THE ROYAL SOCIETY

Received 6 March 2002
Accepted 1 May 2002
Published online 28 June 2002

Going nuclear: gene family evolution and vertebrate phylogeny reconciled

James A. Cotton* **and Roderic D. M. Page**

Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

Gene duplications have been common throughout vertebrate evolution, introducing paralogy and so complicating phylogenetic inference from nuclear genes. Reconciled trees are one method capable of dealing with paralogy, using the relationship between a gene phylogeny and the phylogeny of the organisms containing those genes to identify gene duplication events. This allows us to infer phylogenies from gene families containing both orthologous and paralogous copies. Vertebrate phylogeny is well understood from morphological and palaeontological data, but studies using mitochondrial sequence data have failed to reproduce this classical view. Reconciled tree analysis of a database of 118 vertebrate gene families supports a largely classical vertebrate phylogeny.

Keywords: reconciled trees; gene families; vertebrate phylogeny; gene duplication; gene tree parsimony

1. INTRODUCTION

The central assumption of molecular systematics is that a phylogeny estimated from a set of gene sequences tells us something about the phylogeny of the organisms from which the genes have been isolated. In fact, systematists generally assume that the gene phylogeny (or gene tree) is isomorphic with the organism phylogeny (or species tree), so that a correct estimate of the species tree can be obtained by simply relabelling the leaves of the tree with the appropriate species names. In this case, differences between phylogenies from different loci—or differences between a gene tree and the commonly accepted species tree—are due to either the method by which gene phylogenies have been constructed or sampling error in the estimate of gene phylogeny. In the latter case, more sequence data should produce the correct species tree.

However, gene trees are not species trees and a number of evolutionary processes can introduce differences between a correctly estimated gene phylogeny and the correct species phylogeny (Doyle 1992; Maddison 1997). These processes are horizontal transfer, duplication and loss and deep coalescence (Doyle 1992; Slowinski & Page 1999). Because these events introduce differences between the gene tree and species tree, we can use incongruence between these two trees to infer the past occurrence of the events (Page & Charleston 1997*a*). This is the motivation behind reconciled trees. Reconciled trees are a general method for analysing historical relationships where one entity tracks another, with the fidelity of this 'tracking' dependent on how often events such as duplication, horizontal transfer and lineage sorting occur (Page & Charleston 1998). These events will introduce differences between the trees that describe the hierarchy of the two entities, as in figure 1, where a duplication in the gene tree and three gene losses explain the difference between the gene and species trees. Where all these different events are allowed, it can be very difficult correctly to reconstruct potential evolutionary scenarios (Charleston 1998), but if we restrict the analysis to consider only duplications and losses, then finding the most parsimonious reconstruction of events is relatively trivial and can be computed in linear time (Zhang 1997).

As we consider all of the gene trees to be independent estimates of the underlying species phylogeny, the most parsimonious species tree is that which implies the minimum number of gene duplication (or duplication and loss) events over the set of gene families, and we can use simple and standard heuristic methods to find an optimal species tree topology (Page & Charleston 1997*b*). Using the number of gene duplications as an optimality criterion to choose between competing phylogenetic hypotheses in this way has become known as 'gene tree parsimony' (Slowinski & Page 1999). Gene tree parsimony thus treats gene trees as characters of species, in contrast to conventional phylogenetic methods using molecular sequences as characters of organisms, conflating organismal and gene phylogenies.

The evolution of the vertebrates represents an ideal case for testing the utility of reconciled tree methods (Page 2000). Vertebrate classification has been of interest since antiquity, and a great deal of morphological data from both extant and fossil taxa have produced a well-supported outline of vertebrate phylogeny (figure 2). Vertebrate workers have a keen sense of where the vertebrate tree is fairly robust and where relationships are much less clear, and all of these areas have attracted a great deal of debate. There is thus an opportunity for new techniques both to prove themselves, by successfully reconstructing those parts of the tree that are more or less beyond doubt, and to make a real contribution to resolving areas of contention.

Given the great deal of support for much of the current pattern of vertebrate relationships, it is surprising how poorly molecular methods have fared in reconstructing the broad outline of vertebrate evolution. This is particularly worrying in the case of mitochondrial genome sequences, which are relatively large markers that have been thought of as ideal for phylogenetic work and are certainly very

*Author for correspondence (j.cotton@udcf.gla.ac.uk).

