Report

Current Biology

Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria

Highlights

- Model violations hinder inference of the phylogeny of Bilateria
- Methods predicted to cause errors place Xenacoelomorpha outside Nephrozoa
- Xenacoelomorpha is the sister group of Ambulacraria (hemichordates and echinoderms)
- Monophyletic deuterostomes (Chordata plus Xenambulacraria) are not clearly supported

Authors

Hervé Philippe, Albert J. Poustka, Marta Chiodin, ..., Richard R. Copley, Pedro Martinez, Maximilian J. Telford

Correspondence

m.telford@ucl.ac.uk

In Brief

Philippe et al. use new genetic data to study the phylogenetic affinities of the Xenacoelomorpha, a group of simple marine worms. Tackling potential sources of error strengthens support for a relationship between Xenacoelomorpha and the more complex echinoderms and hemichordates. The simple Xenacoelomorpha probably evolved from a complex ancestor.

Philippe et al., 2019, Current Biology *29*, 1818–1826
June 3, 2019
https://doi.org/10.1016/j.cub.2019.04.009

CellPress

CellPress

Current Biology
Report

Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria

Hervé Philippe,[1,2,29] Albert J. Poustka,[3,4,29] Marta Chiodin,[5,6] Katharina J. Hoff,[7] Christophe Dessimoz,[8,9,10,11,12] Bartlomiej Tomiczek,[8,9,13] Philipp H. Schiffer,[8,9] Steven Müller,[8,9] Daryl Domman,[14,28] Matthias Horn,[14] Heiner Kuhl,[15,16] Bernd Timmermann,[15] Noriyuki Satoh,[17] Tomoe Hikosaka-Katayama,[18] Hiroaki Nakano,[19]

(Author list continued on next page)

[1]Centre de Théorisation et de Modélisation de la Biodiversité, Station d'Ecologie Théorique et Expérimentale, UMR CNRS 5321, 09200 Moulis, France
[2]Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, QC H3C 3J7, Canada
[3]Evolution and Development Group, Max-Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany
[4]Dahlem Centre for Genome Research and Medical Systems Biology Envirnomental and Phylogenomics Group, Max-Planck-Straße 3, 12489 Berlin, Germany
[5]Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain
[6]School of Medicine, New York University, 435 E 30th Street, New York, NY 10016, USA
[7]Bioinformatics Group, Institute for Mathematics and Computer Science, University of Greifswald, Walther-Rathenau-Str. 47, 17487 Greifswald, Germany
[8]Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK
[9]Department of Computer Science, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK
[10]Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland
[11]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland
[12]Swiss Institute of Bioinformatics, Génopode, 1015 Lausanne, Switzerland
[13]Laboratory of Evolutionary Biochemistry, Intercollegiate Faculty of Biotechnology, University of Gdansk and Medical University of Gdansk, 80-307 Gdansk, Poland
[14]Centre for Microbiology and Environmental Systems Science, University of Vienna, 1090 Vienna, Austria
[15]Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany

(Affiliations continued on next page)

SUMMARY

Xenoturbella and the acoelomorph worms (Xenacoelomorpha) are simple marine animals with controversial affinities. They have been placed as the sister group of all other bilaterian animals (Nephrozoa hypothesis), implying their simplicity is an ancient characteristic [1, 2]; alternatively, they have been linked to the complex Ambulacraria (echinoderms and hemichordates) in a clade called the Xenambulacraria [3–5], suggesting their simplicity evolved by reduction from a complex ancestor. The difficulty resolving this problem implies the phylogenetic signal supporting the correct solution is weak and affected by inadequate modeling, creating a misleading non-phylogenetic signal. The idea that the Nephrozoa hypothesis might be an artifact is prompted by the faster molecular evolutionary rate observed within the Acoelomorpha. Unequal rates of evolution are known to result in the systematic artifact of long branch attraction, which would be predicted to result in an attraction between long-branch acoelomorphs and the outgroup, pulling them toward the root [6]. Other biases inadequately accommodated by the models used can also have strong effects, exacerbated in the context of short internal branches and long terminal branches [7]. We have assembled a large and informative dataset to address this problem. Analyses designed to reduce or to emphasize misleading signals show the Nephrozoa hypothesis is supported under conditions expected to exacerbate errors, and the Xenambulacraria hypothesis is preferred in conditions designed to reduce errors. Our reanalyses of two other recently published datasets [1, 2] produce the same result. We conclude that the Xenacoelomorpha are simplified relatives of the Ambulacraria.

RESULTS

Assembling Our Data Matrix

In order to provide the best chance of avoiding artifacts generated by data errors [7, 8], we assembled a new dataset of 1,173 genes (350,088 amino acid positions) from a balanced

Matthew L. Rowe,[20] Maurice R. Elphick,[20] Morgane Thomas-Chollier,[21] Thomas Hankeln,[22] Florian Mertes,[23] Andreas Wallberg,[24] Jonathan P. Rast,[27] Richard R. Copley,[25] Pedro Martinez,[5,26] and Maximilian J. Telford[8,9,30,*]
[16]Department of Ecophysiology and Aquaculture, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 301, 12587 Berlin, Germany
[17]Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan
[18]Natural Science Center for Basic Research and Development, Gene Science Division, Hiroshima University, Higashi-Hiroshima 739-8527, Japan
[19]Shimoda Marine Research Center, University of Tsukuba, Shimoda, Shizuoka 415-0025, Japan
[20]School of Biological & Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[21]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, PSL Université Paris, 75005 Paris, France
[22]Institute of Organismic and Molecular Evolution, Molecular Genetics and Genome Analysis, Johannes Gutenberg University Mainz, J.J. Becher-Weg 30a, 55128 Mainz, Germany
[23]Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Ingolstädter Landstraße 1, 85764 Neuherberg, Gemany
[24]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, 751 23 Uppsala, Sweden
[25]Laboratoire de Biologie du Développement de Villefranche-sur-mer (LBDV), Sorbonne Université, CNRS, 06230 Villefranche-sur-mer, France
[26]ICREA (Institut Català de Recerca i Estudis Avancats), Passeig Lluís Companys 23, 08010 Barcelona, Spain
[27]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, USA
[28]Present address: Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[29]These authors contributed equally
[30]Lead Contact
*Correspondence: m.telford@ucl.ac.uk
https://doi.org/10.1016/j.cub.2019.04.009

and rich selection of 59 taxa with just 23.5% missing data, giving us a matrix that is larger and more complete than any previously used to examine the question. Our new matrix has been carefully curated to minimize potential errors from sources including contamination and non-orthology. Alongside existing data, it includes new gene predictions from 6 partial genomes and 4 new transcriptomes.

New predicted protein sets were derived from partial genomes of *Xenoturbella bocki*, *Symsagittifera roscoffensis*, *Meara stichopi*, *Nemertoderma westbladi*, *Pseudaphanostoma variabilis*, and *Praesagittifera naikaiensis*; from new transcriptomes of *Xenoturbella bocki*, *Symsagittifera roscoffensis*, *Paratomella rubra*, and *Isodiametra pulchra*; and from published data available at the NCBI. To produce a balanced and computationally tractable dataset, we selected approximately equal numbers (6–8) of diverse species from the following clades: Xenacoelomorpha, Hemichordata, Echinodermata, Chordata, Lophotrochozoa, Ecdysozoa, Cnidaria, and Porifera plus the placozoan *Trichoplax adhaerens*. We omitted members of Ctenophora due to their well-documented fast evolutionary rate [9]. From these original sets of predicted protein sequences, we used orthologous matrix (OMA) to identify probable groups of orthologs covering the Metazoa [10, 11]. As OMA is rather stringent and can therefore omit valid orthologs, we added some missing orthologs using the 42 pipeline (https://bitbucket.org/dbaurain/42/downloads/). These putative orthologs were then tested for possible cross-contamination, non-orthology, and other issues likely to affect accurate phylogenetic reconstruction (see STAR Methods). Our final dataset contained 1,173 orthologous genes from 59 species of animals giving a total of 350,088 aligned amino acids.

Comparisons with Existing Recent Data Matrices

We compared our matrix to the two most recent studies addressing the question of the affinities of the Xenacoelomorpha in terms of data quality (percent of clades present in the concatenated tree that are also present in single gene trees) and quantity (number of amino acids present in the supermatrix: this number comes from the total number of amino acids in the matrix; if there were no missing data, this would equal length of alignment multiplied by the number of species). Our dataset is among the largest and of the highest quality: our single-gene trees recover >50% on average of the expected clades, whereas the average for the other datasets is 29% (maximum 39%; see Figure S4D). This indicates that our dataset likely contains fewer erroneous data (e.g., contaminants, paralogs, and frameshifts) than others and is therefore likely to contain more genuine phylogenetic signal: a prerequisite to infer phylogenies accurately [7, 9].

