HAL open science

Multigene Analyses of Bilaterian Animals Corroborate the Monophyly of Ecdysozoa, Lophotrochozoa and Protostomia

Herve Philippe, Nicolas Lartillot, Henner Brinkmann

To cite this version:

Herve Philippe, Nicolas Lartillot, Henner Brinkmann. Multigene Analyses of Bilaterian Animals Corroborate the Monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. Molecular Biology and Evolution, 2005, 22 (5), pp.1246-1253. 10.1093/molbev/msi111 . lirmm-00105354

HAL Id: lirmm-00105354

https://hal-lirmm.ccsd.cnrs.fr/lirmm-00105354v1

Submitted on 15 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Multigene Analyses of Bilaterian Animals Corroborate the Monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia

Hervé Philippe, Nicolas Lartillot,[1] *and Henner Brinkmann*

Canadian Institute for Advanced Research and Département de Biochimie, Université de Montréal, Montréal, Québec, Canada

Almost a decade ago, a new phylogeny of bilaterian animals was inferred from small-subunit ribosomal RNA (rRNA) that claimed the monophyly of two major groups of protostome animals: Ecdysozoa (e.g., arthropods, nematodes, onychophorans, and tardigrades) and Lophotrochozoa (e.g., annelids, molluscs, platyhelminths, brachiopods, and rotifers). However, it received little additional support. In fact, several multigene analyses strongly argued against this new phylogeny. These latter studies were based on a large amount of sequence data and therefore showed an apparently strong statistical support. Yet, they covered only a few taxa (those for which complete genomes were available), making systematic artifacts of tree reconstruction more probable. Here we expand this sparse taxonomic sampling and analyze a large data set (146 genes, 35,371 positions) from a diverse sample of animals (35 species). Our study demonstrates that the incongruences observed between rRNA and multigene analyses were indeed due to long-branch attraction artifacts, illustrating the enormous impact of systematic biases on phylogenomic studies. A refined analysis of our data set excluding the most biased genes provides strong support in favor of the new animal phylogeny and in addition suggests that urochordates are more closely related to vertebrates than are cephalochordates. These findings have important implications for the interpretation of morphological and genomic data.

Introduction

The traditional view of bilaterian animal evolution based on morphological and embryological characters proposed that the phylogeny correlates with a gradual increase in complexity (Adoutte et al. 2000). The most simple organisms emerged first, i.e., acoelomates (e.g., platyhelminths) followed by the pseudocoelomates (e.g., nematodes) and then by the true coelomates (e.g., arthropods and chordates). This view was challenged by a careful analysis of small-subunit ribosomal RNA (rRNA) sequences, sampled from a selected set of animals with slowly evolving rRNAs (Aguinaldo et al. 1997). In this so-called new animal phylogeny, some pseudocoelomates, in particular nematodes, were grouped with some coelomates (e.g., arthropods and tardigrades) in the clade Ecdysozoa, whereas other pseudocoelomates (e.g., rotifers) and acoelomates were grouped with the remaining protostomian coelomates (e.g., annelids and molluscs) in the clade Lophotrochozoa. Ecdysozoa and Lophotrochozoa furthermore form a monophyletic assemblage corresponding to Protostomia *sensu lato*.

Although often taken for granted (Adoutte et al. 2000; Graham 2000; Giribet 2002), the new animal phylogeny has only been confirmed by the analyses of Hox genes (de Rosa et al. 1999, but see Telford 2000), horse radish peroxidase (HRP) antibody staining (Haase et al. 2001), large-subunit rRNA (Mallatt and Winchell 2002), and Na/K adenosine triphosphatase (ATPase) (Anderson, Cordoba, and Thollesson 2004). A sequence signature that was initially proposed to support it (Manuel et al. 2000) turned out to be noninformative (Telford 2004).

In sharp contrast, several multigene analyses (Mushegian et al. 1998; Hausdorf 2000; Blair et al. 2002; Korbel et al. 2002; H. Dopazo, Santoyo, and J. Dopazo 2004; Hugues and Friedman 2004; Wolf, Rogozin, and Koonin 2004) provide strong support in favor of the monophyly of Coelomata, with nematodes and platyhelminths emerging at the base of Bilateria. For example, the analysis of 100 proteins from four taxa (44,214 amino acids) supports the grouping of arthropods with vertebrates to the exclusion of nematodes with extremely high statistical support (Blair et al. 2002). The most exhaustive and careful analysis was performed by Wolf, Rogozin, and Koonin (2004), who studied over 500 sets of orthologous proteins from six species. In summary, the monophyly of Ecdysozoa was only supported by a few single-gene phylogenies based on numerous taxa and strongly rejected by analyses based on a few taxa but numerous genes.

The long-branch attraction (LBA) phenomenon (Felsenstein 1978), according to which divergent (hence long branched) but otherwise unrelated taxa tend to cluster together in the estimated phylogeny, is one of the most pervasive tree reconstruction artifacts (Philippe and Laurent 1998). Typically, the branch leading to the out-group, which is by necessity long, attracts long branches of fast-evolving in-group species, so that, in most cases, an LBA results in an artifactual placement of fast-evolving species at the base of the tree. All tree reconstruction methods, because none are based on an entirely correct model of sequence evolution, are sensitive to LBA, although some, especially the probabilistic ones, are more robust (Lockhart et al. 1996). The LBA artifact has played a central role in the inference of the metazoan phylogeny. The fact that the rapid evolutionary rate of nematode rRNAs prevents a reliable placement of this group had been noticed early on (Philippe, Chenuil, and Adoutte 1994), and only the use of a newly sequenced nematode rRNA that evolved more slowly allowed the recovery of the grouping of nematodes and arthropods (Aguinaldo et al. 1997). Therefore, in recent multigene analyses, the hypothesis that the LBA artifact can be responsible for the nonmonophyly of Ecdysozoa was carefully studied. Several approaches (i.e., the use of slowly evolving genes [Blair et al. 2002] or computer simulations [Wolf, Rogozin, and Koonin 2004]) were carried out and seemed to discard this interpretation (but see Copley et al. 2004).

