



**HAL**  
open science

# Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature

Elie Le Quemener, S. Gribaldo

► **To cite this version:**

Elie Le Quemener, S. Gribaldo. Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution*, 2009, 1, pp.364-381. 10.1093/gbe/evp036 . hal-01604999

**HAL Id: hal-01604999**

**<https://hal.science/hal-01604999>**

Submitted on 31 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## RESEARCH ARTICLE

# Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature

Elie Desmond and Simonetta Gribaldo

Institut Pasteur, Unité de Biologie Moléculaire du gène chez les Extrêmophiles, Paris, France

The availability of complete genomes from a wide sampling of eukaryotic diversity has allowed the application of phylogenomics approaches to study the origin and evolution of unique eukaryotic cellular structures, but these are still poorly applied to study unique eukaryotic metabolic pathways. Sterols are a good example because they are an essential feature of eukaryotic membranes. The sterol pathway has been well dissected in vertebrates, fungi, and land plants. However, although different types of sterols have been identified in other eukaryotic lineages, their pathways have not been fully characterized. We have carried out an extensive analysis of the taxonomic distribution and phylogeny of the enzymes of the sterol pathway in a large sampling of eukaryotic lineages. This allowed us to tentatively indicate features of the sterol pathway in organisms where this has not been characterized and to point out a number of steps for which yet-to-discover enzymes may be at work. We also inferred that the last eukaryotic common ancestor already harbored a large panel of enzymes for sterol synthesis and that subsequent evolution over the eukaryotic tree occurred by tinkering, mainly by gene losses. We highlight a high capacity of sterol synthesis in the myxobacterium *Plesiocystis pacifica*, and we support the hypothesis that the few bacteria that harbor homologs of the sterol pathway have likely acquired these via horizontal gene transfer from eukaryotes. Finally, we propose a potential candidate for the elusive enzyme performing C-3 ketoreduction (ERG27 equivalent) in land plants and probably in other eukaryotic phyla.

## Introduction

Triterpenoids are a large family of lipids that are widely distributed in Bacteria (hopanoids) and Eukaryotes (sterols). Sterols are present in all eukaryotes, where they are essential and are involved in both intra- and intercellular signaling and in the organization of membranes. In membranes they affect fluidity and permeability (London 2002; Tyler et al. 2009) and are major players in the formation of lipid rafts, which are regions of reduced fluidity formed by the close association of sterols with sphingolipids (Jacobson and Dietrich 1999; Brown and London 2000; Anderson and Jacobson 2002). Proteins involved in concerted functions such as cell signaling can be associated through their selective incorporation into lipid rafts (Melkonian et al. 1999; Simons and Toomre 2000). Moreover, sterols play a key role in typical eukaryotic features such as phagocytosis. For example, genes involved in sterol biosynthesis have been shown to be selectively upregulated in the amoebozoan *Dictyostelium discoideum* during phagocytosis (Sillo et al. 2008). Eukaryotes that are not able to synthesize sterols have to obtain them from food, such as is the case, for example, of insects and of most marine invertebrates.

Bacterial and eukaryotic triterpenoids belong to the class of isoprenoids. Isoprenoids are all derived from their universal precursor isopentenyl diphosphate (IPP). IPP synthesis can follow two different routes, either via the mevalonate pathway in Archaea and some Eukaryotes or via the 2-C-methyl-D-erythritol 4-phosphate pathway in Bacteria and other Eukaryotes (Rohmer et al. 1993; Boucher et al. 2004; Volkman 2005). IPP is the precursor of squalene and its cyclization products. In Bacteria, hopanoids are synthesized from the cyclization of squalene by a squalene-hopene cyclase (SHC) in a process that does not require

