

GENOME RESEARCH

Pattern and Timing of Gene Duplication in Animal Genomes

Robert Friedman and Austin L. Hughes

Genome Res. 2001 11: 1842-1847; originally published online Oct 15, 2001;
Access the most recent version at doi:[10.1101/gr.200601](https://doi.org/10.1101/gr.200601)

References This article cites 24 articles, 9 of which can be accessed free at:
<http://www.genome.org/cgi/content/full/11/11/1842#References>

Article cited in:
<http://www.genome.org/cgi/content/full/11/11/1842#otherarticles>

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](#)

Notes

To subscribe to *Genome Research* go to:
<http://www.genome.org/subscriptions/>



Pattern and Timing of Gene Duplication in Animal Genomes

Robert Friedman and Austin L. Hughes¹

Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

Duplication of genes, giving rise to multigene families, has been a characteristic feature of the evolution of eukaryotic genomes. In the case of vertebrates, it has been proposed that an increase in gene number resulted from two rounds of duplication of the entire genome by polyploidization (the 2R hypothesis). In the most extensive test to date of this hypothesis, we compared gene numbers in homologous families and conducted phylogenetic analyses of gene families with two to eight members in the complete genomes of *Caenorhabditis elegans* and *Drosophila melanogaster* and the available portion of the human genome. Although the human genome showed a higher proportion of recent gene duplications than the other animal genomes, the proportion of duplications after the deuterostome–protostome split was constant across families, with no peak of such duplications in four-member families, contrary to the expectation of the 2R hypothesis. A substantial majority (70.9%) of human four-member families and four-member clusters in larger families showed topologies inconsistent with two rounds of polyploidization in vertebrates.

Evolutionary biologists have hypothesized that gene duplication has played an important role in evolution, particularly in eukaryotes, the genomes of which are characterized by the presence of numerous multigene families (Ohno 1970; Li 1983; Lynch and Conery 2000). By creating additional gene copies, gene duplication has permitted the evolution of new protein functions and thus is hypothesized to have played an important role in adaptive evolution (Ohno 1970; Hughes 1999a). Consistent with this hypothesis, there has been a tendency toward an increase in gene number over the course of evolution, with an increased gene number being at least roughly correlated with increased physiological complexity (Miklos and Rubin 1996).

Ohno (1970) argued that tandem duplication is unlikely to lead to new functional gene copies. As a consequence, he emphasized a role for duplication of complete genomes by polyploidization in adaptive evolution, especially in the case of vertebrates (Ohno 1970). In particular, the hypothesis that vertebrates underwent two rounds of genome duplication (the 2R hypothesis) has been widely cited (Lundin 1993; Sidow 1996; Meyer and Scharl 1999). Less frequently, a single round of polyploidization (the 1R hypothesis) has been proposed (Guigo et al. 1996). Thousands of functional genes that have arisen by tandem duplication are now known, thereby removing the initial rationale for Ohno's emphasis on whole-genome duplication. In addition, several recent studies involving phylogenetic analysis of selected gene families have failed to support key predictions of the 2R hypothesis (Hughes 1999b; Martin 1999, 2001; Hughes et al. 2001). However, because the number of gene families examined in these studies has been small, the possibility remains that more extensive analyses will reveal support for this hypothesis (Skrabaneck and Wolfe 1998; Makiłowski 2001).

The purpose of this study was to test these polyploidization hypotheses by a comparative analysis of patterns of gene

duplication in vertebrate and invertebrate animal genomes. We used three approaches: (1) We compared numbers of genes in homologous families in the complete genomes of yeast (*Saccharomyces cerevisiae*), the nematode worm *Caenorhabditis elegans*, and the insect *Drosophila melanogaster* and in the available portion of the human (*Homo sapiens*) genome. (2) We constructed phylogenetic trees of two- to eight-member families in yeast, *C. elegans*, and *Drosophila*, and human and used branching order in the phylogenetic trees to time events of gene duplication relative to three major cladogenic events: the animal–fungus divergence; the coelomate–nematode divergence; and the deuterostome–protostome divergence (Fig. 1a). Because we used branching order to time gene duplication relative to these cladogenic events (Fig. 1b) and because we used phylogenetic methods that do not assume a constant rate of evolution, our conclusions were not dependent on the assumption of a molecular clock or on the accuracy of divergence time estimates either from molecular data or from the fossil record. In addition, in the case of four-member human families, we tested for the consistency of the topology with that expected after two rounds of genome duplication (Fig. 1c,d). (3)