[1] Present address: LIRMM, UMR9928 Université de Montpellier II-CNRS, 161 Rue Ada, 34392 Montpellier Cedex 5, France.

Key words: taxon sampling, phylogenomics, long-branch attraction.

E-mail: herve.philippe@umontreal.ca

Mol. Biol. Evol. 22(5):1246–1253. 2005
doi:10.1093/molbev/msi111
Advance Access publication February 9, 2005

Published by Oxford University Press 2005.

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

However, it should be noted that artifacts such as LBA are systematic, i.e., they tend to be reinforced as more and more data are considered (a property named inconsistency) (Felsenstein 1978; Kim 1996; Lockhart et al. 1996). Multigene analyses are thus expected to be increasingly sensitive to this problem (Phillips, Delsuc, and Penny 2004). There are presently no simple solutions to completely eschew systematic biases, although different approaches have been proposed to reduce their impact (Philippe and Laurent 1998): (1) the use of efficient tree reconstruction methods, (2) the improvement of taxon sampling, and (3) the selection of positions or genes that evolve more slowly.

In the present study, we took advantage of numerous expressed sequence tags and genomic sequencing projects to assemble a very large data set of 146 genes and 49 species. Based on this data set, we demonstrate that LBA indeed affects bilaterian phylogeny and that the basal positioning of platyhelminths and nematodes (i.e., the Coelomata hypothesis) is one of its manifestations. Finally, by a combination of the three methods mentioned above, we show how to overcome the LBA artifact, yielding further molecular support to the Lophotrochozoa-Ecdysozoa hypothesis.

Materials and Methods

To assemble our data set, as detailed in Supplementary Materials, we followed with some modifications the protocol described in Philippe et al. (2004). Most sequences were downloaded from GenBank through National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) except for *Celuca pugilator* (ftp://ftp.genome.ou.edu/pub/fiddlercrab/craball_dir), *Fasciola hepatica* (ftp://ftp.sanger.ac.uk/pub/databases/Trematode/Fhep/), *Fusarium graminearum* (http://www.broad.mit.edu/cgi-bin/annotation/fusarium/download_license.cgi), *Monosiga brevicolis* (King, Hittinger, and Carroll 2003), *Neocallimastix patriciarum* (Brinkmann et al., unpublished data), and *Strongylocentrotus purpuratus* (http://sugp.caltech.edu/ftp_page/). Genes were carefully examined to avoid problems due to hidden paralogy. Importantly, the use of numerous species greatly improves the reliability of orthology assignment. Single-gene phylogenies, shown in Supplementary Materials, were used either to completely discard genes from the analysis for which orthology relationship was difficult to establish (e.g., EF-1α or cytosolic HSP70) or to select the slowest evolving copy of recently duplicated genes (in particular for vertebrates).

Because there is a debate about the relative importance of increasing the number of characters or the number of species to improve phylogenetic accuracy (Hillis et al. 2003), we tried to assemble a data set rich in both species and genes. However, this generally implies allowing for missing or partial sequences of some genes from some species (e.g., the amount of missing data is of 12.5% [Murphy et al. 2001], 20% [Qiu et al. 1999], or 25% [Douzery et al. 2004]). We retained only species for which a sufficiently large number of amino acid residues were available (larger than 6,000). Simulation studies have shown that, under these conditions, the impact of missing data is negligible (Wiens 2003; Philippe et al. 2004). To further verify, analyses without the eight most incomplete sequences were performed (figs. S1 and S2, Supplementary Material online), and, as expected, the results were virtually identical. A large data set comprising 49 species and 146 genes (displaying a mean of ~35% of missing data per species, see Supplementary Materials for more information) was constructed, in which the major animal phyla were represented (echinoderms, urochordates, cephalochordates, vertebrates, arthropods, tardigrades, nematodes, molluscs, annelids, platyhelminths, cnidarians, and ctenophores) as well as two successive outgroups (2 choanoflagellates and 10 diverse fungi). Even if several animal phyla are still not represented (e.g., priapulids, onychophorans, sipunculans, brachiopods, hemichordates) in our alignment consisting of 35,371 positions, it has a much better taxon sampling than any previous multigene analyses of animals, which contained only three to five animal species (Mushegian et al. 1998; Hausdorf 2000; Blair et al. 2002; Wolf, Rogozin, and Koonin 2004).

The phylogeny was inferred by the maximum likelihood (ML) method using concatenated Jones, Taylor, and Thornton (JTT) + F + Γ or separate Whelan and Goldman (WAG) + F + Γ models (Jones, Taylor, and Thornton 1992; Whelan and Goldman 2001). The gamma distribution was used to correct for rate across sites variation and, as expected, significantly improved the fit of the model to the data (from ln L = −908,441 without Γ to ln L = −862,699 with Γ). The separate model (Yang 1996*b*) allows branch lengths and alpha parameter to vary from gene to gene, to take into account heterogeneity of evolutionary rates between genes and lineages. This contrasts with a concatenated model (i.e., considering all the genes as a "supergene") that imposes the same branch lengths and α parameter to all the genes. Despite a serious increase in the number of parameters (16,675 additional parameters), the separate model has a better fit than the concatenated model, according to the Akaike Information Criterion (Akaike 1973) (1,758,748 vs. 1,768,702). When a very large number of positions are used, the problem of local minima is exacerbated because of the height of the potential barriers separating them. We therefore used two approaches. First, a heuristic search was performed with a concatenated JTT + F + Γ model using PHYML (Guindon and Gascuel 2003). For bootstrap analysis (100 replicas), two different starting trees (the BIONJ tree and the PHYML tree obtained for the complete data set) were used to reduce the local minima problem. Second, an exhaustive tree search approach was used by defining several sets of constraints, as explained in the Supplementary Materials, with a separate WAG + F + Γ model. The bootstrap support was computed using the RELL method (Kishino, Miyata, and Hasegawa 1990) based on 1,000 replicates. Only bootstrap values (BVs) obtained for the best-fitting model (separate WAG + F + Γ) are discussed in the text.