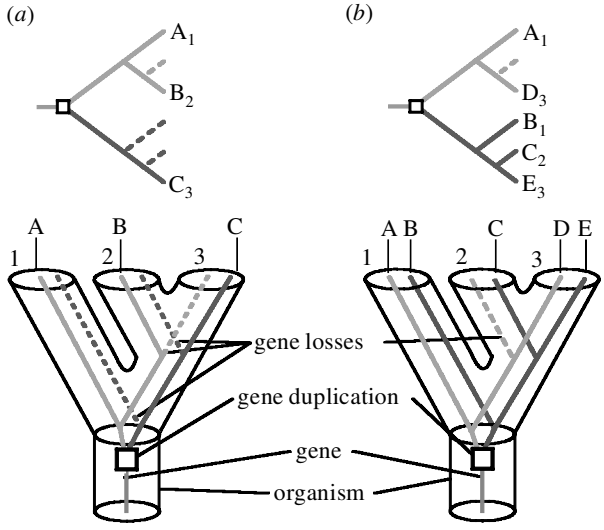

Figure 1. Gene duplication and loss can introduce incongruence between gene phylogenies and species phylogenies. (*a*) With three genes (A–C) sampled from three different species (1–3), the difference in topology between the gene and species trees can be explained by one gene duplication and three losses. The same approach also applies where multiple genes are known from each species: (*b*) shows a gene tree requiring one duplication and one loss. Reconciled trees can be seen as representing the simplest embedding of a gene phylogeny inside a given species phylogeny.

commonly used. Figure 3 shows two recently published phylogenies based on mitochondrial genome sequences, showing the unusual relationships between major groups of basal vertebrates typical of analyses based on these data.

Some of the errors in mitochondrial phylogenies have been due to incorrect rooting of the gnathostome part of the tree (Takezaki & Gojobori 1999), but other unusual placements occur. These errors occur despite mitochondrial loci having increasingly good taxon sampling. Explaining these erroneous results has become a major concern in the literature, particularly because several studies show high bootstrap support for unusual relationships (Zardoya & Meyer 1996; Naylor & Brown 1997), which some have taken at face value as providing strong evidence for these relationships. Other studies have sought to explain the unorthodox relationships as artefacts due to a low signal-to-noise ratio (Zardoya & Meyer 2001*b*) and wide differences in substitution rates between lineages (Takezaki & Gojobori 1999), between classes of amino acids (Naylor & Brown 1997) and between sites (Takezaki & Gojobori 1999). Most authors agree that phylogenetic results from recent analyses of whole mitochondrial genomes need to be confirmed with data from nuclear genes (Curole & Kocher 1999; Takezaki & Gojobori 1999; Zardoya & Meyer 2001*b*).

We have used gene tree parsimony to reconstruct vertebrate phylogeny based on a database of 118 vertebrate gene families. These analyses demonstrate the utility of reconciled trees in inferring phylogenies from gene family data, supporting most of the conventional vertebrate phylogeny and adding to the evidence for some more controversial relationships, such as a monophyletic cyclostome clade of lampreys and hagfish.

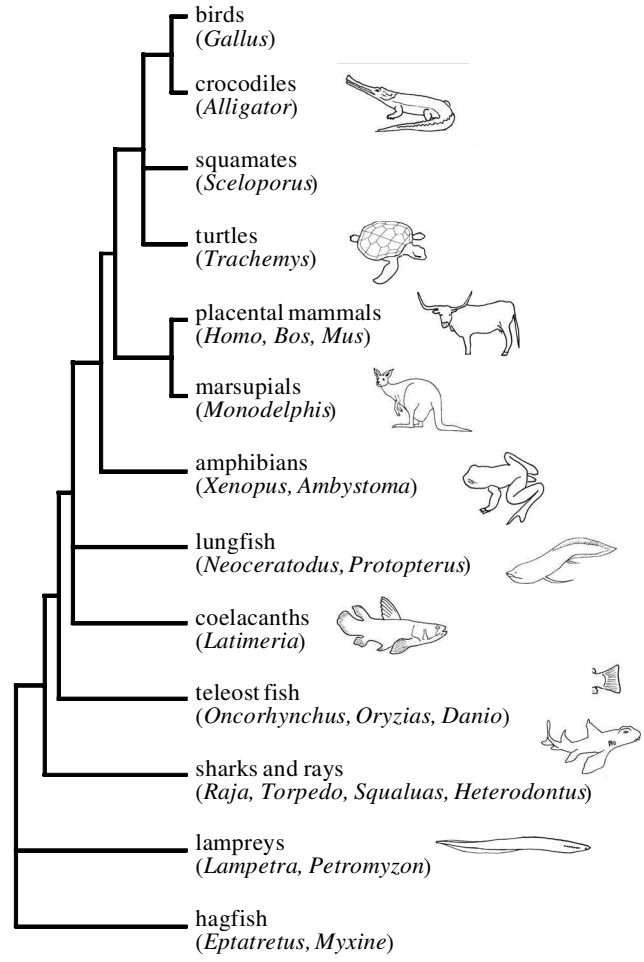

Figure 2. A traditional view of vertebrate phylogeny, based on morphological and palaeontological data. Based on Bishop & Friday (1988). The names of all genera included in the gene tree analysis (see figure 4) are listed.

