Chapter 8 Pp. 119-134 in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, Genetics and Development* (Per E. Ahlberg, Editor). London, Taylor & Francis. 2001.

Molecular evidence for the early history of living vertebrates

S. Blair Hedges

ABSTRACT

Molecular data bearing on the origin and early history of vertebrates are assembled and analysed for phylogeny and times of divergence. The divergence time for cephalochordates and vertebrates is estimated here as 751 ± 30.9 Mya (million years ago) using nine constant-rate nuclear protein-coding genes. This suggests that free-swimming animals with a notochord, neural tube, and metameric lateral muscles were present about 200 million years before the first fossil evidence of bilaterian animals. By inference, urochordates, hemichordates, and echinoderms diverged even earlier in the Proterozoic. It is suggested that the origins of many major lineages of animals at the level of phylum and below were associated with Neoproterozoic glaciation events (Neoproterozoic Refugia Model). A phylogenetic analysis of the major groups of vertebrates, with 10 nuclear genes, supports the traditional tree: (Agnatha, (Chondrichthyes, (Actinopterygii, Tetrapoda))). The monophyly of Gnathostomata (jawed vertebrates) and of Osteichthyes (bony vertebrates) is each supported (100 per cent bootstrap confidence). A separate phylogenetic analysis of seven nuclear protein-coding genes having representatives of both hagfish and lamprey supported cyclostome monophyly (97 per cent) in agreement with analyses of ribosomal genes and some morphological studies. An early Palaeozoic divergence time (499 ± 36.8 Mya) was estimated for hagfish and lamprey.

8.1 Introduction

Vertebrates have left one of the best fossil records of any major group of organisms. From this it is possible, at least in general terms, to trace their evolution starting from primitive jawless fishes, through various lineages of jawed fishes, to terrestrial forms occupying a diversity of habitats (Benton 1997; Carroll 1997). Although the appearance of these taxa in the fossil record suggests a natural continuity of events, our knowledge of the specific branching pattern and times of divergence for many vertebrate groups remains poorly known. New discoveries of fossils continue to refine the picture, but information from molecules has the potential to greatly clarify our understanding of the early history of vertebrates.

The current view of vertebrate phylogeny supported by morphological and fossil data (Figure 8.1) places the tetrapods in a derived position relative to the fishes. The hagfishes (Myxinoidea) are considered to represent the most basal lineage, with the

lampreys (Petromyzontoidea) as closest relatives of the gnathostomes. Among the gnathostomes, the cartilaginous fishes (Chondrichthyes) are believed to be basal (= monophyletic Osteichthyes), with the ray-finned fishes (Actinopterygii) as the closest relatives of the group containing tetrapods and sarcopterygian fishes. The closest relative of the tetrapods among the sarcopterygian fishes remains a controversial question, and there is no real consensus of opinion (Schultze and Trueb 1991; Fritzsch 1992; Ahlberg and Milner 1994; Schultze 1994). All of these divergences mentioned must have occurred before 380–400 Mya based on the fossil record (Benton 1993; 1997).

One limitation of molecular approaches is the inability to sample ancient, extinct taxa (e.g., conodonts and placoderms). Also, the relatively small number of amino acids in a typical gene, about 300, is usually insufficient to significantly resolve most nodes in a phylogenetic tree or to robustly estimate divergence times. To overcome the gene size limitation, multiple nuclear genes can be combined for phylogeny estimation, or individual gene time estimates can be averaged.

Molecular evidence for the early history of vertebrates has come from nuclear and mitochondrial genes, and these two sources of data have produced strikingly different results. In both cases, the traditional tree, based on morphology, has not been supported, although the greatest disagreement has been with mitochondrial data. One of the earliest contributions of nuclear protein data came from sequence analyses of the globin genes which clearly supported the monophyly of gnathostomes (Goodman *et al.* 1975; Goodman *et al.* 1987). Nuclear protein data have also supported a monophyletic Osteichthyes (Goodman *et al.* 1987). However, nuclear protein and ribosomal gene analyses have, in contrast to most recent morphological studies (although, see Janvier 1996), consistently supported the monophyly of the cyclostomes (lampreys and hagfishes) (Goodman *et al.* 1987; Stock and Whitt 1992; Mallatt and Sullivan 1998).

Higher-level vertebrate phylogenies based on mitochondrial protein-coding genes (concatenated) have differed almost completely from those based on morphology

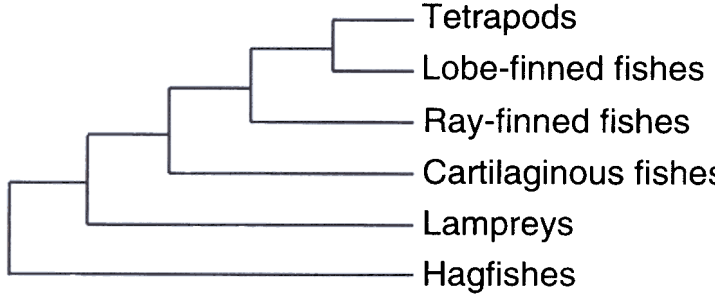

Figure 8.1 The current view of vertebrate phylogeny based on morphology and the fossil record (Benton 1997; Carroll 1997). The jawless fishes (Agnatha; Cyclostomata) are the lampreys (Petromyzontoidea) and hagfishes (Myxionoidea). The jawed vertebrates (Gnathostomata) are all others. Among the gnathostomes, the tetrapods (Tetrapoda), ray-finned fishes (Actinopterygii), and lobe-finned fishes (Sarcopterygii) comprise the bony fishes (Osteichthyes). The remaining gnathostomes are the cartilaginous fishes (Chondrichthyes).

and nuclear gene data. The most surprising result of these studies has been the monophyly of gnathostome fishes, *excluding* tetrapods (Rasmussen *et al.* 1998; Rasmussen and Arnason 1999). Cyclostomes, actinopterygians, and sarcopterygians were each found to be either paraphyletic or polyphyletic, and the lungfishes appear as the most basal lineage of gnathostome fish. According to Rasmussen and Arnason (1999) this phylogeny reflects the true relationships. Although many of the nodes were supported by high bootstrap values, other authors have suggested that these results are due to biased taxon sampling (Cao *et al.* 1998) and limitations in the phylogenetic methods of analysis (Takezaki and Gojobori 1999).

