 doi:10.20944/preprints202004.0186.v5

The eukaryotic last common ancestor was bifunctional for hopanoid and sterol production

Warren R. Francis[1]

[1] Department of Biology, University of Southern Denmark, Odense, Denmark

Abstract

Steroid and hopanoid biomarkers can be found in ancient rocks and may give a glimpse of what life was present at that time. Sterols and hopanoids are produced by two related enzymes, though the evolutionary history of this protein family is complicated by losses and horizontal gene transfers, and appears to be widely misinterpreted. Here, I have added sequences from additional species, and re-analysis of the phylogeny of SHC and OSC indicates a single origin of both enzymes among eukaryotes. This pattern is best explained by endosymbiotic acquisition of both enzymes from a bacterial ancestor, followed by widespread loss of SHC, and two subsequent HGT events to ferns and ascomycetes. Thus, the last common ancestor of eukaryotes would have been bifunctional for both sterol and hopanoid production. Later enzymatic innovations allowed diversification of sterols in eukaryotes. Contrary to previous interpretations, the last eukaryotic common ancestor (LECA) potentially would have been able to produce hopanoids as a substitute for sterols in anaerobic conditions. Without invoking any other metabolic demand, the LECA could have been a facultative aerobe, living in unstable conditions with respect to oxygen level.

Keywords: sterol; hopanoid; proterozoic; biosynthesis; squalene

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

Introduction

Cells maintain fluidity of membranes through hopanoids and sterols (Rohmer et al., 1979), including molecules like cholesterol. Compared to many other biological molecules, the structure of hopenes and sterols makes them extremely stable over long time scales (Ourisson and Albrecht, 1992). Through geological processes, these two types of compounds become chemically modified, being converted into reduced forms called hopanes and steranes, respectively (JJ Brocks and Pearson, 2005). These molecules (often referred to as "biomarkers") may persist for millions of years, and are detectable in some ancient rocks, giving a window into what organisms were present at the time the rock was formed (JJ Brocks and Pearson, 2005).

Hopanoids and sterols are produced by a series of shared enzymatic steps (Figure 1), but diverging at one critical step where oxygen is required to produce sterols but not hopenes (Fischer and Pearson, 2007; Nes, 2011). The enzyme squalene monooxygenase (SQMO, also called SQE/SQLE for squalene epoxidase, or ERG1 from the yeast homolog involved in the ergosterol biosynthetic pathway) requires oxygen to produce an epoxide on squalene, changing the downstream specificity and products (Nes, 2011). One of two enzymes then forms the multi-ring structure, either using squalene as the reactant for squalene-hopene cyclase (SHC) to make hopanoids, or oxidosqualene cyclase (OSC) acting on oxido-squalene to make sterols. Despite the different activity, the two enzymes are nonetheless similar at the sequence level. For instance, the bacterium *Methylococcus capsulatus* is one of the few known organisms that possesses both enzymes, which show 28% identity to each other. This is comparable to the 27% identity measured between human OSC and SHC from *Alicyclobacillus acidocaldarius*, the two members of this protein family with crystal structures (Thoma et al., 2004; Wendt et al., 1999).

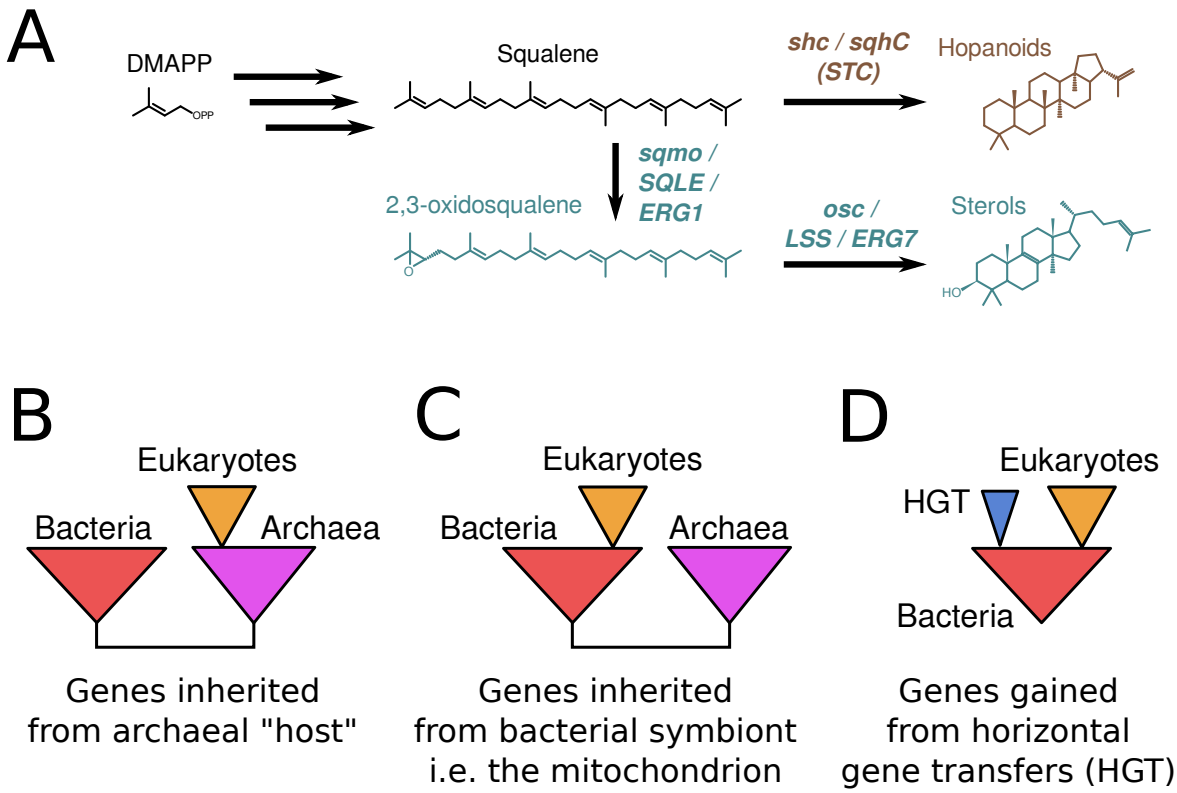
A simplified view postulates that hopanoids are produced by bacteria, while sterols are produced by eukaryotes. However, some bacteria actually produce sterols (Jahnke et al., 1992; Lamb et al., 2007; Pearson et al., 2003; Wei et al., 2016) and some eukaryotes produce hopanoids (Kemp et al., 1984; Mallory et al., 1963), suggesting that the traditional interpretation is an oversimplification (see Figure 2A). The phylogenetic distribution is complicated, involving many gains, losses, and horizontal gene transfers (HGT). Hypotheses have been proposed to explain several aspects of the sterol biosynthesis, albeit referring to different parts of the phylogenetic history, while other parts remain uncertain:

1. HGT of sterol biosynthesis (of OSC and SQMO) from stem-eukaryotes to some bacteria (referred to as Bacterial Group 1 by (Gold et al., 2017) .)
2. HGT of sterol biosynthesis (of OSC and SQMO) from stem-SAR group or stem-Archaeplastida to other bacteria (some gammaproteobacteria, referred to as Bacterial Group 2 by (Gold et al., 2017; Wei et al., 2016) .)
3. HGT of SHC from cyanobacteria to ferns (Takishita, Chikaraishi, Leger, et al., 2012). This reconciled observations of hopanoids in ferns (Pandey and Mitra, 1969; Zander et al., 1969), as well as the presence of both enzymes (Shinozaki et al., 2008).
4. HGT of SHC from proteobacteria (represented only by *Anaeromyxobacter*) to ascomycete fungi (Takishita, Chikaraishi, Leger, et al., 2012).