2. MATERIAL AND METHODS

The data used in this study are available from http://darwin.zoology.gla.ac.uk/~jcotton/vertebrate_data. This includes a complete list of the gene families used in this paper, with phylogenies and alignments for each, along with the GeneTree input file for the analysis.

(a) *Gene family phylogenies*

We chose those representatives of the major vertebrate groups present in the largest number of gene families in the Hovergen (Duret *et al.* 1994) database. We assumed the monophyly of genera, grouping genes from all species in a genus together. Where no genus in a particular group was well represented, an additional genus was used, so that data from both could help to determine accurately the relationship of the larger group. Genera included are listed in figure 2. Gene families sampling at least five vertebrate classes were selected from Hovergen, with additional families chosen if they provided evidence about the relationships of those genera that were poorly sampled in the initial selection. Outgroups for each gene family were found using sequence similarity searches against a number of sequence databases to identify related genes: either invertebrate orthologues or vertebrate paralogues. Due to the size of the dataset, amino acid sequences were aligned in ClustalW (Thompson *et*

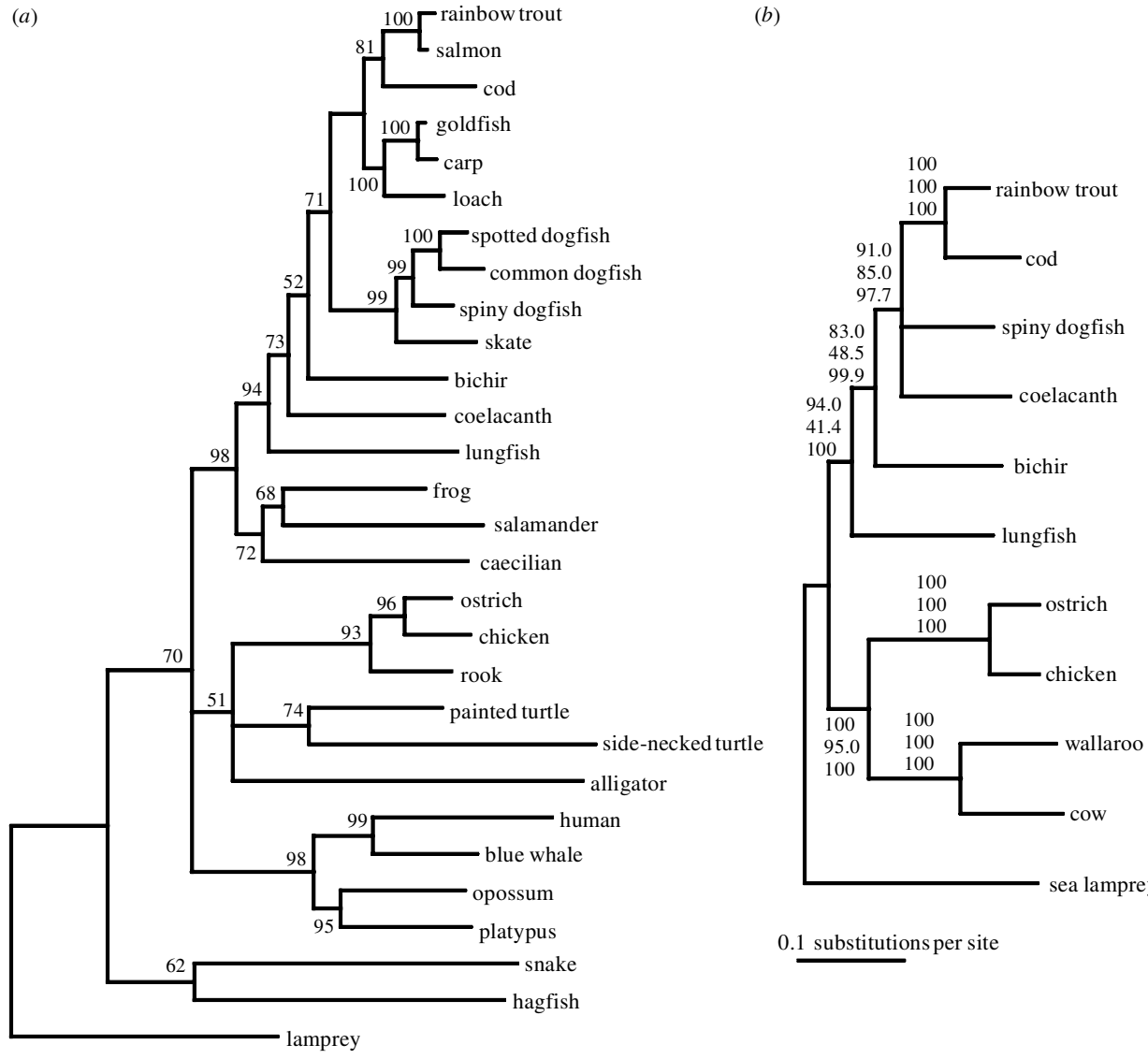

Figure 3. Vertebrate phylogenies based on whole mitochondrial genome data. (*a*) A maximum-likelihood tree from Zardoya & Meyer (2001*b*). Numbers on nodes are bootstrap percentages based on 100 pseudo-replications. Zardoya and Meyer do not accept this tree of vertebrate relationships, but are unable to reconstruct a more reasonable phylogeny. (*b*) The maximum-likelihood tree from Rasmussen & Arnason (1999). Figures on branches are neighbour-joining (top) and maximum-parsimony (middle) bootstrap values based on 100 replicates, and maximum-likelihood (bottom) support values from 1000 puzzle replicates. Both trees were constructed using PUZZLE (Strimmer & von Haeseler 1996) and the mtREV-24 model.

