



Scaphopoda is the sister taxon to Bivalvia: Evidence of ancient incomplete lineage sorting

Hao Song^{a,b,c,1} , Yunan Wang^{a,c,1} , Haojing Shao^{d,1}, Zhuoqing Li^{a,c,1}, Pinli Hu^{d,1} , Meghan K. Yap-Chiongco^e , Pu Shi^{a,c} , Tao Zhang^{a,b,c} , Cui Li^{a,c}, Yiguan Wang^f , Peizhen Ma^{a,c}, Jakob Vinther^{g,h}, Haiyan Wang^{a,b,c,2} , and Kevin M. Kocot^{e,i,2}

Edited by Gonzalo Giribet, Harvard University, Cambridge, MA; received February 13, 2023; accepted August 18, 2023 by Editorial Board Member David Jablonski

The almost simultaneous emergence of major animal phyla during the early Cambrian shaped modern animal biodiversity. Reconstructing evolutionary relationships among such closely spaced branches in the animal tree of life has proven to be a major challenge, hindering understanding of early animal evolution and the fossil record. This is particularly true in the species-rich and highly varied Mollusca where dramatic inconsistency among paleontological, morphological, and molecular evidence has led to a long-standing debate about the group's phylogeny and the nature of dozens of enigmatic fossil taxa. A critical step needed to overcome this issue is to supplement available genomic data, which is plentiful for well-studied lineages, with genomes from rare but key lineages, such as Scaphopoda. Here, by presenting chromosome-level genomes from both extant scaphopod orders and leveraging complete genomes spanning Mollusca, we provide strong support for Scaphopoda as the sister taxon of Bivalvia, revitalizing the morphology-based Diasoma hypothesis originally proposed 50 years ago. Our molecular clock analysis confidently dates the split between Bivalvia and Scaphopoda at ~520 Ma, prompting a reinterpretation of controversial laterally compressed Early Cambrian fossils, including *Anabarella*, *Watsonella*, and *Mellopegma*, as stem diasomes. Moreover, we show that incongruence in the phylogenetic placement of Scaphopoda in previous phylogenomic studies was due to ancient incomplete lineage sorting (ILS) that occurred during the rapid radiation of Conchifera. Our findings highlight the need to consider ILS as a potential source of error in deep phylogeny reconstruction, especially in the context of the unique nature of the Cambrian Explosion.

Scaphopoda | mollusc phylogeny | mollusc fossils | incomplete lineage sorting | Cambrian explosion

The Cambrian Explosion marks a crucial but mysterious point in the history of life on Earth when almost all major animal phyla simultaneously emerged (1). Because of its rich fossil record, Mollusca is potentially one of the most informative clades for understanding the nature of Cambrian Explosion (2, 3). However, the extreme disparity of molluscan body plans, exemplified by well-known representatives (snails, slugs, clams, octopuses, chitons, etc.) and other rare groups (aplacophorans, monoplacophorans, and scaphopods) and incongruent results in molecular phylogenetic studies have given rise to conflicting phylogenetic hypotheses, hindering interpretation of the group's fossil record and understanding of its early evolution (3–5).

Although analyses of morphology and molecular phylogenetic studies based on a small number of loci have rarely found strong support for class-level relationships within Mollusca, phylogenomic studies (6–8) have made significant progress in recent years. These studies have consistently supported a “basal” dichotomy that separates Mollusca into two lineages, Aculifera (Solenogastres + Caudofoveata as a monophyletic Aplacophora, which is the sister taxon of Polyplacophora) and Conchifera (Bivalvia, Gastropoda, Cephalopoda, Monoplacophora, and Scaphopoda; a clade of uni- or bivalved mollusks). However, within Conchifera, higher-level relationships have been difficult to resolve, even with the application of phylogenomic approaches. A major question and long-standing debate is placement of Scaphopoda (“tusk shells”) (4–11). Scaphopods (Fig. 1 *A* and *B*) are among the least commonly encountered and least studied members of Mollusca that display morphological, ontogenetic, and even genomic features resembling those of cephalopods, gastropods, and bivalves (12). At least four competing hypotheses regarding the phylogenetic position of Scaphopoda have been proposed (Fig. 1 *C–F*). Resolution of the phylogenetic position of Scaphopoda is of great interest to paleontologists because several laterally compressed Cambrian mollusk fossils have been interpreted as stem taxa belonging to the clade that gave rise to Scaphopoda and the extinct Rostroconchia, stem bivalves, or stem diasomes (3). Phylogenetic placement of Scaphopoda would also benefit

Significance

Scaphopods are among the rarest and most enigmatic members of Mollusca whose phylogenetic placement has been long disputed, thus impeding understanding early molluscan evolution and the identity of problematic fossil taxa. By sequencing scaphopod genomes and applying robust phylogenomic approaches, we provide strong evidence for a Scaphopoda–Bivalvia clade (Diasoma). This allows us to reinterpret many problematic fossil taxa, including *Anabarella*, *Watsonella*, and *Mellopegma* as stem diasomes. We show that previous uncertainty regarding scaphopod placement in phylogenomic studies was likely due to incomplete lineage sorting (ILS) that arose during the rapid cladogenesis of the Cambrian Explosion, prompting further consideration of ILS when addressing deep recalcitrant nodes in the animal tree of life.

Author contributions: H. Song, H.W., and K.M.K. designed research; H. Song, P.S., T.Z., C.L., P.M., H.W., and K.M.K. performed research; H. Song, T.Z., H.W., and K.M.K. contributed new reagents/analytic tools; H. Song, Yunan Wang, H. Shao, Z.L., P.H., M.K.Y.-C., Yiguan Wang, J.V., H.W., and K.M.K. analyzed data; and H. Song, J.V., H.W., and K.M.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.G. is a guest editor invited by the Editorial Board.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹H. Song, Yunan Wang, H. Shao, Z.L., and P.H. contributed equally to this work.

²To whom correspondence may be addressed. Email: haiyanwang@qdio.ac.cn or kmkocot@ua.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2302361120/-/DCSupplemental>.

Published September 22, 2023.

interpretation of the evolutionary trajectory of key morphological traits that have been interpreted as synapomorphies of various hypothesized clades (10, 11).

Here, we sequenced genomes from both extant scaphopod orders—Dentaliida (*Pictodentalium verneidei*) and Gadilida (*Siphonodentalium dalli*)—and used genome-scale phylogenetic analyses to resolve scaphopod placement, which strongly support a sister taxon relationship between Scaphopoda and Bivalvia. We found evidence for ancient and pervasive incomplete lineage sorting (ILS) in conchiferan genomes that causes gene and species tree incongruency. Using carefully selected fossil calibrations and non-ILS genes under both concatenation- and coalescence-based methods, we inferred the divergence times of major molluscan lineages, which sheds light on the phylogenetic affinities of several problematic laterally compressed fossils. This work resolves a long-standing question in invertebrate zoology, which provides important insight into early molluscan evolution and highlights the impact and possible prevalence of ILS because of the rapid emergence of animals during the Cambrian explosion.

Results and Discussion

Divergent Genomic Architecture of Two Scaphopods. We sequenced the genomes of *S. dalli* from the Southern Ocean and *P. verneidei* from the East China Sea to represent both extant scaphopod orders—Gadilida and Dentaliida, respectively. The 2.30 Gbp genome of *S. dalli* and 6.02 Gbp genome of *P. verneidei* (SI Appendix, Fig. S1) were sequenced with 67× and 85× coverage of PacBio HiFi reads and Illumina short reads, respectively, and resulting assemblies were scaffolded using Hi-C. The *S. dalli* assembly and scaffolding yielded a contig N50 of 2.11 Mb, a scaffold N50 of 234.87 Mbp, and 9 chromosomes, while assembly and scaffolding of *P. verneidei* yielded a contig N50 of 1.83 Mb, a scaffold N50 of 580.33 Mbp, and 10 chromosomes. These are among the most contiguous and complete genomes of mollusks or lophotrochozoans sequenced so far (SI Appendix, Table S1), with 95.4% and 95.0% BUSCO [Benchmarking Universal Single-Copy Orthologs (22)] completeness, respectively.

The *P. verneidei* genome is not only much larger than that of *S. dalli*, but also the largest molluscan genome sequenced so far. *P. verneidei*

and *S. dalli* have 35,615 and 31,723 protein-coding genes, respectively, comparable to other mollusks. However, comparison of gene and genome characteristics of *P. verneidei* and 12 other phylogenetically diverse mollusks revealed that long genes and long introns are strikingly more prevalent in the *P. verneidei* genome (SI Appendix, Table S2). While average coding region lengths in all 13 genomes were similar (ranging from 1,016 to 1,644 bp), *P. verneidei* displayed much longer introns, with an average intron length of 16,808 bp versus 777 to 4,820 bp in the other mollusks (SI Appendix, Fig. S1 and Table S2). Long introns usually have detrimental effects for transcription and splicing (23), as seen in *S. dalli* where genes with the longest total intron length show significantly lower median expression levels. However, in *P. verneidei*, long introns do not seem to adversely affect expression (SI Appendix, Fig. S2), suggesting the evolution of efficient transcription systems for accurate exon recognition.

Whole-Genome-Based Phylogeny Supports Scaphopoda+Bivalvia.