The latter three of these make sense with respect to the tree. They show a restricted distribution (e.g. only in ferns, not to all plants or all archaeaplastida), potentially showing different origins (cyanobacteria vs. proteobacteria), and cannot be adequately explained by endosymbiotic acquisition from a putative bacterial ancestor at the origin of the mitochondrion. Even for the transfer from SAR eukaryotes to bacteria, the position in the phylogenetic tree from (Gold et al., 2017) is unambiguous, that is, the tree cannot be rotated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

Figure 1. Sterol and hopanoid biosynthesis.



(**A**) Schematic of sterol and hopanoid biosynthesis. Steps between dimethylallyl pyrophosphate (DMAPP) and squalene are not shown for clarity. Gene names refer to orthologs across multiple model organisms, so there is no consistent naming scheme. Squalene-tetrahymanol cyclase (STC) does not strictly make hopene, but is included for clarity and would group with SHC rather than OSC. (**B/C/D**) Example trees showing possible relationships between eukaryotic orthologs relative to a proposed endosymbiotic event, with the gene inherited from the archaeal host (**B**), the proto-mitochondrion (**C**), or secondarily gained in select groups from a later horizontal gene transfer (**D**).

in a way to place the bacterial group outside of eukaryotes.

The first HGT, however, fails to explain the origin of OSC and has an equally parsimonious alternative. In that case, HGT is proposed from a stem eukaryote to some bacteria, though as the origin is unknown, the direction of exchange could go either way and would be equally parsimonious, from bacteria to eukaryote, or vice versa. The split between eukaryotic and bacterial OSCs is one node from the root of the tree, so could be explained in either direction. Given that current endosymbiotic theory proposes a major transfer of genes from a bacterium to an “early” eukaryote, that is, the origin of eukaryotes via endosymbiosis of the mitochondrion, it would make the most sense for OSC to be one of the thousands of genes acquired this way from a bacterium. All else being equal, the presence of OSC in some bacteria and stem eukaryotes is nonetheless best explained by OSC originating in bacteria, and then being inherited by a pre-eukaryotic host from a bacterial endosymbiont.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

Material and methods

Sequence searches

The initial dataset combined all SHC, STC, and OSC sequences from Takishita, Chikaraishi, Leger, et al. (2012) (83 sequences) and Gold et al. (2017) (6 sequences). Additional eukaryotic sequences (appx. 130 species/OTUs) were added from the MMETSP transcriptome collection (Keeling et al., 2014), using a database of translated proteins. Additional related sequences were then downloaded from GenBank after being identified on the BLASTP web server, with the query sequence as: human LSS (OSC/ERG7 homolog), *Saccharomyces cerevisiae* ERG7, *Adiantum capillus-veneris* (both SHC and OSC), *Methylococcus capsulatus* (both SHC and OSC), *Gemmata obscuriglobus* (both SHC and OSC), *Tetrahymena thermophila* (STC, `XP_001026696.2`), *Penicillium camemberti* (SHC, `CRL29798.1`), and *Nostoc commune* (SHC, `WP_109008080.1`). A relaxed e-value threshold was initially used (`1e-5`), but all partial sequences were removed. In total, the final gene set contained 488 sequences.

Phylogenetic analysis

Sequences were aligned with MAFFT v7.313 (Katoh et al., 2017) using the option `--ginsi`. The final tree was built using IQ-TREE v1.6.12 (Nguyen et al., 2015), using the options `-m LG+R10` and `-bb 1000 -bnni`, with otherwise default parameters. An additional tree was inferred using a trimmed alignment, where sites with an occupancy under 20 (meaning $> 95\%$ gaps) were removed with the Python script `trim_alignment_by_coverage.py` (Francis, 2022). This trimmed out 1120 out of the original 3764 sites. The trimmed alignment was then used as input for IQ-TREE, as above.

All raw data files are available online (Francis, 2022).

Results

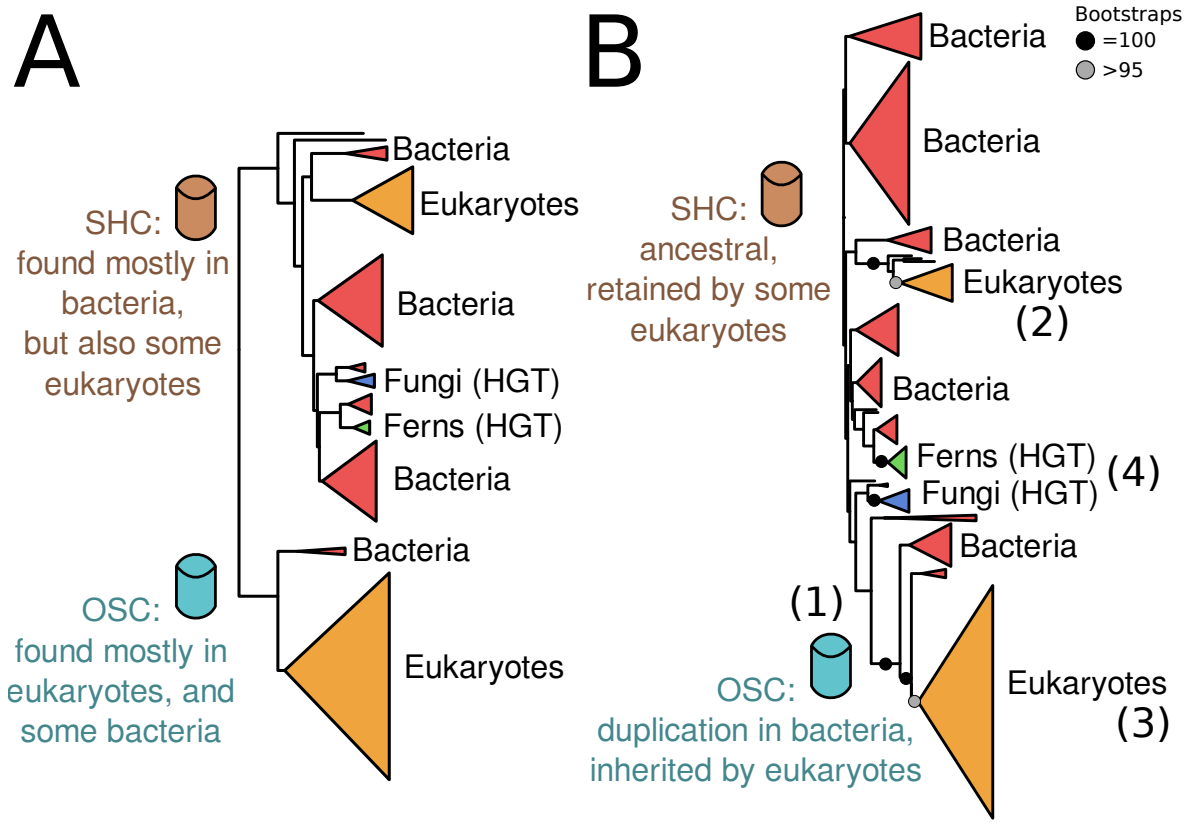
Additional species still recover monophyletic groups for both eukaryotic STC and OSC

