genesis 46:592–604 (2008)

ARTICLE

Additional Molecular Support for the New Chordate Phylogeny

Frédéric Delsuc,[1,2]* Georgia Tsagkogeorga,[1,2] Nicolas Lartillot,[3] and Hervé Philippe[3]

[1]Université Montpellier II, CC064, Place Eugène Bataillon, Montpellier Cedex 5, France
[2]CNRS, Institut des Sciences de l'Evolution (UMR5554), CC064, Place Eugène Bataillon, Montpellier Cedex 5, France
[3]Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal, Québec, Canada

Received 6 May 2008; Revised 16 September 2008; Accepted 22 September 2008

Summary: Recent phylogenomic analyses have suggested tunicates instead of cephalochordates as the closest living relatives of vertebrates. In direct contradiction with the long accepted view of Euchordates, this new phylogenetic hypothesis for chordate evolution has been the object of some skepticism. We assembled an expanded phylogenomic dataset focused on deuterostomes. Maximum-likelihood using standard models and Bayesian phylogenetic analyses using the CAT site-heterogeneous mixture model of amino-acid replacement both provided unequivocal support for the sister-group relationship between tunicates and vertebrates (Olfactores). Chordates were recovered as monophyletic with cephalochordates as the most basal lineage. These results were robust to both gene sampling and missing data. New analyses of ribosomal rRNA also recovered Olfactores when compositional bias was alleviated. Despite the inclusion of 25 taxa representing all major lineages, the monophyly of deuterostomes remained poorly supported. The implications of these phylogenetic results for interpreting chordate evolution are discussed in light of recent advances from evolutionary developmental biology and genomics. genesis 46:592–604, 2008. **© 2008 Wiley-Liss, Inc.**

Key words: phylogenomics; deuterostomes; chordates; tunicates; cephalochordates; olfactores; ribosomal RNA; jackknife; evolution

INTRODUCTION

Besides its fundamental role in systematics, phylogenetic reconstruction is a prerequisite for understanding the evolution of organisms. The essential contribution of phylogenetics for understanding morphological diversity has perhaps been best exemplified in the case of animal evolution (Telford and Budd, 2003). The Cambrian explosion has produced a bewildering diversity of body plans whose origins and evolution can only be apprehended by undertaking an integrative approach through evolutionary developmental biology (Evo-Devo) (Conway-Morris, 2003). The knowledge of phylogenetic relationships, by allowing the polarization of character transformations, sheds light on the extent of morphological convergence and reversal. A phylogenetic framework is therefore required for distinguishing ancestral characters from those representing morphological innovations. Comparative genomics is now providing the opportunity to track these morphological innovations back to the molecular level by revealing the patterns of gene acquisition/loss and giving clues to the molecular adaptations that underline the evolution of body plans (Cañestro *et al.*, 2007).

Animal taxonomy has deep roots. The study of morphological and embryological characters has allowed the definition of the major phyla but left their interrelationships almost unresolved (Nielsen, 2001). The advent of molecular data during the 1990s has revolutionized the traditional classification through a series of phylogenetic analyses of the 18S ribosomal RNA (rRNA) gene for an ever increasing number of key taxa (Aguinaldo *et al.*, 1997; Halanych *et al.*, 1995). This period culminated with the proposition of a new view of animal phylogeny at odds with the traditional paradigm of a steady increase toward morphological complexity, and revealing instead the major role played by secondary simplification from complex ancestors (Adoutte *et al.*, 2000; Lwoff, 1944). Despite these undeniable achievements, the resolving power provided by 18S rRNA and other single genes is nevertheless limited, and a number of open questions in animal phylogeny remained to be answered (Halanych, 2004).

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Frédéric Delsuc, CC064, Institut des Sciences de l'Evolution, UMR5554-CNRS, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. E-mail: frederic.delsuc@univ-montp2.fr

Contract grant sponsor: Contribution ISEM 2008-062 of the Institut des Sciences de l'Evolution (UMR5554-CNRS); Contract grant sponsor: Research Networks Program in Bioinformatics from the High Council for Scientific and Technological Cooperation between France and Israel

Published online 10 November 2008 in
Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/dvg.20450

The most recent advances in animal phylogeny have come from phylogenomics (Delsuc *et al*., 2005), which considerably increases the resolving power by considering numerous concatenated genes from expressed sequence tags (ESTs) and complete genome projects (Philippe and Telford, 2006). Despite some troubled beginnings due to the shortcomings of using only a restricted set of taxa (Philippe *et al*., 2005a), phylogenomics has provided strong corroborating support for the new animal phylogeny, essentially confirming the monophyly of Protostomia, Ecdysozoa, and Lophotrochozoa (Baurain *et al*., 2007; Dunn *et al*., 2008; Lartillot and Philippe, 2008; Philippe *et al*., 2005b). Phylogenomics has also helped solving some longstanding mysteries such as the position of chaetognaths, which finally appear to belong to Protostomia (Marletaz *et al*., 2006; Matus *et al*., 2006) and also proposed unexpected phylogenetic affinities for enigmatic taxa such as *Buddenbrockia plumatellae* recently unmasked as a cnidarian worm (Jimenez-Guri *et al*., 2007), or *Xenoturbella bocki*, representing a fourth deuterostome phylum on its own (Bourlat *et al*., 2006).

Among the most groundbreaking results from recent phylogenomic studies was the identification of tunicates (or urochordates) as the closest living relatives of vertebrates, instead of cephalochordates as traditionally accepted (Delsuc *et al*., 2006). Some hints of this unexpected result had been observed in previous large-scale phylogenetic studies including a single tunicate representative (Blair and Hedges, 2005; Philippe *et al*., 2005b; Vienne and Pontarotti, 2006). However, a substantial increase in taxon sampling turned out to be required for recovering convincing support in favor of such an unorthodox relationship. In particular, the fact that the inclusion of the divergent appendicularian tunicate *Oikopleura dioica* did not disrupt the sister-group relationship between tunicate and vertebrates gave a good indication about the strength of the phylogenetic signal in its favor (Delsuc *et al*., 2006). The grouping of tunicates and vertebrates had already been proposed on morphological grounds by Richard P.S. Jefferies who coined the name Olfactores after the presence a putatively homologous olfactory apparatus in fossils that were proposed to be precursors of tunicates and vertebrates (Jefferies, 1991). This phylogenetic result has nevertheless been the object of some skepticism. One reason for this maybe that it further invalidates the traditional textbook view of chordate evolution as a steady increase toward morphological complexity culminating with vertebrates, as betrayed by the use of the term Euchordates (literally "true chordates") for denoting the grouping of cephalochordates and vertebrates (Gee, 2001). The lack of obvious morphological synapomorphies for Olfactores, apart from the presence of migratory neural crest-like cells (Jeffery, 2007; Jeffery *et al*., 2004), and the apparent conflict with analyses of rRNA data which tend to favor Euchordates (Cameron *et al*., 2000; Mallatt and Winchell, 2007; Winchell *et al*., 2002) might also partly explain the caution with which this result has been considered at first.

Phylogenomics, despite being a powerful approach, is however not immune to potential reconstruction artifacts. The possible pitfalls associated with phylogenomic studies include systematic errors that can be traced back to some kind of model misspecifications (Philippe *et al*., 2005a) and caused mainly by heterogeneity of evolutionary rates among taxa (Lartillot *et al*., 2007; Philippe *et al*., 2005b) and compositional bias (Blanquart and Lartillot, 2008; Jeffroy *et al*., 2006; Lartillot and Philippe, 2008; Phillips *et al*., 2004). Empirical protocols have been designed to detect and reduce the impact of systematic error in genome-scale studies (Rodríguez-Ezpeleta *et al*., 2007) but the ultimate solution lies in the development of improved models of sequence evolution (Felsenstein, 2004; Philippe *et al*., 2005a; Steel, 2005). The reliance of current phylogenomic studies on a relatively limited number of highly expressed genes (Philippe and Telford, 2006) and the potential impact of missing data on phylogenomic inference (Hartmann and Vision, 2008; Philippe *et al*., 2004) are also regularly cited as limitations of the phylogenomic approach.

The aim of this article is to evaluate the current evidence for the new chordate phylogeny by: (1) reanalyzing previous phylogenomic data using improved models of amino-acid replacement, (2) assembling and analyzing an updated phylogenomic dataset with more genes and more taxa, (3) assessing the impact of missing data and gene sampling on phylogenomic results, and (4) performing new analyses of rRNA data taking compositional bias into account.

