

Dealing with saturation at the amino acid level: a case study based on anciently duplicated zebrafish genes

Yves Van de Peer^{a,*}, Tancred Frickey^b, John S. Taylor^b, Axel Meyer^b

^aDepartment of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology, Gent University, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium

^bDepartment of Biology, Universität Konstanz, D-78457 Konstanz, Germany

Received 12 September 2001; received in revised form 2 January 2002; accepted 7 May 2002

Abstract

The ray-finned fishes (Actinopterygii) seem to have two copies of many tetrapod (Sarcopterygii) genes. The origin of these duplicate fish genes is the subject of some controversy. One explanation for the existence of these extra fish genes could be an increase in the rate of independent gene duplications in fishes. Alternatively, gene duplicates in fish may have been formed in the ancestor of all or most Actinopterygii during a complete genome duplication event. A third possibility is that tetrapods have lost more genes than fish after gene or genome duplication events in the common ancestor of both lineages. These three hypotheses can be tested by phylogenetic reconstruction. Previously, we found that a large number of anciently duplicated genes of zebrafish are sister sequences in evolutionary trees suggesting that they were produced in Actinopterygii after the divergence of Sarcopterygii [Phil. Trans. R. Soc. Lond. B 356 (2001) 119]. On the other hand, several well-supported trees showed one of the two fish genes as the sister sequence to a monophyletic clade that included the second fish gene and genes from frog, chicken, mouse and human. These so-called outgroup topologies suggest that the origin of many fish duplicates predates the divergence of the Sarcopterygii and Actinopterygii and support the hypothesis that tetrapods have lost duplicates that have been retained in fish. Here we show that many of these ‘outgroup’ tree topologies are erroneous and can be corrected when mutational saturation is taken into account. To this end, a Java-based application has been developed to visualize the amount of saturation in amino acid sequences. The program graphically displays the number of observed frequent and rare amino acid replacements between pairs of sequences against their overall evolutionary distance. Discrimination between frequent and rare amino acid replacements is based on substitution probability matrices (e.g. PAM and BLOSUM). Evolutionary distances between sequences can be computed from the fraction of unsaturated sites only and evolutionary trees inferred by pairwise distance methods. When trees are computed by omitting the saturated fraction of sites, most fish duplicates are sister sequences. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Duplicate fish genes; Mutational saturation; Amino acid sequences

1. Introduction

For many genes, representatives of Actinopterygii (ray-finned fishes) have two copies whereas members of the sister group Sarcopterygii (lobe-finned fish and tetrapods) have only one (Wittbrodt et al., 1998; Meyer and Schartl, 1999; Taylor et al., 2001; Robinson-Rechavi et al., 2001). These duplicated fish genes are often sister sequences in phylogenetic trees (Fig. 1a) and occur on different linkage groups. Furthermore, different pairs of duplicates seem to have originated at about the same time (Taylor et al., 2001) and often show conserved synteny (Amores et al., 1998; Barbazuk et al., 2000; Postlethwait et al., 2000). These observations all provide support for the ‘fish-specific’

genome duplication hypothesis (Taylor et al., 2001). However, our previous work uncovered numerous examples of gene trees in which one of the duplicated fish genes formed the outgroup to the remaining homologs (Fig. 1b). These trees support the hypothesis that the duplicated fish genes were produced before the divergence of ray-finned and lobe-finned fishes and that the inequity in gene number is due to gene loss (Fig. 1c) in Sarcopterygii (Robinson-Rechavi et al., 2001). Here we show that many of these tree topologies are erroneous and can be corrected by considering mutational saturation at the amino acid (aa) level.

Mutational saturation in DNA and protein sequences occurs when sites have undergone multiple mutations causing sequence dissimilarity (the number of observed differences) to no longer accurately reflect the ‘true’ evolutionary distance, i.e. the number of substitutions that have actually

Abbreviations: aa, amino acid(s)

* Corresponding author. Tel.: +32-9-264-8756; fax: +32-9-264-5349.