Beginning from the dataset used by Takishita, Chikaraishi, Leger, et al. (2012), 405 additional sequences were added. Phylogenetic analysis (see methods) broadly recovers the same topology (Figure 2), though the question of where to place the root became increasingly important for subsequent interpretations. The backbone of the tree is poorly supported at most nodes (as it was for Takishita, Chikaraishi, Leger, et al. (2012)), making it difficult to decide on a root or confidently infer anything from the arrangement of groups.

For SHC and OSC, the root is commonly placed between the two enzymatic groups based on activity (Desmond and Gribaldo, 2009; Frickey and Kannenberg, 2009; Gold et al., 2017; Takishita, Chikaraishi, Leger, et al., 2012). Even if unintended by the authors, this implies a parallel origin of the two enzymes relative to the unknown outgroup, which would be outside of both bacteria and eukaryotes - potentially in the last universal common acenstor (LUCA). Such scenario would be plausible if SHC were present in early Bacteria and OSC were present in early Archaea. However, given the complete absence of either enzyme in known Archaea, this is unlikely. Thus, the tree should not be rooted at this position, or at least should not be drawn this way. The programs that generate phylogenetic trees typically run in a way that produces an unrooted tree (called time-reversible), where the branching and length (broadly speaking) display the weighted number of substitutions to convert one sequence into another. Walking from one end of the tree to the other implicitly is a walk back in time to the root, then forward again to another leaf. The user of the program then decides to place the root based on *a priori* information about the species or genes in question. This may, however, give the impression that the true divide in this enyzme family is between SHC and OSC, which may be the result of functional divisions rather than evolutionary history.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

Figure 2. Schematic tree of SHC and OSC evolution.



(**A**) Tree from (Takishita, Chikaraishi, Leger, et al., 2012), showing the commonly-drawn root position as was done by (Takishita, Chikaraishi, Leger, et al., 2012). This portrays the two proteins as having parallel origins, from an unknown source (that could be LUCA), hence fails to explain the origins of either SHC or OSC. (**B**) Full-alignment tree with added taxa, showing an alternate rooting among bacteria (currently unknown where within bacteria). Key bootstrap values relevant for eukaryotes are shown as black or gray dots. Four key events should be noted: SHC was the original enzyme and (1) a duplication within an ancient bacterial lineage produced OSC from SHC, resulting in some, but not all, bacterial lineages having OSC. (2) Multiple lineages of anaerobic eukaryotes still retain SHC/STC showing a single origin, which is sister group to bacteria. This is explained by acquisition from bacteria at the origin of eukaryotes (probably endosymbiosis). (3) The original eukaryote inherited OSC from the endosymbiont, and the bulk of eukaryotes retain this enzyme. (4) Two additional horizonal gene transfer events gave SHC back to two separate eukaryotic lineages from two different bacterial groups.

Discussion

How did sterol biosynthesis evolve?

If bacteria and hopanoids appeared first on this planet, then it is very likely that the root should be set somewhere within bacteria, as argued by (Santana-Molina et al., 2020). This means that the root of the tree is not the split of SHC and OSC, but somewhere among the bacterial lineages producing hopanoids.

Starting from this root, the evolutionary history may be more adequately imagined as a duplication followed by many losses:

1. SHC was the original enzyme, distributed across many bacterial lineages (though not necessarily all).
2. OSC arose from a gene duplication in some proteobacterial ancestor.
3. Eukaryotes inherited both SHC and OSC from a bacterial endosymbiont, and would have been capable

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

of producing both hopanoids and sterols.

4. Either SHC or OSC was subsequently lost in all eukaryotes, leaving some anaerobic eukaryotes with only SHC and most eukaryotes with only OSC.
5. HGT events occurred in both directions, resulting in a gain of SHC for ferns and ascomycete fungi from two different bacterial groups, and a gain of OSC for some gammaproteobacteria from a stem-SAR eukaryote.
6. Secondary losses of OSC occur widely, as some organisms scavenge or acquire the molecules from their diets.

The steps of this hypothesis are detailed below.

SHC was the original enzyme

Assuming that bacteria existed before the origin of eukaryotes, it therefore makes the most sense to assume that SHC existed first. The time period of this is not clear, though many bacterial lineages likely had this enzyme very early in time (Santana-Molina et al., 2020), though others likely acquired it more recently from horizontal gene transfer between prokaryotic lineages. Nonethless, clearly some ancient bacteria could produce hopanoids from squalene with SHC.

A gene duplication produced OSC from SHC

Next, a duplication of the SHC gene would have produced two copies in some ancient bacterial lineage (marker **(1)** in Figure 2B). Initially, these two copies would produce the same products (diploptene) from the same substrate (squalene). As these two copies diverged in sequence, the preference for substrate changed and so did the reaction mechanism and products, thus one of the two became OSC and allowed for production of sterols. It is very likely that this was coincident with the origin of SQMO. Some other enzymes may have originated later, like the C14-demthylase (called ERG11 in yeast, CYP51A1 in mammals), as these enzymes are absent in some bacteria that produce sterols (Pearson et al., 2003).

The identity of the bacteria with the duplication is uncertain, though some things can be learned from the presence of both genes in some species, and from the tree itself. Several bacteria have been shown to possess both genes in the genomes, including the planctomycete *Gemmata obscuriglobus*, the gamma-proteobacterium *Methylococcus capsulatus*, and the delta-proteobacteria *Plesiocystis pacifica* and *Enhygromyxa salina*. The results of the phylogenetic inference place the SHC from *Gemmata obscuriglobus* firmly within other planctomycetes, but not so for the OSC, which is sandwiched between actinobacteria and proteobacteria. Thus, the presence of OSC in *Gemmata obscuriglobus* may represent a more recent horizontal transfer, rather than inheritance from the original duplication. Thus, this duplication may be better considered to have originated somewhere in the proteobacteria.