In contrast to the results from mitochondrial protein-coding genes, a molecular clock analysis of pairwise divergence times, using 13–107 nuclear protein-coding genes, yielded a monophyletic Osteichthyes (in agreement with morphology) and divergence times only slightly earlier than those in the fossil record (Kumar and Hedges 1998). The split between agnathans (represented by lampreys) and gnathostomes was estimated as 564 ± 74.6 Mya, between Chondrichthyes and Osteichthyes as 528 ± 56.4 Mya, and between Actinopterygii and Tetrapoda as 450 ± 35.5 Mya.

Considering the disparate results obtained with nuclear genes and mitochondrial protein-coding genes, it is worth examining the latest molecular evidence bearing on the early history of vertebrates. Because the mitochondrial results are already based on complete genomes of that molecule, the emphasis here will be on updating the evidence from nuclear genes. The questions that will be addressed are:

1 the time of divergence between the cephalochordates and vertebrates;
2 the relationships of the major lineages of vertebrates (agnathans, chondrichthyans, actinopterygians, and tetrapods); and
3 the relationships of the cyclostomes.

8.2 Materials and methods

8.2.1 Time estimation

To estimate the time of divergence between cephalochordates and vertebrates, all relevant protein sequences in the databases (Entrez/Genbank) were obtained. Genes useful for analysis were those in which sequences were available in a cephalochordate, vertebrate calibration taxa (see below), and an outgroup. If an arthropod lineage was available, it was included along with a more distant outgroup. This was done so that the arthropod/chordate divergence time (Wang *et al.* 1998) could be used as an additional calibration point. Outgroups were necessary for determining rate constancy in the lineages being timed and in the calibration lineages.

The 13 genes and 149 sequences (accession numbers given) analysed are: acetylcholinesterase, U74381, U74380, M32391, P06276, C39768, S70849, Q03311, AF030422, ACRYE, U05036, M55040, JH0314, S47639, A54413, and AJ223965; aldolase, JC4188, JC4189, ADHUB, U85645, ADRTB, ADCHB, S48810, S57270, X82278, AF067796, E27421, ADMSA, ADHUA, D38621, I51247, U36777, AF041454, ADRTC, ADHUC, X60064, and AB005035; bone morphogenetic protein, S45355, S37073, X75914, X63424, AF072456, D30751, M22490, I49541, Q90752, U90122, AF068750, D85464, Q26974, and Z74046; engrailed, M10017,

AF091246, L12705, D48423, E48423, F48423, S30437, X68151, S19005, B48423, C48423, S30438, U82487, and L14730; hedgehog, AB018076, U85610, U58511, U26404, AAB34105, X76290, B53193, L38518, A49424, S56765, U26314, L35248, U30710, A49426, Y13858, Q02936, and U21308; homeotic protein msx (Hox7-8), D31771, A46122, P28362, JS0660, P23410, Q04281, AAB19630, 2107332A, AAB35456, JS0659, AJ130766, and AF042653; insulin receptor factor, J05149, M29014, M32972, AB003362, AJ223164, AF055980, AJ224993, O02466, U72939, and AF012437; phenylalanine hydroxylase, M12337, X51942, U49897, X98116, P17276, AJ001677, Y16353, U26428, L20679, and S51199; ribosomal protein, S6, M20020, Z54209, P47838, AF020551, AF009665, L01658, and Z83268; superoxide dismutase-mn, L35528, P07895, P04179, P41982, L22092, X64061, P41977, and Q00637; triosephosphate isomerase, X69723, ISCZTI, P15426, P00939, P00940, AB00892, L07390, U60870, and AL023828; twist, Y10871, I53066, M27730, AF097914, P10627, and AF037063; whn, X81593, 2022323A, Y11741, X97021, Y11544, and U57029. Sequences were aligned using CLUSTALX (Thompson *et al.* 1994).

Four calibration points were used. The first is the split between mammals and living reptiles at 310 Mya (Benton 1997; Kumar and Hedges 1998). The second is the split between living amphibians and amniotes at 360 Mya, which is a molecular time estimate using 107 nuclear genes (Kumar and Hedges 1998) but is close to the fossil-based time of 338 Mya (Paton *et al.* 1999). The third is the split between actinopterygian fishes and tetrapods at 450 Mya, which is a molecular time estimate based on 44 genes (Kumar and Hedges 1998) but also is close to the fossil time of about 420 Mya (Benton 1993). The last calibration point is the split between chordates and arthropods at 993 Mya based on a molecular time estimate using 50 nuclear genes (Wang *et al.* 1998). Only one calibration point is necessary to estimate time and therefore the lack of complete independence among these four calibration points is not a problem, especially considering that the independent fossil times for the second and third calibration points differ by only 6–7 per cent from the corresponding molecular estimates. The slope of the regression line between these calibration times and their corresponding genetic distances, forced through the origin, was used for estimating the time of divergence between cephalochordates and vertebrates.

Determination of orthology groups was confirmed by visual inspection of the phylogenetic trees. Genetic distances and tests of rate constancy (Takezaki *et al.* 1995) were performed using PHYLTEST (Kumar 1996). All insertion–deletion sites were excluded prior to distance estimation. A gamma distance with rate parameter (alpha) of 2.0 was used. This value of alpha was obtained empirically in two other studies using large numbers of genes (Gu 1997; Wang *et al.* 1998) and corresponds approximately to a Dayhoff correction (Dayhoff *et al.* 1978; Ota and Nei 1994). For distance calculation, sequences representing the same taxonomic group were placed into clusters and average distances between clusters were calculated (Rzhetsky *et al.* 1995; Kumar 1996). Rate differences among lineages were examined for each gene to determine significant rate variation (5 per cent level). The average distance method (Kumar and Hedges 1998) was used to estimate divergence times for each gene (except those violating rate constancy) and these were averaged across all genes. In the case of the lamprey/hagfish divergence time estimate, the lineage-specific method was used (Kumar and Hedges 1998; Schubart *et al.* 1998).