MATERIALS AND METHODS

Phylogenomic Dataset Assembly

We built upon previous phylogenomic datasets assembled in the Philippe lab (Delsuc *et al*., 2006; Jimenez-Guri *et al*., 2007; Lartillot and Philippe, 2008; Philippe *et al*., 2004, 2005b) to select a set of 179 orthologous markers showing sufficient conservation across metazoans to be useful for inferring the phylogeny of metazoans. Alignments were built and updated with available sequences downloaded from the Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/) and the EST Database (http://www.ncbi.nlm.nih.gov/dbEST/) of GenBank at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) using the program ED from the MUST package (Philippe, 1993). Unambiguously aligned regions were identified and excluded for each individual gene using the program GBLOCKS (Castresana, 2000) with a few manual refinements using NET from the MUST package. The complete list of genes with corresponding final numbers of amino-acid sites is available as Supporting Information.

The concatenation of the 179 genes was constructed with the program SCAFOS (Roure *et al*., 2007) by defining 51 metazoan operational taxonomic units (OTUs)

including 25 deuterostomes representing all major lineages. When several sequences were available for a given OTU, the slowest evolving one was selected according to their degree of divergence using ML distances computed by TREE-PUZZLE (Schmidt *et al.*, 2002) under a WAG+F model (Whelan and Goldman, 2001) within SCAFOS. The percentage of missing data per taxon was reduced by creating some chimerical sequences for species belonging to the same OTU. The complete alignment consists of 179 genes and 51 taxa for 53,799 unambiguously aligned amino-acid sites with 32% missing data. To study the potential impact of missing data on phylogenetic inference (Hartmann and Vision, 2008; Philippe *et al.*, 2004; Wiens, 2006), a concatenation of the 106 genes with sequences available for at least 41 of the 51 OTUs was also constructed with SCAFOS. This reduced alignment consists of 106 genes and 51 taxa for 25,321 amino-acid sites and contains only 20% of missing data. The list of defined OTUs, chimerical sequences and percentages of missing data are available as Supporting Information. Individual gene alignments and their concatenations are available upon request.

Phylogenomic Analyses

Bayesian phylogenetic analyses of the two phylogenomic datasets were conducted using the program PHYLOBAYES 2.3c (http://www.atgc-montpellier.fr/phylobayes/) under the CAT+Γ_4 site-heterogeneous mixture model (Lartillot and Philippe, 2004). For each dataset, four independent Monte Carlo Markov Chains (MCMCs) starting from a random topology were run in parallel for 20,000 cycles (1,500,000 generations), saving a point every cycle, and discarding the first 2,000 points as the burnin. Bayesian posterior probabilities (PP) were obtained from the 50% majority-rule consensus tree of the 18,000 MCMC sampled trees using the program READPB of PHYLOBAYES.

Maximum likelihood (ML) reconstruction on the new phylogenomic dataset was also performed using the program TREEFINDER version of March 2008 (Jobb *et al.*, 2004) under the empirical WAG+F+Γ_8 model of amino-acid substitution. The α shape parameter of the Γ distribution was estimated along with the topology and the branch lengths. Reliability of nodes was estimated by bootstrap resampling with 100 pseudo-replicate datasets generated by the program SEQBOOT of the PHYLIP package (Felsenstein, 2001). The 100 corresponding ML heuristic searches were run in parallel on a computing cluster and the majority-rule consensus of the 100 resulting trees was computed using TREEFINDER.

Jackknife Procedure

The robustness of our phylogenomic inference with respect to gene sampling was assessed by applying a jackknife procedure. Fifty jackknife replicates of 50 genes drawn randomly from the full pool of 179 genes were generated. The only condition we imposed to this jackknife procedure was to require that each taxon is represented by at least one gene in each replicate. The 50 jackknife supermatrices ranging from 11,163 to 17,181 amino-acid sites with 27 to 35% missing data were then analyzed using PHYLOBAYES under the CAT+Γ_4 model. To ensure correct convergence of the MCMC on each replicate, an automated stopping rule was used. Specifically, for each jackknife replicate, two independent parallel (and synchronous) MCMC were run, until the posterior probability discrepancy between the two chains was less than 0.15 (maximum discrepancy over all bipartitions), and after removing the first 1,000 sampled trees of each chain as the burnin. A global majority-rule consensus tree was obtained from the 50 replicates as follows: for each jackknife replicate (D_r) taken in turn, we computed the frequency-table of all bipartitions (splits) observed in the sample collected from the posterior distribution p ($T|D\_r$). The frequencies associated to each bipartition were then averaged over the 50 replicates, and the resulting frequency table was used to build a consensus tree. The support values displayed by this Bayesian consensus tree are thus jackknife-resampled posterior probabilities (PP_{JK}). High PP_{JK} values indicate nodes that have high posterior probability support in most jackknife replicates.

Phylogenetic Analyses of Ribosomal RNA

The 46-taxa dataset of combined 18S+28S rRNA genes assembled by Mallatt and Winchell (2007) for studying deuterostome phylogeny was reanalyzed. This alignment contains a total of 3,925 unambiguously aligned nucleotide sites. A principal component analysis (PCA) of nucleotide composition was realized using the R statistical package (R Development Core Team, 2007). The best fitting model of nucleotide sequence evolution was evaluated using MODELTEST 3.7 (Posada and Crandall, 1998). The TIM+Γ_4+I transitional model (Posada and Crandall, 2001) was selected according to the Akaïke information criterion. ML phylogenetic analysis of this nucleotide dataset was conducted with PAUP* 4.0b10 (Swofford, 2002) using a heuristic search with Tree Bisection Reconnection (TBR) branch swapping starting from a Neighbor-Joining (NJ) tree.

The nucleotide dataset was RY-coded by pooling puRines (AG = R) and pYrimidines (CT = Y) in an attempt to alleviate both compositional heterogeneity and substitutional saturation of transition events. This RY-coded dataset was then analyzed by conducting a ML heuristic search with TBR branch swapping on a NJ starting tree using PAUP* under the CF+Γ_8 model for discrete characters (Cavender and Felsenstein, 1987). The α shape parameter of the Γ distribution was previously estimated during a ML heuristic search on the nucleotide dataset conducted with TREEFINDER under the GTR2+Γ_8 two-state model.

Reliability of nodes was estimated for each dataset by nonparametric bootstrap resampling using 100 pseudo-replicates generated by SEQBOOT. The 100 corresponding ML heuristic searches using PAUP* with the previ-

ously estimated ML parameters, NJ starting trees, and TBR branch swapping were parallelized on a computing cluster. ML bootstrap percentages were obtained from the 50% majority-rule consensus tree of the 100 bootstrap ML trees using TREEFINDER.

RESULTS AND DISCUSSION

Effect of an Improved Model of Sequence Evolution

Our initial assessment of deuterostome phylogenetic relationships was based on a phylogenomic dataset encompassing 146 nuclear genes (33,800 amino-acids) from 38 taxa including 14 deuterostomes (Delsuc *et al*., 2006). ML and Bayesian phylogenetic analyses conducted under the standard WAG+F+Γ_4 model provided strong support for grouping tunicates with vertebrates (including cyclostomes), but also disrupted chordate monophyly because cephalochordates grouped with echinoderms, albeit with nonsignificant statistical support (Delsuc *et al*., 2006). The limited taxon sampling available at the time for Ambulacraria (echinoderms and hemichordates), that is, a single echinoderm, prompted us to be cautious about this result and to call for the inclusion of xenoturbellidans, hemichordates, and more echinoderms before drawing definitive conclusions. In fact, a subsequent phylogenomic study did exactly what we pleaded for by adding a representative species for each of these three groups (Bourlat *et al*., 2006). The inclusion of these taxa allowed retrieving the monophyly of chordates in Bayesian analyses using standard amino-acid models, although the alternative hypothesis of chordate paraphyly was still not statistically rejected by ML nonparametric tests (Bourlat *et al*., 2006). Importantly, the strong statistical support for the monophyly of Olfactores was unaffected by taxon addition (Bourlat *et al*., 2006).

Models accounting for site-specific modulations of the amino-acid replacement process, such as the CAT mixture model (Lartillot and Philippe, 2004), seem to offer a significantly better fit to real data than empirical substitution matrices currently used in standard models of amino-acid sequence evolution. Accounting for site-specific amino-acid propensities has also been shown to induce a significant improvement of phylogenetic reconstruction in difficult cases such as long-branch attraction (Baurain *et al*., 2007; Lartillot *et al*., 2007; Lartillot and Philippe, 2008). This improvement essentially lays in the ability of the CAT model to detect multiple conservative substitutions more efficiently than standard amino-acid models (Lartillot *et al*., 2007).