E-mail address: yvdp@gengenp.rug.ac.be (Y. Van de Peer).

occurred since the divergence of two sequences. For the construction of phylogenetic trees, correct estimation of the evolutionary distance is crucial. Therefore, with pairwise distance and maximum-likelihood methods, the number of expected substitutions is estimated by applying a specific evolutionary model that accounts for unobserved substitutions. However, when substitutions are common, when substitution rates differ greatly among sites, or when distantly related sequences are being compared, mutational saturation is expected to be widespread and estimation of the ‘true’ evolutionary distance to be difficult (Swofford et al., 1996; Page and Holmes, 1998). Also with maximum-

parsimony methods, sites, which have been substituted multiple times, are often homoplasious and, therefore, misleading.

The problem of mutational saturation in nucleic acid sequence alignments has been discussed extensively (Page and Holmes, 1998; Nei and Kumar, 2000, and references therein). Usually, protein-coding genes are saturated at third codon positions due to the degenerative nature of the genetic code. Saturation in nucleic acids can also be caused by overabundance of transitions. In mitochondrial genes, for example, transitions occur at a much higher rate than transversions (Brown et al., 1982; Moriyama and Powell, 1997),

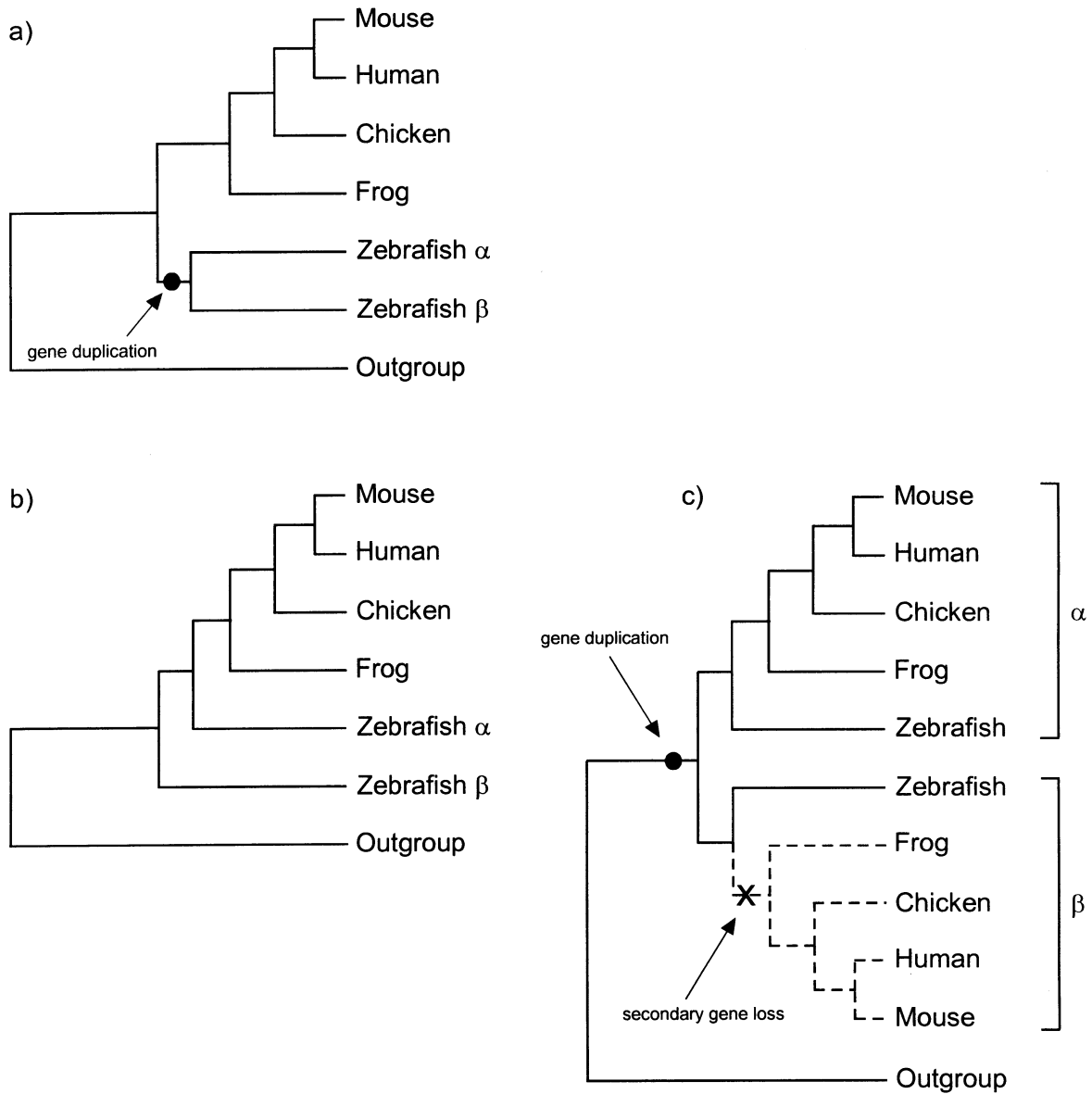


Fig. 1. (a) Expected tree topology for genes formed by a gene/genome duplication event specific to ray-finned fish. (b) Tree topology for some genes for which two copies exist in ray-finned fish and one in other vertebrates (Robinson-Rechavi et al., 2001; Taylor et al., 2001). This ‘outgroup’ topology is expected when genes produced during a duplication event in the ancestor of Actinopterygii plus Sarcopterygii have been secondarily lost in the sarcopterygian lineage as shown in (c). On the other hand, the tree topology shown in (b) might be an artifact due to increased evolutionary rates in one of the duplicates (Taylor and Brinkmann, 2001) or due to saturation (see text).

which can lead to saturation of transitional sites when more distantly related sequences are compared (Wakeley, 1996). Therefore, in many phylogenetic analyses, third codon positions are excluded (Chaw et al., 2000; Nickrent et al., 2000) or transitions are either down-weighted (Tang et al., 1999) or ignored in the computation of evolutionary distances. Alternatively, nucleic acid sequences are translated into aa sequences because they are much more conserved.