RESULTS

Homologous Family Size Ratios

The distributions of homologous family size ratios between the three animal genomes and yeast were all very similar (Table 1). In pairwise comparisons among these ratios, the Kolmogorov–Smirnov two-sample test was used to test the similarity of the two distributions. The hypothesis of an identical distribution could not be rejected in the comparison of *C. elegans*:yeast and *Drosophila*:yeast ratios. However, the hypothesis of identical distributions was rejected when the distributions of *C. elegans*:yeast and *Drosophila*:Yeast ratios were compared with that of human:yeast ratios (Table 1). The most striking difference between the former two distributions and that of human:yeast was the lower proportion of families, with a 1:1 ratio in the latter (Table 1). There was a highly

¹Corresponding author.

E-MAIL austin@biol.sc.edu; FAX (803) 777-4002.

Article published on-line before print: *Genome Res.*, 10.1101/gr.200601.
Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.200601>.

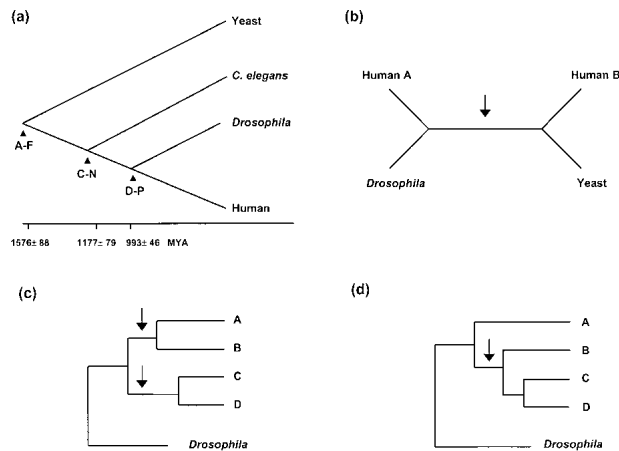


Figure 1 (a) Phylogenetic tree showing the major cladogenetic events used in timing gene duplications: (A–F) the animal–fungus divergence; (C–N), the coelomate–nematode divergence; (D–P), the deuterostome–protostome divergence. The topology of the tree and the divergence time estimates (\pm SE) are from Wang et al. (1999). However, gene duplications were timed relative to cladogenetic events independent of a molecular clock assumption. (b) Hypothetical gene family containing two human members (A and B). If the internal branch (indicated by arrow) is significantly supported, we can conclude (independent of the rooting of the tree) that A and B diverged prior to the deuterostome–protostome divergence. (c) Hypothetical four-member human gene family having a topology of the form (AB)(CD) consistent with the hypothesis of two rounds of genome duplication (the 2R hypothesis). (d) Hypothetical four-member human gene family having a topology of the form (A)(BD) inconsistent with the 2R hypothesis.

significant difference between the distribution of *C. elegans*:*Drosophila* ratios and that of human:*Drosophila* ratios (Table 1). Most of the difference between these two ratios could be attributed to much greater numbers of families with ratios of 2:1, 3:1, and 4:1 in the human:*Drosophila* distribution than in the *C. elegans*:*Drosophila* distribution (Table 1). Although advocates of the 2R hypothesis frequently state that many gene families in human have four times as many members as in *Drosophila* (Sidow 1996; Meyer and Scharl

1999), in our data, the percentage of families having this ratio in the human:*Drosophila* comparison was quite low (4.9%; Table 1).

Interestingly, 1375 families shared between human and *Drosophila* included only one human gene and one or more *Drosophila* genes; this total represents 49.8% of families shared between these two species (Table 1). Further, 1180 (42.7%) of families included a single gene in both human and *Drosophila*. The high proportion of single-gene families in human is very hard to explain on either the 1R or the 2R hypothesis, as both hypotheses require huge numbers of gene deletions after polyploidization to return to a single gene per family.

Timing of Gene Duplications

The human genome differed from that of *Drosophila* in having significantly lower proportions of gene duplication events in two- to eight-member families that could be dated by a significant internal branch prior to the animal–fungus divergence, the coelomate–nematode divergence, or the deuterostome–protostome divergence (Fig. 2). By contrast, the proportions of genes that could be dated prior to the animal–fungus divergence or the coelomate–nematode divergence did not differ significantly between *Drosophila* and *C. elegans* (Fig. 2).