oxygen. On the other hand, in Eukaryotes a first step (squalene monooxygenation, ERG1) adds an oxygen to squalene leading to squalene epoxide, which is then cyclized in a second step either to lanosterol or to cycloartenol by enzymes homologous to SHC (ERG7). Because sterol synthesis is widely distributed in eukaryotes, it can be assumed that the last eukaryotic common ancestor (LECA) already synthesized sterols. However, the origin of this pathway is not clear. The cyclization of squalene and the following steps are very demanding in oxygen. For example, 11 molecules of oxygen are required for the synthesis of one molecule of cholesterol (Summons et al. 2006). Organisms that live under anaerobic conditions therefore resort to external sources of sterol by ingesting other eukaryotes. Moreover, some eukaryotes that are able to synthesize sterols in the presence of oxygen have to acquire them from food when under anaerobic conditions, as it has been shown in yeast (Schneider 2007). A commonly accepted hypothesis is thus that the pathway of sterol biosynthesis appeared after the emergence of oxygenic photosynthesis and the oxygenation of the atmosphere and oceans, thought to have occurred between 2.7 and 2.4 Ga (Summons et al. 2006). It has been recently put forward that the initial role of sterols in eukaryotes may have been that of protection against oxidative stress when oxygen levels rose (Galea and Brown 2009). Moreover, it has been proposed that the ancestral pathway made cycloartenol as a final product and that the rising concentrations of oxygen in the atmosphere would have led to an evolution of the pathway beyond cycloartenol toward more stable sterol compounds (Ourisson and Nakatani 1994).

Intriguingly, although archaea are not known to be able to synthesize sterols, a few bacteria have this ability. *Methylococcus capsulatus* (gamma-Proteobacteria) has been shown to produce different sterols (Bouvier et al. 1976), and *Gemmata obscuriglobus* (Planctomycetales) produces lanosterol and its isomer parkeol (Pearson et al. 2003). According to their sterol-synthesizing abilities, homologous of the first two enzymes of the pathway (ERG1

Key words: sterols, phylogenomics, eukaryotes, evolution.

E-mail: simonetta.gribaldo@pasteur.fr.

*Genome. Biol. Evol.* 1:364–381.

doi:10.1093/gbe/evp036

Advance Access publication September 10, 2009

© The Author(s) 2009. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution*.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

and ERG7) have been found in *M. capsulatus* and *G. obscuriglobus* (Pearson et al. 2003; Lamb et al. 2007). A variety of more or less elaborated sterols have been isolated from a number of Myxobacteria (delta-Proteobacteria), such as cycloartenol in *Stigmatella aurantiaca*, and 7-cholesten-3 $\beta$ -ol synthesized by some strains of *Nannocystis* sp. (Bode et al. 2003). The elaborated sterols synthesized by Myxobacteria imply that these bacteria not only have the first two enzymes of the pathway but also up to four or five more downstream. A gene coding for a homolog of the second enzyme of the pathway (ERG7) has been sequenced from *S. aurantiaca* (Bode et al. 2003). However, mutants where the gene was inactivated showed no noticeable phenotype, indicating that sterol production is not essential in this bacterium (Bode et al. 2003). The function of these lipids in bacteria remains indeed unknown (Bode et al. 2003). Previous phylogenetic analyses have shown that ERG1 homologs from *Methylococcus* and *Gemmata* are specifically related to their eukaryotic counterparts (Pearson et al. 2003). Concerning ERG7, it has been shown that *Methylococcus*, *Gemmata*, and *Stigmatella* homologs are also closely related to their eukaryotic counterparts (Pearson et al. 2003; Chen et al. 2007). The closeness of ERG1 and ERG7 among these sterol-producing bacteria and eukaryotes has been explained by a possible ancient horizontal gene transfer (HGT) (Pearson et al. 2003; Chen et al. 2007). However, it remains unclear whether these two enzymes originated in bacteria or in eukaryotes. It is likely that the assembly of the pathway started with these two initial steps. Subsequently, other enzymatic steps would have added to produce more elaborated sterols. However, it is not known when this would have happened and which type of sterols the LECA would have been potentially able to produce.

