Sterol methyltransferases in annelid worms rewrite the molecular fossil record

David Gold (dgold@ucdavis.edu)
University of California, Davis https://orcid.org/0000-0003-0135-4022

Tessa Brunoir
University of California, Davis

Chris Mulligan
University of California, Davis

Ainara Sistiaga
University of Copenhagen

Khan Vuu
Lawrence Berkeley National Laboratory

Patrick Shih
University of California, Berkeley https://orcid.org/0000-0002-2119-3345

Shane OReilly
Atlantic Technological University

Roger Summons
Massachusetts Institute of Technology https://orcid.org/0000-0002-7144-8537

Biological Sciences - Article

Keywords:

Posted Date: June 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1686449/v1

Sterol methyltransferases in annelid worms rewrite the molecular fossil record.

T. Brunoir[1], C. Mulligan[1], A. Sistiaga[2], K.M. Vuu[3], P.M. Shih[3,4], S.S. OReilly[5], R.E. Summons[6], D.A. Gold[1]*

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

* Corresponding author: dgold@ucdavis.edu

ABSTRACT

Molecular fossils have been used to augment the physical fossil record, charting the expansion of complex life through the Neoproterozoic Era (~1000-541 Ma). This work relies on the hypothesis that C_{27} steranes preserved in sedimentary rocks originated from cholesterol, the predominant sterol produced by red algae and animals. Following the same logic, C_{28} and C_{29} carbon steranes are widely considered to be derived from the sterols of fungi, green algae and some protists. In this study, we demonstrate that the gene *24-C sterol methyltransferase* (*smt*), which is necessary to produce C_{28} and C_{29} sterols, exists in segmented worms, an advanced group of animals. Phylogenetic analysis of the relevant gene suggests it was present in the first animals and lost independently in at least seven major lineages. A molecular clock demonstrates that the SMT specific to animals and their closest ancestors was present through the Neoproterozoic**.** Based on these results, C_{27} steranes cannot be considered indicative of Neoproterozoic animals, and C_{28+} steranes are not solely indicative of fungi or green algae. While our results do not necessarily contradict the emerging picture of Neoproterozoic life informed by molecular fossils, they refute the underlying hypothesis that drives the interpretive paradigm.

INTRODUCTION

Lipids preserved in rocks–also 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 emerging [4]. Steranes, the diagenetic remains of sterol lipids found in eukaryotic cell membranes, have proven particularly informative in Neoproterozoic (~1,000-541 Ma) rocks, where fossils are conspicuously absent (Figure 1). C_{27} steranes–the diagenetic products of cholesterol–are abundant in rocks starting around 850 Ma, while C_{28} and C_{29} steranes–derived from ergosterol, campestrol, stigmasterol and other phytosterols–become prominent in the interglacial period ~663-635 Ma [4,5]. In living organisms cholesterol is commonplace in eukaryotes, particularly red algae and animals, while C_{28} and C_{29} sterols are typically found in fungi, algae and green plants. The first appearance of steranes in the Neoproterozoic is therefore hypothesized to represent the ecological expansion of eukaryotes, with simple eukaryotes (possibly red algae) dominating around 820 Ma and green algae expanding around 663 Ma [4,5]. Significant amounts of the sterane isopropylcholestane also appear around this time, and have been interpreted as biomarkers for sea sponges or rhizarian protists [6,7]. 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 [8,9].

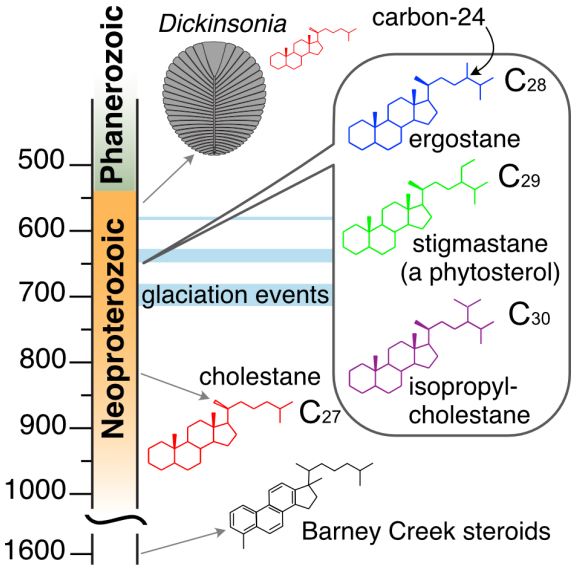

Figure 1: Overview of the Neoproterozoic sterane record. The geologic timescale is provided on the left, with dates in hundreds of millions of years. Arrows indicate the approximate time when various biomarkers become detectable above background thresholds. The site where sterols are normally methylated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