In both *C. elegans* and *Drosophila* genomes, there was a significant nonuniformity among family size classes with respect to the proportion of duplications that could be dated prior to the coelomate–nematode divergence (Fig. 3). In both of these species, two- to three-member families included the highest proportion of duplications that could be dated prior to the coelomate–nematode divergence, whereas the proportion was lower in four-member families and lower still in five- to eight-member families (Fig. 3). By contrast, in the human genome, the proportions of duplications that could be dated prior to the coelomate–nematode divergence was remarkably constant across two- to three-member families, four-member families, and eight-member families (Fig. 3). In neither *Drosophila* nor human was there significant nonuniformity among family size classes with respect to the proportion of duplications that could be dated prior to the deuterostome–protostome divergence (Fig. 3).

Table 1. Summaries of the Distributions of Homologous Family Size Ratios

Ratio	<i>C. elegans</i> :yeast	<i>Drosophila</i> :yeast	Human:yeast	<i>C. elegans</i> : <i>Drosophila</i>	Human: <i>Drosophila</i>
<1:1	92 (16.5%)	93 (14.5%)	163 (22.8%)	250 (18.6%)	195 (7.1%)
1:5	2 (0.4%)	3 (0.5%)	2 (0.3%)	8 (0.6%)	7 (0.3%)
1:4	0 (0.0%)	0 (0.0%)	6 (0.8%)	7 (0.5%)	12 (0.4%)
1:3	6 (1.1%)	8 (1.3%)	12 (1.7%)	30 (2.2%)	16 (0.6%)
1:2	66 (11.9%)	69 (10.8%)	106 (14.8%)	117 (8.7%)	94 (3.4%)
1:1	342 (61.4%)	408 (63.8%)	337 (47.1%)	897 (66.9%)	1180 (42.7%)
>1:1	123 (22.1%)	139 (21.7%)	216 (30.2%)	194 (14.5%)	1386 (50.2%)
2:1	64 (11.5%)	74 (11.6%)	79 (11.0%)	101 (7.5%)	489 (17.7%)
3:1	15 (2.7%)	20 (3.1%)	33 (4.6%)	19 (1.4%)	265 (9.6%)
4:1	5 (0.9%)	5 (0.8%)	10 (1.4%)	9 (0.7%)	136 (4.9%)
5:1	2 (0.4%)	5 (0.8%)	6 (0.8%)	5 (0.4%)	78 (2.8%)
Total	557	640	716	1341	2761
Mean	1.35	1.24	1.56	1.18	2.19
Minimum	1:5	1:5	1:10	1:16	1:15
Maximum	31:1	13:1	167:3	52:1	53:1
Kolmogorov-Smirnov two-sample tests	$P < 0.05^a$	$P < 0.025^a$		$P < 0.001^b$	

Kolmogorov-Smirnov two-sample test of the equality of the underlying distributions with that of human:yeast^a; that of human:*Drosophila*^b.

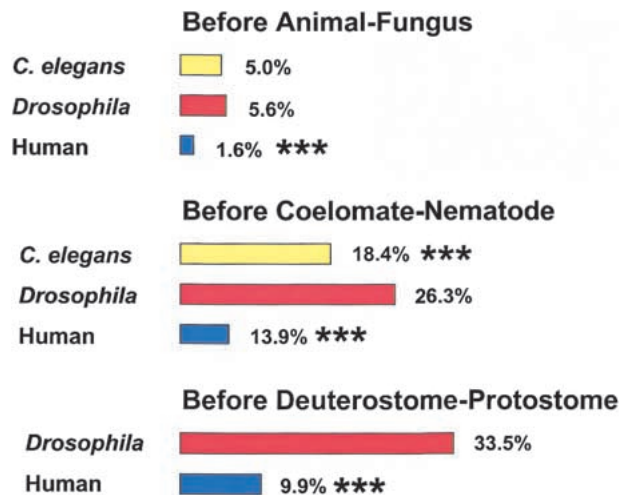


Figure 2 Numbers of gene duplications in two- to eight-member families. A duplication was dated prior to one of the three a major cladogenetic events (the animal–fungus divergence, the coelomate–nematode divergence, and the deuterostome–protostome divergence) if its occurrence prior to the event was supported by a significant internal branch. Chi square tests of the hypothesis that the proportion of duplications prior to a cladogenetic event differed from that in *Drosophila*: (***) $P < 0.001$. Numbers of duplication events were as follows: *Caenorhabditis elegans*, 463; *Drosophila* 567; human, 1760.