Analyses of Our Data Using Site Heterogeneous Models Show Limited Support for Xenambulacraria

We analyzed our complete matrix using a gene jackknife approach, which provides a conservative measure of clade support while being computationally tractable [9]. We used cross-validation to compare the fit of different models of sequence evolution on all datasets and found that the CATGTR model was the best fitting in all cases. We therefore used the CATGTR model of PhyloBayes [12] with a gamma correction for between site rate variability to analyze 100 subsamples each containing ~90,000 positions from the complete dataset. We found weak support (60% jackknife support) for a monophyletic grouping of Xenacoelomorpha and Ambulacraria. The second best supported topology grouped Xenacoelomorpha with Protostomes (24% jackknife support), and Nephrozoa had 13% jackknife support. Other uncontroversial clades in the tree were reconstructed with strong support (Figures 1A and 1B). In common with some previously published results [13, 14], the relationships between Chordata, Xenambulacraria, and Protostomia were

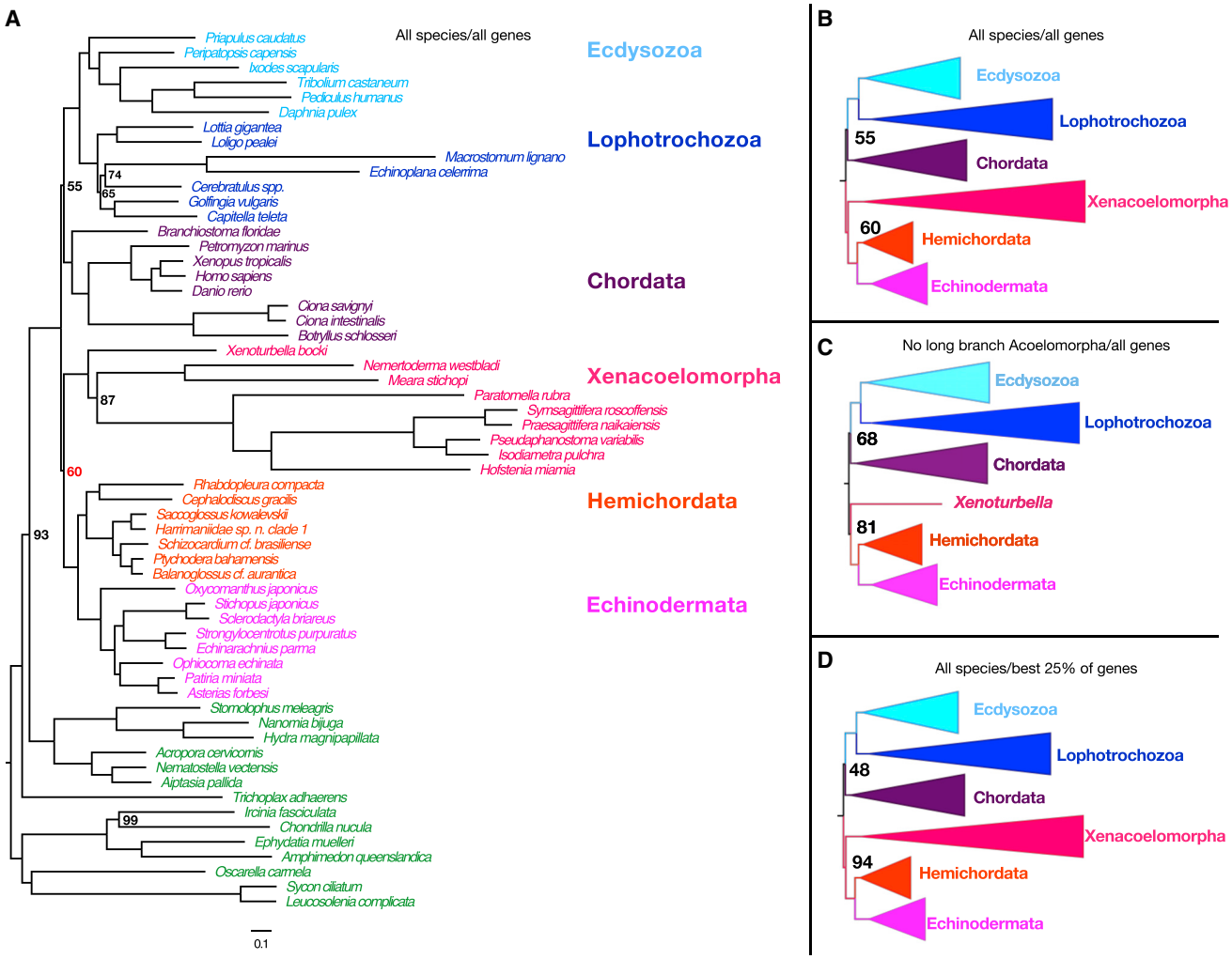

Figure 1. Support for Xenambulacraria Is Strengthened in Experiments Designed to Reduce Systematic Errors

(A) Full dataset using all 1,173 genes and 350,088 positions shows limited support (60% of jackknife replicates highlighted in red) for a sister group relationship between Xenacoelomorpha and Ambulacraria (Xenambulacraria).

(B) Summary figure of result in (A).

(C) Full dataset with long branched Acoelomorpha removed results in increased support for Xenambulacraria (81% jackknife support).

(D) Dataset of all species and the best 25% of genes (as measured by their ability to reconstruct known monophyletic groups) results in increased support for Xenambulacraria (94% jackknife support). Monophyletic deuterostome clade is not supported though the position of the Chordata is not reliably resolved in any analysis. All analyses used 50 or 100 jackknife replicates (support values shown to right of nodes) analyzed with PhyloBayes using the CATGTR+Gamma model. Major clades are indicated with corresponding colors. Jackknife proportions = 100% unless shown. The outgroups are shown in green. The scale bar indicates the inferred substitution per site. See also Figures S1, S2, S3, and S4 and Table S1.

unresolved—we did not reconstruct a monophyletic Deuterostomia (Chordata plus [Xen]ambulacraria).

Removing Fast-Evolving Acoelomorpha Reduces Support for Nephrozoa

Our approach to testing the possible effects of systematic error is to consider situations in which we can predict whether, if the tree is influenced by artifacts, nodal support will increase or decrease using different subsets of data or analytical methods. Manipulations expected to strengthen artifactual signal (less adequate models or subsets of data with an exaggerated systematic bias) are expected to increase support for the artifactual topology and vice versa, and the genuine phylogenetic signal should remain unaffected. One established approach for dealing with long branch attraction (LBA) is to remove the fastest evolving members of the group of interest [6]. If the Nephrozoa signal depends on an LBA artifact, we predict support for Nephrozoa would decrease in favor of Xenambulacraria when fast-evolving members of Xenacoelomorpha are removed. The Acoelomorpha have clearly evolved more rapidly than *Xenoturbella* (Figure 1A), and this difference seems to be mirrored in the more derived gene content of acoelomorph genomes [15, 16].

The validity of this approach requires the Xenacoelomorpha to be monophyletic. In our jackknife tree, and in previous phylogenomic analyses, the Xenoturbellida is strongly supported as the sister group of Acoelomorpha. This conclusion is

Table 1. Comparisons of Characteristics of Best and Worst Quarters of Genes from the Three Datasets

	This Study				Cannon et al.				Rouse et al.			
	Best Genes		Worst Genes		Best Genes		Worst Genes		Best Genes		Worst Genes	
Model	CATGTR	GTR	CATGTR	GTR	CATGTR	GTR	CATGTR	GTR	CATGTR	GTR	CATGTR	GTR
Diversity (*Z* score)	5.7	122.0	7.0	139.7	7.8	132.4	10.3	199.5	3.1	69.7	3.7	84.7
Max heterogeneity (*Z* score)	17.1	37.2	88.5	197.3	9.3	12.1	43.4	106.6	1.4	1.8	2.5	4.7
Mean heterogeneity (*Z* score)	120.0	152.7	208.2	325.7	50.5	68.6	169.0	276.8	6.9	7.9	29.6	39.4
Topology supported	X+A (94%)	X+PCA (100%)	X+PCA (48%)	X+PCA (99%)	X+A (42%)	X+PCA (100%)	X+P (76%)	X+PCA (100%)	X+A (50%)	X+PCA (93%)	X+PCA (50%)	X+PCA (87%)
Congruence score	0.87		0.53		0.80		0.44		0.8		0.44	
% recovered clades	72.58		37.38		60.45		25.17		47.40		3.47	
No. positions	87,791		87,562		84,276		84,462		98,630		98,579	
% missing data	24.75		22.74		39.89		36.39		43.86		40.80	
% constant positions	20.44		24.35		14.66		14.04		20.75		24.05	
Cross-validation	2,078 ± 82		3,539 ± 147		2,914 ± 113		4,960 ± 175		701 ± 62		997 ± 54	
Tree length	28.2		35.1		50.4		63.1		27.9		31.4	
Saturation	0.23		0.19		0.21		0.19		0.17		0.14	

