A new spiralian phylogeny places the enigmatic arrow worms among Gnathiferans

Ferdinand Marletaz, Katja T.C.A. Peijnenburg, Taichiro Goto, Noriyuki Satoh, and Daniel S. Rokhsar

CellPress

Current Biology
Report

A New Spiralian Phylogeny Places the Enigmatic Arrow Worms among Gnathiferans

Ferdinand Marlétaz,[1,2,8,*] Katja T.C.A. Peijnenburg,[3,4] Taichiro Goto,[5] Noriyuki Satoh,[6] and Daniel S. Rokhsar[1,7]
[1]Molecular Genetics Unit, Okinawa Institute of Science & Technology, 1919-1 Tancha, Onna 904-0495, Japan
[2]Department of Zoology, University of Oxford, Mansfield Road, Oxford OX1 3SZ, UK
[3]Marine Biodiversity, Naturalis Biodiversity Center, Vondellaan, Leiden 2300 RA, the Netherlands
[4]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Sciencepark, Amsterdam 1090 GE, the Netherlands
[5]Department of Biology, Faculty of Education, Mie University, Tsu, Mie 514-8507, Japan
[6]Marine Genomics Unit, Okinawa Institute of Science & Technology, 1919-1 Tancha, Onna 904-0495, Japan
[7]Department of Molecular and Cell Biology, University of California, Life Sciences Addition, Berkeley, CA 94720, USA
[8]Lead Contact
*Correspondence: ferdinand.marletaz@gmail.com
https://doi.org/10.1016/j.cub.2018.11.042

SUMMARY

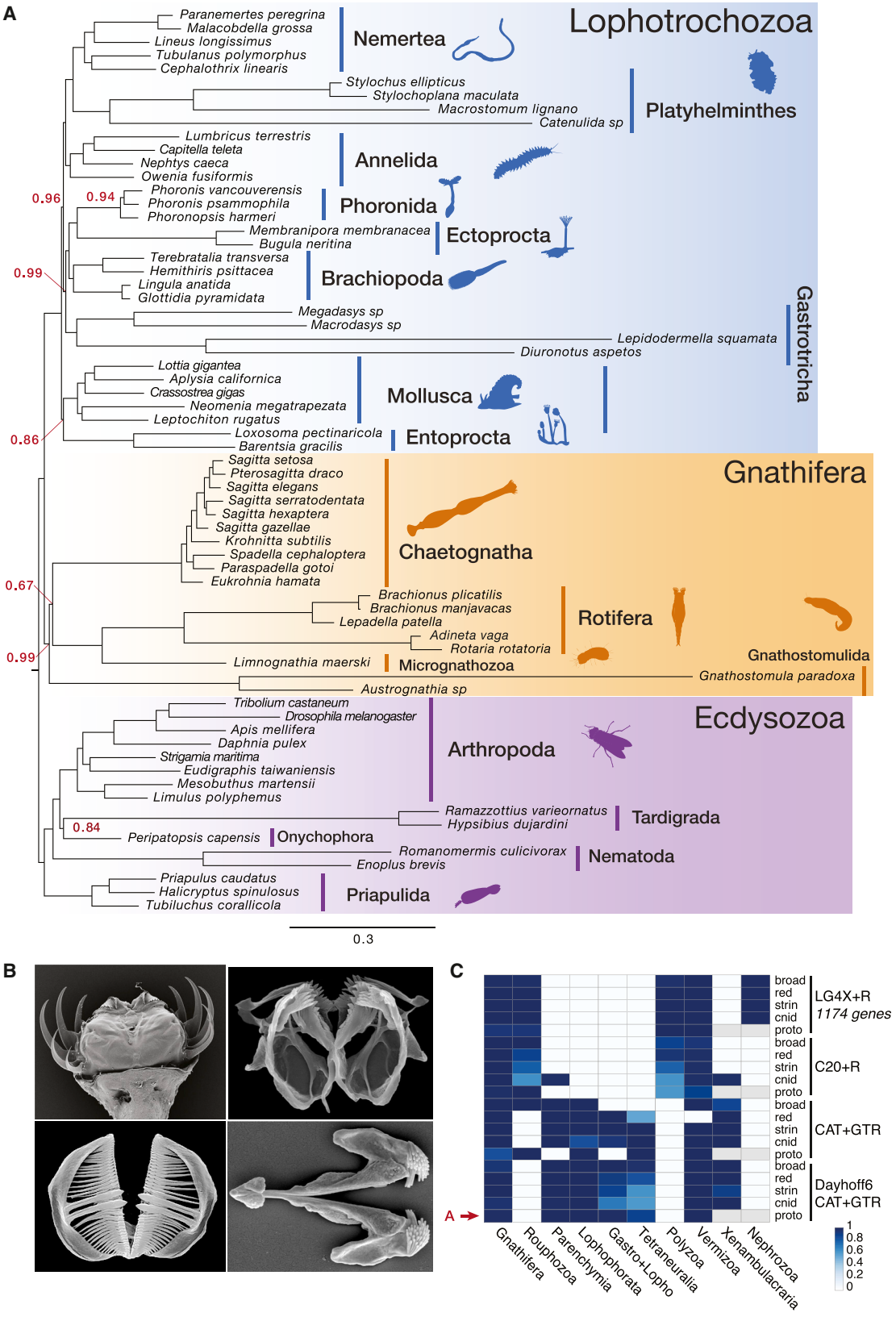
Chaetognaths (arrow worms) are an enigmatic group of marine animals whose phylogenetic position remains elusive, in part because they display a mix of developmental and morphological characters associated with other groups [1, 2]. In particular, it remains unclear whether they are a sister group to protostomes [1, 2], one of the principal animal superclades, or whether they bear a closer relationship with some spiralian phyla [3, 4]. Addressing the phylogenetic position of chaetognaths and refining our understanding of relationships among spiralians are essential to fully comprehend character changes during bilaterian evolution [5]. To tackle these questions, we generated new transcriptomes for ten chaetognath species, compiling an extensive phylogenomic dataset that maximizes data occupancy and taxonomic representation. We employed inference methods that consider rate and compositional heterogeneity across taxa to avoid limitations of earlier analyses [6]. In this way, we greatly improved the resolution of the protostome tree of life. We find that chaetognaths cluster together with rotifers, gnathostomulids, and micrognathozoans within an expanded Gnathifera clade and that this clade is the sister group to other spiralians [7, 8]. Our analysis shows that several previously proposed groupings are likely due to systematic error, and we propose a revised organization of Lophotrochozoa with three main clades: Tetraneuralia (mollusks and entoprocts), Lophophorata (brachiopods, phoronids, and ectoprocts), and a third unnamed clade gathering annelids, nemerteans, and platyhelminthes. Consideration of classical morphological, developmental, and genomic characters in light of this topology indicates secondary loss as a fundamental trend in spiralian evolution.