Topology in Four-Member Families

For four-member families in human and *Drosophila*, topologies of trees were categorized as follows: (1) supporting duplication of at least one gene pair prior to the protostome–deuterostome divergence; (2) supporting duplication after the deuterostome–protostome divergence and having a topology of the form (AB) (CD) (Fig. 1c); and (3) supporting duplication after the deuterostome–protostome divergence and having a topology of the form (A) (BCD) (Fig. 1d). In the case of the human genome, only category 2 supports the 2R hypothesis (Hughes 1999b). In the case of the human genome, 32 of 92 four-member families for which the phylogeny resolved the topology showed a topology supporting duplication of one or more genes prior to the deuterostome–protostome divergence, and, in 25 of these families, the relevant internal branch received significant support (Table 2). In 38 of the remaining families, the topology was of the form (A) (BCD), and, in 17 of these families, the internal branch establishing this topology received significant support (Table 2). Thus, 70 of 92 human four-member families (76.1%) showed topologies different from that predicted by the 2R hypothesis (Table 2).

Likewise, in four-gene clusters within five to eight-member families, the (A) (BCD) topology occurred more frequently than (AB)(CD) (Table 2). Of 42 such clusters in which the topology was resolved, 25 (59.5%) showed topologies inconsistent with the 2R hypothesis (Table 2). Thus, of a total of 134 resolved four-member phylogenies, 95 (70.9%) were not consistent with the 2R hypothesis. Similar results were reported for a smaller number of families by the International Human Genome Sequencing Consortium (2001).

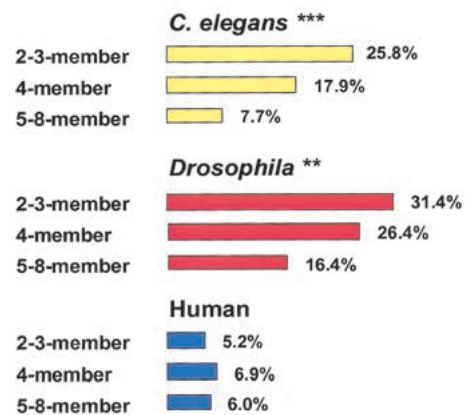
Interestingly, the patterns seen in *Drosophila* were quite similar to those seen in humans. In *Drosophila* 16 of 22 four-member families for which the topology was resolved (72.7%)

showed topologies different from that predicted by the 2R hypothesis. Thus these results suggest that the hypothesis of two rounds of genome duplication is no more likely to be true of vertebrates than of *Drosophila*.

DISCUSSION

Although the exact number of genes in the human genome remains to be determined, vertebrate genomes clearly contain more genes than those of *Drosophila* and *C. elegans* (Bork and Copley 2001). Consistent with the larger gene number in humans, our results showed that a lower proportion of gene duplications in humans than in *Drosophila* could be dated prior to the protostome–deuterostome divergence (Fig. 2). Thus, as expected, more gene duplications have occurred in the human lineage than in the *Drosophila* lineage since their last common ancestor. A number of authors have attributed the increase in gene number to one round (the 1R hypothesis) or two rounds (the 2R hypothesis) of genome duplication by polyploidization early in vertebrate history (Lundin 1993; Sidow 1996). Alternatively, it has been suggested that multiple independent gene duplications would be an alternative mechanism for increased gene numbers in vertebrates (Hughes et al. 2001)

a) Before Coelomate-Nematode



b) Before Deuterostome-Protostome

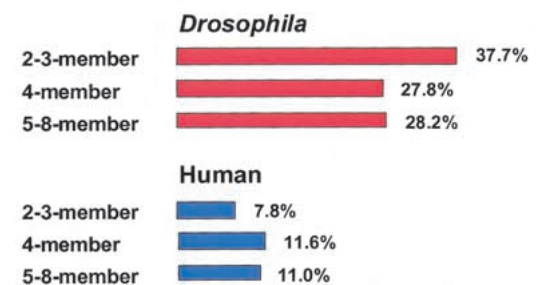


Figure 3 Numbers of gene duplications in three family size classes (two- to three-member families, four-member families, five- to eight-member families). A duplication was dated prior to one of two major cladogenetic events, (a) the coelomate–nematode divergence, and (b) the deuterostome–protostome divergence, if its occurrence prior to the event was supported by a significant internal branch. Chi-square tests of the uniformity across family size classes of the proportion of duplications prior to the cladogenetic event: (**) $P < 0.01$; (***) $P < 0.001$.