For the data from this study, from Cannon et al. [2] and from Rouse et al. [1], we compare several aspects of the best and worst quarter of genes as ranked using our monophyly score. The first five rows show posterior predictive tests of diversity and heterogeneity of best and worst quarters of genes from the three datasets using site homogeneous (GTR) and heterogeneous (CATGTR) models of site evolution. For all three datasets and for all three tests, the CATGTR model provides a closer fit to the observed statistic than the site-homogeneous GTR model as estimated by the *Z* score shown here. There is a slightly better fit of model to data for the best genes compared to the worst genes. The three tests are computed with the readpb_mpi program of the PhyloBayes_mpi suite: diversity (site-specific amino acid preferences); max heterogeneity (maximal compositional heterogeneity observed across the taxa); and mean heterogeneity (mean squared heterogeneity across taxa). The remaining rows show comparisons of best and worst genes made using the CATGTR model: congruence score measures average monophyly score per gene and % recovered clades measures percentage of clans present in the super matrix LG+F+G tree recovered by single genes using the same model; in all cases, the best quarters are better. No. positions, % missing data, and number of constant positions have similar values between best and worst genes. Cross-validation scores show how much better the CATGTR model fits the data compared to the GTR model. For all datasets and partitions, trees based on the best genes are consistently shorter and slightly more saturated (saturation estimated as in [7] from the a0 parameter, using the CATGTR patristic distances) than those based on the worst genes.

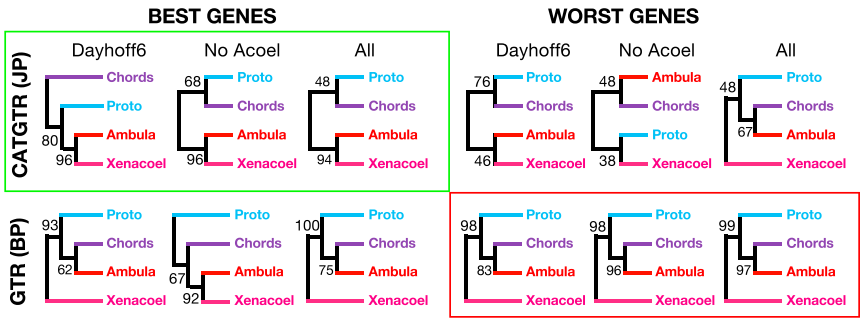

Figure 2. Best Genes and Best-Fitting Model Support Xenambulacraria Hypothesis under Different Conditions (Green Box)

Worst genes and less well-fitting model support the Nephrozoa hypothesis (red box). Summary trees with jackknife support values are shown for relationships between key clades for different methods of analysis. Best genes were selected by their ability to reconstruct known monophyletic groups. Top row is analyzed with better fitting site heterogeneous CATGTR+Gamma model. Bottom row is analyzed with less well-fitting site homogeneous GTR+Gamma model. "Dayhoff6" used Dayhoff recoding to reduce compositional bias. "No Acoel" excluded long-branched Acoelomorpha. "All" included all species with no data recoding. Ambula, Ambulacraria; BP, Bootstrap proportion; Chords, Chordata; JP, jackknife proportion; Proto, Protostomia; Xenacoel, Xenacoelomorpha.

further supported by a *Xenoturbella*/Acoelomorpha-specific rare genomic change involving their caudal/CDX ortholog (Figure S4E). If we therefore accept xenacoelomorphs as monophyletic, it is legitimate to use the slowly evolving member of the clade (*Xenoturbella*) as a representative of the Xenacoelomorpha, so reducing the effects of rapid evolution in the Acoelomorpha. When we removed the long branched Acoelomorpha but included the slower evolving *Xenoturbella* and repeated the jackknifing of the complete dataset, the support for Xenambulacraria increased to 81% (Figure 1C). This result is consistent with the support for Xenacoelomorpha being reduced in part due to LBA caused by the fast-evolving Acoelomorpha.

Stratifying Genes according to Phylogenetic Accuracy: Genes with Difficult-to-Extract Phylogenetic Signal Support Nephrozoa

A given gene is expected to vary in its ability to reconstruct the phylogeny of interest according to the method being used. More accurate genes ("better" genes with respect to the phylogenetic method used) will have more appropriate or more even rates of substitution or, more generally, some genes may fit the assumptions of the models used more closely than others; equally, some alignments may contain non-orthologous—e.g., contaminant—sequences. We reason that the genes that perform best at reconstructing known clades with a given method should be the most reliable when solving a related phylogenetic problem. To stratify the genes in our concatenated alignment according to their ability to reconstruct an accurate tree, we measured the capacity of each gene to reconstruct uncontroversial monophyletic groups of animals using two different methods that gave virtually identical results. After stratifying our genes, we concatenated them in order from best to worst and took the genes covering the first 25% of genes (best) and those covering the last 25% of genes (worst). The proportions of missing data and constant positions were similar for the two sub-datasets, but the worst genes evolved faster and were more saturated (Table 1); CATGTR is the best fitting model in each case, and improvement over GTR seems to be more important for the worst genes (Table 1). Posterior predictive checks show that the best genes violate the models much less than the worst genes (Table 1) but that even the best fitting CATGTR model does not explain the data well. We performed gene jackknife analysis with CATGTR using 50 samples of ~30,000 positions. The best performing genes according to our criterion supported Xenambulacraria (including the long-branched acoelomorphs) with 94% jackknife support (Figure 1D). The worst genes supported Nephrozoa with a weak 48% jackknife support, and we observed lower support for other clades across the tree in agreement with the expected difficulty in extracting phylogenetic signal from these genes. The best genes also support Xenambulacraria (jackknife proportion [JP] = 63%) when the short-branched *Xenoturbella* is removed, leaving just the fast-evolving Acoelomorpha (Figure S1B). Because the genes with the better phylogenetic to non-phylogenetic signal ratio consistently support Xenambulacraria, the likely explanation is that support for Nephrozoa is an artifact caused by the limitations of reconstruction methods when applied to problematic data.

Better-Fitting Models Support Xenambulacraria and Worse Models Support Xenambulacraria if Long-Branch Acoelomorphs Are Removed

Consistent with previous studies [5, 17, 18], the site heterogeneous CATGTR model we used has a better fit to our dataset than the site homogeneous LG and GTR models predominantly used by Cannon et al. [2] and Rouse et al. [1] (cross-validation score of 3,034 ± 152 and 2,001 ± 155, respectively). Although we have shown the best genes analyzed with CATGTR support Xenambulacraria, even with long-branch Acoelomorpha included, analyzing this dataset with less well-fitting site homogeneous GTR models supports Nephrozoa (100% bootstrap support). When reanalyzing the best data after removing the long-branched Acoelomorpha, however, even the less well-fitting GTR model supports Xenambulacraria (92% bootstrap support; Figure 2). For the worst performing genes, all analyses (CATGTR and GTR with or without Acoelomorpha) supported Nephrozoa (Figure 2). Data and analyses that are better by specified, measurable, objective criteria consistently result in increased support for Xenambulacraria.