Despite the fact that aa sequences change relatively slowly when compared with nucleic acid sequences, saturation can become a problem when phylogenies are constructed for distantly related species (Meyer et al., 1986; Hansmann and Martin, 2000; Zardoya and Meyer, 2001). However, to our knowledge, no method has been devised yet to demonstrate and consider mutational saturation at the aa level.

2. Materials and methods

The usual way to demonstrate saturation in nucleotide sequences is to plot the fraction of differences between sequences against the evolutionary distance separating them. When the number of observed differences, for example for the fraction of third codon positions, no longer increases with increasing evolutionary distance, the sequence is said to be saturated (Page and Holmes, 1998). The same technique

can be applied to aa sequences. We have developed a Java application called ASaturA that discriminates aa substitutions with high and low probabilities of occurrence. All aa replacements are defined either as ‘frequent’ or as ‘rare’ depending on their mutation probabilities, which are inferred from substitution probability matrices, such as the well-known PAM (Dayhoff et al., 1978) and BLOSUM (Henikoff and Henikoff, 1992). These 20×20 matrices provide the empirically derived probabilities of one aa being replaced by another one when sequences have diverged over a certain evolutionary distance. ASaturA sorts all substitutions according to these probabilities and a probability ‘cutoff’ value can be chosen that differentiates between frequent and rare substitutions. For each sequence pair, the program plots the number of observed frequent and rare aa replacements against their evolutionary distance (Fig. 2). By modifying the substitution probability ‘cutoff’ value, the number of aa substitutions classified as frequent or rare can be changed. Ideally, careful selection of the ‘cutoff’ value splits the original data set into a saturated and an unsaturated one. Currently, we are investigating how to select the cutoff value automatically.

Besides the most widely used substitution probability matrices, such as PAM, BLOSUM, mtREV24 (Adachi and Hasegawa, 1996), and JTT (Jones et al., 1992), user-defined matrices can be used also. After the fraction of aa replacements is estimated that will probably be saturated,

