Evolution of bacterial steroid biosynthesis and its impact on eukaryogenesis

Yosuke Hoshino[a,1] and Eric A. Gaucher[a,1]

[a]Department of Biology, Georgia State University, Atlanta, GA 30302

Steroids are components of the eukaryotic cellular membrane and have indispensable roles in the process of eukaryotic endocytosis by regulating membrane fluidity and permeability. In particular, steroids may have been a structural prerequisite for the acquisition of mitochondria via endocytosis during eukaryogenesis. While eukaryotes are inferred to have evolved from an archaeal lineage, there is little similarity between the eukaryotic and archaeal cellular membranes. As such, the evolution of eukaryotic cellular membranes has limited our understanding of eukaryogenesis. Despite evolving from archaea, the eukaryotic cellular membrane is essentially a fatty acid bacterial-type membrane, which implies a substantial bacterial contribution to the evolution of the eukaryotic cellular membrane. Here, we address the evolution of steroid biosynthesis in eukaryotes by combining ancestral sequence reconstruction and comprehensive phylogenetic analyses of steroid biosynthesis genes. Contrary to the traditional assumption that eukaryotic steroid biosynthesis evolved within eukaryotes, most steroid biosynthesis genes are inferred to be derived from bacteria. In particular, aerobic deltaproteobacteria (myxobacteria) seem to have mediated the transfer of key genes for steroid biosynthesis to eukaryotes. Analyses of resurrected steroid biosynthesis enzymes suggest that the steroid biosynthesis pathway in early eukaryotes may have been similar to the pathway seen in modern plants and algae. These resurrected proteins also experimentally demonstrate that molecular oxygen was required to establish the modern eukaryotic cellular membrane during eukaryogenesis. Our study provides unique insight into relationships between early eukaryotes and other bacteria in addition to the well-known endosymbiosis with alphaproteobacteria.

eukaryogenesis | steroids | biomarker | cellular membrane | sequence reconstruction

The emergence of modern eukaryotic cells from their archaeal ancestor requires multiple evolutionary steps, most notably the acquisition of mitochondria, nucleus, endomembrane system, and bacterial-like cellular membranes. The acquisition of mitochondria by early eukaryotes before the last eukaryotic common ancestor (LECA) would have been a key step to exploit new resources within an aerobic environment (1). Similarly, the transformation of the eukaryotic cellular membrane system from an archaeal-type to bacterial-type membrane would have also been crucial to develop the dynamic cellular membrane system observed in modern eukaryotes (2). Such a dynamic membrane may have been the prerequisite for a pre-LECA cell to perform endocytosis and, hence, later acquire organelles such as mitochondria and chloroplasts. While the tetraether monolayer membrane in archaea is more rigid and is advantageous to harsh environments, such as high temperature and low pH conditions, the diester bilayer membrane found in bacteria and eukaryotes has been suggested to be adaptable to more general environments (3). Recent phylogenomic analyses of archaea and eukaryotes suggest that eukaryotes evolved from within the archaeal domain (4, 5). However, there is little similarity between the archaeal and the bacterial/eukaryotic cellular membranes. Both the stereochemistry of lipid molecules (i.e., phospholipids) and their lipid composition (e.g., fatty acid chain versus isoprenoid chain and the presence of

sphingolipids and steroids) are fundamentally different between the two types of membranes. As such, the evolutionary history of the eukaryotic cellular membrane is a major gap in our understanding of eukaryogenesis (6).

Here, we focus on the evolution of the steroid component of the eukaryotic cellular membrane. Steroids have indispensable roles in the process of eukaryotic endocytosis because they are embedded in the cellular membrane and are known to regulate membrane fluidity and permeability (7–9). Thus, steroids may also have been integral for the endocytic process to acquire mitochondria during eukaryogenesis. Steroids would also function as hormones to control various cellular processes, including differentiation, morphogenesis, and homeostasis to enable multicellularity (10). In addition to the vital biochemical role in modern eukaryotes, steroids also have a geobiological importance. They can serve as unique biological markers (biomarkers) for eukaryotes in the geological record and thus provide clues to trace the evolutionary history of eukaryotes on a geological time scale. Membrane-bound steroids are modified in a unique manner within each eukaryotic taxon. For instance, cholesterol is a major component of cellular membrane in metazoans, whereas stigmasterol is a major component in plants (11, 12). These taxon-specific modifications can be used as taxonomic markers in the geological record (13–15). No analogous steroid biosynthesis pathway has been observed in archaea and thus steroid biosynthesis is generally inferred to have evolved de novo within eukaryotes (16). Despite this inference,

---

Significance

Steroids are one of three major lipid components of the eukaryotic cellular membrane, along with glycerophospholipids and sphingolipids. Steroids have critical roles in eukaryotic endocytosis and thus may have been structural prerequisites for the endocytic acquisition of mitochondria during eukaryogenesis. The evolutionary history of the eukaryotic cellular membrane is poorly understood and, as such, has limited our understanding of eukaryogenesis. We address the evolution of steroid biosynthesis by combining ancestral sequence reconstruction and phylogenetic analyses of steroid biosynthesis genes. Our results indicate that steroid biosynthesis evolved within bacteria in response to the rise of oxygen and was later horizontally transferred to eukaryotes. Membrane properties of early eukaryotes are inferred to have been different than that of modern eukaryotes.

EVOLUTION

some bacteria are known to produce hopanoids that are structurally similar to steroids (17). Indeed, several genes that are involved in hopanoid and steroid biosynthesis are suggested to have been horizontally transferred from bacteria to eukaryotes (18, 19). These recent observations suggest a more complex evolutionary history of steroid biosynthesis in eukaryotes. Ancestral sequence reconstruction (ASR) enables us to experimentally analyze resurrected enzymes (20, 21) and thus infer evolutionary histories of steroid biosynthesis and their impacts on eukaryogenesis.

Results and Discussion

Steroid Biosynthesis in Bacteria. Steroid biosynthesis is oxygen dependent, and half of the enzymatic reactions require molecular oxygen, while steroid precursor (squalene) biosynthesis requires no oxygen (Fig. 1, Stages 1 and 2). Steroid biosynthesis, particularly postcyclization modification, has almost exclusively been described in eukaryotes and thus has generally been inferred to be characteristic of eukaryotes. However, in our current study, a potential steroid biosynthesis pathway is found in 14 bacterial phyla (steroid-producing bacteria; *SI Appendix*, Fig. S1 and Table S1). The steroid biosynthesis pathway for most of the steroid-producing bacteria is distinct from the eukaryotic pathway, while aerobic deltaproteobacteria (myxobacteria) and Dadabacteria are found to possess a pathway homologous to that of eukaryotes. In the steroid precursor biosynthesis (Fig. 1, Stage 1), eukaryotes utilize the seven-step mevalonate (MVA) pathway for the prenyl unit formation (18). In contrast, most steroid-producing bacteria use the similarly seven-step methylerythritol 4-phosphate (MEP) pathway that is widely distributed in the bacterial domain (22). Only myxobacteria and Dadabacteria are found to possess the

MVA pathway (Fig. 1). Also, prenyl chain elongation after farnesyl diphosphate synthase proceeds by either a single-step pathway (squalene synthase [SQS]) or a three-step pathway (HpnCDE) (19, 23). Eukaryotes possess only the SQS pathway, while both pathways are distributed in steroid-producing bacteria. The majority of myxobacteria and Dadabacteria are found to possess the SQS pathway similar to eukaryotes. Overall, most steroid-producing bacteria utilize the combination of the MEP and HpnCDE pathways or the MEP and SQS pathways. In contrast, only myxobacteria and Dadabacteria have the combination of the MVA and SQS pathways that are homologous to the eukaryotic steroid biosynthesis pathway.