The diversity and nature of sterols and the pathways leading to these compounds have been thoroughly studied in vertebrates, fungi, and land plants. Characterizations of sterols from other organisms have unveiled a wide variety of molecules, among which not only the same sterols known in animals, fungi, and land plants but also other types of sterols with insaturations at various positions of the cycle or in the side chain, and possibly alkylations or inclusion of a cyclopropane ring mostly in C-24 or C-22 (Volkman 2003). Based on these characterizations, fossil steranes found in sediments are used as biomarkers for past eukaryotic life (Kodner, Pearson, et al. 2008). However, relatively little is known about the structure of the pathway in these organisms. A large number of complete genomes are now available from representatives of major eukaryotic phyla, as well as from representatives of sterol-producing bacteria, notably the two Myxobacteria *S. aurantiaca* and *Plesiocystis pacifica*. We have thus carried out an extensive analysis of the taxonomic distribution and phylogeny of the enzymes of the sterol pathway in a large sampling of organisms representing eukaryotic diversity for which complete genomic sequences are available. The use of complete genomes is in fact essential to infer presence or absence of homologs of the different enzymes. We tentatively reconstructed the potential structure of the sterol pathway in diverse organisms where it has not been characterized and point out a number of steps for which yet-to-discover enzymes

may be at work. Moreover, we reconstructed the potential ancestral abilities of sterol production in the LECA and the subsequent evolution of the pathway over the eukaryotic tree, which appears to have occurred by tinkering, mainly by gene losses. We highlight a high capacity of sterol synthesis in the Myxobacterium *P. pacifica*, and we support the hypothesis that the few bacteria that harbor homologs of enzymes of the sterol pathway have likely acquired these via ancient HGT from eukaryotes. Finally, an analysis of phyletic patterns allowed us to propose a potential candidate for the elusive enzyme performing C-3 ketoreduction (ERG27 equivalent) in land plants and other eukaryotic phyla, highlighting the power of phylogenomics approaches to the study of biochemical pathways.

## Materials and Methods

Homologs of the enzymes characterized in fungi, vertebrates, and land plants were identified using the BlastP program version 2.2.1.8 (Altschul et al. 1997) and the HMMER tools (<http://hmmer.janelia.org/>) on a local database of complete genomes (38 eukaryotes: Opisthokonta/metazoans: *Drosophila melanogaster*, *Tribolium castaneum*, *Caenorhabditis elegans*, *Homo sapiens*, *Mus musculus*, *Danio rerio*; Opisthokonta/choanoflagellates: *Monosiga brevicollis*; Opisthokonta/fungi: *Aspergillus fumigatus*, *Neurospora crassa*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Ustilago maydis*, *Cryptococcus neoformans*, *Encephalitozoon cuniculi*; Ameobozoa: *D. discoideum*, *Entamoeba histolytica*; Plantae/land plants: *Arabidopsis thaliana*, *Oryza sativa*; Plantae/green algae: *Chlamydomonas reinhardtii*, *Ostreococcus tauri*; Plantae/red algae: *Cyanidioschyzon merolae*; Alveolata/ciliates: *Tetrahymena thermophila*, *Paramecium tetraurelia*; Alveolata/apicomplexans: *Cryptosporidium parvum*, *Plasmodium falciparum*, *Plasmodium yoelii*, *Theileria annulata*; Heterokonta/oomycetes: *Phytophthora ramorum*; Heterokonta/diatoms: *Thalassiosira pseudonana*; Heterokonta/brown algae: *Aureococcus anophagefferens*; Excavata/kinetoplastids: *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania infantum*, *Leishmania major*, *Leishmania braziliensis*; Excavata/heterolobozoans: *Naegleria gruberi*; Excavata/diplomonads: *Giardia lamblia*; Excavata/parabasalids: *Trichomonas vaginalis*; 586 bacteria, 48 archaea). Most of the complete genomes were retrieved from the Refseq database at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>), except for *N. gruberi*, *O. tauri*, *A. anophagefferens*, *P. ramorum*, and *T. pseudonana* genomes which were retrieved from the JGI database ([http://genome.jgi-psf.org/euk\\_home.html](http://genome.jgi-psf.org/euk_home.html)) and for *C. merolae* genome which was retrieved from <http://merolae.biol.s.u-tokyo.ac.jp>.

When a particular gene was not found in some species, we searched the nucleotide sequences of these complete genomes using TBlastN. The retrieved proteins were aligned using MUSCLE 3.6 (Edgar 2004), the alignments were edited and refined manually using the ED program from the MUST package (Philippe 1993).