To distinguish genes for which nematodes and platyhelminths evolve fast relative to other Bilateria, we computed a distance matrix for each gene with a WAG + F + Γ model using Tree-Puzzle (Schmidt et al. 2002). The evolutionary rate of nematodes and platyhelminths was estimated as the average of all distances between the 16 out-group species and the 15 species of these two groups. We also estimated the evolutionary rates of slowly evolving Bilateria as the average distance between the

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

16 out-group species and the mollusc, the annelid, and the five deuterostome species. Genes were then sorted in decreasing order according to the ratio of these two evolutionary rates. They were removed from the data set five at a time; for each reduced data set, BVs were computed by ML with a separate WAG + F + Γ model. Similar results were obtained if only choanoflagellates, cnidarians, and ctenophores were used as an out-group and/or if nematodes and platyhelminths were considered separately (data not shown).

Results and Discussion

For a set of ~100 evolutionarily conserved orthologous proteins, nematodes and platyhelminths are evolving about two times faster than deuterostomes or arthropods (Philippe et al. 2004). In such a context, where a large data set is used, inconsistent (i.e., strongly supported but erroneous) results are expected to manifest themselves. In particular, the use of a distant out-group (e.g., fungi) should attract nematodes and platyhelminths to the base of the Bilateria. In contrast and provided that the new animal phylogeny is correct, they are expected to emerge higher in the tree when a closer out-group is used because their long branches will no longer be attracted by the now much shorter branch of the out-group.

To test this prediction, we selected subsets of four bilaterian species from our data set of 146 genes and 49 species (see *Materials and Methods*). We followed the first approach to reduce the impact of LBA by inferring the phylogenies by the ML method with a separate WAG + F + Γ model (Yang 1996*a*, 1996*b*), which is among the most efficient tree reconstruction methods currently available. Then, as done in several recent studies (Mushegian et al. 1998; Hausdorf 2000; Blair et al. 2002; Wolf, Rogozin, and Koonin 2004), the yeast *Saccharomyces*, which is quite distant to Bilateria, was used as an out-group. Nematodes and platyhelminths robustly emerged at the base of animals (fig. 1*A*), supporting the monophyly of Coelomata (represented here by *Drosophila* and *Homo*) with a BV of 95%. When the fission yeast *Schizosaccharomyces* and the choanoflagellate *Monosiga* were added to break the very long branch of the out-group (fig. 1*B*), the support for the early emergence of platyhelminths and nematodes decreased markedly (BVs of 43% and 61%). More significantly, when a closer and more slowly evolving out-group, the cnidarian *Hydra*, was added, the topology changed drastically (fig. 1*C*); the nematodes were now a sister-group of platyhelminths (BV of 80%). This group clustered with arthropods, recovering the monophyly of protostomes with high support (98%). These results fit perfectly the prediction that an LBA artifact, caused by the use of a too distant out-group, underlies the early emergence of nematodes and platyhelminths found in previous multigene analyses (Mushegian et al. 1998; Hausdorf 2000; Blair et al. 2002; Wolf, Rogozin, and Koonin 2004).

The addition of a cnidarian essentially prevented the attraction between nematodes/platyhelminths and the out-group but yielded a surprising and somewhat disquieting result: nematodes were now the sister-group of platyhelminths to the exclusion of arthropods, a grouping that has never been proposed and that does not make any biological sense. We suggest that this grouping is also due to an

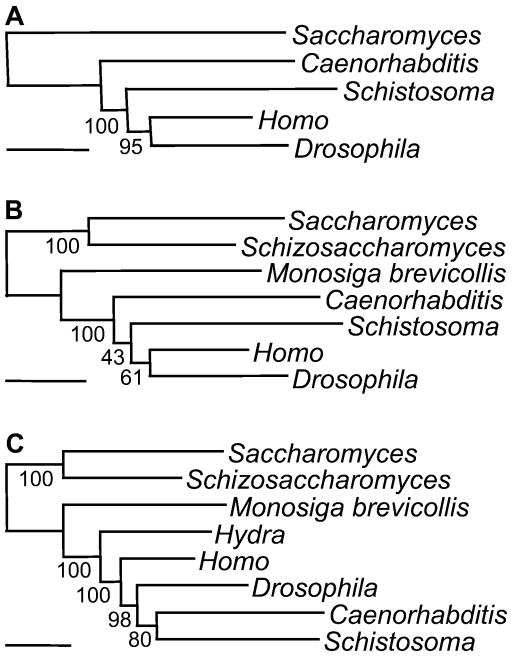

FIG. 1.—LBA and the effect of the out-group. Trees were inferred with an ML method using a separate WAG + F + Γ model. The in-group species remain identical, and the out-group is (*A*) a distantly related yeast, (*B*) two yeasts and a more closely related choanoflagellate, *Monosiga*, or (*C*) two yeasts, a choanoflagellate, and a very closely related cnidarian, *Hydra*. The early emergence of nematodes (*Caenorhabditis*) and platyhelminths (*Schistosoma*) is due to an LBA artifact that disappears when a close out-group is used. BVs are indicated to the left of each node. The scale bar represents 0.1 substitutions per site for a unit branch length.