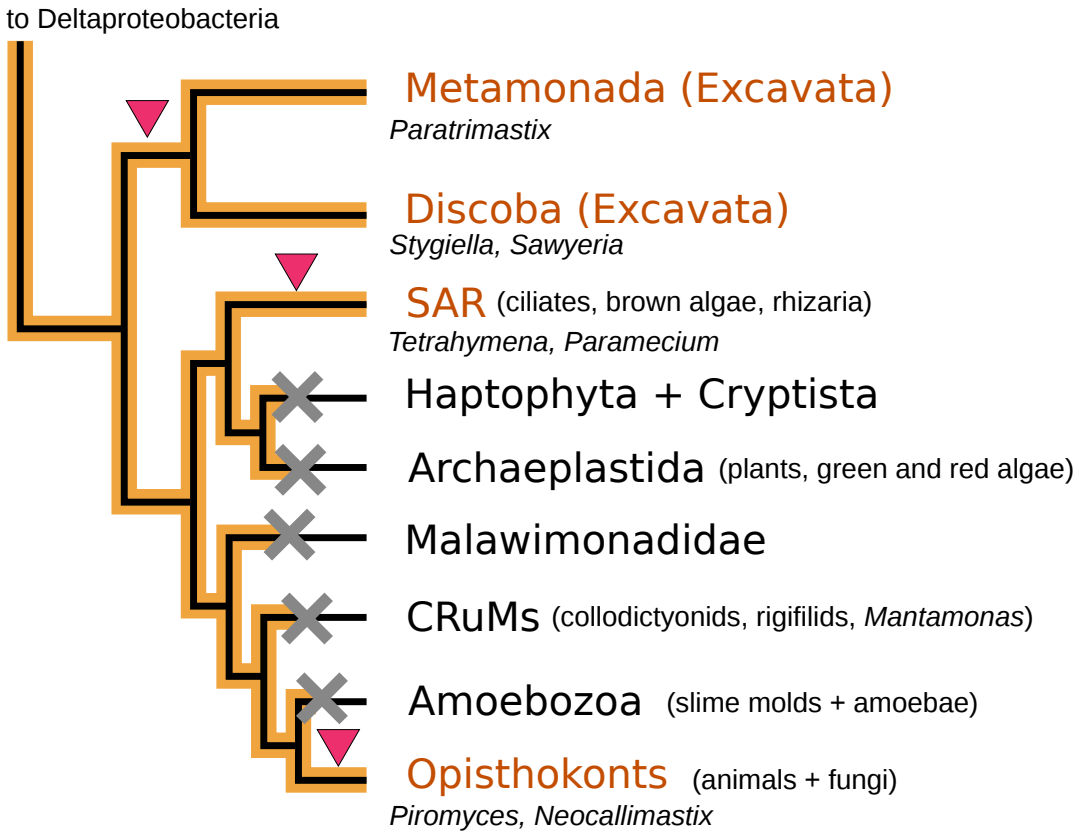
Eukaryotes inherited both enzymes, probably at the same time

The core eukaryotic group of SHC/STC (see marker **(2)** in Figure 2B, while those from HGT are **(4)**) includes representatives from only three eukaryotic lineages yet still shows a single origin; the same goes for eukaryotic OSC. That is, despite the relative paucity of SHC in eukaryotes, the groups that still contain the enzyme include opisthokonts, ciliates, and some excavata. If these three lineages are mapped onto current models of the species tree of eukaryotes (see (Brown et al., 2018; Derelle et al., 2015)), endosymbiotic acquisition of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

SHC in those three lineages would parsimoniously indicate that the last eukaryotic common ancestor had SHC (Figure 3). Even some alternate arrangements (Derelle et al., 2015; Ren et al., 2016) would require that effectively all eukaryotic groups originally had the enzyme.

Figure 3. Schematic tree of presence and absence of SHC/STC in eukaryotes.



Recreated tree from (Brown et al., 2018), showing major clades of eukaryotes. The observed pattern of presence and absence is nonetheless best explained by a single origin of SHC/STC, retention by 3 lineages (4 groups), and 5 or more losses (indicated by gray 'X's). Even alternate topologies (Derelle et al., 2015; Ren et al., 2016) would still require that the LCA of most or all eukaryotic groups originally had SHC/STC, mostly due to the presence of STC in ciliates and Neocallimastigomycota fungi. Otherwise, three HGT events would need to occur at the red triangles. SAR group refers to the clade of Stramenopiles, Alveolates, and Rhizaria. CRuMs group refers to Collodictyonids, Rigifilids, and *Mantamonas*.

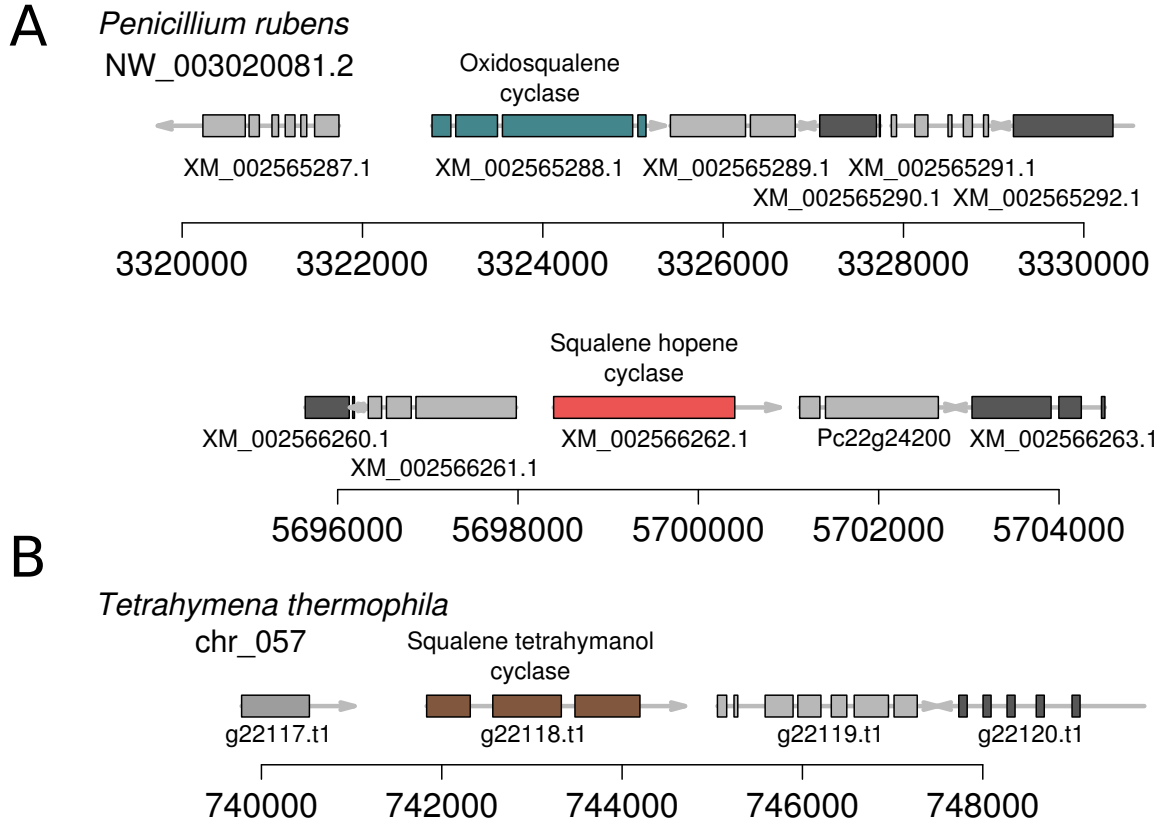
It was proposed that multiple HGT events to eukaryotes allowed survival in oxygen-depleted environments (Takishita, Chikaraishi, Leger, et al., 2012). However, HGT is seen in other parts of this tree (marker **(4)**), as discussed above and below. The pattern seen for eukaryotic STC (marker **(2)**) would therefore require three HGT events from the same bacterial group, close enough in time that the inferred phylogenetic tree would give them a single, strongly-supported origin. That is, three HGT events would have had to occur and give the *illusion* of a single origin by our current phylogenetic models.

