

Common origin of sterol biosynthesis points to a feeding strategy shift in Neoproterozoic animals

Received: 6 March 2023

Accepted: 13 November 2023

Published online: 01 December 2023

 Check for updatesT. Brunoir¹, C. Mulligan¹, A. Sistiaga², K. M. Vuu³, P. M. Shih^{3,4}, S. S. O'Reilly⁵, R. E. Summons⁶ & D. A. Gold¹✉

Steranes preserved in sedimentary rocks serve as molecular fossils, which are thought to record the expansion of eukaryote life through the Neoproterozoic Era (~1000–541 Ma). Scientists hypothesize that ancient C₂₇ steranes originated from cholesterol, the major sterol produced by living red algae and animals. Similarly, C₂₈ and C₂₉ steranes are thought to be derived from the sterols of prehistoric fungi, green algae, and other microbial eukaryotes. However, recent work on annelid worms—an advanced group of eumetazoan animals—shows that they are also capable of producing C₂₈ and C₂₉ sterols. In this paper, we explore the evolutionary history of the *24-C sterol methyltransferase (smt)* gene in animals, which is required to make C₂₈₊ sterols. We find evidence that the *smt* gene was vertically inherited through animals, suggesting early eumetazoans were capable of C₂₈₊ sterol synthesis. Our molecular clock of the animal *smt* gene demonstrates that its diversification coincides with the rise of C₂₈ and C₂₉ steranes in the Neoproterozoic. This study supports the hypothesis that early eumetazoans were capable of making C₂₈₊ sterols and that many animal lineages independently abandoned its biosynthesis around the end-Neoproterozoic, coinciding with the rise of abundant eukaryotic prey.

Organic compounds preserved in rocks—known as molecular fossils or biomarkers—offer a unique window into the early evolution of life. Compared to other biological molecules, such as nucleic acids and proteins, lipids are particularly resistant to degradation, with structural features that can be preserved in the geologic record for hundreds of millions, potentially billions, of years. Despite the ever-present risks of contamination and diagenetic alteration^{1–3} the biomarker field is coalescing around best practices, and clear patterns are emerging⁴. Steranes, the diagenetic remains of sterol lipids found in eukaryotic cell membranes, have proven particularly informative in Neoproterozoic-age rocks (~1000–541 Ma), where animal fossils are vexingly scarce (Fig. 1). Sterols are present in all eukaryotes and perform essential

functions within the cell membrane. In eumetazoan animals and multicellular red algae, these functions are generally performed by the 27-carbon (C₂₇) sterol, cholesterol, while C₂₈ sterols are the dominant sterols in most fungi, and C₂₉ sterols are common in green algae and plants^{5,6}. These sterols are observed widely throughout the eukaryotic tree of life, suggesting their presence in the last common ancestor⁷. In the geologic record, early proto-steranes have been identified in the Barney Creek Formation from ~1640 Ma, and can be found in rocks until ~850 Ma^{8,9}. C₂₇ steranes—the diagenetic products of cholesterol—become abundant in Neoproterozoic-age rocks starting around 850 Ma, while C₂₈ and C₂₉ steranes become prominent in the interglacial period ~663–635 Ma^{4,10}. The first appearance of steranes is hypothesized to

¹Department of Earth and Planetary Sciences, University of California, Davis, Davis, CA, USA. ²Globe Institute, University of Copenhagen, Copenhagen, Denmark. ³Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ⁴Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. ⁵Department of Life Sciences, Atlantic Technological University, ATU Sligo, Ash Lane, Sligo, Ireland. ⁶Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: dgold@ucdavis.edu

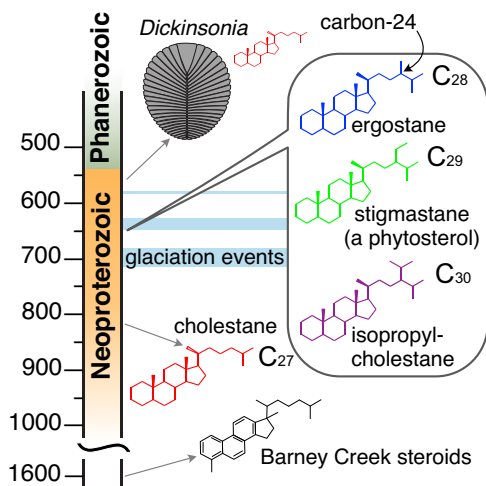


Fig. 1 | Overview of the Neoproterozoic sterane record. The geologic timescale is provided on the left, with dates in hundreds of millions of years. Arrows indicate the approximate time when various biomarkers become detectable above background thresholds. The site where sterols are normally methylated by the gene *24-C sterol methyltransferase* (carbon-24) is noted on ergostane with an arrow.

represent the ecological expansion of early eukaryotes (possibly red algae), with fungi and green algae expanding substantially around 663 Ma^{4,10}. The C₃₀ sterane isopropylcholestane also occurs around this time and could represent a biomarker for sponges or rhizarian protists^{11–13}. Recently, steranes and bacterial triterpenoids have also been used to taxonomically constrain enigmatic fossils from the Neoproterozoic, including the assignment of *Beltanelliformis* as a colonial cyanobacterium and *Dickinsonia* as an animal^{14–16}. Taken together, these steranes paint a picture of increasing eukaryote diversity leading up to the Cambrian (~541–485 Ma) radiation of fossils.

Despite recent advances, the potential to taxonomically constrain various biomarkers is both complex and understudied. For example, there are many living groups of eukaryotes besides green algae that produce C₂₉ sterols whose ancestors could have been responsible for Neoproterozoic steranes—including various fungi, choanoflagellates, brown algae, and ichthyosporeans¹⁴. Thus the linking of C₂₉ steranes in the rock record to prehistoric green algae would be premature. The same can be said of all major steranes in the Neoproterozoic. Comparative genomics can help resolve this uncertainty, as it offers a powerful tool for identifying organisms capable of synthesizing various lipids, and determining when in Earth's history they evolved such abilities¹⁷. In eukaryotes, the gene *24-C sterol methyltransferase* (*smt*) encodes an enzyme that adds methyl groups to carbon-24 of the sterol side chain (signified as C-24; see Fig. 1) and is responsible for the production of C₂₈₊ sterols. With increased genomic sampling the number of candidate *smt* genes has been expanding. This study was initiated based on a putative *smt* from the annelid *Capitella teleta*^{2,18}. Unlike sea sponges, annelid worms are part of the Eumetazoa, a clade of animals whose cell membranes are generally dominated by cholesterol, and who are thought to have lost the ability to make C₂₈₊ sterols^{7,18}. Since we began the study of *smt* in *C. Teleta*, another study by Michellod et al. has demonstrated the synthesis of C_{28/29} sterols in a different annelid, *Olavius algarvensis*, and reported *smt* genes in many animals¹⁹. The goal of this study is to determine the veracity of the *C. teleta smt* and reconstruct its evolutionary history. If *smt* genes have been conserved in annelid worms from the earliest animals, then the history of C₂₈₊ sterol biosynthesis in animals may be far more complex than currently appreciated.