8.2.2 Phylogeny estimation

All protein sequences in the databases (Entrez/Genbank) were obtained bearing on higher-level vertebrate phylogeny. For vertebrate phylogeny, genes were selected if at least one sequence was available for each of the following groups: tetrapod, actinopterygian, chondrichthyan, agnathan, and outgroup (unfortunately there are insufficient genes available for sarcopterygians). For consistency, tetrapods included an amphibian (usually *Xenopus*), a reptile (usually *Gallus*), and a mammal (usually *Homo*). Ten genes and 64 sequences (accession numbers given) were identified that met these criteria (an asterisk denotes sequence used in combined analysis): alpha globin A, HACH2, P07428, HARKJ*, HACA*, HAHU*, S13458*, and S15979*; beta globin, P02023*, P02112, HBCAA*, M32457, HBRKJ*, S13458*, and S15979*; cytochrome c (cyt c), P00001*, CCCH, P00024, P00025*, CCDF*, CCLM*, and P00029*; insulin, AB36057*, IPHF, 1012233A*, HIUB*, INTK, IPXL1, 124688*, and A38422*; insulin-like growth factor (igf2), S82962, M95184*, IGHU2*, Z50082*, P22618*, and Z81098*; large multifunctional protein 7 (lmp7), AF032390*, U17497*, D64056*, D64055*, D64054, and X97729*; neurofilament medium protein (nf-m), U85969*, I50479*, PN0009*, U19361*, and P12036*; neuropeptide Y (npy), L22867*, P01303*, P28673, M87297*, P28674*, and L22868*; proopiomelanocortin (pomc), M38297*, X05940, AB020972*, U59910*, I51117*, and D55629*; and wnt-1, P04628*, X58880*, X55270, M91250*, P28114*, and U58982*.

For agnathan relationships, genes were selected if at least one sequence was available for each of the following groups: lamprey, hagfish, gnathostome (mammal), and outgroup. Seven genes and 49 sequences were identified that met these criteria (an asterisk denotes sequence used in combined analysis): beta adrenoreceptor, M14379, J03019*, Y09213, AJ005436*, AJ005438*, AJ005433*; complement component C3 (CC-C3), K02765*, I50711, AB016213, I50806*, Z11595*, AF025526*; engrailed, S13011, A48423*, S13010, S13012*, S13013*, S18301*; globin, HAHU*, HACH, P07428, HARKJ, HACA, HBRKJ, M32457, HBCAA, P02023, P02112, S13458*, GGHF3G*, S15979*; insulin, AAB36057*, IPHF*, 224208, HIUB*, INTK, IPXL1, 124688, A38422*; lmp7, AF032390, U17497*, D64056, D64055*, D64054*, X97729*; and superoxide dismutase-mn, P04179*, P28762*, X64059*, X64061*.

Sequences were aligned using CLUSTALX (Thompson *et al.* 1994). Phylogenetic analyses were performed with MEGA (Kumar *et al.* 1993), using neighbour-joining (Saitou and Nei 1987) and a gamma distance (alpha = 2.0). Because other methods of analysis yield identical trees for well-supported nodes, they were not used here. All insertion–deletion sites were excluded prior to distance estimation. Confidence values on nodes in the resulting trees were obtained with the bootstrap method (Felsenstein 1985) using 2000 replications (Hedges 1992). Values of 95 per cent and above were considered to be significant.

8.3 Results

Four of the 13 genes analysed for estimating the divergence time for cephalochordates and vertebrates could not be used. Rate constancy was rejected for homeotic protein msx, insulin receptor factor, and ribosomal protein S6. In the case of

phenylalanine hydroxylase, *Branchiostoma* was more closely related to *Drosophila* sequences than to vertebrate sequences suggesting a paralogy problem; thus the gene was omitted from analysis. Times of divergence for the remaining nine constant-rate genes ranged from 553–880 Mya, with a mean of 750.5 Mya and a standard error of 31.9 million years (Table 8.1). If the uppermost and lowermost gene estimates were to be omitted to reduce the probability of including paralogous comparisons (Kumar and Hedges 1998), the mean time would be 760.3 ± 20.1 Mya.

Although the time of divergence between cephalochordates and vertebrates estimated here using nine genes (751 Mya) is superficially similar to that estimated in an earlier study (Nikoh *et al.* 1997) using two genes (700 and 860 Mya), the two studies are not comparable. The calibration times used by those authors were taken from some early studies (Dickerson 1971; Dayhoff 1978) and were lower than those used here. Using their methods (Nikoh *et al.* 1997) with revised calibrations, the divergence time estimate increases to an average of 978 Mya (895 and 1061 Mya) for the two genes in their study. This is nearly the same time as the deuterostome–protostome split (993 Mya), leaving little room for the divergence of echinoderms, hemichordates, and urochordates. Even in this study, using different methods, the time estimates for aldolase (840 Mya) and triosephosphate isomerase (880 Mya) were above average compared with other genes. This is not unusual given the high coefficient of variation of gene-specific time estimates and reinforces the suggestion that large numbers of genes should be used to estimate divergence times (Kumar and Hedges 1998).

All but two of the phylogenetic trees of the ten separate genes analysed for vertebrate relationships resulted in a monophyletic Gnathostomata (Table 8.2). In four genes, the bootstrap support was 100 per cent whereas in two genes (cytochrome c and insulin) there was no significant support for this or any grouping. In general, bootstrap support was highest in the genes having the greatest number of amino

Table 8. Time estimation for the divergence of cephalochordates and vertebrates.

Gene	*No. taxa*	*No. sites*[1]	*Divergence time (Mya)*
Acetylcholinesterase	15	873/523	710.7
Aldolase	21	364/355	840.0
Bone morphogenetic protein	14	555/320	692.2
Engrailed	14	624/170	772.2
Hedgehog	17	602/373	805.4
Homeotic protein msx	12	344/224	–
Insulin receptor factor	10	1936/1259	–
Phenylalanine hydroxylase	10	491/424	–
Ribosomal protein S6	7	250/242	–
Superoxide dismutase-mn	8	234/140	727.0
Triosephosphate isomerase	9	250/209	879.5
Twist	6	519/160	552.8
Winged-helix nude	6	649/127	774.4
Mean			750.5
Standard Error			31.9

[1]Total aligned amino acid sites/sites analysed following removal of insertion–deletion sites.

Table 8.2 Nuclear genes analysed for vertebrate relationships.

Gene	No. taxa	No. sites[1]	Group supported		
			Gnathostomata (%)	Osteichthyes (%)	'Pisces' (%)
Alpha globin A	7	162/129	90	–	–
Beta globin	7	162/130	88	–	–
Cytochrome c	7	105/103	–	61	–
Insulin	8	80/51	–	–	–
Insulin-like growth factor	6	245/116	68	97	–
Large multifunctional protein 7	6	281/253	94	95	
Neurofilament medium protein	5	1173/741	100	58	
Neuropeptide Y	6	105/91	100	–	84
Proopiomelanocortin	6	344/195	100	–	–
Wnt-1	6	383/116	100	79	–
Combined analysis	5	3035/1951	100	100	

[1]Total aligned amino acid sites/sites analysed following removal of insertion–deletion sites.

acid sites. One gene (neuropeptide Y) supported the controversial grouping 'Pisces' (Rasmussen and Arnason 1999) at a bootstrap value of 84 per cent, but five genes supported the traditional grouping Osteichthyes (Table 8.2). As predicted by the separate gene results, the combined analysis of all ten genes (sequences concatenated; 1951 sites) resulted in significant support (100 per cent) for both Gnathostomata and Osteichthyes (Table 8.2; Figure 8.2a).