Despite recent advances, the potential to taxonomically constrain various biomarkers is both complex and understudied. Comparative genomics has become a powerful tool for identifying groups of organisms that are capable of synthesizing various lipids, including the precursors of molecular fossils [10]. The gene *24-C sterol methyltransferase* (*smt*) encodes an enzyme that adds methyl groups to carbon-24 of sterol nuclei (signified as C-24) and is responsible for the production of 28- and 29-carbon sterols (Figure 1). The *smt* gene is broadly absent in animals, which is why their cell membranes are dominated by the C_{27} sterol cholesterol [11,12]. The best-known exception to this are the sea sponges (phylum Porifera), an early-branching group of animals that contain *smt* genes and are known to synthesize a wide variety of unusual sterols [6]. However, with increased genomic sampling the number of candidate animal *smt* genes has been increasing. In particular, a putative *smt* has been identified in the genome of the segmented worm *Capitella teleta* [3,12]. Unlike sponges, segmented worms (phylum Annelida) are one of the “higher” animals (subkingdom Eumetazoa), and the presence of *smt* genes in this group–if real–suggests that the potential for animals to synthesize complex sterols is more commonplace than historically appreciated.

There are reasons to be cautious about overinterpreting *smt* genes in genetic databases. Most putative animal *smt* sequences come from draft genomes and transcriptomes, and it is plausible that they represent contaminants from other organisms that were mistakenly integrated into the assembly, as happened in the microbe *Candidatus poribacteria* [6]. Regarding the *C. teleta* gene specifically, it has not been functionally tested, and it is unknown whether it can perform C-24 methylation. For comparison, nematode worms (phylum Nematoda) have an SMT-like protein that catalyzes a C-4 methylation step, resulting in sterols distinct from the ones described above [13]. Finally, the *C. teleta* gene could be real, but also be derived from a recent horizontal gene transfer event. If this is the case, then the presence of an *smt* gene in annelids may not impact our interpretation of the Neoproterozoic sterane record. 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, it would force a reconsideration of what counts as an animal biomarker, as well as a reinterpretation of the Neoproterozoic sterane fossil record.

RESULTS

Sterol methyltransferases are widespread in annelid worms

Annelids are an ecologically diverse phylum containing ~20,000 described species, yet they are severely undersampled in genetic databases such 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 (Figure 2). In many cases, we discovered multiple SMTs in the same species. Annelid SMTs form a well-supported clade in our tree building analysis, suggesting they are shared from a common ancestor (Figures S1, S2). This demonstrates that sterol methyltransferases are indeed commonplace across the annelid worms.