This is unlikely for several reasons. Firstly, in the scheme proposed by (Takishita, Chikaraishi, Leger, et al., 2012), the only evidence for HGT was that the STC group is nested within bacteria genes. As 7 sequences were present in their analysis, contamination from the same group would be unlikely. This does not rule out problems with tree inference or endosymbiotic acquisition. Horizontal gene transfer should only be invoked when it is clear that there is no possibility of primary inheritance in the LECA from a bacterial endosymbiont at the origin of mitochondria. The broad distribution of eukaryotic groups with a strongly-supported single origin suggests the presence of SHC/STC in the LECA, thus endosymbiotic acquisition cannot be ruled out.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

Endosymbiotic acquisition and horizontal gene transfer occur in different parts of the tree, and have distinct signatures. Oftentimes, intronless genes may result from recent HGT. In the ciliate *Tetrahymena thermophila*, the STC gene still contains introns (Figure 4B), suggesting at least that this gene did not come from a *recent* HGT event. However, the SHC gene in the fungus *Penicillium rubens* (Figure 4A) is a single exon, in agreement with a recent HGT event. However, introns are found in OSC gene in the same species. It should be stated that the absence of introns is not definitely a sign of horizontal gene transfer, or of the ancestral state. For instance, the human ortholog of OSC (LSS, https://www.ncbi.nlm.nih.gov/gene/4047) contains 22 exons, while the yeast ortholog (ERG7, https://www.yeastgenome.org/locus/S000001114) is a single exon.

Figure 4. Genomic context of triterpenoid cyclases



Scale is the same across all loci, showing 10kb windows. (**A**) Genomic loci of OSC and the horizontally transferred SHC in the fungus *Penicillium rubens* (NCBI Assembly: `GCF_000226395.1`), both on the same scaffold. (**B**) Genomic locus of STC in the ciliate *Tetrahymena thermophila* (`v2.2.47` from ciliate.org).

Secondly, prokaryote-eukaryote HGT is not common enough for three parallel HGTs to make sense, nor is there a satisfying explanation as to why these lineages would all take a nearly identical enzyme from the same bacterial source, especially as that clearly was not the case for ferns or fungi. Even if HGT were still invoked only once to a stem eukaryote, this would parsimoniously be equal to primary inheritance in the LECA from a bacterial endosymbiont.

Eukaryotes lost either SHC or OSC

According to the most parsimonious view of the tree (Figure 2B), the ancestor of all eukaryotes then must have had both SHC and OSC. However, either one or the other gene was later lost in all lineages, with two subsequent exceptions (discussed below). In other words, ignoring the HGT events, no known eukaryote

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2022 doi:10.20944/preprints202004.0186.v5

today possesses both enzymes. The lineages with SHC are precisely those that are lacking OSC, and vice versa. This shows that species in each lineage kept either SHC/STC or OSC, but never both.

Two questions may be considered: why was one of the two enzymes lost in all eukaryotic lineages? why did most lineages retain OSC?

1. Given the high similarity of the compounds produced, plausibly some enzyme products could interfere with the production of others. This precise effect was observed in the ciliate *Tetrahymena pyriformis*, where cholesterol was shown to inhibit production of tetrahymanol from labeled squalene,(Conner et al., 1968) possibly by affecting gene expression, while other steroid compounds were shown to inhibit the cyclization in vitro,(Sipe and Holmlund, 1972) likely by directly inhibiting the enzyme. Thus, competitive inhibition of one enzyme with products from the other may have affected the utility of having both enzymes.

   Lanosterol is heavily processed to produce a large variety of different sterols across many different groups (Nes, 2011). The planctomycete *Gemmata obscuriglobus* was shown to produce only lanosterol (Pearson et al., 2003), as it lacks enzymes found in many eukaryotes for demethylation and other functions. Thus, it may be that the disadvantage of having both enzymes is due to subsequent processing of one or both, which may make products that interfere with the functioning of the other.

2. If SHC and OSC were truly equivalent, and the losses were random, then one would expect to find more eukaryotes with STC. For instance, the ciliate *Tetrahymena* normally produces the hopanoid tetrahymenol, but can substitute cholesterol if available (Conner et al., 1968). Of the eukaryotic groups that kept SHC, many are obligate anaerobes, such as the fungus *Piromyces*. However, the vast majority of known eukaryotes have a gene for OSC and not SHC. Is there some advantage of OSC, or of sterols, over SHC and hopanoids? One advantage could be intrinsic properties of the molecules themselves. It was shown that membranes with cholesterol had substantially improved viscosity compared to lanosterol, or no sterol at all (reviewed by (Bloch, 1983)). These studies addressed sterols, but did not consider hopanoids. However, (Saenz et al., 2012) reported that the hopanoid diplopterol is functionally similar to cholesterol in terms of membrane-ordering effects, based on multiple experiments.

   The second advantage may relate to the abundance of oxygen. If environmental oxygen were low, or inconsistently available, then having an oxygen-independent substitute may prove to be useful. However, if environmental oxygen levels were consistently high, such substitute may never be needed, and thus favor the loss of SHC over OSC.

Some lineages gained OSC or SHC by horizontal gene transfer

The only modern eukaryotes with both enzymes (some ascomycete fungi (Hayashi et al., 1996; Isaka et al., 2011) and ferns (Pandey and Mitra, 1969; Zander et al., 1969)) have lost the original STC-type and then re-acquired SHC by horizontal gene transfer from two different bacterial groups (label **(4)** in Figure 2B). Fungi have acquired SHC from an ancestor of *Anaeromyxobacter*, a delta-proteobacterium. On the other hand, ferns have acquired SHC from a cyanobacterial ancestor. Considering this, it should be noted that no modern eukaryote can serve as a "living fossil" proxy for early eukaryotes.