To test for an eventual effect of model misspecification on previous phylogenomic analyses, we reanalyzed our previous dataset (Delsuc *et al*., 2006) under the CAT+Γ_4 model. This analysis provides strong corroborating support for the grouping of tunicates and vertebrates (see Fig. 1). However, in contrast with previous analyses using empirical amino-acid replacement matrices, which favored a sister-group relationship between cephalochordates and echinoderms, the use of the CAT+Γ_4 mixture model strongly supports the classical view of monophyletic chordates and deuterostomes (see Fig. 1). The fact that chordate polyphyly is disrupted both by a richer taxon sampling (Bourlat *et al*., 2006), or upon the use of a more elaborate model, suggests that the previously observed grouping of cephalochordates and echinoderms (Delsuc *et al*., 2006) was probably a phylogenetic reconstruction artifact. On the other hand, the fact that the grouping of tunicates and vertebrates is insensitive to the model used, adds further credence to the Olfactores hypothesis.

An Updated Phylogenomic Dataset

The continuously growing genomic databases allowed us to build an updated phylogenomic dataset that includes both more genes and more taxa than previously considered to address the question of deuterostome phylogeny. This new dataset of 179 genes for 51 taxa includes 25 deuterostomes representing all major lineages: Xenoturbellida (1 taxon), Hemichordata (1), Echinodermata (5), Cephalochordata (1), Tunicata (6), Cyclostomata (2) and Vertebrata (9), plus 26 selected slow evolving metazoan taxa including Cnidarians and Poriferans as the most distant outgroups. Chordates are particularly well sampled with the inclusion, for the first time, of six tunicate species covering the four major clades evidenced by 18S rRNA studies (Swalla *et al*., 2000). This diverse taxon sampling is essential to further test the new chordate phylogeny recently revealed by phylogenomics (Bourlat *et al*., 2006; Delsuc *et al*., 2006).

Bayesian (CAT+Γ_4) and ML (WAG+F+Γ_8) phylogenetic reconstructions conducted on this updated dataset (179 genes, 53,799 amino-acid sites, 51 taxa) resulted in a highly resolved tree (Fig. 2a). These analyses provided strong support for Ambulacraria ($PP_{CAT+\Gamma4}$ = 1.0/$BP_{WAG+F+\Gamma8}$ = 97), chordates (1.0/69) and olfactores (1.0/100). Xenambulacraria (*Xenoturbella* + Ambulacraria) and a basal position for the chaetognath *Spadella* among Protostomia were also moderately supported by our analyses (Fig. 2a). These results are compatible with a recent phylogenomic analysis, which also found strong support for Ambulacraria, chordates, and Olfactores when using the CAT mixture model (Dunn *et al*., 2008). However, the monophyly of Deuterostomes is unresolved in both Bayesian and ML phylogenetic reconstructions (Fig. 2a).

The complete dataset obtained by concatenating all 179 genes contains 32% missing data. Previous studies of the impact of missing data on the accuracy of phylogenetic inference have concluded that probabilistic methods are relatively tolerant to missing data (Hartmann and Vision, 2008; Philippe *et al*., 2004; Wiens, 2003, 2005), the most important factor being the absolute amount of available data for a given taxon. In phylogenomics, even incomplete taxa are usually represented by thousand of sites, and the impact of missing data on accuracy is

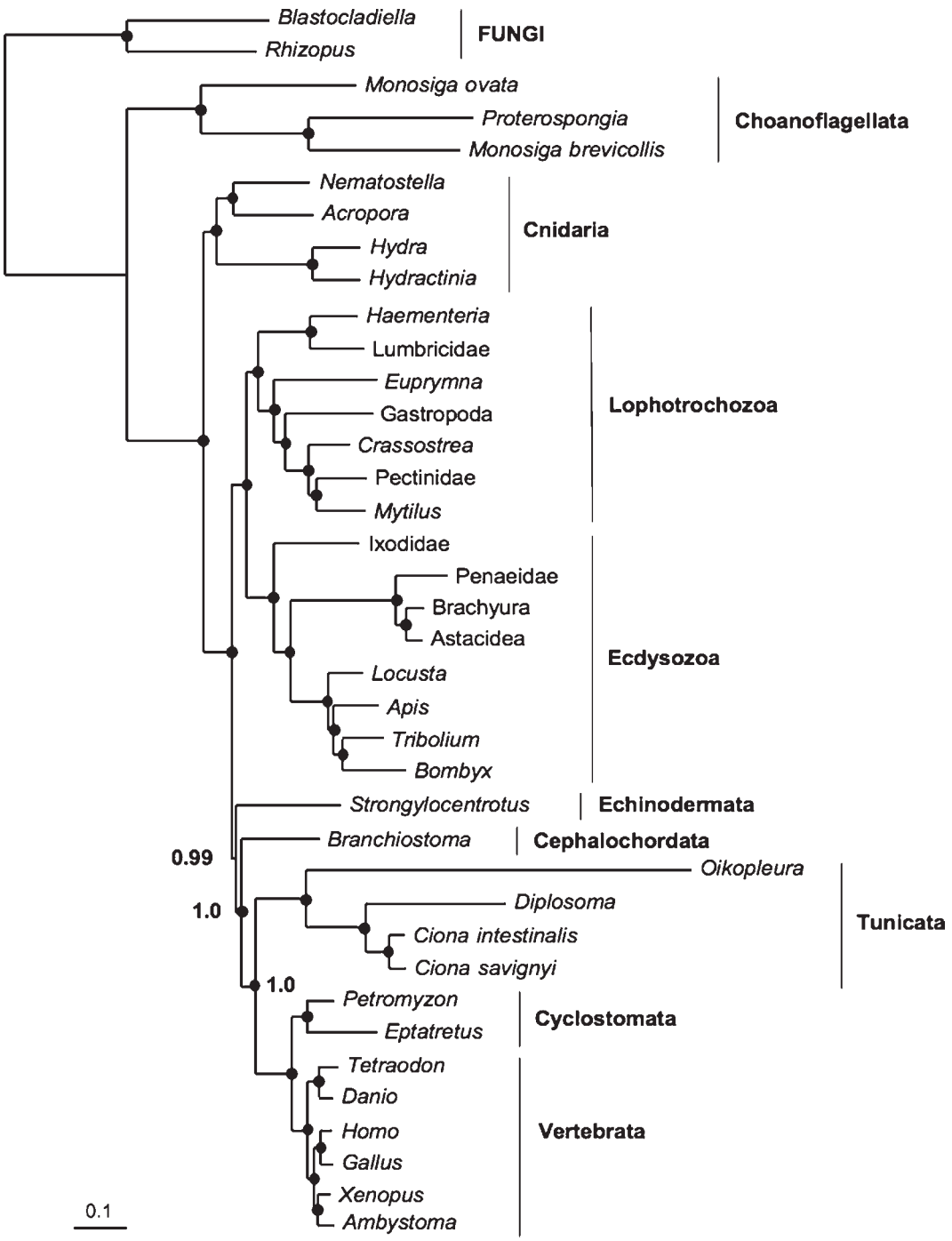

FIG. 1. Reanalysis of previous phylogenomic data using an improved model of sequence evolution. The Delsuc *et al*. (2006) phylogenomic dataset of 146 genes (38 taxa and 33,800 sites) was analyzed under the CAT+Γ_4 site-heterogeneous mixture model of amino-acid replacement. Values at nodes represent Bayesian posterior probabilities (PP). Circles indicate nodes with maximal support PP = 1.0. The scale bar represents the estimated number of substitutions per site.

therefore relatively limited (Philippe *et al*., 2004). Nonetheless, incomplete taxa might still be difficult to place with confidence especially when they represent isolated lineages such as *Xenoturbella* (65% missing data) and *Spadella* (75%) in our dataset. To control for a potential effect of missing data on our phylogenomic results, we restricted our dataset to the 106 genes with sequences available for at least 41 of the 51 taxa. The concatenation of these 106 genes produces a matrix with 25,321 amino-acid sites that contains only 20% of missing data.

