Sterol side chain alkylation by sponge and bacterial methyltransferases

Malory O. Brown[a], Babatunde O. Olagunju[b], José-Luis Giner[b], and Paula V. Welander[a]

[a]Department of Earth System Science, Stanford University, Stanford, CA 94305; and [b]Department of Chemistry, State University of New York-Environmental Science and Forestry, Syracuse, NY 13210

Corresponding author: Paula V. Welander, 473 Via Ortega, Rm 140, Stanford, CA 94305, (650) 723-7341, welander@stanford.edu

Keywords: sterols, biomarkers, sterol methyltransferase, 24-isopropylcholestane

Abstract

Fossilized sterols, or steranes, are broadly interpreted as biomarkers for past eukaryotic life. Steranes with side chain methylations can act as more specific eukaryotic biomarkers given the limited taxonomic distribution of their sterol precursors in extant eukaryotes. One side chain methylated sterane, 24-isopropylcholestane, has been interpreted as a biomarker for marine sponges of the Demospongiae class as they are the only organisms known to produce 24-isopropyl sterols as their major sterols. The presence of 24-isopropylcholestane in late Neoproterozoic rocks may therefore represent some of the earliest evidence for animals on Earth. However, the use of 24-isopropylcholestane as a sponge biomarker has been questioned for a variety of reasons including the potential for alternative sources and a lack of information on the biochemical mechanism behind the formation of its sterol precursors. Specifically, enzymes capable of methylating sterols to give the 24-isopropyl side chain have not been identified, and the function of putative sponge sterol 24-C-methyltransferases has not been demonstrated experimentally. In this study, we verified the functionality of sponge sterol methyltransferases and explored the potential for bacterial sterol side chain methylation through in vitro protein expression and lipid analyses. We identified three bacterial sterol methyltransferases each capable of sequential methylations resulting in the 24-isopropyl side chain structure of the 24-isopropylcholestane biomarker. Two of these side chain propylating enzymes were

identified in a metagenome from a demosponge microbiome, suggesting bacterial symbionts may play a role in 24-isopropyl sterol biosynthesis in the demosponge holobiont.

Significance Statement

According to the sponge biomarker hypothesis, geologic 24-isopropylcholestane can act as an indicator for ancient demosponges and may represent some of the first evidence for animal life on Earth. However, enzymes capable of methylating sterols to give the 24-isopropyl sterol side chain have not been identified. We therefore investigated the functionality of sponge sterol methyltransferases, which alkylate the sterol side chain, through in vitro protein expression. We also analyzed the function of several bacterial sterol methyltransferases, three of which produced 24-isopropyl sterols. Our results indicate yet-uncultured bacteria, including demosponge symbionts, have the genomic capacity to synthesize side chain alkylated sterols and should therefore be considered when interpreting ergostane, stigmastane, and 24-isopropylcholestane biomarkers in the geologic record.

Introduction

Sterols are polycyclic triterpenoid lipids required by most eukaryotes for many critical cellular functions, such as maintaining membrane integrity, phagocytosis, stress tolerance, embryogenesis, and cellular signaling (1–6). In addition, sterols are quite recalcitrant and are preserved as sterane hydrocarbons in sedimentary rocks up to 1.6 billion years old (7). Given their persistence in the geologic record and wide distribution amongst extant eukaryotes, steranes are broadly interpreted as biomarkers for ancient

eukaryotic life (8, 9). But particular modifications to the sterol side chain, such as alkylation at the C-24 and C-26 positions, remain unique to certain eukaryotic lineages and can therefore function as more specific biomarkers (10). Two side chain alkylated steranes, 24-isopropylcholestane (24-ipc) and 26-methylstigmastane (26-mes), have been used as biomarkers for demosponges (11, 12) as they are the only extant organisms known to contain sterols with the 24-ipc and 26-mes side chain structures as their major sterols (12–18). The co-occurrence of 24-ipc and 26-mes in certain late Neoproterozoic rocks (660-540 Ma) predates the earliest sponge macrofossils by ~100 million years (19), potentially representing the first evidence of animals in the geologic record.

However, the use of 24-ipc and 26-mes as biomarkers for demosponges has been controversial and continues to be debated for several reasons (19–26). For example, 24-isopropyl sterols have been identified in pelagophyte algae which produce the structurally related compound 24-*n*-propylcholesterol as their major sterol (27–30). Trace amounts of 24-ipc and 26-mes have also been reported in hydrogenated lipid extracts from Rhizaria (31), although attempts to replicate these results have not been successful (23). Subsequent studies demonstrated 24-ipc and 26-mes form abiotically from certain C_{29} sterols during laboratory pyrolysis under conditions that may mimic diagenesis (26, 32). These studies suggest that the previously observed Rhizarian 24-isopropyl and 26-methyl steranes may have been generated through laboratory pyrolysis rather than by in vivo synthesis and that Neoproterozoic 24-ipc and 26-mes may have formed abiotically from algal sterols. Further, microbial symbionts can constitute up to 60% of sponge biomass (33, 34) and diverse bacteria are known to produce sterols (35). Bacterial symbionts may therefore contribute to 24-isopropyl and 26-methyl sterol biosynthesis in the sponge holobiont – defined as the

sponge host and the various microbes and viruses that reside within it (36). However, bacteria with the genomic capacity to alkylate the sterol side chain have not been identified.

One approach to constrain these uncertainties in the sponge biomarker hypothesis is to better understand the biochemical mechanism underpinning these distinct alkylations at C-24 and C-26. Sterol 24-C-methyltransferases (SMTs), which alkylate the C-24 position of the sterol side chain at an isolated double bond through an *S*-adenosylmethionine (SAM) dependent mechanism (37), are the only enzymes known to methylate the sterol side chain. These proteins have been characterized extensively from fungal and plant sources but only recently has an attempt been made to address this biochemical reaction in sponges. Phylogenetic analyses performed to understand how sponges could be synthesizing 24-propyl sterols suggested that the number of SMT copies in a eukaryotic genome corresponds to the number of carbons added at C-24 in all analyzed lineages except sponges, which contain one less SMT in their genome or transcriptome than expected based on their sterol profile (38). This prompted the hypotheses that at least two SMT copies in a genome are required to produce 24-propyl sterols and that sponge SMTs are capable of multiple methylations at C-24. However, no experimental evidence demonstrating this activity by any sponge SMT currently exists, and SMTs capable of alkylating the sterol side chain to produce 24-propyl sterols have not been identified in any extant organism.

Here, we characterize a variety of putative SMTs from sponge and bacterial sources to better understand the biochemical requirements for generating propylated sterols. We demonstrate that sponge sterol 24-C-methyltransferase homologs are bona fide SMTs that can perform multiple rounds of alkylation to produce sterols methylated and ethylated at

the C-24 position. In addition, we show that putative bacterial SMTs identified in metagenomic datasets, including those from demosponge microbial symbiont communities, are capable of alkylating at the C-24 position. Surprisingly, 3 SMTs of bacterial origin can each perform all 3 of the methylation steps required to produce the 24-isopropyl side chain. This is the first in vitro demonstration of propylation of a sterol side chain by an SMT and the first demonstration that bacteria have the potential to alkylate sterols at C-24. Taken together, these data suggest alternative sources of 24-ipc are plausible and demonstrate the utility of understanding the biochemistry of lipid biosynthesis to allow for more robust biomarker interpretations.