Interestingly, two close relatives of myxobacteria that are not known to produce steroids also possess the MVA and SQS combination—the order Bradymonadales (Deltaproteobacteria) and the phylum Oligoflexia (formerly included in Deltaproteobacteria) (Fig. 1). Bradymonadales and Oligoflexia contain obligately and facultatively aerobic species and share a common ancestor with myxobacteria (Fig. 1, *Inset* and *SI Appendix*, Fig. S2). In contrast to these three aerobic groups, obligately anaerobic deltaproteobacteria only possess genes for the MEP pathway that contains two enzymes inhibited by molecular oxygen (24) (Fig. 1, asterisk). Therefore, the distribution pattern of Stage 1 steroid precursor biosynthesis genes is consistent with the aerobic/anaerobic lifestyle of members in the Deltaproteobacteria/Oligoflexia group. This distinction seems to reflect the adaptation of deltaproteobacteria toward aerobic environments by evolving and/or acquiring MVA and SQS genes.

Recent metagenomic analyses suggest that the presence of uncultured anaerobic taxa near the three aerobic lineages (Oligoflexia, Bradymonadales, and Myxococcales) (25) and thus suggest that
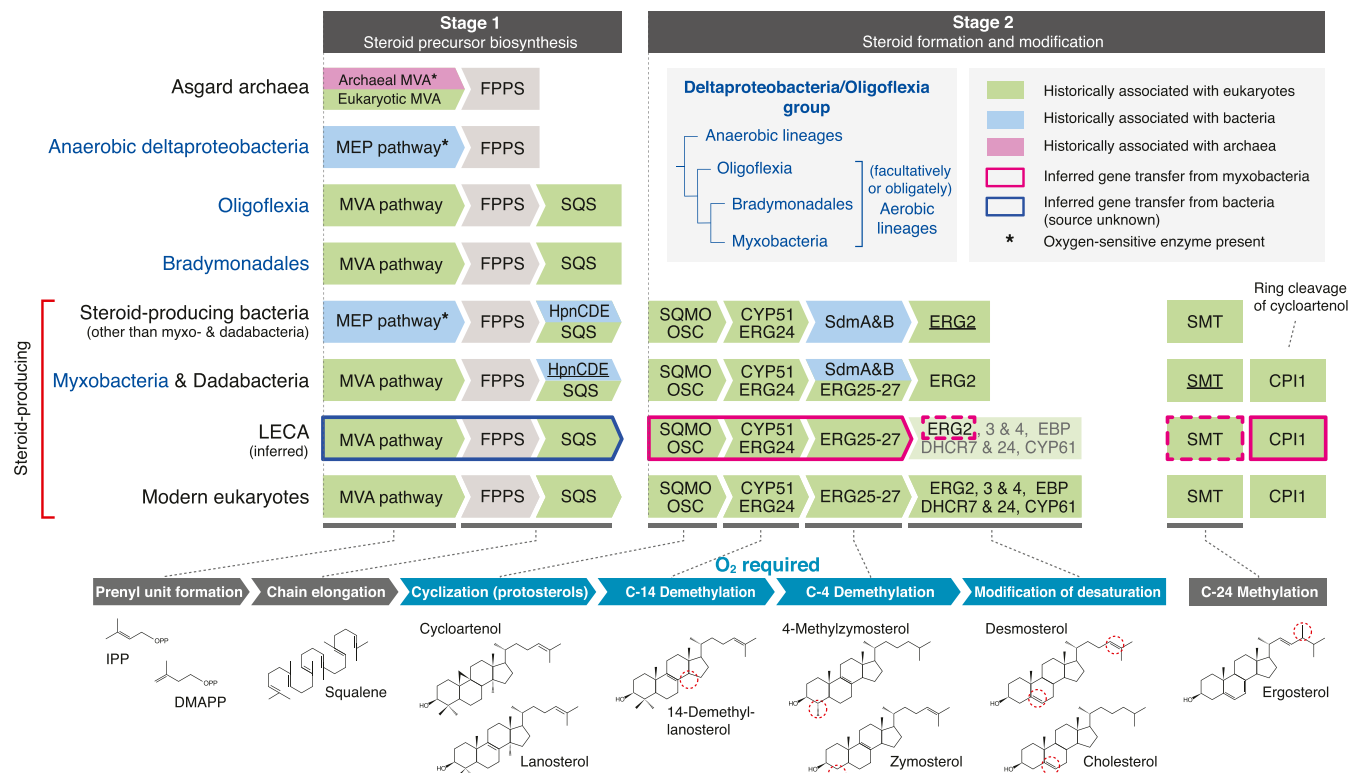
Fig. 1. Distribution of steroid biosynthesis genes in select bacterial taxa and eukaryotes. Representative enzymatic products are shown at each step. The evolutionary relationship within the Deltaproteobacteria and Oligoflexia group (blue color) is shown in the inset. Underlined genes in bacteria indicate that only a small number of species possess the gene within the indicated group. Genes lightly shaded in LECA indicate uncertainty in the presence of the gene. The reaction order within Stage 2 is generally from left (cyclization) to right (modification of desaturation), but there are variations after cyclization depending on some species. The MVA pathway in green color indicates the eukaryotic type, while the pathway in red color indicates the archaeal type. ERG24 is separately treated from other desaturation-modifying genes. See *SI Appendix*, Fig. S1 for gene abbreviations.

these three aerobic lineages evolved independently from anaerobic ancestors. Many of the uncultured species in our analyses are inferred to have secondarily adapted to anaerobic environments (*SI Appendix*, Text 1), yet some of the anaerobic species in our analyses may potentially represent vertical descendants directly from anaerobic ancestors. Thus, the transition between anaerobiosis and aerobiosis seems to have occurred multiple times independently in both directions in the Deltaproteobacteria/Oligoflexia group. The conversion of the gene repertoire from the MEP to MVA gene set for the prenyl unit formation, which is observed for all three aerobic lineages, may be a result of convergent evolution. This would imply that the MVA gene set is critical for the adaptation to aerobic environments for these lineages. Some deltaproteobacteria in anaerobic lineages are known to possess terminal oxidase to metabolize molecular oxygen and are indeed aerotolerant. However, these species do not utilize terminal oxidase for aerobic respiration (e.g., ref. 26). These aerotolerant deltaproteobacteria are, thus, essentially obligate anaerobes but may represent an intermediate metabolism between obligately anaerobic and aerobic lifestyles (27).

Stage 2 has only one variation and that is for C-4 demethylation (Fig. 1). Most steroid-producing bacteria possess a two-step pathway (SdmA&B) only, yet myxobacteria and Dadabacteria additionally possess a three-step pathway (ERG25-27) that is otherwise only found in eukaryotes. As such, myxobacteria and Dadabacteria possess both clusters of SdmA&B and ERG25-27. Together with the observation seen for Stage 1, the steroid biosynthesis pathway in myxobacteria and Dadabacteria is more similar to the eukaryotic pathway than to other bacterial pathways.

Steroid Precursor Biosynthesis in Archaea. Unlike bacteria and eukaryotes, most archaea, including Asgard archaea that are inferred to be the closest relatives of eukaryotes, are found to possess only part of Stage 1 genes (prenyl unit formation) and thus do not even have the ability to produce steroid precursor (Fig. 1). Only some haloarchaea are known to biosynthesize squalene (28), but the SQS gene in haloarchaea is inferred to be derived horizontally from bacteria (*SI Appendix*, Figs. S3 and S12). The distribution of MVA genes in Asgard archaea appears complex, compared to other archaea and the other two domains (bacteria and eukaryotes). The majority of archaea, except for the Asgard group and some others, possess an MVA pathway (archaeal MVA pathway) that is not completely homologous to the eukaryotic MVA pathway (18). Two out of seven enzymes are replaced with two nonhomologous enzymes for those archaea (29). One of the two enzymes is inhibited by oxygen similar to the MEP pathway. In contrast, Asgard archaea are found to possess homologs for both the eukaryotic and archaeal MVA pathways. However, the gene repertoire for the MVA pathway varies depending on species, and only a small number of Asgard archaea possess a complete gene set for either the eukaryotic or the archaeal MVA pathway. Thus, it is not clear which MVA pathway is functioning in most of the Asgard archaea that lack a complete MVA gene set. One possibility is that Asgard archaea might have a chimeric MVA pathway that is a mixture of both the eukaryotic and archaeal pathways. Such a pathway might reflect the divergence of MVA pathway from the archaeal type to eukaryotic type approaching eukaryogenesis.