Addressing the Effects of Compositional Bias Reduces Support for Nephrozoa

After LBA, probably the best known source of systematic error is compositional bias, in which a systematic tendency of substitutions toward certain amino acids in subsets of taxa affects tree

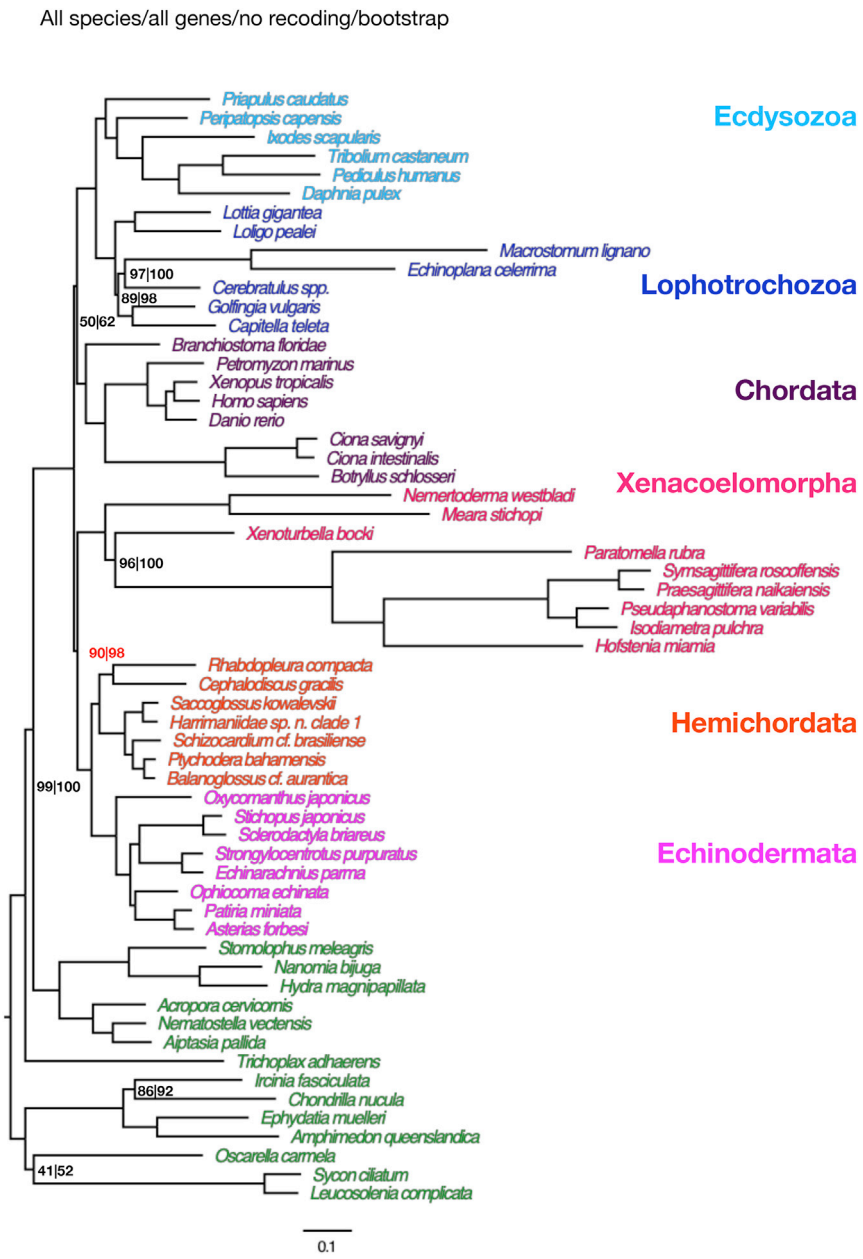

Figure 3. Dayhoff Recoding to Reduce Compositional Bias and Saturation Increases Support for Xenambulacraria

PhyloBayes jackknife and bootstrap analyses of all genes and all taxa using CATGTR and Dayhoff recoding. The jackknife tree is shown though the bootstrap topology was identical, and branch lengths were almost identical. Jackknifing used 50 replicates of 30,000 amino acids. Jackknife proportions (first number) and bootstrap proportions (second number) for nodes with less than 100% support for either measure are shown to the right of node supported. Bootstrap proportions are consistently higher, suggesting jackknifing provides a conservative measure of support. Xenambulacraria support is highlighted in red. The scale bar indicates the inferred substitution per site.

reconstruction [19]. Considering the possibility that compositional biases in the proportions of amino acids found in different species were inadequately accounted for by the models used, we looked for evidence of the existence of compositional bias by using posterior predictive checks in PhyloBayes to compare real amino acid frequencies of the 59 species in our data with their mean values under the null distribution predicted by the best fitting CATGTR model. A strong compositional bias was observed in our data although not specifically in Xenacoelomorpha. Interestingly, part of the superiority of the better genes discussed previously may be explained by the lower compositional bias we observe in the best 25% of data compared to the worst 25% (mean squared heterogeneity—best genes = ~100; worst genes = ~190). If compositional bias is contributing to the support for Nephrozoa, then reducing the effects of this bias would be predicted to lower support for Nephrozoa. To minimize the effects of species-specific compositional bias, we recoded the amino acids in our alignment using a reduced alphabet that gathers similar (and frequently substituted) amino acids into the following 6 "Dayhoff" groups (A,G,P,S,T), (D,E,N,Q), (H,K,R), (F,Y,W), (I,L,M,V), and (C). Recoding also tends to reduce model violations and saturation, as frequently substituting amino acids are consolidated into a single character state [19]. We reran the jackknife analyses of the complete dataset using the recoded data in PhyloBayes [12]. Using all species and all genes, jackknife support for Xenambulacraria increased from 61% to 90%, suggesting that compositional bias affects tree reconstruction and specifically reduces support for Xenambulacraria (Figure 3). We repeated this analysis using a bootstrapping approach instead of jackknifing, and the support for Xenambulacraria was found to be 98%. This increase is in line with other evidence indicating the relatively conservative nature of jackknife support values.

The Effects of Model and Data Testing Are Not Dataset Specific

One possible criticism of our findings is that they depend on the particular subset of genes and taxa used. We repeated our analyses using the datasets of Cannon et al. [2] and Rouse et al. [1]. For each test (removing long-branched taxa, stratifying genes according to phylogenetic accuracy, and recoding to reduce compositional bias), we observed the same direction of change as we observe in our data, albeit with lower support values, especially for the taxon-poor Rouse et al. data [1] (see Figure S4).

Although Cannon et al. [2] analyzed their data with long-branched Acoelomorpha omitted, they used the site-homogeneous LG model and recovered the Nephrozoa tree. Using CATGTR on the same data, we recovered the Xenambulacraria tree (Figure S2). With the same results coming from three large, independently assembled datasets, it is reasonable to conclude that the support for Xenambulacraria cannot be explained by the choices made during dataset assembly.

DISCUSSION

Determining the correct phylogenetic position of the Xenacoelomorpha has significant implications for our understanding of their evolution and that of the Metazoa. If Xenacoelomorpha diverged prior to other bilaterian animals, then this could explain their relative morphological simplicity and lack, for example, of several bilaterian Hox genes and microRNAs [20–22]. Under the assumption of such an ''early diverging'' scenario, xenacoelomorphs were naturally considered to be of particular interest as a branch intermediate between non-bilaterians (such as Cnidaria) and Nephrozoa [23, 24]. If, on the other hand, xenacoelomorphs are the sister group of the Ambulacraria, their simplicity, both morphological and genetic, must have been derived from a more complex ancestor by a process of character loss. If we accept that the Xenambulacraria clade is real, we should expect additional evidence for this relationship to remain in the embryology, morphology, and genomes of these animals, and such evidence would be a valuable corroboration of our results. Although it seems that the branch separating the Xenambulacraria from other Bilateria is short, it would still be predicted that certain characters uniting these taxa exist. Accordingly, the occurrence of neuropeptides in xenacoelomorphs related to echinoderm SALMFamides [25] has been reported previously based on immunohistochemical evidence [26, 27] to add to other known shared molecular characters [5, 28, 29].

One surprising result from our work is the lack of support for a monophyletic clade of deuterostomes when using site-heterogeneous models—the relationships between chordates, Xenambulacraria, and protostomes are essentially unresolved. Although the majority of our analyses recover a monophyletic group of chordates plus protostomes, the support values are very low, meaning there is no solid evidence to refute the traditional protostome and deuterostome dichotomy. All possible relationships between chordates, protostomes, and Xenambulacraria are observed in different analyses (see extended info). This observation nevertheless implies an extremely short branch between the bilaterian common ancestor (Urbilateria) and the deuterostomes. If the deuterostomes are ultimately shown to be monophyletic, then the short branch leading to the deuterostome common ancestor, Urdeuterostomia, suggests it should have much in common with Urbilateria. If the deuterostomes do prove to be paraphyletic, then Urbilateria and Urdeuterostomia must be considered synonymous, and this result has significant implications for our understanding of the characteristics of the common ancestor of Bilateria. Given that the internal branches separating the Xenambulacraria, Chordata, and Protostomia are short, larger datasets and more refined methodologies (e.g., [30]) are required to adequately test the deuterostome monophyly.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Xenoturbella bocki genome
  - Symsagittifera roscoffensis genome
  - Amplifying genomes of small acoels
  - Paratomella rubra transcriptome
  - Isodiametra pulchra transcriptome
  - Initial contaminant cleaning
  - Removing redundancy
  - Initial ortholog predictions using OMA
  - Reducing missing data, adding species and initial cleaning using 42 software
  - Removing potential contaminants
  - Dataset quality
  - Phylogenetic inference
  - Stratifying genes according to support for known monophyletic groups
  - Dayhoff recoding
  - Posterior Predictive Analyses (ppred)
  - Carbon footprint calculations
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Jackknife procedure and tests for reliability
  - Model fit
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.04.009.