RESULTS AND DISCUSSION

A Phylogenomic Resolution of the Bilaterian Tree of Life

The neat tripartite classification of bilaterian animals into deuterostomes, ecdysozoans, and lophotrochozoans imparts considerable uncertainty, especially involving lineages comprising the lophotrochozoan superclade and the relationships among them [9]. To accurately reconstruct bilaterian relationships, we extracted 1,174 single-copy orthologs present in available metazoan complete genomes and transcriptomic datasets, combining existing datasets with newly generated sequences (STAR Methods). We aimed for a balanced taxonomic sampling by retaining four to six species for each phylum to maximize taxonomic representation. To achieve that, we gave priority to species with the highest ortholog recovery and deliberately focused on the slowest-evolving lineages to minimize the impact of long branch attraction (LBA). We evaluated the impact of taxonomic sampling by analyzing five distinct supermatrices that included varying numbers of fast-evolving species and species with marked compositional biases, as well as using alternative rooting taxa (Figure 1C and STAR Methods). We clipped the data matrices for potential contamination and stretches of mistranslated amino acids that can occur in transcriptome data. We also employed several tree inference methods and models known to show distinct robustness toward model violation and reconstruction artifacts. In particular, we used both site-homogeneous models assuming gene partitions (LG4X+R) and site-heterogeneous models (CAT+GTR) for which we restricted the analyses to a set of markers selected for their lower mutational saturation [11, 12]. We also examined how efficiently these different models account for compositional heterogeneity across taxa using posterior predictive analyses (PPAs), which revealed that compositional deviation is affecting model fit (Figures 2B, 2C, and S3) [6, 11]. To reduce the impact of composition heterogeneity, we also performed tree reconstruction using site-heterogeneous models after recoding our datasets in the six broader Dayhoff amino acid functional categories (Dayhoff6) [13].

Using these alternative methods and datasets, we recover the major well-accepted bilaterian clades, obtaining strong support for protostome, ecdysozoan, and a clade of non-ecdysozoan



(legend on next page)