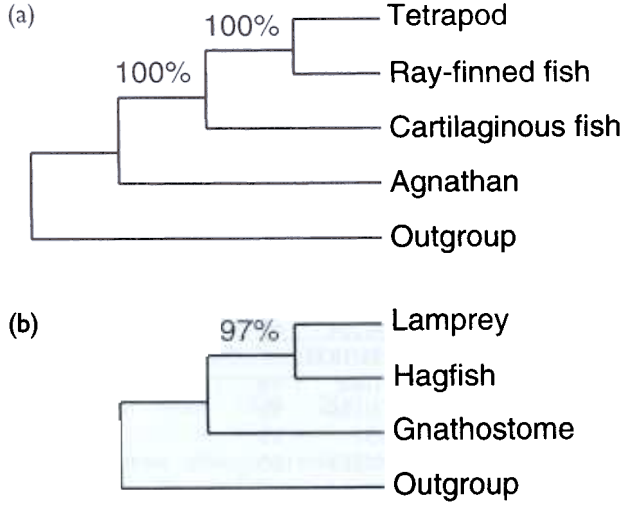

Figure 8.2 Molecular evidence for vertebrate phylogeny. (a) Combined analysis of 10 nuclear protein-coding genes. (b) Combined analysis of seven nuclear protein-coding genes having representatives of hagfish and lamprey. Bootstrap confidence values are shown at nodes.

All but one of the phylogenetic trees of the seven separate genes analysed for agnathan relationships resulted in a monophyletic Cyclostomata (Table 8.3). In three genes, the bootstrap support for this grouping of lamprey and hagfish was significant (99 per cent). The single gene that did not support monophyly of the cyclostomes, superoxide dismutase-mn, instead supported a hagfish + gnathostome grouping (lamprey basal), although not significantly (73 per cent). None of the genes supported the traditional morphological grouping of lampreys with gnathostomes. As expected from this result, the combined analysis of all seven genes (sequences concatenated; 2375 sites) resulted in significant support (97 per cent) for the monophyly of the cyclostomes (Table 8.3; Figure 8.2b).

None of the seven genes used for agnathan relationships could be used for estimating the time of divergence of lampreys and hagfish. In the case of two genes (engrailed and insulin) the number of amino acid sites was too small (< 100 AA) for time estimation. Superoxide dismutase-mn could not be used because it did not result in a monophyletic Cyclostomata. The remaining four genes failed the rate constancy test. However, rate testing of the combined data set showed that the hagfish lineage was 13 per cent longer than the lamprey lineage, and that rate constancy was not rejected if the hagfish was removed. Thus, a lineage-specific method (Schubart *et al.* 1998) was used to estimate the divergence time in the combined data set. The lamprey lineage ($d = 0.444$) and the internal branch between the lamprey/hagfish split and gnathostome/cyclostome split ($d = 0.574$) were used to estimate the divergence time for the lamprey/hagfish split. The agnathan/gnathostome divergence time estimate of 564 Mya (Kumar and Hedges 1998) was used as a calibration, resulting in a time estimate of 499 Mya for the divergence of lamprey and hagfish. The standard error of the lamprey/hagfish distance (0.327) was used to estimate the standard error (36.8 Myr) of the time estimate.

8.4 Discussion

8.4.1 Neoproterozoic Refugia and the origin of vertebrates

The divergence time estimated here for cephalochordates and vertebrates indicates that free-swimming animals (chordates) with a notochord, neural tube, and

Table 8.3 Nuclear genes analysed for agnathan relationships.

Gene	No. taxa	No. sites[1]	Group supported		
			Cyclostomata (%)	*Vertebrata (%)*	*Lamprey basal (%)*
Beta adrenoreceptor	6	554/204	89	–	–
Complement component C3	6	1796/1534	61	–	–
Engrailed	7	401/60	99	–	–
Globin	13	163/126	99	–	–
Insulin	8	80/51	93		
Large multifunctional protein 7	6	281/253	100	–	–
Superoxide dismutase-mn	4	222/144	–	–	73
Combined analysis	4	3499/2375	97		

[1]Total aligned amino acid sites/sites analysed following removal of insertion–deletion sites.

metameric lateral muscles had evolved by 750 Mya (Figure 8.3). It also raises the possibility that some or all of the defining characters of vertebrates (Nielsen 1995) arose deep in the Proterozoic (750–530 Mya). By inference, lineages leading to the urochordates, hemichordates, and echinoderms arose even earlier (751–993 Mya). Molecular clock studies have consistently found early divergences for selected animal phyla (Brown *et al.* 1972; Runnegar 1982b; 1986; Wray *et al.* 1996; Feng *et al.* 1997; Ayala *et al.* 1998; Gu 1998; Wang *et al.* 1998). However, there are more than 30 different phyla and only a few divergent representatives (e.g., chordates and arthropods) have been compared at a significant number of genes, leaving open the possibility of a more recent origin for the derived phyla of protostomes and deuterostomes. This finding of an early divergence between the two most derived phyla of deuterostomes, Cephalochordata and Vertebrata (considered by some authors to be subphyla of the phylum Chordata), provides even stronger evidence of discordance between the fossil record and molecular time estimates.