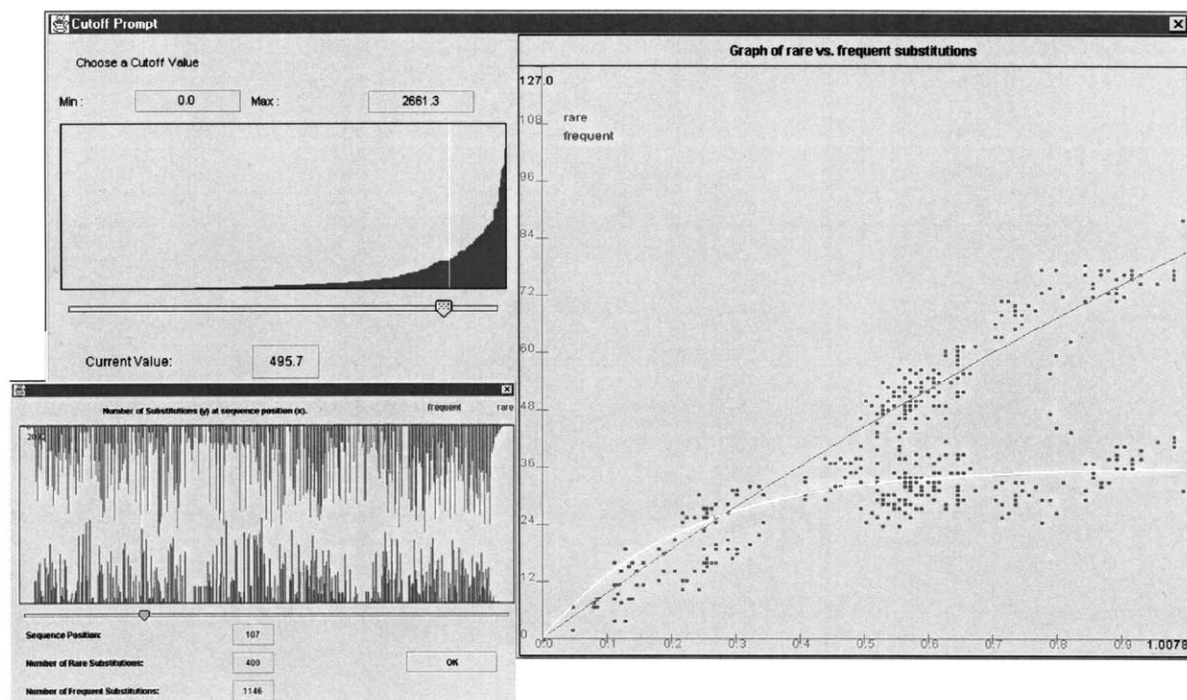


Fig. 2. Window showing the program’s ‘cutoff’ selection subroutine applied to an aa alignment of *Engrailed* (14 sequences). Substitution probabilities of a user-selected substitution probability matrix are sorted from low to high and a probability ‘cutoff’ value is chosen by moving the ruler (window top left). Black and gray (window right) represent the number of observed frequent and rare aa replacements for all pairs of sequences, plotted against their evolutionary distance, respectively. The window on the bottom left shows the number of frequent and rare substitutions over each alignment position for all pairwise comparisons. As can be clearly seen, a considerable fraction of the sequence alignment is saturated for evolutionary distances higher than 0.3.

evolutionary distances between sequences can be computed from the unsaturated fraction of sites (i.e. the ‘rare’ sites). Evolutionary trees can be inferred in ASaturA by pairwise distance methods, such as neighbor joining (Saitou and Nei, 1987). ASaturA is available from the authors upon request. In order to run the ASaturA program, the Java™ Runtime Environment needs to be installed as well.

Trees were reconstructed also by using maximum-parsimony (PAUP*; Swofford, 1998) and maximum-likelihood methods (TREE-PUZZLE; Strimmer and von Haeseler, 1996).

3. Results and discussion

3.1. Application of ASaturA to duplicated fish genes

Fig. 3 shows the application of the approach described above to an aa sequence alignment of *Engrailed*, a gene that codes for a family of homeodomain-containing transcription factors that play important roles in animal segmentation, limb formation and neurogenesis. The phylogenetic tree of

Fig. 3a was constructed by neighbor joining (Saitou and Nei, 1987) and inferred from evolutionary distances computed according to Kimura (1983) and taking into account all sites. The two zebrafish *En2* genes cluster together, although one of the two *En1* genes is the sister sequence to the remaining *En1* genes, i.e. the *En1* tree has the ‘outgroup’ topology. The same topology is obtained with maximum-parsimony and maximum-likelihood methods taking into account among-site rate variation (see below). When a distance tree is computed on the basis of the unsaturated fraction of sites only (see Fig. 2), the *En1* and the *En2* duplicates clearly show a sister-group relationship (Fig. 3b), suggesting that the duplication events that formed both pairs occurred after the divergence of Actinopterygii and Sarcopterygii. Fig. 4 presents four more phylogenetic trees in which a topology consistent with a fish-specific duplication event was reconstructed by considering aa saturation only. Because for all these genes, the duplication is ancient – third codon positions are saturated (Van de Peer et al., 2001) – and the paralogs are found on different linkage groups (Barbazuk et al., 2000; Woods et al., 2000), we strongly believe that the trees with the two fish duplicates as sister sequences are