LBA artifact, this time between the two fastest evolving in-groups (i.e., nematodes and platyhelminths). We therefore applied the second approach to reduce the LBA, i.e., the addition of many taxa to break long branches (Hendy and Penny 1989). We compiled sequences mainly from cDNA-sequencing projects and obtained a data set rich in both species (49) and genes (146) (displaying on average ~35% of missing data; see table S1 [Supplementary Material online] for the detailed distribution of missing data among species), in which most major animal phyla were represented, as well as two successive out-groups (choanoflagellates and fungi).

The ML phylogeny based on this extended data set (fig. 2) was in excellent agreement with the current knowledge. All undisputed groups were strongly supported (e.g., monophyly of ascomycetes, basidiomycetes, choanoflagellates, Bilateria, deuterostomes, nematodes, platyhelminths, arthropods, and insects). The structure of the bilaterian tree was very similar to the new phylogeny of animals (e.g., monophyly of protostomes), with the exception of platyhelminths, which were again a sister-group of nematodes instead of being clustered with the other Lophotrochozoa (represented here by annelids and molluscs). The addition of 42 taxa (from 7 in fig. 1*C* to 49 in fig. 2) was therefore not sufficient to eliminate the attraction between nematodes and platyhelminths. However, it should be noted that many of the added taxa (in particular the nine nematodes and the four platyhelminths) were fast evolving, and the addition of fast-evolving lineages may in fact exacerbate the inconsistency due to the LBA artifact (Kim 1996; Poe 2003).

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

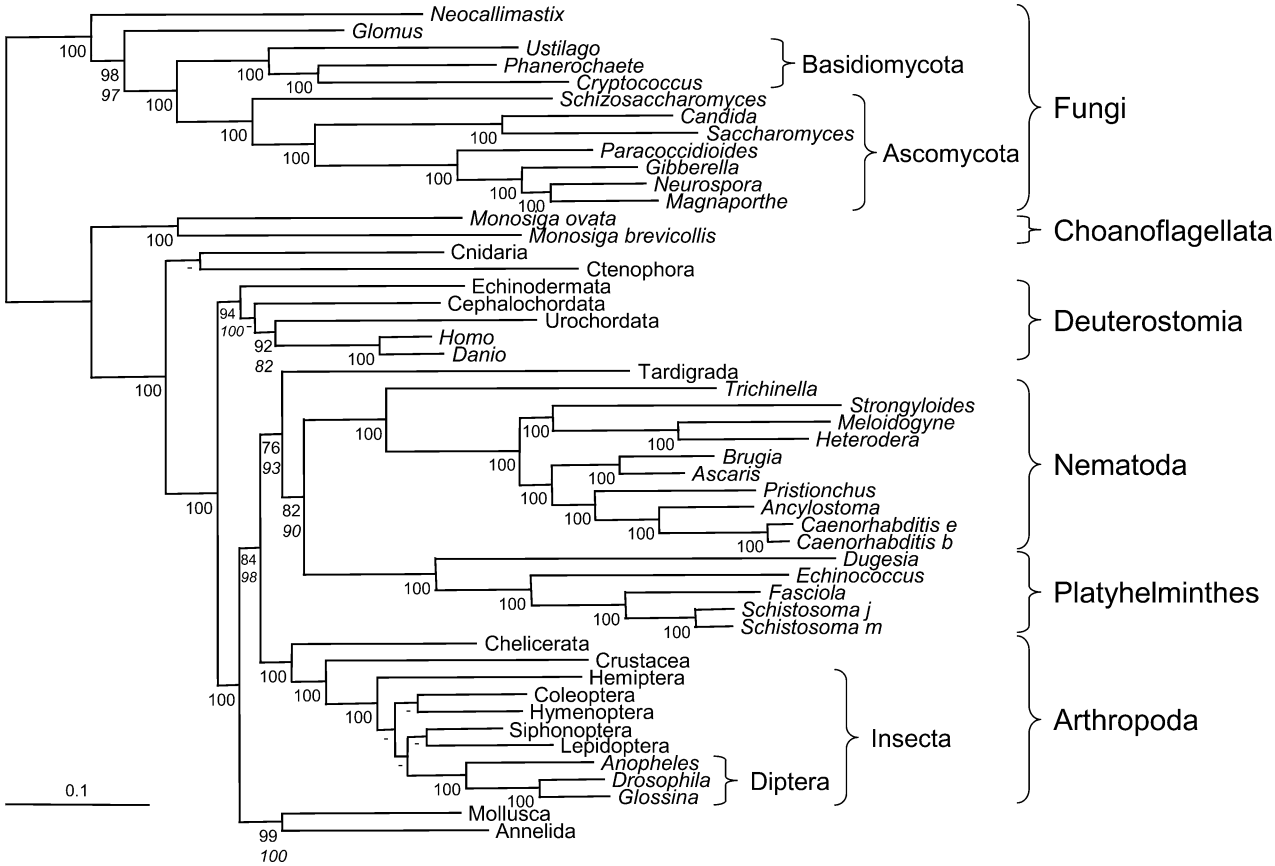

FIG. 2.—Tree based on 146 genes (35,371 amino acid positions). Trees were inferred with an ML method. The same topology (except for a few not supported nodes) was obtained using either a separate WAG + F + Γ model or a concatenated JTT + F + Γ model. The values indicated correspond to bootstrap support values of the separate (upper) or concatenated (lower, in italic) models. When both are equal to 100%, only the first one is indicated, and when at least one is below 75%, the node is indicated by a hyphen.