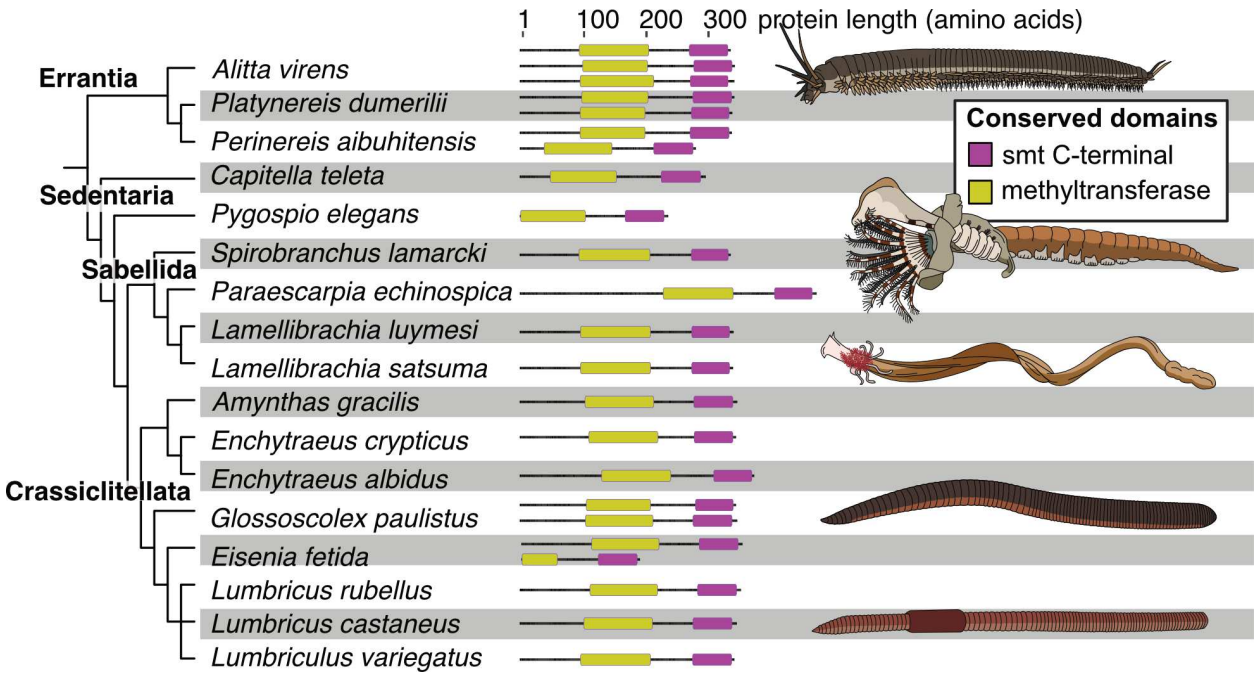

Figure 2. Distribution and structure of annelid SMT proteins. This figure is restricted to annelid SMT proteins containing both conserved domains (i.e. excluding partial sequences). (Left) Phylogeny of annelids in our study with recovered full-length SMT proteins. (Right) Image displaying the number and size of SMT proteins. The position of conserved domains is represented with colored boxes.

We used I-TASSER [14] to model several annelid SMT proteins, comparing their tertiary structure to each other and to better-studied SMTs from yeast and plants (Figure 3). In particular, we analyzed the SMT from *C. teleta* (Figure 3A) as well as two SMTs recovered from the model annelid *Platynereis dumerilii* (Figure 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 C_{28} sterol ergosterol (Figure 3D), and the SMT2 protein in *Arabidopsis thaliana*, a bifunctional enzyme that can generate C_{28} and C_{29} sterols (Figure 3H) [15]. Root-mean-square deviation (RMSD) of atomic positions suggest 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 occurs [16]–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 are more structurally similar to the bifunctional SMT2 than ERG6, as demonstrated 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 C_{28} and C_{29} sterols [17], suggesting this ability is commonplace across annelids.