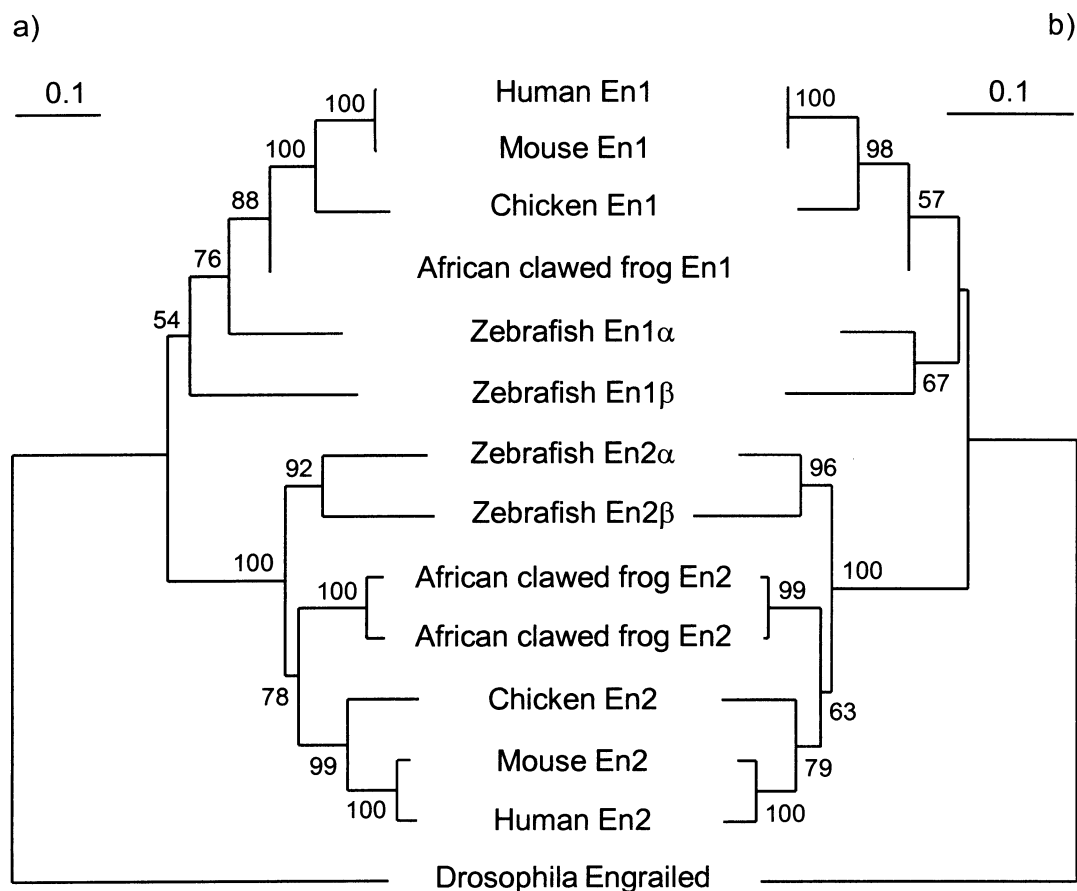


Fig. 3. Phylogenetic tree of En1 and En2. The *Drosophila Engrailed* gene was used to root the tree. Bootstrap values (Felsenstein, 1985) above 50% are shown at the internodes. (a) Neighbor-joining tree inferred from evolutionary distances computed according to Kimura (1983) and taking into account all sites. The same tree topology was inferred by maximum-parsimony and maximum-likelihood methods, even when among-site rate variation was considered (see text). (b) Neighbor-joining tree based on evolutionary distances computed according to Kimura, but taking into account only the unsaturated fraction of sites.

correct and, therefore, reflect a large-scale gene or even complete genome duplication event. Removal of the most variable aa positions might reduce the resolution of the tree, but there is no reason to believe that removal of saturated positions should bias our results in favor of the ‘sister sequences’ topology.