Table 2. Topologies of Four-Member Families and Clusters of Four Genes in 5–8-Member Families of Human and *Drosophila*

	Human	<i>Drosophila</i>
Four-member families		
At least one duplication before deuterostome-protostome divergence	32 (25) ^a	16 (11) ^a
All duplications after deuterostome-protostome divergence		
topology (AB)(CD) ^b	22 (11) ^d	4 (2) ^d
topology (A) (BCD) ^c	38 (17) ^d	2 (1) ^d
Unresolved	14	2
Total	106	24
Four-member clusters in 5–8-member families		
topology (AB) (CD) ^b	17 (5) ^d	5 (3) ^d
topology (A) (BCD) ^b	25 (10) ^d	6 (2) ^d
unresolved	8	0
total	50	11

Numbers are numbers of families showing a given topology.

^aNumbers in parentheses are numbers of families in which the branch establishing duplication prior to the deuterostome-protostome divergence received significant support.

^bAs in Figure 1c.

^cAs in Figure 1d.

^dNumbers in parentheses are cases where the branches establishing the topology (marked with arrows in Figs. 1c and 1d) received significant support.

Our study provided no support for either the 1R or the 2R hypothesis. Because the human proteome available to us is not yet complete, our results must be considered preliminary. Even if the available human proteome represents only 80%–90% of the total, it seems unlikely that the picture will change substantially with additional data.

Comparison of human:*Drosophila* and *C. elegans*:*Drosophila* homologous family size ratios revealed significantly different distributions (Table 1). The difference seemed to lie mainly in the much higher proportion of families falling in the 2:1, 3:1, and 4:1 in the human:*Drosophila* distribution than in the *C. elegans*:*Drosophila* distribution (Table 1). However, contrary to the expectation of the 2R hypothesis, the proportion of families with a 4:1 ratio in the human:*Drosophila* comparison was considerably lower than the proportion with a 2:1 ratio or that with 3:1 ratio. Indeed, <5% of genes families shared between human and *Drosophila* showed a 4:1 ratio, contrary to the concept of a four-to-one rule proposed by advocates of the 2R hypothesis (Meyer and Scharl 1999).

On the 2R hypothesis, we might expect to see evidence of a major burst of duplication in four-member gene families of vertebrates after the deuterostome–protostome divergence but not in families with more or fewer members. Contrary to this expectation, the proportion of duplications in human gene families that could be dated prior to the deuterostome–protostome divergence was remarkably constant across family size categories (Fig. 3).

Furthermore, a substantial majority of phylogenetic trees of four-member families and of four-member clusters within five to eight-member families revealed topologies inconsistent with the 2R hypothesis (Table 2). These results were consistent with those of a previous analysis using a smaller number of families (Hughes 1999b). Interestingly, the topologies of human gene families were no more supportive of two rounds of genome duplication than were those of *Drosophila* gene families (Table 2).

Some authors have taken the existence in four separate chromosomal locations of clusters of paralogous genes belonging to multiple gene families as evidence in favor of the

2R hypothesis. In the human genome, the most widely cited such cases involve chromosomes 1, 6, 9, and 19 (Kasahara et al. 1997) and chromosomes 2, 7, 12, and 17 (Lundin 1993; International Human Genome Sequencing Consortium 2001). However, the existence of such clusters can be taken as support for the 2R hypothesis only if phylogenetic analysis shows that the gene pairs involved were duplicated simultaneously early in vertebrate history (Hughes 1998). This prediction has been falsified by phylogenetic analyses of the gene families in both the clusters on chromosomes 1, 6, 9, and 19 (Hughes 1998; Yeager and Hughes 1999) and those on chromosomes 2, 7, 12, and 17 (Hughes et al. 2001). In both of these cases, the genes involved were duplicated at widely different times over the history of life (Hughes 1998; Yeager and Hughes 1999; Hughes et al. 2001). Venter et al. (2001), using a liberal criterion of homology, identified 1077 potentially duplicated regions in the human genome, each containing at least three pairs of duplicate genes. Application of phylogenetic analysis to all of the gene families in these regions will provide a further test of both the 2R and 1R hypotheses.

