropoda	13,500	Adams et al. 2000
<i>Caenorhabditis elegans</i>	Nematode	19,000	The <i>C. elegans</i> Sequencing Consortium 1998
<i>Ciona intestinalis</i>	Urochordata	15,500	Simmen et al. 1998
<i>Fugu rubiapiens</i>	Actinopterygii	"similar to mammals"	Elgar 1996
<i>Homo sapiens</i>	Mammalia	80,000	Antequera and Bird 1994
		35,000	Ewing and Green 2000; Roest Crolius et al. 2000; Lander et al. 2001; Venter et al. 2001

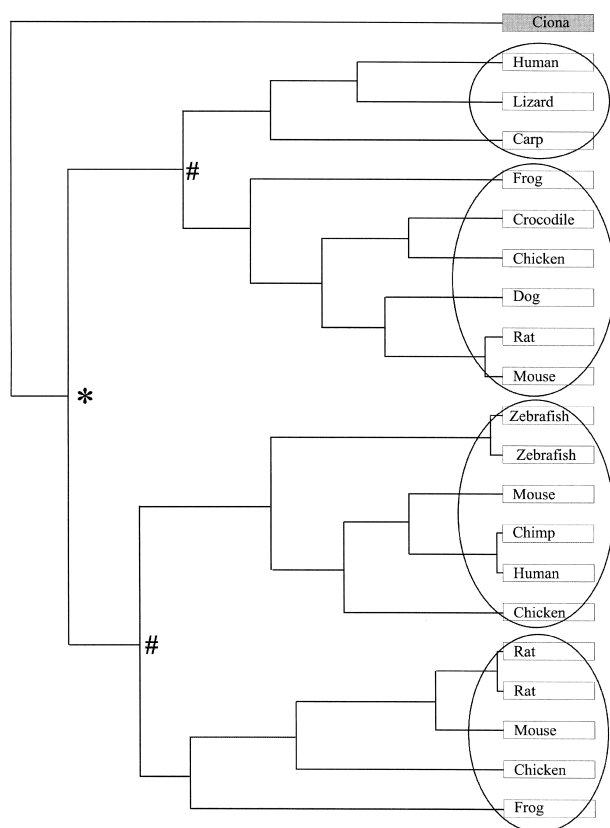


Figure 2 A hypothetical phylogenetic tree of vertebrate gene family according to the 2R hypothesis. The Ascidian gene represents an outgroup, and four clusters of a gene family are encircled. An asterisk (*) marks the first genome duplication and a hash sign (#) marks points of second genome duplication. Different branch lengths suggest different evolutionary rates after ancestral gene duplication.

esis), the fact that, for more than half of the gene families, both hypotheses give equally parsimonious explanations leaves some hope for 2R hypothesis supporters. We shall definitely see more discussion on the subject when more complete vertebrate genomes are available.

Two other articles in this issue refer to vertebrate genome duplications. Ledent and Vervoort analyzed the basic Helix-loop-Helix (bHLH) protein family. As expected, most of the bHLH families comprise more members in vertebrates than in invertebrates. Ledent and Vervoort constructed phylogenetic trees for more than 40 gene families. Nine of them had four or more vertebrate members but only five resulted in reliable trees, none of which had (AB)(CD) topology. Once again, careful phylogenetic analysis did not support the 2R hypothesis.

Ledent and Vervoort also noticed

the existence of extra closely-related genes in the tetraploid *Xenopus* and in Actinopterygia phyla (ray-finned fishes such as the zebrafish). Although no phylogenetic analysis was presented, they state that this observation supports another genome duplication hypothesis, namely the hypothesis of whole genome duplication in the Actinopterygia lineage (euteleost fish) just after the Actinopterygia-Sarcopterygia divergence. Interestingly, this hypothesis is a subject of yet another study reported in this issue (Robinson-Rechavi et al. 2001). Robinson-Rechavi and colleagues compared 251 gene families with at least one copy in mouse and zebrafish. Although the majority (80%) had just one copy in each species, for the rest duplicated genes were more frequent in fish (a 2 : 1 ratio). This is an impressive finding, taking into account the fact that the mouse genome is 25 times more extensively