In this work, we show that *smt* genes responsible for C₂₈₊ sterol biosynthesis are present in diverse annelid worms, as well as several other eumetazoan clades. Using yeast gene rescue experiments, we

demonstrate that the *C. teleta* SMT protein is capable of synthesizing C₂₈ sterols, providing evidence that SMT function has been conserved through annelid evolution. Careful vetting of the animal *smt* tree supports vertical inheritance of the gene from a common ancestor, as opposed to inheritance via horizontal gene transfer¹⁹. Finally, we produce molecular clocks to demonstrate that a vertically inherited SMT was spreading through eumetazoans during the Neoproterozoic, with most lineages losing the gene around this time. Our work suggests that the first eumetazoans were capable of C₂₈₊ sterol biosynthesis and that many lineages independently abandoned the biosynthesis of complex sterols concurrent with the rise of their sterol-rich prey, as evidenced by a rise in algal biomarkers during this time.

Results

Sterol methyltransferases are found across the annelids

Annelids are an ecologically diverse phylum containing ~20,000 described species, yet they are severely undersampled in genetic sequence databases such as the National Center for Biotechnology Information's (NCBI) Genbank. To determine how prevalent *smt* genes are in annelids, we queried a number of unannotated genomes and transcriptomes for candidate genes (see Methods). We ultimately recovered 27 *smt* homologs from 20 species of annelids. These putative genes are found across major clades, and their translated proteins contain the conserved domains expected in functional sterol methyltransferases (Fig. 2). In many cases, we discovered multiple *smt* genes in the same species, indicating multiple rounds of gene duplications within the group. Annelid SMT proteins form a well-supported clade in our tree-building analysis, suggesting they are derived from a common ancestor (Supplementary Figs. 1–3). This demonstrates that sterol methyltransferases are indeed present across diverse annelid worms.

We used I-TASSER²⁰ to model several annelid SMT proteins, comparing their tertiary structure to each other and to better-studied SMTs from yeast and plants (Fig. 3). In particular, we analyzed the SMT from *C. teleta* (Fig. 3A), as well as two SMTs recovered from the model annelid *Platynereis dumerilii* (Fig. 3B, C). These were compared to an SMT from the fungus *Saccharomyces cerevisiae* (also known as ERG6), which is required for the synthesis of the C₂₈ sterol ergosterol (Fig. 3D), and the SMT2 protein in the plant *Arabidopsis thaliana*, a bifunctional enzyme that can generate C₂₈ and C₂₉ sterols (Fig. 3H)²¹. Root-mean-square deviation (RMSD) of atomic positions suggests that annelid SMTs are highly similar to both ERG6 and SMT2, and in fact, are more similar to both proteins than ERG6 and SMT2 are to each other (RMSD = 2.412). The methyltransferase domains—the region where sterol binding occurs²²—demonstrate high conservation across proteins, while the C-terminal domain shows greater variability (3E-G,I-K). The *C. teleta* SMT and one of the two *P. dumerilii* proteins appear more structurally similar to the bifunctional SMT2 than ERG6, as suggested by their lower RMSD scores, and is particularly noticeable in the alpha helices in the C-terminal domain. This supports recent results by Michellod et al. demonstrating that SMTs of the annelids *Olavius* and *Inanidrilus* can bifunctionally generate C₂₈ and C₂₉ sterols¹⁹, and suggests this ability is commonplace across annelids.

Functional analysis of annelid SMTs demonstrates their ability to methylate sterols

After modeling the *Capitella teleta* SMT, we assessed whether or not it is capable of methylating sterols at the C-24 position. To test this we performed a rescue experiment, introducing the *C. teleta smt* into a strain of *S. cerevisiae* yeast lacking the wild-type gene (commonly known as ERG6). ERG6 yeasts are viable but cannot produce ergosterol and have a slower rate of growth than their wild-type counterparts²³. We found that ERG6 *S. cerevisiae* is capable of producing ergosterol with the addition of the *C. teleta* gene (Fig. 4). This functional analysis complements Michellod et al., demonstrating that multiple annelid species can methylate sterols¹⁹.

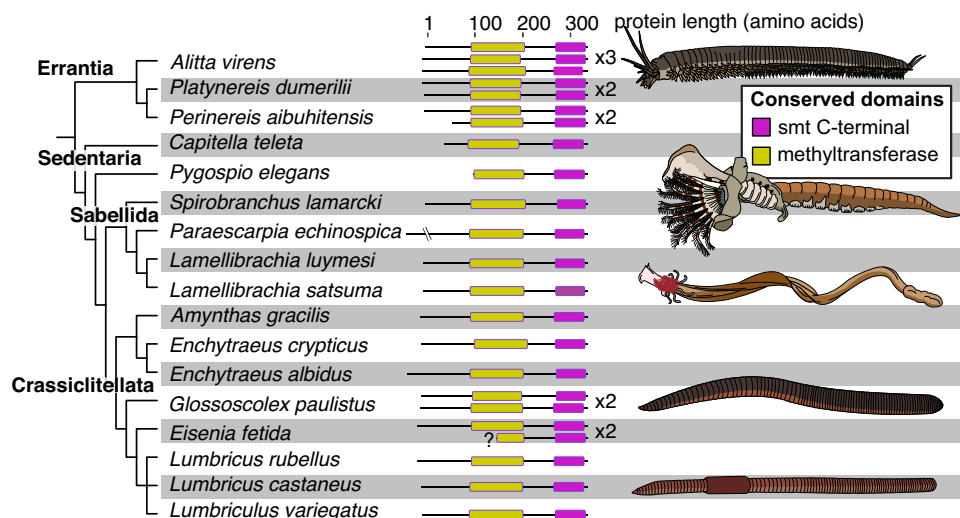


Fig. 2 | Distribution and structure of annelid SMT proteins. This figure is restricted to annelid SMT proteins containing both conserved domains (i.e. excluding partial sequences). (Left) Phylogeny of annelids in our study with

recovered full-length SMT proteins. (Right) Image displaying the number and size of SMT proteins. The position and length of conserved domains is visualized with colored boxes. Source data are provided as a Source Data file.