Nonetheless, for the fungi, this may confer some advantage in anaerobic conditions. Yeast, for instance, can grow under anaerobic conditions if ergosterol is supplied to the medium (Andreasen and Stier, 1953),

 doi:10.20944/preprints202004.0186.v5

suggesting that the only metabolic limitation for yeast under anoxia is the production of sterols. For ferns, however, it seems more likely that SHC would have an ecological role, rather than physiological. Many other flowering plants have multiple paralogs of OSC, named by their different activities: cycloartenol synthase (CAS) and beta-amyrin synthase (BAS). The HGT of SHC to ferns may simply be a way of adding to the diversity of compounds produced, or could serve as toxins against microbial pathogens or herbivorous insects.

Some bacteria appear to have genes for both SHC and OSC in their genomes, such as the planctomycete *Gemmata obscuriglobus*, or the gammaproteobacterium *Methylococcus capsulatus*. Two recent studies (Gudde et al., 2019; Rivas-Marin et al., 2019) have demonstrated that knocking out either OSC or SQMO in *G. obscuriglobus* results in a misshapen membrane structures and an overall loss of viability, which could be rescued by addition of lanosterol to the medium. At first glance, it would appear that this would counter the theory that SHC could serve as a replacement for OSC. However, the *G. obscuriglobus* SHC is around 100 amino acids shorter than the nearest planctomycete neighbors, so it is not clear that this enzyme is functional at all. This may therefore explain the observed requirement for sterols (Gudde et al., 2019; Rivas-Marin et al., 2019), as there is no chance of a substitute.

The gammaproteobacterium *Methylococcus capsulatus*, on the other hand, appears to be making both hopanoids and sterols under normal laboratory conditions (Jahnke et al., 1992). These sterols, however, are still methylated at the C4 (Jahnke et al., 1992), showing that this bacterium possesses enzymes to demethylate at the C14 position, but ineffectively at the C4 position. When compared to yeast, the total pathway would therefore be considered incomplete.

Many other lineages secondarily lost OSC

Considering the animal kingdom, OSC is clearly found in the genomes of species of porifera, placozoans, echinoderms, molluscs, chordates, hemichordate worms, nemertean worms, annelid worms, brachiopods, phoronids, and tardigrades. However, it is reported sterol biosynthesis is absent in arthropods (Zandee, 1964), nematodes (Chitwood, 1999), and platyhelminths (Meyer et al., 1970). Even within molluscs, sterol biosynthesis appears to be absent in some bivalves (Walton and Pennock, 1972) and squid. Neither SHC or OSC are found in the genomes of cnidarians (corals and jellyfish) and ctenophores (comb jellies), though sterols were present in lipids of several cnidarians and ctenophores (Nelson et al., 2000), suggesting that dietary acquisition is sufficient.

A number of microbial eukaryotes appear to lack both OSC and STC (Takishita, Chikaraishi, Tanifuji, et al., 2017). Much like the above schemes, “dietary” acquisition may fulfill this need, or that the function may be dispensible entirely. This could also be a reflection of the overall cost of the pathway. As many animal groups have lost OSC, it is conceivable that many single-celled eukaryotes could have lost OSC as well, and also STC. This further argues for the repeated losses seen in Figure 3.

Early eukaryotes may have anaerobically produced hopanoids

The enzyme OSC is widely assumed to be a feature of the LECA (Desmond and Gribaldo, 2009; Gold et al., 2017), which is the clear and obvious interpretation of the tree. If SHC/STC were a feature of the LECA, then both enzymes would be found ancestral in early eukaryotes. This further suggests that early eukaryotes could have made both hopanoids and sterols.

Contrary to previous hypotheses, this highlights the possibility that early eukaryotes were facultative aerobes capable of living in low-oxygen environments, rather than obligate aerobes living in fully-oxygenated

 doi:10.20944/preprints202004.0186.v5

environments (Desmond and Gribaldo, 2009). Much like the requirement of sterols for anaerobic growth in yeast (Andreasen and Stier, 1953), it is possible that early eukaryotes may have survived in low-oxygen conditions, and only produced sterols when oxygen was available. These conditions are uncertain. Experiments in yeast suggest that the oxidation of squalene could occur at oxygen levels as low as 7nM (Waldbauer et al., 2011), so very low atmospheric levels of oxygen may have been sufficient.

This means that many rocks that contain only hopanes could still derive from ecosystems with abundant eukaryotes, and that those eukaryotes could have made hopanoids in parallel to the bacteria. The clearest case is that of ciliates (Harvey and Mcmanus, 1991; Venkatesan, 1989), where tetrahymanol is found is modern sediments, and its product, gammacerane, is found in many ancient sediments or rocks.

This would also reconcile other evidence from fossils about the timing of the origin of eukaryotes. Many fossils have been found from over 1 billion years ago that appear to be eukaryotes (Bengtson et al., 2017; Butterfield, 2000; Cohen and Macdonald, 2015), yet evidence from the biomarkers would suggest that the environment was dominated by bacteria (J Brocks et al., 2017), and eukaryotes were rare or absent. This would therefore support the interpretation of these fossils, arguing that they are bona fide eukaryotes, but were living in low-oxygen environments where only hopanoid synthesis was possible.

A stark contrast is then found in rocks from 600-700Ma, where steranes are found at appreciable quantites, suggesting that eukaryotic organisms then became globally abundant until modern day (J Brocks et al., 2017). Rather than an ecological shift, this pattern instead could be indicative of a global change in atmospheric oxygen.

The origin of STC in eukaryotes has not received the attention of other parts of the history of this enzyme family. The presence of STC is best explained by primary inheritance in the LECA, and repeated losses. From this, the LECA could have been able to produce both sterols and hopanoids, potentially making it viable in both high and low-oxygen environments. Much of this is still uncertain, and several things may help resolve the biology and implications to the geology. STC enzymes do not appear to be well characterized outside of ciliates. Potential targets for protein expression and characterization include those from *Sawyeria marylandensis*, *Stygiella incarcerata*, or *Paratrimastix pyriformis*. These species may be most informative as to the chemistry of the putative original SHC/STC enzyme in eukaryotes, and when and how STC activity evolved in this branch. Tetrahymanol is predicted to transform into gammacerane by diagenesis. A thorough review of the record of this molecule in rocks would need to be conducted. Clarifying these things may improve our understanding of the capabilities and environment of the earliest eukaryotic organisms on our planet.

Acknowledgements

The author would like to thank DEC and DBM for helpful discussions. Preprint version 5 of this article has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (https://doi.org/10.24072/pci.evolbiol.100144).