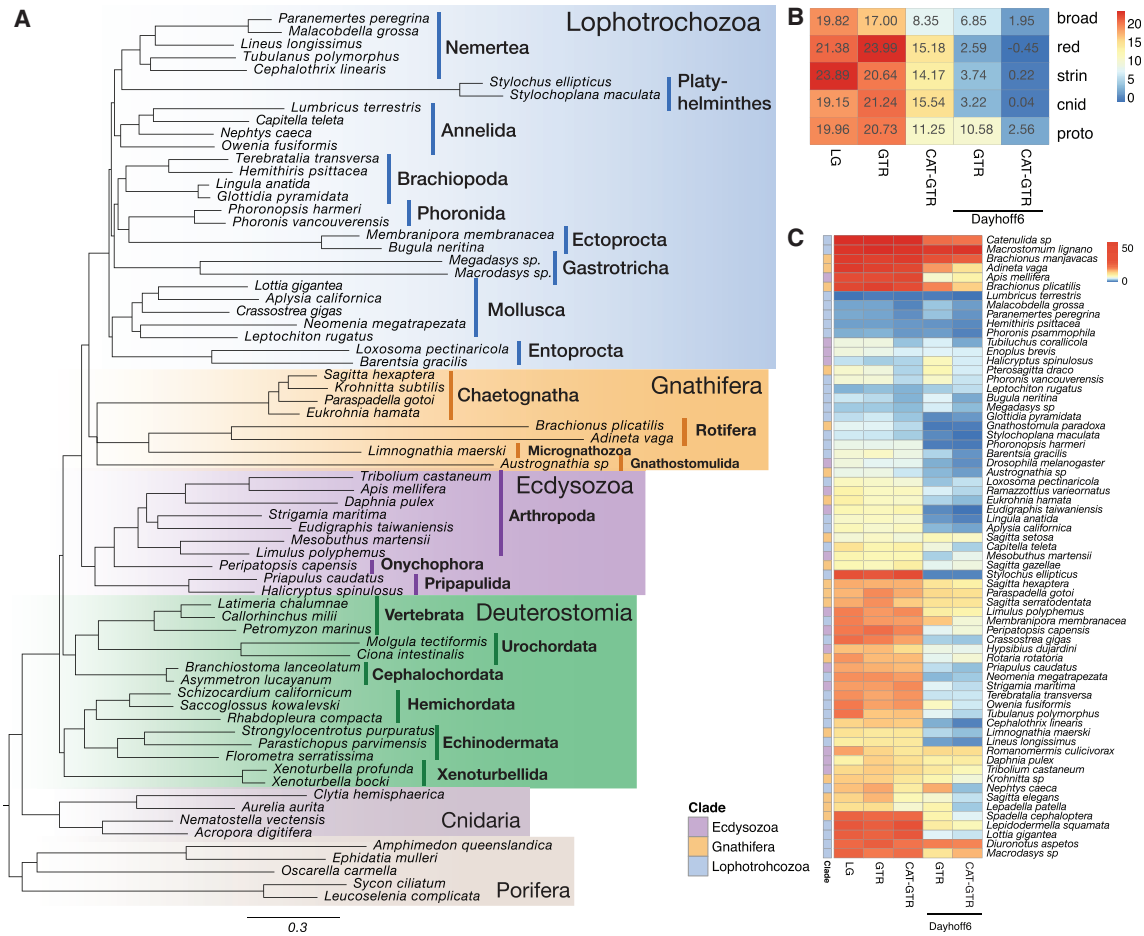


Figure 2. Impact of Compositional Heterogeneity across Taxa on Bilaterian Phylogenetic Reconstruction

(A) Bilaterian phylogeny reconstruction with Phylobayes using the CAT+GTR model using only taxa with slowest evolutionary rates and steady deviating amino acid composition. All posterior probabilities are maximal.

(B and C) *Z* score statistics of posterior predictive analyses (PPAs) to assess compositional heterogeneity. Global scores for different datasets and models (B) and the detail of the *Z* score for each species in the "proto" datasets of Figure 1 (C) are shown. Species order was derived by k-means clustering based on PPA *Z* score.

The scale bar indicates the inferred substitution per site. See Figure S3 for PCA analysis and PPA results on all taxa.

protostomes that include spiralian and lophotrochozoan taxa, and we verify the monophyly of all represented animal phyla (Figures 1A and 2A) [9]. Unexpectedly, the monophyly of deuterostomes does not always receive maximal support, particularly in Dayhoff6 recoded datasets (CAT+GTR; Figures 1C and 2A), with the occasional earlier divergence of Ambulacraria (echinoderms and hemichordates) relative to Chordata, a topology that has been reported previously [14]. Although this question is not the primary focus of our analyses, we found a preferential association of the Xenacoelomorpha group (Xenoturbella and acoelomorph flatworms) with Ambulacraria [15] when more sophisticated molecular evolution models are used (CAT+GTR and Dayhoff6 recoding) and when the fastest-evolving acoel flatworm species are excluded. With simpler site-homogeneous

Figure 1. Chaetognaths are Members of the Gnathifera Clade Together with Rotifers, Micrognathozoans, and Gnathostomulids

(A) Protostome phylogeny reconstructed with Phylobayes using CAT+GTR and Dayhoff6 recoding scheme. All nodes but the ones labeled with red numbers show the maximal posterior probability (silhouettes are from Phylopic).

(B) Pictures of jaw apparatus in gnathiferans; clockwise: chaetognaths, rotifers, micrognathozoans, and gnathostomulids (adapted from [10]). Image credit: M.V. Sørensen (Natural History Museum of Denmark).

(C) Summary of support values obtained using distinct taxonomic sampling, reconstruction methods (site homogeneous and heterogenous), and recoding schemes (see STAR Methods and https://doi.org/10.5281/zenodo.1403005). For the detail of clade and topology names, see STAR Methods. "Gastro+Lopho" means Gastrotriches with Lophophorates.

The scale bar indicates the inferred substitution per site. See Supplemental Information for detailed trees and Figures S1 and S2 for selected maximum-likelihood trees.

models, however, Xenacoelomorpha remain branched as the sister group of bilaterians [16]. We note that acoels show a diverging amino acid composition, which could have impacted earlier studies (Figure S3). In sum, our analyses not only support the "new view" of animal phylogeny but also illustrate the impact of the reconstruction method on the obtained trees.

Chaetognaths Join an Extended Gnathifera Clade

Chaetognaths are a major zooplankton group that has long been a challenge for both morphology- and molecular-based phylogenetics. They display a mosaic of morphological and developmental characters, presenting a secondary blastopore opening reminiscent of deuterostomes, while possessing two ventral nerve cords and a circum-esophageal brain classically associated with protostomes [1, 3]. Interestingly, Cambrian fossil deposits such as the Burgess Shale contain chaetognath representatives that are remarkably similar in body organization to present-day forms [17]. At the molecular level, early evidence from ribosomal RNA rejected deuterostome affinities of chaetognaths [18], and later attempts using multigene datasets indicated a more likely association with protostomes [1–3]. However, these studies did not agree on chaetognath branching within protostomes, suggesting that they could either be the sister group of other protostomes [1, 2] or represent an early lineage within spiralians [3, 4]. Another recent study pointed toward a possible association of chaetognaths with some gnathiferan taxa, but with limited support and discrepancies between analyses [19].

Strikingly, we find that chaetognaths are united with rotifers, gnathostomulids, and the recently described micrognathozoans in a well-supported clade using multiple tree-inference strategies and taxonomic sampling (Figure 1A and 2A). In particular, our results were unchanged when we excluded the fastest-evolving species of gnathostomulid (*Gnathostomula* sp.) and some rotifer species (Figures 1A, 1C, 2A, S1, and S2). Since rotifers present a derived amino acid content compared to that of other bilaterian taxa (Figure 2B), we generated a set of trees after Dayhoff6 recoding, which alleviates compositional heterogeneity across taxa [6, 13], and we still obtained strong support for a Gnathifera clade including chaetognaths (Figure 1A).