To resolve the internal phylogenetic relationships of Conchifera and placement of Scaphopoda, we obtained another 20 high-quality genomes spanning the higher-level diversity of Mollusca plus two other lophotrochozoans (*Lingula* and *Eisenia*) as outgroups. All classes of Mollusca were sampled except for Monoplacophora; the published Illumina-only monoplacophoran genome (8) was not included here because of its incompleteness and high level of fragmentation. We identified orthologous sequences using a conservative bioinformatic pipeline and conducted phylogenomic analyses on five datasets including a 92% occupancy (i.e., each gene was sampled for at least 92% of the taxa) supermatrix composed of 663 genes (92_pct), a 75% occupancy supermatrix (3,825 genes; 75_pct), a 50% occupancy supermatrix (6,430 genes; 50_pct), and matrices with the top 250 and top 500 genes scored by genesortR (24). Analyses were performed under multiple concatenation-based and coalescent-aware methodologies, with the former analyzed using both maximum likelihood (ML) and Bayesian inference (BI) implementations of site-heterogeneous and site-homogeneous models, as these approaches are known to differ in their susceptibility to model violations (25, 26). Results were remarkably stable across all datasets and methods, with all nodes identically resolved, including support for placement of

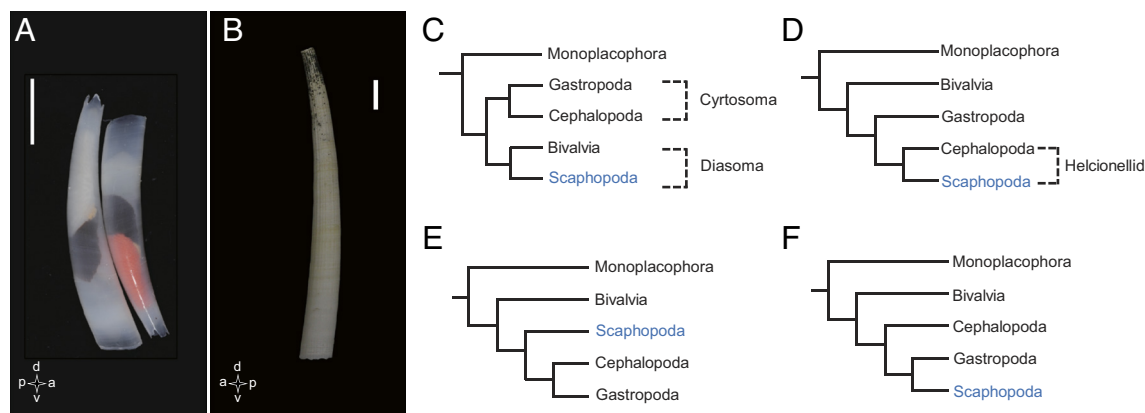


Fig. 1. Scaphopods and competing hypotheses with respect to Scaphopoda placement. (A) *Siphonodentalium dalli* male (Left) and female (Right). The anatomical orientation of the male is indicated in the lower left corner. d/v, dorsal/ventral; a/p, anterior/posterior. (B) *Pictodentalium verneidei*. (C) The Diasoma-Cyrtosoma hypothesis proposing a Scaphopoda-Bivalvia (Diasoma) clade, which is supported by similarities in the weakly developed head, pedal morphology, formation of the mantle and shell, and lateral compression of the body with the Paleozoic group Rostroconchia (13–16). (D) The helcionellid concept which places Scaphopoda and Cephalopoda as sister taxa, which is established on interpretation of shell coiling direction making Helcionellida a plesiomorphic total group (17, 18) and additionally some analyses of 18S rDNA sequences (10) and neurophylogenetic interpretations (11). (E) The hypothesis that places Scaphopoda sister to a Gastropoda-Cephalopoda clade based on shared features including fewer than three dorsoventral muscle pairs and the presence of the hydrostatic muscular system (19). (F) The hypothesis that places Scaphopoda sister to a Bivalvia-Cephalopoda clade based on similarities of branched head tentacles, prominent dorsoventral body axes, and the occurrence of shell slits (20, 21) as well as by some phylogenomic analyses (7, 8).

Scaphopoda as the sister taxon of Bivalvia (Fig. 2 *A* and *B* and *SI Appendix*, Fig. S3), bolstering support for the Diasoma (13) concept—a relationship previously supported by morphology but hardly recovered by molecular phylogenetics.

Analyses of the 50% occupancy dataset unambiguously supported placement of scaphopods as the sister group of bivalves regardless of the chosen method of inference, while only analyses of smaller datasets (ASTRAL analysis of 92_pct, LG+C60+F+G analysis of 75_pct, and analyses of the most strict subsampling by genesortR) resulted in weakened support for placement of Scaphopoda (Fig. 2*B* and *SI Appendix*, Fig. S1). Contentious phylogenetic relationships can be driven by a handful of genes with strong nonphylogenetic signal in phylogenomic datasets (27). We investigated this possibility by conducting a series of sensitivity analyses removing a handful of genes (1, 10, 50, and 100) with the highest difference in site-wise log-likelihood scores (Δ SLS) (27) from the 92%, 75%, and 50% occupancy matrices, which did not alter placement of Scaphopoda (*SI Appendix*, Fig. S4).

Uncertainty regarding scaphopod placement, as revealed using likelihood mapping (28), does not stem from a lack of phylogenetic signal but rather from the presence of conflicting signal in the dataset (Fig. 2*C*). This could explain why scaphopod placement has been difficult to determine in previous studies (5–8, 10, 29, 30). A careful dissection of our data revealed many gene trees that are incongruent with the inferred species tree with respect to the internal relationships of Conchifera (Fig. 2*D*). For 92_pct, 75_pct, and 50_pct, the proportion of genes supporting unrooted Tr1 (Scaphopoda+Bivalvia), Tr2 (Scaphopoda+Cephalopoda), and Tr3 (Scaphopoda+Gastropoda) were quite close, ranging from 35.37 to 37.25%, 29.93 to 30.48%, and 32.28 to 34.69%, respectively. We also summarized gene tree discordance (Fig. 2*E*), which reinforced our finding that a substantial fraction of gene trees are incongruent with the inferred species tree regarding scaphopod placement.

Deeply Conserved Synteny and Synteny-Based Orthology Inference. The large variability of both chromosome number (8 to 46) (31) and genome size ($C = 0.3$ to 7.98) (32) among mollusks implies that interchromosomal rearrangements have occurred frequently over the course of molluscan evolution. Unexpectedly, comparative genomic studies have revealed gene linkages across diverse animal lineages, from sponges to bilaterians, which can be used to infer ancestral karyotypes. A previous study showed that the scallop *Patinopecten yessoensis* has retained a near-perfect correspondence to ancestral bilaterian linkage groups (BLGs), indicating the ancestral molluscan chromosomes likely resembled chromosomes of their bilaterian progenitors in gene content and organization (33). Follow-up studies also found that retention of ancient BLGs exists at various levels in both gastropods and *Nautilus* (34). However, the phylogenetic extent of chromosome-scale synteny among mollusks remains elusive.

We retrieved 23 BLGs (34) and inferred chromosome-scale synteny and karyotype evolution among the assembled scaphopod genomes and previously published chromosome-level genomes of the bivalve *P. yessoensis*, the gastropod *Achatina fulica*, and the cephalopod *N. pompilius*, a small but diverse set of chromosome-scale molluscan genomes. As evident from Fig. 3*A*, scallop chromosomes show clear correspondence to BLGs, with 15 chromosomes having 1:1 correspondence to BLGs, while the remaining four chromosomes (Chr 1, 2, 3, and 4) have a 1:2 correspondence, indicating that these four chromosomes are the result of fusion-with-mixing. By comparing the chromosomes of the scallop to the chromosomes of other conchiferans, we see that three of those four fusions-with-mixing (scallop Chr 2, 3, and 4) are shared by all sampled conchiferan

mollusks, implying that they occurred before their divergence and represent a plesiomorphy of the sampled conchiferans. Meanwhile, in scallop Chr 1, a fusion-with-mixing result of two BLGs [BLG-M (brown) and BLG-B1 (cobalt blue), Fig. 3*A*] can be seen as an autapomorphy that is unique to the scallop lineage. This suggests that the last common ancestor of scaphopods and the other sampled conchiferans had an $n=20$ karyotype, the same number as has been inferred for the common ancestor of bivalves and gastropods (34).

The scaphopod genomes show a near-perfect 1:2 correspondence to the scallop genome (Fig. 3*A*). Most scaphopod chromosomes are the result of fusion-with-mixing of two ancestral linkages (where scallop Chr 1 represents two conchiferan ancestral linkage groups). Only Chr 3, 4, and 6 of the scaphopods underwent a more complex process. We infer that the BLG-J1+L (metallic blue+ bud green) and BLG-J2 (brownish orange; represented by *P. yessoensis* Chr 4 and 18, respectively) had a reciprocal translocation first, then one chromosome fused-and-mixed with BLG-B1 (cobalt blue), which gave rise to Chr 4 of *P. verednei*, while another fused-and-mixed with BLG-C1 (grayish blue), which led to Chr 6 of *P. verednei*. Chr 3 in *P. verednei* likely resulted from fusion-and-mixing of BLG-N (dark green; scallop Chr 14), BLG-P (grayish pink; scallop Chr 15), and a segment from breakage of BLG-A1 (dark pink; scallop Chr 5). Despite 300 million years of divergence, we find a striking near 1:1 correspondence between Gadilida and Dentaliida in terms of BLGs with the only notable difference being that Chr 3 of *S. dalli* corresponds to Chr 3 and 7 in *P. verednei*, indicating a fusion-with-mixing during *S. dalli*'s more recent evolutionary past.

While most genes participate in deeply conserved syntenies (*SI Appendix*, Fig. S5*A*), some genes have “jumped” between chromosomes (gray dots). These nonsyntenic genes account for between 5.92% (*Pictodentalium-Siphonodentalium*) and 15.07% (snail-scallop) of recognized orthologs (*SI Appendix*, Table S3) that are dispersed across noncorresponding chromosomes (omitted for clarity in Fig. 3*A*). Such variable synteny can be caused by the cumulative effect of numerous small-scale interchromosomal translocations over time (35). The rate of gene translocation among mollusks was estimated to be ~1% per 46 million years, which is similar to the translocation rate estimated among metazoans (36), and is around 10-fold slower compared to the typical gene-duplication rate in eukaryotes (37).

Recent studies have demonstrated the utility of synteny information in phylogenomics. By enabling the assignment of gene orthology relationships, synteny blocks can provide an invaluable framework for inferring the shared ancestry of genes, particularly for large multigene families where phylogenetic methods may be inconclusive (38). While this approach has proven effective in marsupials (38) and teleost fishes (39), its application has been limited in more anciently diverged lineages such as mollusks. We applied a likelihood-mapping analysis (28) to visualize the phylogenetic information content of the aligned syntenic sequences. Results (Fig. 3*B*) show that the Tr1 topology (Bivalvia+Scaphopoda) has 100% tree-likeness among all quartets, suggesting that this orthology inference strategy is suitable for phylogenetic reconstruction. Because most other extant mollusks have not retained deep ancestral synteny, we thus developed a second approach to infer orthology based on synteny information from a broader sampling of Mollusca. These analyses recovered a branching order (*SI Appendix*, Fig. S5*B*) consistent with that recovered in our analyses using more routine orthology inference methods (e.g., Fig. 2*A*), with maximal support for nodes. Although both approaches showed a similar proportion of gene tree discordance (Fig. 3*C*), they converged on a single scenario that unambiguously places Scaphopoda and Bivalvia in a monophyletic clade (Diasoma).

