Chapter 9

Vertebrate phylogeny: limits of inference of mitochondrial genome and nuclear rDNA sequence data due to an adverse phylogenetic signal/noise ratio

Rafael Zardoya and Axel Meyer

ABSTRACT

The phylogenetic relationships among the main lineages of vertebrates (mammals, birds, reptiles, amphibians, and fishes) were analysed using nuclear 28S rDNA, and mitochondrial (combined tRNA genes and concatenated protein-coding genes) sequence data. The comparatively slowly-evolving nuclear 28S rRNA gene was able to recover a vertebrate phylogeny which is in agreement with palaeontological and morphological evidence. Cartilaginous fishes were placed basal to a clade including bony fishes and tetrapods with a high bootstrap support. Lobe-finned fishes showed an unusually high rate of evolution for the 28S rRNA gene. The mitochondrial tRNA data set showed an extensive among-site rate variation, and a limited number of sites containing phylogenetic signal, unable to resolve the short nodes on the base of the vertebrate tree. As a result, the recovered tRNA tree, although congruent with the morphology-based vertebrate phylogeny, remained largely unresolved. The phylogenetic analyses of the protein data set at the amino acid level using hagfish and lamprey as outgroups arrived at rather unorthodox topologies in which bizarre vertebrate groupings were found such as e.g. snake + hagfish, amphibians + bony fishes, teleosts + cartilaginous fishes. The biologically incorrect phylogenetic estimates were identified to be artefacts stemming from non-random misleading noise in the protein data set. The adverse phylogenetic signal/noise ratio of the protein mitochondrial data set was likely due to several causes including saturation, heterogeneous rates of evolution among different vertebrate lineages, among-site rate variation, and the selection of distant taxa as outgroups.

9.1 Introduction

Two sets of phylogenetic markers, mitochondrial DNA (mtDNA) and nuclear rRNA genes, are the most widely used in molecular systematics. Generally, mtDNA is considered a rapidly evolving molecule (Brown *et al.* 1979) and, in the past, it was mainly used to infer phylogenetic relationships among closely related species (but see Meyer and Wilson 1990). Similarly, nuclear rRNA genes, because of their slow on average evolutionary rates, are usually used in phylogenetic studies among distantly related taxa (Sogin 1989; Hillis and Dixon 1991). However, the advent of the

polymerase chain reaction (PCR, Saiki *et al.* 1988), the advances of sequencing techniques (Kocher *et al.* 1989), and the sophistication of methods of phylogenetic reconstruction (Swofford *et al.* 1996) have significantly extended the phylogenetic scope of the application of these two molecular markers, which now are widely used to infer phylogenies at any level of divergence.

In particular, phylogenetic analyses of mitochondrial genome and nuclear rRNA sequence data have been incorporated into the study of vertebrate evolution with great success (Russo *et al.* 1996; Zardoya and Meyer 1996b; Zardoya and Meyer 1996c; Cao 1998; Naylor and Brown 1998). Molecular studies have largely corroborated the traditional morphology-based phylogeny of vertebrates (Figure 9.1) that was firmly established based on the analyses of the comparatively complete fossil record of vertebrates (e.g. Romer 1966; Carroll 1988; Benton 1990; Cloutier and Ahlberg 1996; Carroll 1997). In the traditional phylogeny of vertebrates, the agnathans (lampreys and hagfishes) are basal to gnathostomes (jawed vertebrates) (Figure 9.1). Within the latter, cartilaginous fishes are basal to a clade including bony fishes and tetrapods (Figure 9.1). Among bony fishes, the ray-finned fishes (Actinopterygii) are the most basal clade, and lobe-finned fishes (Sarcopterygii) are the sister group of tetrapods (Figure 9.1). The amphibians, which are the most basal tetrapods, are the sister group of the amniotes i.e. reptiles + birds and mammals (Figure 9.1). Molecular-based analyses have contributed particularly to studying further some of the remaining puzzles in vertebrate phylogeny, e.g. the relative phylogenetic positions of lobe-finned fishes (reviewed in Zardoya *et al.* 1998), whales (Milinkovitch *et al.* 1993), and monotremes (Janke *et al.* 1996).

Surprisingly, however, in some recent molecular studies based on mitochondrial sequence data, highly unorthodox hypotheses of phylogenetic relationships among the major lineages of vertebrates were supported when highly divergent taxa such as lamprey, hagfish, or echinoderms were used as outgroups (Russo *et al.* 1996; Zardoya and Meyer 1996a; Cao 1998; Naylor and Brown 1998; Zardoya *et al.* 1998; Rasmussen and Arnason 1999b; Takezaki and Gojobori 1999). For example, depending on which method and data set was used, the lungfish was placed in, at least, five different positions, (e.g. basal to the rest of the taxa, basal to a group including the frog, the bichir, the coelacanth, and teleosts, etc.) (see fig. 1 in Zardoya *et al.* 1998), however, none of which were as a lobe-finned fish in the expected place. Likewise, the presumed phylogenetic position of the frog as the sister group of amniotes was hardly ever recovered correctly (e.g. Naylor and Brown 1998; Zardoya *et al.* 1998; Takezaki and Gojobori 1999). Moreover, sharks were typically misplaced as the sister group of teleosts (Rasmussen and Arnason 1999b). Interestingly, in many cases, these incorrect groupings were nonetheless supported by high bootstrap values (e.g. Naylor and Brown 1998; Zardoya *et al.* 1998; Takezaki and Gojobori 1999).

Two different sets of explanations are possible for such odd results. Either these trees reflected the 'true' phylogenetic relationships among vertebrates (Rasmussen *et al.* 1998; Rasmussen and Arnason 1999b) or noise in the data set rather than the phylogenetic signal was responsible for these unexpected groupings (Naylor and Brown 1997; Cao 1998; Naylor and Brown 1998; Zardoya *et al.* 1998; Takezaki and Gojobori 1999). Here, we present new analyses of molecular data that successfully recover the traditional phylogeny of vertebrates (Figure 9.1). Hence, we reject

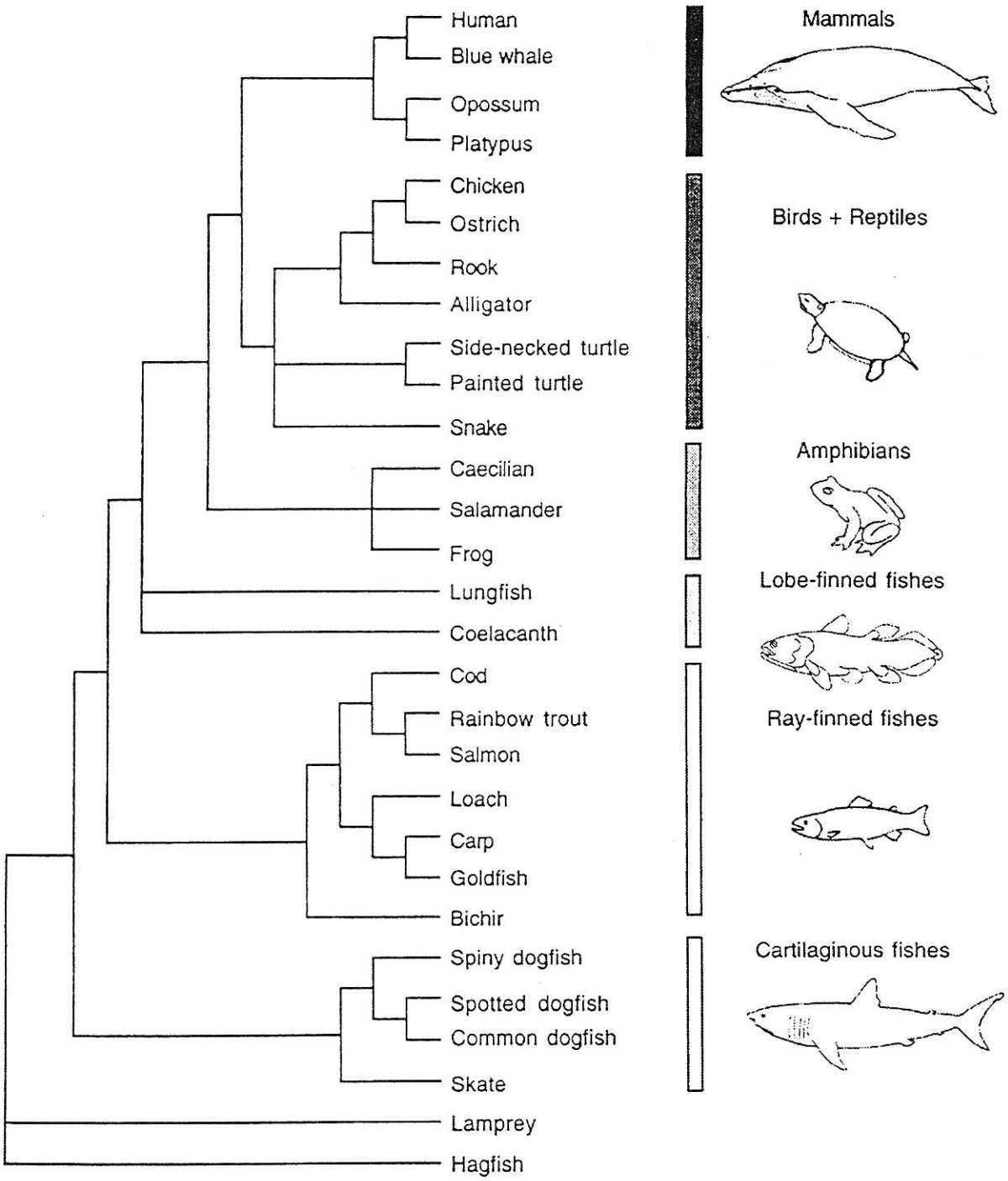

Figure 9.1 Phylogenetic relationships of the main lineages of vertebrates based on morphological and palaeontological evidence (e.g. Carroll 1988; Cloutier and Ahlberg 1997).

the hypothesis that the unorthodox results (Rasmussen *et al.* 1998; Rasmussen and Arnason 1999b) reflect the true phylogenetic relationships among vertebrates. We further investigated the effect of among-site rate variation in the mitochondrial and nuclear rRNA gene data sets, and characterized the phylogenetic utility and limits of resolution of these molecular markers, to determine the reasons underlying the recovery of biologically nonsensical results in some of the analyses.