Bayesian and ML phylogenetic inference on this reduced dataset produced a tree fully congruent with

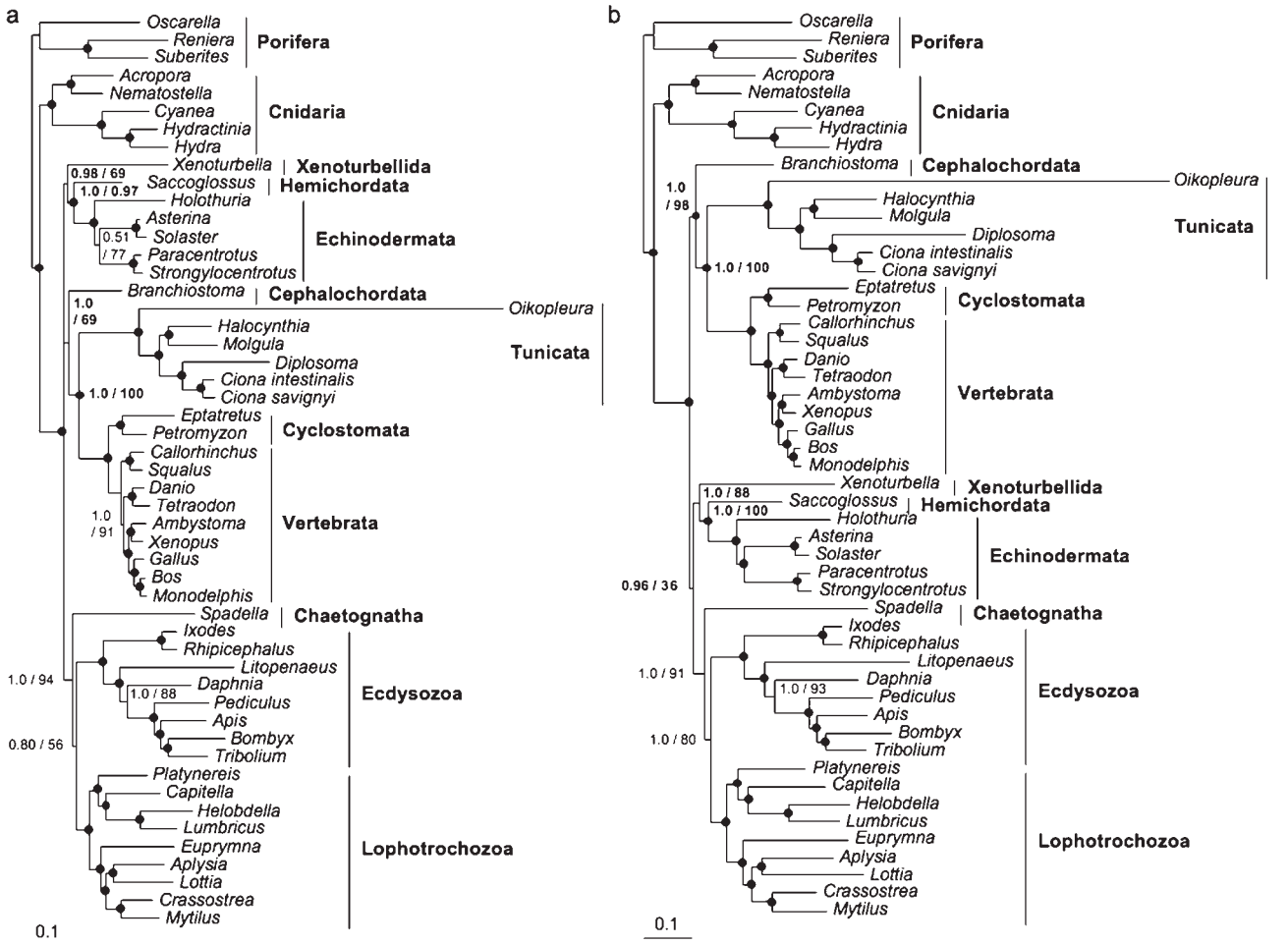

FIG. 2. Phylogenetic analyses of an updated phylogenomic dataset. (**a**) Bayesian consensus tree obtained using the CAT+Γ_4 mixture model on the complete dataset based on the concatenation of 179 genes (51 taxa and 53,799 amino-acid sites) containing 32% missing data. (**b**) Bayesian inference using the CAT+Γ_4 mixture model on the dataset reduced to the concatenation of the 106 genes for which sequences were available for at least 41 of the 51 taxa (25,321 amino-acid sites) containing only 20% of missing data. Values at nodes indicate Bayesian posterior probabilities (PP)/Maximum-likelihood bootstrap percentages (BP; 100 replicates) obtained under the WAG+Γ_8. Circles indicate strongly supported nodes with PP $\geq$0.95 and BP $\geq$95. The scale bar represents the estimated number of substitutions per site.

the phylogenetic picture given by the complete dataset (Fig. 2b). In particular, the support for the monophyly of chordates was still maximal in terms of $PP_{CAT+\Gamma4}$, but $BP_{WAG+F+\Gamma8}$ increased from 69 to 88%. The monophyly of Olfactores received maximal support in both cases and appeared not affected by missing data. Statistical support in terms of $PP_{CAT+\Gamma4}$ and $BP_{WAG+F+\Gamma8}$ was generally increased especially for locating incomplete taxa such as the *Xenoturbella* as the sister-group to Ambulacraria (from 0.98/69 to 1.0/88) and *Spadella* at the base of Protostomia (from 0.80/56 to 1.0/80). Altogether, reducing the amount of missing data, despite also reducing the total number of available sites, seems to result in a slight increase in bootstrap proportions. The only real noticeable difference between the two methods concerns the monophyly of deuterostomes. The Bayesian inference under the CAT mixture model suggests deuterostome paraphyly by supporting a basal position of chordates within Bilateria (Fig. 2b) as previously reported (Lartillot and Philippe, 2008). However, ML retrieved the monophyly of deuterostomes, but with $BP_{WAG+F+\Gamma8}$ of only 50%, leaving the monophyly of deuterostomes unresolved by our data.

Robustness of Phylogenomics to Gene Sampling

A legitimate question that can be directed to the phylogenomic approach is the degree to which the results are robust to the sample of genes used to infer phylogenetic trees. This potential concern was addressed by applying a jackknife statistical resampling protocol: fifty datasets were assembled by randomly drawing 50 genes from the total 179 genes, and subjected to Bayesian phylogenetic reconstruction using the CAT+Γ_4 mixture model (see Methods). The resulting majority-rule consensus tree shows that the vast majority of inferred phy-