Results

Identification of functional sponge sterol methyltransferases. Previous studies identified SMT homologs in a variety of sponge genomes and transcriptomes (38). However, there is currently no experimental evidence demonstrating that these putative SMTs alkylate the sterol side chain much less methylate multiple times to generate 24-propyl sterols. Our first goal was to heterologously express these putative SMTs from sponge sources and show methylation activity. We tested 8 SMT homologs identified in the genomes and transcriptomes of 8 sponges of the classes Demospongiae, Homoscleromorpha, and Calcarea which have each been shown to contain 24-ethyl sterols at the species (39–42) or genus (43, 44) level. SMTs were codon-optimized and artificially synthesized for expression in *Escherichia coli*. Cell extracts were generated from *E. coli* overexpressing each sponge SMT and incubated with two potential SMT substrates, desmosterol or 24-methylenecholesterol, and resulting sterol products were analyzed using

GC-MS. All 8 of the analyzed sponge SMTs were functional in our system and alkylated desmosterol to produce 24-methyl sterols including 24-methylenecholesterol and codisterol (Table 1, Fig. 1). Six of the sponge SMTs also alkylated 24-methylenecholesterol to 24-ethyl sterols including fucosterol, isofucosterol, and (epi)clerosterol demonstrating that some sponge SMTs can perform two sequential side chain methylations. None of the analyzed sponge SMTs produced 24-propyl sterols in our expression system.

Bacterial SMTs are functional and produce 24-isopropyl sterols. Although we demonstrated sponge SMT homologs are true sterol methyltransferases, none of the proteins we tested produced 24-isopropyl sterols–the diagenetic precursors to the 24-ipc biomarker. Our analyses of sponge microbiomes, however, revealed several potential sterol biosynthesis proteins, including SMT homologs, from bacterial symbionts. We hypothesized that these bacterial SMTs may contribute to side chain alkylation in the sponge holobiont. However, no putative bacterial SMTs, from either free-living or host-associated sources, have been experimentally shown to alkylate the sterol side chain.

We chose 10 putative bacterial SMTs identified in the microbiome of 4 demosponge species for in vitro expression and lipid analyses as previously performed with the sponge SMTs. Of these 10 symbiont SMTs, 8 methylated desmosterol to 24-methyl sterols including codisterol, 24-methylenecholesterol, and 24-methyldesmosterol (Table 1, Fig. 1). Of the 8 SMTs that methylated desmosterol, 3 also methylated 24-methylenecholesterol to 24-ethyl sterols including fucosterol, isofucosterol, and (epi)clerosterol. Two SMTs, both identified in a metagenome from the demosponge *Aplysina aerophoba*, also methylated 24-methylenecholesterol to two distinct C_{30} sterols.

NMR analysis confirmed the first SMT produced 24*S*-isopropenylcholesterol while the second produced 24-isopropylidenecholesterol. Both *A. aerophoba* metagenome SMTs were capable of all three methylation steps required to synthesize the 24-ipc side chain. One propylating SMT was identified in a gene cluster with several other homologs of sterol biosynthesis genes, including squalene monooxygenase, oxidosqualene cyclase, sterol 14α-demethylase, delta(14)-sterol reductase, and the bacterial C-4 demethylation genes *sdmA* and *sdmB*, suggesting an unknown bacterial sponge symbiont has the genomic capacity to synthesize 24-isopropyl sterols de novo (Fig. 2).

Identification of bacterial symbiont SMTs capable of producing side chain alkylated sterols, including 24-isopropyl sterols, led us to question if bacteria from environments outside a sponge host have the potential to alkylate the sterol side chain. Through additional searches, we identified potential bacterial SMTs from a variety of publicly available environmental metagenomes and metagenome-assembled genomes (MAGs). We tested the function of 14 of these SMT homologs from marine, freshwater, thermal spring, and coral reef metagenomes and MAGs. Twelve of these SMTs alkylated the sterol side chain to produce 24-methyl and/or 24-ethyl sterols (Table 1). One SMT, identified in a drinking water filter MAG assigned to the phylum Chlamydiae, was capable of alkylating desmosterol to both 24*S*-isopropenylcholesterol and 24-isopropylidenecholesterol (Fig. 1).

Phylogenetic analyses of bacterial SMTs. The identification of functional SMTs in metagenomes and MAGs assigned to the bacterial domain, including the three SMTs that propylate the sterol side chain, is quite significant given that no bacterial proteins have ever been confirmed to alkylate the sterol side chain. To provide further evidence that these

SMTs are truly of bacterial origin, we first performed phylogenetic analyses of 10 sponge SMTs and the 24 metagenomic SMTs tested in our in vitro system. We added these SMTs and additional eukaryotic SMT homologs identified in GenBank to a previously published SMT dataset (38) and repeated their analysis to generate a maximum-likelihood phylogenetic tree from a concatenated alignment of conserved methyltransferase and C-terminal domains as identified in Pfam. Through this analysis, we see that 23 out of the 24 putative bacterial SMTs, including those from sponge symbionts, form two clades separate from the sponge SMT clade suggesting bacterial symbionts did not acquire their SMTs from their sponge host via horizontal gene transfer. While the observed bootstrap values are not robust enough to draw any meaningful evolutionary conclusions, the observed clustering suggests that the majority of the metagenomic SMTs we tested are distinct and more closely related to each other than to the SMTs of eukaryotes known to produce side chain alkylated sterols (Fig. 3).

The genomic context of the putative bacterial SMTs we tested can also provide evidence for their bacterial origins. Of the 24 metagenomic SMTs tested here, 9 were in gene clusters or MAGs with a homolog of oxidosqualene cyclase (OSC), a key enzyme in sterol biosynthesis responsible for cyclizing oxidosqualene to lanosterol (35). Previous work has demonstrated that OSCs cluster into robust phylogenetic clades with high bootstrap support (45). Phylogenetic analyses of environmental metagenomic OSCs from unknown sources can therefore provide insight into the types of uncultured organisms that may be producing sterols (35, 46). Thus, phylogenetic analyses of OSCs associated with the SMTs we tested could provide more context of the bacterial origins of these proteins. To generate the OSC tree, we identified 485 OSC sequences from bacterial and eukaryotic

genomes and MAGs using the OSC from the γ-proteobacterium *Methylococcus capsulatus* BATH as a search query against the JGI IMG database. These OSC protein sequences were aligned and used to generate a maximum-likelihood phylogenetic tree using the squalene-hopene cyclase (SHC) protein as an outgroup. Seven of the OSCs, including 4 from demosponge metagenomes, fell into the Group 1 clade which contains only bacterial OSCs (Fig. S2). The remaining 2 OSCs, from soil and hot spring sediment metagenomes, fell into bacterial Group 2. The position of these metagenomic OSCs, which are associated with SMTs we tested, provides additional evidence that these SMTs originate from a bacterial source.

To further demonstrate these SMTs originate from a bacterial source, we used the recently developed classifier Whokaryote (47), which predicts whether a metagenomic contig is likely eukaryotic or prokaryotic based on intergenic distance, gene density, gene length, and k-mer frequencies. Whokaryote predicts all 19 of the contigs containing the metagenomic and MAG SMTs tested here to be of prokaryotic origin (Table S3).

