Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., ... Wörheide, G. (2015). Genomic data do not support comb jellies as the sister group to all other animals. Proceedings of the National Academy of Sciences of the United States of America, 112(50), 15402-15407. DOI: 10.1073/pnas.1518127112

Publisher's PDF, also known as Version of record
License (if available):
Unspecified
Link to published version (if available):
10.1073/pnas.1518127112

Link to publication record in Explore Bristol Research
PDF-document

This is the final published version of the article (version of record). It first appeared online via PNAS at http://www.pnas.org/content/112/50/15402.abstract. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights
This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
Genomic data do not support comb jellies as the sister group to all other animals

Davide Pisani, Walker Pett, Martin Dohrmann, Roberto Feuda, Omar Rota-Stabelli, Hervé Philippe, Nicolas Lartillot, and Gert Wörheide

*School of Earth Sciences, University of Bristol, Bristol BS8 1TG, United Kingdom; 1School of Biological Sciences, University of Bristol, Bristol BS8 1TG, United Kingdom; 1Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, CNRS, UMR 5558, 69622 Villeurbanne cedex, France; 1Department of Earth & Environmental Sciences & GeoBio-Center, Ludwig-Maximilians-Universität München, Munich 80333, Germany; 1Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125; 1Department of Sustainable Agro-Ecosystems and Bioresources, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all’Adige 38010, Italy; 1Centre for Biodiversity Theory and Modelling, USR CNRS 2936, Station d’Ecologie Expérimentale du CNRS, Moulis 82200, France; 1Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, QC, Canada H3C 3J7; and 1Bayrische Staatsanstalt für Paläontologie und Geologie, Munich 80333, Germany

Edited by Neil H. Shubin, The University of Chicago, Chicago, IL, and approved November 2, 2015 (received for review September 11, 2015)

Understanding how complex traits, such as epithelia, nervous systems, muscles, or guts, originated depends on a well-supported hypothesis about the phylogenetic relationships among major animal lineages. Traditionally, sponges (Porifera) have been interpreted as the sister group to the remaining animals, a hypothesis consistent with the conventional view that the last common animal ancestor was relatively simple and more complex body plans arose later in evolution. However, this premise has recently been challenged by analyses of the genomes of comb jellies (Ctenophora), which, instead, found ctenophores as the sister group to the remaining animals (the “Ctenophora-sister” hypothesis). Because ctenophores are morphologically complex predators with true epithelia, nervous systems, muscles, and guts, this scenario implies these traits were either present in the last common ancestor of all animals and were lost secondarily in sponges and placozoans (Trichoplax) or, alternatively, evolved convergently in comb jellies. Here, we analyze representative datasets from recent studies supporting Ctenophora-sister, including genome-scale alignments of concatenated protein sequences, as well as a genomic gene content dataset. We found no support for Ctenophora-sister and conclude it is an artifact resulting from inadequate methodology, especially the use of simplistic evolutionary models and inappropriate choice of species to root the metazoan tree. Our results reinforce a traditional scenario for the evolution of complexity in animals, and indicate that inferences about the evolution of Metazoa based on the Ctenophora-sister hypothesis are not supported by the currently available data.