Phylogenies were reconstructed for each protein. Regions where homology was doubtful were manually

removed from the data sets before phylogenetic analysis using the NET program of the MUST package (Philippe 1993), and phylogenetic trees were computed with PHYML (Guindon and Gascuel 2003) and MrBayes (Ronquist and Huelsenbeck 2003; Huelsenbeck and Ronquist 2001). PHYML trees were calculated using the Whelan and Goldman model, a gamma correction to take into account the heterogeneity of evolutionary rates across sites (four discrete classes of sites, an estimated alpha parameter, and an estimated proportion of invariable sites), and subtree pruning and regrafting topology searches. The robustness of each branch was estimated by a nonparametric bootstrap procedure implemented in PHYML (100 replicates of the original data set and the same parameters). MrBayes trees were calculated with an estimated fixed rate model and a gamma correction as for the PHYML trees. The Markov Chain Monte Carlo runs were performed with four chains on 1,000,000 generations sampled every 100 generation, and the final tree was summarized discarding the first 1,000 generations.

## Results

### Canonical Pathways

The pathway of sterol synthesis has been thoroughly studied in vertebrates, fungi, and land plants (fig. 1). A number of enzymatic reactions are found in the three pathways (fig. 1A). In these three “canonical pathways,” squalene is oxygenated in a first step and cyclized in a second step either to lanosterol in vertebrates and fungi or to cycloartenol in land plants (fig. 1A). Cycloartenol includes a cyclopropane ring between carbon C-9 and C-10 (fig. 1A and B). This cycle is cleaved at a later step of the pathway by a specific land plants enzyme (CPI1) leading to typical sterols with four-carbon cycles (fig. 1A and B). Lanosterol or cycloartenol is then converted to cholesterol in vertebrates, ergosterol in fungi, and to campesterol, sitosterol, and stigmasterol in land plants by a succession of oxidations, reductions, and demethylations (fig. 1A). Most of these steps are shared in the three pathways, although they do not occur in the same order. C-4 demethylation occurs twice in the three canonical pathways and is performed both times via three concerted steps (C-4 methyl oxidation, C-3 dehydrogenation/C-4 decarboxylation, and C-3 ketoreduction; fig. 1C). In vertebrates and fungi, the two C-4 demethylations occur one after the other (fig. 1A). Conversely, in land plants the two C-4 demethylations do not occur one after the other, and two C-4 methyl oxidases (SMO1 and SMO2) perform these steps (fig. 1C), each being specific for a particular substrate (SMO1 acting on 24-methylenecycloartenol, and SMO2 acting several steps later on 24-ethylenelophenol or 24-methylenelophenol; fig. 1A). In land plants, the enzyme responsible for C-3 ketoreduction is unknown (fig. 1C and Bouvier et al. 2005). It has been shown in *Saccharomyces* that C-4 demethylation takes place at the endoplasmic reticulum membrane through the anchoring of ERG26 and ERG27 by ERG28 (Gachotte et al. 2001). It is likely that this anchoring also occurs in vertebrates and land plants because they have a copy of ERG28 (fig. 1C). However, it has been shown that ERG28 is not essential in *Saccharomyces* (Gachotte et al. 2001), indicating that this

enzyme may be dispensable. The fungi pathway is very similar to that of vertebrates. However, the ergosterol produced contains 28 carbons contrary to cholesterol, which contains only 27 carbons (fig. 1A). The additional carbon present in fungi ergosterol is added by the C-24 methylase ERG6, which is in fact absent in vertebrates (fig. 1C). On the other hand, the sitosterol and stigmasterol produced by land plants contain 29 carbons (fig. 1A). These two additional carbons are added in two steps catalyzed by two copies of ERG6: SMT1 performing C-24 methylation and SMT2 performing C-28 methylation (fig. 1A and C).