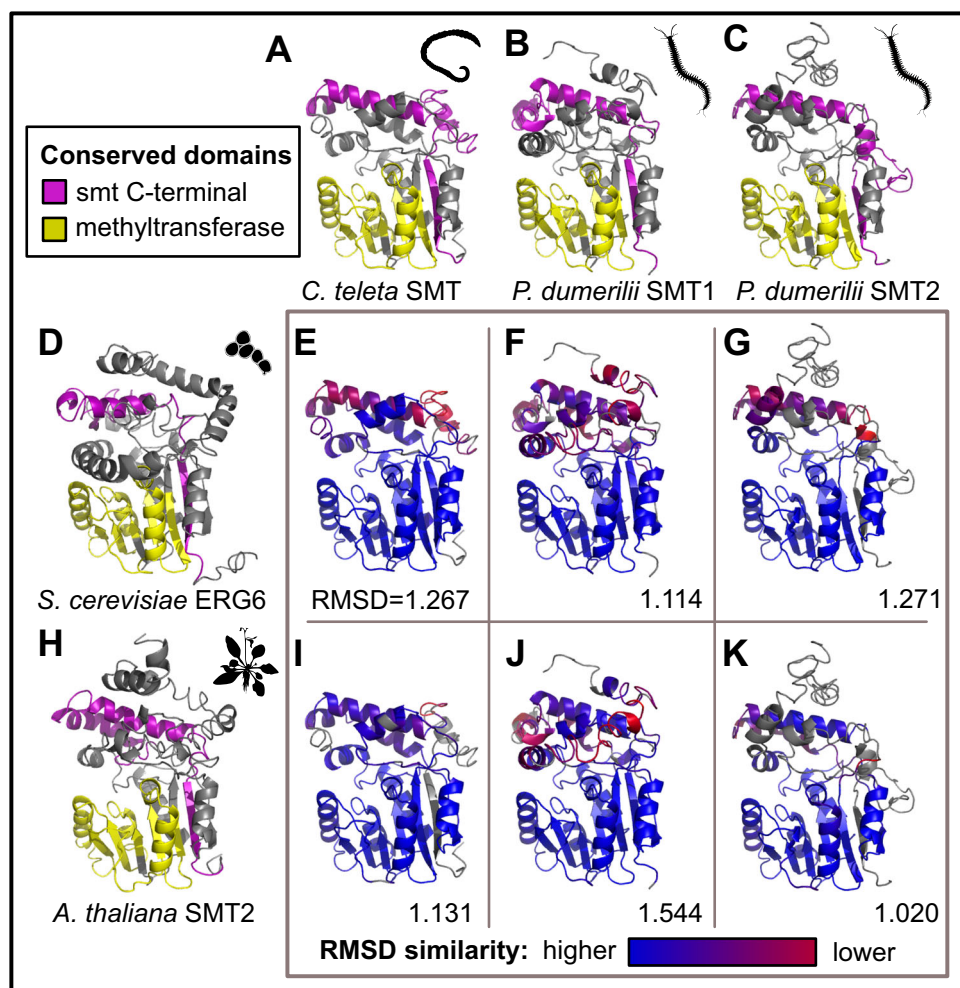


Fig. 3 | Modeling of select annelid SMT proteins. **A–C** best-scoring models for the SMT protein in *C. teleta* and two SMTs in *P. dumerilii*, with conserved domains highlighted. **D** Model of *S. cerevisiae* ERG6. **E–G** Alignment of the three above annelid proteins against ERG6, colored by RMSD values. The overall RMSD score for each comparison is shown in the lower right corner. **H** Model of *A. thaliana* SMT2

protein. **I–K** Alignment of the three above annelid proteins against *A. thaliana* SMT2, colored by their RMSD values. The overall RMSD score for each comparison is shown in the lower right corner. Protein models were generated using the I-TASSER server. Source data are provided as a Source Data file.

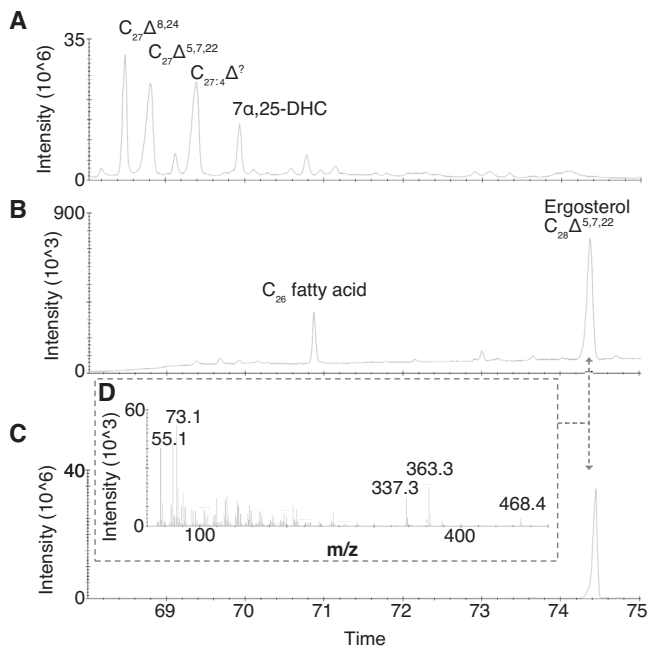


Fig. 4 | Production of ergosterol is rescued in yeast when ERG6 is replaced by its annelid homolog. **A** Partial total gas chromatogram from an ERG6 knockout yeast without transformation of the *C. teleta smt* gene. **B** Partial total ion chromatogram from an ERG6 knockout yeast with transformation of *C. teleta smt*. **C** An ergosterol standard mix. **D** The ergosterol mass spectrum from comparable peaks in **(B)** and **(C)**, with characteristic fragment ions at m/z 468, 363, 337 for trimethylsilyl-derived samples. Source data are provided as a Source Data file.