sampled than the zebrafish. Furthermore, they found that this gene number expansion is not specific to the zebrafish nor to its order. Although all major Actinopterygian groups have an elevated number of gene paralogs, phylogenetic analysis showed that these gene duplications occurred after the divergence of different Actinopterygian lineages. Consequently, the hypothesis of whole genome duplication in the evolution of Actinopterygia must seemingly be rejected. This has some implication for the 2R hypothesis, because the idea of a whole genome duplication in fish is invoked as an argument for possible genome duplications in early stages of vertebrate evolution; in other words, if whole genome duplications occurred recently in the fish lineage, it could happen in early vertebrate evolution as well (Ohno 1970).

Conclusions

Reconstruction of the history of living organisms is a very difficult task. We are not able to reconstruct it with certainty because of its complexity. Many evolutionary events become obscure with the time; hence, inferences about early evolution of vertebrate genome remain in a scientific “gray zone”. It is probable that we never will be able to say *how this happened* but only *how it could have happened*. Among proposed scenarios we ought to choose the most likely, applying some rules of demarcation. One such rule, the principal of parsimony, was formulated six centuries ago: “Pluralitas non est ponenda sine necessitate” (William of Ockham 1957).

Formulated three decades ago, the hypothesis of whole genome duplications in the early stages of vertebrate evolution has as many adherents as opponents. It seems that the current data do not support the 2R hypothesis, and the existence of more paralogs in vertebrates than in invertebrates can be explained by waves of tandem duplications of single genes or larger chromosomal fragments. Applying “Ockham’s Razor”, we must refute the 2R hypothesis, at least for now. However, the history of science teaches that falsification

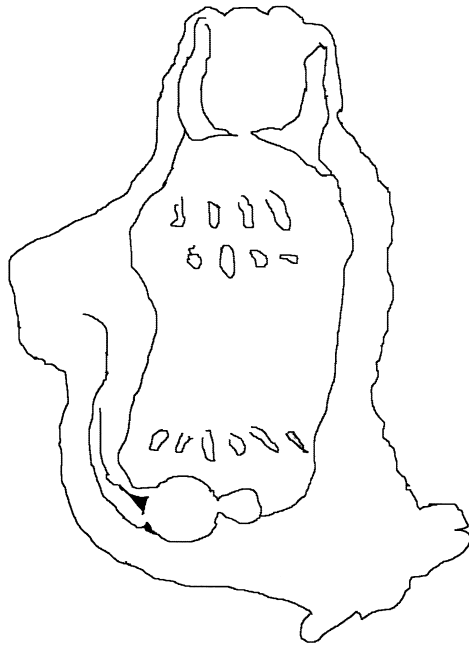


Figure 3 The invertebrate chordate *Ciona intestinalis* is a small marine creature. Its 165-Mb genome codes for ~15,500 protein genes.

of a hypothesis does not always lead to its refutation, especially if the hypothesis was put forward by an influential personality. Supporters very often bring ad hoc adjustments to a hypothesis and the 2R hypothesis is no exception. For example, Gibson and Spring in their recent review wrote: "Although gene trees have been held as evidence against the octaploidy because they often do not show the expected topology, they cannot be used to assess the evolutionary history of an octavalent octaploid. The sequence trees can be taken as evidence against models with a long period of diploidization between genome duplications or models of allooctaploidy (interspecific hybridization), unless the parental species were so close that octavalents resulted. (. . .) We conclude, therefore,

that there is still no strong evidence against the ancestral octaploidy, but that the second round of genome duplication must have followed rapidly upon the first" (Gibson and Spring 2000).