Bacterial Origin of Steroid Biosynthesis in Eukaryotes. Phylogenetic analyses of steroid biosynthesis genes also support a close evolutionary relationship between myxobacteria and eukaryotes. Eukaryotic homologs cluster with bacterial homologs for most of the steroid biosynthesis genes from Stages 1 and 2, except for the majority of desaturation-modifying genes that are largely absent from bacteria (Fig. 2 and *SI Appendix*, Figs. S3–S5). Outgroup sequences always root our trees within the bacterial lineage, and the outgroups are composed generally of bacteria. More importantly,

the observed global tree topologies mostly remain identical for each of the four different substitution models (maximum likelihood and Bayesian inferences with and without a profile mixture model) (see *SI Appendix*, Text 2 for details). Thus, the majority of the steroid biosynthesis genes in eukaryotes are inferred to be derived from bacteria, contrary to previous inferences that steroid biosynthesis genes were transferred from eukaryotes to bacteria (16, 30).
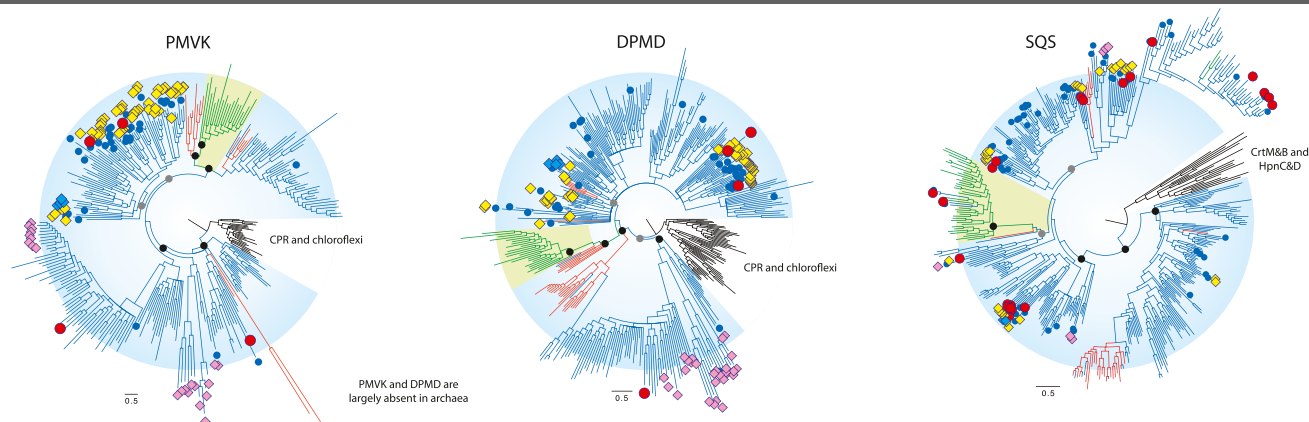
In particular, eight Stage 2 genes display a similar tree topology, with eukaryotic homologs always clustering within myxobacterial homologs (SQMO, oxidosqualene cyclase [OSC], CYP51, ERG24, ERG25-27, and CPI1) (Fig. 2 and *SI Appendix*, Fig. S4). These eight genes are, therefore, inferred to have been horizontally transferred from myxobacteria to eukaryotes. The myxobacterial origin of the eukaryotic OSC gene was previously suggested (19), and our results further suggest that the transfer of steroid biosynthesis genes was more extensive. These genes are mostly in a syntenic relationship with each other and thus are likely to have been transferred as a syntenic set (Fig. 2, filled red circles). In previous studies, the availability of genomic data was limited and thus bacterial steroid production was inferred to be an exception. Also, only two steroid biosynthesis genes (SQMO and OSC) were analyzed based on a single-substitution model (which can be prone to under parameterization). Our study includes a larger number of genomes and all 14 Stage 2 genes that apparently carry more phylogenetic signal than the two genes used in previously published work. Furthermore, four different substitution models are utilized to test the tree robustness. Thus, the inferred tree topology presented here is more robustly supported than trees from previous studies and supports gene transfer of eight Stage 2 genes from bacteria to eukaryotes. One additional Stage 2 gene (SMT) may similarly have been transferred horizontally from myxobacteria to eukaryotes (Fig. 2), but the distribution of the SMT gene in myxobacteria is sparse, and the individual node support for myxobacterial and eukaryotic clades is not high. Thus, the detailed evolutionary trajectory of SMT from bacteria to eukaryotes is not conclusive. Interestingly, the SMT gene is frequently accompanied by three uncharacterized genes (Fig. 2), and these genes might be involved in an unknown steroid biosynthesis pathway (*SI Appendix*, Text 3 and Table S2).

The tree topology of myxobacterial homologs for eight Stage 2 genes broadly follows the species tree of myxobacteria (Fig. 3 and *SI Appendix*, Fig. S6). Thus, the common ancestor of myxobacteria is inferred to have had these eight genes and hence a steroid biosynthesis pathway. The absence of most Stage 2 genes, except SQMO and OSC, in the subclass Cystobacterineae of myxobacteria is inferred to be a result of gene loss. Eukaryotic homologs form a monophyletic clade in most cases, except for ERG24, and the tree topology within the eukaryotic clade roughly follows the species tree. Thus, the gene transfer from bacteria to eukaryotes is likely to have occurred before the divergence of LECA. Most of the steroid-producing bacteria, except for myxobacteria and Dadabacteria, possess bacterial-type steroid biosynthesis genes (MEP, HpnCDE, and SdmA&B genes; Figs. 1 and 3). Among those genes, SdmA&B are almost exclusively distributed only in one clade that is solely composed of steroid-producing bacteria, except for only one eukaryotic species (MEP and HpnCDE genes are widespread also in bacteria that do not produce steroids). In contrast, ERG25-27 genes (functional equivalent of SdmA&B) are distributed in two major clades that include both bacteria and eukaryotes (Fig. 3). Therefore, SdmA&B genes may have evolved within steroid-producing bacteria after the divergence of three myxobacterial subclasses, while ERG25-27 genes are inferred to have already been present in the common ancestor of myxobacteria.
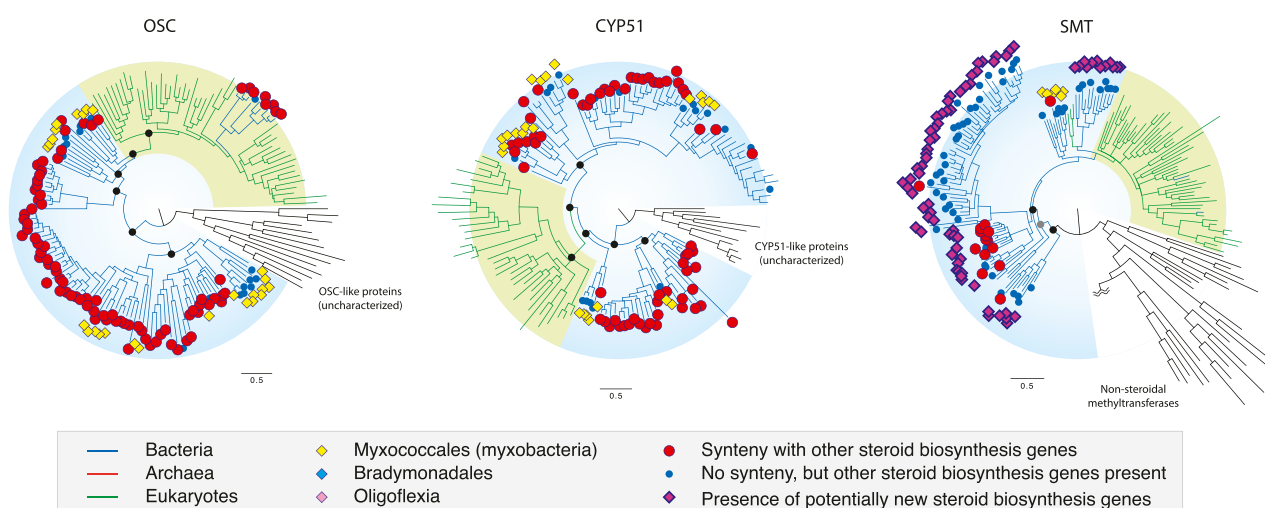
In contrast to the eight Stage 2 genes, most of the desaturation-modifying genes, except for ERG2 and ERG24, are inferred to have evolved within eukaryotes (ERG3, ERG4, DHCR7, DHCR24,