Most discussions of the Cambrian Explosion and early evolution of animals concern the origin of animal *phyla*. However, these results suggest that perhaps many major groups of animals below the level of phylum arose during the Neoproterozoic. Among extant protostomes, likely candidates include (but are not limited to) the molluscan classes Bivalvia, Cephalopoda, Gastropoda, Monoplacophora, and Polyplacophora, and the arthropod taxa Chelicerata, Ostracoda, Cirripedia, and Malacostraca. All are represented in the Cambrian fossil record (Benton 1993). Among deuterostomes, the echinoderm classes and subclasses Echinoidea, Holothuroidea, Asteroidea, Ophiuroidea, Somasteroidea, and Crinoidea all have a fossil record extending back into the Cambrian or Ordovician (Benton 1993). If one

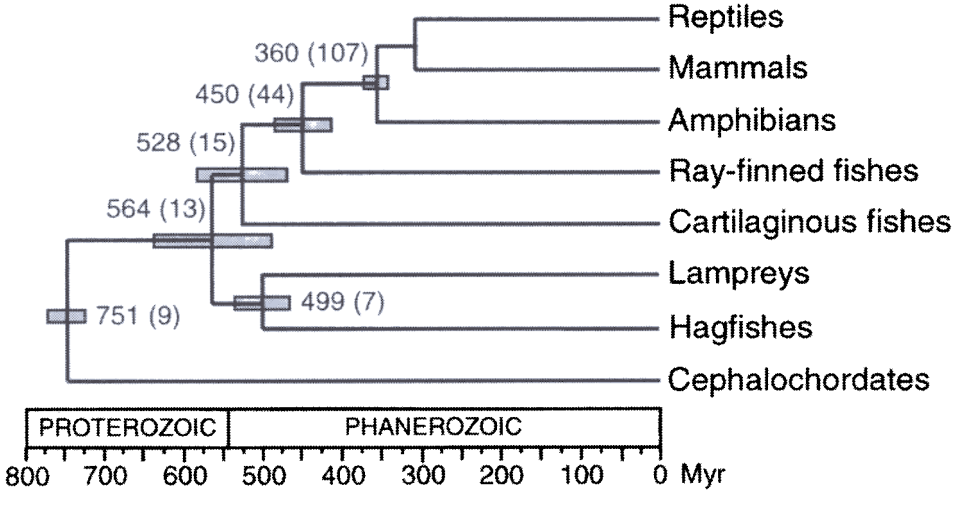

Figure 8.3 Timetree of vertebrate phylogeny and times of divergence estimated from nuclear protein-coding genes. Times of divergence are indicated at nodes (number of genes in parentheses) along with standard errors (gray bars). The divergence between reptiles and mammals was used as a calibration point. The divergence times for amphioxus (Cephalochordata) and between lampreys (Petromyzontoidea) and hagfishes (Myxinoidea) were estimated here; other times are from elsewhere (Kumar and Hedges 1998).

considers the subgroups of other animal phyla (e.g. sponges, cnidarians) already known to have a long fossil record, and those poorly fossilized groups believed to have arisen early, then it is possible that more than 100 lineages of extant metazoans arose in the Neoproterozoic.

Why is there no clear evidence in the fossil record for the existence of metazoans prior to about 600 Mya (Li *et al.* 1998; Xiao *et al.* 1998)? A variety of explanations have been proposed (Bengston and Lipps 1992; Lipps *et al.* 1992) although the one most frequently mentioned is that early animals were smaller and soft-bodied (Runnegar 1982a; Bengston 1994; Davidson *et al.* 1995; Fedonkin 1994; Weiguo 1994; Seilacher *et al.* 1998). There is evidence from trace fossils of a size increase in bilaterian animals and for the acquisition of hard parts occurring at the Proterozoic/Phanerozoic boundary (Bengston and Farmer 1992; Lipps *et al.* 1992; Valentine *et al.* 1999). Nearly one-third of animal phyla (e.g., Gastrotricha, Placozoa), all small in size and most soft-bodied, have virtually no fossil record (Valentine *et al.* 1999) yet most of those have existed at least since the early Phanerozoic based on phylogenetic evidence. This fact in itself argues that the absence of metazoan fossils prior to 600 Mya should not be taken as a challenge to the molecular time estimates. However, the body plans of most metazoan phyla, such as the arthropod limb and molluscan foot, are adapted to a bottom dwelling (macrobenthic) lifestyle. It is not yet clear how these body plans arose during a small and soft-bodied stage of evolution (Conway Morris 1998; Valentine *et al.* 1999). The evolution of animal body plans would be more compatible with the molecular time estimates if early animals were macroscopic, and this cannot yet be ruled out. For example, conodont vertebrates were abundant in the Palaeozoic, based on fossils of their mineralized feeding apparatus but their soft, eel-like bodies (~40 mm long) were unknown until relatively recently (Benton 1997).

The divergence time estimate for the origin of the vertebrate lineage is about the same time as the onset of the first major Neoproterozoic glaciation event (Sturtian; 750–700 Mya) or 'snowball Earth' episode (Hoffman *et al.* 1998). It is possible that this and the other major glaciation (Varanger; 610–570 Mya) of the late Neoproterozoic led to considerable speciation as a result of contraction of ranges and genetic isolation for long periods of time (~10 million years). An association between these major glaciations and the presumed origin of metazoans in the latest Neoproterozoic (~600 Mya) has been proposed elsewhere (Kirschvink 1992; Knoll 1994; Kaufman *et al.* 1997; Hoffman *et al.* 1998). However, molecular time estimates and phylogenetic constraints suggest that at least 10 lineages of metazoans (ancestors of extant phyla) were already present prior to the first glaciation at 750 Mya.

The time required for speciation varies with taxonomic group and is not well understood, although it is often less than one million years (Mayr 1963) and 10 million years presumably would be sufficient for nearly any two populations to evolve into different species. Two populations separated by a short amount of time, such as hundreds of years or a few thousand years, will most likely interbreed upon contact and speciation will not occur. Each major glaciation event could have led to many small refugia and thus may have generated many new species, the latter being potential precursors to major animal groups.

Climatic cycles also have been suggested as a mechanism for the generation of species during the Pleistocene (Haffer 1969). However, the existence of Pleistocene

refugia has been debated (Colinvaux *et al.* 1996) and intervals during Pleistocene glaciations may have been too short to have caused speciation in most groups, as evidenced by molecular clock studies of vertebrates (Maxson and Roberts 1984; Maxson and Heyer 1988; Klicka and Zink 1997). The Neoproterozoic refugia probably were an order of magnitude (or more) longer in duration, extending beyond the length of time needed for speciation. Such refugia were not necessary for geographic isolation, because any reproductive barrier resulting in prolonged isolation can lead to speciation. However, isolation of populations would have been greater than usual during the Neoproterozoic glaciations.

The extreme environmental conditions associated with the major Neoproterozoic glaciations, including a post-glacial greenhouse period, would have provided a strong selective force on populations surviving in refugia. Such refugia probably were associated with rift zones near the surface (Kasting 2000); other possible but less likely sites were deep-sea vents or thermal springs on continents. Indirect evidence against deep-sea vents as Neoproterozoic refugia for animals is that there are no known extant phyla endemic to deep-sea vents, with the exception of Vestimentifera (if regarded as a phylum). Also, most phyla are not known to be associated with deep-sea vents and phyla that occur in those areas (Hessler and Kaharl 1995) are phylogenetically derived, not basal. An argument against continental thermal springs as refugia is that the greatly reduced precipitation on continents during glaciation events probably would have been insufficient to charge the springs.