Discussion

The sponge biomarker hypothesis proposes that demosponges are the predominant source of 24-isopropylcholestane biomarkers preserved in the sedimentary rock record. Support for this hypothesis was put forth through phylogenetic and molecular clock analyses of putative sponge sterol methyltransferases (SMTs) thought to catalyze the formation of propylated sterols (38). These studies concluded that sponge SMTs are not only bona fide sterol methyltransferases but that they are capable of multiple rounds of

methylation. However, no experimental evidence has been provided to confirm that sponge-derived SMTs methylate at the C-24 position much less methylate multiple times.

The results presented here experimentally demonstrate that sponge genomes and transcriptomes do indeed contain functional SMTs capable of multiple rounds of methylation at the sterol side chain. However, we did not observe this activity across all sponge SMTs analyzed in this study. Sterol profiles from *Oscarella carmela* (Homoscleromorpha) and *Sycon ciliatum* (Calcarea) have not been reported, but 24-ethyl sterols have been identified in other members of *Oscarella* and *Sycon* (43, 48). Yet, the *O. carmela* and *S. ciliatum* SMTs tested here did not produce 24-ethyl sterols. SMTs capable of multiple rounds of methylation may therefore be restricted to certain sponge species, although we recognize that these SMTs may be more selective and require a different substrate than we have provided in our assays to produce 24-ethyl sterols. Further, we did not observe the production of 24-propyl sterols by any sponge SMT, although this was not surprising given that 24-propyl sterols have not been identified in any of the sponge species analyzed here. One of the difficulties in trying to ascertain the potential for propylation by sponge-derived SMTs is a lack of genomic data from demosponges that have been confirmed to produce 24-ipc. Future sequencing of genomes or transcriptomes from demosponges known to contain sterols with the 24-ipc carbon skeleton may reveal sponge-derived SMTs sufficient for the biosynthesis of 24-propyl sterols.

While we were unable to demonstrate the production of propylated sterols through the expression of sponge SMTs, we did show that bacterial symbionts have the genomic capacity to synthesize these lipids and may therefore contribute to the 24-isopropyl sterol pool in the demosponge holobiont. Studies performed prior to the development of 16S

community sequencing attempted to demonstrate de novo sterol biosynthesis in demosponges. However, these experiments were performed using cell-free sponge extracts or on whole sponge tissue from which the presence of symbiont-derived enzymes cannot be excluded (49–51). Without confirmation that the bacterial symbionts in these samples do not contain functional SMTs, the possibility remains that bacterial symbionts contribute to sterol biosynthesis in demosponges. Further, holo-transcriptomic sequencing of the demosponge *Vaceletia* sp. found that several key sterol biosynthesis transcripts, including those for squalene synthase, squalene monooxygenase, and oxidosqualene cyclase, were bacterially-derived (52). Both sponge and bacterial symbiont proteins may therefore be required to produce 24-isopropyl sterols in vivo. However, the metagenomic gene clusters containing functional SMTs and several other key genes for sterol biosynthesis identified here suggest bacterial demosponge symbionts may produce sterols de novo without demosponge input. The 2 symbiont SMTs sufficient for side chain propylation were both identified in metagenomes from the demosponge *A. aerophoba*, which appears to contain only trace amounts of 24-isopropyl sterols (18) . Our work highlights the need for robust sequencing of both the sponge host and its microbiome coupled to sterol analysis in demosponge species that produce 24-isopropyls as their major sterols.

From a biomarker perspective, our demonstration that a single bacterial SMT is sufficient for propylating a sterol side chain is quite significant. Previous molecular clock analyses suggested demosponges were the most likely source of Cryogenian 24-ipc biomarkers given that demosponges acquired a second SMT copy prior to their deposition, while pelagophyte algae did not acquire a third SMT copy until the Phanerozoic (38). Further, researchers have argued against Rhizaria as a source of 24-isopropyl sterols given

that *Bigelowiella natans* contains only 2 SMT copies in its genome and is thus unlikely to synthesize 24-propyl sterols (23). These interpretations rely on the hypothesis that multiple SMT copies are required to produce 24-isopropyl sterols. However, we identified 3 SMTs each capable of producing the 24-isopropyl side chain via a single protein. SMT copy number may therefore be an unreliable predictor of an organism's ability to produce 24-propyl sterols. Now that propylating SMTs have been identified, site-directed mutagenesis studies like those performed with yeast (53–55), algal (56), and plant (57) SMTs may reveal specific amino acid changes that allow an SMT protein to perform a third methylation. If such key residues are identified, they could act as a more reliable predictor of 24-isopropyl sterol biosynthesis directly from sequencing data than SMT copy number alone.

In addition to its use as a demosponge biomarker, 24-ipc has also been used to indicate marine input into terrestrial systems (58–60) given that 24-ipc precursor sterols have not been identified in freshwater demosponges or algae. However, the production of 24-isopropyl sterols by an SMT identified in a freshwater Chlamydiae MAG suggests the occurrence of these sterols may not be limited to marine environments. Members of the phylum Chlamydiae have not been shown to produce sterols, and homologs of other sterol biosynthesis genes were not present in the Chlamydiae MAG. It is therefore unlikely that this organism synthesizes 24-isopropyl sterols de novo. However, some endosymbiotic Chlamydiae are known to recruit sterols from their eukaryotic host into their membranes (61). The possibility therefore remains that certain freshwater Chlamydiae alkylate the side chain of host-acquired sterols to give the 24-isopropyl side chain via their SMTs for reasons yet unknown. If this is the case, it may explain the presence of 22-dehydro-24-

isopropylcholestanol throughout a freshwater sediment core from Motte Lake, France (62). We would therefore exercise caution when, for example, interpreting the presence of 24-ipc as evidence for seawater incursion events.

Finally, although we identified SMTs capable of 24-isopropyl sterol biosynthesis, proteins capable of methylating sterols to give other geologically important side chain structures remain undiscovered. Identification of enzymes capable of methylating sterols at the C-26 position to give the 26-mes side chain structure remains especially important when considering the biomarker evidence for Precambrian sponges. Studies in which radiolabeled sterol precursors were fed to whole sponges or sponge extracts have demonstrated codisterol or epicodisterol are the preferred substrates for methylation at C-26 via a hypothetical 26-SMT protein (49, 63, 64). We identified (epi)codisterol as a product of several SMTs, from both sponges and bacteria, suggesting both groups have the genomic capacity to produce these potential 26-SMT substrates. However, it remains to be seen if sponges and/or bacteria, sponge-associated and otherwise, encode SMTs capable of methylation at C-26. Once enzymes with this functionality are identified, in-depth kinetics studies with purified SMT proteins and multiple sterol substrates may reveal if SMTs from sponges and/or bacteria are likely to produce 24-ipc and 26-mes sterol precursors in ratios that correspond to the C_{30} sterane distributions observed in the Cryogenian. Similarly, confirmed side chain alkylated sterol biosynthesis in a bacterial isolate along with identification of any unique isotopic signatures resulting from the alkylation of sterol side chains by bacteria could help determine if bacteria may have been a potential source of Cryogenian 24-ipc. Until then, the co-occurrence of 24-ipc and 26-mes in high ratios relative to other steranes continues to provide strong evidence for derived sponges in the

Precambrian, consistent with molecular clock analyses (65–67). However, sterol side chain alkylation by symbiotic and free-living bacteria should now be considered both when analyzing the sterols of extant eukaryotes and when interpreting the presence of geologic steranes including ergostane, stigmastane, and 24-isopropylcholestane.