Phylogenetic analysis of animal SMTs supports vertical inheritance

In addition to annelids, we recovered SMT proteins from several other animal clades. Many of our SMTs came from shotgun transcript sequencing projects, which can easily be contaminated by an animal's diet and symbionts. To minimize likely contaminants in our final dataset, we made an initial phylogenetic tree and rejected any animal sequences that did not form a clade with one or more species from the same phylum (see Methods; Supplementary Figure 1). While this methodology may have inadvertently pruned real instances of horizontal gene transfer from our dataset, it also removed many genes we were able to independently confirm were contaminants. For example, some of the rejected SMTs come from animals that have publicly available genomes. In each case we failed to find the putative *smt* transcript in the relevant genome (Figure S1). Our analysis of the data in Michellod et al. suggests contamination was a problem in their phylogenetic tree as well. The molluscs provide a good case study: Michellod et al. report 13 mollusc *smt* transcripts, which evolved through a minimum of 11 separate horizontal gene transfer events (see Fig S26 from their paper). Three of the molluscs in their tree have genomes on NCBI (*Biomphalaria pfeifferi*, *Crassostrea hongkongensis*, and *Pecten maximus*); none of their genomes contains evidence of an *smt* (see Methods). An additional two species do not have genome assemblies on NCBI (*Haliotis tuberculata* and *Elysia cornigera*), but other members of the same genus do. Again, we found no evidence for *smt* genes in these genomes. This strongly suggests that many, perhaps all, of the putative mollusc *smt* transcripts are contaminants from shotgun sequencing projects. No mollusc *smt* passed vetting in our analysis. It is possible that future work will demonstrate that some of the *smt* genes we rejected represent true animal sequences, but false positives caused by contamination are clearly a major issue. Our results challenge the hypothesis that rampant horizontal gene transfer explains the presence of *smt* genes in annelids and other eumetazoans¹⁹.

Following our vetting process we were left with *smt* genes from three clades of eumetazoans—annelids, rotifers and stony corals. Nematode worms also have an SMT-like protein¹⁸, but we did not include it in our analysis because the gene did not cluster with other eukaryote SMTs in our phylogeny, and because the protein is known to catalyze a C-4 methylation step that is distinct from the C-24 methylation seen in genuine SMTs²⁴. Rotifers are a group of near-microscopic animals that are distant relatives to annelids. A recent project assembling 31 rotifer genomes resulted in a large number of *smt* genes from the genera *Adineta*, *Rotaria*, and *Didymodactylos*²⁵. Regarding stony corals (clade Scleractinia), one SMT-like protein has been annotated in the *Orbicella faveolata* genome (NCBI Accession: [XP_020604180.1](https://ncbi.nlm.nih.gov/nuccore/XP_020604180.1)), but it lacks the C-terminal domain found in functional SMTs. However, we recovered multiple SMTs from other coral transcriptomes that retain both domains. We were able to map one of these genes (NCBI Accession: [FX438716.1](https://ncbi.nlm.nih.gov/nuccore/FX438716.1)) to the genome of the coral *Porites australiensis* (Supplementary Figure 4). A second candidate SMT from *P. australiensis* did not map to the genome, and shows high sequence similarity to the coral symbiont *Symbiodinium*, again demonstrating that our methods can distinguish between authentic and contaminating sequences.

After removing questionable sequences from our dataset, all animal SMT proteins, except rotifers, formed a monophyletic clade (Supplementary Fig. 2). While rotifers did not cluster with the rest of the animals, the nodes separating these two groups are poorly supported. Following gene tree / species tree reconciliation (a process where nodes with low statistical support in a gene tree are rearranged to parsimoniously reflect the known species tree), all animal SMTs formed a single clade (Supplementary Figure 3). We therefore find no compelling support for horizontal gene transfer in animal SMTs, where we would expect to find a clade of animal SMTs that clusters with non-animals with strong statistical support.

Notably, the animal SMT tree does not replicate the species tree exactly. Instead, our final tree (Figure S3) prefers two clades: one containing sea sponges and corals, and another containing sea sponges, rotifers, and annelids. Our result is driven by strong statistical support for a clade including Heteroscleromorph sponges and cnidarians to the exclusion of other sponges. It is possible that better taxon sampling of sponges will ultimately demonstrate that this duplication inference is driven by an error in the gene tree. Alternatively, this could suggest that the sponge gene duplication that was first hypothesized in Gold et al. (2016) is actually more ancient than anticipated, predating the divergence of living animals. Given the data currently available, we conclude that the best interpretation of animal SMTs is vertical inheritance from a common ancestor, with a gene duplication event occurring before sea sponges diverged from the other living animals.

A molecular clock suggests eumetazoan SMTs diversified in the Neoproterozoic

To test whether the diversification of eumetazoan SMTs coincides with the Neoproterozoic biomarker record, we generated a gene-centered molecular clock using 107 vetted SMT sequences and nine fossil calibrations (see Supplementary Methods for details and justification of calibrations). The results are summarized in Fig. 5, with a more detailed output provided in Supplementary Figure 5. As this clock is based on a single gene, the error bars are large and the exact dates should be interpreted with caution. However, our results are consistent with multigene, species-level molecular clocks, which place the origin of eukaryotes between -1800 and 1600 Ma and animals between -828 and 572 Ma^{26–28}. As mentioned earlier, many annelids and rotifers have multiple SMTs. Our tree suggests this is primarily due to lineage-specific gene duplication events (marked by circles in Fig. 5), which have occurred throughout the Phanerozoic (<541 Ma). Our results demonstrate that the two eumetazoan *smt* genes overlap with the appearance of C_{28} and C_{29} steranes in the fossil record -663–635 Ma. We re-ran the molecular clock using an alternate topology where all