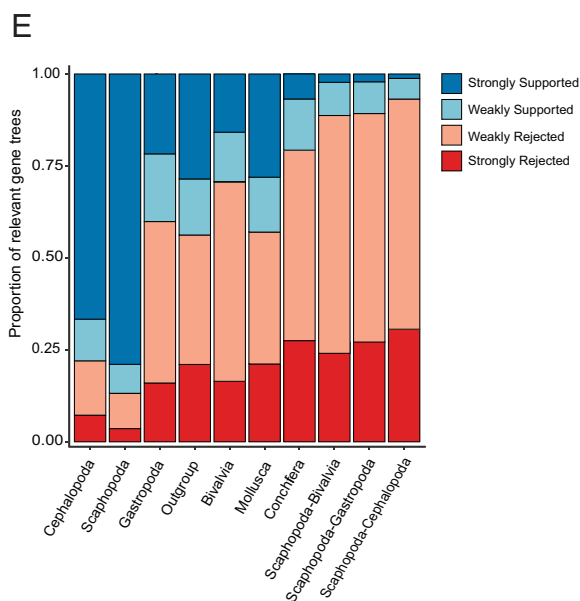
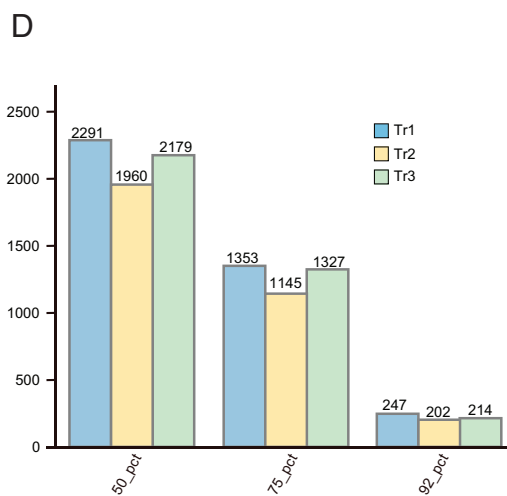
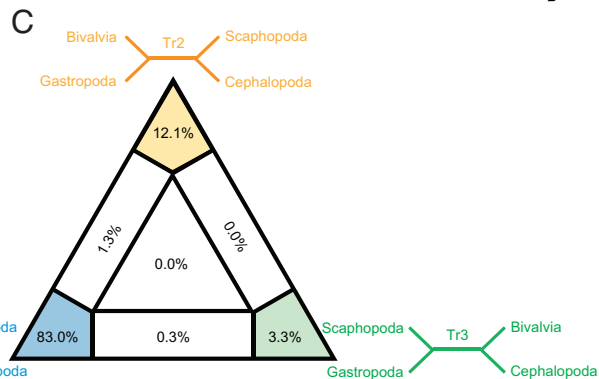
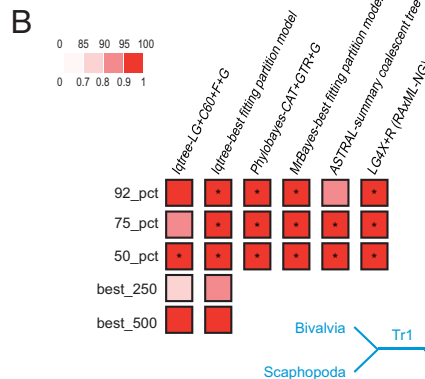
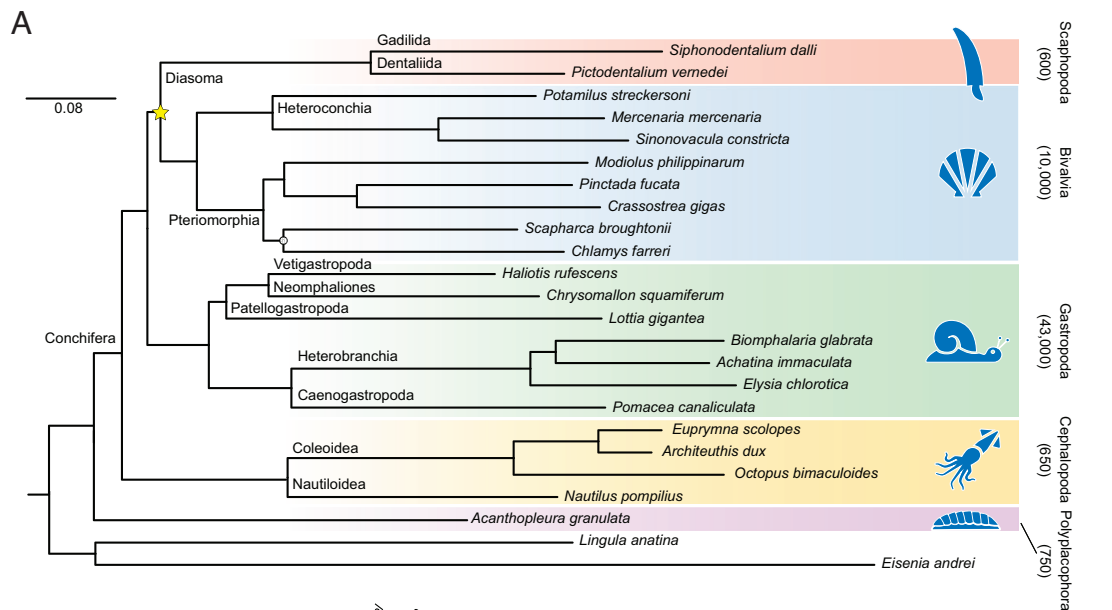


Fig. 2. Phylogenetic relationships among major clades of Mollusca inferred from genome assemblies. (A) Topology obtained analyzing 92_pct with best-fit partitioning scheme in IQ-Tree 2. All recovered clades received maximal support in all analyses with the exceptions of placement of Scaphopoda (marked with a yellow star) and the *Scapharca broughtonii*-*Chlamys farreri* clade (marked with a hollow circle). (B) Support for the Scaphopoda-Bivalvia sister group across analyses. The asterisk indicates BS = 100 or pp = 1. (C) Likelihood mapping analysis showing the proportion of quartets supporting different placements of Scaphopoda. The vast majority of quartets support the topology depicted in a (shown in blue, Tr1). The 92_pct supermatrix was used for likelihood mapping. (D) Discordance between gene trees and species trees in the 92_pct, 75_pct, 50_pct orthologous data matrices based on Astral results. Color coding and topological scenarios are as in C, with the blue column representing gene tree topologies that are the same as the species tree (Scaphopoda-Bivalvia). (E) Gene tree discordance inferred from DiscoVista. Splits compatible with gene trees are categorized as highly (weakly) supported based on bootstrap support values above (below) the 75% threshold. Weakly rejected splits are incompatible with the original tree but become compatible if low support branches are collapsed, while strongly rejected splits remain incompatible even after collapsing low support branches.

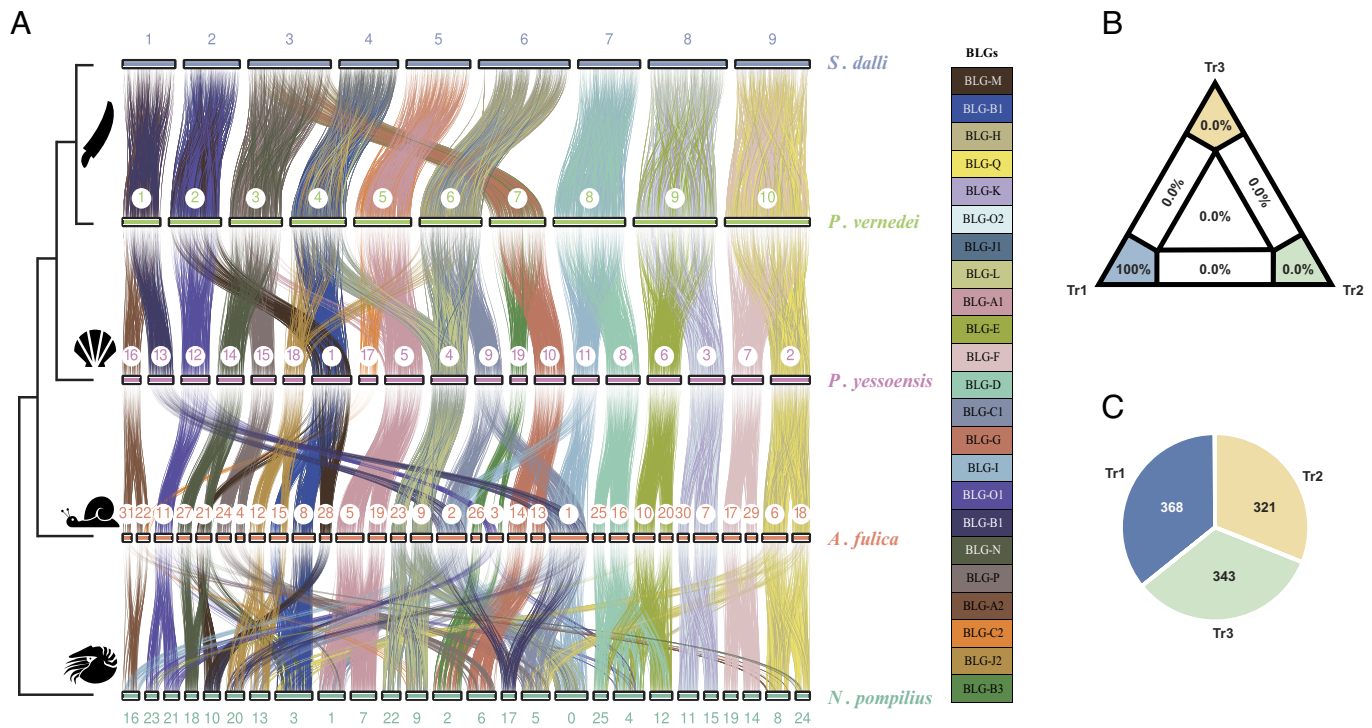


Fig. 3. Anciently conserved synteny across Conchifera. (A) Cladogram shown on the left. In the middle, numbered horizontal bars represent the chromosomes of five species. Colored vertical ribbons represent the orthologous genes among species. Only orthologous connections between chromosome pairs with statistically significant enrichment of conserved chromosomal synteny are shown. Between-chromosome synteny was colored according to the 23 bilaterian ancient linkage groups (BLGs) identified by ref. 34 using the same nomenclature (on the right). (B) Likelihood mapping to visualize the phylogenetic content of the synteny sequences, with five species in A included in the analysis. Tr1, Tr2, and Tr3 correspond to three tree topologies presented in Fig. 2C. (C) Discordance between gene trees and species trees in the synteny-based phylogenetic analysis.

We therefore used this tree as the phylogenetic framework for all downstream analyses.