The Gnathifera clade was originally proposed based on pharyngeal hard parts and protonephridial structures [20] found to be shared by Rotifera, Gnathostomulida [7], and later Micrognathozoa as well [21]. Our phylogenetic analyses expand this clade to include chaetognaths, which is corroborated by additional morphological and molecular characters. First, the presence of a complex jaw apparatus with hardened parts in chaetognaths is compatible with the primary morphological character defining this group (Figure 1B) [7]. The association of chaetognaths with Gnathifera was originally suggested on this basis in the second edition of Nielsen (2001) [20], but not in the subsequent third edition published 10 years later. The chaetognath grasping spines, the mastax of rotifers, and the jaws of gnathostomulids possibly share a composite organization at the ultrastructural level with alternating layers of material that is opaque or dense to electrons disposed in tubular fashion [22, 23]. Such an organization is absent from other spiralians (e.g., annelids and mollusks) [23]. Second, several authors have suggested that chaetognaths might share an unusual intracellular mode of cuticle formation with rotifers and acanthocephalans, although this was never confirmed [24]. These observations require further investigation as chaetognaths are notable for their multilayered epidermis [25]. Finally, the recent discovery in the rotifer *Brachionus plicatilis* of a plausible ortholog of a MedPost Hox gene would constitute a remarkable molecular synapomorphy for an extended Gnathifera clade including chaetognaths [26]. This class of Hox genes shows intermediate residues between median and posterior Hox and was originally considered to be specific to chaetognaths [27].

The relationships within the newly extended Gnathifera clade are more elusive. Micrognathozoans robustly group with rotifers, corroborating the morphology and previous molecular studies [8, 21]. Gnathostomulids sometimes branch as the sister group to other gnathiferans (Figure 1A), while in other trees they are closer to Rotifera and Micrognathozoa (Figure S1) or even chaetognaths (Figure S2). Our broad chaetognath sampling also provides some insights into the intraphyletic relationships of chaetognaths. Eukrohniidae are positioned as the sister group to other chaetognath species, which supports the paraphyly of Phragmophora. In contrast, Aphragmophora is monophyletic with an early divergence of Krohnittidae and with a paraphyletic *Sagitta* genus due to the nested position of the monospecific genus *Pterosagitta* [28, 29]. Although sampling of additional taxa would be required to examine the details of chaetognath taxonomy, the recovered relationships are in broad agreement with previous schemes [28, 29].

Revised Spiralian Relationships

Our phylogenetic reconstructions also suggest a new scheme for the relationships among non-ecdysozoan protostome taxa, which is a notoriously difficult problem [9]. Many early phylogenomic studies recovered a Platyzoa clade that collected many morphologically simple, fast-evolving lineages, including platyhelminthes, rotifers, gastrotrichs, gnathostomulids, and others [4]. However, recent studies instead found this assembly to be paraphyletic, leading to the proposal of a Rouphozoa clade that unites platyhelminthes and gastrotrichs, sometimes closely related to a clade consisting of entoprocts and ectoprocts (Polyzoa) [5, 8, 19]. This latter topology receives support in some of our analyses, particularly the ones relying on site-homogeneous and empirical mixture (C20) models (Figures S1 and S2), but an improved taxonomic sampling, the usage of a site-heterogenous model, and the reduction of missing data rejects the Rouphozoa (Figures 1A and 2A). More sophisticated infinite mixture models (CAT-GTR), however, suggest a novel view of spiralian relationships. We find three distinct spiralian subclades (Figures 1A and 2A): (1) an entoproct group with mollusks to recover the Tetraneuralia clade, previously proposed based on muscle system and larval characteristics [30]; (2) a monophyletic Lophophorata that includes brachiopods, phoronids, and ectoprocts (but not Entoprocta), possibly associated with gastrotrichs, and (3) a new unnamed clade that gathers nemerteans and platyhelminthes with annelids. Platyhelminthes are very quickly evolving and show marked compositional deviation compared to that of other bilaterians (Figures 2B and S3). Hence, we particularly scrutinized both the recoded dataset (Figures 1A and 1C) and analyses restricted to the slowest-evolving available Platyhelminthes (Figure 2C and "strin" dataset), which both confirmed

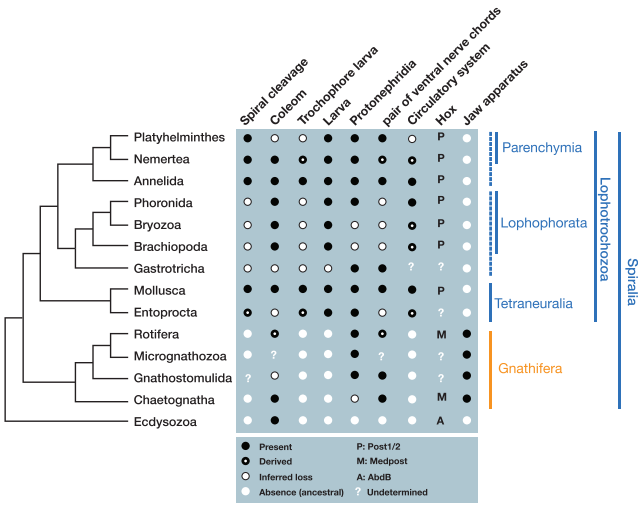


Figure 3. Distribution of Morpho-Developmental Characters in Spiralians

Morphological characters of interest are mapped along the proposed topology. In the legend, "derived" means that the character homology has been a subject of debate in the literature, and "inferred loss" indicates a possible character loss following a parsimony reasoning, but other scenarios are not excluded. Accepted clade names are indicated on the right along with novel unnamed spiralian clades (indicated by the dashed line).

this new and as-yet-unnamed clade. The association of platyhelminthes and nemerteans supports a century-old view mostly based on the abundant parenchyma between body wall and internal organs (Parenchymia). Similarities of ciliary band organization in the Götte's larva of platyhelminthes and in the *pilidium* larva of nermerteans, as well as the shared absence of chitin in both groups, have been noted as possible synapomorphies for Parenchymia. However, the Parenchymia clade has not been supported by recent molecular phylogenies [20, 31]. Similarly, the association of annelids and nemerteans has previously been argued based on similarities in their circulatory systems under the name Vermizoa [32].