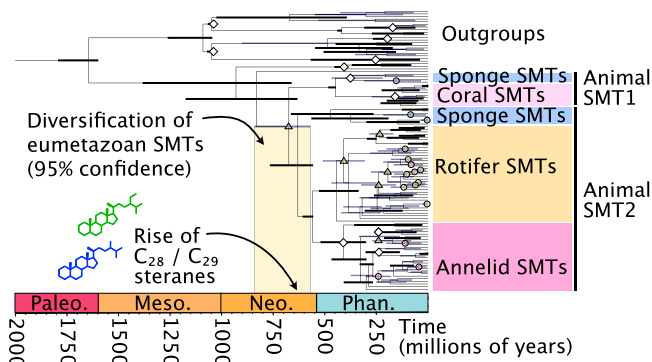


Fig. 5 | The molecular clock of animal SMT proteins. This SMT-centered molecular clock was generated using BEAST. White diamonds represent fossil calibrations, which are described in detail in the Supplementary Methods. Putative gene duplication events in the animals have been labeled at relevant nodes with colored shapes: circles represent gene duplication events that are specific to one species in our study; triangles represent events that include multiple species. White diamonds represent fossil calibrations. Phan. = Phanerozoic; Neo. = Neoproterozoic; Meso. = Mesoproterozoic; Paleo. = Paleoproterozoic. Source data are provided as a Source Data file.

sponge sequences are monophyletic (i.e. no ancestral gene duplication). The results of this alternate topology still suggest the diversification of eumetazoan SMTs coincides with the rise of $C_{28/29}$ steranes (Supplementary Fig. 6). We can therefore conclude that eumetazoans had functional SMTs concurrent with, and likely prior to, the rise of complex steranes in the Neoproterozoic.

Discussion

Our research suggests that the *smt* gene necessary to synthesize complex sterols existed in the ancestor of Eumetazoa, and was retained in some lineages long after their diversification in the Cambrian. These results contradict the hypothesis that *smt* genes evolved in eumetazoans as a result of horizontal gene transfer¹⁹. If we are correct, then the observation that most living eumetazoans lack *smt* genes is a function of extensive gene loss¹⁸. The retention of *smt* genes in annelids via vertical inheritance necessitates its loss in at least seven major animal groups (Fig. 6). The actual number of losses is likely an order of magnitude higher, given the large number of coral and annelid genomes that lack SMT proteins, as well as the possible loss of SMTs in the minor animal phyla, which are poorly represented in genetic sequence databases. The presence of one or more *smt* genes in the ancestor of Eumetazoa raises important questions about the sterols biosynthesized by the earliest animals. What lipid(s) did the first animal SMT—or, if our scenario is correct, the first pair of SMTs—synthesize? Was it bifunctional like those seen in some sea sponges and annelids? Answers to these questions will help determine whether early eumetazoans could biosynthesize some of the exotic steranes found in sedimentary rocks.

Additionally, why the *smt* gene has been retained and independently duplicated in many eumetazoan lineages remains an open question. For rotifers, the large number of *smt* duplicates may lay in part with their unusual reproductive strategy; the species in our dataset reproduce asexually and exhibit tetraploidy²⁹. Consequently, their genomes are more like plants than other animals, and like plants, duplicate SMT proteins are likely to be partially redundant³⁰. The presence of multiple SMTs in some annelids is harder to explain. For example, two SMT copies have been retained in hesionoid worms (*Alitta virens*; *Perinereis aibuhitensis*, and; *Platynereis dumerilii*) for over 300 million years, and must therefore confer some function. There is little work on the lipids of annelids, and fewer where the environment and diet is well constrained. Most annelids studied have cholesterol as their

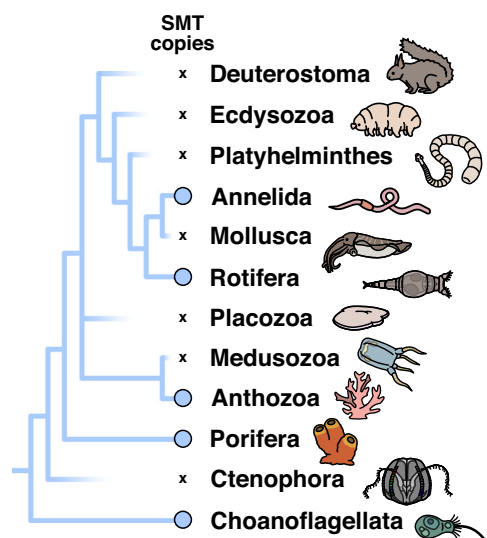


Fig. 6 | Summary of *smt* gene loss in the animals. A simplified animal phylogeny, with the hypothesized presence of *smt* visualized with blue lines. This tree provides a conservative estimate of the number of *smt* losses, as it excludes many understudied animal phyla that may also lack the protein.

dominant sterol, but many contain complex C_{28+} sterols as well^{31–34}. In *C. teleta*, the abundance of sterols can vary dramatically depending on diet and life stage, in some cases demonstrating higher percentages of C_{28+} sterols in their bodies than exists in their food³⁵. This variation suggests a strong biological control on endogenous sterol uptake. Therefore, annelid SMTs may play specific roles based on the animal's development and/or environment. While there is a general relationship between SMT copy number and the type of sterols synthesized, we know of enough lineages with bifunctional enzymes to prevent any a priori predictions about the sterols produced by these animals^{19,30,36,37}. The selective advantage of retaining SMTs in some eumetazoan lineages, when the vast majority have abandoned the protein, will require further research on their use and function in specific species.

Regardless of the reason why *smt* genes were retained and expanded in some animals, our results suggest that the biosynthesis of complex sterols is an ancestral and ancient trait. This work does not discount fossil evidence of C_{27} -dominated early animals, such as dickinsoniamorphs, but when C_{28+} steranes are found in eumetazoan fossils, the possibility these organisms were synthesizing higher sterols should be considered^{15,16}. In addition, our conclusions have distinct implications for the evolution of animal sterols and feeding strategies. In our scenario of vertical inheritance, multiple animal groups lost the ability to synthesize higher sterols across the Neoproterozoic Cambrian transition, which we hypothesize was caused by environmental changes. With the radiation of microbial eukaryotes after the Sturtian glaciation—as documented by the rise in $C_{28/29}$ steranes—early animals had a novel and abundant source for sterols¹⁰. The Neoproterozoic was also a time of fluctuating oxygen levels, and it has been demonstrated that some eukaryotes will switch to exogenous sterols under anaerobic conditions³⁸. As ocean chemistry shifted and feeding strategies diversified, many Precambrian animal groups independently abandoned sterol modification (or sterol biosynthesis altogether). In the alternative scenario where annelids, rotifers, and cnidarians each received their *smt* genes through horizontal gene transfer, our molecular clock suggests these events occurred in the Phanerozoic (Fig. 5). The presence of *smt* genes in these animals would therefore have no relevance to our interpretation of the biomarker record or Precambrian evolution. Adjudicating between these two competing hypotheses is therefore critical to interpreting the significance of eumetazoan SMTs. If animal SMTs have been vertically inherited, then

understanding their evolutionary history and usage will reveal insights into the origins of animal feeding strategies.