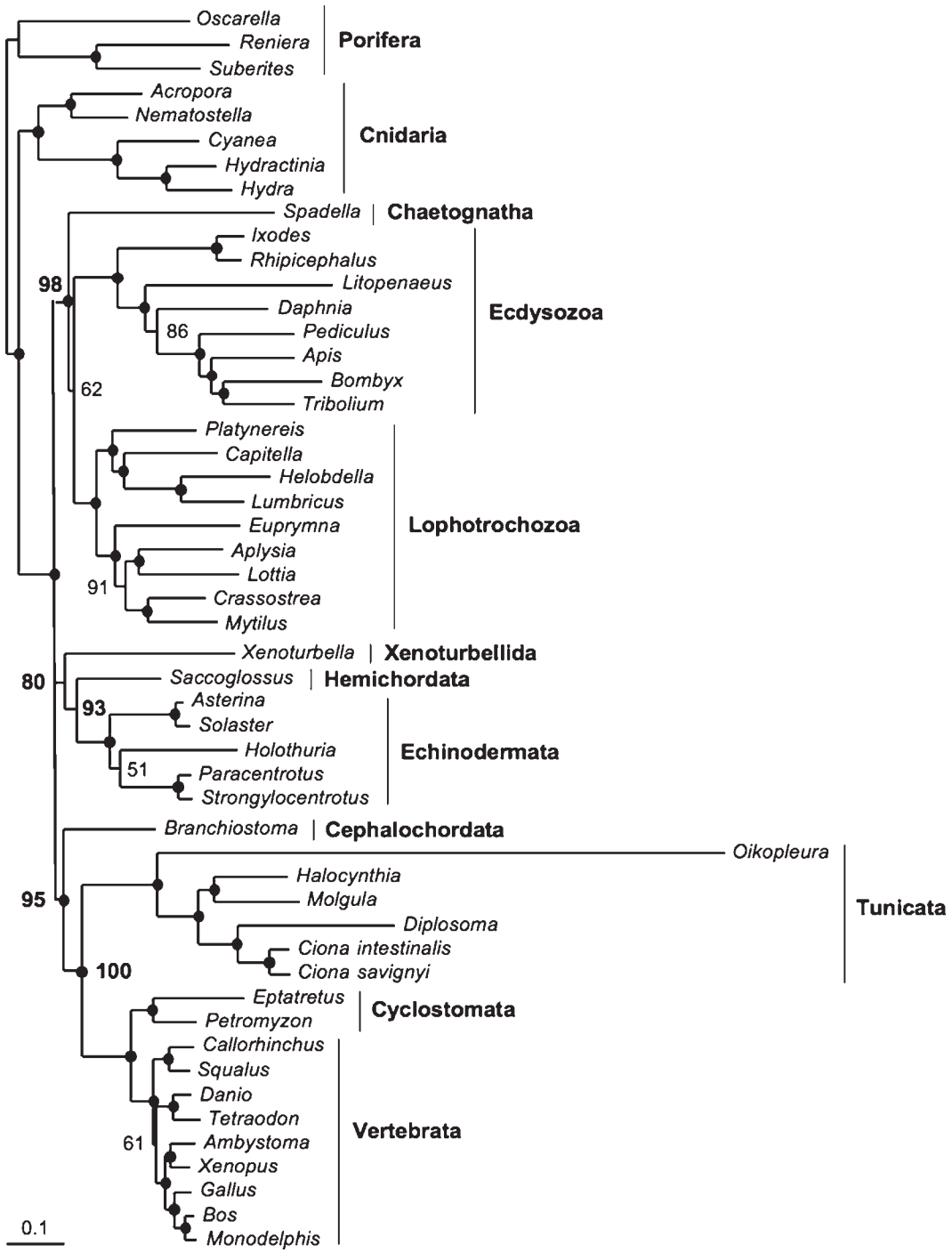

FIG. 3. Assessing the robustness of phylogenetic results to gene sampling using a jackknife procedure. The Bayesian phylogenetic inference was conducted under the CAT+Γ_4 mixture model on 50 jackknife replicates of 50 genes over a total of 179. The tree presented is the weighted majority-rule consensus of all trees sampled every 10 cycles across the 50 replicates after removing the first 1000 trees in each MCMC as the burnin. Values at nodes represent corresponding jackknife-resampled posterior probabilities indices (PP_{JK}). Circles indicate highly repeatable nodes with $PP_{JK} \geq 95\%$. The scale bar represents the number of substitutions per site.

logenetic relationships are highly repeatable across the 50 jackknife replicates (see Fig. 3). Olfactores, Chordata, and Ambulacraria all received PP_{JK} of more than 90% indicating that phylogenetic support is not dependent upon a particular gene combination. Xenambulacraria appears slightly more affected by gene sampling (PP_{JK} = 80%), but this relative instability might be explained by the poor gene representation available for *Xenoturbella* with only 98 genes over 179 (55%). The same kind of reasoning could apply to the relatively unstable posi-

tions of the chaetognath *Spadella* within protostomes (PP_{JK} = 62%) and of *Holothuria* within echinoderms (PP_{JK} = 51%) (see Fig. 3).

In fact, the only major clade whose monophyly appears to be influenced by gene sampling is deuterostomes for which PP_{JK} was less than 50% (see Fig. 3). In practice, this means that depending on the particular combination of 50 genes considered, deuterostomes might appear either monophyletic or paraphyletic, with the three possible topological alternatives retrieved in similar proportions: Deuterostomes (38%), basal chordates (28%), and basal Xenambulacraria (22%). Despite the inclusion of 25 taxa representing all major lineages in our dataset, these results confirm deuterostomes as one of the most difficult groups to resolve in the animal phylogeny despite its wide acceptance (see Lartillot and Philippe, 2008).

New Analyses of Ribosomal RNA Genes

The sister-group relationship between tunicates and vertebrates (Olfactores) observed in phylogenomics is in conflict with most (if not all) analyses of rRNA which favor cephalochordates as the closest relatives of vertebrates (Euchordates) (Cameron *et al*., 2000; Mallatt and Winchell, 2007; Swalla *et al*., 2000; Wada and Satoh, 1994; Winchell *et al*., 2002). However, the statistical support for Euchordates in rRNA-based phylogenetic studies is moderate. Indeed, the first 18S rRNA study, based on a limited taxon sampling of deuterostomes, reported a bootstrap value of only 45% for Euchordates (Wada and Satoh, 1994). A subsequent 18S rRNA study considering only slowly evolving sequences for 16 deuterostomes found only a moderate bootstrap support of 71% for grouping cephalochordates and vertebrates (Cameron *et al*., 2000). A study focused on tunicates also obtained moderate support for Euchordates (58 to 85% depending on the dataset and reconstruction method) but failed to support chordate monophyly likely because tunicate 18S rRNA sequences are rapidly evolving (Swalla *et al*., 2000).

The next studies used the combination of 18S and 28S rRNAs. An investigation using 28 taxa for the two rRNA subunits found strong boostrap support (89 to 97% depending on the method) for Euchordates (Winchell *et al*., 2002). However, this study again failed to support chordate monophyly. Detailed analyses confirmed that tunicate genes have evolved rapidly and showed that they are compositionally biased toward AT, rendering tunicates virtually impossible to locate convincingly in the tree on the basis of rRNA data (Winchell *et al*., 2002). Finally, increasing the sampling to 46 taxa for this 18S+28S rRNA data did not helped in further resolving the relationships among the major groups of deuterostomes and even decreased the ML bootstrap support for Euchordates from 97% in the previous study to 50% (Mallatt and Winchell, 2007).

To gauge the extent to which the rRNA data conflicts with our phylogenomic results, we reanalyzed the 46-taxa dataset of Mallatt and Winchell (2007). The heterogeneity of base composition in this dataset is well illustrated by the PCA presented in Figure 4a. At one extreme, tunicates (especially *Oikopleura*) are particularly AT-rich, and at the other extreme, Myxinidae (*Myxine* and *Eptatretus*) and the pterobranch hemichordate *Cephalodiscus* are highly GC-rich. We therefore compared phylogenetic reconstructions conducted on nucleotides and on RY-coded data, a coding scheme allowing reducing both substitutional saturation and nucleotide compositional bias (Fig. 4b). The two inferred ML trees appear mostly congruent except for two major topological shifts.

The strongest topological change occurred within hemichordates (Fig. 2b). Although the use of a standard DNA model strongly supports the paraphyly of enteropneusts by grouping the pterobranch *Cephalodiscus* with *Saccoglossus* and *Harrimania* (BP = 95), RY-coding allows recovering the monophyly of enteropneusts with high bootstrap support (BP = 90). This helps in understanding the conflict between 18S rRNA that supports enteropneust paraphyly (Cameron *et al*., 2000; Halanych, 1995) and 28S rRNA that rather favors their monophyly (Mallatt and Winchell, 2007; Winchell *et al*., 2002). This result is of particular importance because it potentially invalidates the controversial hypothesis that pterobranchs evolved from an enteropneust (Cameron *et al*., 2000; Halanych, 1995) by suggesting that it is likely an artifact of 18S rRNA-based phylogenetic reconstructions due to shared nucleotide compositional bias between pterobranchs and Harrimaniidae (Fig. 4a).

Second, the support for the monophyly of Euchordates observed with nucleotides (BP = 84) disappeared in favor of the monophyly of Olfactores in the RY-coding dataset, albeit with no statistical support (BP = 44). This nevertheless strongly suggests that the high composition bias of tunicate sequences has blurred the phylogenetic signal for Olfactores in previous analyses. Thus, according to our interpretation, reducing compositional bias and substitutional saturation by RY-recoding allows recovering a limited signal for Olfactores in agreement with our phylogenomic analysis of amino-acid data. It is worth noting however that rRNA does not statistically support chordate monophyly in both cases (Fig. 2b).