It might be argued that our results are consistent with the 1R hypothesis rather than the 2R hypothesis. However, our results are problematic for the 1R hypothesis as well. A high proportion (42.7%) of families shared between human and *Drosophila* were found to be represented by a single human gene (Table 1). Li et al. (2001) noted the large number of singletons in the human genome, and our results show that singletons also constitute a high proportion of the gene families humans share with *Drosophila*. Furthermore, in 85.7% of families, the human:*Drosophila* ratio of gene number was <4:1. Given these data, if early vertebrates underwent even a single polyploidization event, it must have been followed by deletion of the vast majority of duplicated genes. It is often assumed that polyploidization, because it duplicates numerous genes simultaneously, is a more parsimonious explanation of an increase in gene number than multiple independent events of tandem duplication, but this is not necessarily the case (Hughes et al. 2001). In the case of vertebrates, the numbers of events of gene deletion that must be assumed

under either the 1R or the 2R hypothesis far exceeds the number of events of tandem duplication that must be assumed if polyploidization is not evoked. Therefore, the hypothesis that the increase in gene number in vertebrates occurred as a result of multiple independent gene duplications, as well as occasional duplication of chromosomal blocks, is far more parsimonious given our results than any hypothesis invoking polyploidization.

METHODS

Sequences and Homologous Families

Sequences of proteome members were obtained from the following resources: for yeast, <http://genome-www.stanford.edu/Saccharomyces>; for *C. elegans* http://www.sanger.ac.uk/C_elegans (Wormpep 27); for *Drosophila*, ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_sets; and for human, the publicly available IPI database (International Human Genome Sequencing Consortium 2001) from <http://genome.cse.ucsc.edu>. The human database included known and predicted 31,778 proteins (International Human Genome Sequencing Consortium 2001). Using the BLASTP program (Altschul et al. 1997) to search for homology at the amino acid sequence level, we identified all shared families in pairwise comparisons between yeast, *C. elegans*, *Drosophila*, and human proteomes. To ensure that only genes homologous throughout their length were used, rather than those showing homology in only one or a few domains, we used a conservative Expect (E) value of 10^{-50} . We identified 557 families shared by *C. elegans* and yeast, 640 shared by *Drosophila* and yeast, 716 shared by human and yeast, 1341 shared by *C. elegans* and *Drosophila*, and 2761 shared by human and *Drosophila*. To compare family size in the different genomes, we examined the frequency distributions of the ratios of homologous family sizes for the following comparisons: *C. elegans*:yeast; *Drosophila*:yeast; human:yeast; *C. elegans*:*Drosophila*; and human:*Drosophila*. We refer to these ratios as homologous family size ratios.

Phylogenetic Analyses

We conducted phylogenetic analyses of two- to eight-member families in the three animal species. For each species, we included only families for which at least two sequences were available from one or more of the other two animal species or from yeast. We constructed 1330 such phylogenies (238 for *C. elegans*, 313 for *Drosophila*, and 779 for human). Phylogenetic trees were constructed by two methods: (1) the neighbor-joining (NJ) method (Saitou and Nei 1987) based on the uncorrected proportion (p) of amino acid difference; (2) the quartet maximum-likelihood (ML) method (Strimmer and von Haeseler 1996) as implemented in TREEPUZZLE 5.0, using the JTT (Jones et al. 1992) model of amino acid evolution and assuming that rate variation among sites followed a gamma distribution. NJ based on p is a simple method making minimal assumptions, whereas ML assumes an explicit evolutionary model (Nei and Kumar 2000). In the present case, the two methods yielded essentially identical results; thus, only the ML results are presented in the following. All trees were treated as unrooted, and no attempt was made to assign an outgroup to root any tree.