The small size of each refugium would have resulted in reduced environmental variability (abiotic and biotic) within the refugium and greater environmental differences among refugia. Such strong selection could have resulted in rapid organismal change eventually leading to different body plans. It is tempting to suggest that living phyla showing resistance to extreme environmental conditions, such as the tardigrades, represent evolutionary products of the Neoproterozoic glaciations. Although it may be true in some cases, the relationship between current adaptations of the phyla and Neoproterozoic environmental conditions may be complicated, if a relationship exists. Genetic drift must have occurred to some degree in the isolated populations, although its importance in speciation (Mayr 1954; Carson 1975) is not widely accepted (Coyne 1992). It is more likely that geographic isolation and natural selection were the only two factors needed for speciation and organismal change in Neoproterozoic refugia.

The prediction of this model is that molecular estimates of divergence time should cluster around the time of those Neoproterozoic glaciations (Figure 8.4). Currently two major glaciations are recognized (Kennedy *et al.* 1998) but additional glaciation events may be discovered in the future. Although such a model would lessen the importance of the Cambrian Explosion in generating animal diversity (lineages of metazoans), it is easy to envision a scenario where Neoproterozoic refugia and the Cambrian Explosion both played a part in generating diversity. Moreover, lineage-splitting and phenotypic change, especially concerning major body plans, do not have to be coupled. The Cambrian Explosion may have been the result of an environmental trigger, such as the rise in atmospheric oxygen above a critical level, that permitted larger body size and the evolution of hard parts (Bengston and Lipps 1992; Bengston 1994; Knoll 1994). Adaptive radiations and additional lineage-splitting almost certainly accompanied such an event. Also, some major lineages probably arose during intervening times (Figure 8.4). This model only suggests that

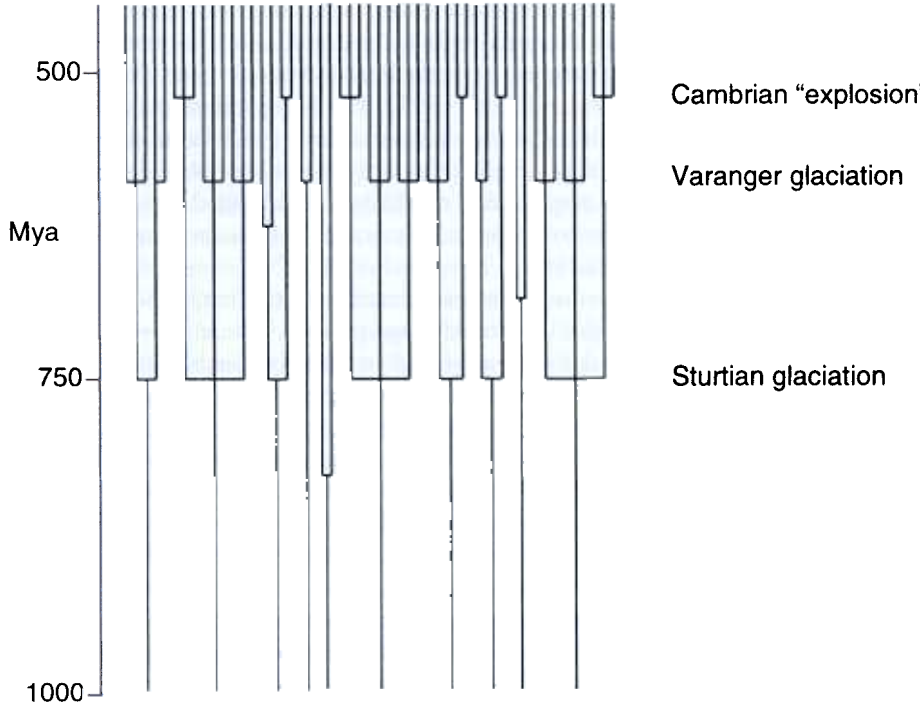

Figure 8.4 Neoproterozoic Refugia model of animal evolution. A predicted pattern of lineage-splitting is shown in ten hypothetical metazoan lineages during the Neoproterozoic and early Phanerozoic. The model predicts that most phylogenetic divergences were concentrated during two Neoproterozoic glaciation events: Sturtian (~750 Mya) and Varanger (~600 Mya). Some additional lineage-splitting would have occurred during the Cambrian Explosion (~530 Mya) associated with adaptive radiation, and during intervening periods.

a normal evolutionary process, speciation and organismal change, was *accelerated* during the Neoproterozoic glaciations.

Given that most major lineages of animals are not yet represented adequately in the sequence databases, the finding that the cephalochordate/vertebrate divergence (~750 Mya) corresponds to the Sturtian glaciation may only be coincidental. A test of whether there was any association between the major Neoproterozoic glaciations and animal evolution will come when divergence times for most or all major metazoan lineages are estimated with large numbers of genes.

8.4.2 Vertebrate relationships

The phylogenetic analysis of nuclear protein coding genes (Table 8.2; Figure 8.2) is concordant with evidence from morphology (Benton 1997; Carroll 1997) and with the topology inferred from a molecular clock analysis of 13–107 nuclear protein-coding genes (Kumar and Hedges 1998). The latter analysis included many genes not used here, and the methods were different. The monophyly of Gnathostomata and of Osteichthyes is each strongly supported. Cyclostome monophyly is also

consistently well supported, both with individual genes and in the combined analysis (Table 8.3; Figure 8.2). None of the seven genes supported a basal position for the hagfish ('Vertebrata'). This agrees with analyses of nuclear ribosomal genes (Stock and Whitt 1992; Mallatt and Sullivan 1998), early considerations based on morphology (Dumeril 1806), and with some recent morphological studies (Løvtrup 1977; Janvier 1996; Mallat 1997a, b). The early Palaeozoic divergence time estimated for the lamprey and hagfish (Figure 8.3) predates the earliest cyclostome fossils, from the Carboniferous (Benton 1997).

These results regarding vertebrate phylogeny stand in contrast to phylogenetic studies of concatenated mitochondrial protein sequences (Rasmussen *et al.* 1998; Rasmussen and Arnason 1999). However, previous mitochondrial protein sequence analyses have yielded phylogenies known to be incorrect when deep divergences among vertebrates were examined (Nei 1996; Naylor and Brown 1997). Recent studies exploring the reason for this have identified taxon-sampling and rooting (Cao *et al.* 1998) and among-site rate variation (Takezaki and Gojobori 1999) as important factors, rather than structural constraints associated with hydrophobic amino acids (Naylor and Brown 1997). Whatever the cause, the unconventional mitochondrial-based trees of higher-level vertebrate relationships do not have support from either morphological or nuclear gene phylogenies.