Stage 1 — PMVK, DPMD, SQS

Stage 2 — OSC, CYP51, SMT

Legend:
— Bacteria
— Archaea
— Eukaryotes
◆ Myxococcales (myxobacteria)
◆ Bradymonadales
◆ Oligoflexia
● Synteny with other steroid biosynthesis genes
● No synteny, but other steroid biosynthesis genes present
◆ Presence of potentially new steroid biosynthesis genes

Fig. 2. Bayesian phylogenetic trees for select Stage 1 and 2 steroid biosynthesis genes. Filled black circles indicate that the node support is >75% for maximum likelihood inference and >75% for Bayesian inference for all four trees (node support is shown only for major clades). Filled gray circles indicate that the node support is above the same threshold for three out of four trees. The nodes with less than 0.5 support are collapsed. The black line indicates the outgroup. (Scale bar represents 0.5 amino acid replacements per site per unit evolutionary time.) See *SI Appendix*, Figs. S3–S5 for other gene trees and *SI Appendix*, Figs. S9–S19 for the complete trees with the species annotation. Abbreviations: CYP51, cytochrome P450 family 51 (sterol C-14 demethylase); DPMD, diphosphomevalonate kinase; OSC; PMVK, phosphomevalonate kinase; SQMO, squalene monooxygenase; and SQS.

and EBP) (*SI Appendix*, Text 2 and Figs. S4 and S5). These desaturation-modifying genes are generally involved in the last steps of steroid biosynthesis after C-4 demethylation steps. The distribution of these desaturation-modifying genes depends on the steroid biosynthesis pathway in individual species, and thus, individual species have only a subset of known desaturation-modifying genes. It is likely that individual steroid biosynthesis pathways evolved in certain lineages independently from an ancestral pathway, and desaturation-modifying genes that are specific to individual pathways evolved within each lineage. The uneven distribution of desaturation-modifying genes within eukaryotes may suggest that not all of the desaturation-modifying genes were present in LECA. The evolution of ERG2 is complex (see *SI Appendix*, Text 2 for details). Although the ERG2 gene may have been transferred from myxobacteria to eukaryotes similar to eight Stage 2 genes, it is not conclusive.

The transfer of Stage 1 genes from bacteria to eukaryotes may have been more complex than the transfer of Stage 2 genes. For instance, the closest sister clade to eukaryotes contains a small number of archaea for MVA genes, including Asgard archaea for some of MVA genes, even though eukaryotes and those archaea

cluster with bacteria (Fig. 2 and *SI Appendix*, Fig. S3). This observation may indicate that MVA genes were transferred not directly from bacteria to eukaryotes but rather first to archaea and then to eukaryotes either horizontally or vertically (Asgard archaeal lineage). The three aerobic lineages in the Deltaproteobacteria/Oligoflexia group (Oligoflexia, Bradymonadales, and Myxococcales) fail to form a monophyletic clade for Stage 1 genes (Fig. 2 and *SI Appendix*, Fig. S3). While Bradymonadales and Myxococcales (myxobacteria) generally cluster together, Oligoflexia form separate clades. This observation suggests at least two separate origins of Stage 1 genes in the three aerobic lineages.

A mitochondrial origin of steroid biosynthesis genes is unlikely (31). Steroid biosynthesis genes are sporadically distributed in bacteria and are likely to have been horizontally transferred within the bacterial domain in most cases. Steroid biosynthesis genes are nearly absent in Alphaproteobacteria, and thus, none of those genes are inferred to have been present in the common ancestor of Alphaproteobacteria and thus in ancestral mitochondria (*SI Appendix*, Text 4). Hence, the origin of steroid biosynthesis is likely to be either in eukaryotes or in myxobacteria, and our data support a myxobacterial origin.
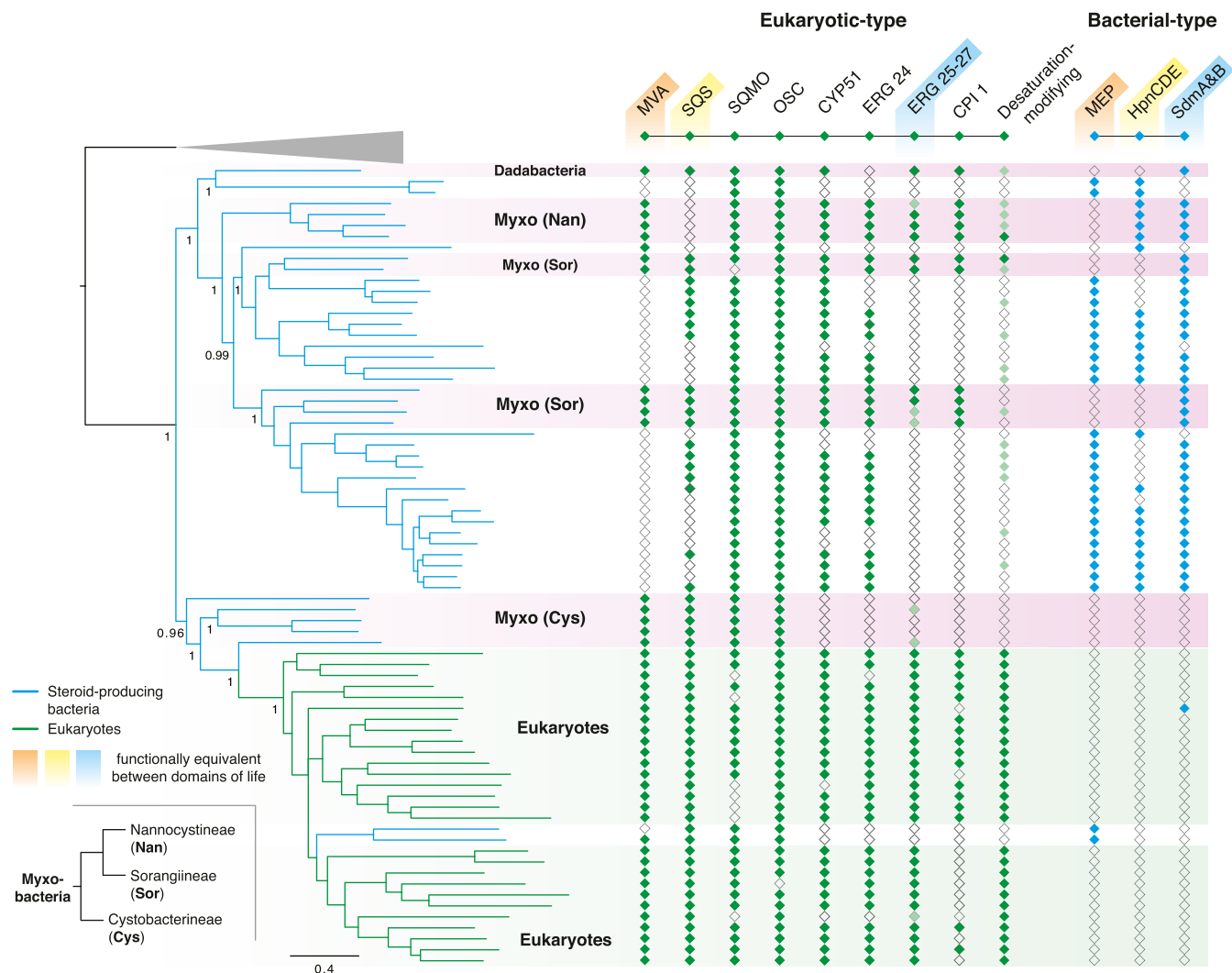
Fig. 3. Concatenated Bayesian phylogenetic tree of eight Stage 2 steroid biosynthesis genes (SQMO, OSC, CYP51, ERG24, ERG25, ERG26, ERG27 analog, and CPI1). Sequences are selected from species that have a potential steroid biosynthesis pathway for at least protosterols. The distribution of Stage 1 and 2 steroid biosynthesis genes is also shown for individual species. The tree topology remains nearly identical under the four different substitution models. The nodes with less than 0.5 support are collapsed. (Scale bar represents 0.4 amino acid replacements per site per unit evolutionary time.) See *SI Appendix*, Fig. S6 for the complete tree with the species annotation and *SI Appendix*, Table S3 for the source data.