Molecular Phylogenetic Conclusions

Our aim was to reanalyze the phylogenetic relationships among chordates. The revision of the position of tunicates proposed by recent phylogenomic studies (Bourlat *et al*., 2006; Delsuc *et al*., 2006; Dunn *et al*., 2008) by concluding in favor of the monophyly of Olfactores, has not yet been considered as totally convincing, essentially because it is at odds with both the traditional view based on embryological and morphological characters (Rowe, 2004; Schaeffer, 1987), and with earlier molecular phylogenetic analyses based on rRNA (Cameron *et al*., 2000; Mallatt and Winchell, 2007; Swalla *et al*., 2000; Wada and Satoh, 1994; Winchell *et al*., 2002). The

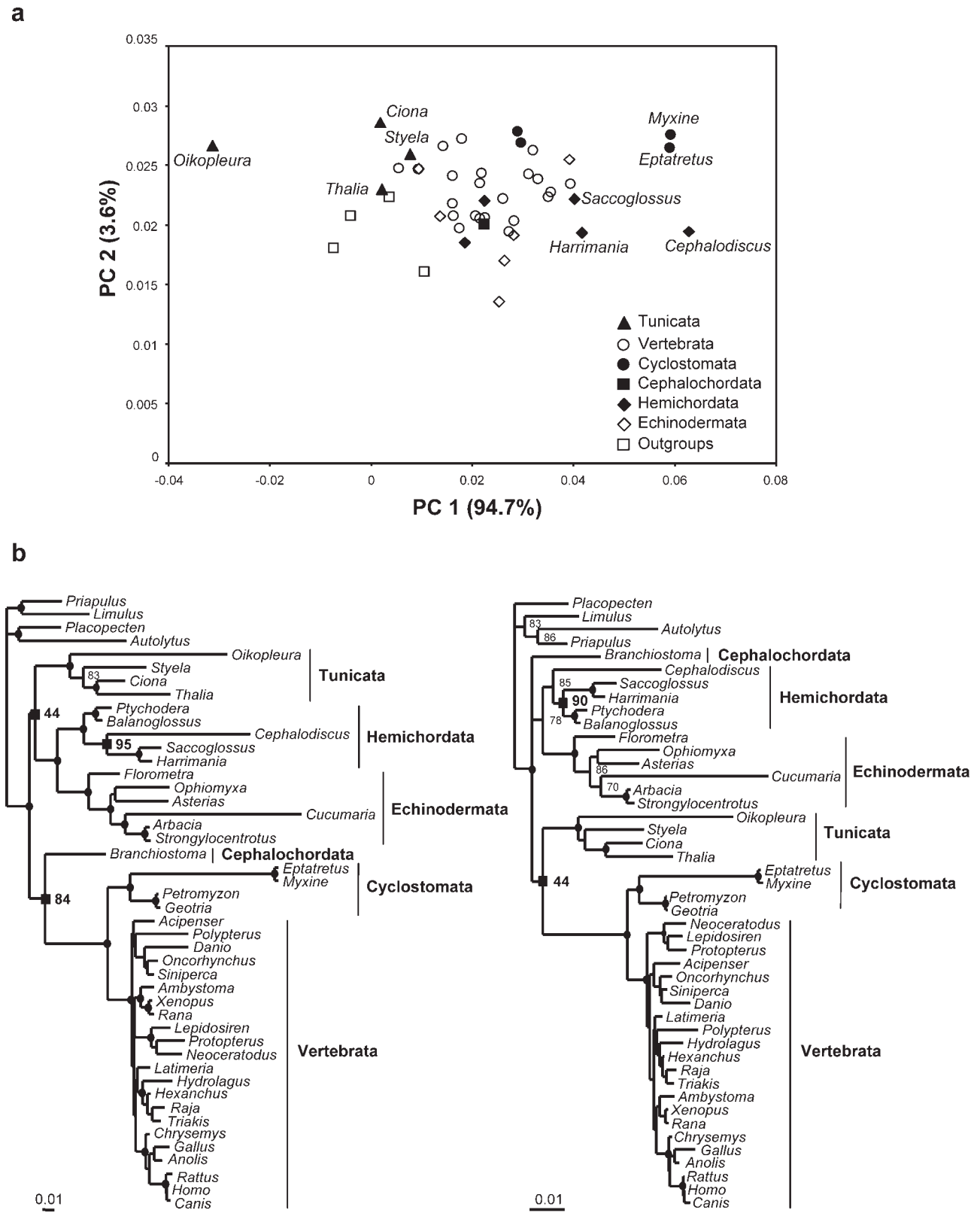

FIG. 4. New phylogenetic analyses of ribosomal RNA genes. (**a**) Principal component analysis of nucleotide composition of the combined 18S+28S rRNA dataset. The graph represents the projection of individuals on the first two axes, which explain more than 98% of the total variance. (**b**) Maximum-likelihood analyses of the 18S+28S dataset using the best-fitting standard DNA model (TIM+Γ_4+I) on nucleotides (left) and a two-state model (CF+Γ_8) after RY-coding of nucleotides (right). ML bootstrap percentages are given at nodes when greater than 70 except within vertebrates. Circles indicate strongly supported nodes with BP $\geq$95. Squares point to shifting nodes of interest between the two ML trees. Scale bars represent the number of substitutions per site.

unexpected sister-group relationship between echinoderms and cephalochordates observed in one of these studies (Delsuc *et al*., 2006) may also have suggested the possibility that the monophyly of Olfactores was due to an artefactual attraction of cephalochordates with echinoderms (Bourlat *et al*., 2006).

In this analysis, we have tried to address these points, essentially by reanalyzing both phylogenomic and rRNA data, under better taxonomic sampling and using more elaborate methods and probabilistic models. First, we demonstrate that, although the grouping of echinoderms and cephalochordates was indeed a probable artifact, disappearing upon the addition of several taxa or using an improved model of sequence evolution, the monophyly of Olfactores appears to be robust with respect to taxon sampling and model choice. Second, our reanalysis of rRNA data using RY-recoding also reveals a weak signal in favor of Olfactores, and suggests that the grouping of vertebrates and cephalochordates in former studies may have been an artifact driven by compositional biases. Altogether, our analyses allows a coherent interpretation of all empirical results observed thus far concerning chordate phylogeny, yielding further evidence in favor of the monophyly of both chordates and Olfactores.

At larger scale, however, we observe an overall lack of support for the monophyly of deuterostomes. Deuterostomes have nearly unanimously been considered as an unquestionable monophyletic group, a hypothesis backed up by traditional comparative analyses of embryological characters such as the fate of the blastopore (Nielsen, 2001), and morphological traits such as gill slits (Schaeffer, 1987). However, in our analyses, the status of deuterostomes seems to be sensitive to the model used, with CAT slightly favoring the paraphyletic configuration, and WAG the more traditional monophyly. In either case, the support measured by nonparametric resampling procedures (site-wise bootstrap or gene-wise jackknife) is weak.

Other phylogenomic studies (Dunn *et al*., 2008; Lartillot and Philippe, 2008) also failed to obtain strong support for the relative phylogenetic positions of chordates and Ambulacraria. Moderate support for the monophyly of Deuterostomes was only obtained under empirical matrix models, the support disappearing when the CAT model was used instead (Dunn *et al*., 2008; Lartillot and Philippe, 2008). Although profile mixture models such as CAT, whereas having a better fit than empirical matrices such as WAG, may have some inherent weaknesses as to their phylogenetic accuracy, the WAG empirical matrix fails in many cases, particularly when confronted to a high level of saturation (Lartillot *et al*., 2007). This casts doubts on results that seem to receive support exclusively under this model, as is the case for deuterostome monophyly. More observations are needed to better gauge the relative merits of either type of model. Overall, although deuterostome monophyly still remains a reasonable working hypothesis to date, more work is needed before the question can be settled.

Corroborative Evidence for Olfactores Monophyly

The monophyly of Olfactores receives strong support from sequence-based phylogenomic inference. Rare genomic changes have also provided some evidence in its favor: the domain structure of cadherins (Oda *et al*., 2002), a unique amino-acid insertion in fibrillar collagen (Wada *et al*., 2006), and the distribution of micro RNAs (miRNAs) (Heimberg *et al*., 2008). Finally, the *Branchiostoma floridae* genome helps confirming the sister-group relationship between tunicates and vertebrates in offering additional evidence from analyses of intron dynamics (Putnam *et al*., 2008).

Cadherins are a superfamily of highly conserved adhesion molecules mediating cell communication and signaling that are pivotal for developmental processes of multicellular organisms. Their recent detection in the closest unicellular relatives of metazoans, the choanoflagellates, has highlighted their potential role in the origin of multicellularity (Abedin and King, 2008). Comparative studies on the classic cadherin subfamily has revealed that the structural element called Primitive Classic Cadherin Domain (PCCD) complex, otherwise termed non-chordate classic cadherin domain, is also present in cephalochordates, but has been lost in both tunicates and vertebrates (Oda *et al*., 2002). The most parsimonious scenario is that this particular protein domain complex has been lost in the common ancestor of tunicates and vertebrates and constitutes a synapomorphy of Olfactores. However, cephalochordates possess two classic cadherin genes which originated by lineage-specific tandem duplication and that have a particular structure in lacking extracellular repeats found in all other investigated metazoans (Oda *et al*., 2004). This derived state renders difficult to ascertain domain homology among chordate classic cadherin genes and casts doubt on its phylogenetic significance.