Metazoan | Ctenophora | Porifera | phylogenomics | evolution

Resolving the phylogenetic relationships close to the root of the animal tree of life, which encompasses the phyla Porifera (sponges), Cnidaria (jellyfish, corals, and their allies), Ctenophora (comb jellies), Placozoa (the “plate animals” of the genus Trichoplax), and Bilateria (the group containing all remaining phyla), is fundamental to understanding early animal evolution and the emergence of complex traits [reviewed by Dohrmann and Wörheide (1)]. Traditionally, sponges have been recognized as the sister group to the remaining animals (the “Porifera-sister” hypothesis). Under this scenario, true epithelia (with belt desmosomes connecting neighboring cells) and extracellular digestion are conventionally thought to have been primitively absent in sponges, having evolved in the common ancestor of Placozoa, Ctenophora, Cnidaria, and Bilateria. Within this group, gap junctions between neighboring cells, ectodermal and endodermal germ layers, sensory cells, nerve cells, and muscle cells evolved only once in the common ancestor of Ctenophora, Cnidaria, and Bilateria. Thus, Porifera-sister is consistent with the view that the last common ancestor of the animals was relatively simple and more complex body plans evolved after sponges had separated from the other animal lineages. However, a series of recent papers (2–6) have challenged this view, arguing the earliest split in the animal phylogeny separated ctenophores from all other animals (the “Ctenophora-sister” hypothesis), implying a group uniting Porifera, Placozoa, Cnidaria, and Bilateria, for which no shared derived morphological characters (synapomorphies) are known. The Ctenophora-sister hypothesis, if correct, would require a major revision of our understanding of animal evolution because it would imply a more complicated evolutionary history, dominated by multiple independent gains and/or losses, of key metazoan characters (7, 8). Indeed, this hypothesis has already stirred a controversial discussion about multiple origins of nervous systems (9–11).

Although results from the first study supporting Ctenophora-sister (2) were questioned soon thereafter and suggested to be an artifact stemming from the inclusion of too few nonbilaterian species (12) and the use of too rapidly evolving genes (13), this hypothesis has recently been revived in several studies, including analyses of the first two complete ctenophore nuclear genomes, as well as transcriptomic datasets from numerous other ctenophore species (4–6). Here, we present analyses of key datasets from Ryan et al. (4), Moroz et al. (5), and Whelan et al. (6), and identify several problems in these studies, specifically the combined use of relatively simplistic models of molecular

**Significance**

Clarifying the phylogeny of animals is fundamental to understanding their evolution. Traditionally, sponges have been considered the sister group of all other extant animals, but recent genomic studies have suggested comb jellies occupy that position instead. Here, we analyzed the current genomic evidence from comb jellies and found no convincing support for this hypothesis. Instead, when analyzed with appropriate methods, recent genomic data support the traditional hypothesis. We conclude that the alternative scenario of animal evolution according to which ctenophores evolved morphological complexity independently from cnidarians and bilaterians or, alternatively, sponges secondarily lost a nervous system, muscles, and other characters, is not supported by the available evidence.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freedly available online through the PNAS open access option.

Data deposition: The scripts to run our gene content analyses have been deposited in Github, github.com/willpett/ctenophora-gene-content (apart from implementing the methods in R Bayes). 1To whom correspondence may be addressed. Email: davide.pisani@bristol.ac.uk or woeheide@lmu.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1518127112/-/DCSupplemental.
evolution and distantly related outgroups (the species used to root the animal tree), and not accounting for a data acquisition bias in the analysis of a gene presence/absence matrix (4). Our analyses correcting for these issues consistently failed to support Ctenophora as the sister group to the root (25, 27–31). We therefore conclude that previous support for Ctenophora-sister arose from uncorrected systematic biases. Given the absence of convincing evidence in support of Ctenophora-sister, downstream inferences based on this hypothesis should be considered with caution.

Addressing Biases in Phylogenetic Reconstruction

Potential Biases in Phylogenomic Datasets. When analyzing phylogenomic datasets, proper modeling of the amino acid substitution process is crucial because the use of overly simplistic models can lead to inaccurate phylogenetic inferences (reviewed in 13–17). For example, the monophyly of Chordata was not confidently resolved from phylogenomic data until sophisticated substitution models were applied (18, 19). The most commonly used models assume the substitution process is the same in all sites of a protein (site-homogeneous) (e.g., 20). Although these models have the advantage of allowing for fast computation, site homogeneity is biologically unrealistic because biochemical constraints (e.g., polarity, hydrophobicity) tend to limit the set of amino acids allowed at different sites in a protein. By not accounting for this effect, site-homogeneous models tend to overestimate the number of amino acids a site can accept, and therefore underestimate the probability of convergent evolution toward identical amino acids in unrelated species (17). This underestimation can lead to the misidentification of some convergent substitutions as evidence of shared common ancestry (reviewed in 21). To address this issue, site-heterogeneous models have been developed (22), which relax the homogeneity assumption to account for site-specific biochemical constraints. Although computationally more demanding, their increased capacity to identify convergent evolution is reflected in the better statistical fit these models generally provide to many empirical datasets (e.g., 23, 24). Here, we used a common statistical technique, Bayesian cross-validation, to compare the fit of site-homogeneous and site-heterogeneous models, and investigate whether previous studies that recovered Ctenophora-sister were influenced by the use of poorly fitting substitution models.