9.2 Material and methods

9.2.1 Sequence data

To recover the phylogenetic relationships among the main lineages of vertebrates we analysed nuclear and mitochondrial sequence data. The nuclear data set comprises the following 19 nearly complete nuclear 28S rRNA nucleotide sequences: tunicate, *Herdmania momus* (X53538; Degnan *et al.* 1991); lancelet, *Branchiostoma floridae* (AF061796; Mallat and Sullivan 1998); hagfish, *Eptatretus stouti* (AF061796; Mallat and Sullivan 1998); lamprey, *Petromyzon marinus* (AF061797; Mallat and Sullivan 1998); chimaera, *Hydrolagus colliei* (AF061799; Mallat and Sullivan 1998); shark, *Squalus acanthias* (AF061800; Mallat and Sullivan 1998); bichir, *Polypterus ornatipinnis* (AF154052; this paper); sturgeon, *Acipenser brevirostrum* (U34340; Zardoya and Meyer 1996b); eel, *Anguilla rostrata* (U34342; Zardoya and Meyer 1996b); rainbow trout, *Oncorhynchus mykiss* (U34341; Zardoya and Meyer 1996b); coelacanth, *Latimeria chalumnae* (U34336; Zardoya and Meyer 1996b); Australian lungfish, *Neoceratodus forsteri* (U34338; Zardoya and Meyer 1996b); African lungfish, *Protopterus aethiopicus* (U34339; Zardoya and Meyer 1996b); South American lungfish, *Lepidosiren paradoxa* (U34337; Zardoya and Meyer 1996b); clawed frog, *Xenopus laevis* (X59734; Ajuh *et al.* 1991); Kenyan clawed frog, *Xenopus borealis* (X59733; Ajuh *et al.* 1991); rat, *Rattus norvegicus* (V01270; Hadjiolov *et al.* 1984); mouse, *Mus musculus* (X00525; Hassouna *et al.* 1984); human, *Homo sapiens* (U13369; Gonzalez *et al.* 1985).

The mitochondrial data set includes the following representative vertebrate mitochondrial genomes: hagfish, *Myxine glutinosa* (Y15180-Y15192; Rasmussen *et al.* 1998); lamprey, *Petromyzon marinus* (U11880; Lee and Kocher 1995); starry skate, *Raja radiata* (AF106038; Rasmussen and Arnason 1999b); common dogfish, *Scyliorhinus canicula* (Y16067; Delarbre *et al.* 1998); spotted dogfish, *Mustelus manazo* (AB015962; Cao *et al.* 1998); spiny dogfish, *Squalus acanthias* (Y18134; Rasmussen and Arnason 1999a); bichir, *Polypterus ornatipinnis* (U62532; Noack *et al.* 1996); cod, *Gadus morhua* (X99772; Johansen and Bakke 1996); salmon, *Salmo salar* (U12143; Hurst, Bartlett, Bruce, and Davidson, unpublished); rainbow trout, *Oncorhynchus mykiss* (L29771; Zardoya *et al.* 1995); carp, *Cyprinus carpio* (X61010; Chang *et al.* 1994); goldfish, *Carassius auratus* (AB006953; Murakami *et al.* 1998); loach, *Crossostoma lacustre* (M91245, Tzeng *et al.* 1992); African lungfish, *Protopterus dolloi* (L42813; Zardoya and Meyer 1996a); coelacanth, *Latimeria chalumnae* (U82228; Zardoya and Meyer 1997); clawed frog, *Xenopus laevis* (M10217; Roe *et al.* 1985); caecilian, *Typhlonectes natans* (AF154051; Zardoya and Meyer 2000); salamander, *Mertensiella luschani* (AF154053; Zardoya, Malaga-Trillo, Veith, García-Paris, and Meyer, in preparation), Akamata snake, *Dinodon semicaritanus* (AB008539; Kumazawa *et al.* 1998); side-necked turtle, *Pelomedusa subrufa* (AF039066; Zardoya and Meyer 1998); painted turtle, *Chrysemys picta* (AF069423; Mindell *et al.* 1999); alligator, *Alligator mississippiensis* (Y13113; Janke and Arnason 1997); chicken, *Gallus gallus* (X52392; Desjardins and Morais 1990); ostrich, *Struthio camelus* (Y12025; Härlid *et al.* 1997); rook, *Corvus frugilegus* (Y18522; Härlid and Arnason 1999); platypus, *Ornithorhynchus anatinus* (X83427; Janke *et al.* 1996); opossum, *Didelphis virginiana* (Z29573; Janke *et al.* 1994); blue whale, *Balaenoptera musculus* (X72204; Arnason and Gullberg 1993); Human, *Homo sapiens* (D38112; Horai *et al.* 1995).

9.2.2 *Phylogenetic analyses*

Homologous sequences were aligned using CLUSTAL W (Thompson *et al.* 1994) followed by refinement by eye. Gaps resulting from the alignment were treated as missing data. Ambiguous alignments were excluded from the phylogenetic analyses (aligned sequences and exclusion sets are available at http://www.mncn.csic.es/investigacion/bbe/zardoya.htm).

Three distinct sequence data sets were analysed separately:

1 28S rRNA gene
2 All 22 tRNA gene sequences combined, and
3 All protein-coding genes combined (except ND6 because it is encoded by the L-strand, and thus, has a very different base composition) at the amino acid level.

We did not use mitochondrial rRNA sequences to recover phylogenetic relationships among vertebrates because they have proven to lack, due to extensive among-site rate variation, enough sites that contain phylogenetic signal at this level of divergence (Zardoya and Meyer 2000). Moreover, at this level of divergence, the alignment of mitochondrial rRNA sequences turns out to be highly subjective due to ambiguity in the fast evolving portions of the molecule. Each data set was subjected to maximum parsimony (MP), neighbour-joining (NJ), and maximum likelihood (ML) phylogenetic analyses. MP analyses were conducted with PAUP* version d65 (Swofford 1997), using heuristic searches (TBR branch swapping; MULPARS option in effect), and 10 random stepwise additions of taxa. Unless specified, transitions and transversions were given equal weight. NJ (Saitou and Nei 1987) analyses were based on HKY85 (Hasegawa *et al.* 1985) and LogDet (Lockhart *et al.* 1994) distance matrices (PAUP* version d65; Swofford 1997). ML analyses were performed with PAUP* version d65 (HKY 85 model; Hasegawa *et al.* 1985), and PUZZLE version 4.0.1 (Strimmer and von Haeseler 1996). In the DNA ML analyses, transition/transversion ratios were optimized to maximize the likelihood, and empirical base frequencies were used. In the protein ML analyses, the ML tree was inferred with the mtREV model (Adachi and Hasegawa 1996), using PUZZLE version 4.0.1 (Strimmer and von Haeseler 1996).

Robustness of the phylogenetic results was tested by bootstrap analyses (Felsenstein 1985) (as implemented in PAUP* version d65) with 100 pseudo-replications each, and quartet puzzling (QP, as implemented in PUZZLE version 4.0.1 (Strimmer and von Haeseler 1996) with 1000 puzzling steps.

9.2.3 *Among-site rate variation and statistical analyses*

The number of nucleotide substitutions or the consistency index at each site of the different sequence data sets were calculated using the CHART STATE CHANGES AND STASIS option in MacClade (Maddison and Maddison 1992). Parameters were estimated from the traditional morphology-based tree (Figure 9.1) using non-overlapping 20-bp windows, and the maximum parsimony method with 1000 random resolutions of the polytomies contained within the tree.

Statistical support of the different mitochondrial protein subsets (constructed

based on their CI; see results) for the traditional vertebrate tree (Figure 9.1) versus the MP tree recovered by the protein data set (Figure 9.9), was assessed by calculating the standard deviation of the difference in number of steps between both alternative trees using a two-tailed Wilcoxon signed-ranks test (Templeton 1983). If the difference in number of steps between two competing phylogenetic hypotheses were more than 1.96 times the standard deviations then the two phylogenies were declared significantly different at the $p < 0.05$ level. Statistical tests were peformed in PAUP* version d65 (Swofford 1997).

9.3 Results

9.3.1 Vertebrate phylogeny based on the 28S rRNA gene

A total of 5462 positions were aligned, of which 2730 were gapped positions that were excluded due to ambiguity. Of the remaining sites, 1648 were constant sites, and 493 were phylogenetically informative sites using the parsimony criterion. The 28S data set showed a high among-site rate variation (a = 0.36; Yang and Kumar 1996) which could interfere in the phylogenetic reconstruction (Figure 9.2). An