"If it is to form a part of science, an hypothesis must be falsifiable" (Chalmers 1999). Is the 2R hypothesis falsifiable? Is it scientific? The answer is not very simple; most likely the 2R hypothesis is not falsifiable directly. It is not a question of available data; I don't think that mammalian comparative genomics by itself will solve the problem. Long syntenic regions could be created by local chromosome duplications just before mammalian radiation, and the human genome analysis showed that local chromosome duplications have occurred even recently (Venter et al. 2001). The hope is in sequencing one of the invertebrate chordates. We, the evolutionary community, need full genomic information at the root of the vertebrate phyla. The small tunicata, *Ciona intestinalis* (Fig. 3), seems to be an ideal candidate for this effort. Its 15,500 genes are coded by a genome similar in size to one of *Drosophila* (165 Mb). Moreover, there are >10,000 ESTs already deposited in GenBank. Complete information on tunicata and other chordate genomes would enable us to address many evolutionary questions, including the frequency and extent of genomic duplications. In the meantime, we are restricted to the level of anecdote and we shouldn't make, from a few case studies, far-ranging inferences about entire genomes and their evolutionary history.

ACKNOWLEDGMENTS

I thank Jakub Makałowski and Maciej

Makałowski for preparation of the figures.

REFERENCES

- Adams, M.D., Celniker, S.E., et al. 2000. *Science* **287**: 2185–2195.
- Amores, A., Force, A., et al. 1998. *Science* **282**: 1711–1714.
- Antequera, F. and Bird, A. 1994. *Nat. Genet.* **8**: 114.
- Chalmers, A. F. 1999. *What is this thing called science?* Hackett, Indianapolis.
- The *C. elegans* Sequencing Consortium 1998. *Science* **282**: 2012–2018.
- Elgar, G. 1996. *Hum. Mol. Genet.* **5**(Spec. No): 1437–1442.
- Ewing, B. and Green, P. 2000. *Nat. Genet.* **25**: 232–234.
- Gibson, T.J. and Spring, J. 2000. *Biochem. Soc. Trans.* **28**: 259–264.
- Holland, P.W., Garcia-Fernandez, J., et al. 1994. *Development Suppl.*: 125–133.
- Hughes, A.L. 1999. *J. Mol. Evol.* **48**: 565–576.
- Kasahara, M., Hayashi, M. et al. 1996. *Proc. Natl. Acad. Sci.* **93**: 9096–9101.
- Kuhn, T.S. 1970. *The structure of scientific revolutions.* University of Chicago Press.
- Kumar, S. and Hedges, S.B. 1998. *Nature* **392**: 917–920.
- Lander, E.S., Linton, L.M., et al. 2001. *Nature* **409**: 860–921.
- Ledent, V. and Vervoort, M. 2001. *Genome Res.* **11**: 754–770.
- Lundin, L.G. 1993. *Genomics* **16**: 1–19.
- Meyer, A. and Scharl, M. 1999. *Curr. Opin. Cell Biol.* **11**: 699–704.
- . 1973. *Nature* **244**: 259–262.
- . 1998. *The notion of the Cambrian pananimalia genome and a genomic difference that separated vertebrates from invertebrates.* *Molecular Evolution: Evidence for Monophyly of Metazoa (Progress in Molecular and Subcellular Biology, vol. 21)*, Springer-Verlag, NY.
- Ohno, S. 1970. *Evolution by gene duplication.* Springer-Verlag, NY.
- Robinson-Rechavi, M., Marchand, O., et al. 2001. *Genome Res.* **11**: 781–788.
- Roest Crolius, H., Jaillon, O., et al. 2000. *Nat. Genet.* **25**: 235–238.
- Sidow, A. 1996. *Curr. Opin. Genet. Dev.* **6**: 715–722.
- Simmen, M.W., Leitgeb, S., et al. 1998. *Proc. Natl. Acad. Sci.* **95**: 4437–4440.
- Spring, J. 1997. *FEBS Lett.* **400**: 2–8.
- Venter, J.C., Adams, M.D. et al. 2001. *Science* **291**: 1304–1351.
- William of Ockham. 1957. *Philosophical writings; a selection.* Nelson, Edinburgh/New York.