We further explored the issue of species sampling by specifically and independently removing nematodes and platyhelminths because they constitute potential major attractors. When platyhelminths were removed (fig. S3, Supplementary Material online), the topology remained exactly the same but the support was higher, suggesting that Ecdysozoa are monophyletic. By contrast, when nematodes were removed (fig. S4, Supplementary Material online), a single but major topological change occurred. Platyhelminths moved from inside the Ecdysozoa to a sister-group position of molluscs + annelids. Interestingly, the monophyly of Lophotrochozoa was recovered with a high support (BV of 95%). No obvious artifact could explain this result because a fast-evolving lineage was clustered with two slowly evolving ones. These analyses strongly suggest that the fast-evolving nematodes constitute a potent attractor to platyhelminths.

The drastic approach used above has the disadvantage that all groups of interest cannot be present simultaneously in our analysis. We therefore turned to an alternative method, and selected the slowest evolving taxa among nematodes and platyhelminths, as previously done in the case of rRNA (Aguinaldo et al. 1997). Interestingly, the inferred phylogeny is now identical to the new animal phylogeny (fig. S5, Supplementary Material online). However, even if the monophyly of protostomes remained highly supported (BV of 100%), support for the monophyly of Ecdysozoa and Lophotrochozoa was weak (BV of 55%).

Because artifactual attraction between nematodes and platyhelminths appears to be extremely strong, we applied the third approach to overcome LBA, i.e., the use of slowly evolving characters. We proceeded by the selective elimination of the most biased genes, made feasible by the large size of our data set, with the hope that some of the 146 genes used here are sufficiently slowly evolving for both nematodes and platyhelminths. For each gene, we computed the ratio of the mean evolutionary rates of nematodes and platyhelminths to those of short-branch organisms, i.e., annelids, deuterostomes, and molluscs. As expected from the branch lengths of figure 2, these ratios were greater than 1 for the vast majority of genes (fig. 3*A*). Genes with the highest ratio would be expected to contribute most to the artifactual grouping of nematodes and platyhelminths. We therefore progressively removed these genes and recomputed the phylogeny. As shown in figure 3*B*, bootstrap support for the grouping of nematodes and platyhelminths decreased continuously, reaching <10% when 75 genes were discarded. Remarkably, in parallel, support for the monophyly of both Ecdysozoa and Lophotrochozoa increased steadily, up to 90%–95%. It should be noted that the monophyly of protostomes was virtually unaffected by gene removal (BV around 100%), until more than 100 genes

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

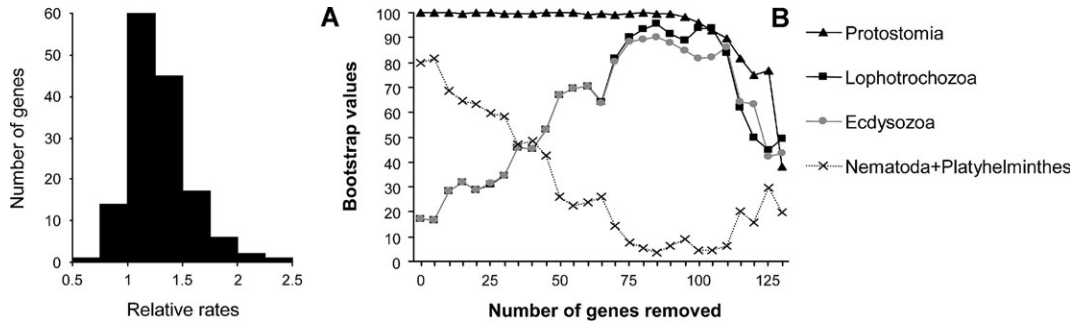

FIG. 3.—Evolutionary rates, gene removal, and the new animal phylogeny. For each of the 146 genes, the ratio of the mean evolutionary rate in nematodes and platyhelminths to the mean evolutionary rate in annelids, molluscs, and deuterostomes is displayed (*A*). A ratio higher than one indicates that nematodes and platyhelminths evolve faster than annelids, molluscs, and deuterostomes. Only 14 genes have a ratio below one. Genes with the highest ratio are removed five at a time, and the evolution of BVs for the four nodes of interest is monitored (*B*). The monophyly of Protostomia is used to indicate when gene removal leads to a significant decrease in the phylogenetic signal.

were discarded. At this stage, no sufficient phylogenetic signal is present in the remaining genes (fewer than 40) and, in consequence, the support decreased for all the nodes (data not shown). To reduce the impact of the LBA artifact without decreasing the resolution too significantly, the removal of 75 genes was an acceptable trade-off. The phylogeny based on the 71 remaining genes (20,705 positions, fig. 4) was in excellent agreement with the new animal phylogeny: the Ecdysozoa and the Lophotrochozoa were both monophyletic (BVs of 87% and 88%) and formed together the clade Protostomia (BV of 100%). In conclusion, the combination of the three approaches given above (i.e., efficient tree reconstruction method, large species sampling, and selection of slowly evolving features) was necessary to overcome the artifactual attraction of platyhelminths by nematodes.

The position of urochordates as the sister-group of vertebrates to the exclusion of cephalochordates (BVs of 92% and 97% in figs. 2 and 4) deserves special attention. Although this grouping has been proposed by Jefferies (1986) based on the interpretation of unusual fossils called mitrates and cornutes, the current consensus favors the alternative grouping of cephalochordates and vertebrates, following the seminal work of Garstang (1928). However, this consensus relies neither on particularly strong morphological evidence nor on molecular evidence (Oda et al. 2002; Winchell et al. 2002; Mallatt and Chen 2003), and the great similarities between cephalochordates and vertebrates probably represent chordate symplesiomorphies. Our results seemed to be robust, particularly when considering that urochordates were fast evolving and should therefore be attracted to the base of the deuterostomes by an LBA artifact but not toward the slowly evolving vertebrates. However, possible inconsistency of tree reconstruction when few species are used (Philippe and Laurent 1998) (here only five deuterostomes) argues for caution before making any firm conclusions. At any rate, the grouping of urochordates with vertebrates constitutes a reasonable working hypothesis that must be tested with a much larger deuterostomian taxon sampling. The migratory neural crestlike cells, recently found in the ascidian urochordate *Ecteinascidia turbinate* (Jeffery, Strickler, and Yamamoto 2004), could potentially constitute a synapomorphy for this hypothetical group.