Methods

This research complies with all relevant ethical regulations of the University of California.

Collection of putative animal SMTs

The ERG6 protein from *Saccharomyces cerevisiae* (NCBI accession: P25087.4) was used as a query for all database searches. We queried the NCBI Transcriptome Shotgun Archive (TSA) using tBLASTn, restricting our analysis to animals (taxid:33208). A similar search was performed on the non-redundant protein database using BLASTp. Transcripts from the TSA search were converted into proteins using the TransDecoder (v5.5.0) program packaged with Trinity³⁹. Sequences were then separated by species, and redundant proteins were removed using CD-HIT (v4.8.1) with a 90% similarity cutoff⁴⁰. Conserved domains were identified in the remaining proteins using the pfam_scan Perl script included in BioConda, which uses HMMER (v3.4) to compare proteins against the PFAM-A database^{41,42}. Methyltransferase and C-terminal domains were extracted independently using SAMTOOLS (v1.13) and aligned with MAAFT (v7.490); they were combined again with FASconCAT-G and cleaned with trimAl (v1.4)^{43,44}. These aligned sequences were used for further annotation and tree building. Relevant files and code can be found in folder 2 on GitHub.

Vetting of SMTs

Contamination was a major concern with data downloaded from the NCBI TSA database, so a thorough vetting process was used to find real SMT sequences for the analysis. Additional data was appended to sequence IDs to help interpret the data, including (1) NCBI's TaxIdentifier was used to add taxonomic information in the dataset, (2) top reciprocal BLASTp hit when the proteins were compared against the Uniprot Swissprot dataset, which was augmented with additional SMT sequences (XP_003387525.1, XP_004346937.1, ELU07827.1, CAF1633859.1). A selection of SMT and methyltransferase outgroups were chosen from across the eukaryotes and added to our SMT protein alignment. An initial phylogenetic tree was generated from the alignment using FastTree⁴⁵. We then went carefully through the tree to determine which genes were likely to be genuine and which were likely to be contaminants (results illustrated and annotated in Supplementary Fig. 1). When possible, we used BLASTn to query putative animal transcripts against their respective genomes. Transcripts that were not identifiable in genomes are annotated in Supplementary Figure 1, and were treated as contaminants. Sequences were only considered if two or more genera from the same phylum formed a monophyletic clade. Once clades were defined, individual sequences were removed from a clade for the following reasons: (1) the sequence was redundant, meaning it was highly similar to another sequence from the same species, or (2) the sequence came from a species in a different phylum than the clade, making it a probable contaminant. Following this process, animal SMT-like genes were restricted to the Rotifera, Annelida, Nematoda, Porifera and Cnidaria. These sequences were extracted from the original alignment, the outgroups were added back in, and a final tree was generated using IQTree (v1.6.12)⁴⁶, which is visualized in Supplementary Figure 2. In this tree, the nematode sequences fell outside of the SMT clade, but all other putative animal SMTs remained. This clade of SMTs (highlighted in Supplementary Figure 2) Was used for downstream species tree reconciliation and molecular clock analysis. Relevant files and code can be found in folder 2 on GitHub.

Vetting of putative mollusc sequences from Michellod et al.

We used the tBLASTn algorithm on NCBI, using yeast ERG6 (accession: NP_013706.1) as a query. The following genomes were searched as

databases: *Crassostrea hongkongensis* (GCA_015776775.1), *Biomphalaria pfeifferi* (GCA_030265305.1), *Pecten maximus* (GCF_902652985.1), *Elysia chlorotica* (GCA_003991915.1), *Elysia marginata* (GCA_019649035.1), *Haliotis laevigata* (GCA_008038995.1), *Haliotis cracherodii* (GCA_022045235.1), *Haliotis rufescens* (GCA_023055435.1), and *Haliotis rubra* (GCA_003918875.1). No significant matches were returned in any of these analyses. To verify that we could recover SMTs from genomes using tBLASTn, we also queried *Monosiga brevicollis* (GCA_000002865.1), *Amphimedon queenslandica* (GCA_000090795.2), and *Capitella teleta* (GCA_000328365.1). In each of these genomes, we recovered a matching genome contig. BLASTp was also used on the *Pecten maximus* genome assembly since the genome was sufficiently annotated to perform this additional analysis. The top hit in this search was a phosphoethanolamine N-methyltransferase-like protein, again suggesting no sterol methyltransferase proteins are present in this species.

Species tree reconciliation and molecular clock analysis

The species included in our SMT clade was downloaded from the NCBI Taxonomy Browser website. The tree was manually edited using Mesquite (v3.6.1)⁴⁷ to reflect updated taxonomic affinities based on references^{48–50}. The species tree and gene were passed to NOTUNG (v2.1.5), which rearranged poorly supported branches in the gene tree to provide the most parsimonious reconciliation with the species tree⁵¹. A molecular clock was generated from the reconciled tree using BEAST (v1.10.5)⁵². An LG amino acid substitution model was used with a 4-site gamma heterogeneity model. An uncorrelated relaxed clock with a lognormal distribution was chosen for the clock model. The tree prior followed the speciation: yule process. Fossil calibrations were modeled as lognormal priors, and are detailed in the Supplementary Methods. BEAST was run for 10,000,000 generations, with sampling every 1,000 states. A consensus tree was generated from the results using a 25% burnin and median node heights. We then re-ran the molecular clock analysis using a different starting tree. In this second analysis sponges are monophyletic, meaning there is no SMT duplication event at the origin of animals. All input files, including the XML used for the BEAST runs, are provided in folder 5 on GitHub.

Protein modeling

Protein structure and function predictions were performed on the I-TASSER server²⁰. Sequences submitted include the *Capitella teleta* SMT protein (accession: ELU07827.1), two SMTs translated from *Platynereis dumerilii* transcripts (accessions: HALR01229039.1; HALR01261698.1), ERG6 from *Saccharomyces cerevisiae* (accession: P25087.4), and SMT2 from *Arabidopsis thaliana* (Accession: NP_173458.1). The resulting Protein Data Bank models were visualized in open-source Pymol (v2.5.0). The results from I-TASSER and the Pymol code used to generate the figures are provided in folder 3 on GitHub.