Outgroup selection (the species used to root the tree) can also strongly affect phylogenetic results (13, 25, 26). In particular, the inclusion of outgroups very distant from the ingroup can cause reconstruction artifacts by attracting fast-evolving (long-branched) ingroup species toward the root (25, 27–31). A typical solution is to introduce more closely related outgroups to “break up” the long branch leading to the ingroup, but long-branch attraction artifacts can be further minimized by also removing the distant outgroups. This effect has previously been documented, for example, in the case of the nematode worms in the context of testing the Ecdysozoa hypothesis against Cocolomata (32), as well as for nonbilaterian relationships (33), where the removal of distant outgroups stabilized ingroup relationships. Although the effect of outgroup composition was investigated in some previous studies supporting Ctenophora-sister, this test was done only in combination with site-homogeneous models (5, 6) or results obtained under site-heterogeneous models were considered unreliable (4). Here, we performed outgroup subsampling experiments under the best-fitting models and compared our results with previous studies to clarify whether the use of distant outgroups in combination with poorly fitting models might have influenced previous analyses that found support for Ctenophora-sister.

Table 1. Cross-validation likelihood scores under the models GTR, CAT, and CAT-GTR (relative to WAG, used as a reference model)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>GTR</th>
<th>CAT</th>
<th>CAT-GTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ryan-Choano</td>
<td>342 ± 32</td>
<td>1,282 ± 110</td>
<td>1,654 ± 93</td>
</tr>
<tr>
<td>Moroz-3D</td>
<td>342 ± 25</td>
<td>701 ± 85</td>
<td>1,060 ± 71</td>
</tr>
<tr>
<td>Whelan-6-Choano</td>
<td>560 ± 50</td>
<td>1,472 ± 153</td>
<td>2,376 ± 100</td>
</tr>
</tbody>
</table>

This analysis used three exemplar datasets taken from the studies of Ryan et al. (4), Moroz et al. (5), and Whelan et al. (6). Higher scores indicate a better empirical fit. In each case, the mean and SD are calculated over 10 independent replicates (Methods).
strongly supported Porifera-sister instead (Fig. 1A–C). In other words, under the better-fitting site-heterogeneous model, cteno-
phores emerge as sister to all other animals only when the most
distantly related outgroup, Fungi, is included, suggesting Cteno-
phora-sister most likely represents a long-branch attraction artifact.
Repeating the analyses under CAT-GTR also gave preliminary
support for Porifera-sister, but we were unable to run this analysis
to convergence within the time frame of this study (Fig. S1D).

Analysis of the Moroz et al. Phylogenomic Datasets. In the Pleuro-
brachia bachei genome study (5), the Ctenophora-sister hy-
pothesis was obtained from the analysis of two datasets, one of
which was constructed to maximize the number of species and the
other to maximize the number of proteins. Whereas the dataset
emphasizing protein sampling was broadly comparable to the
dataset of Ryan et al. (4), the dataset emphasizing species sampling
(Moroz-3D; Methods) was unique because it included the largest
number of ctenophores sampled thus far. Given that the same
authors have now assembled new datasets (6) that supersede the
protein-rich datasets of Moroz et al. (5) (discussed in the next
section), we only analyzed the species-rich dataset Moroz-3D.