From a methodological point of view, apart from the question of the position of nematodes and platyhelminths, the phylogeny of opisthokonts is well recovered by our multigene analyses, as evidenced by its good agreement with morphologically based trees. This indicates, first, that an ancient phylogenetic signal is still present in the genomic data, even for the deepest nodes of the tree and, second, that current tree reconstruction methods are rather efficient, provided some care is brought to reduce the impact of potential artifacts. For instance, the use of a close out-group is sufficient for current ML methods to discover the LBA between fungi and fast-evolving bilaterians (fig. 1).

However, more difficult phylogenetic issues, such as the position of platyhelminths and nematodes, highlight the limits of the currently available procedures and point to the urgent need for better tree reconstruction methods, in particular through the development of better models of sequence evolution. Heterotachy (shifts in position-specific evolutionary rates) has been recently proposed as an important cause of these limitations (Lockhart et al. 1996; Philippe and Germot 2000; Inagaki et al. 2004; Kolaczkowski and Thornton 2004). In this respect, the separate model, allowing branch lengths to be different for each gene, deals with a particular case of heterotachy, that between genes. In addition, it is based on a partitioning of the data and is thus particularly well suited to the kind of situations investigated by Kolaczkowski and Thornton (2004). We indeed found a significant level of heterotachy between genes, as demonstrated by the better fit to the data of the separate model over the concatenated one. However, its impact on the phylogeny is limited because the very same reconstructions are obtained with both separate and concatenated models (data not shown), suggesting that, more generally, heterotachy might not be a major cause of phylogenetic artifacts. Nevertheless, apart from heterotachy, quite a few other model violations, such as the nonindependence of sites or the nonstationarity of the evolutionary process, could be an important source of systematic errors in tree reconstructions.

At any case, until better tree reconstruction methods are available, the specific removal of the data that are the most responsible for the tree reconstruction artifact (here, the fastest evolving genes) will surely constitute a simple and efficient heuristic approach to improve the

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

FIG. 4.—ML tree inferred from the separate analysis of 71 genes that evolve slowly in nematodes and platyhelminths (20,705 amino acid positions). For phylogenetic methods, see legend of figure 2. It should be noted that the use of slowly evolving genes alone is not sufficient to overcome the LBA artifact. If the same taxon sampling as in figure 1*A* and *B* was used, nematodes and platyhelminths still artifactually emerged at the base of the Bilateria (fig. S6, Supplementary Material online). The monophyly of Ecdysozoa was only recovered when the cnidarian sequence was used as an out-group, however, with a weak support (fig. S6C, Supplementary Material online). A reduced support, when only a few species were used (51% instead of 87% here), demonstrates the important effect of a large species sampling. The support for grouping nematodes and tardigrades to the exclusion of arthropods decreased from 76% to 65% when the fast-evolving nematode genes were removed. This suggests that this weakly supported grouping (instead of the expected sister-group relationship of arthropods and tardigrades) is rather the result of an LBA artifact, in agreement with the very long branches of these two groups. In fact, the monophyly of Ecdysozoa could be the result of an LBA artifact because arthropods, nematodes, and tardigrades were all fast evolving. To test this hypothesis, only the slowest evolving arthropod (chelicerate) was retained. Even in this case, the fast-evolving nematodes still clustered with arthropods (fig. S7, Supplementary Material online) with high support. Therefore, the monophyly of Ecdysozoa is most likely correct, albeit LBA could potentially increase its support (Siddall 1998).

accuracy of inferred phylogenies. Such data removal methods (Brinkmann and Philippe 1999; Lopez, Forterre, and Philippe 1999; Pisani 2004) are well suited for phylogenomic analyses because the remaining data set is sufficiently large to yield highly supported results.

In summary, three lines of evidence argue in favor of the new animal phylogeny (fig. 4) and suggest that the grouping of nematodes and platyhelminths (figs. 1*C* and 2) is the result of an LBA artifact. The monophyly of Ecdysozoa and Lophotrochozoa was recovered when (1) the two attractors were separately discarded (figs. S3 and S4, Supplementary Material online), (2) the two attractors were represented each by the slowest evolving representative (fig. S5, Supplementary Material online), and (3) the genes for which the two attractors evolved the fastest were removed (fig. 4). Several independent lines of molecular evidence (rRNA [Aguinaldo et al. 1997; Mallatt and Winchell 2002], Hox cluster [de Rosa et al. 1999], HRP staining [Haase et al. 2001], Na/K ATPase [Anderson, Cordoba, and Thollesson 2004], and our 71 protein-encoding genes) now support the hypothesis of the new animal phylogeny.

Strikingly, this phylogeny has until now received only limited support from a morphological or embryological point of view. The most commonly cited characters are the molt in Ecdysozoa, which relies on a partially conserved hormonal triggering pathway (Gissendanner et al. 2004), spiral cleavage for Lophotrochozoa (but see Anderson 1973), and the fate of the blastopore in Protostomia, although this last character might not be reliable, many protostomes having in fact a deuterostomous (brachiopods) or an amphiostomous (annelids) gastrulation (Nielsen 2001). Nevertheless, it should be kept in mind that most of the organisms for which development is well studied are highly derived (e.g., *Caenorhabditis* or *Drosophila*), which could have blurred ancient characteristics, and a more detailed analysis of molecular mechanisms of development and body-plan formation, in particular in less derived groups (e.g., priapulids and onychophorans), might reveal more striking shared derived developmental mechanisms supporting these clades.