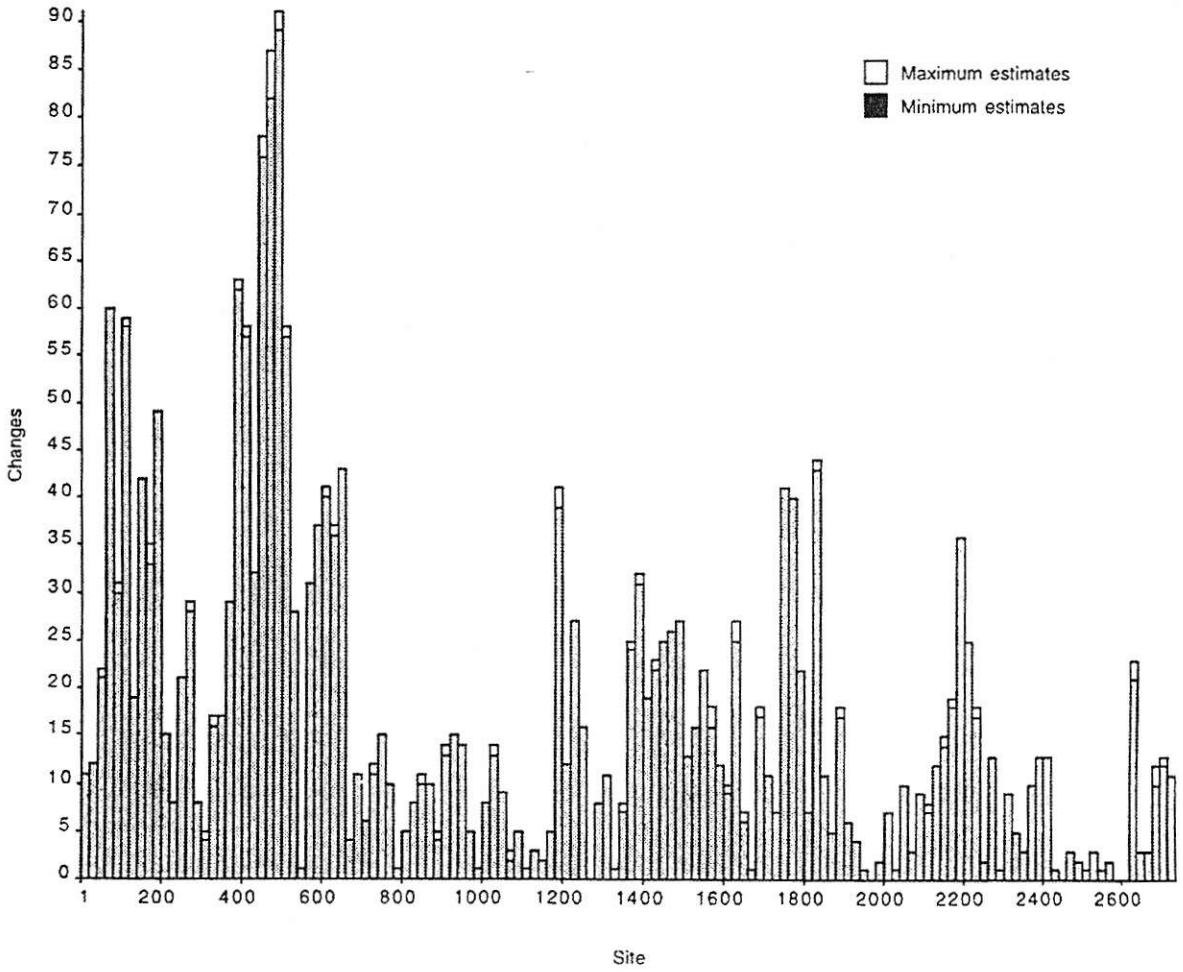

Figure 9.2 The inferred number of nucleotide substitutions per site over the entire alignment of the 28S rRNA data set. Maximum (open bars) and minimum (filled bars) estimates of the number of nucleotide changes were calculated for non-overlapping 20-bp windows using maximum parsimony, and the CHART STATE CHANGES AND STASIS option in MacClade (Maddison and Maddison 1992).

overall Ts/Tv ratio of 1.39 was estimated for this data set. Uncorrected *p* distances between taxa varied from 0.02 to 0.2. In this range, substitutions (both transitions and transversions) increased proportionally to sequence divergence, and no obvious saturation effects were detected (Figure 9.3).

Phylogenetic analyses of the 28S rRNA gene sequence data set with MP, NJ, and ML phylogenetic methods of inference, recovered trees with identical branching patterns, using the tunicate and the lancelet as outgroups (Figure 9.4). In these trees, the living jawless vertebrates (lamprey and hagfish) are the sister group to gnathostomes (jawed vertebrates). Within the latter, cartilaginous fishes (chimaera and shark) are basal to a clade including bony fishes and tetrapods (Figure 9.4). This phylogenetic relationship is supported by high bootstrap values (MP, 100 per cent; NJ, 100 per cent; ML, 99 per cent). The position of the bichir as the most basal of the ray-finned fishes and that of the frogs within tetrapods are supported by the phylogenetic analyses of the 28S data set (Figure 9.4a). Interestingly, the resolution of the phylogenetic relationships of bony fishes and tetrapods is dependent on the inclusion of lobe-finned fishes (coelacanth and lungfishes) in the analyses (Figure 9.4b). These phylogenetic relationships are fully resolved in the absence of lobe-finned fishes (Figure 9.4a), but become unstable (bootstrap values below 50 per cent) when these taxa are included (Figure 9.4b).

To further understand the effect of the inclusion of the lobe-finned fish 28S rRNA gene sequences, rate variation among jawed vertebrate lineages was estimated by calculating the genetic distances from their most recent common ancestor (MRCA) to the tip of each branch (Figure 9.5) (Farias *et al.* 1999). Lungfishes and the bichir showed significantly higher rates of evolution than cartilaginous fishes, teleosts, the coelacanth, and tetrapods (Figure 9.5).

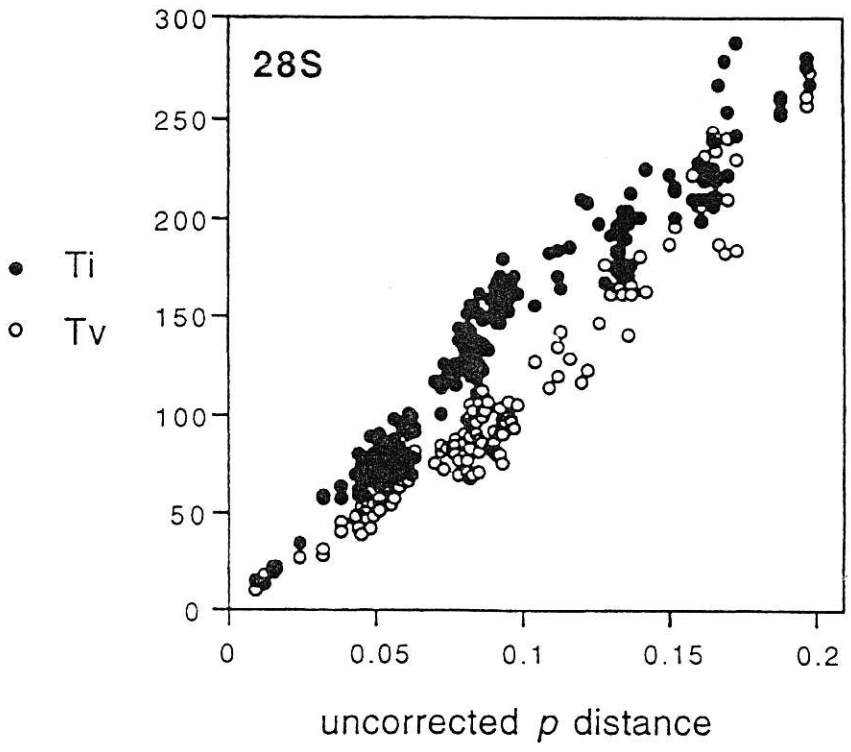

Figure 9.3 Scatter plot of transitions (filled circles) and transversions (open circles) over uncorrected *p* distances for the 28S rRNA gene data set.

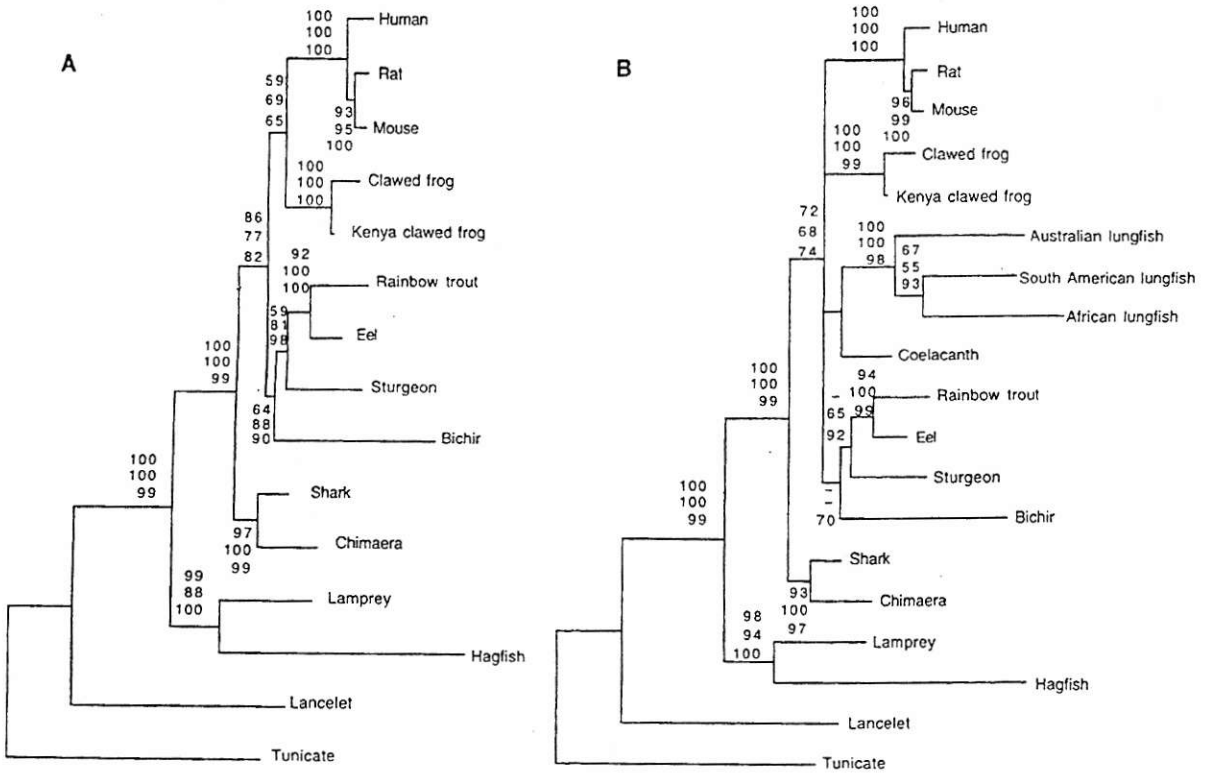

Figure 9.4 Phylogenetic relationships among the main lineages of vertebrates based on the nuclear 28S rRNA gene using a tunicate as outgroup. 50 per cent majority-rule consensus bootstrap trees obtained (a) excluding, and (b) including sarcopterygians in the phylogenetic analyses. Numbers indicate bootstrap values based on 100 pseudo-replications. The 28S rRNA sequence data set was subjected to MP (bootstrap values upper of each triplet of numbers), NJ (bootstrap values in the middle of each triplet of numbers), and ML (bootstrap values lower of each triplet of numbers) analyses.