Pervasive Signatures of ILS in Scaphopod Genomes. Even though Scaphopoda+Bivalvia was confidently recovered (Fig. 2A and B and SI Appendix, Fig. S5B), we observed substantial discordances between the species tree (Figs. 2D and 3C, Tr1) and gene trees, with the latter generating two alternative topologies, Tr2 and Tr3. The high proportion of genes, regardless of the matrix examined (50_pct, 75_pct, and 92_pct), in support of incongruent topologies could be explained by systematic error, ILS, or hybridization (40).

To test for systematic error, we plotted the three alternative topologies in the 92% occupancy dataset by a range of factors that may cause systematic error (e.g., saturation, compositional heterogeneity, proportion of variable sites) and observed no apparent clustering (SI Appendix, Fig. S6). We therefore can reject the possibility that topological conflict may be caused by site-based inconsistencies among the genes. To evaluate the potential influence of different orthology assessment methods, orthology was additionally inferred using BUSCO with the Metazoa_odb10 and Mollusca_odb10 datasets. Analyses based on this orthology inference approach also led to a consistent species tree with variable BS support on the Bivalvia–Scaphopoda node (SI Appendix, Fig. S7A), and orthogroups inferred using BUSCO showed no apparent clustering in terms of the three alternative topologies by the above-mentioned factors (SI Appendix, Figs. S8 and S9). The consistency in trees produced among the different approaches used for identifying orthogroups suggests that our phylogenetic inferences are robust across methods and that discordance in gene tree topologies is prevalent regardless of the method of orthology inference. This was probed further by excluding genes whose trees

do not recover incontrovertible class-level clades (i.e., monophyletic Bivalvia, Gastropoda, and Cephalopoda), as those genes are likely cases of inadvertent paralogy (41). This stringent subsampling to retain only orthogroups whose gene trees could recapitulate known, uncontested clades (RIC_pass genes in SI Appendix, Fig. S7B) recovered the same branching order (SI Appendix, Fig. S7C), although the BS value for Diasoma is lower than the original dataset. The retained RIC_pass dataset in 92_pct, 75_pct, and 50_pct also showed similar composition of genes supporting Tr1, Tr2, and Tr3, indicating the prevalent presence of conflicting signals in orthologs regardless of the orthology inference method. Because errors in alignment trimming could cause errors in tree estimation (42), we also compared two tools for trimming ambiguously aligned positions which yielded the same branching order (SI Appendix, Fig. S10).

This leaves ILS and hybridization as viable explanations for the discordance between gene trees and the species tree. A typical feature of ILS-caused incongruence is that the conflicting gene trees are usually equal in frequency, while in the case of hybridization, they are not (43). In this study, genes supporting Tr1, Tr2, and Tr3 are quite equal in frequency (36%/30%/34% on average), similar to ILS in marsupials (38) (Fig. 2D and 3C). The best way to distinguish genome-wide signatures of ILS and hybridization is to inspect the coalescence times of genome-wide orthologous genes, as ILS occurs before speciation events, whereas occurrence of hybridization is later than speciation (Fig. 4A). To test whether the observed incongruence is caused by ILS or hybridization, we partitioned the genomic orthologs into three paired-topology categories (Bivalvia–Scaphopoda, Bivalvia–Gastropoda, and Scaphopoda–Gastropoda) and reconstructed phylogenetic trees for each category and estimated divergence times using the same calibrated root age. The estimated

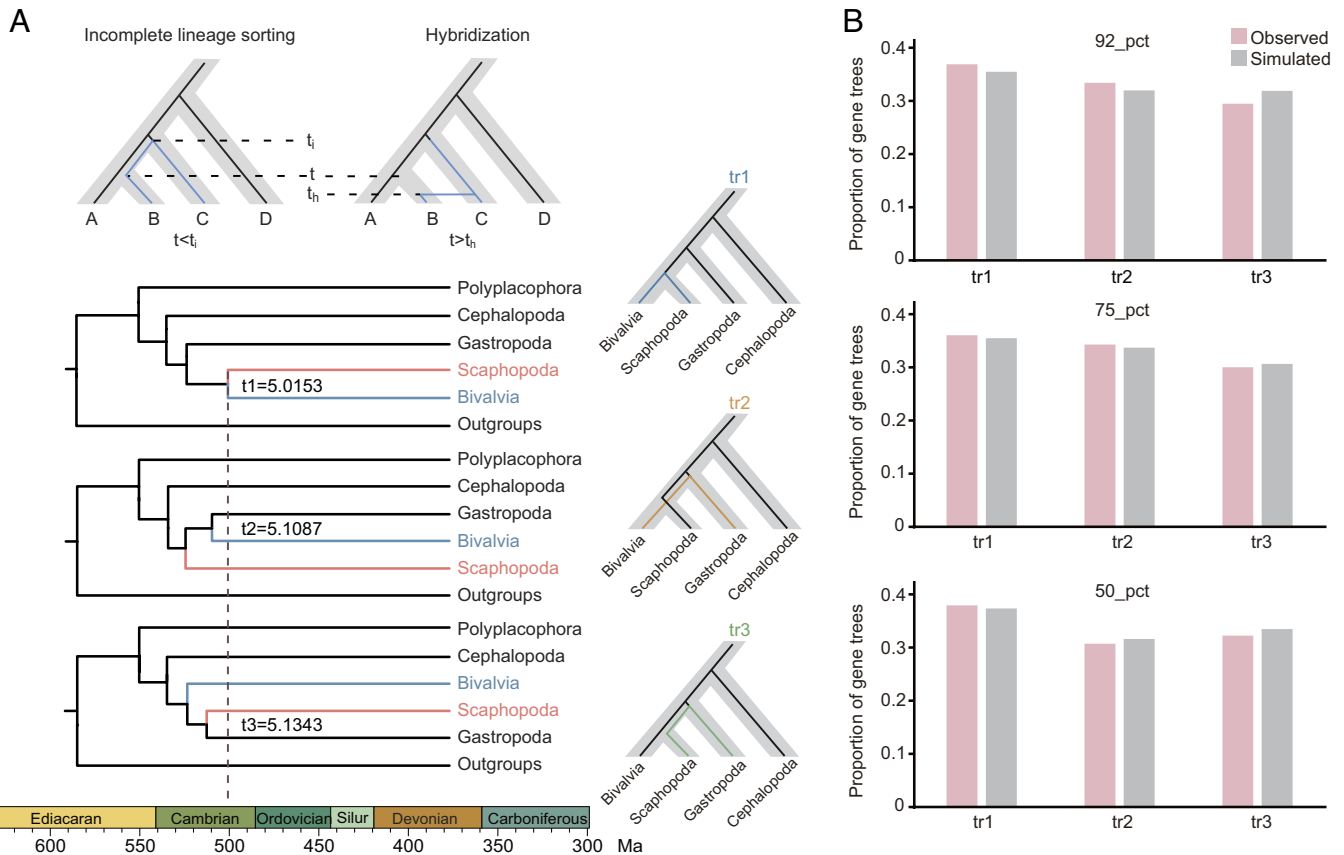


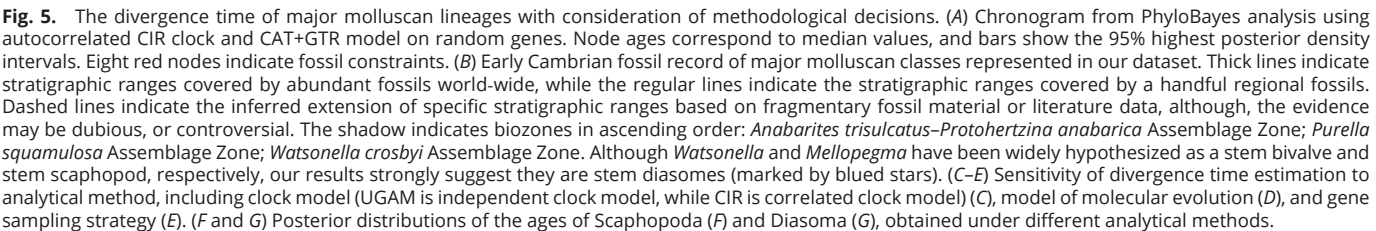
Fig. 4. Pervasive signatures of ILS in Scaphopoda genomes. (A) Using the estimated divergence times to distinguish between ILS and hybridization scenarios. As the upper schematic trees illustrate, the coalescence time under ILS (t_i) should be earlier than the speciation event (t), whereas the expected divergence time under hybridization (t_h) should be later than the speciation event (t). The lower trees show the divergence times of major molluscan clades estimated by MCMCTree with three potential genealogy scenarios as presented on the right. The species tree (non-ILS assumed) was designated as hypothesis 1 (Tr1), and two alternative genealogies, hypothesis 2 (Scaphopoda–Gastropoda sister group, Tr2) and hypothesis 3 (Bivalvia–Gastropoda sister group, Tr3), represent two alternative ILS scenarios. Color coding is the same as in Fig. 2C. (B) The topology frequency of observed gene trees and the corresponding simulated trees.

divergence time (t) between Scaphopoda and Bivalvia is expected to reflect the time of speciation, meaning the estimated divergence time between Gastropoda and Bivalvia or Scaphopoda from the other two categories should correspond to a younger expected divergence time under hybridization (t_h) or to a longer expected coalescence time under ILS (t_i). As shown in Fig. 4A, MCMCTree (44) inferred more ancient divergence times for alternative hypotheses Tr2 and Tr3 compared to the Bivalvia–Scaphopoda hypothesis (Tr1). Given that Tr1 was strongly favored across all datasets and methods (Figs. 2 and 3), this strongly suggests that pervasive signatures of incongruence across conchiferan genomes were caused by ILS, rather than hybridization. To corroborate this, we simulated 20,000 gene trees under the multispecies coalescent model on the basis of the ASTRAL trees (45). We observed high consistency between simulated and empirical gene trees, and the relative frequencies of various topologies (Tr1, Tr2, and Tr3) were in accordance with frequencies of ILS (Fig. 4B) as estimated from our coalescent analyses. These results indicate that ILS accounts for the incongruent placement of scaphopods rather than systematic error or ancient hybridization. While some have argued that ILS mainly is a concern for more recent divergences, it has been argued to be just as important and, of concern, in ancient diversification events (46).

ILS May Have Been Prevalent during the Cambrian Explosion.