The analysis of Moroz et al. (5) was conducted under the site-
homogeneous Whelan and Goldman (WAG) model (20), which
gave a tree congruent with the Ctenophora-sister hypothesis,
albeit with weak statistical support. However, analyzing the
Moroz-3D dataset using the similar but generally better-fitting
site-homogeneous Le and Gascuel (LG) model (44), we found a
different tree with a better likelihood score (Fig. S2A). This tree
united demosponges and glass sponges as the sister group of all
other animals, followed by ctenophores and then by calcareous
and homoscleromorph sponges. Although statistical support for
this branching order is very low (Fig. S2A), the same is true for
the tree found by Moroz et al. (5). Finally, an analysis of this
data set using the better-fitting site-heterogeneous CAT-GTR
model (45) supported demosponges, glass sponges, and homo-
scleromorphs as the sister group of all other animals, followed by
tenophores. However, in this tree, the calcareous sponges are
deeply nested within cnidarians (Fig. S2B), and, furthermore,
this analysis did not converge. The high dissimilarity between
these three trees and the uniformly low support obtained across
all analyses suggest the phylogenetic signal in this dataset is very
weak. This weakness of signal might, among other factors, be re-
lated to massive amounts of missing data, which reach 98% for the
calcareous sponges, the most unstable lineage in this dataset.
Furthermore, Moroz et al. (5) reported that using a subset of their
data consisting only of the most conserved proteins, they were
unable to resolve relationships of the major animal lineages and
could not reject Porifera-sister with statistical tests. Accordingly, we
conclude the Moroz-3D dataset does not provide sufficient signal
for resolving the position of Ctenophora.

Analysis of the Whelan et al. Phylogenomic Datasets. Whelan et al.
(6) assembled 25 datasets differing in protein and species selec-
tion, and recovered Ctenophora-sister with strong support from all
of them. Although they pointed out the importance of using site-
heterogeneous substitution models, as well as the impact of out-
group composition, they did not examine the combined effect of
these factors. That is, all of the outgroup-subsampled datasets
were analyzed exclusively using site-homogeneous substitution
models, whereas the analyses using the better-fitting site-heterogeneous
model were exclusively performed using the full set of outgroups, which
included distantly related Fungi.
We chose to base our analyses on their two most stringent datasets (Whelan-6 and Whelan-16; details are provided in Methods), because Whelan et al. (6) argue that these datasets are the most robust to systematic errors. Furthermore, these datasets were the only ones they analyzed with a site-heterogeneous model of sequence evolution (CAT-GTR). We performed outgroup subsampling analogous to Ryan et al. (4) on both of these datasets and analyzed the resulting six datasets under the site-heterogeneous CAT model (Methods). Consistent with our results from the Ryan et al. (4) datasets, analysis of the Whelan et al. (6) datasets gave decreased support for Ctenophora-sister because distantly related outgroups were excluded (Fig. 2 and Figs. S3 A–C and S4 A–C). At the same time, support for Porifera-sister increased (Fig. 2 B and C). These analyses were repeated for Whelan-6-Choa and Whelan-16-Choa under the computationally more demanding CAT-GTR model, which confirmed the lack of support for Ctenophora-sister with Whelan-6-Choa (Fig. S3D) and found strong support for Porifera-sister with Whelan-16-Choa (Fig. 1 and Fig. S4D). Although strong support for Porifera-sister is only provided by Whelan-16, this dataset is more conservative than the Whelan-6 dataset in that it has undergone an additional data filtering step in which further potentially paralogous sequences were removed. Because the inclusion of ctenophore paralogs would have the net effect of pushing Ctenophora toward the root of the tree, the stronger support for Porifera-sister after removing these sequences is consistent with the artifactual nature of Ctenophora-sister. Taken together, our results show the datasets of Whelan et al. (6) do not support Ctenophora-sister when both distantly related outgroups are excluded and better-fitting substitution models are used.