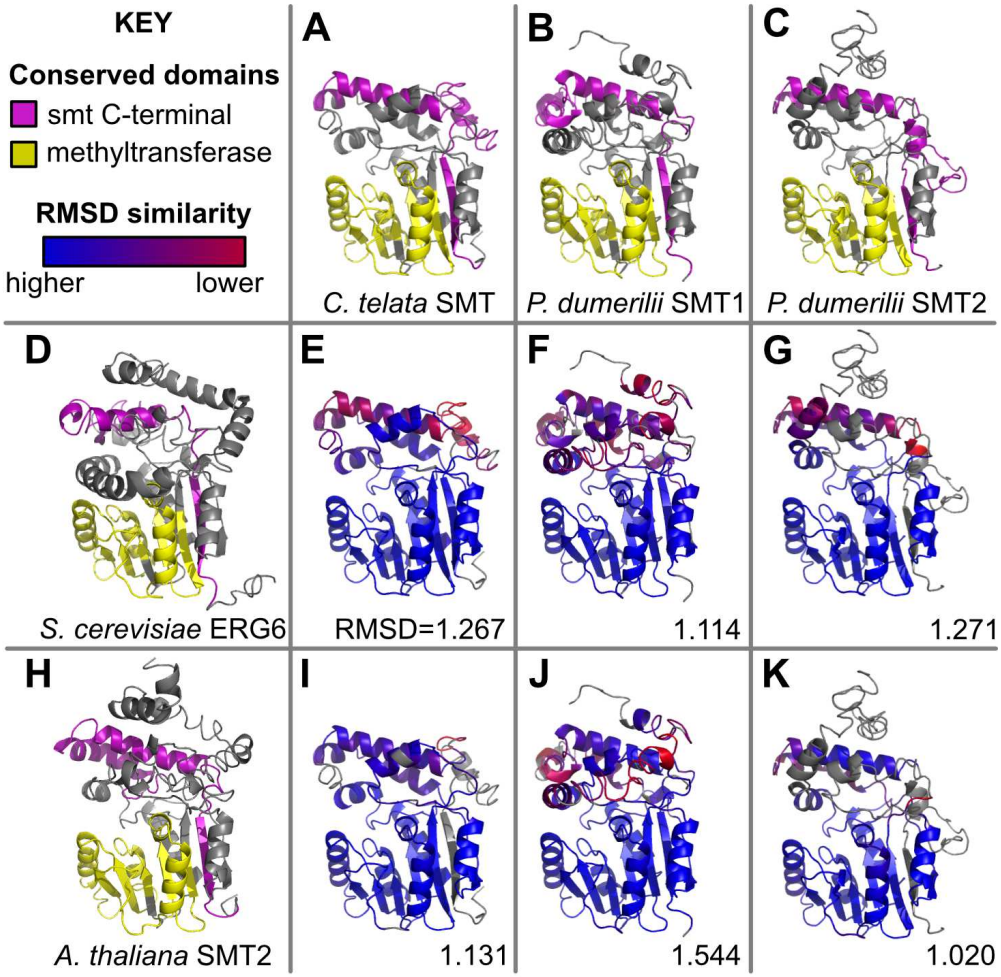

Figure 3. Modeling of annelid SMT proteins. (A-C) best-scoring models for the SMT protein in *C. telata*, as well as two SMTs in *P. dumerilli*, with conserved domain highlighted. (D) Model of *S. cerevisiae* ERG6. (E-F) Alignment of annelid proteins against ERG6, colored by RMSD values. The overall RMSD score is shown in the lower right corner. (H) Model of *A. thaliana* SMT2 protein. (I-J) Alignment of annelid proteins against *A. thaliana* SMT2, colored by RMSD values. The overall RMSD score is shown in the lower right corner. Protein models were generated using I-TASSER server.

Functional analysis of annelid SMTs demonstrate their ability to methylate sterols
After characterizing the *C. telata* SMT, we asked whether or not it is capable of methylating sterols at C-24. To test this we performed a rescue experiment, introducing the *Capitella teleta smt* into a strain of *S. cerevisiae* lacking the wild type gene (commonly known as ERG6 in yeast). ERG6$^-$ yeast are viable but cannot produce ergosterol and have a slower rate of growth than their wild type counterparts [18]. We found that ERG6$^-$ *S. cerevisiae* are capable of producing ergosterol with the addition of the *C. teleta* gene (**Figure 4**). This work complements Michellod et al., demonstrating that multiple annelid species can synthesize phytosterols *de novo* [17].