A Reappraisal of Spiralian Character Orientation

Our new spiralian phylogeny argues for a reappraisal of the evolution of a number of clade-defining traits [20, 33]. The respective branching of annelids, mollusks, and platyhelminthes in our trees indicates a common origin for spiral cleavage and trochophore larvae, followed by subsequent loss in lophophorates (Figure 3) [34]. The observed position of nemerteans also corroborates observations in paleonemerteans, which suggests that a trochophore-like larva is likely the ancestral condition in nemerteans [35]. In our extended Gnathifera, spiral cleavage and larval stage are absent in chaetognaths and rotifers, which are both direct developers, while gnathostomulids have only been briefly mentioned as undergoing possible spiral cleavage in a 1969 publication [36]. Until future investigation clarifies the ancestral cleavage type in Gnathifera, for instance by applying lineage tracing or 4D microscopy techniques, it seems reasonable to continue using the name Spiralia to refer to the clade formed by all non-ecdysozoan protostomes, which itself is subdivided into Gnathifera and Lophotrochozoa. Under this scheme, platyhelminthes and gastrotrichs would be considered members of the Lophotrochozoa clade, although they were not originally included. If future characterization of embryonic development in gnathiferans—especially gnathostomulids—were to reject the hypothesis that spiral cleavage is ancestral to these taxa, then the name Spiralia, as a synapomorphy-based name, would become a synonym of the clade referred to here as Lophotrochozoa. In that case, a new name would be needed for the non-ecdysozoan protostome clade comprised of Gnathifera and the restricted Spiralia. Here, we propose that such a clade could be named Gnathospiralia following the same associative reasoning as was used for naming the Lophotrochozoa [34, 36].

The distribution of morphological characters shows a patchy distribution across Spiralia, which is consistent either with repeated character losses from a complex ancestor or repeated character acquisition [5, 37] (Figure 3). This remark also applies to genome evolution with traits such as intron positions, gene families, and genome architecture that are shared among bilaterians but are experiencing dramatic loss in some lineages [38] (Figure 3). Earlier claims that the relative simplicity encountered in the previously proposed Platyzoa assemblage could be indicative of a simple acoelomate bilaterian ancestor are therefore questioned [5] (Figure 3). Finally, some clades proposed here, such as Annelida with Parenchymia or Gastrotricha with Lophotrochozoa, do not have acknowledged synapomorphies that could be used as a basis for a name, which pleads for further investigation.

Conclusion

Our phylogenetic analyses support the inclusion of chaetognaths in the Gnathifera clade, which strengthens the importance of this subdivision of the animal tree of life. Our analyses reject the previously proposed Platyzoa assemblage that gathered fast-evolving lineages. Some of its members (e.g., rotifers and gnathostomulids) are incorporated in Gnathifera; others (e.g., Platyhelminthes) are now members of the Lophotrochozoa clade. The inclusion of chaetognaths, a coelomate phylum with a complex nervous system, among Gnathifera implies that subsequent character loss took place in the other members of this group or that chaetognaths independently evolved a number of traits, such as a condensed nervous system or deuterostomic development [37, 39]. Further study of genomes and development in gnathiferans is essential to better understand the ancestral condition in this clade, in protostomes, and in bilaterians.

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
  - Sample collection
- METHOD DETAILS
  - Sample preparation for sequencing
  - Transcriptome assembly and filtering
  - Gene family reconstruction and homology search

- Supermatrix assembly, marker selection and taxonomic sampling
- Phylogenetic analyses

- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.11.042.

ACKNOWLEDGMENTS

This study was supported by internal funding from the Okinawa Institute of Science and Technology Graduate University to the Molecular Genetics and Marine Genomics Units, by the European Research Council (ERC; grant 268513 to Peter W.H. Holland), and by Netherlands Organization for Scientific Research (NWO) Veni grant 863.08.024 and Vidi grant 016.161.351 to K.T.C.A.P. Fieldwork was also supported under NSF grants OCE-1029478 and OCE-1338959 (to Erica Goetze). The Atlantic Meridional Transect program is supported by the UK NERC National Capability funding. We would like to thank Peter W.H. Holland for his support in the early stages of this study; Jordi Paps and Daria Gavriouchkina for their comments; Nicolas Lartillot and Roberto Feuda for their advice on phylogenetic analyses; Hervé Philippe for providing a development version of the Hmmclean tool; Maximilian Telford for his advice on Illumina cross-contamination filtering; Christine Hörnlein, Lene Friis Møller, Alice Burridge, and Erica Goetze for help with collecting zooplankton samples; and Emma Otto and Annelies Pierrot-Bults for their assistance in the species identification of chaetognath specimens. We thank Martin Vinther Sørensen for letting us use the SEM pictures in Figure 1B.

AUTHOR CONTRIBUTIONS

F.M., K.T.C.A.P., N.S., and D.S.R. designed the study. K.T.C.A.P., F.M., and T.G. collected and identified animal samples. F.M. and N.S. prepared and sequenced samples. F.M. performed phylogenetic analyses and wrote the manuscript. F.M., D.S.R., and K.T.C.A.P revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 31, 2018
Revised: September 12, 2018
Accepted: November 14, 2018
Published: January 10, 2019