Whelan et al. (6) further argued that support for Coelenterata, a sister-group relationship of Ctenophora and Cnidaria, in the phylogenomic study of Philippe et al. (33), resulted from a bias caused by excessive reliance on ribosomal proteins. They illustrate the effect of this putative bias by reanalyzing the dataset of Philippe et al. (33) after excluding all ribosomal proteins, which yielded a tree that did not support Coelenterata and showed only moderate support for Porifera-sister. Here, we performed the same analysis, but excluded all nonchoanoflagellate outgroups, and recovered Coelenterata (albeit with weak support) and strong support for Porifera-sister (Fig. S5). These results suggest that the lack of support for Coelenterata and decreased support for Porifera-sister in Whelan et al.’s (6) reanalysis was not caused by the absence of a misleading signal specific to the generally slowly evolving ribosomal proteins but, instead, by a bias introduced by distant outgroups that becomes dominant when only the faster evolving nonribosomal proteins are retained.

Analysis of the Ryan et al. Gene Content Dataset. We analyzed the gene content dataset of Ryan et al. (4) both before (Fig. S6A) and after (Fig. 3 and Fig. S6B) applying an ascertainment bias correction to account for the fact that genes present in fewer than two species were not included in this dataset. Our estimate for the ratio of gene loss and gain rates was two orders of magnitude higher after accounting for unobserved losses (posterior mean = 189.4) compared with the uncorrected estimate (posterior mean = 1.94), indicating the original analysis of Ryan et al. (4) was severely biased. Indeed, we found the magnitude of this bias had a major impact on the inference of animal relationships. First, several well-established groups, such as Protostomia, Deuterostomia, Lophotrochozoa, Chordata, and Amnelida, which the original analysis of Ryan et al. failed to recover (figure 4 of ref. 4), were resolved with strong statistical support once a corrected model was used (Fig. 3 and Fig. S6D). Second, the strong support for Ctenophora-sister found in the uncorrected analysis (Fig. S6A) entirely disappeared, and strong support was obtained for Porifera-sister instead (Fig. 3 and Fig. S6B). Thus, our results show that the gene content dataset of Ryan et al. (4) contains strong signal in favor of Porifera-sister, and the Ctenophora-sister hypothesis only emerges, together with a number of other erroneous groups, when an uncorrected model of gene gain and loss is applied.

Discussion

We have analyzed representative genomic datasets presented by recent studies in support of the Ctenophora-sister hypothesis, which proposes that the first split on the metazoan tree of life was between comb jellies (Ctenophora) and all other animals (4–6), rather than between sponges (Porifera) and all other animals (the Porifera-sister hypothesis). We found that support for Ctenophora-sister disappears once steps are taken to minimize systematic errors, including the exclusion of distantly related outgroups and the use of better-fitting substitution models. The results of our phylogenomic analyses were further corroborated by our analysis of gene content data (4), which, after accounting for the data acquisition and filtering process, found strong support for Porifera-sister. Beyond our results, another recent study including only data from published whole-genome sequences (46) found support for Ctenophora-sister, but support for this hypothesis became insignificant when the data were analyzed under a biologically more realistic, site-heterogeneous model. Taken together, these results demonstrate the current lack of support for Ctenophora-sister, and therefore indicate that inferences about the origin of complex anatomical and genomic features in animals should not be based on an assumed position of Ctenophora as the sister group to all of the remaining animals.

Ctenophores are morphologically complex predators with true epithelia, nervous systems, muscle cells, and a digestive tract. These
characters are absent from sponges, and in light of our results, this
assumption should be interpreted as an ancestral condition, contrary
in the alternative scenario in which sponges lost these characters sec-
ondarily from a complex common ancestor of all animals [a discussion
regarding nervous systems is provided elsewhere (47)]. An alternative
interpretation under the Ctenophora-sister hypothesis would be that
some or all of these characters evolved convergently in ctenophores.
However, resolving the exact phylogenetic positions of Ctenophora
and Placozoa [discussions are provided elsewhere (1, 48, 49)] will be
also crucial to reconstruct the evolution of key characters, such as nervous
systems, muscles, and digestive tracts, in more detail. Although re-
solving the relationships among these taxa will require further re-
search, our results support a clade uniting all nonsponge animals,
which is consistent with a scenario in which the last common meta-
zooan ancestor was a relatively simple, possibly filter-feeding organism,
and complex traits related to a predatory lifestyle originated later.