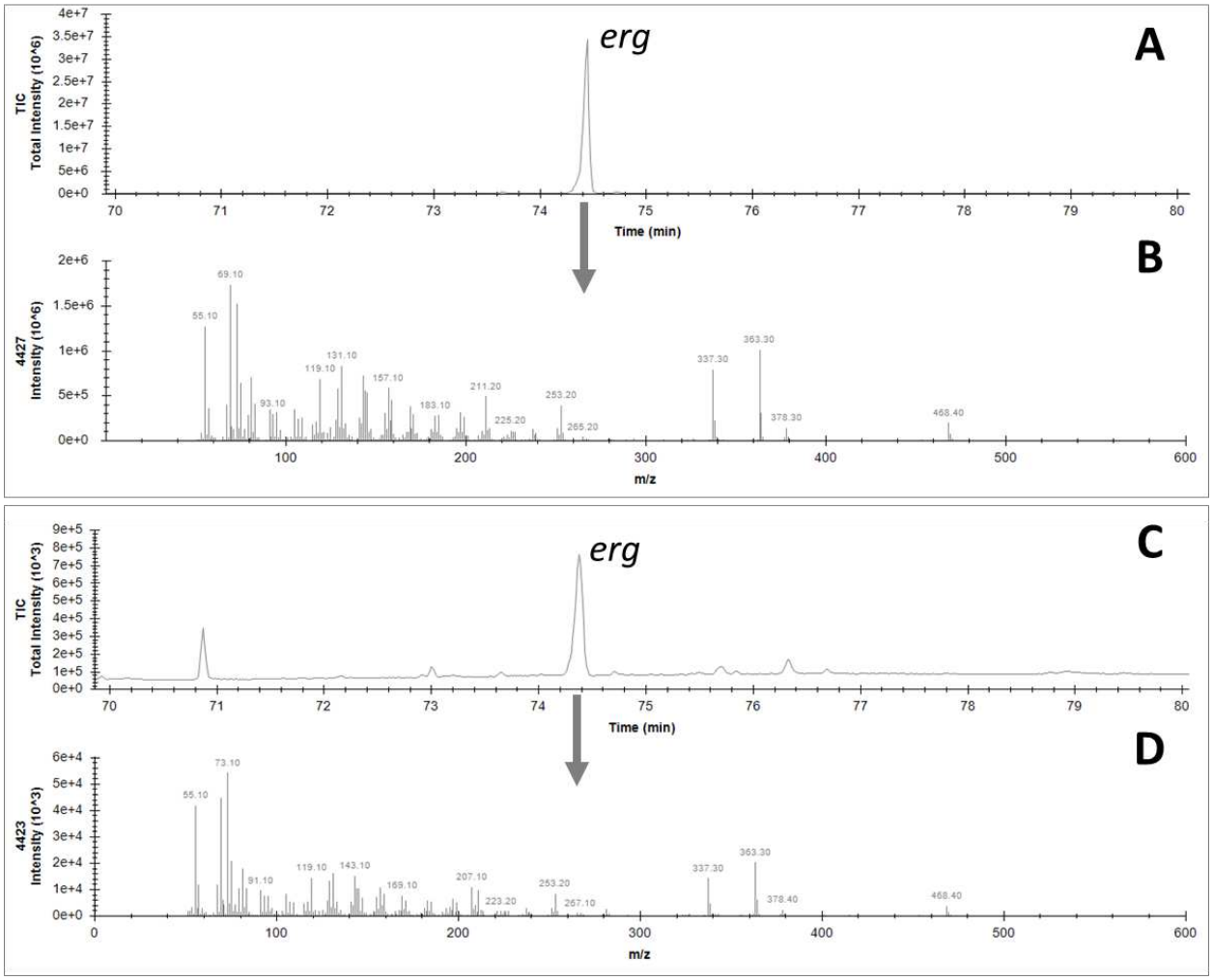

Figure 4. **Production of ergosterol is rescued in yeast with the wild type ERG6 replaced by its annelid homolog.** (A) partial total ion chromatogram of an authentic commercial ergosterol standard (peak labeled ‘erg’) and the confirmatory mass spectrum for this peak, showing characteristic fragment ions at *m*/*z* 468, 363, 337 for the TMS derivative. Partial total ion chromatogram of a sample total lipid extract from

yeast with the wild type ERG6 replaced by its annelid homolog (A), confirming the occurrence of ergosterol based on retentional time and the mass spectrum (D)

Phylogenetic analysis of animal SMTs supports vertical inheritance

In addition to annelids, we recovered SMT proteins from several other animal clades. Many of our sequences came from transcriptomes which are known to be fraught with contaminations from food and symbionts. Still, our careful vetting process indicates that *smt* genes are present in a minimum of two additional clades of eumetazoans–rotifers and stony corals. Rotifers (clade are near-microscopic animals that are distant relatives to annelids. A recent project assembling 31 rotifer genomes has resulted in a large number of *smt* genes in the genera *Adineta*, *Rotaria*, and *Didymodactylos* [19]. Regarding stony corals (clade Scleractinia), one SMT-like protein has been predicted from the *Orbicella faveolata* genome (NCBI Accession: 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, and were able to map one of these genes (NCBI Accession: FX438716.1) to the genome of the coral *Porites australiensis* (Figure S3). A second candidate SMT from *P. australiensis* did not map to the genome, and shows high sequence similarity to the coral symbiont *Symbiodinium*, demonstrating that we can distinguish between host and microbial sequences (see Methods). Similar to Michellod et al. [17] our putative animal sequences are scattered across the SMT gene tree (Figure S1), which, if taken literally, would support multiple horizontal gene transfer events. But after removing probable contaminants and performing species tree reconciliation (see Methods), the most parsimonious scenario is that all animal SMTs are monophyletic. We therefore conclude that vertical inheritance from a common ancestor best explains the distribution of SMTs in animals.

A molecular clock suggests animals had SMTs in the Neoproterozoic

To test whether animal SMTs coincide with the Neoproterozoic biomarker record, we generated a gene-centered molecular clock using 107 vetted SMT sequences and nine fossil calibrations (see supplementary text for details and justification of calibrations). The results are summarized in Figure 5, with a more detailed output provided in Figure S4. 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 remarkably consistent with multi-gene, species-level molecular clocks–placing the origin of eukaryote SMTs between 1800 and 1600 Ma and animal SMTs between 828 and 572 Ma[20–22]. The division of sponges into two clades supports the hypothesis that a gene duplication event occurred early in animal evolution, consistent with previous results [6]. The SMT found in corals represents the retention of one of those duplicated genes while the SMT found in annelids and rotifers evolved from the second copy. 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 Figure 5), which have occurred throughout the Phanerozoic (<541 Ma). Our results are ambiguous as to whether the first animal SMTs existed by the time C_{28} and C_{29} steranes appear in the fossil record ~663-635 Ma. However, when considering the animal stem lineage (i.e. extinct organisms more closely related to animals than choanoflagellates), even the younger end of the 95% confidence range (1207-623 Ma) overlaps with the rise of C_{28} and C_{29} steranes. We can therefore conclude that animals or their closest extinct relatives had functional SMTs concurrent with the rise of complex steranes in the Neoproterozoic.