Materials and Methods

Bioinformatics Analysis. SMT homologs were identified in the Joint Genome Institute Integrated Microbial Genomes & Microbiomes (JGI IMG), GenBank, and Compagen (http://www.compagen.org) databases using the *Saccharomyces cerevisiae* S299C SMT amino acid sequence as the BLASTP search query (locus tag: YML008C, maximum e-value: $1e^{-50}$, minimum percent identity: 30%). Additional demosponge SMT sequences from *C. nucula*, *S. fasciculatas*, and *P. ficiformis* transcriptomes were provided by Prof. David A. Gold (38). Identifying sequence information is available in Table S1. Conserved sterol methyltransferase C-terminal and methyltransferase 11 domains were identified using the HmmerWeb tool (version 2.41.1).

Oxidosqualene cyclase homologs were identified through BLASTP searches of the metagenome scaffolds and MAGs containing the SMTs analyzed in this study and all bacterial and eukaryotic genomes available in JGI IMG as of October 2019 using the *Methylococcus capsulatus* Bath OSC protein sequence (locus tag: MCA2873, maximum e-value: $1e^{-50}$, minimum percent identity: 20%, minimum sequence length: 350 aa). Redundancy was decreased through manual removal of OSC homologs with 100% similarity at the amino acid level. The N-termini of fused SMO/OSC homologs were

trimmed out of the dataset to begin the sequences at the squalene-hopene cyclase N-terminal domain as identified using HmmerWeb.

All sequence datasets were aligned via MUSCLE. Conserved SMT domains were aligned separately, then concatenated using Geneious. Maximum-likelihood trees were generated using IQ-Tree on XSEDE with 1,000 ultra-fast bootstrap replicates using the best amino acid substitution models, gamma shape parameters, and invariable sites proportions under the Bayesian Information Criterion and Akaike Information Criterion using ModelTest-NG on XSEDE (Table S2). Resulting phylogenetic trees were visualized in iTOL (68).

All contigs containing metagenomic SMTs were retrieved from JGI IMG or GenBank and analyzed with Whokaryote using the default inputs with the minimum contig size decreased to 1000 base pairs. Calculated features and predictions are shown in Table S3.

Gene Synthesis and Molecular Cloning. SMT DNA sequences were codon-optimized for expression in *E. coli* and artificially synthesized by GeneArt (ThermoFisher Scientific) or through the Department of Energy Joint Genome Institute (DOE JGI) DNA Synthesis Science Program. SMT sequences from DOE JGI were obtained in the IPTG-inducible plasmid pSRKGm-*lac*UV5-rbs5 in *E. coli* TOP10. SMT sequences from GeneArt were subcloned into pSRKGm-*lac*UV5-rbs5 by sequence and ligase independent cloning (SLIC) (69) and transformed into *E. coli* DH10B as described in *SI Appendix*.

Bacterial Culture and Heterologous Expression. *E. coli* expression strains were cultured in 50 mL terrific broth (TB) supplemented with gentamycin (15 μg/mL) at 37 °C while shaking at 225 rpm. Cultures were induced with 500 μM isopropyl *β*-D-1-

thiogalactopyranoside (IPTG) at an OD_{600} of ~0.6, then incubated an additional 4 hours at 30 °C while shaking at 225 rpm. Cells were harvested by centrifugation at 4,500 × *g* for 10 minutes at 4 °C. Cell pellets were stored at -80 °C until sonication.

SMT assay. Cells pellets were resuspended in 5 mL buffer containing the following: 50 mM Tris-HCl, 2 mM $MgCl_2$, 20% glycerol (v/v), and 0.1% β-mercaptoethanol (v/v), pH 7.5. Cells were then lysed on ice via a Qsonica Sonicator Q500 equipped with a 3.2 mm probe at 30% amplitude pulsing at 5 seconds on, 15 seconds off for 8 minutes of total on time. Lysates were then partially clarified by centrifugation at 4,500 × *g* for 20 minutes at 4 °C. Total protein concentration in the resulting supernatant was quantified using a Coomassie (Bradford) Protein Assay Kit (Thermo Scientific) according to the manufacturer's Standard Microplate Protocol and a BioTek Synergy HT Microplate Reader. SMT assays were then immediately prepared using enough supernatant to give 4500 μg total protein. Other reaction components included 100 μM desmosterol (Sigma-Aldrich) or 24-methylenecholesterol (Avanti Polar Lipids, Inc.), 100 μM *S*-(5′-sdenosyl)-*L*-methionine chloride dihydrochloride (Sigma-Aldrich), and 0.1% Tween-80 (v/v) to a total reaction volume of 400 μL. Reactions were held at 30 °C for 20 hours and then stored at -20 °C until lipid extraction.

Lipid Extraction. Lipids were extracted using a modified Bligh-Dyer method (70, 71) with the completed in vitro reactions as the water phase. Reactions were sonicated in 10:5:4 (vol:vol:vol) methanol:dichloromethane:water for 1 hour. The organic phase was then separated with twice the volume of 1:1 (vol:vol) dichloromethane:water followed by storage at -20 °C for >1 hr. Following centrifugation at 2,800 x *g* for 10 min at 4 °C, the organic phase was transferred and evaporated under N_2 to give total lipid extracts (TLEs).

TLEs were derivatized to trimethylsilyl ethers in 1:1 (vol:vol) pyridine:Bis(trimethylsilyl)trifluoroacetamide for 1 hr at 70 °C prior to GC-MS analysis.

GC-MS Analysis. Lipids were separated with an Agilent 7890B Series GC equipped with two Agilent DB-17HT columns (30 m x 0.25 mm i.d. x 0.15 μm film thickness) in tandem with helium as the carrier gas at a constant flow of 1.1 ml/min and programmed as follows: 100°C for 2 min, then 12°C/min to 250°C and held for 10 min, then 10°C/min to 330°C and held for 17.5 min. 2 uL of each sample was injected in splitless mode at 250°C. The GC was coupled to an Agilent 5977A Series MSD with the ion source at 230°C and operated at 70 eV in EI mode scanning from 50 to 850 Da in 0.5 s. Lipids were identified based on retention time and comparison to laboratory standards and previously published spectra deposited in the National Institute of Standards and Technology (NIST) database and/or NMR analysis.

Acknowledgments

We thank Prof. David A. Gold for providing demosponge SMT protein sequences, Jeremy H. Wei for technical assistance, and Hanon McShea for helpful discussions. We also thank the Department of Energy Joint Genome Institute DNA Synthesis Science Program (Project 503267) for synthesis and cloning of the majority of the SMTs analyzed in this study. Funding for this study was provided by National Science Foundation Grant EAR-1752564 (to P.V.W).

References

1. T. M. Swan, K. Watson, Stress tolerance in a yeast sterol auxotroph: role of ergosterol, heat shock proteins and trehalose. *FEMS Microbiol Lett* 169, 191–197 (1998).

2. L. Miao, *et al.*, From lanosterol to cholesterol: structural evolution and differential effects on lipid bilayers. *Biophys J* 82, 1429–1444 (2002).