In each phylogeny, we timed each gene duplication event relative to the animal–fungus divergence, the coelomate–nematode divergence, and the deuterostome–protostome divergence (Fig. 1a) on the basis of the tree topology. This process is illustrated in Figure 1b. In the hypothetical family illustrated, there are two human genes (A and B; Fig. 1b). Given the topology of the tree, assuming that there

is significant support for the internal branch (indicated by arrow), we can conclude that these two human genes duplicated at least prior to the deuterostome–protostome divergence. We can make this conclusion independently of how the tree might be rooted. In the ML analyses, we concluded that a branch was significantly supported if it was supported in 95% or more of 10,000 puzzling steps; this represents a highly conservative test for significance of an internal branch (Strimmer and von Haeseler 1996). The 1330 trees analyzed included 2790 gene duplication events (463 in *C. elegans*, 567 in *Drosophila*, and 779 in human); we tallied the numbers of these for which there was significant support for gene duplication prior to each of the three cladogenetic events (Fig. 1a).

In four-member families in vertebrates, only one of the possible topologies is consistent with the 2R hypothesis (Hughes 1999b); this is a topology showing two clusters of two sequences, designated (AB) (CD) (Fig. 1c). Obviously, if one of more duplications in a vertebrate four-member family occurred prior to the deuterostome–protostome divergence, that family does not support the 2R hypothesis (Hughes 1999b). Likewise, even if all genes duplicated within the vertebrates, a topology in which one vertebrate gene falls outside the others, designated (A) (BCD) (Fig. 1d), is inconsistent with the 2R hypothesis (Hughes 1999b).

ACKNOWLEDGMENTS

This research was supported by grants to A.L.H. from the National Institutes of Health and the South Carolina Commission on Higher Education.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997 **25**: 3389–3402.
- Bork, R.P. and Copley, R. 2001. Filling in the gaps. *Nature* **409**: 818–820.
- Guigo, R., Muchnik, I., and Smith, T.F. 1996. Reconstruction of ancient molecular phylogeny. *Mol. Phylog. Evol.* **46**: 189–213.
- Hughes, A.L. 1998. Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1. *Mol. Biol. Evol.* **15**: 854–870.
- . 1999a. *Adaptive evolution of genes and genomes*. Oxford University Press, New York.
- . 1999b. 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.
- Hughes, A.L., da Silva, J., and Friedman, R. 2001. Ancient genome duplications did not structure the human *Hox*-bearing chromosomes. *Genome Res.* **11**: 771–780.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.
- Kasahara, M., Navaka, J., Sayya, Y., and Takahata, N. 1987. Chromosomal duplication and the emergence of the adaptive immune system. *Trends Genet.* **13**: 90–92.
- Li, W.-H. 1983. Evolution of duplicate genes and pseudogenes. In *Evolution of genes and proteins* (eds. M. Nei, and R.K. Koehn) pp. 14–37. Sinauer, Sunderland, MA.
- Li, W.-H., Gu, Z., Wang, H., and Nekrutenko, A. 2001. Evolutionary analyses of the human genome. *Nature* **409**: 847–852.
- Lundin, L.G. 1993. Evolution of the vertebrate genome as reflected

- in paralogous chromosome regions in man and the house mouse. *Genomics* **16**: 1–19.
- Lynch, M., and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- Makalowski, W. 2001. Are we polyploids? A brief history of one hypothesis. *Genome Res.* **11**: 667–670.
- Martin, A.P. 1999. Increasing genomic complexity by gene duplication and the origin of vertebrates. *Amer. Nat.* **154**: 111–128.
- . 2001. Is tetralogy true? Lack of support for the “one-to-four rule.” *Mol. Biol. Evol.* **18**: 89–93.
- Meyer, A., and Schartl, M. 1999. Gene and genome duplication 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.
- Miklos, G.L.G. and Rubin, G.M. 1996. The role of the genome project in determining gene function: Insights from model organisms. *Cell* **86**: 521–529.
- Nei, M. and Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer Verlag, New York.
- Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Sidow, A. 1996. Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* **6**: 715–722.
- Skrabanek, L. and Wolfe, K.H. 1998. Eukaryotic gene duplication—where's the evidence? *Curr. Opin. Genet. Dev.* **8**: 694–700.
- Strimmer, K. and von Haeseler, A. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**: 964–969.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Wang, D.Y., Kumar, S., and Hedges, S.B. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. R. Soc. Lond. B* **266**: 163–171.
- Yeager, M. and Hughes, A.L. 1999. Evolution of the mammalian MHC: Natural selection, recombination, and convergent evolution. *Immunol. Rev.* **167**: 45–58.

Received June 14, 2001; accepted in revised form August 7, 2001.115