Reconstruction of Ancestral Steroid Biosynthesis Pathway in Pre-LECA Eukaryotes. Given the robust tree topology for steroid biosynthesis genes, we next applied ASR to infer the ancestral steroid profile from pre-LECA eukaryotes. The steroid carbon structure is generated by the second Stage 2 enzyme—OSC (protosterol formation; Fig. 1). OSC is evolutionarily homologous to squalene cyclase (SC) that is utilized by bacteria to produce steroid-like compounds called hopanoids (Fig. 4) (17). Unlike steroid synthesis, hopanoid synthesis does not require molecular oxygen. OSC is mostly found in eukaryotes, while SC is mostly found in bacteria. SC is widespread in bacteria, and its origin is inferred to be earlier than the evolution of alphaproteobacteria and thus crown group eukaryotes (19).

Our phylogenetic analyses indicate that the OSC gene originated from a SC gene in bacteria (Fig. 4). Archaea do not have any enzymes homologous to OSC or SC, and it seems unlikely that the OSC gene evolved de novo in eukaryotes separately from bacterial SC. The bacterial origin of the OSC gene supports the inference that steroid biosynthesis originated in bacteria independently of the evolutionary origin of downstream steroid modification genes. Myxobacteria possess not only OSC but also

uncharacterized distant OSC homologs that form sister clades to the OSC clade (Fig. 4). These OSC-like proteins cluster with other terpene cyclases between the SC and OSC clades. Thus, these OSC-like proteins may be evolutionary remnants from the functional divergence of SC into OSC (32). The presence of both OSC and OSC-like proteins in myxobacteria may suggest that steroid biosynthesis evolved within myxobacteria. Interestingly, OSC-like protein clades demarcate O_2-independent and O_2-dependent terpene cyclase families (Fig. 4). SC and nearby cyclase families cyclize squalene and are O_2 independent, while bacterial and eukaryotic OSC families cyclize only oxidosqualene that requires molecular oxygen for its biosynthesis. Therefore, the functional divergence of SC into OSC mirrors the adaptation of terpene cyclase toward aerobic environments.

ASR at three internal nodes of the OSC clade indicates that the major protosterol produced by LECA was cycloartenol (Fig. 4, Node C and *SI Appendix*, Fig. S8). Our results experimentally validate the hypothesis that LECA was aerobic and utilized molecular oxygen. The production of cycloartenol is consistent with our inference that the CPI1 gene was indeed in LECA (CPI1 is specific to cycloartenol modification). The cycloartenol-based
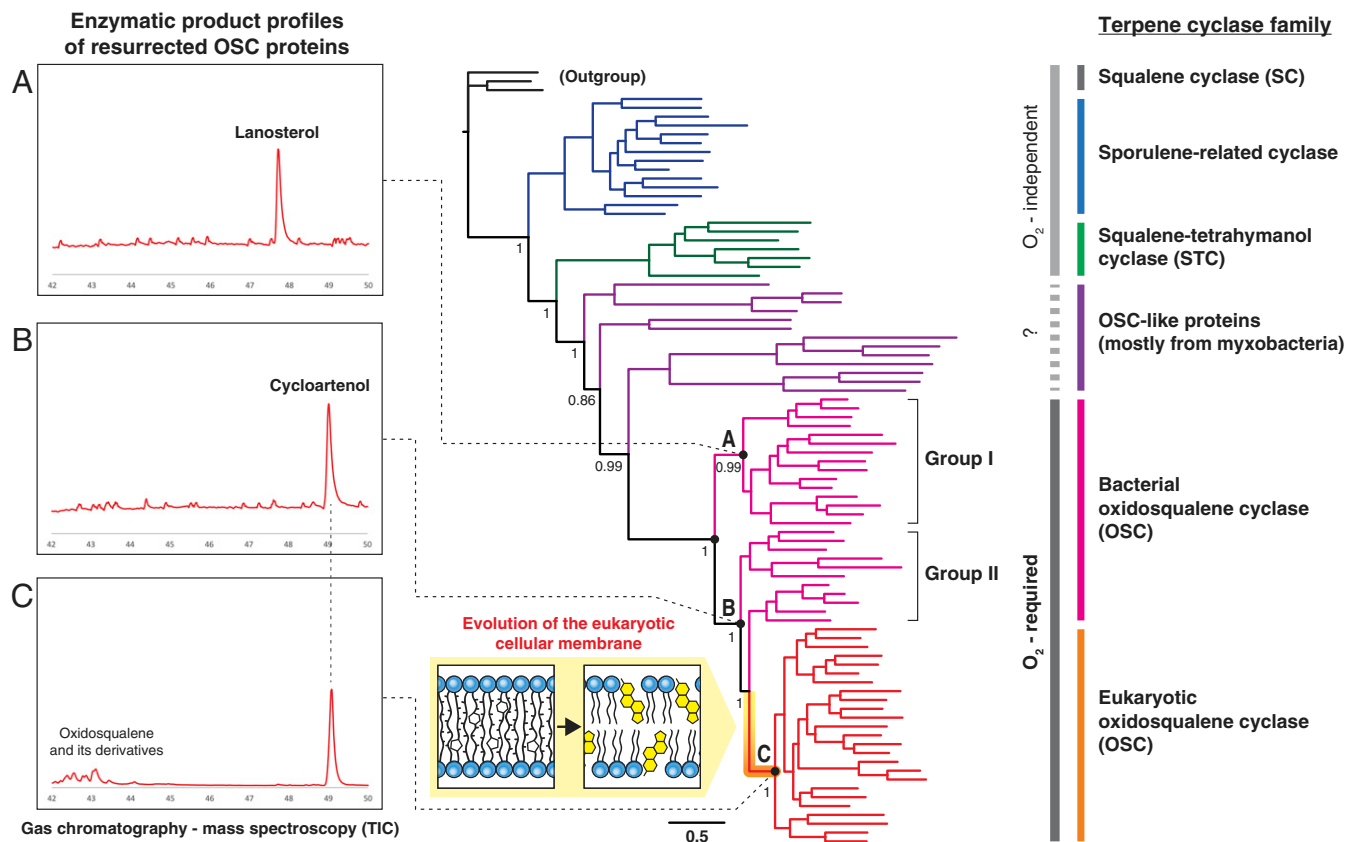
Fig. 4. Bayesian phylogenetic tree of terpene cyclase family and product profiles for three resurrected OSCs (nodes A, B, and C). The tree topology remains nearly identical under the four different substitution models. The nodes with less than 0.5 support are collapsed. (Scale bar represents 0.5 amino acid replacements per site per unit evolutionary time.) See *SI Appendix*, Fig. S7 for the complete tree with the species annotation.

steroid biosynthesis is today observed in Archaeplastida (e.g., land plants and algae). In contrast, the steroid biosynthesis in Opisthokonta (e.g., animals and fungi) proceeds via a different protosterol (lanosterol). Hence, the ancestral steroid biosynthesis pathway in LECA may have been similar to the cycloartenol-based pathway in Archaeplastida rather than the lanosterol-based pathway in Opisthokonta.