Acknowledgements

I thank Michael R. Tracy, Marcel van Tuinen, and an anonymous reviewer for comments on the manuscript. This research was supported by grants from NASA (NCC2-1057) and NSF (DEB 9615643).

References

Ahlberg, P. E. and Milner, A. R. (1994) 'The origin and early diversification of tetrapods', *Nature*, 368, 507–14.

Ayala, F. J., Rzhetsky, A. and Ayala, F. J. (1998) 'Origin of the metazoan phyla: molecular clocks confirm paleontological estimates', *Proceedings of the National Academy of Sciences of the United States of America*, 95, 606–11.

Bengston, S. (1994) 'The advent of animal skeletons', in Bengston, S. (ed.) *Early life on Earth*, New York: Columbia University Press, pp. 412–25.

Bengston, S. and Farmer, J. D. (1992) 'The evolution of metazoan body plans', in Schopf, J. W. and Klein, C. (eds) *The Proterozoic biosphere*, Cambridge: Cambridge University Press, pp. 443–6.

Bengston, S. and Lipps, J. R. (1992) 'The Proterozoic-Early Cambrian evolution of metaphytes and metazoans', in Schopf, J. W. and Klein, C. (eds) *The Proterozoic biosphere*, Cambridge: Cambridge University Press, pp. 427–8.

Benton, M. J. (1993) *The fossil record 2*, London: Chapman and Hall.

Benton, M. J. (1997) *Vertebrate paleontology*, New York: Chapman and Hall.

Brown, R. H., Richardson, M., Boulter, D., Ramshaw, J. A. M. and Jeffries, R. P. S. (1972) 'The amino acid sequence of cytochrome c from *Helix aspera* Müeller (Garden Snail)', *Biochemical Journal*, 128, 971–4.

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. (1997) *Patterns and processes of vertebrate evolution*, Cambridge Palaeobiology Series, Cambridge: Cambridge University Press.

Carson, H. L. (1975) 'The genetics of speciation at the diploid level', *American Naturalist*, **109**, 83–92.

Colinvaux, P. A., De Oliveira, P. E., Moreno, J. E., Miller, M. C. and Bush, M. B. (1996) 'A long pollen record from lowland Amazonia: forest and cooling in glacial times', *Science*, **274**, 85–8.

Conway Morris, S. (1998) 'Early metazoan evolution: reconciling paleontology and molecular biology', *American Zoologist*, **38**, 867–77.

Coyne, J. A. (1992) 'Genetics and speciation', *Nature*, **355**, 511–15.

Davidson, E. H., Peterson, K. J. and Cameron, R. A. (1995) 'Origin of bilaterian body plans: evolution of developmental regulatory mechanisms', *Science*, **270**, 1319–25.

Dayhoff, M. O. (1978) 'Survey of new data and computer methods of analysis', in Dayhoff, M. O. (ed.) *Atlas of protein sequence and structure*, Washington, DC: National Biochemical Research Foundation, pp. 1–8.

Dayhoff, M. O., Schwarts, R. M. and Orcutt, B. C. (1978) 'A model of evolutionary change in proteins', in Dayhoff, M. O. (ed.) *Atlas of protein sequence and structure*, Washington, DC: National Biomedical Research Foundation, pp. 345–52.

Dickerson, R. E. (1971) 'The structures of cytochrome c and the rates of molecular evolution', *Journal of Molecular Evolution*, **1**, 26–45.

Dumeril, A. M. C. (1806) *Zoologie analytique ou methode naturelle de classification des animaux*, Paris: Didot.

Fedonkin, M. A. (1994) 'Vendian body fossils and trace fossils', in Bengston, S. (ed.) *Early life on Earth*, New York: Columbia University Press, pp. 370–88.

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

Feng, D. -F., Cho, G. and Doolittle, R. F. (1997) 'Determining divergence times with a protein clock: update and reevaluation', *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 13028–33.

Fritzsch, B. (1992) 'The water-to-land transition: evolution of the tetrapod basilar papilla, middle ear, and auditory nuclei', in Webster, D. B., Fay, R. R. and Popper, A. N. (eds) *The evolutionary biology of hearing*, New York: Springer, pp. 351–75.

Goodman, M., Miyamoto, M. M. and Czelusniak, J. (1987) 'Pattern and process in vertebrate phylogeny revealed by coevolution of molecules and morphologies', in Patterson, C. (ed.) *Molecules and morphology in evolution: conflict or compromise?*, Cambridge: Cambridge University Press, pp. 141–76.

Goodman, M., Moore, G. W. and Matsuda, G. (1975) 'Darwinian evolution in the genealogy of haemoglobin', *Nature*, **253**, 603–8.

Gu, X. (1997) 'The age of the common ancestor of eukaryotes and prokaryotes: statistical inferences', *Molecular Biology and Evolution*, **14**, 861–6.

Gu, X. (1998) 'Early metazoan divergence was about 830 million years ago', *Journal of Molecular Evolution*, **47**, 369–71.

Haffer, J. (1969) 'Speciation in Amazonian forest birds', *Science*, **165**, 131–7.

Hedges, S. B. (1992) 'The number of replications needed for accurate estimation of the bootstrap p-value in phylogenetic analysis', *Molecular Biology and Evolution*, **9**, 366–9.

Hessler, R. R., Kaharl, V. A. (1995) 'The deep-sea hydrothermal vent community: an overview', in Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. and Thomson, R. E. (eds) *Seafloor hydrothermal systems*, Washington, DC: American Geophysical Union, pp. 72–84.

Hoffman, P. F., Kaufman, A. J., Halverson, G. P. and Schrag, D. P. (1998) 'A Neoproterozoic Snowball Earth', *Science*, **281**, 1342–6.

Janvier, P. (1996) *Early vertebrates*, Oxford: Clarendon Press.

Kasting, J. F. (2000) 'Long-term stability of Earth's climate: the faint young Sun problem revisited', in *Proceedings of the IGBP Workshop on Geosphere-Biosphere Interactions and Climate,* Vatican City: in press.

Kaufman, A. J., Knoll, A. H. and Narbonne, G. M. (1997) 'Isotopes, ice ages, and terminal Proterozoic Earth history', *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 6600–5.