Fundings

This work was supported by a VILLUM Experiment grant (no. 00028022) to WRF.

Conflict of interest disclosure

The author declares that he complies with the PCI rule of having no financial conflicts of interest in relation to the content of the article. In addition, the author declares that he has no non-financial conflict of interest with the content of this article.

Data, script and code availability

Data, scripts and codes are available online: https://doi.org/10.5281/zenodo.7137265

References

Andreasen AA and TJB Stier (1953). Anaerobic nutrition of Saccharomyces cerevisiae. I. Ergosterol requirement for growth in a defined medium. *Journal of Cellular and Comparative Physiology* 41, 23–36. https://doi.org/10.1002/jcp.1030410103.

Bengtson S, T Sallstedt, V Belivanova, and M Whitehouse (2017). Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biology* 15, 1–38. https://doi.org/10.1371/journal.pbio.2000735.

Bloch KE (1983). Sterol. Structure and Membrane Function. *Critical Reviews in Biochemistry and Molecular Biology* 14, 47–92.

Brocks J, A Jarrett, M Sirantoine, C Hallmann, Y Hoshino, and T Liyanage (2017). The rise of algae in Cryogenian oceans and the emergence of animals. *Nature*, in press. https://doi.org/10.1038/nature23457.

Brocks JJ and A Pearson (2005). Building the Biomarker Tree of Life. *Reviews in Mineralogy and Geochemistry* 59, 233–258. https://doi.org/10.2138/rmg.2005.59.10.

Brown MW, AA Heiss, R Kamikawa, Y Inagaki, A Yabuki, AK Tice, T Shiratori, KI Ishida, T Hashimoto, AGB Simpson, and AJ Roger (2018). Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biology and Evolution* 10, 427–433. https://doi.org/10.1093/gbe/evy014.

Butterfield NJ (2000). Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26, 386–404. https://doi.org/10.1666/0094-8373(2000)026<0386:BPNGNS>2.0.CO;2.

Chitwood DJ (1999). Biochemistry and function of nematode steroids. *Critical Reviews in Biochemistry and Molecular Biology* 34, 273–284. https://doi.org/10.1080/10409239991209309.

Cohen PA and FA Macdonald (2015). The Proterozoic Record of Eukaryotes. *Paleobiology* 41, 610–632. https://doi.org/10.1017/pab.2015.25.

Conner RL, JR Landrey, CH Burns, and FB Mallory (1968). Cholesterol Inhibition of Pentacyclic Triterpenoid Biosynthesis in Tetrahymena pyriformis. *The Journal of Protozoology* 15, 600–605. https://doi.org/10.1111/j.1550-7408.1968.tb02178.x.

Derelle R, G Torruella, V Klimeš, H Brinkmann, E Kim, Č Vlček, BF Lang, and M Eliáš (2015). Bacterial proteins pinpoint a single eukaryotic root. *Proceedings of the National Academy of Sciences of the United States of America* 112, E693–E699. https://doi.org/10.1073/pnas.1420657112.

Desmond E and S Gribaldo (2009). Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution* 1, 364–381. https://doi.org/10.1093/gbe/evp036.

Fischer WW and A Pearson (2007). Hypotheses for the origin and early evolution of triterpenoid cyclases. *Geobiology* 5, 19–34. https://doi.org/10.1111/j.1472-4669.2007.00096.x.

Francis W (2022). *wrf/euk-sterol-hopene: v5 for publication*. Version V5. https://doi.org/10.5281/zenodo.7137265.

Frickey T and E Kannenberg (2009). Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. *Environmental Microbiology* 11, 1224–1241. https://doi.org/10.1111/j.1462-2920.2008.01851.x.

Gold DA, A Caron, GP Fournier, and RE Summons (2017). Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* 543, 420–423. https://doi.org/10.1038/nature21412.

Gudde LR, M Hulce, AH Largen, and JD Franke (2019). Sterol synthesis is essential for viability in the planctomycete bacterium Gemmata obscuriglobus. *FEMS Microbiology Letters* 366, 1–7. https://doi.org/10.1093/femsle/fnz019.

Harvey HR and GB Mcmanus (1991). Marine ciliates as a widespread source of tetrahymanol and hopan-3β-ol in sediments. *Geochimica et Cosmochimica Acta* 55, 3387–3390. https://doi.org/10.1016/0016-7037(91)90496-R.

Hayashi H, Y Asabu, K Naito, M Nakayama, H Nozaki, and M Arai (1996). New oleanane triterpene with three ketones produced by penicillium simplicissimum ATCC 90288. *Bioscience, Biotechnology and Biochemistry* 60, 1732–1734. https://doi.org/10.1271/bbb.60.1732.

Isaka M, P Chinthanom, M Sappan, R Chanthaket, JJ Luangsa-Ard, S Prabpai, and P Kongsaeree (2011). Lanostane and hopane triterpenes from the entomopathogenic fungus hypocrella sp. BCC 14524. *Journal of Natural Products* 74, 2143–2150. https://doi.org/10.1021/np200429b.

Jahnke LL, H Stan-Lotter, K Kato, and LI Hochstein (1992). Presence of methyl sterol and bacteriohopanepolyol in an outer-membrane preparation from Methylococcus capsulatus (Bath). *Journal of General Microbiology* 138, 1759–1766. https://doi.org/10.1099/00221287-138-8-1759.

Katoh K, J Rozewicki, and KD Yamada (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 1, 1–7. https://doi.org/10.1093/bib/bbx108.

Keeling PJ, F Burki, HM Wilcox, B Allam, EE Allen, LA Amaral-Zettler, EV Armbrust, JM Archibald, AK Bharti, CJ Bell, B Beszteri, KD Bidle, CT Cameron, L Campbell, DA Caron, RA Cattolico, JL Collier, K Coyne, SK Davy, P Deschamps, ST Dyhrman, B Edvardsen, RD Gates, CJ Gobler, SJ Greenwood, SM Guida, JL Jacobi, KS Jakobsen, ER James, B Jenkins, U John, MD Johnson, AR Juhl, A Kamp, LA Katz, R Kiene, A Kudryavtsev, BS Leander, S Lin, C Lovejoy, D Lynn, A Marchetti, G McManus, AM Nedelcu, S Menden-Deuer, C Miceli, T Mock, M Montresor, MA Moran, S Murray, G Nadathur, S Nagai, PB Ngam, B Palenik, J Pawlowski, G Petroni, G Piganeau, MC Posewitz, K Rengefors, G Romano, ME Rumpho, T Rynearson, KB Schilling, DC Schroeder, AG Simpson, CH Slamovits, DR Smith, GJ Smith, SR Smith, HM Sosik, P Stief, E Theriot, SN Twary, PE Umale, D Vaulot, B Wawrik, GL Wheeler, WH Wilson, Y Xu, A Zingone, and AZ Worden (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biology* 12. https://doi.org/10.1371/journal.pbio.1001889. arXiv: 0402594v3 [arXiv:cond-mat].