One major result of the first whole-genome analyses of cteno-
phores (4, 5) was the finding that these organisms apparently lack
many genes or use different genes involved in the development of
anatomical structures, such as nervous systems, in other animal
groups. In light of the Ctenophora-sister hypothesis, this result has
been interpreted as evidence for convergent evolution, especially for
nervous systems (5, 11). However, other authors have interpreted the
same data differently, concluding they actually are consistent with a
single origin of nervous systems (9, 10). Likewise, analyses of the
ospin gene family, which is involved in light detection in animals,
as well as ion-channel proteins involved in mechanoreception, are
consistent with a close relationship between Ctenophora, Chordata,
and Bilateria (50, 51). Finally, the absence of many gene families,
coupled with massive lineage-specific expansions in others (6), sug-
gests ctenophore genomes may be extremely derived compared with
genomes of other animals. Thus, it may be difficult to draw con-
clusions about the homology or nonhomology of anatomical struc-
tures and cell types between ctenophores and other animals based
on the genes involved in their development. Future studies focused
on the evolution of gene content in animals will help to clarify the
relationship between the homology of similar structures and their
underlying genetic mechanisms (52-54).

Conclusions
The Ctenophora-sister hypothesis originally emerged as a
surprising byproduct of a study aimed at resolving bilaterian
relationships (2), and it has continued to grow in popularity fol-
lowing the recent publication of the first ctenophore nuclear ge-
onomes and accompanying phylogenetic results (4, 5). In our
assessment of these previous studies (4-6), we found that support
for Ctenophora-sister vanishes when steps are taken to minimize
systematic error. Thus, while strong support for Ctenophora-sister
may be obtained from phylogenomic datasets (2-6, 46, 55), our
analysis suggests these results are caused by undetected systematic
bias. Therefore, several recent studies whose conclusions are based
on the assumed accuracy of Ctenophora-sister (e.g., 26-29) should
be re-evaluated in light of these alternative phylogenetic hypotheses.
Our results do not support the currently emerging point of view
according to which the origin of complex characters, such as ner-
vous systems, was far more complicated than previously thought
(e.g., 7, 8). More broadly, our study highlights the danger of relying
solely on the presumed power of large datasets rather than on the
best possible modeling of the data and carefully designed phylo-
genetic analyses aimed at correcting systematic errors.

Methods
Dataset Selection. We considered a representative selection of datasets from
the studies of Ryan et al. (4), Moroz et al. (5), and Whelan et al. (6):

i) EST datasets of Ryan et al. (4), called est.choanimalia, est.holozoa, and est.
opisthokonta in the original study but, for consistency, called Ryan-
Choano, Ryan-Holo, and Ryan-Opisto here. These datasets include the
same set of genes but differ in the composition of outgroup species. Ryan-
Choano only includes choanoflagellates; Ryan-Holo includes additional,
more distantly related holozoa; and Ryan-Opisto also includes Fungi.

iv) Dataset of Moroz et al. (5) associated with their extended data figure
3D (Moroz-3D). This dataset was chosen because it has a substantially
improved sampling of ctenophores (11 vs. three) compared with the data-
sets of Ryan et al. (4), as well as other datasets presented by Moroz et al. (5).

v-x) Datasets 6 and 16 of Whelan et al. (6), each with a different outgroup
composition analogous to the Ryan et al. datasets (Whelan-6-Opisto,
-Holo, -Choano; Whelan-16-Opisto-Holo, -Choano). These datasets were
chosen because the authors stated that they maximize the number of
slowly evolving genes and minimize the number of certain paralogs (data-
sets 16 and the number of certain paralogs [dataset 3D]).

xii) Datasets composed of all nonribosomal proteins extracted with Whelan
et al. (6) from the Philippe et al. (33) dataset, with all choanofla-
gellate outgroups removed.