ACKNOWLEDGMENTS

This research was supported by European Research Council grant (ERC-2012-AdG 322790) to M.J.T., by the Labex TULIP (ANR-10-LABX-41) to H.P., by the Deutsche Forschungsgemeinschaft (DFG Ha2103/4) and Johannes Gutenberg University Center of Computational Sciences Mainz (CSM) to T.H., by the Swiss National Science Foundation (150654) to C.D., by the OIST internal fund to N.S., and by the JSPS Grant-in-Aid for Young Scientists (A) (JP26711022) to H.N. Computations were performed at MPI Berlin, at the University College London Computer Science cluster, at the University College London Legion and Grace supercomputers, at the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics, and on the Mp2 and Ms2 supercomputers of the Université de Sherbrooke, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), the Ministère de l'Économie, de la Science et de l'Innovation du Québec (MESI), and the Fonds de Recherche du Québec — Nature et Technologies (FRQ-NT). A rough estimation of the carbon footprint for this work is >7 tons of CO_2 for travel and >260 tons CO_2 for computations (see STAR Methods for details). We thank the staff of Sven Lovén Centre for Marine Infrastructure, University of Gothenburg for help collecting specimens.

AUTHOR CONTRIBUTIONS

Initial Concept, A.J.P., M.J.T., R.R.C., and P.M.; Collecting Specimens, H.N., A.W., M.C., and P.M.; Sequencing of New Genomes and Transcriptomes,

A.J.P., T.H., P.M., M.C., N.S., T.H.-K., M.J.T., R.R.C., B. Timmermann, and H.K.; Genome and Transcriptome Assembly and Gene Predictions, A.J.P., K.J.H., and H.K.; Curation of Sequence Dataset to Remove Bacterial and Algal Contigs, D.D. and M.H.; OMA Orthology Analyses, C.D., B. Tomiczek, and S.M.; "42" Analyses and Alignment Cleaning, H.P.; Phylogenetic Analyses, H.P., M.J.T., A.J.P., C.D., and P.H.S.; Additional Comparative Analysis of Sequence Data, R.R.C., M.R.E., M.L.R., M.T.-C., T.H., J.P.R., and F.M.; Drafting Manuscript, M.J.T. and H.P.; Commenting on Manuscript, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 13, 2019
Revised: March 19, 2019
Accepted: April 3, 2019
Published: May 16, 2019

REFERENCES

1. Rouse, G.W., Wilson, N.G., Carvajal, J.I., and Vrijenhoek, R.C. (2016). New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. Nature *530*, 94–97.
2. Cannon, J.T., Vellutini, B.C., Smith, J., 3rd, Ronquist, F., Jondelius, U., and Hejnol, A. (2016). Xenacoelomorpha is the sister group to Nephrozoa. Nature *530*, 89–93.
3. Bourlat, S.J., Nielsen, C., Lockyer, A.E., Littlewood, D.T., and Telford, M.J. (2003). *Xenoturbella* is a deuterostome that eats molluscs. Nature *424*, 925–928.
4. Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., et al. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature *444*, 85–88.
5. Philippe, H., Brinkmann, H., Copley, R.R., Moroz, L.L., Nakano, H., Poustka, A.J., Wallberg, A., Peterson, K.J., and Telford, M.J. (2011). Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. Nature *470*, 255–258.
6. Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., and Lake, J.A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature *387*, 489–493.
7. Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. *9*, e1000602.
8. Laurin-Lemay, S., Brinkmann, H., and Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. Curr. Biol. *22*, R593–R594.
9. Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., et al. (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr. Biol. *27*, 958–967.
10. Altenhoff, A.M., Glover, N.M., Train, C.M., Kaleb, K., Warwick Vesztrocy, A., Dylus, D., de Farias, T.M., Zile, K., Stevenson, C., Long, J., et al. (2018). The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. Nucleic Acids Res. *46* (D1), D477–D485.
11. Train, C.M., Glover, N.M., Gonnet, G.H., Altenhoff, A.M., and Dessimoz, C. (2017). Orthologous Matrix (OMA) algorithm 2.0: more robust to asymmetric evolutionary rates and more scalable hierarchical orthologous group inference. Bioinformatics *33*, i75–i82.
12. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.
13. Lartillot, N., and Philippe, H. (2008). Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 1463–1472.
14. Marlétaz, F., Peijnenburg, K.T.C.A., Goto, T., Satoh, N., and Rokhsar, D.S. (2019). A new spiralian phylogeny places the enigmatic arrow worms among gnathiferans. Curr. Biol. *29*, 312–318.e3.
15. Gavilán, B., Perea-Atienza, E., and Martínez, P. (2016). Xenacoelomorpha: a case of independent nervous system centralization? Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150039.
16. Perea-Atienza, E., Gavilán, B., Chiodin, M., Abril, J.F., Hoff, K.J., Poustka, A.J., and Martinez, P. (2015). The nervous system of Xenacoelomorpha: a genomic perspective. J. Exp. Biol. *218*, 618–628.
17. Schiffer, P.H., Robertson, H.E., and Telford, M.J. (2018). Orthonectids are highly degenerate annelid worms. Curr. Biol. *28*, 1970–1974.e3.
18. Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., Škunca, N., Rawlinson, K.A., Cameron, C.B., Beli, E., et al. (2015). A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Curr. Biol. *25*, 1347–1353.
19. Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Wörheide, G., and Pisani, D. (2017). Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. Curr. Biol. *27*, 3864–3870.e4.
20. Fritzsch, G., Böhme, M.U., Thorndyke, M., Nakano, H., Israelsson, O., Stach, T., Schlegel, M., Hankeln, T., and Stadler, P.F. (2008). PCR survey of Xenoturbella bocki Hox genes. J. Exp. Zool. B Mol. Dev. Evol. *310*, 278–284.
21. Cook, C.E., Jiménez, E., Akam, M., and Saló, E. (2004). The Hox gene complement of acoel flatworms, a basal bilaterian clade. Evol. Dev. *6*, 154–163.
22. Sempere, L.F., Martinez, P., Cole, C., Baguñà, J., and Peterson, K.J. (2007). Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. Evol. Dev. *9*, 409–415.
23. Hejnol, A., and Martindale, M.Q. (2009). Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel *Convolutriloba longifissura*. BMC Biol. *7*, 65.
24. Hejnol, A., and Martindale, M.Q. (2008). Acoel development indicates the independent evolution of the bilaterian mouth and anus. Nature *456*, 382–386.
25. Elphick, M.R. (2014). SALMFamide salmagundi: the biology of a neuropeptide family in echinoderms. Gen. Comp. Endocrinol. *205*, 23–35.
26. Dittmann, I.L., Zauchner, T., Nevard, L.M., Telford, M.J., and Egger, B. (2018). SALMFamide2 and serotonin immunoreactivity in the nervous system of some acoels (Xenacoelomorpha). J. Morphol. *279*, 589–597.
27. Stach, T., Dupont, S., Israelson, O., Fauville, G., Nakano, H., Kånneby, T., and Thorndyke, M. (2005). Nerve cells of *Xenoturbella bocki* (phylum uncertain) and *Harrimania kupfferi* (Enteropneusta) are positively immunoreactive to antibodies raised against echinoderm neuropeptides. J. Mar. Biol. Assoc. U.K. *85*, 1519–1524.
28. de Mendoza, A., and Ruiz-Trillo, I. (2011). The mysterious evolutionary origin for the GNE gene and the root of bilateria. Mol. Biol. Evol. *28*, 2987–2991.
29. Chang, Y.C., Pai, C.Y., Chen, Y.C., Ting, H.C., Martinez, P., Telford, M.J., Yu, J.K., and Su, Y.H. (2016). Regulatory circuit rewiring and functional divergence of the duplicate admp genes in dorsoventral axial patterning. Dev. Biol. *410*, 108–118.
30. Dang, T., and Kishino, H. (2019). Stochastic variational inference for Bayesian phylogenetics: A case of CAT model. Mol. Biol. Evol. *36*, 825–833.
31. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. *62*, 611–615.
32. Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

33. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience *1*, 18.
34. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. *268*, 78–94.
35. Sommer, D.D., Delcher, A.L., Salzberg, S.L., and Pop, M. (2007). Minimus: a fast, lightweight genome assembler. BMC Bioinformatics *8*, 64.
36. Brady, A., and Salzberg, S.L. (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nat. Methods *6*, 673–676.
37. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.
38. Amemiya, C.T., Alföldi, J., Lee, A.P., Fan, S., Philippe, H., Maccallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. Nature *496*, 311–316.
39. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
40. Altenhoff, A.M., Gil, M., Gonnet, G.H., and Dessimoz, C. (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. PLoS ONE *8*, e53786.
41. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
42. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. *10*, 210.
43. Roure, B., Rodriguez-Ezpeleta, N., and Philippe, H. (2007). SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. *7* (*Suppl 1*), S2.
44. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. *21*, 1095–1109.
45. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Xenoturbella bocki	Gullmarsfjord, Sweden	NCBI:txid242395
Symsagittifera roscoffensis	Beaches off Roscoff, France	NCBI:txid84072
Meara stichopi	From pharynx of *Stichopus*, off Bergen Norway	NCBI:txid84115
Nemertoderma westbladi	Gullmarsfjord, West coast Sweden	NCBI:txid172109
Pseudaphanostoma variabilis	Hållö close to Smögen, West coast Sweden	NCBI:txid2510493
Praesagittifera naikaiensis	Onomichi, Hiroshima, Japan	NCBI:txid31270
Paratomella rubra	Sand from Filey bay, Yorkshire, UK	NCBI:txid90914
Isodiametra pulchra	Lab strain from Innsbruck, Austria	NCBI:txid504439
Deposited Data		
Alignments, software and trees	GitHub	https://github.com/MaxTelford/Xenacoelomorpha2019
Genome and transcriptome assemblies	https://figshare.com/search project number	PRJNA517079
Raw data for novel sequences.	Sequence Read Archive BioProject	PRJNA517079
Software and Algorithms		
PhyloBayes	[31]	http://www.atgc-montpellier.fr/phylobayes
Flash	[32]	http://ccb.jhu.edu/software/FLASH/index.shtml
SOAPfilter_v2.0	[33]	https://github.com/tanghaibao/jcvi-bin/blob/master/SOAP/SOAPfilter_v2.0
SOAPGapcloser v1.12	[33]	http://soap.genomics.org.cn/soapdenovo.html
Genescan	[34]	http://genes.mit.edu/GENSCAN.html
Soapdenovo2	[33]	https://github.com/aquaskyline/SOAPdenovo2
minimus2	[35]	https://github.com/sanger-pathogens/circlator/wiki/Minimus2-circularization-pipeline
PhymmBL	[36]	https://ccb.jhu.edu/software/phymmbl/index.shtml
CD-Hit	[37]	http://weizhongli-lab.org/cd-hit/
42		https://bitbucket.org/dbaurain/42/downloads
HmmCleaner version 1.8	[38]	https://metacpan.org/pod/HmmCleaner.pl
Mafft	[39]	https://mafft.cbrc.jp/alignment/software/
OMA	[40]	https://omabrowser.org
RAxML	[41]	https://cme.h-its.org/exelixis/software.html
BMGE	[42]	ftp://ftp.pasteur.fr/pub/gensoft/projects/BMGE/
SCaFoS	[43]	http://megasun.bch.umontreal.ca/Software/scafos/scafos.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Maximilian J. Telford (m.telford@ucl.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Xenoturbella bocki were collected from mud dredged at approx. 60 m depth in Gullmarsfjord, Sweden.

Symsagittifera roscoffensis were collected from intertidal regions of beaches in region of Roscoff, France.

Meara stichopi were collected by dissection from the pharynx of the sea cucumber *Stichopus sp*. The sea cucumbers were collected in the sea close to Bergen, Norway.

Pseudaphanostoma variabilis were found in sediment collected close to the island of Hållö close to Smögen, West coast Sweden.

Praesagittifera naikaiensis were collected from intertidal sand on the island of Mukaishima, Onomichi, Hiroshima, Japan.

Paratomella rubra were collected from intertidal sands of Filey bay, Yorkshire, United Kingdom.

Isodiametra pulchra came from a lab strain from the University of Innsbruck, Austria.

METHOD DETAILS

Xenoturbella bocki genome

Independent Illumina fragment libraries were made from two single animals, which had been starved for at least 3 months in the presence of Penicillin, Streptomycin and Gentamycin antibiotics to minimize environmental bacterial contaminations. The fragment libraries had insert sizes of ~200bp and ~150 bp and were sequenced as single paired reads with read length of 36-100bp. Overlapping paired reads were joined using flash [32]. The majority of sequences were obtained from these two libraries of which 10 lanes were sequenced.

Mate pair libraries were constructed from DNA isolated from several animals with insert sizes of 700, 1,000, 1,500 and 2,000 bp. After standard Illumina filtering all sequences shorter than 31bp were discarded. All reads were subsequently filtered for adaptor sequences, PCR duplicates and quality with SOAPfilter_v2.0 (https://github.com/tanghaibao/jcvi-bin/blob/master/SOAP/SOAPfilter_v2.0) using standard settings except setting the insert sizes and the appropriate asci quality shifts. A total of 731,057,046 reads were assembled simultaneously using SOAPdenovo (v2) [33] using settings –K 31 -M3 –F –U -g200. A total of 108,063,238 bp were assembled in a total of 21,594 scaffolds. The average scaffold length was 5004 bp, the longest scaffold had a size of 317,597 bp. Including contigs not merged into scaffolds the total sequence size was 119,097,168 bp with an average length of 1210 bp an N50 of 22,208 and an N90 of 443bp. Additional gaps were filled using SOAP Gapcloser v1.12 (http://soap.genomics.org.cn/soapdenovo.html).

Using the human matrix, Genescan [34] was used to generate predictions of coding regions resulting in 23 Mb of protein coding sequence (N50: 1872 bp) in 21,769 predicted protein or peptide sequences, which were subsequently used for phylogenomic analyses.

Symsagittifera roscoffensis genome

A standard fragment Illumina library was made from a pool of symbiont free hatchlings, which were raised in artificial sea water in the presence of antibiotics. Reads were processed as described for *Xenoturbella* above. 526,232,442 reads were assembled using SOAPdenovo2 (-M3, -R, –d1, -K31) and the Celera assembler using the settings for large and heterozygous genomes. Single gene analyses indicated that the two assemblers had different qualities in different regions of the genome. Hence the entire Soap assembly and the Celera assembly using its contigs and degenerate contigs larger than 500 bp were jointly assembled using minimus2 [35]. Although the total assembled genome size of about 1 Gbp from the SOAPdenovo assembly was reduced to about 450 Mb of assembled sequence many single gene analyses and PCR amplifications indicated that many more genes are represented in the joint assembly in significantly longer gene models. The joint assembly had an N50 of 2,905bp and a N90 of 587bp. Analysis of missing sequences indicated that most of the removed part is composed of repetitive sequence. The total number of predictions for coding sequences is 113,993 and comprising a total of 52Mb. A transcriptome was also sequenced from *S. roscoffensis* mixed stage embryos using standard methods.

Amplifying genomes of small acoels

Due to their small sizes one whole animal each of *Meara stichopi*, *Nemertoderma westbladi*, and *Pseudaphanostoma variabilis* were used without prior DNA extraction to directly amplify genomic DNA using the illustra Genomphi V2 DNA amplification Kit (GE Healthcare Nr.: 25-6600-30). Amplified DNA was cleaned by Isopropanol precipitation and shared to 1.5-3 kb fragments using speed code SC6 on the Hydroshear DNA Shearing Device (Thermo Fisher Scientific). After additional cleaning and quantification 1 μg DNA from each animal was used to generate standard illumina fragment libraries and these were sequenced as paired end with sequence length 100 bp. Sequence data have been submitted to the European Nucleotide Archive (ENA) under accession number PRJEB25577.

Nemertoderma westbladi was collected from mud at the site "Telekabeln" in the Gullmarsfjord in July 2009. For *Nemertoderma westbladi*, 800,863,374 reads equalling ~80 Gb of sequence were used for the genome assembly using SOAPdenovo2. The best results were obtained using the settings -K39 -d0 –M 3 –map 45. The assembly comprised about 205 Mb with an N50 of about 380 bp. 80,966 gene predications resulted in 38Mb of coding sequence.