### Phylogenomics of the Sterol Pathway

Starting from the characterized genes in fungi, vertebrates, and land plants, we investigated their taxonomic distribution and phylogeny in complete genomes from a broad taxonomic sample of eukaryotic diversity (38 genomes containing representatives of main eukaryotic supergroups: one choanoflagellate, six animals, seven fungi, two amoebozoans, two land plants, two green algae, one red algae, two ciliates, four apicomplexans, one brown algae, one diatom, one oomycete, five kinetoplastids, one heterolobosean, one diplomonad, one parabasalid, for a complete list see Materials and Methods) as well as from complete genomes from 586 bacteria and 48 archaea. Phylogenetic analysis allowed us assigning each of these homologs to an orthology group and thus to a potential function (table 1 and Supplementary Material 1, Supplementary Material online). Consistently with the oxygen demand of the pathway, orthologs are totally absent in strict anaerobic organisms (*Encephalitozoon*, *Giardia*, *Entamoeba*, *Cryptosporidium*, and *Trichomonas*) but also in aerobic ones (*Plasmodium*, *Theileria*, *Drosophila*, and *Tribolium*) (not shown). We found only a partial number of orthologs in the oomycete *P. ramorum* (ERG3 and DHCR7), and the two ciliates (ERG3 and FK) (not shown), consistent with the notion that these organisms do not synthesize sterols (Trigos et al. 2005). These orthologs may be in fact used to process sterols obtained from diet.

In fungi, vertebrates, and land plants, the same enzymatic steps are generally performed by homologous enzymes (table 1). Two exceptions are the delta-8, delta-7 isomerization step performed by the nonhomologous ERG2 in fungi and EBP/HYD1 in vertebrates/land plants and the delta-24 reduction performed by the nonhomologous ERG4 in fungi and DHCR24/DWF1 in vertebrates/land plants (table 1). A few enzymes performing different steps in fungi, vertebrates, and land plants are evolutionarily related (table 1). For example, ERG4-(DHCR7/DWF5)-(ERG24/TM7SF2/FK) all belong to a large protein family. In a global phylogeny of this family, members with different functions segregate into different orthology groups (Supplementary Material 1, Supplementary Material online) indicating that these functions arose by gene duplication. It has to be noted that ERG24/TM7SF2 and FK have the same function but belong to paralogous clusters (Supplementary Material 1, Supplementary Material online). Finally, ERG3/ERG25 are paralogs within a large family of oxidases (not shown), as well as ERG5/ERG11