3.2. Complex versus simple models

For many of the genes for which two copies are present in zebrafish, one of the duplicates showed an increased rate of evolution after duplication (Van de Peer et al., 2001), possibly leading to a so-called long-branch attraction in phylogenetic trees. Long-branch artifacts have gained much interest the last few years and are now generally considered to be a major source of distorted tree topologies (Forterre and Philippe, 1999). This is especially true when there is considerable among-site rate variation in the sequences (Van de Peer et al., 2000). The existence of among-site rate variation implies that the majority of mutations take place at the same, relatively small fraction of positions, which are probably saturated. Naturally, this effect is even more pronounced when distances between sequences are large. Consequently, genetic distances are seriously underestimated when site-to-site rate variation is not taken into account, which often results in an artificial clustering of long branches, or in long branches that are pulled more closely to the base of the tree in the presence of a distantly related outgroup (Felsenstein, 1978; Olsen, 1987, Yang, 1996; Brinkmann and Philippe, 1999). Over the last few years, much effort has been put in considering among-site rate variation in the estimation of evolutionary distances or tree likelihoods. Most of these approaches are based on the observation that substitution rates are usually gamma distributed (Yang, 1996), involving a parameter α that describes the extent of the rate variation, from which an equation can be derived to compute the evolutionary distance from the observed sequence dissimilarities. Nowadays, the estimation of evolutionary distances or likelihoods based on gamma-distributed substitution rates is used extensively and has been implemented in many tree inference programs. In theory, the preferable method to deal with mutational saturation accounts for the effect of saturation by applying a model that describes rate heterogeneity. This method has the advantage that the maximum amount of information is used to estimate the evolutionary distance, which is not the case when the saturated fraction of sites is omitted from the analysis, as is the approach described here. Therefore, we have reconstructed trees by considering all positions but by applying methods that take into account among-site rate variation. Nevertheless, in most cases, the ‘correct’ tree topologies (i.e. those reconstructed using only unsaturated aa sites) could not be recovered.

It is generally known that modeling the evolution of aa sequences is much harder than modeling that of nucleic acid sequences, mainly because of many more character states.

This is particularly true when the number of sequences is limited and/or are short. Therefore, the removal of saturated sites as discussed should be viewed as a suitable approach when more adequate models of evolution do not exist or are unreliable. Furthermore, the ASaturA program can be simply used to visualize the amount of saturation in the aa sequence alignment.

3.3. Tree topology and inferred duplication events

When one tries to reconstruct gene and genome duplication events, the inference of reliable phylogenetic trees is crucial, because inaccurate tree topologies can lead to the wrong conclusion as shown in Fig. 1b,c. The occurrence of complete genome duplication events in eukaryotic evolu-