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

Note

After the acceptance of this manuscript, Philip et al. published a work on the very same subject (G. K. Philip, C. J. Creevey, and J. O. McInerney. 2005. The Opisthokonta and the Ecdysozoa may not be Clades: Stronger Support for the Grouping of Plant and Animal than for Animal and Fungi and Stronger Support for the Coelomata than Ecdysozoa. *Mol. Biol. Evol.*, doi:10.1093/molbev/msi102). Based on an analysis of 780 single-copy genes from 10 species, they proposed that both Ecdysozoa and Opisthokonta are not monophyletic. We believe that their results concerning Ecdysozoa are due to a long branch attraction artifact not correctly handled by the current tree reconstruction methods. This artifact is especially problematic since their taxonomic sampling is too limited, as illustrated in our Figure 1. The nonmonophyly of Opisthokonta is more puzzling. First, contrary to the claims of Philip et al., the plant-animal-fungi relationships have been tested with more than 23 proteins, since we addressed precisely this question using 129 genes (Philippe et al. 2004. Phylogenomics of Eukaryotes: Impact of Missing Data on Large Alignments. *Mol. Biol. Evol.* **9**:1740–1752). Second, using this data-set we found an extremely high support for the monophyly of opisthokonts, a signal that is recovered whatever the species sampling and the tree reconstruction methods used. Indeed, with the same species as Philip et al. except for Mus and Anopheles, the topologies that do not include the clade Opisthokonta (numbers 1–3 and 7–9 in Table 3 of Philip et al.) were significantly rejected by the AU test (p-value between 2×10^{-45} and 3×10^{-113}, with a WAG + F + Γ model; Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**: 492−508). Additional work is required to understand the reasons for these conflicting results.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online (www.mbe.oupjournals.org).

Acknowledgments

We thank Frédéric Delsuc, Peter Holland, Franz Lang, Denis Lavrov, David Moreira, Miklòs Müller, and Nicolas Rodrigue for critical comments on the manuscript. We acknowledge the contributions of genome and cDNA projects that have generated some sequences used in these analyses. *Celuca pugilator* sequences were obtained by Doris Kupfer, Sunkyoung So, Yuhong Tang, Ieva Zumbakyte, Bruce Roe, and David Durica in the Department of Chemistry and Biochemistry and the Department of Zoology at the University of Oklahoma; *Fasciola hepatica* sequence data were produced by the *Schistosoma mansoni* Sequencing Group at the Sanger Institute and can be obtained from ftp://ftp.sanger.ac.uk/pub/databases/Trematode/Fhep/; *Fusarium graminearum* sequences were generated by the Sequencing Project, Center for Genome Research (http://www.broad.mit.edu); *Cryptococcus neoformans* data from *C. neoformans* Sequencing Project, NIH-NIAID grant number AI147079, and Bruce A. Roe, Doris Kupfer, Heather Bell, Sun So, Yuong Tang, Jennifer Lewis, Sola Yu, Kent Buchanan, Dave Dyer, and Juneann Murphy were supported by an NIH-NIAID grant number AI147079; *C. neoformans* genome data courtesy of the Stanford Genome Technology Center, funded by the NIH-NIAID under cooperative agreement AI47087, and The Institute for Genomic Research, funded by the NIH-NIAID under cooperative agreement U01 AI48594. H.P. was supported by Canada Research Chair Program, the Université de Montréal, and a Bioinformatics Grant of Génome Québec.

Literature Cited

Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prud' homme, and R. de Rosa. 2000. The new animal phylogeny: reliability and implications. Proc. Natl. Acad. Sci. USA **97**:4453–4456.

Aguinaldo, A. M., J. M. Turbeville, L. S. Linford, M. C. Rivera, J. R. Garey, R. A. Raff, and J. A. Lake. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature **387**:489–493.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pp. 267–281 *in* B. N. Petrov and F. Csaki, eds. Proceedings of the 2nd International Symposium on Information Theory. Akademia Kiado, Budapest, Hungary.

Anderson, D. T. 1973. Embryology and phylogeny in annelids and arthropods. Pergamon Press, Oxford.

Anderson, F. E., A. J. Cordoba, and M. Thollesson. 2004. Bilaterian phylogeny based on analyses of a region of the sodium-potassium ATPase beta-subunit gene. J. Mol. Evol. **58**: 252–268.

Blair, J. E., K. Ikeo, T. Gojobori, and S. B. Hedges. 2002. The evolutionary position of nematodes. BMC Evol. Biol. **2**:7.

Brinkmann, H., and H. Philippe. 1999. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol. Biol. Evol. **16**:817–825.

Copley, R. R., P. Aloy, R. B. Russell, and M. J. Telford. 2004. Systematic searches for molecular synapomorphies in model metazoan genomes give some support for Ecdysozoa after accounting for the idiosyncrasies of *Caenorhabditis elegans*. Evol. Dev. **6**:164–169.

de Rosa, R., J. K. Grenier, T. Andreeva, C. E. Cook, A. Adoutte, M. Akam, S. B. Carroll, and G. Balavoine. 1999. Hox genes in brachiopods and priapulids and protostome evolution. Nature **399**:772–776.

Dopazo, H., J. Santoyo, and J. Dopazo. 2004. Phylogenomics and the number of characters required for obtaining an accurate phylogeny of eukaryote model species. Bioinformatics **20**:i116–i121.

Douzery, E. J., E. A. Snell, E. Bapteste, F. Delsuc, and H. Philippe. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. USA **101**:15386–15391.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**: 401–410.

Garstang, W. 1928. The morphology of the tunicata and its bearing on the phylogeny of the Chordata. Q. J. Microsc. Sci. **72**: 51–187.

Giribet, G. 2002. Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion? Mol. Phylogenet. Evol. **24**:345–357.