Kemp P, DJ Lander, and CG Orpin (1984). The lipids of the rumen fungus Piromonas communis. *Journal of general microbiology* 130, 27–37.

Lamb DC, CJ Jackson, AG Warrilow, NJ Manning, DE Kelly, and SL Kelly (2007). Lanosterol biosynthesis in the prokaryote Methylococcus Capsulatus: Insight into the evolution of sterol biosynthesis. *Molecular Biology and Evolution* 24, 1714–1721. https://doi.org/10.1093/molbev/msm090.

Mallory FB, JT Gordon, and RL Conner (1963). The isolation of a pentacyclic triterpenoid alcohol from a protozoan. *Journal of the American Chemical Society* 85, 1362–1363. https://doi.org/10.1021/ja00892a042.

Meyer F, H Meyer, and E Bueding (1970). Lipid metabolism in the parasitic and free-living flatworms, Schistosoma mansoni and Dugesia dorotocephala. *Biochimica et Biophysica Acta (BBA)/Lipids and Lipid Metabolism* 210, 257–266. https://doi.org/10.1016/0005-2760(70)90170-0.

Nelson MM, CF Phleger, BD Mooney, and PD Nichols (2000). Lipids of gelatinous antarctic zooplankton: Cnidaria and Ctenophora. *Lipids* 35, 551–559. https://doi.org/10.1007/s11745-000-555-5.

Nes WD (2011). Biosynthesis of cholesterol and other sterols. *Chemical Reviews* 111, 6423–6451. https://doi.org/10.1021/cr200021m.

Nguyen LT, HA Schmidt, A Von Haeseler, and BQ Minh (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Ourisson G and P Albrecht (1992). Hopanoids. 1. Geohopanoids: The Most Abundant Natural Products on Earth? *Accounts of Chemical Research* 25, 398–402. https://doi.org/10.1021/ar00021a003.

Pandey GN and CR Mitra (1969). Constituents of Oleandra wallichii. *Phytochemistry* 8, 327–330. https://doi.org/10.1016/S0031-9422(00)85836-7.

Pearson A, M Budin, and JJ Brocks (2003). Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. *Proceedings of the National Academy of Sciences* 100, 15352–15357. https://doi.org/10.1073/pnas.2536559100.

Ren R, Y Sun, Y Zhao, D Geiser, H Ma, and X Zhou (2016). Phylogenetic Resolution of Deep Eukaryotic and Fungal Relationships Using Highly Conserved Low-Copy Nuclear Genes. *Genome Biology and Evolution* 8, evw196. https://doi.org/10.1093/gbe/evw196.

Rivas-Marin E, S Stettner, EY Gottshall, C Santana-Molina, M Helling, F Basile, NL Ward, and DP Devos (2019). Essentiality of sterol synthesis genes in the planctomycete bacterium Gemmata obscuriglobus. *Nature Communications* 10, 3–8. https://doi.org/10.1038/s41467-019-10983-7.

Rohmer M, P Bouvier, and G Ourisson (1979). Molecular evolution of biomembranes: Structural equivalents and phylogenetic precursors of sterols. *Proceedings of the National Academy of Sciences of the United States of America* 76, 847–851. https://doi.org/10.1073/pnas.76.2.847.

Saenz JP, E Sezgin, P Schwille, and K Simons (2012). Functional convergence of hopanoids and sterols in membrane ordering. *Proceedings of the National Academy of Sciences* 109, 14236–14240. https://doi.org/10.1073/pnas.1212141109.

Santana-Molina C, E Rivas-Marin, AM Rojas, and DP Devos (2020). Origin and Evolution of Polycyclic Triterpene Synthesis. *Molecular Biology and Evolution*, 1–17. https://doi.org/10.1093/molbev/msaa054.

Shinozaki J, M Shibuya, K Masuda, and Y Ebizuka (2008). Squalene cyclase and oxidosqualene cyclase from a fern. *FEBS Letters* 582, 310–318. https://doi.org/10.1016/j.febslet.2007.12.023.

Sipe J and C Holmlund (1972). A Comparison of the effects of some hypocholesteremic compounds on squalene metabolism in Tetrahymena pyriformis and rat liver. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 280, 145–160. https://doi.org/10.1016/0005-2760(72)90220-2.

Takishita K, Y Chikaraishi, MM Leger, E Kim, A Yabuki, N Ohkouchi, and AJ Roger (2012). Lateral transfer of tetrahymanol-synthesizing genes has allowed multiple diverse eukaryote lineages to independently adapt to environments without oxygen. *Biology Direct* 7, 5. https://doi.org/10.1186/1745-6150-7-5.

Takishita K, Y Chikaraishi, G Tanifuji, N Ohkouchi, T Hashimoto, K Fujikura, and AJ Roger (2017). Microbial Eukaryotes that Lack Sterols. *Journal of Eukaryotic Microbiology* 64, 897–900. https://doi.org/10.1111/jeu.12426.

Thoma R, T Schulz-Gasen, B D'Arcy, J Benz, J Aebi, H Dehmlow, M Hennig, M Stihle, and A Ruf (2004). Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* 432, 118–122. https://doi.org/10.1038/nature02993.

Venkatesan MI (1989). Tetrahymanol: Its widespread occurrence and geochemical significance. *Geochimica et Cosmochimica Acta* 53, 3095–3101. https://doi.org/10.1016/0016-7037(89)90190-7.

Waldbauer JR, DK Newman, and RE Summons (2011). Microaerobic steroid biosynthesis and the molecular fossil record of Archean life. *Proceedings of the National Academy of Sciences* 108, 13409–13414. https://doi.org/10.1073/pnas.1104160108.

Walton MJ and JF Pennock (1972). Some studies on the biosynthesis of ubiguinone, isoprenoid alcohols, squalene and sterols by marine invertebrates. *The Biochemical journal* 127, 471–479. https://doi.org/10.1042/bj1270471.

Wei JH, X Yin, and PV Welander (2016). Sterol synthesis in diverse bacteria. *Frontiers in Microbiology* 7, 1–19. https://doi.org/10.3389/fmicb.2016.00990.

 doi:10.20944/preprints202004.0186.v5

Wendt KU, A Lenhart, and GE Schulz (1999). The structure of the membrane protein squalene-hopene cyclase at 2.0 Å resolution. *Journal of Molecular Biology* 286, 175–187. https://doi.org/10.1006/jmbi.1998.2470.

Zandee DI (1964). Absence of Sterol Synthesis in some Arthropods. *Nature* 202, 1335–1336. https://doi.org/10.1038/2021335a0.

Zander JM, E Caspi, GN Pandey, and CR Mitra (1969). The presence of tetrahymanol in Oleandra wallichii. *Phytochemistry* 8, 2265–2267. https://doi.org/10.1016/S0031-9422(00)88195-9.