For *Meara stichopi* 1,167,743,394 reads (~110 Gb) were read. An assembly was generated using standard settings and –K –M 3. The assembly had a total size of about 1.4 Gbp and an N50 of 1.1 Kb. A total of 130,115 protein or peptide fragments were predicted comprising 37Mb of coding sequence.

Pseudaphanostoma variabilis was collected from shell gravel near the island Hållö close to Smögen in July 2009. The *Pseudaphanostoma variabilis* genome was assembled from 672,950,533 reads with the SOAPdenovo2 settings –K 31 –d 0 –M 3 –map 36 and resulted in an assembly size of about 413 Mb. 115,245 gene predictions comprised 45 Mb of coding sequence.

The *Praesagittifera naikaiensis* genome was sequenced and assembled at the Okinawa Institute of Science and Technology. 1,148,317 sequences with a total size of about 1.2 Gb and an N50 of 4,452 bp resulted in 400,106 gene predictions comprising 233Mb of coding sequence.

Paratomella rubra transcriptome

Specimens of the acoel *Paratomella rubra* were collected from intertidal sand in Filey Bay, Yorkshire, UK. RNA was prepared and sequenced, the transcriptome was assembled and cross-contaminants were removed and proteins predicted as described in [18]. Data available in the NCBI Short Read archive: SRX3470480.

Isodiametra pulchra transcriptome

Specimens of the acoel *Isodiametra pulchra* were harvested from a lab stock provided by B Egger, Innsbruck. RNA was prepared and sequenced, the transcriptome was assembled and cross-contaminants were removed and proteins predicted as described in [18]. Data available in the NCBI Short Read archive SRX3469680.

Initial contaminant cleaning

All sequences were scanned for contaminating bacterial sequences using the PhymmBL program [36]. Sequences were additionally clustered based on tetranucleotide frequencies using an emergent self-organizing map (ESOM).

Removing redundancy

We translated gene predictions from genomes and transcriptomes into protein sequence and, when both present from a given species, we joined both predictions and clustered using CD-HIT with a 97% identity threshold [37], resulting in non-redundant proteomes for each species. We obtained 32,456 complete gene predictions in *Symsagittifera roscoffensis,* 35,867 complete gene predictions in *Meara stichopi,* 23,233 complete gene predictions in *Nemertoderma westbladi,* 27,378 complete gene predictions in *Pseudophanostoma variabilis,* 24,329 complete gene predictions in *Paratomella rubra,* 19,206 complete gene predictions in *Xenoturbella bocki*.

Initial ortholog predictions using OMA

Non-redundant peptide datasets from 67 species including 9 Xenacoelomorpha species, 8 Chordata, 15 Ambulacraria, and 13 Protostomia and 22 non-Bilateria organisms were processed by the OMA standalone software version 0.99w [40], using default settings. This identified 245,524 Orthologous Groups (OGs)—sets of genes in which all members are orthologous to all other members. From these, we selected the 3,683 OGs which had a minimum of 34 species represented (at least 50% of all species), and further filtered 1,665 OGs containing at least one member of Xenoturbellida and Nemetodermatida and Acoela.

Reducing missing data, adding species and initial cleaning using 42 software

Transcriptomic data from 77 species were then incorporated into the 1,665 previously assembled core orthologous clusters using a multiple Best Reciprocal Hit approach implemented in the newly designed Forty-Two software (https://bitbucket.org/dbaurain/42/downloads). First, we removed the most divergent sequences, which are the most likely to be paralogs or contaminants. More precisely for each species having multiple sequences, each sequence was BLASTed against the rest of the alignment and the best hit identified; a sequence was removed if it overlapped with the best hit sequence by $\geq$ 95% and if its BLAST score was below the best hit score by a given threshold. Using a threshold of 10%, 17,480 sequences were removed. The resulting clusters were cleaned using HmmCleaner version 1.8 [38] and the same process was repeated, this time removing 4,267 additional sequences. Most of these sequences were sequencing variants of the same transcripts (due to sequencing errors or to *in vivo* transcript degradation).

Removing potential contaminants

As in [9], alignments of ribosomal proteins containing a large eukaryotic taxonomic diversity were used to detect contaminations. We used BLASTP against several custom databases to detect and remove the contaminants. An additional screening was done using BLASTN to remove the few remaining contaminants from *Homo sapiens* and *Danio rerio*. The case of homoscleromorph and calcareous sponges was analyzed differently, because of the absence of clean complete genomes that can serve as a reference for decontamination. For each alignment, we BLASTed each poriferan sequence against the other sequences and removed the 2,434 sequences that had a BLAST bit score to the 'wrong' clade that was 5% higher than to the expected clade (i.e., Calcarea, Demospongiae, or Homoscleromorpha).

To discard genes for which orthology/paralogy relationships are difficult to infer, we made alignments using Mafft [39] (mafft–quiet–localpair–maxiterate 5000 —reorder), cleaned alignments with HmmCleaner and constructed RAxML trees [41] using the LG+Gamma+F model. We then computed the number of taxonomic groups (among the 14 clades displaying a long basal branch: Acoela, Anthozoa, Calcarea, Chordata, Demospongiae, Ecdysozoa, Echinodermata, Hemichordata, Homoscleromorpha, Lophotrochozoa, Medusozoa, Nemertodermatida, Rotifera and Xenoturbellida) displaying paralogous copies (see [9]) and eliminated the 157 genes with > = 5 cases of paralogy.

To reduce the amount of missing data and the computational burden, we removed 21 species (highly incomplete, taxonomically redundant or fast-evolving) and then the 137 genes in which more than one of the following 8 groups (Acoela, Nemertodermatida, Xenoturbellida, Echinodermata, Hemichordata, Chordata, Protostomia and outgroup) is missing. We had three criteria for choosing which taxa to retain: 1. Taxonomic diversity with the aim of picking a member of each of the major groups of a given clade (i.e., not all arthropods for Ecdysozoa). 2. Avoiding taxa with known issues such as extreme branch lengths or compositional biases (e.g., picking a shorter branch nematode rather than the familiar but rapidly evolving *Caenorhabditis elegans*). 3. choosing a species with fewest missing data.

Our last quality check was based on the rationale that non-orthologous sequences (being either a contaminant or a paralog and thus misplaced) typically display very long branches when constrained on the species tree. First, alignments were cleaned with HmmCleaner version 1.8 [38] and BMGE [42], and concatenated using SCaFoS [43]. The phylogeny inferred using RAxML [41] from the supermatrix under the LG+Gamma$_4$+F model was considered as a proxy of the species tree (note that xenacoelomorphs were sister to all other bilaterians in this tree). Then, for each alignment, the reference topology was pruned of the species missing in that alignment, and branch lengths on this constrained topology were estimated using RAxML (LG+Gamma$_4$+F model). This allowed us to compare terminal branch lengths observed in the single-gene tree to those observed in the pruned supermatrix tree, and to remove sequences for which the branch-length ratio was > 5, hence eliminating 642 questionable sequences.

Finally, we only kept the 1173 alignments in which at most 16 species were missing. We used SCaFoS to assemble the supermatrix, build chimeras of closely-related species (*Oscarella carmela*/*Oscarella* SN2011, *Saccoglossus kowalevskii*/*Saccoglossus mereschkowskii* and *Cephalodiscus gracilis*/*Cephalodiscus hodgsoni*) and retained only the slowest-evolving sequence when multiple copies were available for a given species (using Tree-Puzzle and the WAG+F model to compute distances). This produced a supermatrix containing 350,088 amino acid positions for 59 species, with an overall amount of 23.5% missing data.

Dataset quality

To compare of our dataset with those of Cannon and Rouse [1, 2], for each gene separately we computed a phylogeny using RAxML (LG+Gamma$_4$+F model) [41]. We then computed the number of tree bipartitions observed in the supermatrix tree (constructed with the same model) that are recovered by each gene. We assume that the majority of partitions in the supermatrix tree are likely to be correct and the percent of recovered bipartitions in the single gene trees is thus an estimation of dataset quality. Dataset quantity was measured as total amino acids.

Phylogenetic inference

The supermatrix was analyzed with the site-heterogeneous CATGTR model [44] using PhyloBayes-MPI version 1.8 [31] after the removal of constant positions ('-dc' option) and with the site-homogeneous GTR model using raxml version 8.2.8 [41]. The use of LG or LG4X models gave virtually the same results as GTR. The robustness of phylogeny was inferred with 100 rapid bootstraps in the case of the GTR model and with 100 gene jackknifes in the case of the CATGTR model.