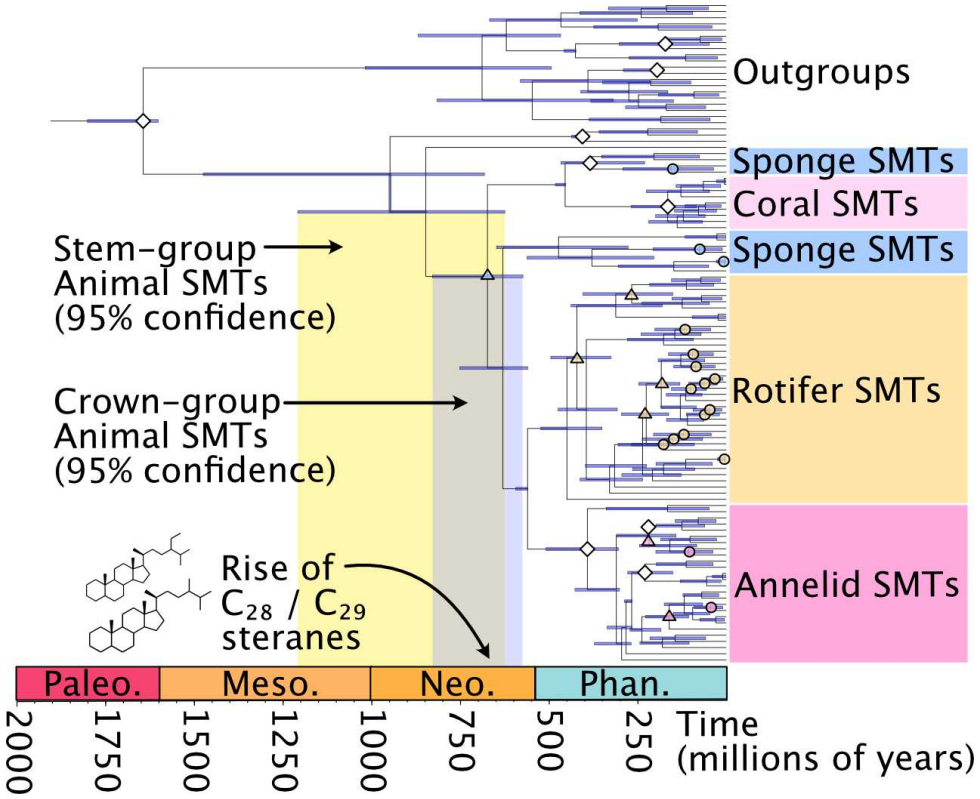

Figure 5. Molecular clock of animal SMT proteins. An SMT-centered molecular clock generated using BEAST. White diamonds represent fossil calibrations, which are described in detail in the Supplementary Text. Putative gene duplication events in the animals have been labeled at relevant nodes with colored shapes: circles represent gene duplication events that are specific to one species in our study; triangles represent events that include multiple species. Phan. = Phanerozoic; Neo. = Neoproterozoic; Meso. = Mesoproterozoic; Paleo. = Paleoproterozoic.

DISCUSSION

The analyses in this study demonstrate that the genes necessary to synthesize complex sterols existed in the first animals, and were retained long after the diversification of the Eumetazoa. We therefore reject the hypothesis that cholesterol was the ancestral sterol synthesized by higher animals, or that cholestanes can serve as biomarkers for animals. If our interpretations are correct, then the observation that most living animals utilize cholesterol as their sterol lipid is a function of extensive gene loss. The retention of SMTs in annelids via vertical inheritance necessitates its loss in at least seven major lineages (**Figure 6**). The actual number of losses is likely an order of magnitude higher, given the large number of coral and annelid genomes that lack SMTs, as well as the possible loss of SMTs in the minor animal phyla, which are poorly represented in genetic databases. The loss of SMTs in so many lineages is likely due to the evolution of complex feeding strategies. Many early animal groups were able to acquire the necessary sterols from their diets, and lost the ability to make complex sterols or any sterols *de novo* [12]. As to why the *smt* gene has been retained in many annelids, corals, and rotifers–nonetheless why it has been independently duplicated in many species–is more puzzling. In many eukaryotes, the number of SMT proteins correlates with the number of methylene groups added to the C-24 site on sterols [6]. This seems unlikely to hold in eumetazoans, as a single SMT in the annelids *Olavius* and *Inanidrilus* is sufficient for adding two methylene groups to the sterol side-chain [17], and unlike sponges, exotic sterols have not been recovered in these groups. For rotifers, the answer may lay in part with unusual reproductive strategy; the bdelloid rotifers from our dataset reproduce asexually and exhibit tetraploidy [23]. As a consequence, their genomes are more like plants than other animals, and like plants, rotifer SMTs are likely to be partially redundant [24]. Protein modeling suggests there are subtle differences in shape of the two SMTs from *P. dumerilli* (Figure 3) but why two copies of the gene have been preserved in the Errantia, or why *Alitta virens* has a third copy, is currently unknown. The selective advantage of retaining SMTs in some lineages, when the vast majority of animal lineages have abandoned the protein, will require further study.