REFERENCES

1. Marlétaz, F., Martin, E., Perez, Y., Papillon, D., Caubit, X., Lowe, C.J., Freeman, B., Fasano, L., Dossat, C., Wincker, P., et al. (2006). Chaetognath phylogenomics: a protostome with deuterostome-like development. Curr. Biol. *16*, R577–R578.
2. Marlétaz, F., Gilles, A., Caubit, X., Perez, Y., Dossat, C., Samain, S., Gyapay, G., Wincker, P., and Le Parco, Y. (2008). Chaetognath transcriptome reveals ancestral and unique features among bilaterians. Genome Biol. *9*, R94.
3. Matus, D.Q., Copley, R.R., Dunn, C.W., Hejnol, A., Eccleston, H., Halanych, K.M., Martindale, M.Q., and Telford, M.J. (2006). Broad taxon and gene sampling indicate that chaetognaths are protostomes. Curr. Biol. *16*, R575–R576.
4. Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., et al. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. Nature *452*, 745–749.
5. Struck, T.H., Wey-Fabrizius, A.R., Golombek, A., Hering, L., Weigert, A., Bleidorn, C., Klebow, S., Iakovenko, N., Hausdorf, B., Petersen, M., et al. (2014). Platyzoan paraphyly based on phylogenomic data supports a noncoelomate ancestry of spiralia. Mol. Biol. Evol. *31*, 1833–1849.
6. 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.
7. Ahlrichs, W.H. (1997). Epidermal ultrastructure of Seison nebaliae and Seison annulatus, and a comparison of epidermal structures within the Gnathifera. Zoomorphology *117*, 41–48.
8. Laumer, C.E., Bekkouche, N., Kerbl, A., Goetz, F., Neves, R.C., Sørensen, M.V., Kristensen, R.M., Hejnol, A., Dunn, C.W., Giribet, G., and Worsaae, K. (2015). Spiralian phylogeny informs the evolution of microscopic lineages. Curr. Biol. *25*, 2000–2006.
9. Telford, M.J., Budd, G.E., and Philippe, H. (2015). Phylogenomic insights into animal volution. Curr. Biol. *25*, R876–R887.
10. Gross, L. (2007). Who needs sex (or males) anyway? PLoS Biol. *5*, e99.
11. Rodrigue, N., and Lartillot, N. (2014). Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. Bioinformatics *30*, 1020–1021.
12. Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? Trends Genet. *22*, 225–231.
13. Susko, E., and Roger, A.J. (2007). On reduced amino acid alphabets for phylogenetic inference. Mol. Biol. Evol. *24*, 2139–2150.
14. Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., Hisata, K., Bredeson, J., Shoguchi, E., Gyoja, F., et al. (2015). Hemichordate genomes and deuterostome origins. Nature *527*, 459–465.
15. 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.
16. 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.
17. Briggs, D.E.G., and Caron, J.-B. (2017). A large cambrian chaetognath with supernumerary grasping spines. Curr. Biol. *27*, 2536–2543.e1.
18. Telford, M.J., and Holland, P.W. (1993). The phylogenetic affinities of the chaetognaths: a molecular analysis. Mol. Biol. Evol. *10*, 660–676.
19. Kocot, K.M., Struck, T.H., Merkel, J., Waits, D.S., Todt, C., Brannock, P.M., Weese, D.A., Cannon, J.T., Moroz, L.L., Lieb, B., and Halanych, K.M. (2017). Phylogenomics of Lophotrochozoa with consideration of systematic error. Syst. Biol. *66*, 256–282.
20. Nielsen, C. (2001). Animal Evolution: Interrelationships of the Living Phyla, Second Edition (Oxford University Press).
21. Sørensen, M.V. (2003). Further structures in the jaw apparatus of Limnognathia maerski (Micrognathozoa), with notes on the phylogeny of the Gnathifera. J. Morphol. *255*, 131–145.
22. Bone, Q., Ryan, K.P., and Pulsford, A.L. (1983). The structure and composition of the teeth and grasping spines of chaetognaths. J. Mar. Biol. Assoc. U. K. *63*, 929–939.
23. Rieger, R.M., and Tyler, S. (1995). Sister-group relationship of Gnathostomulida and Rotifera-Acanthocephala. Invertebr. Biol. *114*, 186–188.
24. Van Der Land, J., and Nørrevang, A. (1985). Affinities and intraphyletic relationships of the Priapulida. In the Origins and Relationships of Lower Invertebrates, S. Conway Morris, ed. (Clarendon Press), pp. 261–273.
25. Ahnelt, P. (1984). Chaetognatha. In Biology of the Integument: Invertebrates, J. Bereiter-Hahn, A.G. Matoltsy, and K.S. Richards, eds. (Springer Berlin Heidelberg), pp. 746–755.
26. Fröbius, A.C., and Funch, P. (2017). Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans. Nat. Commun. *8*, 9.