al. 1994) using default parameters and neighbour-joining phylogenies constructed in CLUSTALW, including gapped positions and using uncorrected distances. Alignments were also examined by eye to ensure that they were reasonably sensible, and so that small sequence fragments that might reduce alignment quality and be difficult to place phylogenetically were removed. Several gene families were excluded at this stage and some large gene families split into subsets. This rapid approach was chosen to allow our methods to be scaled up to much larger amounts of data. It is important to note that many gene families only contained sequences from a few species and that some pairs of genera never co-occurred in the same gene family.

(b) *Gene tree parsimony*

The species phylogeny minimizing the total number of duplications on the gene family trees was found using GENETREE (Page 1998), constrained to consider only trees supporting the monophyly of the two genera each of lampreys, hagfish, lungfish and rays. Fifty heuristic searches were performed from random starting trees, with the 'steepest ascent' option and using alternate nearest-neighbour interchange and subtree pruning and regrafting branch swapping (Page & Charleston 1997*b*). The same analysis, but minimizing the total numbers of duplications and losses, was also performed. Note that because each of the gene family trees is rooted, the species tree found by this procedure is also a rooted tree.

(c) *Confidence in species tree nodes*

Current implementations of reconciled trees have lacked any method to take account of uncertainty in gene family trees and express confidence levels in the reconciled species tree (Page & Cotton 2000). To calculate support values on nodes, 100 pseudoreplicate alignments were generated for each gene family using the bootstrap (Felsenstein 1985) and phylogenies for each replicate constructed exactly as described above. The species tree minimizing the number of gene duplications was then found