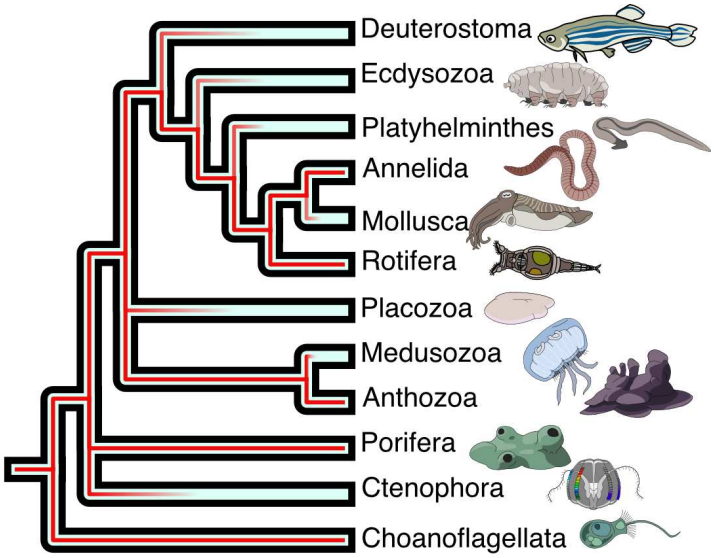

Figure 6 Summary of SMT loss in the animals. This tree provides a conservative estimate of the number of SMT losses, as it excludes many understudied animal phyla that may also lack the protein.

Regardless of the reason why *smt* genes were retained and expanded in some animals, our results, combined with complementary work by Michellod et al. [17], clearly demonstrate that the biosynthesis of complex sterols is an ancestral and ancient trait in animals. This puts our current interpretation of the early eukaryotic biomarker record into question. Despite the challenge this presents to the overarching paradigm, there is also promise that a more refined understanding of sterol evolution will provide fresh insight into the meaning of molecular fossils. As an example, there are other lines of evidence besides biomarkers supporting an animal affinity for *Dickinsonia* [25,26], so the dominance of cholestane in this fossil could help constrain where in the animal tree it belongs. As the fields of comparative genomics and molecular fossils continue to mature, dialogue between the two is the surest way to reconstruct the history of complex life.

METHODS

Data Accessibility: All code used in this study, including input datasets, intermediate files, and output, are provided on GitHub at https://github.com/DavidGoldLab/2022_Annelid_SMTs.

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 program packaged with Trinity [27]. Sequences were then separated by species, and redundant proteins were removed using CD-HIT with a 90% similarity cutoff [28]. Conserved domains were identified in the remaining proteins using the pfam_scan Perl script included in BioConda, which uses HMMER v3 to compare proteins against the PFAM-A database [29,30]. Methyltransferase and C-terminal domains were extracted independently using SAMTOOLS v.1.13 and aligned with MAAFT; they were combined again with FASconCAT-G and cleaned with trimAl v.1.4 [31,32]. 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 for finding 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 FastTree [33]. 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 Figure S1). Sequences were only considered if three 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 v.1.6.12 [34], which is visualized in Figure S2. 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 Figure S2) Was used for downstream species tree reconciliation and molecular clock analysis. Relevant files and code can be found in folder 2 on GitHub.

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 v.3.6.1 [35] to reflect updated taxonomic affinities based on references [36–38]. The species tree and gene were passed to NOTUNG v.2.1.5,

which rearranged poorly supported branches in the gene tree to provide the most parsimonious reconciliation with the species tree [39]. A molecular clock was generated from the reconciled tree using BEAST v.1.10.5 [40]. 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 Test. 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. All input files, including the XML used for the BEAST run, are provided in folder 5 on GitHub.