Cycloartenol-based steroid biosynthesis is likely to have occurred within bacteria before LECA. OSC homologs from steroid-producing bacteria comprise two distinct clades (Fig. 4, Groups I and II). The common ancestor of each group seems to have produced lanosterol and cycloartenol, respectively (Fig. 4, Nodes A and B and *SI Appendix*, Fig. S8). Group II bacteria that form sister clades to eukaryotes include the cycloartenol-producing myxobacteria (33). Thus, pre-LECA eukaryotes that acquired steroid biosynthesis genes from myxobacteria are inferred to have produced cycloartenol similar to LECA. Several Group II bacteria (that are not myxobacteria) are known to produce nonsteroidal polycyclic isoprenoids (34), but these species form a subcluster within Group II. The amino acid residues that are involved in the enzymatic formation of the steroid structure are not conserved in this subcluster. In contrast, resurrected OSC sequences retain the amino acid residues that are critical to steroid formation. Correspondingly, nonsteroidal isoprenoids are not detected in resurrected OSC products (Fig. 4). Thus, the OSC gene in bacteria that produce nonsteroidal isoprenoids is inferred to have diverged from a steroid-producing, ancestral OSC. In contrast to Group II, Group I contains mostly lanosterol-producing bacteria. The observation that myxobacteria are distributed in both Groups I and II indicates that myxobacterial OSC functionally diverged into two groups—lanosterol and cycloartenol producers.

By contrast, cyclized products were not detected from the resurrected myxobacterial OSC before this functional divergence. This resurrected protein possessed the conserved XXDC motif that is necessary for the initiation of cyclization (17). The protein also retained the conserved amino acid residues that are involved in the formation of steroid carbon structure (34). This observation indicates that there are additional amino acid residues that are critical to the catalytic activity of OSC. Distant OSC homologs from myxobacteria (OSC-like proteins) that form sister clades to the OSC clade (Fig. 4) are generally shorter than OSC and lack several motifs that are conserved in the OSC clade. These OSC-like proteins lack the XXDC motif and thus are unlikely to cyclize oxidosqualene. OSC-like proteins may either have a different function from the conventional terpene cyclization or take a different substrate other than oxidosqualene and squalene. One possibility is that the common ancestor of the OSC clade partially retained the traits of OSC-like proteins, even though the XXDC motif was present. In this scenario, the common ancestor of the OSC clade may not have yet evolved OSC-enzymatic activity. The OSC function then evolved twice independently in Group I and Group II, each generating a different steroid product (lanosterol and cycloartenol, respectively) (Fig. 4).

Membrane Property of Pre-LECA Eukaryotes. A cycloartenol-producing OSC gene is inferred to have been transferred from myxobacteria to pre-LECA eukaryotes together with seven other Stage 2 genes, while Stage 1 genes were probably transferred to eukaryotes separately from Stage 2 genes. Together, this initial steroid biosynthesis gene set theoretically had the ability to produce C-14 and C-4 demethylated steroids (i.e., zymosterol and 4-methylzymosterol) as end products (Fig. 5). Two additional

Stage 2 genes (SMT and ERG2) may have also been transferred from myxobacteria to pre-LECA eukaryotes, but it is not conclusive (Fig. 1). The initial gene set and the extended gene set that additionally includes SMT and ERG2 genes are overlapping with part of the modern steroid biosynthesis pathway from Archaeplastida and Opisthokonta (Fig. 5, light blue and yellow colors). Interestingly, the inferred steroid profile of pre-LECA eukaryotes is a mixture of steroids from these two eukaryotic groups, while the biosynthesis pathway itself is more similar to the Archaeplastida pathway. SMT and ERG2 genes enable pre-LECA to convert zymosterol/4-methylzymosterol into episterol/24-methylenelophenol. These four steroids may have been among the earliest steroids in pre-LECA eukaryotes.

Our inferred pre-LECA gene sets nearly completely lack the desaturation-modifying genes that are utilized to further adjust the biophysical property of steroids beyond demethylation. Thus, modern steroids, such as cholesterol and campesterol, are inferred to have been absent in pre-LECA eukaryotes. This observation implies that the cellular membrane of pre-LECA eukaryotes may not have been fully established relative to the modern eukaryotic membrane. Subtle structural differences in steroids are known to have large functional affects (9, 35). For instance, cholesterol is universally distributed in Metazoa and is well known for its ability to form a lipid raft that is critical for membrane regulation (36). In contrast, other steroids, such as desmosterol and zymosterol, have low-to-intermediate, raft-forming abilities, even though the difference between cholesterol and those steroids is the presence/absence (and position) of a double bond (35, 37). Therefore, steroid-based cellular processes, such as membrane regulation and cell signaling in pre-LECA eukaryotes, are likely to have been still evolving and/or substantially different when compared to LECA and modern eukaryotes.