Functional analysis

The *C. teleta smt* mRNA was codon optimized for yeast and integrated into a pPMS090 plasmid with chloramphenicol resistance and p15a origin of replication. The cassette incorporating the gene was flanked by 5' and 3' homology arms at the yeast ERG6 location and contained a LEU2 auxotroph marker. Target knockout of wild-type ERG6 in *S. cerevisiae* strain BY4742 (MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0) was achieved by restriction digest of the *C. teleta smt* and transforming the linearized DNA using Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA). Yeast colonies were selected on SC-LEU agar plates and grown at 30 °C for -1-2 weeks. The presence of the *C. teleta smt* at the ERG6 location was screened by yeast colony PCR and subsequently grown in SC-LEU liquid media at 30 °C. Relevant data, including the plasmid design, are provided in folder 4 on GitHub.

Lipid extraction

The complemented yeast cultures were extracted twice with dichloromethane (DCM)/methanol (MeOH) (9:1 v/v). After the addition of water, the DCM layers were combined, washed several times with H₂O, and dried over Na₂SO₂. The lipid extracts were then concentrated and transferred to a vial for derivatization (silylation) by reaction with N, O-Bistrifluoroacetamide (BSTFA) + 1% trimethylchlorosilane in pyridine (2 h at 70 °C). All glassware, aluminum foil, silica, quartz wool, and quartz sand were combusted at 500 °C for at least 12 hours to remove organic contamination, whereas metal tools were rinsed in MeOH and DCM.

GC-MS analysis of sterols

1 µL of silylated samples were analyzed by gas chromatography-mass spectrometry on an Agilent 5890 GC hyphenated to an Agilent 5975 C Mass Selective Detector. The GC inlet was operated in splitless mode and fitted with a J&W DB5-MS 60 m capillary column (0.25 mm inner diameter, 250 µm film thickness). The GC temperature program was 80 °C for 2 min, ramp at 3.5 °C min⁻¹ to 315 °C, and a final hold time of 31 min. The mass spectrometer was operated in electron impact ionization mode (70 eV), with a mass scan range from m/z 50 to 700. All solvents used were high-purity (OmniSolv), all aqueous solutions were cleaned with dichloromethane prior to use, and procedural blanks were run to monitor background contamination. MSD data is provided in folder 4 on GitHub. The two ergosterol spectra from Fig. 4B and C are illustrated in Supplementary Figure 7 for comparative purposes.

Mapping of putative SMTs to coral genomes

Putative *smt* transcripts from the corals *Acropora* and *Porites* were downloaded along with their respective genomes from NCBI. Transcripts were mapped to the genome using GMAP (v2019-09-12)⁵³. The only sequence to map to a genome was a single *smt* (Accession: [FX438716.1](https://.ncbi.nlm.nih.gov/nucl/FX438716.1)) in *P. australiensis*. To produce a gene model, the reverse complement was mapped to the genome. We then generated a genome-based coding region annotation file with TransDecoder³⁹. The input files and results of this analysis are available in folder S3 on GitHub.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this study, including scripts executed, input files, intermediate files, and output, are provided on GitHub at https://github.com/DavidGoldLab/2022_Annelid_SMTs. The accession numbers for all previously published genetic data can be found on GitHub in Supplementary Data file 1. The data generated in this study have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.10063989>⁵⁴. Source data are provided with this paper.

Code availability

All code used is provided on GitHub at https://github.com/DavidGoldLab/2022_Annelid_SMTs. The data generated in this study have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.10063989>⁵⁴.

References

- Brocks, J. J. Millimeter-scale concentration gradients of hydrocarbons in Archean shales: Live-oil escape or fingerprint of contamination? *Geochim. et. Cosmochim. Acta* **75**, 3196–3213 (2011).
- Gold, D. A., O'Reilly, S. S., Luo, G., Briggs, D. E. & Summons, R. E. Prospects for sterane preservation in sponge fossils from museum collections and the utility of sponge biomarkers for molecular clocks. *Bull. Peabody Mus. Nat. Hist.* **57**, 181–189 (2016).
- French, K. L. et al. Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proc. Natl Acad. Sci.* **112**, 5915–5920 (2015).
- Love, G. D. & Zumberge, J. A. *Emerging patterns in proterozoic lipid biomarker records*. (Cambridge University Press, 2021).
- Summons, R. E., Bradley, A. S., Jahnke, L. L. & Waldbauer, J. R. Steroids, triterpenoids and molecular oxygen. *Philos. Trans. R. Soc. B: Biol. Sci.* **361**, 951–968 (2006).
- Weete, J. D., Abril, M. & Blackwell, M. Phylogenetic distribution of fungal sterols. *PLoS one* **5**, e10899 (2010).
- Desmond, E. & Gribaldo, S. Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.* **1**, 364–381 (2009).
- Brocks, J. J. & Schaeffer, P. Okenane, a biomarker for purple sulfur bacteria (Chromatiaceae), and other new carotenoid derivatives from the 1640 Ma Barney Creek Formation. *Geochim. et. Cosmochim. Acta* **72**, 1396–1414 (2008).
- Brocks, J. J. et al. Lost world of complex life and the late rise of the eukaryotic crown. *Nature* **618**, 767–773 (2023).
- Brocks, J. J. et al. The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578 (2017).
- Gold, D. A. et al. Sterol and genomic analyses validate the sponge biomarker hypothesis. *Proc. Natl Acad. Sci.* **113**, 2684–2689 (2016).
- Nettersheim, B. J. et al. Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. *Nat. Ecol. Evol.* **3**, 577–581 (2019).
- Love, G. D. et al. Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718–721 (2009).
- Bobrovskiy, I., Hope, J. M., Krasnova, A., Ivantsov, A. & Brocks, J. J. Molecular fossils from organically preserved Ediacara biota reveal cyanobacterial origin for Beltanelliformis. *ENat. Ecol. Evol.* **2**, 437–440 (2018).
- Bobrovskiy, I. et al. Ancient steroids establish the Ediacaran fossil Dickinsonia as one of the earliest animals. *Science* **361**, 1246–1249 (2018).
- Bobrovskiy, I., Nagovitsyn, A., Hope, J. M., Luzhnaya, E. & Brocks, J. J. Guts, gut contents, and feeding strategies of Ediacaran animals. *Curr. Biol.* **32**, 5382–5389 (2022).
- Summons, R. E., Welander, P. V. & Gold, D. A. Lipid biomarkers: molecular tools for illuminating the history of microbial life. *Nat. Rev. Microbiol.* **20**, 174–185 (2022).
- Zhang, T. et al. Evolution of the Cholesterol Biosynthesis Pathway in Animals. *Mol. Biol. Evol.* **36**, 2548–2556 (2019).
- Michellod, D. et al. De novo phytosterol synthesis in animals. *Science* **380**, 520–526 (2023).
- Yang, J. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
- Husselstein, T., Gachotte, D., Desprez, T., Bard, M. & Benveniste, P. Transformation of *Saccharomyces cerevisiae* with a cDNA encoding a sterol C-methyltransferase from *Arabidopsis thaliana* results in the synthesis of 24-ethyl sterols. *FEBS Lett.* **381**, 87–92 (1996).
- Nes, W. D. et al. Active Site Mapping and Substrate Channeling in the Sterol Methyltransferase Pathway*. *J. Biol. Chem.* **277**, 42549–42556 (2002).
- Gaber, R. F., Copple, D. M., Kennedy, B. K., Vidal, M. & Bard, M. The yeast gene ERG6 is required for normal membrane function but is not essential for biosynthesis of the cell-cycle-sparking sterol. *Mol. Cell. Biol.* **9**, 3447–3456 (1989).
- Zhou, W. et al. A nematode sterol C4 α -methyltransferase catalyzes a new methylation reaction responsible for sterol diversity. *J. Lipid Res.* **61**, 192–204 (2020).
- Nowell, R. W. et al. Evolutionary dynamics of transposable elements in bdelloid rotifers. *Elife* **10**, e63194 (2021).
- Parfrey, L. W., Lahr, D. J., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci.* **108**, 13624–13629 (2011).