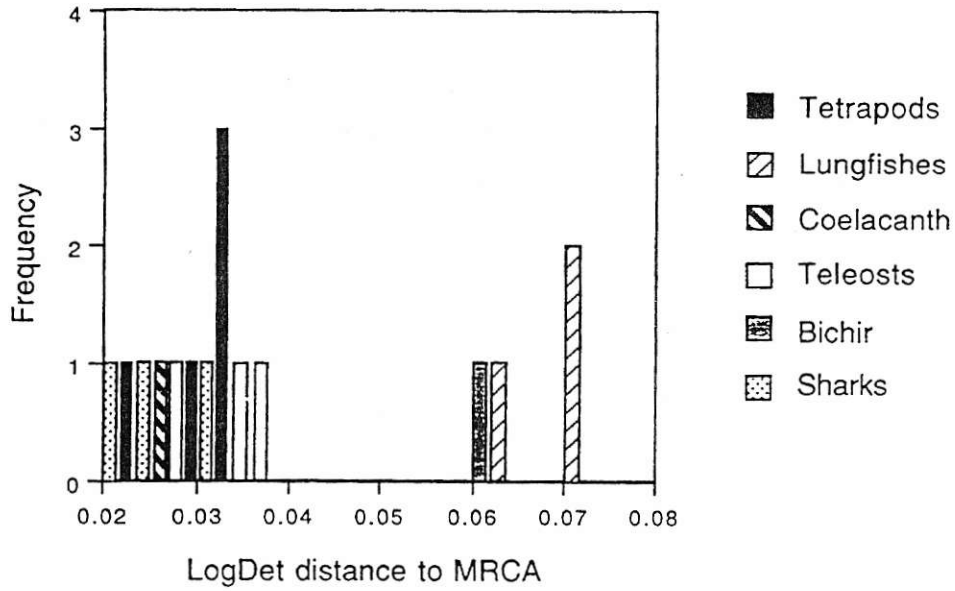

Figure 9.5 Frequency distribution of LogDet distances among the main lineages of jawed vertebrates. Path lengths from the most recent common ancestor (MRCA) to sharks, the bichir, teleosts, the coelacanth, the lungfishes and tetrapods were calculated across the NJ tree recovered from the 28S rRNA gene data set. A faster evolutionary rate is observed for the lungfishes and the bichir.

9.3.2 Phylogenetic analyses of the vertebrate mitochondrial tRNA data set

The nucleotide sequences of the 22 tRNAs encoded by the mitochondrial genome were combined, and aligned for several representative vertebrate taxa. A final data set of 1635 positions was assembled, of which 469 were excluded because of ambiguity. Of the remaining sites, 22 per cent were invariant, and 737 were parsimony-informative. An overall Ts/Tv ratio of 2.74 was estimated for this data set. Sequence divergence between taxa varied from 2 per cent to 48 per cent, and no saturation was observed for transitions and transversions (Figure 9.6). MP (using a 3:1 Tv:Ti weighting scheme), NJ (with HKY85 distances), and ML (with the HKY85 model) analyses with the lamprey as outgroup, arrived at congruent, but largely unresolved, trees (most of the nodes in the 50 per cent majority-rule bootstrap trees are collapsed) (Figure 9.7). As expected, the branches that connect nodes which were collapsed due to low bootstrap support, are extremely short (0.43–1.3 per cent HKY85 distances; not shown).

9.3.3 Phylogenetic signal and noise in the vertebrate mitochondrial protein data set

The deduced amino acid sequences of the 12 mitochondrial protein-coding genes of 29 vertebrate taxa were combined into a single alignment of 3694 positions. A total of 1158 positions were excluded from the phylogenetic analyses due to ambiguity in the alignment. Of the remaining, 949 sites were constant (i.e. 37 per cent) and 1079 were informative under the parsimony criterion. The average uncorrected p distance for the ingroup data set was 0.20 ± 0.06. The mean uncorrected p distance for the

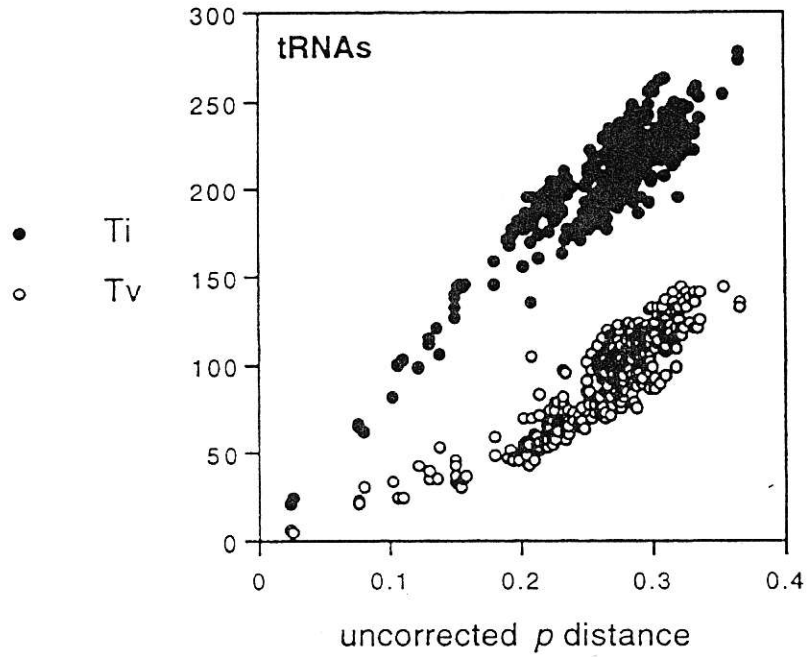

Figure 9.6 Scatter plot of transitions (open circles) and transversions (filled circles) over uncorrected p distances for the mitochondrial tRNA data set.

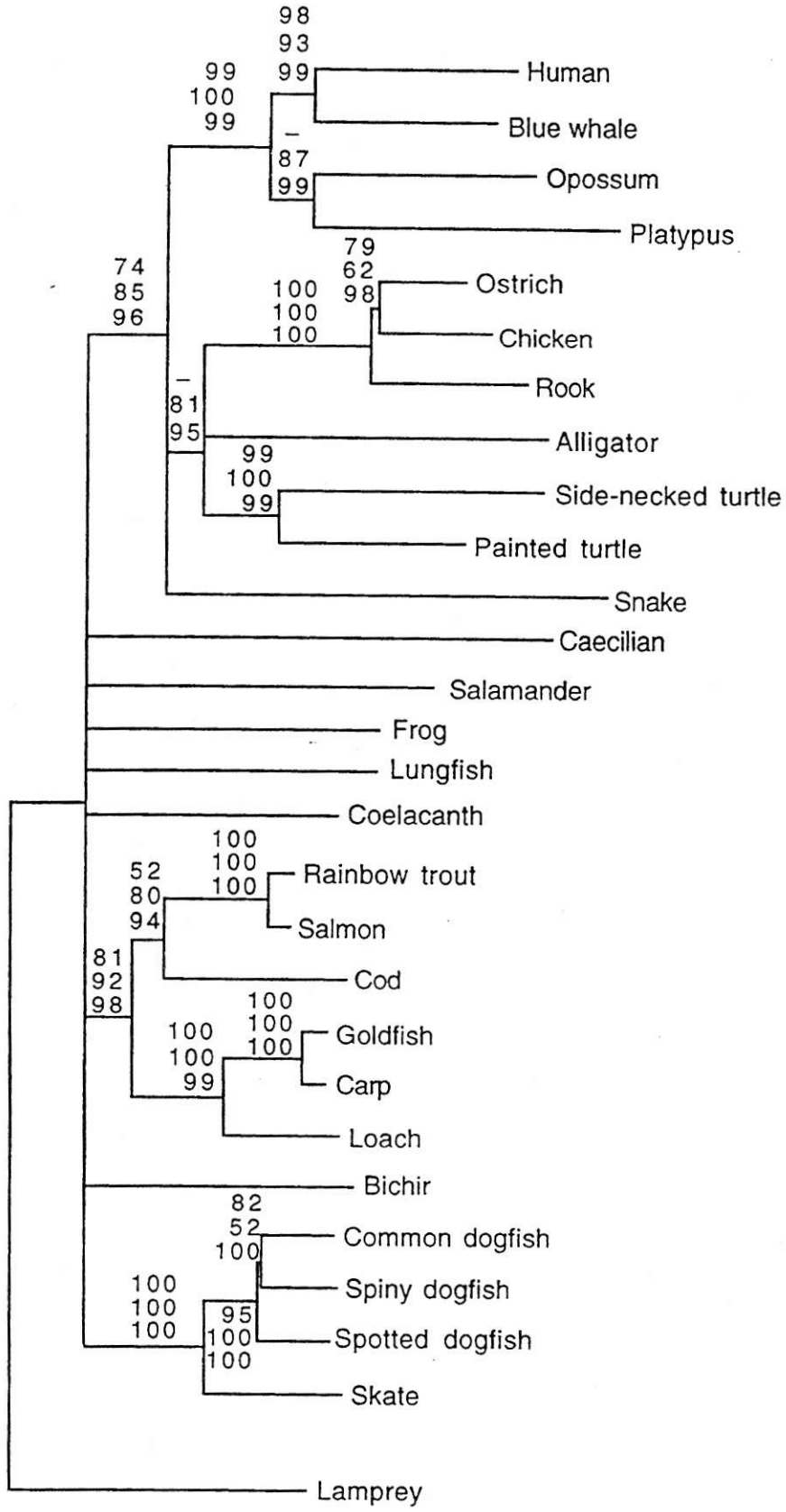

Figure 9.7 50 per cent majority-rule consensus bootstrap trees based on the mitochondrial tRNA data set using the lamprey as outgroup. Numbers indicate bootstrap values based on 100 pseudo-replications. The mitochondrial tRNA sequence data set was subjected to MP (bootstrap values upper of each triplet of numbers), NJ (bootstrap values in the middle of each triplet of numbers), and ML (bootstrap values lower of each triplet of numbers) analyses.

lamprey and hagfish amino acid sequences was 0.29 ± 0.04. To detect putative saturation processess in the amino acid sequence data, we plotted uncorrected *p* distances between pairs of vertebrate taxa over mtREV (Adachi and Hasegawa 1996) distances (this method was inspired by Philippe and Adoutte 1998). The relationship between both distances demonstrates that there is a certain level of saturation in the mitochondrial aminoacid data set (Figure 9.8). The effect of saturation is particularly strong for the outgroups (lamprey and hagfish) (Figure 9.8).