The Cambrian witnessed a marked radiation of animal life, recorded in trace fossil diversity and disparity (47), skeletal fossils (48), and small carbonaceous fossils (49). From these records,

a picture emerges within which essentially all phyla diversified and led to complex macroscopic marine ecosystems during the Cambrian. Mollusca is one of the most informative clades for understanding the nature of this event. Thanks to secondarily phosphatized skeletal elements, the timing of the establishment of the major clades is well constrained (Fig. 5B) (3). Small shelly fossils appear near the Cambrian boundary at the onset of the Fortunian (538.8 Ma) (50), but no molluscan sclerites or shells are found. Those are soon followed by molluscan fossils including isolated sclerites of siphonogonuchitids and their sclerite-covered shell plates (*Maikhanella*) (51). Those taxa resemble aculiferan mollusks in several respects (52, 53) but could be stem/total group mollusks (54), while less equivocal aculiferans, such as *Halkieria* (3, 54), appeared in the late Fortunian [between 532 Ma and 529 Ma within the *Purella* biozone (51)]. Diverse taxa considered the oldest conchiferans appeared later within the same biozone, represented by a diversity of taxa, most notably the helcionellid *Oelandiella* (55). A range of laterally compressed limpet-like or monoplacophoran-grade taxa have been debated as either stem scaphopod or stem bivalve taxa (56, 57). Key characteristics are their laterally compressed shells and either internal thickenings of the shell (pegma) or the presence of laterally divided shells connected by a toothed hinge (56) and include among others *Anabarella*, *Watsonella*, and *Mellopegma*. A complicating factor in establishing the affinity of these taxa can potentially be attributed to the uncertainty of how scaphopods and bivalves are interrelated. The univalved taxon *Watsonella* has been hypothesized as a stem bivalve sharing a number of features with the slightly younger,



bivalved *Fordilla* and *Pojetaia* (56, 57) and hence constraining the divergence of Diasoma and offering a timeline for the divergence of diasomes from the oldest known conchiferans. *Watsonella* is a widespread taxon and defines a biozone that coincides with the onset of Cambrian Stage 2 about 529 Ma (50). *Anabarella* shares characteristics with both stem bivalves and early rostroconchs and occurs in the upper levels of the latest Fortunian *Purella* biozone (51). Therefore, based on this interpretation of the fossil record, the time between the divergences of Conchifera and Diasoma is narrowly constrained. Molluscan groups inferably evolved rapidly, explaining the very short internode branches in molecular phylogenetic studies. Such phenomena render phylogenetic inference problematic (58) and, combined with ILS due to pervasive polymorphisms in large populations, which is likely for mollusks given their near-cosmopolitan distributions during their early evolution (59), would complicate the issue further. Such conflating phenomena likely influenced many other bilaterian nodes that radiated during the Cambrian explosion.

Divergence Time Estimation and Re-Evaluation of the Fossil Record. In light of strong support for Diasoma, we revisited the Early Cambrian molluscan fossil record along with Paleozoic records that constrain each of the major molluscan classes represented in our dataset to establish a set of well-justified node calibrations (Dataset S1). We left the divergence time of scaphopods unconstrained to compare our divergence time estimates with hypotheses of their origin based on the fossil record.

Molecular divergence time estimation depends on methodological choices and confident calibrations from the fossil record. We analyzed the sensitivity of node ages to alternative methodological choices including models of molecular evolution (site-heterogeneous vs. site-homogeneous), gene sampling strategies (five different datasets), and different molecular clocks (correlated vs. independent). Considering all these factors, 20 different calibration settings were investigated using a Bayesian approach under a constrained tree topology (shown in Fig. 5A). We examined between-group principal component analyses (bgPCAs) and the distribution of posterior probabilities for node ages to measure the overall effect of these decisions on inferred dates (Fig. 5C–E and SI Appendix, Figs. S11–S13). Our results reveal that the choice between alternative clock models explained 40.11% of the total variance in node ages across all analyses (Fig. 5C). In contrast, the choice of amino acid substitution model and gene sampling had much lesser effects, explaining 8.82% and 6.07% of the total variance, respectively (Fig. 5D and E). A similar result was recently found in a study of echinoids, emphasizing clock models and calibrations rather than other factors have the strongest impact on inferred divergence times (60), in which the autocorrelated clock model generally resulted in more unbiased results congruent with the fossil record.

We establish that the sampled conchiferans radiated in the Early Cambrian, which is largely informed by the molecular clock calibrations used (Fig. 5A). While we relied on the assumption that *Watsonella* is a stem bivalve and hence placed a minimum constraint for the age of Diasoma at 530.7 Ma, our analysis finds this calibration to be in violation of molecular clock, and, with our use of soft-bound calibrations, Diasoma was recovered with a median age of 520 Ma (Fig. 5G). This age precisely matches the occurrences of unequivocal bivalves represented by *Fordilla* and *Pojetaia* (Fig. 5B). Our molecular clock analysis therefore suggests that the oldest laterally compressed conchiferans (*Anabarella*, *Watsonella*, and *Mellopegma*) should be reconsidered as stem diasomes, rather than stem bivalves or stem scaphopods (Fig. 5B) (3, 61). To further evaluate this hypothesis, we performed a divergence time estimation

without a constraint on the Diasoma node, which resulted in a very similar inferred split time for Diasoma (519.41 Ma, SI Appendix, Fig. S14A).

Our analyses (Fig. 5A) also suggest a divergence of the unconstrained scaphopod crown group around 355.1 Ma (95% HPD = 376–325). Scaphopods are hypothesized to have evolved from conocardioid rostroconchs (14, 15, 62). While different rostroconch groups evolved tusk-like shells independently (62), modern scaphopods are thought to have evolved in the Carboniferous as represented by various taxa documented from at least the Viséan (346 to 330 Ma) (63). While resembling members of the macroscopic Dentaliida, this group may be anatomically ancestral and hence represent total group scaphopods. Noteworthy is the Late Carboniferous (Kassimovian: 307 to 303 Ma) scaphopod *Minodentalium kansasense* (64), which exhibits a constricted aperture occurring in members of Gadilida, the other major clade of living scaphopods (65). These fossils tentatively offer a minimal bracket of 346 to 303 Ma for the origin of the scaphopod crown group, consistent with our estimates.

The phenomenon of deep coalescence can give rise to gene tree heterogeneity, thereby influencing estimates of divergence (66). Given the high incidence of ILS, we took care to minimize its impact through appropriate gene sampling by focusing on genes that supported the Tr1 topology. As an alternative approach to mitigate ILS in divergence time estimation, we also employed a coalescence-based approach, namely, StarBeast3 (67). Although this method produced a wide 95% HPD interval for younger nodes, the estimates for older nodes, such as Conchifera and Diasoma, were quite consistent with the results obtained from concatenation-based methods (SI Appendix, Fig. S14). These findings lend further support to our interpretations regarding the early cladogenesis of major molluscan lineages.

Evolution of Diasome Body Plans. Our phylogenetic framework also enables reappraisal of the evolution of several clade-defining traits, for example, the differentiation of a distinct head and the affiliated cephalic retractor muscles. Whereas Gastropoda and Cephalopoda have well-developed heads and a single pair of cephalic retractors, the head of Scaphopoda is simplified and not very distinctive. It does not protrude from the shell and is not separated from mantle or visceral sac clearly. However, scaphopods have retained one pair of cephalic retractors and are thus intermediate between forms exhibiting a well-developed head (Gastropoda and Cephalopoda) and the “headless” Bivalvia (10). In light of our reconstructed topology, we pose that the clade Ganglionata [Conchifera minus Monoplacophora (19)] was ancestrally cephalate, and dedifferentiation of the head region is a synapomorphy of Diasoma. Additionally, a lateroventral extension of the mantle and a shell enclosing the body, a burrowing foot, and an epiatroid nervous system (i.e., nervous system with closely associated or fused cerebral and pleural ganglia) can also be interpreted as a synapomorphy of Scaphopoda and Bivalvia (17).

Homeobox genes are central to body patterning and tissue segmentation during metazoan evolution (68, 69). These genes display widely conserved synteny among diverse animals, generally occurring in ordered clusters with the paraHox cluster beginning with Gsx and ending with Cdx and the Hox cluster starting with Hox1 and terminating with Post1. Both sets of these transcription factor-encoding genes exhibit collinear (staggered) expression along the anterior-to-posterior axis in most metazoans. The molluscan ancestor has been inferred to have intact homeobox clusters, but deviation from the canonical staggered homeobox gene expression is well documented in conchiferans (33, 70). In this study, we show that both scaphopod genomes retain a complete set of molluscan paraHox and Hox genes but differ from each other in

the location or orientation of *Gsx*, *Cdx*, *Lox5*, and *Post1* (*SI Appendix, Fig. S15*). *S. dalli* has an inversion of *Gsx* and a translocation of *Post1* with respect to the ancestral order inferred for Mollusca [i.e., the arrangement observed in the scallop (33)]. *P. vernedei* has a translocation of *Cdx*, and translocation-and-inversion of *Lox5*. Interestingly, the dentaliid scaphopod *Antalis entails*, a close relative of *P. vernedei*, exhibits staggered expression of Hox genes in early mid-stage trochophore larvae, with exception of *Lox5* and *Post1* (12). Disrupting this staggered expression pattern of *Lox5* (and *Post1*) could be correlated with the translocation and/or inversion of this gene.

Conclusion

This study presents two genomes for Scaphopoda, an enigmatic group rarely investigated yet critical for understanding early mollusc evolution. Although the phylogenetic position of Scaphopoda has been long contentious, our whole genome-based phylogenomic analyses consistently place Scaphopoda as the sister taxon of Bivalvia, consistent with the Diasoma concept, with a split that dates to 520.6 Ma (95% HDP = 522–517 Ma). The subphylum Diasoma, originally proposed approximately five decades ago, posits a common evolutionary origin, ecology, and general morphological bauplan uniting the Bivalvia, Scaphopoda, and the extinct Rostroconchia (14, 15). However, this hypothesized subphylum has seldom been recovered in prior molecular phylogenetic studies. Thus, our present phylogenomic findings provide critical corroboration for the paleontologically derived Diasoma hypothesis, demonstrating the explanatory power of integrating fossil and whole-genome-wide phylogenomic approaches and datasets to resolve evolutionary relationships.

Moreover, our phylogenetic framework for Mollusca resolved here sheds light on not only the identity of many important but controversial Cambrian fossils that show similarities to both bivalves and scaphopods (such as *Anabarella*, *Watsonella*, and *Mellopegma*), but also the evolutionary history of several clade-defining traits. Notably, by inspecting the coalescence times of thousands of genes across the genome, we show that previous incongruence among phylogenomic studies concerning placement of Scaphopoda is likely due to ILS spurred by the rapid radiation of conchiferans in a narrow interval of time during the Earliest Cambrian (~520 to 534 Ma). Such ancient but prevalent ILS in scaphopod genome argues for further consideration of ILS when addressing deep recalcitrant nodes in the animal tree of life.

Methods and Materials

Sampling and Sequencing. *P. vernedei* and *S. dalli* were collected from the east China Sea (27°31'32.73"N, 22°30'39.13"E) and Southern Ocean (63°56'06.7"S, 56°34'13.9"W), respectively.