3. H. Schaller, The role of sterols in plant growth and development. *Prog Lipid Res* 42, 163–175 (2003).

4. A. B. Castoreno, *et al.*, Transcriptional regulation of phagocytosis-induced membrane biogenesis by sterol regulatory element binding proteins. *Proc Natl Acad Sci U S A* 102, 13129–13134 (2005).

5. P. Huang, *et al.*, Cellular cholesterol directly activates smoothened in hedgehog signaling. *Cell* 166, 1176-1187.e14 (2016).

6. K. E. Bloch, Sterol, structure and membrane function. *Crit Rev Biochem Mol Biol* 14, 47–92 (1983).

7. J. J. Brocks, *et al.*, Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. *Nature* 437, 866–870 (2005).

8. R. E. Summons, A. S. Bradley, L. L. Jahnke, J. R. Waldbauer, Steroids, triterpenoids and molecular oxygen. *Philos Trans R Soc Lond B Biol Sci* 361, 951–968 (2006).

9. I. Bobrovskiy, *et al.*, Ancient steroids establish the Ediacaran fossil *Dickinsonia* as one of the earliest animals. *Science* 361, 1246–1249 (2018).

10. J. J. Brocks, *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* 548, 578–581 (2017).

11. G. D. Love, *et al.*, Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* 457, 718–721 (2009).

12. J. A. Zumberge, *et al.*, Demosponge steroid biomarker 26-methylstigmastane provides evidence for Neoproterozoic animals. *Nat Ecol Evol* 2, 1709–1714 (2018).

13. W. Hofheinz, G. Oesterhelt, 24-isopropylcholesterol and 22-dehydro-24-isopropylcholesterol, novel sterols from a sponge. *Helv Chim Acta* 62, 1307–1309 (1979).

14. H. Q. Do, *et al.*, Unusual sterolic mixture, and 24-isopropylcholesterol, from the sponge *Ciocalypta* sp. reduce cholesterol uptake and basolateral secretion in Caco-2 cells. *J Cell Biochem* 106, 659–665 (2009).

15. M. Bortolotto, J. C. Braekman, D. Daloze, B. Tursch, Chemical studies of marine invertebrates. XXXVI. Strongylosterol, a novel C-30 sterol from the sponge *Strongylophora durissima* Dendy. *Bull Soc Chim Belges* 87, 539–543 (1978).

16. G. Barnathan, *et al.*, Unusual sterol composition and classification of three marine sponge families. *Boll Mus Ist Biol Univ Genova* 68, 201–208 (2004).

17. J.-L. Giner, Biosynthesis of marine sterol side chains. *Chem Rev* 93, 1735–1752 (1993).

18. J. A. Zumberge, "A lipid biomarker investigation tracking the evolution of the Neoproterozoic marine biosphere and the rise of eukaryotes," University of California, Riverside, Riverside, CA. (2019).

19. J. B. Antcliffe, R. H. T. Callow, M. D. Brasier, Giving the early fossil record of sponges a squeeze: the early fossil record of sponges. *Biol Rev* 89, 972–1004 (2014).

20. J. B. Antcliffe, Questioning the evidence of organic compounds called sponge biomarkers. *Palaeontology* 56, 917–925 (2013).

21. G. D. Love, R. E. Summons, The molecular record of Cryogenian sponges - a response to Antcliffe (2013). *Palaeontology* 58, 1131–1136 (2015).

22. J. B. Antcliffe, The oldest compelling evidence for sponges is still early Cambrian in age - reply to Love and Summons (2015). *Palaeontology* 58, 1137–1139 (2015).

23. G. D. Love, *et al.*, Sources of C30 steroid biomarkers in Neoproterozoic–Cambrian rocks and oils. *Nat Ecol Evol* 4, 34–36 (2020).

24. C. Hallmann, *et al.*, Reply to: Sources of C30 steroid biomarkers in Neoproterozoic–Cambrian rocks and oils. *Nat Ecol Evol* 4, 37–39 (2020).

25. I. Bobrovskiy, *et al.*, Algal origin of sponge sterane biomarkers negates the oldest evidence for animals in the rock record. *Nat Ecol Evol* 5, 165–168 (2021).

26. L. M. van Maldegem, *et al.*, Geological alteration of Precambrian steroids mimics early animal signatures. *Nat Ecol Evol* 5, 169–173 (2021).

27. W. C. M. C. Kokke, J. N. Shoolery, W. Fenical, C. Djerassi, Biosynthetic studies of marine lipids. 4. Mechanism of side chain alkylation in E-24-propylidenecholesterol by a Chrysophyte alga. *J Org Chem* 49, 3742–3752 (1984).

28. J.-L. Giner, C. Djerassi, Biosynthetic studies of marine lipids. 31. Evidence for a protonated cyclopropyl intermediate in the biosynthesis of 24-propylidenecholesterol. *J Am Chem Soc* 113, 1386–1393 (1991).

29. J.-L. Giner, X. Li, G. L. Boyer, Sterol composition of *Aureoumbra lagunensis*, the Texas brown tide alga. *Phytochemistry* 57, 787–789 (2001).

30. J.-L. Giner, H. Zhao, G. L. Boyer, M. F. Satchwell, R. A. Andersen, Sterol chemotaxonomy of marine pelagophyte algae. *C&B* 6, 1111–1130 (2009).

31. B. J. Nettersheim, *et al.*, Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. *Nat Ecol Evol* 3, 577–581 (2019).

32. I. Bobrovskiy, *et al.*, Algal origin of sponge sterane biomarkers negates the oldest evidence for animals in the rock record. *Nat Ecol Evol* 5, 165–168 (2021).

33. H. Gross, J. Reitner, GabrieleM. König, Isolation and structure elucidation of azoricasterol, a new sterol of the deepwater sponge *Macandrewia azorica*. *Naturwissenschaften* 91, 441–446 (2004).

34. M. W. Taylor, R. Radax, D. Steger, M. Wagner, Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *MMBR* 71, 295–347 (2007).

35. J. H. Wei, X. Yin, P. V. Welander, Sterol synthesis in diverse bacteria. *Front Microbiol* 7, 990 (2016).

36. N. S. Webster, M. W. Taylor, Marine sponges and their microbial symbionts: love and other relationships. *Environ Microbiol* 14, 335–346 (2012).

37. W. D. Nes, Enzyme mechanisms for sterol C -methylations. *Phytochemistry* 64, 75–95 (2003).

38. D. A. Gold, *et al.*, Sterol and genomic analyses validate the sponge biomarker hypothesis. *Proc Natl Acad Sci U S A* 113, 2684–2689 (2016).

39. D. A. Gold, *et al.*, Lipidomics of the sea sponge *Amphimedon queenslandica* and implication for biomarker geochemistry. *Geobiology* 15, 836–843 (2017).

40. W. Bergmann, F. H. McTigue, Contributions to the study of marine products. XXI. Chondrillasterol. *J Org Chem* 13, 738–741 (1948).

41. Y. Venkateswarlu, M. V. R. Reddy, M. R. Rao, A new epoxy sterol from the sponge *Ircinia fasciculata*. *J Nat Prod* 59, 876–877 (1996).

42. J. Fromont, S. Kerr, R. Kerr, M. Riddle, P. Murphy, Chemotaxonomic relationships within, and comparisons between, the orders Haplosclerida and Petrosida (Porifera: Demospongiae) using sterol complements. *Biochem Syst Ecol* 22, 735–752 (1994).