Steroid Biosynthesis and Eukaryogenesis. The vast majority of modern eukaryotes utilize steroids to perform endocytosis. The near ubiquity of steroids and steroid-required endocytosis in eukaryotes suggests that LECA likely possessed a similar ability for endocytosis. Steroid biosynthesis is an oxygen-dependent process and is currently found only in aerobes, although the required oxygen concentration is low (38). Analyses of resurrected OSC demonstrate the presence of ancestral steroid biosynthesis and the utilization of molecular oxygen in both LECA and pre-LECA eukaryotes that acquired steroid biosynthesis genes from myxobacteria. This is consistent with a previous inference for the
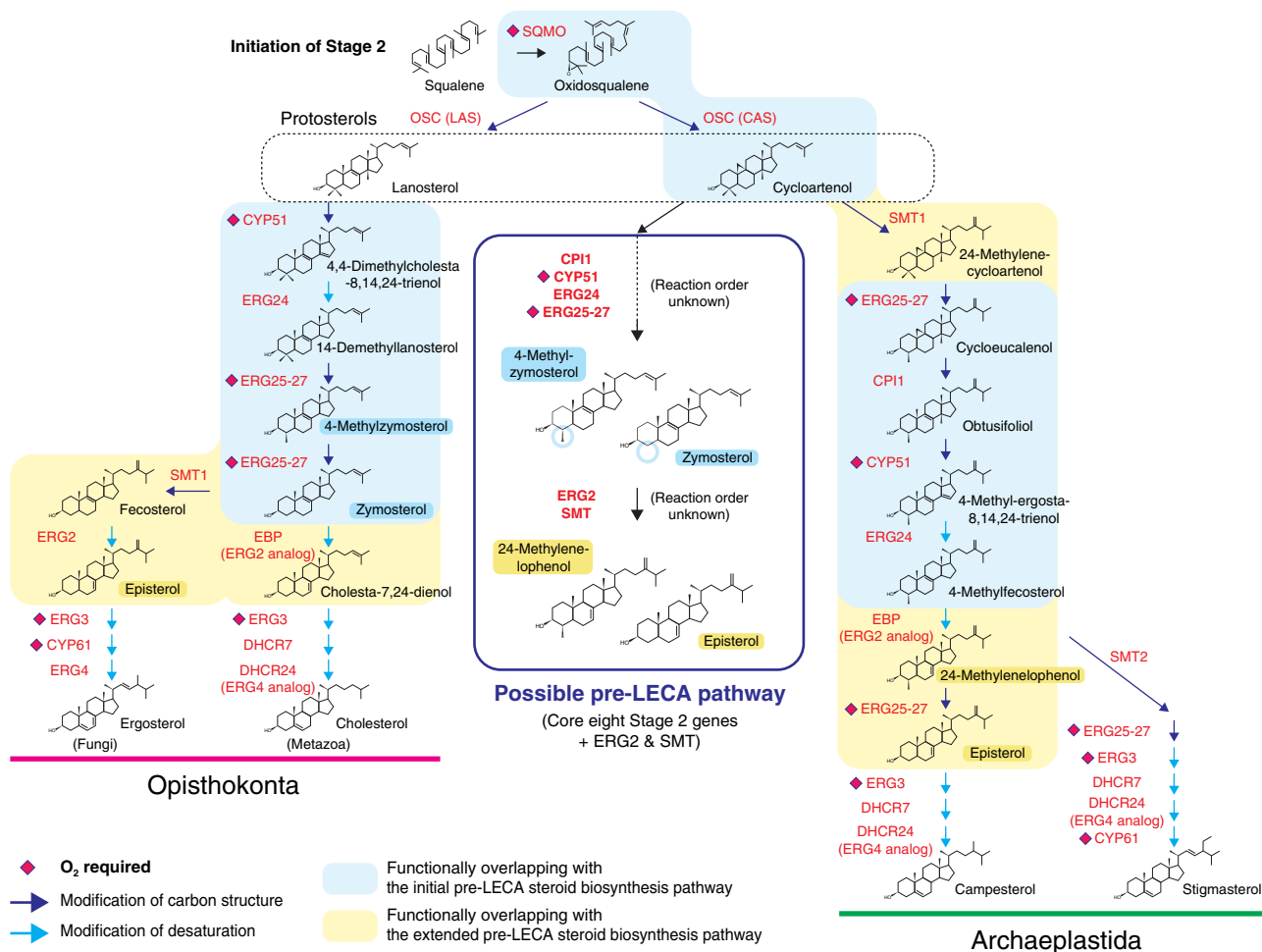
Fig. 5. Hypothetical steroid biosynthesis pathway for pre-LECA eukaryotes that acquired eight Stage 2 genes and, additionally, ERG2 and SMT genes (Stage 1 not displayed). Steroid biosynthesis pathways in modern eukaryotes other than Opisthokonta and Archaeplastida are largely unknown. The inferred pre-LECA pathway functionally overlaps with part of the modern steroid biosynthesis pathways for Archaeplastida and Opisthokonta (light blue and yellow colors). Since the reaction order of the pre-LECA pathway is unknown, the identity of intermediate steroid compounds is not clear and thus only possible end products are shown. The C-24 ethylation of steroids for the stigmasterol biosynthesis requires a pair of paralogous SMT genes (SMT1 and SMT2) and thus is inferred to have been not possible in both pre-LECA eukaryotes and LECA, because they had only a single copy of the gene. Abbreviations: CAS, cycloartenol synthase and LAS, lanosterol synthase. See *SI Appendix*, Fig. S1 for gene abbreviations.

origin of steroid biosynthesis shortly after the Great Oxidation Event at 2.4 Ga (16).

The aerobiosis of eukaryotes is generally inferred to have been obtained through the acquisition of mitochondria. However, a widely accepted hypothesis that the acquisition of mitochondria proceeded through endocytosis requires steroids in the cellular membrane of eukaryotes prior to mitochondria acquisition. This implies that eukaryotes were already aerobic before mitochondria acquisition and contradicts the current assumption for the origin of aerobiosis in eukaryotes. This paradoxical scenario implies that either early eukaryotes were actually aerobic before mitochondria acquisition or mitochondria acquisition was not through a steroid-required mechanism (e.g., E^3 model and inside-out model) (39–41). In either case, the cellular membrane of pre-LECA eukaryotes that acquired mitochondria was likely to be different from the modern, eukaryotic cellular membranes that have a fine-tuned steroid composition. Thus, models for mitochondria acquisition are enhanced when they account for the membrane properties of pre-LECA cells.

Among current hypotheses for eukaryogenesis, the evolutionary relationship between myxobacteria and early eukaryotes is discussed only in the so-called syntrophic model (42, 43). In this model, anaerobic ancestors of myxobacteria had a syntrophic relationship with archaea. Since the last common ancestor of modern myxobacteria is inferred to have been aerobic based on our analyses, anaerobic ancestors before the evolution of aerobic myxobacteria may have had a syntrophic relationship to early eukaryotes. The phylogenetic presence of uncultured (obligately) anaerobic species branching near the aerobic, myxobacterial clade has been suggested (25). Some of these anaerobic taxa may retain ancient traits of preaerobic, myxobacterial ancestors. Further addition of molecular data, and the culturing of those enigmatic, anaerobic species, may provide information about their metabolisms (sulfate reduction in particular) and their relationships to early eukaryotes. Deltaproteobacteria and eukaryotes may have coevolved in close proximity for a particular amount of time, during which a deltaproteobacterial, syntrophic partner gradually evolved into myxobacteria, acquired (or evolved) steroid biosynthesis genes, and transferred those genes to a eukaryotic partner. Our study suggests necessary modifications of the current eukaryogenesis models to accommodate the important relationship between aerobic deltaproteobacteria (myxobacteria) and early eukaryotes.

Methods

Bioinformatics Analysis. There are 30 genes involved in the eukaryotic-type steroid biosynthesis. Some genes are not specific to isoprenoid biosynthesis and/or distributed in a limited number of species. These genes were excluded from our dataset, and eventually, 24 genes were selected (*SI Appendix, Methods*). Representative sequences for individual steroid biosynthesis enzymes were identified from UniProt (https://www.uniprot.org/) and references cited therein (see *SI Appendix, Methods* for accession numbers). Obtained protein sequences were utilized as seeds to identify additional homologous sequences. Homologous protein sequences were retrieved for all three domains of life from GenBank (https://www.ncbi.nlm.nih.gov/), using the protein Basic Local Alignment Search Tool (BLAST) and position-specific–iterated BLAST (PSI-BLAST; 44, 45), with the cutoff threshold of $<1 \times 10^{-5}$.

Sequences were aligned using Muscle version 3.8.31 (46) and Profile Multiple Alignment with Local Structure 3D (PROMALS3D) for structural-based alignments (47). Phylogenetic trees were constructed by maximum likelihood inference using Randomized Axelerated Maximum Likelihood (RAxML) version 8.2.11 and IQ-TREE version 2.1.06 (48) and by Bayesian inference using MrBayes version 3.2.6 (49) and PhyloBayes version 4.1c (50) (see *SI Appendix, Methods* for details). Substitution models were selected using ModelFinder in IQ-TREE. The top tree using each of the four models was generated for individual genes, and the tree topologies were compared among one another. For ASR, representative sequences were selected from individual terpene cyclase families, and the total number of 88 sequences were utilized. Ancestral sequences were reconstructed using FastML (version 3.1) (see *SI Appendix, Methods* for details) (51). Three internal nodes were selected for laboratory resurrection (Fig. 4).