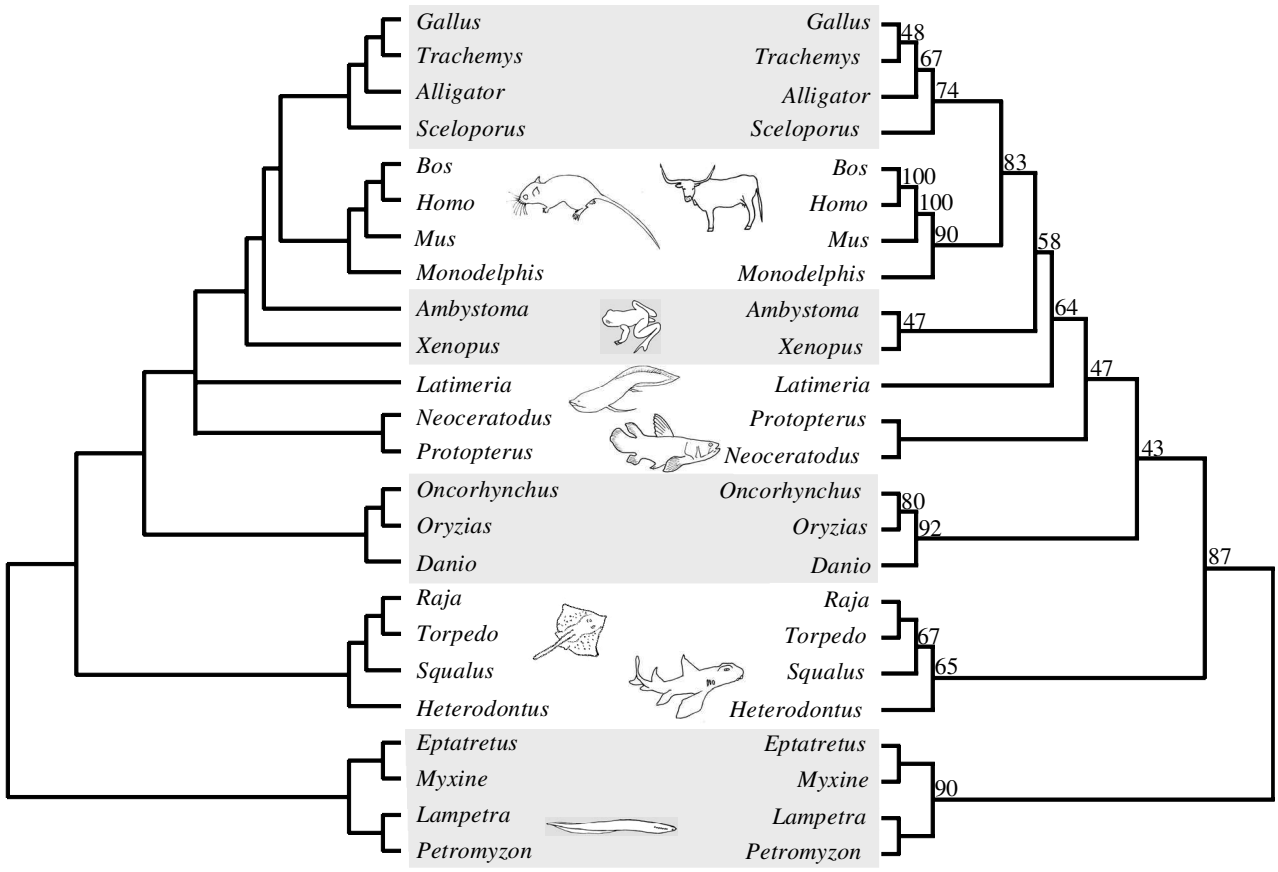

Figure 4. Phylogenies of vertebrates reconstructed using gene tree parsimony on a set of 118 nuclear genes. Alternate bands of shading and non-shading identify traditional higher taxonomic groups of vertebrates. (*a*) The strict consensus of three most parsimonious trees, each requiring 1380 gene duplications to fit the gene family trees. (*b*) The majority-rule consensus of 100 bootstrap replicates as described in § 2c. Figures on nodes are bootstrap percentages from this analysis.

for successive trees from the bootstrap profile of each gene family, producing 100 species trees. Each search was performed from a single random starting tree, using the same options as the main gene tree parsimony analysis but only finding a single shortest tree for each replicate. Support values analogous to standard bootstrap values could then be calculated for nodes in the species tree.

3. RESULTS

The results of our gene tree parsimony analysis are shown in figure 4. Fifty heuristic searches found the same island of three equally parsimonious shortest trees 19 times. Figure 4 also shows the majority rule consensus tree of the 100 species trees from gene tree parsimony analysis of the bootstrap profile of gene trees. Our phylogenies differ very little from traditional views of vertebrate relationships. Relationships within the major terminal groups are reconstructed identically to recent phylogenetic analyses for the teleosts (Nelson 1994) and chondrichthyes (Maisey 1984). Interestingly, we get very good support for the three-taxon relationship between *Mus*, *Bos* and *Homo*, agreeing with the largest study of mammalian phylogeny (Liu *et al.* 2001) but disagreeing with a recent molecular study (Murphy *et al.* 2001). There is ongoing difficulty in resolving many ordinal-level relationships within the placental mammals (Waddell *et al.* 1999).

There are two main competing hypotheses about the relationship between hagfish, lampreys and the higher, jawed vertebrates or gnathostomes. Our analysis very strongly supports a close relationship of hagfish and lampreys, with these groups together forming a sister clade to the gnathostomes, called the cyclostomes. The other popular alternative unites lampreys and vertebrates as a 'Vertebrata' group, which together with the hagfish forms the 'Craniata'. Traditional classifications included the cyclostome group, but the first cladistic studies of the group led to a new view of the group (Løvtrup 1977; Janvier 1981) and eventually to a consensus among morphologists supporting the alternative Vertebrata group (Forey & Janvier 1993; Janvier 1996). By contrast, molecular phylogenies have consistently supported a cyclostome group, with evidence from 18S and 28S rRNA molecules (Stock & Whitt 1992; Mallatt & Sullivan 1998) and a number of nuclear loci (Kuraku *et al.* 1999). Evidence from mitochondrial genomes has been somewhat equivocal: a maximum-likelihood analysis of the hagfish mitochondrial genome sequence (Rasmussen *et al.* 1998) supported the lamprey and gnathostome clade, and a subsequent analysis (Delarbre *et al.* 2000) found that the position of the hagfish depended on the method of analysis used. Recent evidence from additional sequence data strongly supports cylcostome monophyly (Delarbre *et al.* 2002). There is also some other molecular evidence supporting a lamprey and gnathostome clade (Suzuki *et al.* 1995; Gursoy *et al.* 2000; Page 2000), but our results show that nuclear gene loci strongly support a cyclostome clade, adding weight to a recent morphological re-evaluation of basal vertebrate relationships (Mallatt 1997).