Further potential evidence for the clade Olfactores has been inferred from the evolution of fibrillar collagen genes within chordates. These genes represent important components of the notochord, the cartilage and mineralized bones in vertebrates. Phylogenetic analyses suggested that three ancestral fibrillar collagens gave rise to the gene diversity observed in living deuterostomes (Wada *et al*., 2006). Comparative sequence analyses showed that tunicates and vertebrates share a molecular signature in the form of a six to seven amino-acid insertion in the C-terminus noncollagenous domain of one type of fibrillar collagens, that is absent in cephalochordates and echinoderms (Wada *et al*., 2006). This insertion was interpreted as supporting the idea that vertebrates are more closely related to tunicates than to cephalochordates (Wada *et al*., 2006). The homology of the insertion appears nevertheless difficult to assert with certainty given the high degree of sequence divergence observed in this region of the molecule. More tunicate fibrillar collagen sequences might help in better understanding the dynamics of this peculiar amino-acid insertion and the phylogenetic signal it conveys.

The comparison of miRNA repertoires in metazoans has also recently unearthed some potential signatures for the sister-group relationship of tunicates and vertebrates (Heimberg *et al.*, 2008). miRNAs are small noncoding RNAs involved in regulation of gene expression in eukaryotes and play an important role in the development of metazoans. Comparative genomic studies of miRNAs underlined that, during the evolution of metazoans, major body-plan innovations seemed to coincide with dramatic expansions of miRNA repertoires, suggesting a potential role in the increase of morphological complexity (Hertel *et al.*, 2006; Sempere *et al.*, 2006). The most recent study unveiled that three miRNA families (*mir-126*, *mir-135,* and *mir-155*) were likely acquired in the common ancestor of tunicates and vertebrates (Heimberg *et al.*, 2008). Taking into consideration that miRNAs might be only rarely secondarily lost once they have been recruited, this finding provides corroborative evidence for the clade Olfactores. It should be noted however that, of these three miRNA families, only *mir-126* constitutes an exclusive synapomorphy for Olfactores without subsequent secondary lost in descendant taxa confirmed by Northern analysis. Moreover, the profound reorganization of miRNA repertoire undergone by tunicates requires being cautious when interpreting acquisition of miRNAs as potential signatures for reconstructing their phylogenetic relationships (Fu *et al.*, 2008).

Additional sequence-based phylogenomic reconstructions and analyses of rare genomic changes have been issued along with the recently published draft sequence of a cephalochordate (*Branchiostoma floridae*) genome (Putnam *et al.*, 2008). The phylogenetic analysis of a concatenation of 1,090 orthologs from 12 complete genomes retrieved maximal Bayesian support for Olfactores and chordates, whereas the corresponding bootstrap support was maximal for Olfactores but of only 78% for chordate monophyly (Putnam *et al.*, 2008). Moreover, the analysis of individual gene phylogenies revealed twice more cases where Olfactores was favored over Euchordates than the reverse (Putnam *et al.*, 2008). Further evidence was obtained by analyzing the phylogenetic signal deduced from the dynamics of intron gain and loss among chordate genomes. Despite extensive intron losses along the tunicate lineage, a number of shared intron gain/loss events can be identified as a signature of tunicates and vertebrates common ancestry (Putnam *et al.*, 2008). Overall, the new evidence brought by the analysis of the *Branchiostoma floridae* genome essentially corroborates our phylogenetic results.

Implications for Chordate Evo-Devo

The additional evidence presented for the new chordate phylogeny provides a robust phylogenetic framework for (re)interpreting the evolution of morphological characters and developmental features. Inverting the phylogenetic position of tunicates and cephalochordates within monophyletic chordates highlights the prevalence of morphological simplification with characters that are likely ancestral for chordates, such as metameric segmentation, being lost secondarily in the tunicate lineage. On the other hand, the loss of preoral kidney and the presence of multiciliated epithelial cells might in fact constitute morphological synapomorphies for olfactores (Ruppert, 2005). The new chordate phylogeny further portrays tunicates as highly derived chordates with specialized lifestyles and developmental modes, whereas cephalochordates might have retained more ancestral chordate characteristics. We will use two examples to illustrate the importance of considering the new phylogenetic status of tunicates as the sister-group of vertebrates in the context of evolutionary developmental biology.

The first illustration concerns the evolutionary origin of such fundamental structures as the neural crest and olfactory placodes. Migratory neural crest cells and sensory placodes have long been considered as vertebrate innovations. Implicated respectively in the development of major tissues and sensory organs, their origin is generally correlated with the increase in morphological complexity of vertebrates. However, recent molecular developmental studies have revealed the presence in tunicates of migratory neural crest-like cells (Jeffery, 2006; Jeffery *et al.*, 2004) and olfactory placodes (Bassham and Postlethwait, 2005; Mazet *et al.*, 2005). When reinterpreted in light of the new chordate phylogeny, these results implied that both of these features did not evolve *de novo* in the vertebrate lineage, but rather evolved from specialized pre-existing structures in the common ancestor of vertebrates and tunicates.

The second example illustrates how the new phylogenetic context helps in understanding the genomic and developmental peculiarities of tunicates within chordates. The new phylogenetic picture implied that tunicate genomes have undergone significant genome reduction from the ancestral chordate genome (Holland, 2007). This genome compaction is also associated with a high rate of genomic evolution at the levels of both primary sequences (Delsuc *et al.*, 2006; Edvardsen *et al.*, 2004) and genome organization (Holland and Gibson-Brown, 2003). One of the most spectacular rearrangements of tunicate genomes is the lost of several Hox genes, the disintegration of the Hox cluster, and the lost of temporal colinearity in Hox gene expression during development (Ikuta *et al.*, 2004; Seo *et al.*, 2004). These observations raise the question of how tunicates, with their altered Hox clusters, are still able to develop a chordate body plan. In chordates, and deuterostomes more generally, temporal colinearity is regulated by the Retinoic-Acid (RA) signaling pathway which controls the antero-posterior patterning of the embryo (Cañestro *et al.*, 2006; Marlétaz *et al.*, 2006). However, axial patterning in tunicates seems to have become independent of RA-signaling, with the genes of the RA machinery even being lost in *Oikopleura* (Cañestro and Postlethwait, 2007). Functional studies have shown that if "*Oikopleura* can be considered as a classical RA-signal-

ing knock-down mutant naturally produced by evolution," it is still capable of developing a typical chordate body plan (Cañestro and Postlethwait, 2007). With cephalochordates, which possess the RA genomic toolkit, being basal among chordates, RA-signalling must have been present in the tunicate ancestor and secondarily lost in *Oikopleura* suggesting that appendicularians use alternative mechanisms for the development of chordate features (Cañestro *et al*., 2007; Holland, 2007).

The new chordate phylogeny strengthens the view that tunicates and cephalochordates represent complementary models for studying vertebrate Evo-Devo (Schubert *et al*., 2006). Tunicates are phylogenetically closer to vertebrates but appear both morphologically and molecularly highly derived. The diversity of their developmental modes offers the opportunity to study the evolution of alternative adaptive solutions to the typical chordate development. In having retained more ancestral features, cephalochordates provide an ideal outgroup for polarizing evolutionary changes that occurred in tunicates and vertebrates. With the cephalochordate *Branchiostoma floridae* genome (Putnam *et al*., 2008) and the upcoming genome sequence of the appendicularian *Oikopleura dioica*, the newly established phylogenetic framework makes chordate comparative genomics appearing full of promises for the Evo-Devo community as exemplified in a recent work on the origin and evolution of the Pax gene family (Bassham *et al*., 2008).

ACKNOWLEDGMENTS

The authors thank the associate editors Billie Swalla and José Xavier-Neto for giving the opportunity to write this article. The authors also thank John Mallatt for kindly providing his 18S-28S rRNA alignment, Julien Claude for help in using R statistical package, and two anonymous reviewers for their comments. The extensive phylogenomic calculations benefited from the ISEM computing cluster.