Protein Expression and Lipid Analysis. Genes encoding resurrected OSC sequences were synthesized and codon optimized for expression in *Escherichia coli* cells. Steroid biosynthesis genes were cloned into three different plasmids BbA5c-MevT(CO)-MBIS (Addgene #35152) (52), pTrc99a derivative (53), and pSRKGm derivative (54) (see *SI Appendix, Methods* for details). These three plasmids were transformed onto *E. coli* BL21 (DE3) (New England Biolabs #C2527H). Cells were incubated in terrific broth medium at 37 °C/250 rpm. Expression was induced by adding 500 µM isopropyl β-D-1-thio-galactopyranoside when OD_{600} reached 0.6. Cells were further incubated for 24 h. Lipids were extracted by a modified Bligh–Dyer extraction method (55) (*SI Appendix, Methods*). In short, harvested cells were suspended in the mixture of water and organic solvents and were sonicated. The organic phase was separated by centrifugation. The alcohol fraction that contains steroids were purified using silica gel column chromatography. Steroids were derivatized and were subject to gas chromatography mass spectroscopy measurements (see *SI Appendix, Methods* for details). Steroids were identified based on their retention time and comparison with laboratory standards and published mass spectra.

Data Availability. All study data are included in the article and/or *SI Appendix*.

1. A. J. Roger, S. A. Muñoz-Gómez, R. Kamikawa, The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
2. G. van Meer, D. R. Voelker, G. W. Feigenson, Membrane lipids: Where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**, 112–124 (2008).
3. M. F. Siliakus, J. van der Oost, S. W. M. Kengen, Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* **21**, 651–670 (2017).
4. K. Zaremba-Niedzwiedzka *et al.*, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
5. T. A. Williams, P. G. Foster, C. J. Cox, T. M. Embley, An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).
6. L. Eme, A. Spang, J. Lombard, C. W. Stairs, T. J. G. Ettema, Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
7. H. Pichler, H. Riezman, Where sterols are required for endocytosis. *Biochim. Biophys. Acta Biomembr.* **1666**, 51–61 (2004).
8. E. Sezgin, I. Levental, S. Mayor, C. Eggeling, The mystery of membrane organization: Composition, regulation and roles of lipid rafts. *Nat. Rev. Mol. Cell Biol.* **18**, 361–374 (2017).
9. A. Heese-Peck *et al.*, Multiple functions of sterols in yeast endocytosis. *Mol. Biol. Cell* **13**, 2664–2680 (2002).
10. A. C. B. Cato, A. Nestl, S. Mink, Rapid actions of steroid receptors in cellular signaling pathways. *Sci. STKE* **2002**, re9 (2002).
11. 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).
12. E. Desmond, S. Gribaldo, Phylogenomics of sterol synthesis: Insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biol. Evol.* **1**, 364–381 (2009).
13. G. Luo, H. Yang, T. J. Algeo, C. Hallmann, S. Xie, Lipid biomarkers for the reconstruction of deep-time environmental conditions. *Earth Sci. Rev.* **189**, 99–124 (2019).
14. J. J. Brocks *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578–581 (2017).
15. Y. Hoshino *et al.*, Cryogenian evolution of stigmasteroid biosynthesis. *Sci. Adv.* **3**, e1700887 (2017).
16. D. A. Gold, A. Caron, G. P. Fournier, R. E. Summons, Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* **543**, 420–423 (2017).
17. D. W. Christianson, Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442 (2006).
18. Y. Hoshino, E. A. Gaucher, On the origin of isoprenoid biosynthesis. *Mol. Biol. Evol.* **35**, 2185–2197 (2018).
19. 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).
20. E. A. Gaucher, S. Govindarajan, O. K. Ganesh, Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–707 (2008).
21. R. N. Randall, C. E. Radford, K. A. Roof, D. K. Natarajan, E. A. Gaucher, An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.* **7**, 12847 (2016).

22. M. Rohmer, The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **16**, 565–574 (1999).

23. J.-J. Pan et al., Biosynthesis of squalene from farnesyl diphosphate in bacteria: Three steps catalyzed by three enzymes. *ACS Cent. Sci.* **1**, 77–82 (2015).

24. J. I. Martien et al., Systems-level analysis of oxygen exposure in *Zymomonas mobilis*: Implications for isoprenoid production. *mSystems* **4**, e00284-18 (2019).

25. K. I. Mohr, Diversity of myxobacteria-we only see the tip of the iceberg. *Microorganisms* **6**, 84 (2018).

26. F. Ramel et al., Membrane-bound oxygen reductases of the anaerobic sulfate-reducing Desulfovibrio vulgaris Hildenborough: Roles in oxygen defence and electron link with periplasmic hydrogen oxidation. *Microbiology* **159**, 2663–2673 (2013).

27. M. Schoeffler et al., Growth of an anaerobic sulfate-reducing bacterium sustained by oxygen respiratory energy conservation after O_2-driven experimental evolution. *Environ. Microbiol.* **21**, 360–373 (2019).

28. M. Y. Kellermann, M. Y. Yoshinaga, R. C. Valentine, L. Wörmer, D. L. Valentine, Important roles for membrane lipids in haloarchaeal bioenergetics. *Biochim. Biophys. Acta* **1858**, 2940–2956 (2016).

29. H. Hayakawa et al., Modified mevalonate pathway of the archaeon *Aeropyrum pernix* proceeds via *trans*-anhydromevalonate 5-phosphate. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10034–10039 (2018).

30. A. Pearson, M. Budin, J. J. Brocks, Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15352–15357 (2003). Correction in: *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3991 (2004).

31. T. Thiergart, G. Landan, M. Schenk, T. Dagan, W. F. Martin, An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol. Evol.* **4**, 466–485 (2012).

32. E. A. Gaucher, X. Gu, M. M. Miyamoto, S. A. Benner, Predicting functional divergence in protein evolution by site-specific rate shifts. *Trends Biochem. Sci.* **27**, 315–321 (2002).

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

34. A. B. Banta, J. H. Wei, C. C. C. Gill, J.-L. Giner, P. V. Welander, Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 245–250 (2017).

35. T. Róg, M. Pasenkiewicz-Gierula, I. Vattulainen, M. Karttunen, Ordering effects of cholesterol and its analogues. *Biochim. Biophys. Acta* **1788**, 97–121 (2009).

36. D. Lingwood, K. Simons, Lipid rafts as a membrane-organizing principle. *Science* **327**, 46–50 (2010).

37. T. J. LaRocca et al., Proving lipid rafts exist: Membrane domains in the prokaryote Borrelia burgdorferi have the same properties as eukaryotic lipid rafts. *PLoS Pathog.* **9**, e1003353 (2013).

38. J. R. Waldbauer, D. K. Newman, R. E. Summons, Microaerobic steroid biosynthesis and the molecular fossil record of Archean life. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13409–13414 (2011).

39. H. Imachi et al., Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519–525 (2020).

40. D. A. Baum, B. Baum, An inside-out origin for the eukaryotic cell. *BMC Biol.* **12**, 76 (2014).

41. W. Martin, M. J. Russell, On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 59–83, discussion 83–85 (2003).

42. D. Moreira, P. López-García, Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).

43. P. López-García, D. Moreira, The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* **5**, 655–667 (2020).

44. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

45. S. F. Altschul et al., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

46. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

47. J. Pei, B.-H. Kim, N. V. Grishin, PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).

48. B. Q. Minh et al., IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). Correction in: *Mol. Biol. Evol.* **37**, 2461 (2020).

49. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

50. N. Lartillot, H. Philippe, A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).

51. H. Ashkenazy et al., FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584 (2012).

52. P. P. Peralta-Yahya et al., Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2**, 483 (2011).

53. E. Amann, B. Ochs, K. J. Abel, Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**, 301–315 (1988).

54. S. R. Khan, J. Gaines, R. M. Roop II, S. K. Farrand, Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. *Appl. Environ. Microbiol.* **74**, 5053–5062 (2008).

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

EVOLUTION