DNA was extracted from foot tissue for *Pictodentalium* and the whole animal for *Siphonodentalium*. Illumina paired-end sequencing was performed on HiSeq X Ten (Illumina, Cambridge, MA) for *Pictodentalium* at Novogene Ltd. (Tianjin, China) and NovaSeq S4 flowcell at Psomagen (Cambridge, MA) for *Siphonodentalium*. SMRTbell libraries for PacBio sequencing were constructed using the SMRTbell Template Prep Kit 12.0 (PacBio, Menlo Park, CA, USA). *Pictodentalium* were sequenced on six cells of the PacBio RS II platform, while *Siphonodentalium* were sequenced on the Sequel II platform with five cells. Hi-C libraries were prepared using the Animal Hi-C Kit (Phase Genomics, Seattle, WA). The library for *Pictodentalium* was sequenced on a HiSeq X Ten at Novogene (Tianjin, China), and the library for *Siphonodentalium* was sequenced on a NovaSeq S4 at Psomagen, both with 2X150 bp reads.

For transcriptome sequencing, multiple tissues of living *Pictodentalium* were dissected and used for RNA extraction. An entire specimen of *Siphonodentalium* preserved in RNAlater was dissected to remove the digestive system, and the rest of the body was used for RNA extraction. Transcriptomes of both species were sequenced on an Illumina NovaSeq S4 flowcell with 2X150 bp reads.

Genome Assembly and Annotation. The *Pictodentalium* genome was assembled de novo based on PacBio subreads using FALCON (<https://github.com/PacificBiosciences/FALCON/>). The assembled sequences were then polished using Quiver (SMRT Analysis v2.3.0) based on the alignment of PacBio reads to the assembly. Several rounds of iterative error correction were performed using the Illumina data. The *Siphonodentalium* genome was assembled de novo based on HiFi reads using hifiasm v0.13 (71) with the option -l 3 to exclude redundant haplotigs. To scaffold contigs, we used Juicer (72) to compute the interaction matrix, then scaffolded contigs using 3D-DNA pipeline in Juicebox (73). To evaluate genome quality, we first mapped Illumina reads onto the assemblies with BWA (74). Next, genome completeness was assessed by BUSCO v5.4.2b (22). Because an entire animal was used for DNA extraction for *Siphonodentalium*, contamination was screened for and removed with BlobTools2 (75). We removed scaffolds with fewer than 5× Illumina read coverage, scaffolds not annotated as Metazoa, and scaffolds with a GC content <0.0.3 or >0.5, which appeared as clear outliers when GC content was plotted against coverage.

Repetitive sequences were identified through a combination of homologous comparison and de novo prediction as per Ma et al. (76). Gene annotation was carried out by a combination of de novo prediction, homolog-based prediction, and transcriptome-based prediction. De novo prediction was carried out as per Ma et al. (76). For homologous annotation, a protein library including *Acanthopleura granulata*, *Crassostrea gigas*, *Lottia gigantea*, and *Octopus sinensis* was used to search against the scaphopod genomes using TBLASTN (RID: SCR 011822). For transcriptome-based prediction, RNA-seq data were mapped against the assembly using HISAT2 v2.2.1 (77), and the transcripts were converted to gene models using Cufflinks v2.3.1 (78). All gene models were integrated via EvidenceModeler (79).

Site-Based Phylogenetic Inference. We selected 24 available genomes broadly spanning the diversity of Mollusca and other lophotrochozoans based on their high quality in terms of contiguity and completeness according to BUSCO (*SI Appendix, Table S1*). Homologous groups of sequences ("homogroups") among those species were identified using OrthoFinder v2.4.0 (80) with an inflation parameter of 2.1. Our approach for orthology inference, sequence selection, and matrix assembly and husbandry followed the bioinformatic pipeline of Krug et al. (81). Sequences <100 amino acids (a.a.) were removed from OrthoFinder output fasta files, and those sampled for ≥50% of taxa were aligned with MAFFT v7.453 (82). Putatively mistranslated regions were then removed with HmmerCleaner (83), and aligned homogroups were trimmed to remove ambiguously aligned regions with BMGE v1.12.2 (84). Approximately, ML trees were constructed for each alignment with FastTree 2 (85), and PhyloPyPruner v0.9.5 (86) was used to identify strictly orthologous sequences. Orthologous genes sampled from at least 92%, 75%, and 50% of the taxa were retained and concatenated into three data matrices, named 92_pct, 75_pct, and 50_pct. The 92_pct dataset was further filtered using genesortR (24) to quantify phylogenetic usefulness. A ML tree produced in IQ-Tree 2 (version 2.1.3) (87) based on 92_pct with the best-fitting model for each partition was used as the reference. The best 250 and 500 genes were concatenated into another two data matrices, best_250 and best_500.

Maximum likelihood analyses were conducted using IQ-Tree 2 and RAxML (88). For IQ-TREE analyses, the best-fitting partition model found by ModelFinder (89) and the LG+C60+F+G model were used. Topological support was assessed with 1,000 ultrafast bootstrap replicates. For RAxML analyses, the LG4X+R model was used, and topological support was assessed with 100 replicates of rapid bootstrapping. Bayesian inference analyses were performed using MrBayes (90) v3.2.0 and PhyloBayes MPI v1.6 (91). For MrBayes analyses, the best-fitting partition model was found by ModelFinder (89). For PhyloBayes analyses, the CAT-GTR+G model was used. Two independent runs were practiced in parallel. Using a burn-in of 1,000, and sampling every 10 trees, the bpcomp produced the largest (maxdiff) and mean (meandiff) bipartition discrepancy and the consensus tree. Chains were considered to have converged when maxdiff reached 0.1. Single-gene trees were reconstructed using IQ-Tree 2 based on all orthogroups of all three datasets. Coalescent-based analyses were estimated by ASTRAL v5.7.1 (92) with species trees inferred.

A likelihood-mapping analysis was implemented in IQ-Tree 2 to visualize the phylogenetic signal for alternative resolutions of the quartet including these three lineages (Cephalopoda, Gastropoda, and Bivalvia) and their sister clade (Scaphopoda). The 92_pct supermatrix was used for likelihood mapping. To further determine whether the branching order is influenced by a handful of genes with strong phylogenetic signals, we calculated Δ GLS values for each locus and excluded 1, 10, 50, 100, and outlier genes with the highest absolute Δ GLS values to see whether removal of these genes affected the branching order following the approach of ref. 27. Conflict between the reconstructed species tree and single gene trees was examined using Phylogenetic_signal_parser v1.1 (27). For each dataset, all genes were divided into three groups, that supported the Scaphopoda-Bivalvia tree, the Scaphopoda-Gastropoda tree, and the Bivalvia-Gastropoda tree, respectively. Gene tree discordance was also examined using DiscoVista (93). Ten groups were considered, including eight groups that are identified in the species tree (Mollusca, Conchifera, Bivalvia, Gastropoda, Scaphopoda, Cephalopoda, outgroups, and Scaphopoda-Bivalvia) and two conflictive groups (Scaphopoda-Gastropoda, and Scaphopoda-Cephalopoda). Bootstrap support values higher than 75% were considered as highly supported.

To test whether hidden paralogy in our phylogenomic dataset is driving alternative topologies in the placement of Scaphopoda, we employed two strategies. First, comparison of the different orthogroups inferred using OrthoFinder, BUSCO with the Metazoa odb10 dataset, and BUSCO with the Mollusca odb10 dataset was conducted to observe any changes in tree branching order. Orthogroups selected using BUSCO with both the Metazoa and Mollusca datasets were processed as described above, and three matrices with varying levels of tolerance of missing data were assembled (92_pct, 75_pct, and 50_pct) and used for tree construction in IQ-Tree 2 using the best-fitting model for each partition. The second strategy involved subsampling genes as per Mulhair et al. (41) with small modifications. Briefly, we used 92_pct supermatrix and employed DiscoVista to screen out orthogroups whose constructed gene trees could recapitulate known, incontestable clans (Bivalvia, Gastropoda, and Cephalopoda), while orthogroups whose gene trees did not recover these predefined clans as monophyletic were removed. After removing those orthogroups that failed to recover incontestable relationships, we constructed species tree using IQ-Tree 2 with the best-fitting model for each partition. A chi-squared test was applied to determine whether the number of genes supporting each of the three topologies were significantly different.

To test whether errors in alignment trimming is driving alternative topologies, we also employed Clipkit (94) as an alternative to BMGE to trim ambiguously aligned positions of the 92_pct,

75_pct, and 50_pct supermatrices before concatenating and then used IQ-Tree 2 to analyze the resulting trimmed supermatrices as described above.

Identification of Conserved Chromosome Synteny and Synteny-Based Phylogenetic Inference. To identify ancestral linkage groups, we selected several published mollusc genomes, including *N. pompilius* (95) from Cephalopoda, *A. fulica* (96) from Gastropoda, and *P. yessoensis* (33) from Bivalvia, in addition to our two scaphopod assemblies. A reciprocal best BLASTp (version 2.13.0+) search was performed to identify high-confidence homologous genes among molluscan genomes. Combined with the reciprocal best hits (MBHs, e-value cutoff of $1e^{-5}$), a total of 5,491 core homologous sequences were obtained, which means that any one gene in a set of homologous genes is homologous to all other genes. Following Simakov et al. (36), bilaterian linkage group (BLG) genes in *P. yessoensis* were used to filter the core homologs, resulting in 2,837 gene combinations. To determine the best chromosome-chromosome associations between species, we applied Fisher's exact test against a null model of random gene permutation. Bonferroni correction was further applied for the number of chromosome-chromosome tests in each species pair. Syntenic genes among those five species were used for a quartet likelihood-mapping analysis (28) following the methods described above. Because most other mollusks do not retain conserved ancestral synteny, we used the syntenic genes from the five representative mollusks as baits, then blasted them against 26 genomes (24 species in Fig. 2A plus *P. yessoensis* and *A. fulica* from Fig. 3A). The best-hitting sequences were concatenated and analyzed with ML in IQ-Tree 2 with the best-fitting model for each partition and with BI in PhyloBayes with the CAT+GTR+G model.