Kennedy, M. J., Runnegar, B., Prave, A. R., Hoffman, K. -H. and Arthur, M. A. (1998) 'Two or four Neoproterozoic glaciations?', *Geology*, **26**, 1059–63.

Kirschvink, J. L. (1992) 'Late Proterozoic low-latitude global glaciation: the Snowball Earth', in Schopf, J. W. and Klein, C. (eds) *The Proterozoic biosphere*, Cambridge: Cambridge University Press, pp. 51–2.

Klicka, J., and Zink, R. M. (1997) 'The importance of recent ice ages in speciation: a failed paradigm', *Science*, **277**, 1666–9.

Knoll, A. H. (1994) 'Neoproterozoic evolution and environmental change', in Bengston, S. (ed.) *Early life on Earth*, New York: Columbia University Press, pp. 439–49.

Kumar, S. (1996) *Phyltest: a program for testing phylogenetic hypotheses*, University Park, Pennsylvania: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.

Kumar, S. and Hedges, S. B. (1998) 'A molecular timescale for vertebrate evolution', *Nature*, **392**, 917–20.

Kumar, S., Tamura, K. and Nei, M. (1993) *MEGA: Molecular Evolutionary Genetic Analysis*, University Park: Pennsylvania State University.

Li, C. -W., Chen, J. -Y. and Hua, T. -E. (1998) 'Precambrian sponges with cellular structures', *Science*, **279**, 879–82.

Lipps, J. H., Bengston, S. and Farmer, J. D. (1992) 'The Precambrian-Cambrian evolutionary transition', in Schopf, J. W. and Klein, C. (eds) *The Proterozoic biosphere*, Cambridge: Cambridge University Press, pp. 453–7.

Løvtrup, S. (1977) *The phylogeny of Vertebrata*, London: John Wiley & Sons.

Mallat, J. (1997a) 'Crossing a major morphological boundary: the origin of jaws in vertebrates', *Zoology*, **100**, 128–40.

Mallat, J. (1997b) 'Hagfish do not resemble ancestral vertebrates', *Journal of Morphology*, **232**, 293.

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.

Maxson, L. R. and Heyer, W. R. (1988) 'Molecular systematics of the frog genus *Leptodactylus* (Amphibia: Leptodactylidae)', *Fieldiana Zoology*, **41**, 1–13.

Maxson, L. R. and Roberts, J. D. (1984) 'Albumin and Australian frogs: molecular data a challenge to speciation model', *Science*, **225**, 957–8.

Mayr, E. (1954) 'Change of genetic environment and evolution', in Huxley, J. S., Hardy, A. C. and Ford, E. B. (eds) *Evolution as a process*, Allen and Unwin: London, pp. 156–80.

Mayr, E. (1963) *Animal species and evolution*, Cambridge, Massachusetts: Harvard University Press.

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

Nei, M. (1996) 'Phylogenetic analysis in molecular evolutionary genetics', *Annual Review of Genetics*, **30**, 371–403.

Nielsen, C. (1995) *Animal evolution: interrelationships of the living phyla*, Oxford: Oxford University Press.

Nikoh, N., Iwabe, N., Kuma, K., Ohno, M., Sugiyama, T., Watanabe, Y., Yasui, K., Shi-cui, Z., Hori, K., Shimura, Y. and Miyata, T. (1997) 'An estimate of divergence time of Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by aldolase and triose phosphate isomerase clocks', *Journal of Molecular Evolution*, **45**, 97–106.

Ota, T. and Nei, M. (1994) 'Estimation of the number of amino acid substitutions per site

when the substitution rate varies among sites', *Journal of Molecular Evolution*, **38**, 642–3.

Paton, R. L., Smithson, T. R. and Clack, J. A. (1999) 'An amniote-like skeleton from the Early Carboniferous of Scotland', *Nature*, **398**, 508–13.

Rasmussen, A.-S. and Arnason, U. (1999) '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., 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.

Runnegar, B. (1982a) 'The Cambrian explosion: animals or fossils?', *Journal of the Geological Society of Australia*, **29**, 395–411.

Runnegar, B. (1982b) 'A molecular-clock date for the origin of the animal phyla', *Lethaia*, **15**, 199–205.

Runnegar, B. (1986) 'Molecular palaeontology', *Palaeontology*, **29**, 1–24.

Rzhetsky, A., Kumar, S. and Nei, M. (1995) 'Four-cluster analysis: a simple method to test phylogenetic hypotheses', *Molecular Biology and Evolution*, **12**, 163–7.

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

Schubart, C. D., Diesel, R. and Hedges, S. B. (1998) 'Rapid evolution to terrestrial life in Jamaican crabs', *Nature*, **393**, 363–5.

Schultze, H.-P. (1994) 'Comparison of hypotheses on the relationships of sarcopterygians', *Systematic Biology*, **43**, 155–73.

Schultze, H.-P. and Trueb, L. (eds) (1991) *Origins of the higher groups of tetrapods: controversy and consensus*, Ithaca, NY: Cornell University Press.

Seilacher, A., Bose, P. K. and Pfluger, F. (1998) 'Triploblastic animals more than 1 billion years ago: trace fossil evidence from India', *Science*, **282**, 80–3.

Stock, D. W. and Whitt, G. S. (1992) 'Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group', *Science*, **257**, 787–9.

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.

Takezaki, N., Rzhetsky, A. and Nei, M. (1995) 'Phylogenetic test of the molecular clock and linearized tree', *Molecular Biology and Evolution*, **12**, 823–33.

Thompson, J. D., Higgins, D. G. and 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 Research*, **22**, 4673–80.

Valentine, J. W., Jablonski, D. and Erwin, D. H. (1999) 'Fossils, molecules and embryos: new perspectives on the Cambrian explosion', *Development*, **126**, 851–9.

Wang, D. Y. -C., Kumar, S. and Hedges, S. B. (1998) 'Divergence time estimates for the early history of animal phyla and the origin of plants, animals, and fungi', *Proceedings of the Royal Society of London* B, **266**, 163–71.

Weiguo, S. (1994) 'Early multicellular fossils', in Bengston, S. (ed.) *Early life on Earth*, New York: Columbia University Press, pp. 358–69.

Wray, G. A., Levinton, J. S. and Shapiro, L. H. (1996) 'Molecular evidence for deep Precambrian divergences', *Science*, **274**, 568–73.

Xiao, S., Zhang, Y. and Knoll, A. H. (1998) 'Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite', *Nature*, **391**, 553–8.