The vertebrate mitochondrial protein data set was analysed with MP, NJ, and ML using hagfish and lamprey as outgroup taxa (more basal taxa such as, for example, sea urchins were not included in the analyses because they have been shown to have significatively different amino acid composition; Takezaki and Gojobori 1999). Three different, but largely congruent, trees were recovered (Figure 9.9). Interestingly, the inferred trees show groupings that are inconsistent with the traditional morphology-based tree (Figure 9.1). For instance, the amphibians are placed as the sister group of fishes, the sharks group with teleosts, and the snake clusters with the hagfish at the base of all vertebrates (Figure 9.9).

In trying to better understand the unsatisfactory phylogenetic performance of the mitochondrial protein data set, the among-site rate variation along the 12 mitochondrial proteins was examined. The inferred number of amino acid changes per site were calculated for non-overlapping 20-bp windows (Figure 9.10), and we detected a considerable among-site rate variation which could potentially mislead the phylogenetic inference (Figure 9.10). Furthermore, those sites which evolved more rapidly

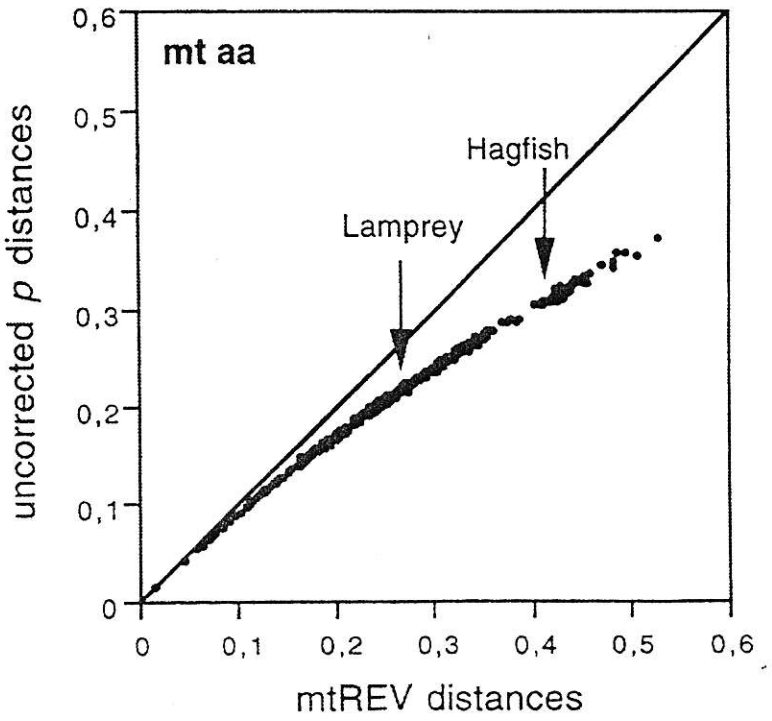

Figure 9.8 Scatter plot of uncorrected *p* distances for the mitochondrial protein data set at the amino acid level over distances determined for the same data set using the mtREV model. The resulting curve departs from the diagonal line (no saturation) indicating some level of saturation in the mitochondrial protein data set. This effect is particularly evident for hagfish and lamprey pairwise distances (arrows indicate the minimum pairwise distances for these two taxa).

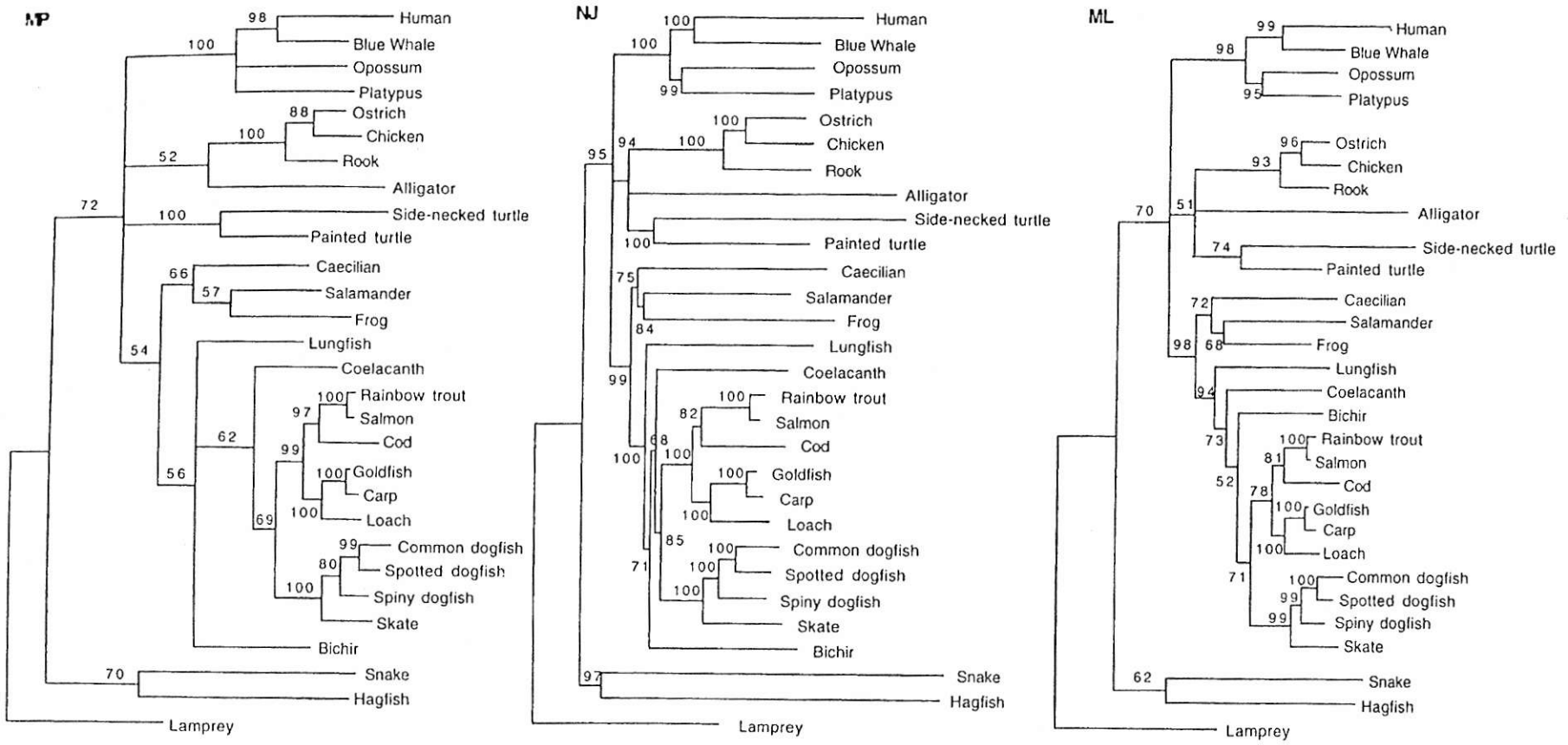

Figure 9.9 50 per cent majority-rule consensus bootstrap trees based on the mitochondrial protein data set using lamprey and hagfish as outgroup. MP, NJ, and ML phylogenetic analyses are shown. Numbers indicate bootstrap values based on 100 pseudo-replications. Biologically incorrect relationships such as, e.g. snake + hagfish, frog + bony fishes, are recovered regardless of the phylogenetic method of inference.