27. dos Reis, M. et al. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 2939–2950 (2015).
28. Strasser, J. F., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 1879 (2021).
29. Flot, J.-F. et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* **500**, 453–457 (2013).
30. Carland, F., Fujioka, S. & Nelson, T. The sterol methyltransferases SMT1, SMT2, and SMT3 influence *Arabidopsis* development through nonbrassinosteroid products. *Plant Physiol.* **153**, 741–756 (2010).
31. KOBAYASHI, M., NISHIZAWA, M., TODO, K. & MITSUHASHI, H. Marine sterols. I. Sterols of annelida, *Pseudopotamilla ocellata* Moore. *Chem. Pharm. Bull.* **21**, 323–328 (1973).
32. Sica, D. & Di Giacomo, G. Sterols from two marine sedimentary annelids. *Comp. Biochem. Physiol. Part B: Comp. Biochem.* **70**, 719–723 (1981).
33. Ballantine, J. A. et al. Marine sterols—VII. The sterol compositions of oceanic and coastal marine Annelida species. *Comp. Biochem. Physiol. Part B: Comp. Biochem.* **61**, 43–47 (1978).
34. Masaru, K. & Hiroshi, M. Marine sterols. V. 1 isolation and structure of ocellasterol, a new 27-norergostane-type sterol, from an annelida, *Pseudopotamilla ocellata*. *Steroids* **24**, 399–410 (1974).
35. Marsh, A. G., Harvey, H. R., Gremare, A. & Tenore, K. R. Dietary effects on oocyte yolk-composition in *Capitella* sp. I (Annelida: Polychaeta): fatty acids and sterols. *Mar. Biol.* **106**, 369–374 (1990).
36. Brown, M. O., Olagunju, B. O., Giner, J.-L. & Welander, P. V. Sterol methyltransferases in uncultured bacteria complicate eukaryotic biomarker interpretations. *Nat. Commun.* **14**, 1859 (2023).
37. Gold, D. A. et al. Lipidomics of the sea sponge *Amphimedon queenslandica* and implication for biomarker geochemistry. *Geobiology* **15**, 836–843 (2017).
38. Schneiter, R. Intracellular sterol transport in eukaryotes, a connection to mitochondrial function? *Biochimie* **89**, 255–259 (2007).
39. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
40. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
41. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
42. Grüning, B. et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. methods* **15**, 475–476 (2018).
43. Kück, P. & Meusemann, K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* **56**, 1115–1118 (2010).
44. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
45. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
46. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
47. Maddison, W. P. Mesquite: a modular system for evolutionary analysis. *Evolution* **62**, 1103–1118 (2008).
48. Plese, B. et al. Mitochondrial evolution in the Demospongiae (Porifera): Phylogeny, divergence time, and genome biology. *Mol. Phylogenet. Evol.* **155**, 107011 (2021).
49. Tilic, E., Stiller, J., Campos, E., Pleijel, F. & Rouse, G. W. Phylogenomics resolves ambiguous relationships within Aciculata (Errantia, Annelida). *Mol. Phylogenet. Evol.* **166**, 107339 (2022).
50. Weigert, A. & Bleidorn, C. Current status of annelid phylogeny. *Org. Diver. Evol.* **16**, 345–362 (2016).
51. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).
52. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
53. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
54. Gold, D. A., 2022_Annelid_SMTs, Zenodo, <https://doi.org/10.5281/zenodo.10063990> (2023).

Acknowledgements

D.A.G. was supported by the National Science Foundation grant (grant no. 2044871). R.E.S. was supported by the Simons Collaboration on the Origin of Life (grant no. 290361FY18). S.S.O.R. was supported by the Irish Research Council (grant no. ELEVATEPD/2014/47) and in part by a grant from Science Foundation Ireland (grant no. 21/FFP-A/9153).

Author contributions

D.A.G. and R.E.S. conceived the project. T.B., D.A.G., and C.M. analyzed computational genetic data. P.M.S. and K.M.V. performed yeast genetics experiments. A.S., S.S.O.R., and R.E.S. performed lipid extractions and GC-MS analysis. D.A.G. drafted the original manuscript with help from T.B. and C.M. All authors read and approved the manuscript before submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43545-z>.

Correspondence and requests for materials should be addressed to D. A. Gold.

Peer review information *Nature Communications* thanks Gabriel Markov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023