Gissendanner, C. R., K. Crossgrove, K. A. Kraus, C. V. Maina, and A. E. Sluder. 2004. Expression and function of conserved nuclear receptor genes in *Caenorhabditis elegans*. Dev. Biol. **266**:399–416.

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021

Graham, A. 2000. Animal phylogeny: root and branch surgery. Curr. Biol. **10**:R36–R38.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**:696–704.

Haase, A., M. Stern, K. Wachtler, and G. Bicker. 2001. A tissue-specific marker of Ecdysozoa. Dev. Genes Evol. **211**:428–433.

Hausdorf, B. 2000. Early evolution of the Bilateria. Syst. Biol. **49**:130–142.

Hendy, M., and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. **38**:297–309.

Hillis, D. M., D. D. Pollock, J. A. McGuire, and D. J. Zwickl. 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. **52**:124–126.

Hugues, A. L., and R. Friedman. 2004. Differential loss of ancestral gene families as a source of genomic divergence in animals. Proc. R. Soc. Lond. B **271**(Suppl.):S107–S109.

Inagaki, Y., E. Susko, N. M. Fast, and A. J. Roger. 2004. Covarion shifts cause a long-branch attraction artifact that unites Microsporidia and Archaebacteria in EF-1α phylogenies. Mol. Biol. Evol. **21**:1340–1349.

Jefferies, R. P. S. 1986. The ancestry of the vertebrates. British Museum (Natural History), London.

Jeffery, W. R., A. G. Strickler, and Y. Yamamoto. 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature **431**:696–699.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Bio. Sci. **8**:275–282.

Kim, J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. Syst. Biol. **45**:363–374.

King, N., C. T. Hittinger, and S. B. Carroll. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. Science **301**:361–363.

Kishino, H., T. Miyata, and M. Hasegawa. 1990. Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. J. Mol. Evol. **31**:151–160.

Kolaczkowski, B., and J. W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature **431**:980–984.

Korbel, J. O., B. Snel, M. A. Huynen, and P. Bork. 2002. SHOT: a web server for the construction of genome phylogenies. Trends Genet. **18**:158–162.

Lockhart, P. J., A. W. Larkum, M. Steel, P. J. Waddell, and D. Penny. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. USA **93**:1930–1934.

Lopez, P., P. Forterre, and H. Philippe. 1999. The root of the tree of life in the light of the covarion model. J. Mol. Evol. **49**: 496–508.

Mallatt, J., and J. Y. Chen. 2003. Fossil sister group of craniates: predicted and found. J. Morphol. **258**:1–31.

Mallatt, J., and C. J. Winchell. 2002. Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. Mol. Biol. Evol. **19**:289–301.

Manuel, M., M. Kruse, W. E. Muller, and Y. Le Parco. 2000. The comparison of beta-thymosin homologues among Metazoa supports an arthropod-nematode clade. J. Mol. Evol. **51**:378–381.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001. Molecular phylogenetics and the origins of placental mammals. Nature **409**: 614–618.

Mushegian, A. R., J. R. Garey, J. Martin, and L. X. Liu. 1998. Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. Genome Res. **8**:590–598.

Nielsen, C. 2001. Animal evolution, interrelationships of the living phyla. Oxford University Press, Oxford.

Oda, H., H. Wada, K. Tagawa, Y. Akiyama-Oda, N. Satoh, T. Humphreys, S. Zhang, and S. Tsukita. 2002. A novel amphioxus cadherin that localizes to epithelial adherens junctions has an unusual domain organization with implications for chordate phylogeny. Evol. Dev. **4**:426–434.

Philippe, H., A. Chenuil, and A. Adoutte. 1994. Can the Cambrian explosion be inferred through molecular phylogeny? Development **120**:S15–S25.

Philippe, H., and A. Germot. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol. Biol. Evol. **17**:830–834.

Philippe, H., and J. Laurent. 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. **8**:616–623.

Philippe, H., E. A. Snell, E. Bapteste, P. Lopez, P. W. Holland, and D. Casane. 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol. Biol. Evol. **21**:1740–1752.

Phillips, M. J., F. Delsuc, and D. Penny. 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. **21**:1455–1458.

Pisani, D. 2004. Identifying and removing fast evolving sites using compatibility analysis: an example from the Arthropoda. Syst. Biol. **53**:978–989.

Poe, S. 2003. Evaluation of the strategy of long-branch subdivision to improve the accuracy of phylogenetic methods. Syst. Biol. **52**:423–428.

Qiu, Y. L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, E. A. Zimmer, Z. Chen, V. Savolainen, and M. W. Chase. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature **402**:404–407.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502–504.

Siddall, M. E. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris zone. Cladistics **14**:209–220.

Telford, M. J. 2000. Turning Hox "signatures" into synapomorphies. Evol. Dev. **2**:360–364.

———. 2004. The multimeric beta-thymosin found in nematodes and arthropods is not a synapomorphy of the Ecdysozoa. Evol. Dev. **6**:90–94.

Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. Mol. Biol. Evol. **18**:691–699.

Wiens, J. J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst. Biol. **52**:528–538.

Winchell, C. J., J. Sullivan, C. B. Cameron, B. J. Swalla, and J. Mallatt. 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol. Biol. Evol. **19**:762–776.

Wolf, Y. I., I. B. Rogozin, and E. V. Koonin. 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. Genome Res. **14**:29–36.

Yang, Z. 1996*a*. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. **11**:367–370.

Yang, Z. 1996*b*. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. **42**:587–596.

Mark Springer, Associate Editor

Accepted January 25, 2005

Downloaded from https://academic.oup.com/mbe/article/22/5/1246/1066923 by Bibliothèque Universitaire de médecine - Nîmes user on 15 June 2021