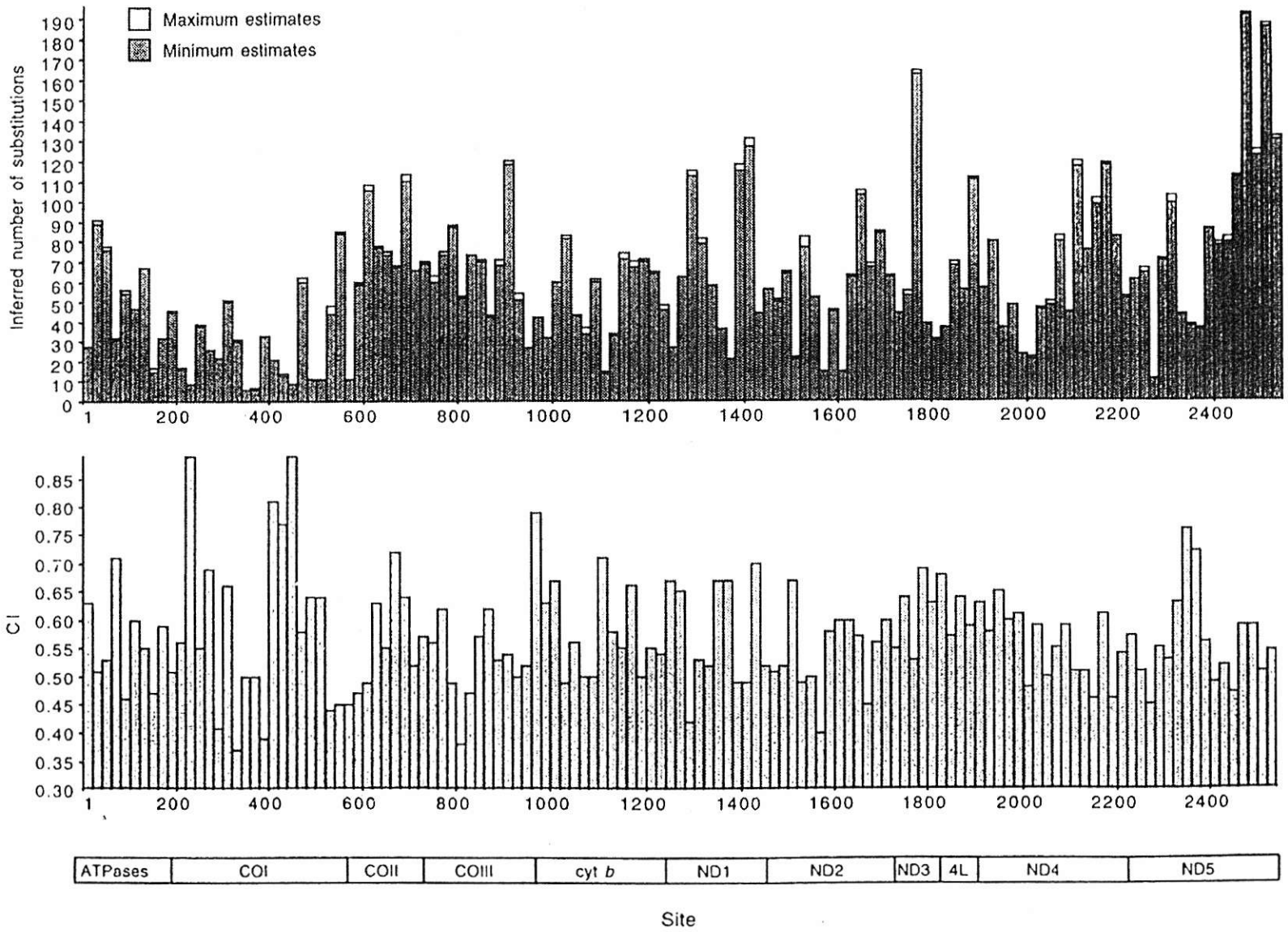

Figure 9.10 Maximum (open bars) and minimum (filled bars) inferred number of amino acid changes and the consistency index per site over the entire alignment of the mitochondrial proteins calculated for non-overlapping 20-bp windows using maximum parsimony, and the CHART STATE CHANGES AND STASIS option in MacClade (Maddison and Maddison 1992).

were shown to have a lower consistency index, and hence, to be more noisy (Figure 9.10). Taking this relationship into account, and to assess the effects of among-site-rate variation, up to seven data subsets of the mitochondrial protein data set were specified based on their different consistency indexes. The first subset included those positions with a CI > 0.75; the second subset covered positions with a CI > 0.70; the third, sites with a CI > 0.65, and so on (Figure 9.10). MP analyses were conducted based on each of these subsets, and the number of steps of the resulting MP trees for each subset were plotted against the number of positions of each subset (Figure 9.11) (Brinckmann and Philippe, 1999). The number of positions included in successive subsets increases steadily (Figure 9.11). However, the number of steps of the resulting MP trees shows a dramatic increase when positions with CI < 0.65 are included in the analyses (Figure 9.11), i.e. the addition of positions with CI < 0.65 seem to contribute a lot of noise rather than phylogenetic signal to the recovery of the MP trees. Moreover, a Templeton test (Templeton 1983) shows that the traditional morphology-based vertebrate tree (Figure 9.1) is not significantly different from the atypical MP tree (Figure 9.9) when positions with a CI > 0.55 are included in the analyses (Figure 9.11). A statistically significant support for the MP tree (Figure 9.9) is only achieved when the noisy positions with a CI < 0.55 are included in the analyses.

9.4 Discussion

Phylogenetic reconstruction based on molecular sequences can be incorrect if the effect of molecular evolutionary processes, such as, for example, saturation, heterogeneity of rates of substitution among lineages, and among-site rate variation within a molecule, is not taken into account (Takezaki and Gojobori 1999). The extent of such molecular evolutionary processes in shaping the data determines the divergence range in which molecular sequences are useful for phylogenetic inference (Naylor and Brown 1998); outside that divergence range, results may be strongly influenced by noise rather than be based on a robust phylogenetic signal.

In this work, we have explored the divergence range in which mtDNA and nuclear rRNA sequences provide reliable phylogenetic inferences of vertebrate phylogenetic relationships, as well as the causes underlying the limits and pitfalls of current methods of phylogenetic inference. Three sequence data sets i.e. the nuclear 28S rRNA gene, the combined mitochondrial tRNA gene, and the combined mitochondrial protein-coding gene data sets, have shown significantly different performance in recovering the traditional morphology-based vertebrate phylogeny (Figure 9.1). The 28S data set successfully recovers the traditional vertebrate tree in which cartilaginous fishes are the sister group of a clade including bony fishes and tetrapods (Figure 9.4). The mitochondrial tRNA data set is unable to recover vertebrate phylogenetic relationships, and renders a rather unresolved tree (Figure 9.7). The mitochondrial protein data set not only does not recover the traditional vertebrate phylogeny, but also supports with high bootstrap values a biologically erroneous topology (Figure 9.9) (Russo *et al.* 1996; Zardoya and Meyer 1996a; Cao 1998; Naylor and Brown 1998; Zardoya *et al.* 1998; Rasmussen and Arnason 1999b; Takezaki and Gojobori 1999).

To understand the phylogenetic behaviour of each of the molecular data sets, several analyses were conducted. Our results suggest that the different rates of

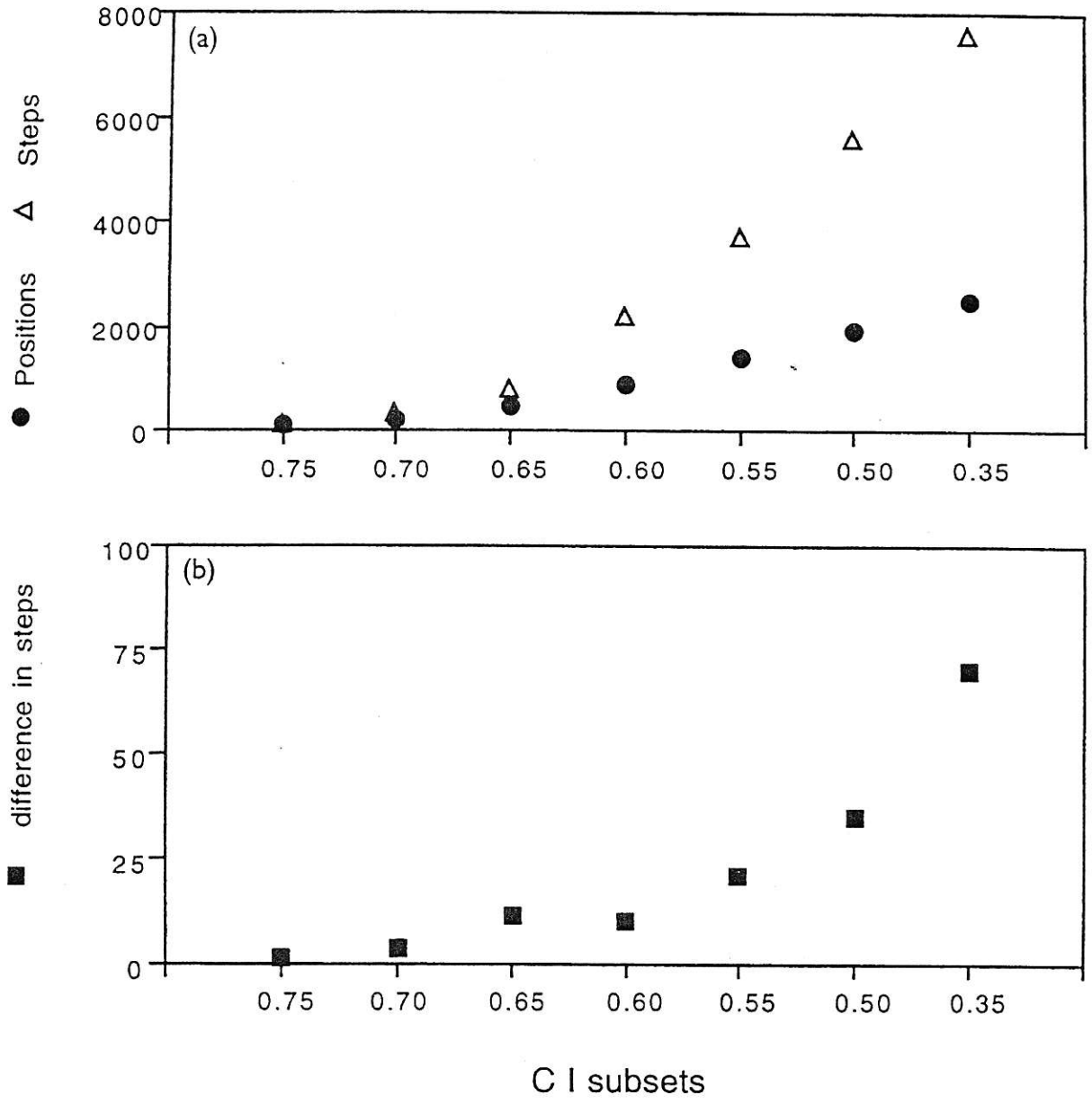

Figure 9.11 Phylogenetic signal and noise content of the mitochondrial protein data set. The mitochondrial protein data set was divided in subsets based on the consistency index (e.g. 0.75 includes positions with a CI of 0.75 or higher; 0.70 includes positions with a CI of 0.70 or higher, etc.). (a) For each subset, an MP analysis was performed, and the number of steps of the MP tree(s) was plotted along with the number of positions included in each phylogenetic analysis. The number of positions in each subset increases steadily, but the number of steps of the MP tree(s) shows important increases when positions with CI below 0.65 are included in the analyses. This indicates that positions with a CI below 0.65 add noise to the analysis rather than phylogenetic signal. (b) The difference in number of steps between the biologically incorrect MP tree (Figure 9.9) and the expected morphologically-based tree (Figure 9.1) for vertebrate relationships were calculated for each subset using a Wilcoxon signed-rank test. A statistically significative support for the biologically incorrect vertebrate tree is achieved only for subsets including positions with a CI below 0.5, i.e. those that add more noise to the phylogenetic analyses.

evolution of the three types of molecules is one of the causes affecting their phylogenetic performance. Both the nuclear 28S rRNA gene (an overall slowly-evolving gene; e.g. Mallat and Sullivan 1998) and the mitochondrial tRNA genes (the slowest in the mitochondrial genome; e.g. Kumazawa and Nishida 1993), show an adequate rate of substitution to study phylogenetic relationships among the main lineages of vertebrates (Figures 9.2 and 9.6). However, the mitochondrial protein-coding genes (relatively fast-evolving genes; e.g. Zardoya *et al.* 1998) appear to be saturated within the divergence range studied (Figure 9.8), and this particularly affects the outgroup taxa (hagfish and lamprey). In the mitochondrial protein data set, it is evident that too many substitutions have accumulated along the branch connecting the hagfish and the lamprey to vertebrate ingroup taxa (Figure 9.8) (Zardoya *et al.* 1998). This contributes to effectively randomize the sequences (Swofford *et al.* 1996), and significantly reduces the performance of the outgroups (Lyons-Weiler *et al.* 1998; Milinkovitch and Lyons-Weiler 1998).