ILS Analysis. After having tested for gene tree topology conflict due to systematic error (amino acid site or orthology-identification based), we considered hybridization and ILS as possible explanations. Compared to speciation events, coalescence times for regions under ILS should be older, whereas hybridization should be younger. To distinguish between ILS and hybridization, we followed the methods of Feng et al. (38). We partitioned the genomic sequences into three paired-topology categories (Bivalvia_Scaphopoda, Gastropoda_Scaphopoda, and Cephalopoda_Scaphopoda) and reconstructed phylogenetic trees using concatenated genome sequences whose single-gene trees supported each grouping. Species divergence time was calculated with MCMCTree with an approximate likelihood calculation algorithm in PAML v4.9 (44). Before using gHmatrix to produce an out.BV file (containing the Hessian matrix), we applied Baseml in the PAML package to determine alpha and the substitution rate. Based on these parameters, we used MCMCTree to estimate divergence times. To improve the accuracy of the estimate, we constrained the root node at the same age, while other nodes were not constrained. Besides, we further performed ILS simulation using Phybase (97) as previously described (98). We then simulated 20,000 gene trees using the R function sim.coaltree.sp implemented in the Phybase package. The internal branch length of the ASTRAL tree was used, and all terminal branches were converted to equal lengths.

Divergence Time Calibration. We reviewed the molluscan fossil record in light of the topology recovered in this study, particularly with respect to Diasoma. Following the proposed practice for justifying fossil calibrations, we established eight calibrations for the major conchiferan crown groups and internodes (Dataset S1). To test the effect of these calibrations and the sensitivity and validity of the molecular clock under different permutations, we did not constrain the scaphopod crown group to derive estimates and compare those with the fossil record and hypotheses derived from those for the timing of scaphopod radiation.

The sensitivity of divergence time estimation to gene selection, clock type, and molecular evolution model was tested. We sampled one hundred of the non-ILS genes from the 75_pct dataset (i.e., concatenating orthogroups supporting Bivalvia+Scaphopoda) according to usefulness, RF distance, root-to-tip variance, and matrix occupancy as calculated in genesortR. We also sampled one hundred random non-ILS genes. Two substitution models (site-heterogeneous CAT+GTR+G and site-homogeneous LG+G) were also included. All datasets were analyzed in PhyloBayes under both independent and correlated clock models (UGAM and CIR). In all cases, we applied an unconstrained birth–death prior and soft bounds on divergence times. In total, twenty analyses were conducted based on the combination of these settings. Five hundred random iterations were sampled from the last 10,000 iterations for each analysis. Node ages were subjected to bgPCA using the Morpho package in R. Separate bgPCAs were performed for each of the factors explored, and the proportion of total variance explained by bgPCA axes was taken as an estimate of the relative impact of these choices on divergence times. We then selected five nodes, Mollusca-outgroup, Cephalopoda–Gastropoda–Scaphopoda–Bivalvia, Gastropoda–Scaphopoda–Bivalvia, Scaphopoda–Bivalvia, and the origin of Scaphopoda, and generated density distribution curves for each node under both clock types and all five gene selection strategies in Origin v.2022b.

According to the sensitivity of major nodes (SI Appendix, Figs. S11–S13), we thus estimated divergence times using PhyloBayes with the autocorrelated CIR clock model, CAT+GTR+G model of sequence evolution, and one hundred random non-ILS genes (parameters: pb -d xxxx.phy -T species.tree -cal calib -r outgroups -cir -cat -gtr -bd -sb -rp 580 30 chain_name). The first 5000 iterations were discarded as burn-in. Then, PhyloBayes sampled every two iterations until it gathered 10,000 samples. Two schemes for fossil calibrations were applied, one constrained eight nodes (the eight calibrations in Dataset S1) while the other constrained seven nodes (with the Diasoma node excluded).

StarBeast3 (67), a coalescence-based approach, was also applied to verify the divergence time estimation. One hundred of the non-ILS genes in the 75_pct dataset were analyzed using the best-fitting model for each gene as determined with ModelFinder (89) under a relaxed clock model with seven fossil calibrations (without Diasoma) with birth–death priors. The chain-length was set to 200,000,000 with the first 10% discarded as burn-in. We sampled every 500 generations until the ESS values in Tracer v1.7.1 (99) reached no less than 300.

Hox Gene Analysis. Based on available Hox gene models, arrangement of Hox genes in the genomes of the scaphopods *P. vernelei* and *S. dalli*, the bivalve *P. yessoensis* (33), the gastropod *L. gigantea* (96), and the cephalopod *N. pompilius* (95) were investigated under GeMoMa v.1.4.2 (100) with default parameters. We performed prediction by GeMoMa annotation filter under default parameters, with exception of the evidence percentage filter (e-value = 0.1). These were further manually scrutinized to reach a single high-confidence transcript prediction for each locus. Following the methods of Wang et al. (33), phylogenetic analysis was performed to confirm the exact annotations of each Hox gene.

Data, Materials, and Software Availability. The *P. vernelei* genome including all whole-genome and transcriptome sequencing data have been deposited with the NCBI under BioProject PRJNA903467 (101). The *S. dalli* genome including all whole-genome and transcriptome sequencing data have been deposited with the NCBI under BioProject PRJNA916950 (102). Annotations of both scaphopod genomes are available from Figshare (103). The 92_pct, 75_pct, and 50_pct data matrices as well as the scripts used for orthology inference and genome synteny analysis, are also available from Figshare (104).

ACKNOWLEDGMENTS. We thank the crew and scientists of the Icy Inverts cruises aboard RV Lawrence M. Gould and RVIB Nathaniel B. Palmer, especially Ken Halanych, for supporting specimen collection. We also thank Ximing Guo from Rutgers University, Jin Sun from Ocean University of China, and Pin Huan from IOCAS for helpful discussions during manuscript preparation. This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB42000000), NSF of China (Grant Nos. 32002409, 42376088, 42076092, 41776179), the National Key R&D Program of China (Grant Nos. 2022YFD2401301 and 2019YFD0901303), and the earmarked fund for CARs (CARs-49). H. Song was funded by Youth Innovation Promotion Association CAS and Young Elite Scientists Sponsorship Program by CAST (2021QNRC001). K.M.K. was funded by NSF DEB-1846174. The funders had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

Author affiliations: ^aInstitute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; ^bLaboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China; ^cUniversity of Chinese Academy of Sciences, Beijing 100049, China; ^dShenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518000, China; ^eDepartment of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487; ^fInstitute of Ecology and Evolution, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom; ^gSchool of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom; ^hSchool of Earth Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom; and ⁱAlabama Museum of Natural History, University of Alabama, Tuscaloosa, AL 35487

1. C. R. Marshall, Explaining the Cambrian "explosion" of animals. *Annu. Rev. Earth Planet. Sci.* **34**, 355–384 (2006).
2. W. F. Ponder, D. R. Lindberg, J. M. Ponder, *Biology and Evolution of the Mollusca* (CRC Press, Boca Raton, FL, ed. 1, 2020), vol. **1**, pp. 924.
3. J. Vinther, The origins of molluscs. *Palaeontology* **58**, 19–34 (2015).
4. K. M. Kocot, Recent advances and unanswered questions in deep molluscan phylogenetics. *Am. Malacol. Bull.* **31**, 195–208 (2013).
5. A. Wanner, T. Wollesen, The evolution of molluscs. *Biol. Rev.* **94**, 102–115 (2019).
6. K. M. Kocot et al., Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456 (2011).
7. S. A. Smith et al., Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367 (2011).
8. K. M. Kocot, A. J. Poustka, I. Stöger, K. M. Halanych, M. Schrödl, New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci. Rep.* **10**, 1–8 (2020).
9. P. D. Reynolds, The scaphopoda. *Adv. Marine Biol.* **42**, 137–236 (2002).
10. G. Steiner, H. Dreyer, Molecular phylogeny of Scaphopoda (Mollusca) inferred from 18S rDNA sequences: Support for a Scaphopoda–Cephalopoda clade. *Zool. Scripta* **32**, 343–356 (2003).
11. L. H. Sumner-Rooney et al., A neurophylogenetic approach provides new insight to the evolution of Scaphopoda. *Evol. Dev.* **17**, 337–346 (2015).
12. T. Wollesen, S. V. Rodríguez Monje, A. Luiz de Oliveira, A. Wanner, Staggered Hox expression is more widespread among molluscs than previously appreciated. *Proc. R. Soc. B* **285**, 20181513 (2018).
13. J. Pojeta Jr., The paleontology of rostronch mollusks and the early history of the Phylum Mollusca. US Geological Survey Professional Paper 968 (1976), pp. 1–54.
14. J. Pojeta Jr., B. Runnegar, "The early evolution of diosome molluscs" in *Evolution*, E. R. Clarke, M. R. Trueman, Eds. (Academic Press, Boston, 1985), vol. **10**, pp. 295–336.
15. B. Runnegar, J. Pojeta Jr., Molluscan Phylogeny: The Paleontological Viewpoint: The early Paleozoic fossil record shows how living and extinct molluscan classes originated and diversified. *Science* **186**, 311–317 (1974).
16. H. de Lacaze Duthiers, *Histoire de l'organisation, du développement, des mœurs et des rapports zoologiques du dentale* (Librairie de Victor Masson, Paris, France, 1858).
17. T. R. Waller, P. Johnston, J. Haggart, "Origin of the molluscan class Bivalvia and a phylogeny of major groups" in *Bivalves: An Eon of Evolution*, P. A. Johnston, J. W. Haggart, Eds. (University of Calgary Press, Calgary, 1998), vol. **1**, pp. 1–45.
18. B. Hatschek, *Lehrbuch der Zoologie, eine morphologische Übersicht des Tierreiches zur Einführung in das Studium dieser Wissenschaft* (Gustav Fischer Verlag, Jena, Germany, 1888).
19. G. Haszprunar, Is the Aplousobranchia monophyletic? A cladistic point of view. *Am. Malacol. Bull.* **15**, 115–130 (2000).
20. L. Plate, Über den Bau und die Verwandtschaftsbeziehungen der Solenococonchen. *Zoologische Jahrbücher der Anatomie* **5**, 301–386 (1892).
21. H. G. Bronn, *Dr. H. G. Bronn's Klassen und Ordnungen des Thier-Reichs, wissenschaftlich dargestellt in Wort und Bild. Dritter Band. Mollusca (Weichthiere)* (CF Winter'sche-Verlagshandlung, Leipzig, 1894).
22. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness" in *Gene Prediction*, M. Kollmar, Ed. (Springer, Humana, New York, NY, 2019), pp. 227–245.
23. D. F. Grabski et al., Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. *Wiley Interdiscip. Rev. RNA* **12**, e1631 (2021).
24. N. Mongiardino Koch, Phylogenomic subsampling and the search for phylogenetically reliable loci. *Mol. Biol. Evol.* **38**, 4025–4038 (2021).
25. M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).