Another area of considerable debate is the relationship between lungfish, coelacanths and the tetrapods. The traditional taxonomy placed the fossil coelacanths as the closest relative of tetrapods, uniting them in the paraphyletic group Crossopterygii along with a number of other fossil taxa, but the discovery of the extant coelacanth *Latimeria* revealed many untetrapod-like features (Forey 1988), casting doubt on how conclusive the morphological data really are (Janvier 1998). We find the coelacanths as closest relatives to the tetrapods, but bootstrap support below 50% shows that this node is essentially unresolved. Evidence from mitochondrial genome sequences has been ambiguous, depending on the phylogenetic method used (Zardoya & Meyer 1997) and often misplacing both lungfish and tetrapods completely (see figure 3*a*,*b*).

Finally, we have an unusual result for the phylogeny of the reptiles (taken to include the birds). The bulk of morphological and palaeontological evidence groups alligators and birds with the extinct dinosaurs as the archosauria, with lizards forming the sister group to this clade and turtles most basal. This has been challenged by data placing turtles as the sister group to the lepidosaurs (Rieppel & deBraga 1996) and molecular data, which seem unanimously to place turtles as relatives of archosaurs (Hedges & Poling 1999; Rieppel 2000). A number of recent reviews (Rieppel 2000; Zardoya & Meyer 2001*a*) have concluded that relationships within the reptiles are still uncertain. The results of our analysis are unconventional in placing turtles as the closest relative of birds, but add to the molecular evidence placing turtles within crown-group diapsids.

4. DISCUSSION

The gene tree parsimony method makes a number of assumptions about the process of gene duplication that may be important in this context. First, the correct inference of gene duplications and losses on a gene tree requires that the gene tree be known without error. This is a potentially important problem that has been widely recognized (Page 2000; Page & Cotton 2000) which we have dealt with by using a bootstrap profile of trees for each gene family.

We also make some assumptions about the process of gene duplication, as the number of duplications and losses is assumed to be the minimum required to fit the gene tree into the species tree. If duplications and losses are frequent, there may be lineages that originated in a duplication event and were then lost, leaving no trace in extant genomes. These numbers could thus be a significant underestimate of the true number of duplication and loss events, but should not introduce any systematic bias in the optimal species tree.

Another important issue is that failure to sample (where a gene has simply not been sequenced from an organism) is conflated with gene loss (where the gene is actually deleted from the genome). This has no effect on the optimal species tree under a duplication-only criterion, but could lead to artefacts under the duplication and loss criterion, where species can cluster on the basis of this failure to sample (Page & Charleston 1997*a*; Page 2000). We would advise against duplication and loss as an optimality criterion in data where this problem is likely to be very significant; although in fact the optimal species tree under the duplication and loss criterion for our data differs little from the minimum-duplications tree, placing *Latimeria* as sister taxon to an amphibian clade at the base of the tetrapods and grouping *Trachemys* with *Alligator* rather than *Gallus*.

Finally, our method assumes that gene duplication and gene loss are the only processes introducing disparity between gene and species trees. Gene duplications have clearly been important in vertebrates, as shown by the existence of many complex gene families in vertebrate genomes (Page 2000), but we cannot rule out that other processes might introduce incongruence between gene and species trees. The frequency with which genes will fail to coalesce between speciation events (deep coalescence) will depend on both the effective size of the population in which the alleles are present and the time between speciations. If we imagine the width of branches to be effective population size, long, thin branches should show few, if any, failures to coalesce, while short, fat branches should show many failures to coalesce (Pamilo & Nei 1988). We have no information about effective population sizes, but all of the branches on our phylogeny are very long in population genetics terms: molecular clock divergence dates indicate that the split between *Homo* and *Bos* is probably *ca.* 92 Myr ago, and that between birds and crocodilians *ca.* 222 Myr ago (Kumar & Hedges 1998). There are very few reliable reports of horizontal gene transfer in eukaryotes (Syvanen 1994), so we can rule out any large-scale effect from horizontal transfer in our dataset.