43. A. Aiello, E. Fattorusso, S. Magno, M. Menna, Isolation of five new 5α-hydroxy-6-keto-Δ7 sterols from the marine sponge *Oscarella lobularis*. *Steroids* 56, 337–340 (1991).

44. P. R. Bergquist, W. Hofheinz, W. Hofheinz, G. Oesterhelt, Sterol composition and the classification of the demospongiae. *Biochem Syst Ecol* 8, 423–435 (1980).

45. C. Santana-Molina, E. Rivas-Marin, A. M. Rojas, D. P. Devos, Origin and evolution of polycyclic triterpene synthesis. *Mol Biol Evol* 37, 1925–1941 (2020).

46. L. Villanueva, W. I. C. Rijpstra, S. Schouten, J. S. S. Damsté, Genetic biomarkers of the sterol-biosynthetic pathway in microalgae. *Environ Microbiol Rep* 6, 35–44 (2014).

47. L. J. U. Pronk, M. H. Medema, Whokaryote: distinguishing eukaryotic and prokaryotic contigs in metagenomes based on gene structure. *bioRxiv*, 2021.11.15.468626 (2021).

48. P. R. Bergquist, A. Lavis, R. C. Cambie, Sterol composition and classification of the porifera. *Biochem Syst Ecol* 14, 105–112 (1986).

49. J.-L. Giner, C. Djerassi, Biosynthetic studies of marine lipids-XXVIII. Use of sponge cell-free extracts in the study of marine sterol biosynthesis. *Tetrahedron Lett* 31, 5421–5424 (1990).

50. R. G. Kerr, I. L. Stoilov, J. E. Thompson, C. Djerassi, Biosynthetic studies of marine lipids 16. De novo sterol biosynthesis in sponges. Incorporation and transformation of cycloartenol and lanosterol into unconventional sterols of marine and freshwater sponges. *Tetrahedron* 45, 1893–1904 (1989).

51. C. J. Silva, L. Wünsche, C. Djerassi, Biosynthetic studies of marine lipids 35. The demonstration of de novo sterol biosynthesis in sponges using radiolabeled isoprenoid precursors. *Comp Biochem Physiol B* 99B, 763–773 (1991).

52. J. Germer, N. Cerveau, D. J. Jackson, The holo-transcriptome of a calcified early branching metazoan. *Front Mar Sci* 4, 81 (2017).

53. K. Ganapathy, *et al.*, Molecular probing of the *Saccharomyces cerevisiae* sterol 24-C methyltransferase reveals multiple amino acid residues involved with C2-transfer activity. *Biochim Biophys Acta Mol Cell Biol Lipids* 1781, 344–351 (2008).

54. W. D. Nes, *et al.*, Site-directed mutagenesis of the sterol methyl transferase active site from *Saccharomyces cerevisiae* results in formation of novel 24-ethyl sterols. *J Org Chem* 64, 1535–1542 (1999).

55. W. D. Nes, *et al.*, Active site mapping and substrate channeling in the sterol methyltransferase pathway. *J Biol Chem* 277, 42549–42556 (2002).

56. B. A. Haubrich, *et al.*, Characterization, mutagenesis and mechanistic analysis of an ancient algal sterol C24-methyltransferase: implications for understanding sterol evolution in the green lineage. *Phytochemistry* 113, 64–72 (2015).

57. W. D. Nes, *et al.*, Probing the sterol binding site of soybean sterol methyltransferase by site-directed mutagenesis: functional analysis of conserved aromatic amino acids in region 1. *Arch Biochem Biophys* 448, 23–30 (2006).

58. J. F. Hu, *et al.*, Seawater incursion events in a Cretaceous paleo-lake revealed by specific marine biological markers. *Sci Rep* 5, 9508 (2015).

59. M. McCaffrey, *et al.*, Paleoenvironmental implications of novel C30 steranes in Precambrian to Cenozoic age petroleum and bitumen. *Geochim Cosmochim Acta* 58, 529–532 (1994).

60. J. M. Moldowan, C30-steranes, novel markers for marine petroleums and sedimentary rocks. *Geochim Cosmochim Acta* 48, 2767–2768 (1984).

61. C. A. Elwell, J. N. Engel, Lipid acquisition by intracellular Chlamydiae. *Cell Microbiol* 14, 1010–1018 (2012).

62. L. Wünsche, F. O. Gülaçar, A. Buchs, Several unexpected marine sterols in a freshwater sediment. *Org Geochem* 11, 215–219 (1987).

63. I. L. Stoilov, J. E. Thompson, J. Ho. Cho, Carl. Djerassi, Biosynthetic studies of marine lipids. 9. Stereochemical aspects and hydrogen migrations in the biosynthesis of the triply alkylated side chain of the sponge sterol strongylosterol. *J Am Chem Soc* 108, 8235–8241 (1986).

64. C. A. N. Catalan, J. E. Thompson, W. C. M. C. Kokke, C. Djerassi, Biosynthetic studies of marine lipids—3: experimental demonstration of the course of side chain extension in marine sterols. *Tetrahedron* 41, 1073–1084 (1985).

65. K. J. Peterson, J. A. Cotton, J. G. Gehling, D. Pisani, The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philos Trans R Soc Lond B Biol Sci* 363, 1435–1443 (2008).

66. D. H. Erwin, *et al.*, The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334, 1091–1097 (2011).

67. M. Dohrmann, G. Wörheide, Dating early animal evolution using phylogenomic data. *Sci Rep* 7, 3599 (2017).

68. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49, W293–W296 (2021).

69. M. Z. Li, S. J. Elledge, SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol* 852, 51–59 (2012).

70. P. V. Welander, *et al.*, Identification and characterization of *Rhodopseudomonas palustris* TIE-1 hopanoid biosynthesis mutants. *Geobiology* 10, 163–177 (2012).