Model Testing. We used Bayesian cross-validation (36, 37) implemented in
PhyloBayes 3.3 (59) to compare the fit of the site-homogeneous WAG and GTR
models and the site-heterogeneous CAT and CAT-GTR models (20, 22). To
alleviate computational burden, we restricted these analyses to three exemplar
datasets: Ryan-Choano, Moroz-3D, and Whelan-6-Choano. Cross-validation
scores were computed by comparison with the WAG model. In addition, all
tests were trained under the tree topology favored by WAG, thus making the
test conservative in favor of the WAG model. Ten replicates were con-
sidered, each consisting of a random subsample of 10,000 sites for training the
model and 2,000 sites for calculating the cross-validation likelihood score.

Phylogenetic Reconstruction. We analyzed the Ryan et al. (4) datasets under
CAT either including or excluding X. bocki. Ryan-Choano was also analyzed
under CAT-GTR. All CAT and CAT-GTR analyses were performed using Phy-
loBayes MPI 1.5a (59). We analyzed Moroz-3D in RAxML 8.2.60 (60) using
WAG (20) and LG (44) with empirical amino acid frequencies (+F), as well as
under CAT-GTR with PhyloBayes MPI. We analyzed each of the Whelan et al.
(6) datasets under CAT in PhyloBayes MPI. To minimize computational bur-
den, only Whelan-6-Choano and Whelan-16-Choano were also analyzed
under CAT-GTR. The nonribosomal protein dataset of Philippe et al. (33) was
stripped of all nonchoanoflagellate outgroups and analyzed with CAT-GTR.
In all Bayesian analyses, among-site rate variation was modeled using a gamma
distribution (+G) discretized into four rate categories. In maximum likelihood
analyses, the 25-category CAT approximation (61) was used instead (note that
the CAT approximation in RAxML is unrelated to the CAT mixture model used
in PhyloBayes). Node support was evaluated using posterior probabilities in
Bayesian analyses and bootstrapping (100 replicates) in maximum likelihood
analyses. Convergence of Bayesian analyses was assessed by running two in-
dependent Markov chains and using the tspcount and tracerecomp tools from
PhyloBayes to monitor the maximum discrepancy in clade support (maxdiff),
the effective sample size (effsize), and the relative difference in posterior
mean estimates (rel_diff) for several key parameters and summary statistics of
the model. The appropriate number of samples to discard as “burnin” was
determined first by visual inspection of parameter trace plots, and then by optimizing convergence criteria. With the exception of the CAT-GTR analyses of Ryan-Chao and Moroz-3D, the maxdiff statistic was always <0.1 under the CAT model (<0.25 under the computationally more intensive CAT-GTR model); the minimum effective sample size was >50, and the maximum rel_diff statistic was <0.3 in all but one case (the CAT-GTR analysis of Whelan-6-Chao), which had a maximum rel_diff statistic <0.45.

**Gene Content Analysis.** We analyzed Ryan et al.'s (4) binary gene content dataset after applying a correction we developed specifically for the exclusion of genes present in fewer than two taxa, which we implemented in MrBayes 3.2.6 (version 3.2.6 r1067 62). We also analyzed this dataset after applying a correction for the exclusion of parsimony uninformative sites, which was already available in MrBayes (more details are provided in SI Methods and Fig. 5).

**ACKNOWLEDGMENTS.** We are indebted to the computational resources at the University of Bristol and the Iowa State University High Performance Computing Group. We thank the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities for the provisioning and support of Cloud computing infrastructure essential to this publication. Prof. Neumer is highly acknowledged for setting up and maintaining computational resources at Ludwig-Maximilians-Universität München Geobiology. We thank the associate editor and two anonymous reviewers for their constructive comments. We are also indebted to Prof. Eric Davidson for his help and encouragement while composing the manuscript. G.W. was funded by the German Research Foundation [Deutsche Forschungsgemeinschaft (DFG)] and the Ludwig-Maximilians-Universität München LMUexcellent program (Project MODELSPONGE) through the German Excellence Initiative. M.D. was funded through DFG Grants DO 1742/1-2, W.P. and N.L. were funded by the Agence Nationale de la Recherche (ANR) grant Ancestrone ANR-10-BINF-01-01.