Any study attempting to infer species phylogenies from gene phylogenies of multiple loci needs to take into account the potential problem of paralogy. As large-scale sequencing projects produce genomic sequence data from an increasing number of taxa, we believe that the issues discussed in this paper will become of increasing importance to systematists and that reconciled tree methods will become more widely used. Gene tree parsimony is fast enough to scale-up to analysis of whole genomes and even whole genetic databases, raising the possibility of effective automated phylogenetic reconstructions from molecular data (Page & Cotton 2000).

5. CONCLUSION

We have shown that reconciled trees can successfully reconstruct phylogeny in the presence of a mixture of orthologous and paralogous genes. In contrast to evidence from mitochondrial sequences, our results largely agree with traditional views on vertebrate phylogeny, but add new evidence to support some controversial ideas, such as a monophyletic cyclostome group. The techniques described in this paper should scale-up to genome-scale comparisons, so we hope that this success will encourage systematists struggling to reconstruct credible phylogenies from the vast amounts of genomic data that is now accumulating (Brown 1996).

This work was supported by the NERC, the Wolfson Foundation and the EMBO. We thank Mike Charleston, Trevor Cotton, Martyn Kennedy, Gavin Naylor and the late Joe Slowinski for constructive comments on various versions of this manuscript.

REFERENCES

Bishop, M. J. & Friday, A. E. 1988 Estimating the interrelationships of tetrapod groups on the basis of molecular sequence data. In *The phylogeny and classification of the tetrapods*, vol. 1 (ed. M. J. Benton), pp. 33–58. Oxford: Clarendon.

Brown, J. R. 1996 Preparing for the flood: evolutionary biology in the age of genomics. *Trends Ecol. Evol.* **11**, 510–513.

Charleston, M. A. 1998 Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Math. Biosci.* **149**, 191–223.

Curole, A. P. & Kocher, T. D. 1999 Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol. Evol.* **14**, 394–398.

Delarbre, C., Escriva, H., Gallut, C., Barriel, V., Kourilsky, P., Janvier, P., Laudet, V. & Gachelin, G. 2000 The complete nucleotide sequence of the mitochondrial DNA of the agnathan *Lampetra fluviatilis*: bearings on the phylogeny of cyclostomes. *Mol. Biol. Evol.* **17**, 519–529.

Delarbre, C., Gallut, C., Barriel, V., Janvier, P. & Gachelin, G. 2002 Complete mitochondrial DNA of the hagfish, *Eptatretus burgeri*: the comparative analysis of mitochondrial DNA sequences strongly supports the cyclostome monophyly. *Mol. Phylogenet. Evol.* **22**, 184–192.

Doyle, J. J. 1992 Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* **17**, 144–163.

Duret, L., Mouchiroud, D. & Gouy, M. 1994 HOVERGEN: a database of homologous vertebrate genes. *Nucleic Acids Res.* **22**, 2360–2365.

Felsenstein, J. 1985 Confidence-limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Forey, P. L. & Janvier, P. 1993 Agnathans and the origin of jawed vertebrates. *Nature* **361**, 129–134.

Forey, P. L. 1988 Golden jubilee for the coelacanth *Latimeria chalumnae*. *Nature* **336**, 727–732.

Gursoy, H. C., Koper, D. & Benecke, B. J. 2000 The vertebrate 7S K RNA separates hagfish (*Myxine glutinosa*) and lamprey (*Lampetra fluviatilis*). *J. Mol. Evol.* **50**, 456–464.

Hedges, S. B. & Poling, L. L. 1999 A molecular phylogeny of reptiles. *Science* **283**, 998–1001.

Janvier, P. 1981 The phylogeny of the craniata, with particular reference to the significance of fossil 'agnathans'. *J. Vertebrate Palaeontol.* **1**, 121–159.

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

Janvier, P. 1998 A cold look at odd vertebrate phylogenies. *J. Mol. Evol.* **46**, 375–377.

Kumar, S. & Hedges, S. B. 1998 A molecular time-scale for vertebrate evolution. *Nature* **392**, 917–920.

Kuraku, S., Hoshiyama, D., Katoh, K., Suga, H. & Miyata, T. 1999 Monophyly of lampreys and hagfishes supported by nuclear DNA-coded genes. *J. Mol. Evol.* **49**, 729–735.

Liu, F.-G. R., Miyamoto, M., Freire, N. P., Ong, P. Q., Tennant, M. R., Young, T. S. & Gugel, K. F. 2001 Molecular and morphological supertrees for eutherian (placental) mammals. *Science* **291**, 1786–1789.

Løvtrup, A. 1977 *The phylogeny of the vertebrata*. New York: Wiley.

Maddison, W. P. 1997 Gene trees in species trees. *Syst. Biol.* **46**, 523–536.

Maisey, J. G. 1984 Higher elasmobranch phylogeny and biostratigraphy. *Zool. J. Linn. Soc.* **82**, 33–54.