71. E. G. Bligh, W. J. Dyer, A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37, 911–917 (1959).

Figures and Tables

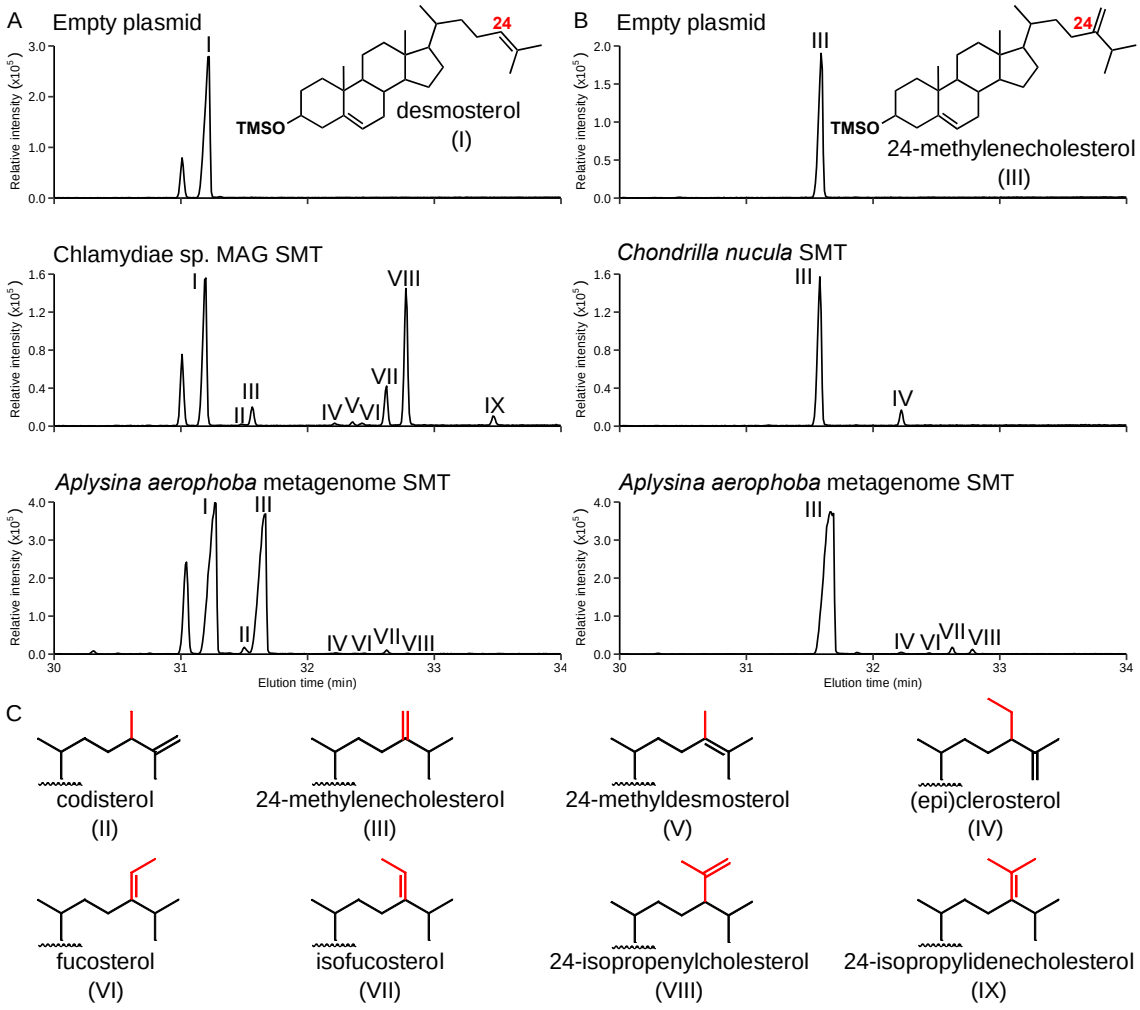

Figure 1. Sterol methyltransferases from sponges and bacteria perform multiple side chain methylations. (*A*) Representative extracted ion chromatograms (*m/z* 456, 470, 484, 498) of total lipid extracts from in vitro reactions performed with *E. coli* lysates containing an empty plasmid (*Upper*), a sterol methyltransferase (SMT) identified in a freshwater metagenome-assembled genome assigned to the phylum Chlamydiae (*Middle*), and an SMT identified in a metagenome from the microbiome of the demosponge *Aplysina aerophoba* (*Lower*) with desmosterol as the substrate. The Chlamydiae MAG SMT produces codisterol, 24-methylenecholesterol, 24-methyldesmosterol, epiclerosterol,

fucosterol, isofucosterol, 24-isopropenylcholesterol, and 24-isopropylidenecholesterol. The *A. aerophoba* metagenome SMT produces codisterol, 24-methylenecholesterol, epiclerosterol, fucosterol, isofucosterol, and 24-isopropenylcholesterol from desmosterol. (*B*) Representative extracted ion chromatograms (*m/z* 456, 470, 484, 498) of total lipid extracts from in vitro reactions performed with *E. coli* lysates containing an empty plasmid (*Upper*), an SMT identified in a transcriptome of the demosponge *Chondrilla nucula* (*Middle*), and the same SMT identified in the *A. aerophoba* metagenome as in (A) (*Lower*) with 24-methylenecholesterol as the substrate. The *C. nucula* SMT produces (epi)clerosterol, and the *A. aerophoba* metagenome SMT produces the same sterols as in (A). (*C*) Side chain structures of sterols identified in (A) and (B). All lipids were derivatized to trimethylsilyls. Mass spectra of identified sterols are shown in *SI Appendix*, Fig. S1.

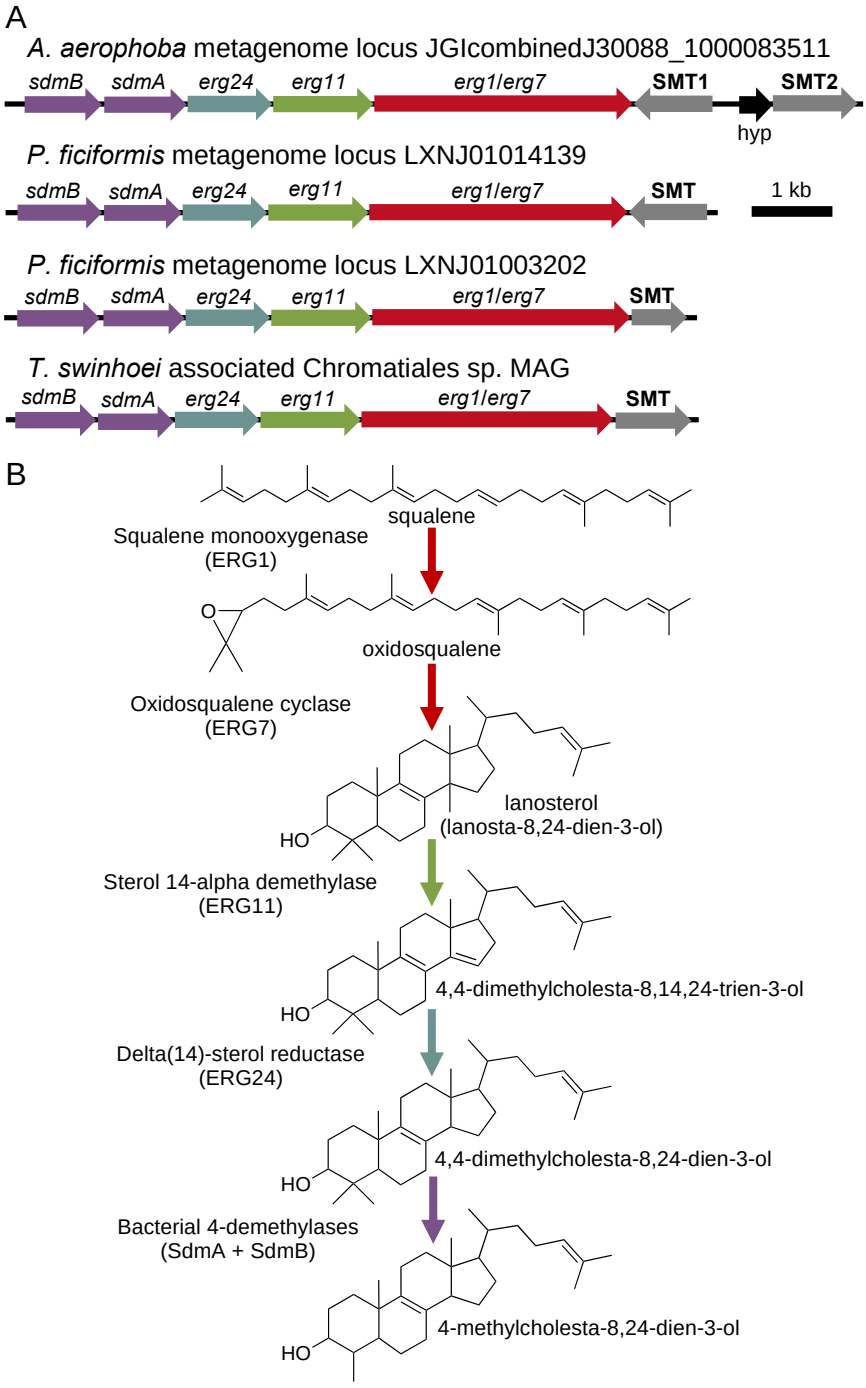

Figure 2. Sterol gene clusters in demosponge metagenomes suggest de novo biosynthesis by bacterial symbionts. (*A*) Putative sterol biosynthesis gene clusters identified in metagenomes from the microbiomes of the demosponge species *Aplysina aerophoba*, *Petrosia ficiformis*, and *Theonella swinhoei*, each containing a sterol methyltransferase