Stratifying genes according to support for known monophyletic groups

To select the genes from all three datasets (this study, Rouse et al. [1] and the larger 881 genes dataset of Cannon et al. [2]) most likely to contain easy to extract phylogenetic signal, we used two different approaches. First, each gene was analyzed separately to find their individual level of support for known monophyletic groups. All Xenacoelomorph sequences were removed such that the monophyly measure was independent of the presence of this clade. For each aligned and trimmed gene, a tree was reconstructed using phyml [45] (settings -d aa -o tlr -a e -c 5). Each resulting tree was analyzed using a custom perl script that measured the support for the following uncontroversial monophyletic groups: Cnidaria, Ambulacraria, Hemichordata, Echinodermata, Chordata, Ecdysozoa, Lophotrochozoa, Porifera, Ctenophora (where present) Protostomia and Bilateria The monophyly score for each clade was calculated as the size of the largest clade on the tree containing species from the monophyletic group in question divided by the total number of species from that monophyletic group in the dataset. For example, if there were five chordates in the dataset and the largest chordate-only grouping on the tree contained four of them, the monophyly score for chordates would be $^4/_5$ = 0.8. The total score for the tree was calculated as the monophyly score averaged over all clades. Clades with fewer than two species in the tree were ignored. The datasets were then ranked by monophyly score and concatenated (with Xenacoelomorphs now included) in order from best (highest monophyly score) to worst.

For each of the three stratified datasets (ours, Cannon et al. [2] and Rouse et al. [1]) we took the genes representing the first 25% of positions (best) and the last 25% positions (worst)

and performed jackknife resampling to produce 50 jackknife replicates each containing ~30,000 positions. Each jackknife replicate dataset was analyzed using PhyloBayes-MPI and a CATGTR+Gamma model with a single run and stopping after 1500 cycles. The jackknife summary tree was produced using a bpcomp analysis using all 50 replicates with a burnin discarding the first 1000 cycles. We also inferred Maximum LIkelihood trees using the GTR+Gamma model with RAxML [41] based on the concatenations of the best and worst 25% of genes.

In a second closely related approach, we sorted the genes according to the percentage of bipartitions observed in the supermatrix tree that are recovered by each gene and took the 25% genes with the highest (lowest) values as the best (worst) genes this time including all species. These approaches gave congruent results and we present only those from the first approach.

Dayhoff recoding

This was performed using the "-recode Dayhoff6" command in PhyloBayes-MPI.

Posterior Predictive Analyses (ppred)

These were conducted using PhyloBayes ppred command as described in [19].

Carbon footprint calculations

The carbon footprint for travel was computed only for flights for the three meetings specifically organized for this project, so constitute a small underestimate. We used the calculator of the International Civil Aviation Organization (https://www.icao.int/environmental-protection/Carbonoffset/Pages/default.aspx), which did not include radiative forcing, so seriously underestimating the impact on global warming (Table S2).

The carbon footprint for computation was more difficult to compute since analyses were done in multiple labs, using various computers. More importantly, we did not archive all computations done for this work (e.g., preliminary analyses). We used the reasonable hypothesis that the jackknife analyses with the CATGTR model are by far the largest contributor and compute their footprint only. This certainly leads to an underestimation (ignoring for example assembly of genomes/transcriptomes, dataset building, dataset curation, RAxML analyses and Dayhoff analyses were ignored). For simplicity we also assumed that all the computations were done on a single computer, mp2 of ComputeCanada (https://wiki.calculquebec.ca/w/Accueil).

For 3 taxon sampling experiments, the 100 jackknife replicates of ~90,000 positions were performed on 6 nodes of 24 cores. The average CPU time for a single replicate was 520.5 hours, giving a total of 936,900 hours (= 520.5*6*100*3). The 50 jackknife replicates of ~30,000 positions were performed on 2 nodes of 24 cores, for 3 datasets (Our data, Cannon and Rouse), 2 taxon samples, 2 data samples (best/worse) and 2 methods. The average time for a single replicate is 188.8 hours, so a total of 453,120 hours of a single node (= 188.8*2*50*3*2*2*2*2). Total time for all jackknife experiments assuming a single node is 1,390,020 hours.

A node of mp2 consumes 300 W, to which we add cooling (22,75%) and other components (~5%) (Suzanne Talon, personal communication), so one hour of computation corresponds to ~0.38 kWh (= 0.3*1.2775). Total electric energy consumption for our CATGTR jackknife replicates was 531,683 kWh (= 1,390,020*0.38). To convert this into CO_2 emissions, we used the world average carbon intensity of power generation in 2017 (https://www.iea.org/tcep/power/), 491 gCO2/kWh, which leads to an estimate of 261 tonnes of CO_2 (= 531,683*0.000491).

QUANTIFICATION AND STATISTICAL ANALYSIS

Jackknife procedure and tests for reliability

A jackknife replicate was generated by randomly sampling single-gene alignments without replacement until > 90,000 positions (~390 genes per replicate for most) or > 30,000 positions (~130 genes per replicate for the analyses of best and worst genes) depending on analysis were selected. For PhyloBayes-MPI analysis of jackknife replicates, 3000 cycles were performed and consensus tree and jackknife support were obtained as in [9].

To see whether the number of cycles gives an accurate measure, we experimented by extending our chains. Increasing the number of cycles did not alter jackknife proportions (Table S1.).

Similarly, running two chains of each jackknife replicate until convergence also strengthens our results. We performed an experiment where we ran two chains for each of 100 jackknife samples of 30k positions for the 'best' quarter of positions of our data with all taxa. Of these, 51 pairs of chains converged (maxdiff < 0.3) and 49 pairs did not (maxdiff > 0.3) - we compared the results from converged and imperfectly converged sets (Table S1.).

50 of 59 nodes received 100% support (Jackknife Proportion JP = 100%) in both converged and non-converged datasets and all but 4 received > 90% support in both converged and non-converged pairs of chains. For all nodes that did not receive maximum support, the level of support is very similar for the converged and the imperfectly converged set. Interestingly, for 7 out of 9 nodes, the level of support in the converged set of runs was higher. Xenambulacraria support increased from 0.91 to 0.96. Chordata + Protostomia from 0.45 to 0.58. Only support for monophyly of Acoelomorpha and sister-group of *Ircinia* and *Chondrilla* was lower in the converged data (0.5 and 0.98) than in non-converged (0.65 and 1).

We also compared the results from Jackknifing to those from Bootstrapping (which uses full sized datasets as opposed to jackknifing which uses a smaller subsample). Bootstrapping can be applied in some of the less CPU intensive analyses (reduced alphabet analyses which are significantly quicker). When we do this (100 replicates) for our full dataset with all species, the supports were very similar to those of the jackknife based on 90K positions, and, as expected, slightly higher (see below). Interestingly, the support value for monophyletic Xenambulacraria increases from 90% jackknife to 98% bootstrap support (Table S1.). This supports our contention that jackknifing provides a conservative estimate of support.

Due to the relatively small size of the main Cannon et al. [2] dataset (~45k positions) we managed to run a full PhyloBayes analysis to convergence on a complete dataset. We used the CATGTR site heterogeneous model on a dataset from which the long branched Acoelomorpha had been removed. We found *Xenoturbella* + Ambulacraria supported with a value of 1.0 posterior probability showing that our jackknife analysis of the same was conservative (Figure S3B).

Model fit

To assess the fit of different models, we performed 10-fold model cross-validations. Model fit tests were done using training datasets of 10,000 amino acids and test datasets of 2,000 amino acids we used PhyloBayes version 4.1 [12] to perform cross-validation for the following models: LG+Γ, GTR+Γ, CAT+Γ and CAT-GTR+Γ. PhyloBayes was run for 1100 (LG and GTR) or 3100 (CAT and CATGTR) cycles and we kept the last 1000 cycles for following likelihood computations. Cross validation was run for full datasets as well as for the best and worst genes from the gene stratification experiments. The model cross-validations in all cases clearly favored CAT-GTR+Γ > CAT+Γ > GTR+Γ > LG+Γ (for our principal, complete dataset likelihood scores with respect to LG are 3034 ± 152, 2270 ± 151 and 268 ± 40).

DATA AND SOFTWARE AVAILABILITY

The sequence alignments, phylogenetic trees that support the findings of this study, as well as the script for measuring monophyletic groups, are available on GitHub (https://github.com/MaxTelford/Xenacoelomorpha2019). The accession number for the genome and transcriptome assemblies reported in this paper is are available at https://figshare.com/search Figshare: PRJNA517079 and raw data for novel sequences are available at the Sequence Read Archive, SRA: PRJNA517079.