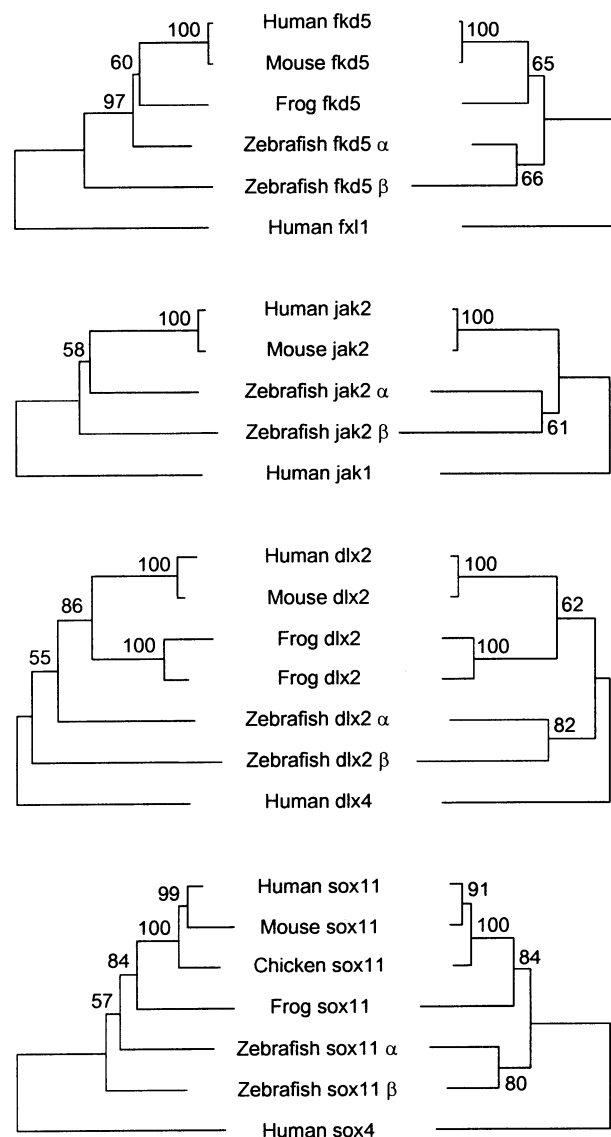


Fig. 4. Phylogenetic trees computed on the basis of all sites versus phylogenetic trees computed on the basis of the unsaturated fraction only. Genes analyzed are *fkd5*, *jak2*, *dlx2*, and *sox11*. See Taylor et al. (2001) and Van de Peer et al. (2001) for more details.

tion is highly debated. For example, two genome duplication events have been suggested to have happened during the evolution of vertebrates (the 2R hypothesis) from an invertebrate *amphioxus*-like ancestor (e.g. Spring, 1997; Holland, 1998). The 2R hypothesis can be tested by tree construction. When two rounds of genome duplications actually are found early in the vertebrate lineage, phylogenetic trees for four paralogous sequences of vertebrates plus an (invertebrate) outgroup should have the (A,B)(C,D) tree topology (Skrabanek and Wolfe, 1998). The first genome duplication produced the common ancestors of sequences (A,B) and (C,D) and later these two lineages split simultaneously in the second genome duplication. Such trees should have two testable characteristics: a '2 + 2' topology, and equal ages for the later two divergences. If the widespread occurrence of topologies of the (A,B)(C,D) type could be demonstrated, this would provide strong evidence for two rounds of genome duplications. In contrast, a high proportion of gene families with paralogs that diverged before the origin of vertebrates or with topologies of, for example, the (A)(B,C,D) type would argue against the one-to-four hypothesis. So far, the 2R hypothesis is refuted by some on the basis of inferred tree topologies (Hughes, 1999) but accepted by others (Wang and Gu, 2000). It is our intention to apply the method described here to the published data sets to see whether the topology predicted by 2R is inferred when aa saturation is taken into account.

Acknowledgements

The authors thank Rafa Zardoya for pointing out the problem of saturation in amino acid sequences and for proposing a method to deal with it, and Tomas Hrbek is acknowledged for critical reading of an earlier version of the paper. This work was supported by the University of Konstanz and the German Science Foundation (DFG PE 842/21). J.S.T. is indebted to the National Sciences and Engineering Research Council of Canada for a Postdoctoral Fellowship. Y.V.d.P. is a Postdoctoral Fellow of the Fund for Scientific Research (Flanders).