27. Papillon, D., Perez, Y., Fasano, L., Le Parco, Y., and Caubit, X. (2003). Hox gene survey in the chaetognath Spadella cephaloptera: evolutionary implications. Dev. Genes Evol. *213*, 142–148.
28. Gasmi, S., Nève, G., Pech, N., Tekaya, S., Gilles, A., and Perez, Y. (2014). Evolutionary history of Chaetognatha inferred from molecular and morphological data: a case study for body plan simplification. Front. Zool. *11*, 84.
29. Papillon, D., Perez, Y., Caubit, X., and Le Parco, Y. (2006). Systematics of Chaetognatha under the light of molecular data, using duplicated ribosomal 18S DNA sequences. Mol. Phylogenet. Evol. *38*, 621–634.
30. Wanninger, A. (2009). Shaping the things to come: ontogeny of lophotrochozoan neuromuscular systems and the tetraneuralia concept. Biol. Bull. *216*, 293–306.
31. Struck, T.H., and Fisse, F. (2008). Phylogenetic position of Nemertea derived from phylogenomic data. Mol. Biol. Evol. *25*, 728–736.
32. Cavalier-Smith, T. (1998). A revised six-kingdom system of life. Biol. Rev. Camb. Philos. Soc. *73*, 203–266.
33. Giribet, G. (2008). Assembling the lophotrochozoan (=spiralian) tree of life. Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 1513–1522.
34. Hejnol, A. (2010). A twist in time–the evolution of spiral cleavage in the light of animal phylogeny. Integr. Comp. Biol. *50*, 695–706.
35. Maslakova, S.A., Martindale, M.Q., and Norenburg, J.L. (2004). Vestigial prototroch in a basal nemertean, Carinoma tremaphoros (Nemertea; Palaeonemertea). Evol. Dev. *6*, 219–226.
36. Riedl, R.J. (1969). Gnathostomulida from America. Science *163*, 445–452.
37. Martín-Durán, J.M., Pang, K., Børve, A., Lê, H.S., Furu, A., Cannon, J.T., Jondelius, U., and Hejnol, A. (2018). Convergent evolution of bilaterian nerve cords. Nature *553*, 45–50.
38. Simakov, O., Marletaz, F., Cho, S.-J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.-H., Larsson, T., Lv, J., Arendt, D., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. Nature *493*, 526–531.
39. Martín-Durán, J.M., Janssen, R., Wennberg, S., Budd, G.E., and Hejnol, A. (2012). Deuterostomic development in the protostome Priapulus caudatus. Curr. Biol. *22*, 2161–2166.
40. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
41. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. *9*, R7.
42. Roth, A.C.J., Gonnet, G.H., and Dessimoz, C. (2008). Algorithm of OMA for large-scale orthology inference. BMC Bioinformatics *9*, 518.
43. Davidson, N.M., and Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol. *15*, 410.
44. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.
45. Liu, Y., Schmidt, B., and Maskell, D.L. (2010). MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. Bioinformatics *26*, 1958–1964.
46. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
47. 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.
48. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
49. Simion, P., Belkhir, K., François, C., Veyssier, J., Rink, J.C., Manuel, M., Philippe, H., and Telford, M.J. (2018). A software tool 'CroCo' detects pervasive cross-species contamination in next generation sequencing data. BMC Biol. *16*, 28.
50. 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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Alignments and phylogenetic analyses	Zenodo	https://zenodo.org/record/1403005
Transcriptome sequencing	SRA	SRA: SRR7754742-SRR7754750
Software and Algorithms		
Sickle (v1.33)		https://github.com/najoshi/sickle
Trinity (v2.3.2)	[40]	https://github.com/trinityrnaseq/trinityrnaseq/wiki
Trans-decoder (v5.0.2)	[41]	https://github.com/TransDecoder/TransDecoder
OMA (v1.0.5)	[42]	https://omabrowser.org/standalone/
Corset (v1.04)	[43]	https://github.com/Oshlack/Corset
Kallisto (v0.43.1)	[44]	https://github.com/pachterlab/kallisto
Msaprobs (v0.9.7)	[45]	http://msaprobs.sourceforge.net
Hmmer (v3.1b2)		http://www.hmmer.org
Hmmclean		https://metacpan.org/release/Bio-MUST-Apps-HmmCleaner
RAxML (v8.2.4)	[46]	https://github.com/stamatak/standard-RAxML
BMGE (v1.12)	[47]	ftp://ftp.pasteur.fr/pub/gensoft/projects/BMGE/
IQ-TREE (v1.6.2)	[48]	https://github.com/Cibiv/IQ-TREE
Phylobayes-mpi (v1.7)	[11]	https://github.com/bayesiancook/pbmpi
Analysis scripts	N/A	https://github.com/fmarletaz/phylogenomics/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and data should be directed to and will be fulfilled by the Lead Contact, Ferdinand Marlétaz (ferdinand.marletaz@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sample collection

Samples from multiple chaetognath species were collected during the Atlantic Meridional Transect cruises 22 and 24 in the Atlantic Ocean, in the Gullmarfjord (Sweden), in Amakusa (Japan) and in Marseille (France) (details about collection locations are given in Table S2). Chaetognaths were examined under a microscope while alive, identified, and preserved in RNAlater (Sigma or Invitrogen) and kept frozen. Reference samples were preserved in 4% formaldehyde solution to verify species identifications.

METHOD DETAILS

Sample preparation for sequencing

RNA was extracted using the RNAeasy micro or mini kits (QIAGEN) after homogenization with the Tissuelyser device (QIAGEN). RNA integrity was verified on an Experion instrument (Biorad), and RNA-seq polyA+ libraries were prepared using the TruSeq RNA kit (Illumina) at the WTCHG (Oxford) and Ovation RNA-seq system (NuGen) at SQC (OIST, Japan). Libraries were sequenced on Illumina instruments (HiSeq2000 and HiSeq4000). Detail and accession numbers for each sample are provided on Table S1.

Transcriptome assembly and filtering

We processed newly sequenced samples and datasets downloaded from the NCBI Short Read Archive (SRA) similarly. After trimming reads with Sickle, transcriptomes were assembled with Trinity (v2.3.2) using default parameters and a k-mer of 25 [40]. Open reading frame (ORF) for each transcript was determined using Trans-decoder [41] using a blast against a version of Swissprot restricted to metazoan taxa as clue (e-value 10-5). For newly sequenced samples, we sometimes noticed occurrence of cross-contamination during the sequencing process. To remove transcripts derived from mis-barcoded reads, we followed the approach described in [49]. To do this, we estimated the read counts of each transcript demultiplexed as one sample for each other sample

sequenced on the same lane and excluded the transcripts with (1) two times more coverage in another sample than the one they belonged to, and (2) a coverage less than 2x in their sample of origin. Python implementation of this approach is available together with other scripts used in this study at https://github.com/fmarletaz/phylogenomics/.