The differing rates of evolution of the main vertebrate lineages also hinder the ability to reliably establish phylogenetic relationships, just as they hamper the performance of nuclear rRNA (Figure 9.4b) and mitochondrial (Figure 9.9; Zardoya and Meyer 1996c; Takezaki and Gojobori 1999) genes. In the case of the 28S data set, lobe-finned fishes show a significantly faster rate of variation which adversely affects the successful recovery of their expected phylogenetic position within vertebrates (Figure 9.5). In the mitochondrial protein data set, the restricted choice of living representatives of potential outgroups to the vertebrate ingroup (they are all too distantly related) randomizes ingroup relationships due to the well-known long-branch attraction effect (Felsenstein 1978; Maddison and Maddison 1992; Swofford *et al.* 1996). The use of the hagfish and lamprey as outgroup Taxa directly attracts to basal positions those lineages of vertebrates which are known to have long branches, i.e. frog, bichir, lungfish, and snake, (Roe *et al.* 1985; Noack *et al.* 1996; Zardoya and Meyer 1996a; Kumazawa *et al.* 1998). This attraction effect is exacerbated in this particular case because the internal branches which connect cartilaginous fishes, the bichir, teleosts, lobe-finned fishes, and amphibians are extremely short (Figure 9.9) (Cao 1998; Zardoya *et al.* 1998). The result is the tendency of taxa such as cartilaginous fishes and teleosts, which are known to have relatively slow rates of evolution (Martin *et al.* 1992; Cantatore *et al.* 1994) to presumably incorrectly cluster together (Figure 9.9) (Zardoya *et al.* 1998).

Another important phenomenon that adversely affects phylogenetic reconstruction is among-site rate variation within the molecule analysed (Takezaki and Gojobori 1999). Both the nuclear 28S rRNA (Figure 9.3) and the mitochondrial tRNA genes (e.g. Kumazawa and Nishida 1993) show extensive among-site rate variation. This rate heterogeneity significantly reduces the number of positions that contain phylogenetic signal at any level of divergence, and partially explains the polytomies in the 28S (Figure 9.2) and tRNA (Figure 9.7) trees. The protein data set also shows a considerable amount of among-site rate variation (Figure 9.10). Moreover, there is a clear negative correlation between the variability and the phylogenetic signal (as measured by the consistency index) at each site (Figure 9.10) (Zardoya and Meyer 1999). Our results demonstrate that the biologically unexpected phylogeny of vertebrates recovered by the mitochondrial protein data set is strongly supported by those positions that are more variable, and hence, show a lower consistency index as measured over the entire data set (Figure 9.11). Therefore, the assumption (Ras-

mussen *et al.* 1998; Rasmussen and Arnason 1999b) that high bootstrap values validate the results is probably incorrect because the strength of the phylogenetic signal in the data is overwhelmed by nonrandom noise that adds false confidence (Naylor and Brown 1998).

In conclusion, our results suggest that the range of utility of a molecular sequence is determined by the ratio between phylogenetic signals and noise at a given divergence level. The exact relationship depends on factors such as saturation, heterogeneity of substitution rates among different lineages, among-site rate variation, and the selection of distant outgroups which randomize phylogenetic signals at the base of the phylogeny and will have the effect of attracting the long-branched taxa within the ingroup. In the best scenario, an adverse ratio will prompt the recovery of unresolved topologies (e.g. as appears to be the case in the mitochondrial tRNA data set), but in the worst case, when misleading nonrandom noise accumulates in the data set, highly biased and biologically incorrect phylogenies will be recovered (e.g. as in the case of the mitochondrial protein data set: Russo *et al.* 1996; Zardoya and Meyer 1996a; Cao 1998; Naylor and Brown 1998; Rasmussen and Arnason 1999b; Zardoya *et al.* 1998; Takezaki and Gojobori 1999).

In numerous studies DNA sequences have been demonstrated to contain reliable phylogenetic information, and to be particularly useful in recovering phylogenies among taxa where high levels of morphological convergence or lack of phenotypic synapomorphies made the morphological approach problematic. In these cases, current available methods of phylogenetic inference are capable of recovering the putatively correct phylogeny based on a favourable phylogenetic signal/noise ratio. The flourishing of molecular systematics prompted studies that expanded the range of utility of widely used phylogenetic markers such as mtDNA or nuclear rRNAs into 'deeper' and 'shallower' zones, respectively. In these new zones of enquiry, however, the phylogenetic signal of the molecules is considerably reduced and often seemingly over-ridden by noise. In this context, contradictory results to well-known phylogenies are then suspect, and can be demonstrated to be caused by molecular biases rather than to reflect the correct phylogeny. Future efforts should concentrate on characterizing the limits of resolution of currently widely-used phylogenetic markers. This can be accomplished by elaborating more complex models of phylogenetic inference that are capable of maximizing the phylogenetic signal that better fits the actual pattern of evolution of sequences at deep levels of divergence (Cao 1998; Naylor and Brown 1998). Importantly, to resolve challenging phylogenetic questions, we will also have to search for new nuclear phylogenetic markers that have complementary rates of evolution to the widely-used mitochondrial markers (Takezaki and Gojobori 1999).

Acknowledgements

We thank Dave Mindell and Henner Brinckmann for insightful comments. Dave Swofford granted permission to publish results based on the test version (d65) of his most excellent PAUP* program. R.Z. was sponsored by a postdoctoral contract of the Ministerio de Educacion y Cultura of Spain. This work received partial financial support from grants from the Lion Foundation, the Deutsche Forschung Gemeinschaft, the University of Konstanz, and the US National Science Foundation (DEB-9615178) to A.M.

References

Adachi, J. and Hasegawa, M. (1996) 'Model of amino acid substitution in proteins encoded by mitochondrial DNA', *Journal of Molecular Evolution*, 42, 459–68.

Ajuh, P. M., Heeney, P. A. and Maden, B. E. H. (1991) '*Xenopus borealis* and *Xenopus laevis* 28S ribosomal DNA and the complete 40S ribosomal precursor RNA coding units of both species', *Proceedings of the Royal Society of London* B, 245, 65–71.

Arnason, U. and Gullberg, A. (1993) 'Comparison between the complete mtDNA sequences of the blue and fin whale, two species that can hybridize in nature', *Journal of Molecular Evolution*, 37, 312–22.

Benton, M. J. (1990) 'Phylogeny of the major tetrapod groups: morphological data and divergence dates', *Journal of Molecular Evolution*, 30, 409–24.

Brinckmann, H. and Philippe, H. (1999) 'Archaea sister group of bacteria? Indications for tree reconstruction artifacts in ancient phylogenies', *Molecular Biology and Evolution*, 16, 817–25.

Brown, W. M., George, M. J. and Wilson, A. C. (1979) 'Rapid evolution of animal mitochondrial DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 76, 1967–71.

Cantatore, P., Roberti, M., Pesole, G., Ludovico, A., Milella, F., Gadaleta, M. N. and Saccone, C. (1994) 'Evolutionary analyses of cytochrome b sequences in some perciformes: evidence for a slower rate of evolution than in mammals', *Journal of Molecular Evolution*, 39, 589–97.

Cao, Y. (1998) *Molecular Phylogeny and Evolution of Vertebrates*, PhD Thesis. Dept of Bioscience. Tokyo Institute of Technology, Tokyo, pp. 1–155.

Cao, Y., Waddell, P. J., Okada, N. and Hasegawa, M. (1998) 'The complete mitochondrial DNA sequence of the shark *Mustelus manazo*: evaluating rooting contradictions to living bony vertebrates', *Molecular Biology and Evolution*, 15, 1637–46.

Carroll, R. L. (1988) *Vertebrate paleontology and evolution*, New York: Freeman.

Carroll, R. L. (1997) *Patterns and processes of vertebrate evolution*, Cambridge: Cambridge University Press.

Chang, Y. S., Huang, F. L. and Lo, T. B. (1994) 'The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome', *Journal of Molecular Evolution*, 38, 138–55.

Cloutier, R. and Ahlberg, P. E. (1996) 'Morphology, characters, and the interrelationships of basal sarcopterygians', in Stiassny, M. L. J., Parenti, L. R. and Johnson, G. D. (eds) *Interrelationships of fishes*, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto: Academic Press, pp. 445–79.

Degnan, B. M., Yan, J., Hawkins, C. J. and Lavin, M. F. (1991) 'rRNA genes from the lower chordate *Herdmania momus*: structural similarity with higher eukaryotes', *Nucleic Acids Research*, 18, 7063–70.

Delarbre, C., Spruyt, N., Delmarre, C., Gallut, C., Barriel, V., Janvier, P., Laudet, V. and Gachellin, G. (1998) 'The complete nucleotide sequence of the mitochondrial DNA of the dogfish, *Scyliorhinus canicula*', *Genetics*, 150, 331–44.