(gray arrows). Each gene cluster contains homologs of a fused squalene monooxygenase/oxidosqualene cyclase (red arrows), sterol 14-alpha demethylase (green arrows), delta(14)-sterol reductase (blue arrows), and the bacterial sterol demethylases *sdmA* and *sdmB* (purple arrows). *(B)* Schematic of the bacterial sterol biosynthesis pathway from squalene to 4-methylcholesta-8,24,-dien-3-ol. Arrow colors at each step correspond to the putative homologs identifiied in (A). Detailed information on the source of each metagenomic locus is available in *SI Appendix*, Table S1.

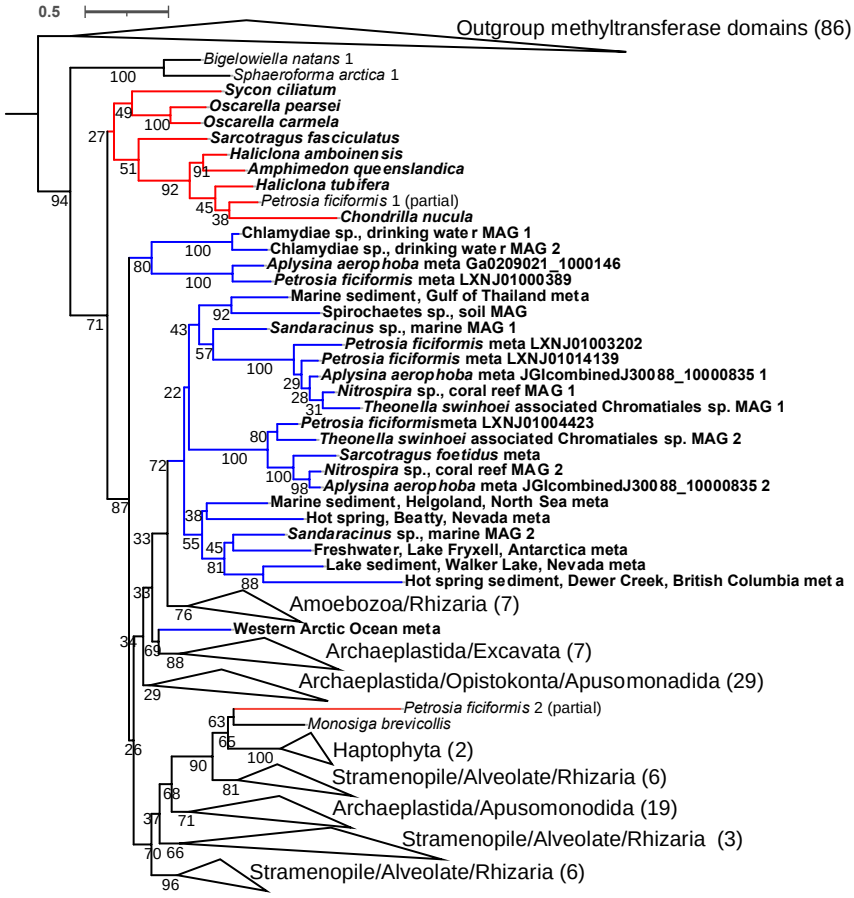

Figure 3. Maximum likelihood tree of sterol methyltransferases. The majority (23 out of 24) of the putative bacterial SMTs tested here (blue) form two clades distinct from sponge SMTs (red) and other eukaryotic SMTs (black), although bootstrap support is low. The numbers in parentheses indicate the number of proteins in collapsed clades. Bolded labels indicate SMTs tested in this study.

Table 1. Sterols identified as products of sponge and bacterial sterol methyltransferases.

	C_{28}			C_{29}			C_{30}	
	codisterol (II)	24-methylenecholesterol (III)	24-methyldesmosterol (V)	(epi)clerosterol (IV)*	fucosterol (VI)	isofucosterol (VIII)	24-isoprenylcholesterol (VIII)	24-isopropylidenecholesterol (IX)
DEMOSPONGIAE								
Amphimedon queenslandica		+			+			
Chondrilla nucula	+	+		+				
Sarcotragus fasciculatus	+	+			+			
Haliclona amboinensis	+	+		+	+	+		
Haliclona tubifera	+	+		+				
HOMOSCLEROMORPHA								
Oscarella pearsei	+	+		+	+	+		
Oscarella carmela	+	+						
CALCAREA								
Sycon ciliatum	+	+						
DEMOSPONGE METAGENOME								
Aplysina aerophoba JGIcombinedJ30088_10000835 1 of 2	+	+		+	+	+	+	
Aplysina aerophoba JGIcombinedJ30088_10000835 2 of 2								
Aplysina aerophoba Ga0209021_1000146	+	+	+		+	+		+
Petrosia ficiformis LXNJ01014139		+						
Petrosia ficiformis LXNJ01003202		+						
Petrosia ficiformis LXNJ01000389	+	+	+		+			
Petrosia ficiformis LXNJ01004423								
Sarcotragus foetidus		+						
DEMOSPONGE METAGENOME-ASSEMBLED GENOME								
Theonella swinhoei associated Chromatiales sp. 1 of 2		+						
Theonella swinhoei associated Chromatiales sp. 2 of 2								
OTHER METAGENOME								
Marine sediment, Gulf of Thailand		+						
Marine sediment, Helgoland, North Sea		+						
Western Arctic Ocean	+	+		+	+	+		
Freshwater, Lake Fryxell, Antarctica	+	+			+	+		
Lake sediment, Walker Lake, Nevada		+						
Hot spring, Beatty, Nevada								
Hot spring sediment, Dewer Creek, British Columbia		+						
OTHER METAGENOME-ASSEMBLED GENOME								
Chlamydiae sp., drinking water 1 of 2	+	+	+	+	+	+	+	+
Chlamydiae sp., drinking water 2 of 2	+		+					
Nitrospira sp., coral reef 1 of 2	+	+		+				
Nitrospira sp., coral reef 2 of 2								
Sandaracinus sp., marine 1 of 2	+	+						
Sandaracinus sp., marine 2 of 2	+	+		+	+	+		
Spirochaetes sp., soil		+						

*NMR confirmed epiclerosterol and not clerosterol as products of the "*Aplysina aerophoba* JGIcombinedJ30088_10000835 1 of 2" and "Chlamydiae sp., drinking water 1 of 2" sterol methyltransferases.