Gene family reconstruction and homology search

We used OMA (v1.0.5) to reconstruct a set of single-copy orthologous gene families [42] from proteomes derived from 20 metazoan genomes (Table S1) including a set of proteins derived from the *Paraspadella gotoi* genome (in preparation). To avoid performing clustering on transcript isoforms, transcripts were clustered using Corset [43] and we picked the most highly expressed transcript for each cluster, after measuring expression with Kallisto (v0.43.1) [43, 44]. We retained all families with 16 representatives or more, yielding a set of 1174 single-copy orthologs suitable to infer metazoan phylogeny. For each marker, we generated a protein alignment using Msaprobs (v0.9.7) [45] and built a hmm profile using hmmbuild of the hmmer package (v3.1b2). We then searched the collection of translated transcriptome using these hmm profiles using hmmsearch (e-value 1e-1). To eliminate any wrongly assigned sequence, we performed a reciprocal Blast search (e-value $1e^{-9}$) against the proteomes and excluded any sequence whose best hit did not belong to the corresponding orthologous group.

Supermatrix assembly, marker selection and taxonomic sampling

To assemble a comprehensive phylogenomic dataset for bilaterians, we incorporated our novel transcriptomes from chaetognaths and we included transcriptomic data from previously published studies, in particular [5, 8, 16]. Some taxa are sometimes represented in databases by a large number of RNA-seq datasets. To take into account the computational limitations, our philosophy was to assemble a dataset with a balanced number of taxa in each phylum (4–6 species), to minimize missing data, and to preferentially retained the slowest evolving species. Our initial dataset includes 103 taxa (Figure 1C and Figure S1) and we alternatively analyzed datasets with subsets of 83, 70 and 65 taxa (Figure 1C).

Proteins extracted from transcriptomes and genomes were independently aligned using Msaprobs (v0.9.7) [45]. To detect and remove mis-translated regions, we used Hmmclean with a threshold of 20 [50]. Hmmclean compares each sequence to a Hidden Markov Model profile derived from the alignment without this sequence and removes highly divergent stretches of amino acid. Alignments were further trimmed with BMGE to exclude blocks of highly variable misaligned residues with a maximum gap rate of 0.9 (-g option) [47]. After these steps, a ML tree was reconstructed for each individual marker alignment with RAxML (v8.2.4) assuming a LG+Γ_4 model [46]. To further exclude possible residual contaminations, we calculated the median absolute deviation of the distance to the root for all taxa, and we excluded those showing a 20-fold higher distance than this deviation, leading to the exclusion of 104 sequences in total. To perform marker gene stratification, we calculated the saturation for each of them as the linear regression coefficient between the ML p-distance and the percentage identity for each pair of taxa.

Phylogenetic analyses

Concatenation of all 1174 genes yielded a 416,663 amino acid supermatrix with 34.59% missing data. We analyzed the whole matrix using a per-gene partition scheme and a LG4X+R model in IQ-TREE (v1.6.2) [48]. Support values were estimated by ultrafast bootstrapping for 1000 replicates with the UFBoot option to account for model violation (-bnni -bb 1000) (Figures 1C, S1, and S2). Such an alignment is computationally too expensive to examine using a site-heterogenous model. Therefore, we selected the 267 marker genes with the lowest levels of saturation computed as explained before. This yielded a 74,014 positions matrix (Table S3). This matrix was examined using Phylobayes-mpi assuming a CAT+GTR+Γ_4 model with 4 chains running for more than 5000 generations and the first 1500 cycles discarded as burn-in [11]. Total computation represented more than 9 weeks of computation with 50 cores per chain.To evaluate the possible impact of compositional heterogeneity across taxa, we performed a CAT+GTR+Γ_4 analysis after recoding the data in broad protein categories using Dayhoff6 scheme (Figure 1 and Table S3). Convergence was assessed using the bpcomp command and by visually inspecting parameters values for the multiple runs. Alternatively, to approximate the CAT model in a maximum-likelihood framework, we also applied IQTREE with a C20 mixture of profiles, the LG matrix of exchange rates and free-rates heterogeneity (LG+C20+R4+FO) (Table S3). Composition heterogeneity was asses using a posterior predictive analysins in Phylobayes-mpi on bayesian samples for alternative datasets and models of evolution [11]. A Z-score was used as a measure of the deviation to the null-hypothesis (homogeneous compositional distribution) across replicates of PPA; the higher the absolute z-score, the stronger the deviation.

We examined trees with several taxonomic sampling: 'broad' corresponded to most exhaustive datasets (103 taxa), 'red' (reduced) excluded taxa showing strongest compositional biases and fastest evolutionary rate (83 taxa), 'strin' (stringent) excluded further taxa (70 taxa), 'cnid' used only cnidarians (and not sponges) as an outgroup (65 taxa) and 'proto' included all protostome taxa (67 taxa). Observed topologies as detailed in Figure 1C are the following: Rouphozoa (Platyhelmintha + Gastrotricha), Parenchymia (Nemertea+Platyheminthes), Lophophorata (Brachiopoda + Phoronida + Ectoprocta [Bryozoa]), Gastrotriches with Lophophorata (Gastro+Lopho), Tetraneuralia (Mollusca+Entoprocta), Polyzoa (Ectoprocta [Bryozoa] + Entoprocta), Vermizoa (Annelida + Nemertea disregarding platyhelminthes branching), Xenambulacraria (Xenacoelomorpha + Ambulacraria) and Nephrozoa (for Xenoacoelomorpha sister-group of other bilaterians) (Table S3).

DATA AND SOFTWARE AVAILABILITY

The alignments, phylogenetic trees, bayesian samples have been deposited as a zenodo dataset (https://doi.org/10.5281/zenodo.1403005). Datasets are available on Zenodo as https://doi.org/10.5281/zenodo.1403005.

The sequencing reads for newly sequenced species have been deposited under the bioproject NCBI: PRJNA487918 with the accessions NCBI: SRR7754742-SRR7754750 and NCBI: SRR8149062.