LITERATURE CITED

Abedin M, King N. 2008. The premetazoan ancestry of cadherins. Science 319:946–948.

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

Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493.

Bassham S, Cañestro C, Postlethwait JH. 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. BMC Biol 6:35.

Bassham S, Postlethwait JH. 2005. The evolutionary history of placodes: A molecular genetic investigation of the larvacean urochordate Oikopleura dioica. Development 132:4259–4272.

Baurain D, Brinkmann H, Philippe H. 2007. Lack of resolution in the animal phylogeny: Closely spaced cladogeneses or undetected systematic errors? Mol Biol Evol 24:6–9.

Blair JE, Hedges SB. 2005. Molecular phylogeny and divergence times of deuterostome animals. Mol Biol Evol 22:2275–2284.

Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. Mol Biol Evol 25:842–858.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444:85–88.

Cameron CB, Garey JR, Swalla BJ. 2000. Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. Proc Natl Acad Sci USA 97:4469–4474.

Cañestro C, Postlethwait JH. 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev Biol 305:522–538.

Cañestro C, Postlethwait JH, Gonzalez-Duarte R, Albalat R. 2006. Is retinoic acid genetic machinery a chordate innovation? Evol Dev 8:394–406.

Cañestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. Nat Rev Genet 8:932–942.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552.

Cavender JA, Felsenstein J. 1987. Invariants of phylogenies in a simple case with discrete states. J Classif 4:57–71.

Conway-Morris S. 2003. The Cambrian "explosion" of metazoans and molecular biology: Would Darwin be satisfied? Int J Dev Biol 47:505–515.

Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965–968.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6:361–375.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749.

Edvardsen RB, Lerat E, Maeland AD, Flat M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo HC, Chourrout D. 2004. Hypervariable and highly divergent intron-exon organizations in the chordate *Oikopleura dioica*. J Mol Evol 59:448–457.

Felsenstein J. 2001. PHYLIP (Phylogenetic Inference Package) version 3.6. (distributed by the author). Seattle: Department of Genetics, University of Washington.

Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA, USA: Sinauer Associates, Inc. 645 p.

Fu X, Adamski M, Thompson EM. 2008. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. Mol Biol Evol 25:1067–1080.

Gee H. 2001. Deuterostome phylogeny: the context for the origin and evolution of the vertebrates. In: Ahlberg PE, editor. Major events in early vertebrate evolution: Palaeontology, phylogeny, genetics, and development. London: Taylor and Francis. pp 1–14.

Halanych KM. 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. Mol Phylogenet Evol 4:72–76.

Halanych KM. 2004. The new view of animal phylogeny. Annu Rev Ecol Evol Syst 35:229–256.

Halanych KM, Bacheller JD, Aguinaldo AM, Liva SM, Hillis DM, Lake JA. 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science 267:1641–1643.

Hartmann S, Vision TJ. 2008. Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? BMC Evol Biol 8:95.

Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ. 2008. MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci USA 105:2946–2950.

Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. 2006. The expansion of the metazoan microRNA repertoire. BMC Genomics 7:25.

Holland LZ. 2007. Developmental biology: A chordate with a difference. Nature 447:153–155.

Holland LZ, Gibson-Brown JJ. 2003. The *Ciona intestinalis* genome: When the constraints are off. Bioessays 25:529–532.

Ikuta T, Yoshida N, Satoh N, Saiga H. 2004. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. Proc Natl Acad Sci USA 101:15118–15123.

Jefferies RPS. 1991. Two types of bilateral symmetry in the Metazoa: Chordate and bilaterian. In: Bock GR, Marsh J, editors. Biological asymmetry and handedness. Chichester: Wiley. pp 94–127.

Jeffery WR. 2006. Ascidian neural crest-like cells: Phylogenetic distribution, relationship to larval complexity, and pigment cell fate. J Exp Zoolog B Mol Dev Evol 306:470–480.

Jeffery WR. 2007. Chordate ancestry of the neural crest: New insights from ascidians. Semin Cell Dev Biol 18:481–491.

Jeffery WR, Strickler AG, Yamamoto Y. 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431:696–699.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: The beginning of incongruence? Trends Genet 22:225–231.

Jimenez-Guri E, Philippe H, Okamura B, Holland PWH. 2007. Buddenbrockia is a Cnidarian worm. Science 317:116–118.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4:18.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol 7 Suppl 1:S4.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21:1095–1109.

Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos Trans R Soc Lond B Biol Sci 363:1463–1472.

Lwoff A. 1944. L'évolution physiologique: Étude des pertes de fonctions chez les microorganismes. Paris: Hermann. 308 p.

Mallatt J, Winchell CJ. 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. Mol Phylogenet Evol 43:1005–1022.

Marlétaz F, Holland LZ, Laudet V, Schubert M. 2006a. Retinoic acid signaling and the evolution of chordates. Int J Biol Sci 2:38–47.

Marlétaz F, Martin E, Perez Y, Papillon D, Caubit X, Lowe CJ, Freeman B, Fasano L, Dossat C, Wincker P, Weissenbach J, Le Parco Y. 2006b. Chaetognath phylogenomics: A protostome with deuterostome-like development. Curr Biol 16:R577–R578.

Matus DQ, Copley RR, Dunn CW, Hejnol A, Eccleston H, Halanych KM, Martindale MQ, Telford MJ. 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. Curr Biol 16:R575–R576.

Mazet F, Shimeld SM. 2005. Molecular evidence from ascidians for the evolutionary origin of vertebrate cranial sensory placodes. J Exp Zoolog B Mol Dev Evol 304:340–346.

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

Oda H, Akiyama-Oda Y, Zhang S. 2004. Two classic cadherin-related molecules with no cadherin extracellular repeats in the cephalochordate amphioxus: Distinct adhesive specificities and possible involvement in the development of multicell-layered structures. J Cell Sci 117:2757–2767.

Oda H, Wada H, Tagawa K, Akiyama-Oda Y, Satoh N, Humphreys T, Zhang S, Tsukita S. 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. 1993. MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res 21:5264–5272.

Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005a. Phylogenomics. Annu Rev Ecol Evol Syst 36:541–562.

Philippe H, Lartillot N, Brinkmann H. 2005b. Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. Mol Biol Evol 22:1246–1253.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PW, Casane D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol Biol Evol 21:1740–1752.

Philippe H, Telford MJ. 2006. Large-scale sequencing and the new animal phylogeny. Trends Ecol Evol 21:614–620.

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

Posada D, Crandall KA. 1998. MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817–818.

Posada D, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. Syst Biol 50:580–601.

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071.

R Development Core Team. 2007. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol 56:389–399.

Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007. SCaFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol Biol 7 Suppl 1:S2.

Rowe T. 2004. Chordate phylogeny and development. In: Cracraft J, Donoghue MJ, editors. Assembling the tree of life. Oxford: Oxford University Press. pp 384–409.

Ruppert EE. 2005. Key characters uniting hemichordates and chordates: Homologies or homoplasies? Can J Zool 83:8–23.

Schaeffer B. 1987. Deuterostome monophyly and phylogeny. Evol Biol 21:179–235.

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

Schubert M, Escriva H, Xavier-Neto J, Laudet V. 2006. Amphioxus and tunicates as evolutionary model systems. Trends Ecol Evol 21:269–277.

Sempere LF, Cole CN, McPeek MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. J Exp Zoolog B Mol Dev Evol 306:575–588.

Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431:67–71.

Steel M. 2005. Should phylogenetic models be trying to "fit an elephant"? Trends Genet 21:307–309.

Swalla BJ, Cameron CB, Corley LS, Garey JR. 2000. Urochordates are monophyletic within the deuterostomes. Syst Biol 49:52–64.

Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony and other methods version 4.0b10. Sunderland, MA: Sinauer.

Telford MJ, Budd GE. 2003. The place of phylogeny and cladistics in Evo-Devo research. Int J Dev Biol 47:479–490.

Vienne A, Pontarotti P. 2006. Metaphylogeny of 82 gene families sheds a new light on chordate evolution. Int J Biol Sci 2:32–37.

Wada H, Okuyama M, Satoh N, Zhang S. 2006. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. Evol Dev 8:370–377.

Wada H, Satoh N. 1994. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. Proc Natl Acad Sci USA 91:1801–1804.

Whelan S, Goldman N. 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 JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst Biol 52:528–538.

Wiens JJ. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Syst Biol 54:731–742.

Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. J Biomed Inform 39:34–42.

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