References

- Adachi, J., Hasegawa, M., 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* 42, 459–468.
- Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., Postlethwait, J.H., 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711–1714.
- Barbazuk, W.B., Korf, I., Kadavi, C., Heyen, J., Tata, S., Wun, E., Bedell, J.A., McPherson, J.D., Johnson, S.L., 2000. The syntenic relationships of the zebrafish and human genomes. *Genome Res.* 10, 1351–1358.
- Brinkmann, H., Philippe, H., 1999. Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol. Biol. Evol.* 16, 817825.
- Brown, W.M., Prager, E.M., Wang, A., Wilson, A.C., 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* 18, 225–239.
- Chaw, S.M., Parkinson, C.L., Cheng, Y., Vincent, T.M., Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc. Natl. Acad. Sci. USA* 97, 4086–4091.
- Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C., 1978. A model of evolutionary change in proteins. In: Dayhoff, M.O. (Ed.). *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, DC, pp. 345–352.
- Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Forster, P., Philippe, H., 1999. Where is the root of the universal tree of life? *BioEssays* 21, 871–879.
- Hansmann, S., Martin, W., 2000. Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis. *Int. J. Syst. Evol. Microbiol.* 50, 1655–1663.
- Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89, 10915–10919.
- Holland, P.W.H., 1998. Major transitions in animal evolution: a developmental genetic perspective. *Am. Zool.* 38, 829–842.
- Hughes, A., 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.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- Kimura, M. (Ed.), 1983. *The Neutral Theory of Molecular Evolution* Cambridge University Press, Cambridge, UK.
- Meyer, A., Schartl, M., 1999. Gene and genome duplications in vertebrates: the one-to-four (to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* 11, 699–704.
- Meyer, T.E., Cusanovich, M.A., Kamen, M.D., 1986. Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. *Proc. Natl. Acad. Sci. USA* 83, 217–220.
- Moriyama, E.N., Powell, J.R., 1997. Synonymous substitution rates in *Drosophila*: mitochondrial versus nuclear genes. *J. Mol. Evol.* 45, 378–391.
- Nei, M., Kumar, S. (Eds.), 2000. *Molecular Evolution and Phylogenetics* Oxford University Press, Oxford, UK.
- Nickrent, D.L., Parkinson, C.L., Palmer, J.D., Duff, R.J., 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* 17, 1885–1895.
- Olsen, G.J., 1987. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harbor Symp. Quant. Biol.* 52, 825–837.
- Page, R.D.M., Holmes, E.C. (Eds.), 1998. *Molecular Evolution, A Phylogenetic Approach* Blackwell Science, Oxford, UK.
- Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., Talbot, W.S., 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* 10, 1890–1902.
- Robinson-Rechavi, M., Marchand, O., Escriva, H., Bardet, P.L., Zelus, D., Hughes, S., Laudet, V., 2001. Euteleost fish genomes are characterized by expansion of gene families. *Genome Res.* 11, 781–788.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Spring, J., 1997. Vertebrate evolution by interspecific hybridization – are we polyploid? *FEBS Lett.* 400, 2–8.
- Skrabanek, L., Wolfe, K.H., 1998. Eukaryote genome duplication where's the evidence? *Curr. Opin. Gen. Dev.* 8, 694–700.
- Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969.

- Swofford, D.L. (Ed.), 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 Sinauer Associates, Sunderland, MA.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.). *Molecular Systematics*, Sinauer Associates, Sunderland, MA, pp. 407–514.
- Tang, K.L., Berendzen, P.B., Wiley, E.O., Morrissey, J.F., Winterbottom, R., Johnson, G.D., 1999. The phylogenetic relationships of the suborder Acanthuroidei (Teleostei: Perciformes) based on molecular and morphological evidence. *Mol. Phylogenet. Evol.* 11, 415–425.
- Taylor, J.S., Brinkmann, H., 2001. 2R or not 2R? *Trends Genet.* 17, 488–489.
- Taylor, J.S., Van de Peer, Y., Braach, I., Meyer, A., 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. *Phil. Trans. R. Soc. Lond. B* 356, 119.
- Van de Peer, Y., Ben Ali, A., Meyer, A., 2000. Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* 246, 1–8.
- Van de Peer, Y., Taylor, J.S., Braasch, I., Meyer, A., 2001. The ghost of selection past: rates of evolution and functional divergence in anciently duplicated genes. *J. Mol. Evol.* 53, 434–444.
- Wang, Y., Gu, X., 2000. Evolutionary patterns of gene families generated in the early stage of vertebrates. *J. Mol. Evol.* 51, 88–96.
- Wakeley, J., 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol. Evol.* 11, 158–163.
- Wittbrodt, J., Meyer, A., Schartl, M., 1998. More genes in fish? *BioEssays* 20, 511–512.
- Woods, I.G., Kelly, P.D., Chu, F., Ngo-Hazelett, P., Yan, Y.L., Huang, H., Postlethwait, J.H., Talbot, W.S., 2000. A comparative map of the zebrafish genome. *Genome Res.* 10, 1903–1914.
- Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11, 367–372.
- Zardoya, R., Meyer, A., 2001. Vertebrate phylogeny: limits of inference of mitochondrial genome and nuclear rDNA data due to an adverse phylogenetic signal/noise ratio. In: Ahlberg, P. (Ed.). *Major Events in Early Vertebrate Evolution*, Taylor and Francis, London, pp. 135–155.