Desjardins, P. and Morais, R. (1990) 'Sequence and gene organization of the chicken mitochondrial genome', *Journal of Molecular Biology*, 212, 599–634.

Farias, I. P., Orti, G., Sampaio, I., Schneider, H. and Meyer, A. (1999) 'Mitochondrial DNA phylogeny of the Cichlidae: monophyly and fast molecular evolution of the Neotropical assemblage', *Journal of Molecular Evolution*, 48, 703–11.

Felsenstein, J. (1978) 'The number of evolutionary trees', *Systematic Zoology*, 27, 27–33.

Felsenstein, J. (1985) 'Confidence limits on phylogenies: an approach using the bootstrap', *Evolution*, 39, 783–91.

Gonzalez, I. L., Gorski, J. L., Campen, T. J., Dorney, D. J., Erickson, J. M., Sylvester, J. E.

and Schmickel, R. D. (1985) 'Variation among human 28S ribosomal RNA genes', *Proceedings of the National Academy of Sciences of the United States of America*, 82, 7666–70.

Hadjiolov, A. A., Georgiev, O. I., Nosikov, V. V. and Yavachev, L. P. (1984) 'Primary and secondary structure of rat 28S ribosomal RNA', *Nucleic Acids Research*, 12, 3677–93.

Härlid, A. and Arnason, U. (1999) 'Analyses of mitochondrial DNA nest ratite birds within the Neognathae: supporting a neotenous origin of ratite morphological characters', *Proceedings of the Royal Society of London* B, 266, 305–9.

Härlid, A., Janke, A. and Arnason, U. (1997) 'The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds', *Molecular Biology and Evolution*, 14, 754–61.

Hasegawa, M., Kishino, H. and Yano, T. (1985) 'Dating of the human–ape splitting by a molecular clock of mitochondrial DNA', *Journal of Molecular Evolution*, 22, 160–74.

Hassouna, N., Michot, B. and Bachellerie, J. -P. (1984) 'The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryocytes', *Nucleic Acids Research*, 12, 3563–83.

Hillis, D. M. and Dixon, M. T. (1991) 'Ribosomal DNA: molecular evolution and phylogenetic inference', *Quarterly Review of Biology*, 66, 411–53.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. and Takahata, N. (1995) 'Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs', *Proceedings of the National Academy of Sciences of the United States of America*, 92, 532–6.

Janke, A. and Arnason, U. (1997) 'The complete mitochondrial genome of *Alligator mississippiensis* and the separation between Recent Archosauria', *Molecular Biology and Evolution*, 14, 1266–72.

Janke, A., Feldmaier-Fuchs, G., Thomas, K., Von Haeseler, A. and Pääbo, S. (1994) 'The marsupial mitochondrial genome and the evolution of placental mammals', *Genetics*, 137, 243–56.

Janke, A., Gemmell, N. J., Feldmaier-Fuchs, G., von Haeseler, A. and Pääbo, S. (1996) 'The mitochondrial genome of a monotreme – the platypus (*Ornithorhynchus anatinus*)', *Journal of Molecular Evolution*, 42, 153–9.

Johansen, S. and Bakke, I. (1996) 'The complete mitochondrial DNA sequence of Atlantic cod, *Gadus morhua*: relevance to taxonomic studies among cod-fishes', *Molecular Marine Biology*, 5, 203–14.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X. and Wilson, A. C. (1989) 'Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers', *Proceedings of the National Academy of Sciences of the United States of America*, 86, 6196–200.

Kumazawa, Y. and Nishida, M. (1993) 'Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics', *Journal of Molecular Evolution*, 37, 380–98.

Kumazawa, Y., Ota, H., Nishida, M. and Ozawa, T. (1998) 'The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions', *Genetics*, 150, 313–29.

Lee, W. J. and Kocher, T. D. (1995) 'The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions', *Genetics*, 139, 873–87.

Lockhart, P. J., Steele, M. A., Hendy, M. D. and Penny, D. (1994) 'Recovering evolutionary distances under a more realistic model of sequence evolution', *Molecular Biology and Evolution*, 11, 605–12.

Lyons-Weiler, J., Tausch, R. J. and Hoelzer, G. A. (1998) 'Optimal outgroup analysis', *Biological Journal of the Linnean Society*, 64, 493–511.

Maddison, W. P. and Maddison, D. R. (1992) *MacClade: analysis of phylogeny and character evolution*, Sunderland, Massachusetts: Sinauer Associates Inc.

Mallatt, J. and Sullivan, J. (1998) '28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes', *Molecular Biology and Evolution*, 15, 1706–18.

Martin, A. P., Naylor, G. J. P. and Palumbi, S. R. (1992) 'Rates of mitochondrial DNA evolution in sharks are slow compared with mammals', *Nature*, 357, 153–5.

Meyer, A. and Wilson, A. C. (1990) 'Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish', *Journal of Molecular Evolution*, 31, 359–64.

Milinkovitch, M. C. and Lyons-Weiler, J. (1998) 'Finding optimal outgroup topologies and convexities when the choice of outgroups is not obvious', *Molecular Phylogenetics and Evolution*, 9, 348–57.

Milinkovitch, M. C., Orti, G. and Meyer, A. (1993) 'Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences', *Nature*, 361, 346–8.

Mindell, D. P., Sorenson, M. D., Dimcheff, D. E., Hasegawa, M., Ast, J. C. and Yuri, T. (1999) 'Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes', *Systematic Biology*, 48, 138–52.

Murakami, M., Yamashita, Y. and Fujitani, H. (1998) 'The complete sequence of mitochondrial genome from a gynogenetic triploid "Ginbuna" (*Carassius auratus langsdorfi*)', *Zoological Science*, 15, 335–7.

Naylor, G. J. P. and Brown, W. M. (1997) 'Structural biology and phylogenetic estimation', *Nature*, 388, 527–8.

Naylor, G. J. P. and Brown, W. M. (1998) 'Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences', *Systematic Biology*, 47, 61–76.

Noack, K., Zardoya, R. and Meyer, A. (1996) 'The complete mitochondrial DNA sequence of the bichir (*Polypterus ornatipinnis*), a basal ray-finned fish: ancient establishment of the consensus vertebrate gene order', *Genetics*, 144, 1165–80.

Philippe, H. and Adoutte, A. (1998) 'The molecular phylogeny of Eukaryota: solid facts and uncertainties', in Coombs, G., Vickerman, K., Sleigh, M. and Warren, A. (eds) *Evolutionary relationships among Protozoa*, London: Chapman & Hall, pp. 25–56.

Rasmussen, A. -S. and Arnason, U. (1999a) 'Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes', *Journal of Molecular Evolution*, 48, 118–23.

Rasmussen, A. -S. and Arnason, U. (1999b) 'Molecular studies suggest that cartilaginous fishes have a terminal position in the piscine tree', *Proceedings of the National Academy of Sciences of the United States of America*, 96, 2177–82.

Rasmussen, A -S., Janke, A. and Arnason, U. (1998) 'The mitochondrial DNA molecule of the hagfish (*Myxine glutinosa)* and vertebrate phylogeny', *Journal of Molecular Evolution*, 46, 382–8.

Roe, B. A., Din-Pow, M., Wilson, R. K. and Wong, J. F. (1985) 'The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome', *Journal of Biological Chemistry*, 260, 9759–74.

Romer, A. S. (1966) *Vertebrate paleontology*, 3rd edition, Chicago: The University of Chicago Press.

Russo, C. A. M., Takezaki, N. and Nei, M. (1996) 'Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny', *Molecular Biology and Evolution*, 13, 525–36.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) 'Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase', *Science*, 239, 487–91.

Saitou, N. and Nei, M. (1987) 'The neighbour-joining method: a new method for reconstructing phylogenetic trees', *Molecular Biology and Evolution*, 4, 406–25.

Sogin, M. L. (1989) 'Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs', *American Zoologist*, 29, 487–99.

Strimmer, K. and von Haeseler, A. (1996) 'Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies', *Molecular Biology and Evolution*, 13, 964–9.

Swofford, D. L. (1997) *PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0*, Sunderland, Massachusetts: Sinauer Associates, Inc.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M. (1996) 'Phylogenetic inference', in Hillis, D. M., Moritz, C. and Mable B. K. (eds) *Molecular systematics*, 2nd edition, Sunderland, Massachusetts: Sinauer Associates, Inc., pp. 407–514.

Takezaki, N. and Gojobori, T. (1999) 'Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences', *Molecular Biology and Evolution*, 16, 590–601.

Templeton, A. R. (1983) 'Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes', *Evolution*, 37, 221–44.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*, 22, 4673–80.

Tzeng, C. S., Hui, C. F., Shen, S. C. and Huang, P. C. (1992) 'The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variation among vertebrates', *Nucleic Acids Research*, 20, 4853–8.

Yang, Z. and Kumar, S. (1996) 'Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites', *Molecular Biology and Evolution*, 13, 650–9.

Zardoya, R., Cao, Y., Hasegawa, M. and Meyer, A. (1998) 'Searching for the closest living relative(s) of tetrapods through evolutionary analyses of mitochondrial and nuclear data', *Molecular Biology and Evolution*, 15, 506–17.

Zardoya, R., Garrido-Pertierra, A. and Bautista, J. M. (1995) 'The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*', *Journal of Molecular Evolution*, 41, 942–51.

Zardoya, R. and Meyer, A. (1996a) '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, 1249–63.

Zardoya, R. and Meyer, A. (1996b) 'Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene', *Proceedings of the National Academy of Sciences of the United States of America*, 93, 5449–54.

Zardoya, R. and Meyer, A. (1996c) 'Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates', *Molecular Biology and Evolution*, 13, 933–42.

Zardoya, R. and 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. and Meyer, A. (1998) 'Complete mitochondrial genome suggests diapsid affinities of turtles', *Proceedings of the National Academy of Sciences of the United States of America*, 95, 14226–31.

Zardoya, R. and Meyer, A. (2000) 'Mitochondrial evidence on the phylogenetic position of caecilians (Amphibia: Gymnophiona)', *Genetics*, 155, 765–75.