26. F. Marlétaz, K. T. Peijnenburg, T. Goto, N. Satoh, D. S. Rokhsar, A new spiralian phylogeny places the enigmatic arrow worms among gnathiferans. *Curr. Biol.* **29**, 312–318.e3 (2019).
27. X.-X. Shen, C. T. Hittinger, A. Rokas, Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 1–10 (2017).
28. K. Strimmer, A. Von Haeseler, Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6815–6819 (1997).
29. Z. Yang, L. Zhang, J. Hu, J. Wang, S. Wang, The evo-devo of molluscs: Insights from a genomic perspective. *Evol. Dev.* **22**, e12336 (2020).
30. K. M. Kocot *et al.*, Phylogenomics of Lophotrochozoa with consideration of systematic error. *Syst. Biol.* **66**, 256–282 (2017).
31. N. M. Hallinan, D. R. Lindberg, Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol. Evol.* **3**, 1150–1163 (2011).
32. T. R. Gregory Animal Genome Size Database (2022). <http://www.genomesize.com>. Accessed 3 November 2022.
33. S. Wang *et al.*, Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **1**, 1–12 (2017).
34. C. B. Albertin *et al.*, Genome and transcriptome mechanisms driving cephalopod evolution. *Nat. Commun.* **13**, 1–14 (2022).
35. J. Lv, P. Havlak, N. H. Putnam, Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC Bioinformatics* **12**, 1–12 (2011).
36. O. Simakov *et al.*, Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).
37. M. Lynch, J. S. Conery, The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
38. S. Feng *et al.*, Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* **185**, 1646–1660.e18 (2022).
39. E. Parey *et al.*, Genome structures resolve the early diversification of teleost fishes. *Science* **379**, 572–575 (2023).
40. A. Som, Causes, consequences and solutions of phylogenetic incongruence. *Brief. Bioinformatics* **16**, 536–548 (2015).
41. P. O. Mulhair, C. G. McCarthy, K. Siu-Ting, C. J. Creevey, M. J. J. C. B. O'Connell, Filtering artifactual signal increases support for Xenacoelomorpha and Ambulacraria sister relationship in the animal tree of life. *Curr. Biol.* **32**, 5180–5188 (2022).
42. G. Tan *et al.*, Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* **64**, 778–791 (2015).
43. X. Guo *et al.*, Chloranthus genome provides insights into the early diversification of angiosperms. *Nat. Commun.* **12**, 1–14 (2021).
44. Z. Yang, PAML: A program package for phylogenetic analysis by maximum likelihood. *Comp. Appl. Biosci.* **13**, 555–556 (1997).
45. Y. Yang *et al.*, Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat. Plants* **6**, 215–222 (2020).
46. J. C. J. E. Oliver, Microevolutionary processes generate phylogenomic discordance at ancient divergences. *Evolution* **67**, 1823–1830 (2013).
47. M. G. Mángano, A. L. Buatois, Decoupling of body-plan diversification and ecological structuring during the Ediacaran-Cambrian transition: Evolutionary and geobiological feedbacks. *Proc. R Soc. B Biol. Sci.* **281**, 20140038 (2014).
48. A. Y. Zhuravlev, A. R. Wood, The two phases of the Cambrian Explosion. *Sci. Rep.* **8**, 1–10 (2018).
49. N. Butterfield, T. J. G. Harvey, Small carbonaceous fossils (SCFs): A new measure of early Paleozoic paleobiology. *Geology* **40**, 71–74 (2012).
50. S. C. Peng, L. E. Babcock, P. Ahlberg, "The Cambrian period" in *Geologic Time Scale*, F. M. Gradstein, J. M. Ogg, M. D. Schmitz, G. M. Ogg, Eds. (Elsevier, 2020), pp. 565–629.
51. A. Kouchinsky *et al.*, Terreneuvian stratigraphy and faunas from the Anabar Uplift, Siberia. *Acta Palaeontol. Pol.* **62**, 311–440 (2017).
52. S. Bengtson, The cap-shaped Cambrian fossil *Maikhanella* and the relationship between coelocleritophorans and molluscs. *Lethaia* **25**, 401–420 (1992).
53. Y. Pang *et al.*, Morphometric analysis of stem-group mollusks from the northern Yangtze Craton, China. *J. Paleontol.* **96**, 1024–1036 (2022).
54. J. Vinther, L. Parry, D. E. Briggs, P. Van Roy, Ancestral morphology of crown-group molluscs revealed by a new Ordovician stem aculiferan. *Nature* **542**, 471–474 (2017).
55. A. Kouchinsky, S. Bengtson, S. Clausen, M. J. Vendrasco, An early Cambrian fauna of skeletal fossils from the Emyaksin Formation, northern Siberia. *Acta Palaeontol. Pol.* **60**, 421–512 (2013).
56. M. J. Vendrasco, A. V. Kouchinsky, S. M. Porter, C. Z. Fernandez, Phylogeny and escalation in *Mellopegma* and other Cambrian molluscs. *Palaeontol. Electronica* **14**, 1–44 (2011).
57. J. S. Peel, Pseudomyona from the Cambrian of North Greenland (Laurentia) and the early evolution of bivalved molluscs. *Bull. Geosci.* **96**, 195–215 (2021).
58. D. Esquerré *et al.*, Rapid radiation and rampant reticulation: Phylogenomics of South American Liolaemus lizards. *Syst. Biol.* **71**, 286–300 (2022).
59. L. Na, Á. T. Kocsis, Q. Li, W. J. P. Kiessling, Coupling of geographic range and provincialism in Cambrian marine invertebrates. *Paleobiology* **49**, 284–295 (2022).
60. N. M. Koch *et al.*, Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *Elife* **11**, e72460 (2022).
61. P. J. Wagner, Patterns of morphologic diversification among the Rostroconchia. *Paleobiology* **23**, 115–150 (1997).
62. J. S. Peel, Scaphopodization in Palaeozoic molluscs. *Palaeontology* **49**, 1357–1364 (2006).
63. E. Yochelson, Carboniferous Scaphopoda (Mollusca) and non-scapopods from Scotland. *Scottish J. Geol.* **47**, 67–79 (2011).
64. R. J. Gentile, A new species of Dentalium from the Pennsylvanian of eastern Kansas. *J. Paleontol.* **48**, 1213–1216 (1974).
65. L. S. de Souza, C. H. S. Caetano, Morphometry of the shell in Scaphopoda (Mollusca): A tool for the discrimination of taxa. *Mar. Biol. Assoc. UK* **100**, 1271–1282 (2020).
66. S. V. Edwards, Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1–19 (2009).
67. J. Douglas, C. L. Jiménez-Silva, R. Bouckaert, StarBast3: Adaptive parallelized Bayesian inference under the multispecies coalescent. *Syst. Biol.* **71**, 901–916 (2022).
68. J. C. Pearson, D. Lemons, W. McGinnis, Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893–904 (2005).
69. M. A. Biscotti, A. Canapa, M. Forconi, M. Barucca, Hox and ParaHox genes: A review on molluscs. *Genesis* **52**, 935–945 (2014).
70. O. Simakov *et al.*, Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526–531 (2013).
71. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
72. N. C. Durand *et al.*, Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
73. J. T. Robinson *et al.*, Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256–258.e1 (2018).
74. J. M. Abujin, J. C. Pichel, T. F. Pena, J. Amigo, BigBWA: Approaching the Burrows-Wheeler aligner to Big Data technologies. *Bioinformatics* **31**, 4003–4005 (2015).
75. R. Challis, E. Richards, J. Rajan, G. Cochrane, M. Blaxter, BlobToolKit: interactive quality assessment of genome assemblies. *G3 Genes Genomes Genetics* **10**, 1361–1374 (2020).
76. Z. Ma *et al.*, High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. *Nat. Genet.* **53**, 1385–1391 (2021).
77. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
78. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
79. B. J. Haas *et al.*, Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, 1–22 (2008).
80. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
81. P. J. Krug *et al.*, Phylogenomic resolution of the root of Panpulmonata, a hyperdiverse radiation of gastropods: New insight into the evolution of air breathing. *Proc. R Soc. B* **289**, 20211855 (2022).
82. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
83. A. Di Franco, R. Poujol, D. Baurain, H. Philippe, Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol. Biol.* **19**, 1–17 (2019).
84. A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 1–21 (2010).
85. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
86. K. M. Kocot, M. R. Citarella, L. L. Moroz, K. M. Halanych, PhyloTreePruner: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinform. Online* **9**, 429–435 (2013).
87. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
88. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
89. S. Kalyaanamoorthy, B. Q. Minh, T. K. Wong, A. Von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
90. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
91. N. Lartillot, "PhyloBayes: Bayesian phylogenetics using site-heterogeneous models" in *Phylogenetics in the Genomic Era*, C. Scornavacca, F. Delsuc, N. Galtier, Eds. (2020), pp. 1.5:1–1.5:16.
92. M. Rabiee, E. Sayyari, S. Mirarab, Multi-allele species reconstruction using ASTRAL. *Mol. Phylogenet. Evol.* **130**, 286–296 (2019).
93. E. Sayyari, J. B. Whitfield, S. Mirarab, DiscoVista: Interpretable visualizations of gene tree discordance. *Mol. Phylogenet. Evol.* **122**, 110–115 (2018).
94. J. L. Steenwyk, T. J. Buida III, Y. Li, X.-X. Shen, A. Rokas, ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, e3001007 (2020).
95. Z. Huang *et al.*, Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving "living fossil". *Mol. Ecol. Resources* **22**, 15–27 (2022).
96. Y. Guo *et al.*, A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *Gigascience* **8**, giz124 (2019).
97. L. Liu, L. Yu, Phybase: An R package for species tree analysis. *Bioinformatics* **26**, 962–963 (2010).
98. K. Wang *et al.*, Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun. Biol.* **1**, 1–9 (2018).
99. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
100. J. Keilwagen, F. Hartung, J. Grau, "GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data" in *Methods in Molecular Biology*, M. Kollmar, Ed. (Humana Press, Clifton, N.J. 2019), vol. **1962**, pp. 161–177.
101. H. Song *et al.*, *Pictodentalium venedei*. NCBI BioProject. <https://www.ncbi.nlm.nih.gov/bioproject/903467>. Deposited 20 November 2022.
102. H. Song *et al.*, Genome sequencing of *Siphonodentalium dalli* (Mollusca, Scaphopoda). NCBI BioProject. <https://www.ncbi.nlm.nih.gov/bioproject/916950>. Deposited 30 December 2022.
103. H. Song *et al.*, Annotation of two scaphopod genomes. Figshare. <https://figshare.com/s/b9a58037afb9cd3b60d5>. Deposited 31 December 2022.
104. H. Song *et al.*, Phylogenetic placement of Scaphopoda. Figshare. <https://doi.org/10.6084/m9.figshare.22758134.v1>. Deposited 3 May 2023.