Mallatt, J. 1997 Hagfish do not resemble ancestral vertebrates. *J. Morphol.* **232**, 293.

Mallatt, J. & Sullivan, J. 1998 28S and 18S rDNA sequences support the monophyly of lampreys and hagfish. *Mol. Biol. Evol.* **15**, 1706–1718.

Murphy, W. J. (and 10 others) 2001 Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351.

Naylor, G. J. & Brown, W. M. 1997 Structural biology and phylogenetic estimation. *Nature* **388**, 527–528.

Nelson, J. S. 1994 *Fishes of the world*. New York: Wiley.

Page, R. D. M. 1998 GENETREE: comparing gene and species phylogenies using reconciled trees. *Bioinformatics* **14**, 819–820.

Page, R. D. M. 2000 Extracting species trees from complex gene trees: reconciled trees and vertebrate phylogeny. *Mol. Phylogenet. Evol.* **14**, 89–106.

Page, R. D. M. & Charleston, M. A. 1997*a* From gene to organismal phylogeny: reconciled trees and the gene tree species tree problem. *Mol. Phylogenet. Evol.* **7**, 231–240.

Page, R. D. M. & Charleston, M. A. 1997*b* Reconciled trees and incongruent gene and species trees. In *Mathematical hierarchies in biology*, vol. 37 (ed. B. Mirkin, F. R. McMorris, F. S. Roberts & A. Rzhetsky), pp. 57–71. Providence, RI: American Mathematical Society.

Page, R. D. M. & Charleston, M. A. 1998 Trees within trees: phylogeny and historical associations. *Trends Ecol. Evol.* **13**, 356–359.

Page, R. D. M. & Cotton, J. A. 2000 GENETREE: a tool for exploring gene family evolution. In *Comparative genomics: empirical and analytical approaches to gene order dynamics, map alignment and the evolution of gene families* (ed. D. Sankoff & J. H. Nadeau), pp. 525–536. Dordrecht, The Netherlands: Kluwer.

Pamilo, P. & Nei, M. 1988 Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583.

Rasmussen, A. S. & Arnason, U. 1999 Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. *J. Mol. Evol.* **48**, 118–123.

Rasmussen, A. S., Janke, A. & Arnason, U. 1998 The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa*) and vertebrate phylogeny. *J. Mol. Evol.* **46**, 382–388.

Rieppel, O. 2000 Turtles as diapsid reptiles. *Zoologica Scripta* **29**, 199–212.

Rieppel, O. & deBraga, M. 1996 Turtles as diapsid reptiles. *Nature* **384**, 453–455.

Slowinski, J. B. & Page, R. D. M. 1999 How should phylogenies be inferred from sequence data? *Syst. Biol.* **48**, 814–825.

Stock, D. & Whitt, G. 1992 Evidence from 18S ribosomal RNA sequences that lampreys and hagfish form a natural group. *Science* **257**, 787–789.

Strimmer, K. & von Haeseler, A. 1996 Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969.

Suzuki, M., Kubokawa, K., Nagasawa, H. & Urano, A. 1995 Sequence analysis of vasotocin cDNAs of the lamprey *Lampetra japonica*, and the hagfish, *Eptratretus burgeri*: evolution of cyclostome vasotocin precursors. *J. Mol. Endocrinol.* **14**, 67–77.

Syvanen, M. 1994 Horizontal gene transfer: evidence and possible consequences. *A. Rev. Genet.* **28**, 237–261.

Takezaki, N. & Gojobori, T. 1999 Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. *Mol. Biol. Evol.* **16**, 590–601.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994 CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Waddell, P. J., Okada, N. & Hasegawa, M. 1999 Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48**, 1–5.

Zardoya, R. & Meyer, A. 1996 The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. *Genetics* **142**, 1249–1263.

Zardoya, R. & Meyer, A. 1997 The complete DNA sequence of the mitochondrial genome of a 'living fossil', the coelacanth (*Latimeria Chalumnae*). *Genetics* **146**, 995–1010.

Zardoya, R. & Meyer, A. 2001*a* The evolutionary position of turtles revised. *Naturwissenschaften* **88**, 193–200.

Zardoya, R. & Meyer, A. 2001*b* Vertebrate phylogeny: limits of inference of mitochondrial genome and nuclear rDNA sequence data due to an adverse phylogenetic signal/noise ratio. In *Major events in early vertebrate evolution* (ed. P. E. Ahlberg), pp. 135–156. London: Taylor and Francis.

Zhang, L. 1997 On a Mirchkin–Muchnik–Smith conjecture for comparing molecular phylogenies. *J. Comput. Biol.* **4**, 177–187.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.