Protein Modeling. Protein structure and function predictions were performed on the I-TASSER server [14]. 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. 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 wildtype ERG6 in *S. cerevisiae* strain BY4742 (MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) 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. 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 addition of water, the DCM layers were combined, washed several times with H_2O, and dried over Na_2SO_2. 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 5975C 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.

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 [41]. The only sequence to map to a genome was a single *smt* (Accession: 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 TransDecoder [27]. The input files and results of this analysis are available in folder S3 on GitHub.

REFERENCES

1. Brocks, J. J. Millimeter-scale concentration gradients of hydrocarbons in Archean shales: Live-oil escape or fingerprint of contamination? *Geochim. Cosmochim. Acta* **75**, 3196–3213 (2011).
2. French, K. L. *et al.* Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proc. Natl. Acad. Sci.* **112**, 5915–5920 (2015).
3. 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).
4. Love, G. D. & Zumberge, J. A. *Emerging patterns in proterozoic lipid biomarker records*. (Cambridge University Press, 2021).
5. Brocks, J. J. *et al.* The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578–581 (2017).

6. Gold, D. A. *et al.* Sterol and genomic analyses validate the sponge biomarker hypothesis. *Proc. Natl. Acad. Sci.* **113**, 2684–2689 (2016).
7. Nettersheim, B. J. *et al.* Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. *Nat. Ecol. Evol.* **3**, 577–581 (2019).
8. Bobrovskiy, I., Hope, J. M., Krasnova, A., Ivantsov, A. & Brocks, J. J. Molecular fossils from organically preserved Ediacara biota reveal cyanobacterial origin for Beltanelliformis. *Nat. Ecol. Evol.* **2**, 437–440 (2018).
9. Bobrovskiy, I. *et al.* Ancient steroids establish the Ediacaran fossil Dickinsonia as one of the earliest animals. *Science* **361**, 1246–1249 (2018).
10. Summons, R. E., Welander, P. V. & Gold, D. A. Lipid biomarkers: molecular tools for illuminating the history of microbial life. *Nat. Rev. Microbiol.* 1–12 (2021).
11. 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).
12. Zhang, T. *et al.* Evolution of the cholesterol biosynthesis pathway in animals. *Mol. Biol. Evol.* **36**, 2548–2556 (2019).
13. Zhou, W. *et al.* A nematode sterol C4α-methyltransferase catalyzes a new methylation reaction responsible for sterol diversity [S]. *J. Lipid Res.* **61**, 192–204 (2020).
14. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 1–8 (2008).
15. 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).
16. Nes, W. D. *et al.* Active Site Mapping and Substrate Channeling in the Sterol Methyltransferase Pathway*. *J. Biol. Chem.* **277**, 42549–42556 (2002).
17. Michellod, D. *et al.* De novo phytosterol synthesis in animals. *bioRxiv* (2022).

18. 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).
19. Nowell, R. W. *et al.* Evolutionary dynamics of transposable elements in bdelloid rotifers. *Elife* **10**, e63194 (2021).
20. 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).
21. Strassert, J. F. H., 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).
22. 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).
23. Flot, J.-F. *et al.* Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. *Nature* **500**, 453–457 (2013).
24. 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).
25. Evans, S. D., Droser, M. L. & Gehling, J. G. Dickinsonia liftoff: evidence of current derived morphologies. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **434**, 28–33 (2015).
26. Gold, D. A., Runnegar, B., Gehling, J. G. & Jacobs, D. K. Ancestral state reconstruction of ontogeny supports a bilaterian affinity for Dickinsonia. *Evol. Dev.* **17**, 315–324 (2015).
27. 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).
28. 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).

29. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
30. Grüning, B. *et al.* Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* **15**, 475–476 (2018).
31. Kück, P. & Meusemann, K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* **56**, 1115–1118 (2010).
32. 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).
33. 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).
34. 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).
35. Maddison, W. P. Mesquite: a modular system for evolutionary analysis. *Evolution* **62**, 1103–1118 (2008).
36. Plese, B. *et al.* Mitochondrial evolution in the Demospongiae (Porifera): Phylogeny, divergence time, and genome biology. *Mol. Phylogenet. Evol.* **155**, 107011 (2021).
37. 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).
38. Weigert, A. & Bleidorn, C. Current status of annelid phylogeny. *Org. Divers. Evol.* **16**, 345–362 (2016).
39. 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).
40. 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).

41. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GoldAnnelidSMTsSI.pdf