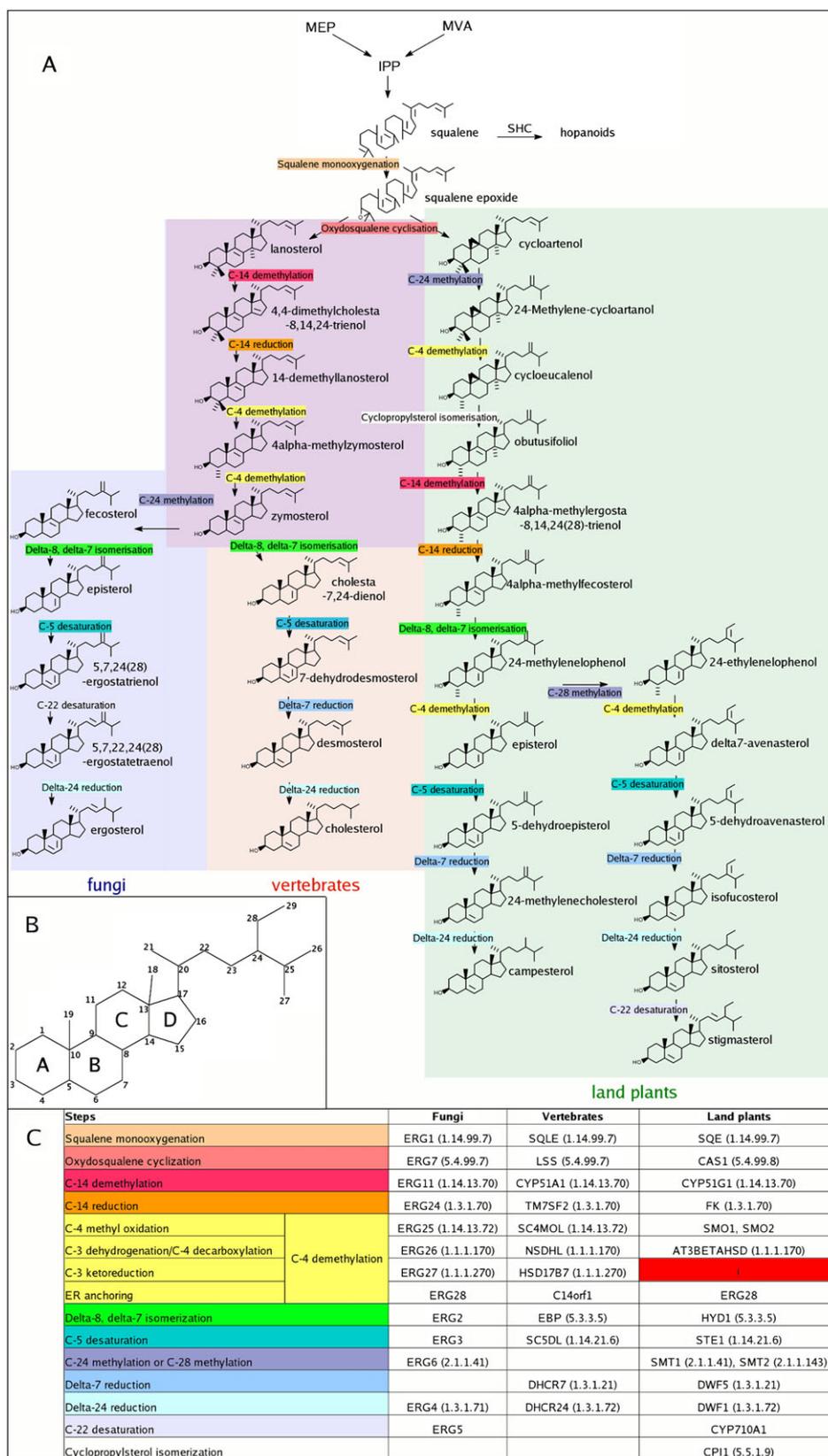


FIG. 1.—(A) Canonical pathways of sterol synthesis leading to land plants, fungi, and vertebrate sterols. Upstream of squalene, the mevalonate (MVA) and 2-C-methyl-D-erythritol 4-phosphate (MEP) ways leading to IPP are shown. Downstream of squalene, the bacterial pathway of hopanoid synthesis *via* SHC is also indicated. (B) Numbering of carbons and cycles of steroids. (C) Table indicating the names and EC numbers of fungi, vertebrates, and land plants genes corresponding to each step of the pathway shown in (A). The red box indicates that the gene performing C-3 ketoreduction is still unknown in land plants.

**Table 1**  
**Distribution of Orthologs of Enzymes of the Sterol Pathway in the Eukaryotic Genomes of Our Data Set**

Steps	Fungi	Vertebrates	Land plants	<i>M. brevicollis</i>	<i>D. discoideum</i>	<i>O. tauri</i>	<i>C. reinhardtii</i>	<i>C. merolae</i>	<i>T. pseudonana</i>	<i>A. anophagefferens</i>	Kinetoplastids	<i>N. gruberi</i>	LECA	
squalene monooxygenation	ERG1	SQLE	SQE	!	X	X	!	X	!	!	X	X	X	
oxydosqualene cyclization	ERG7	LSS	CAS1	X	X	X	X	X	X	X	X	X	X	
C-14 demethylation	ERG11	CYP51A1	CYP51G1	X	X	X	X	X	X	X	X	X	X	
C-14 reduction	ERG24	TM7SF2	FK*	XX*	X*	X	X	X*	X	X	X	X*	XX*	
C-4 methyl oxidation	C-4 demethylation	ERG25	SC4MOL	SMO1, SMO2	!	XX	X	!	XX	!	!	!	X	XX
C-3 dehydrogenation/C-4 decarboxylation		ERG26	NSDHL	AT3BETAHSD	X	X	X	X	X	X	X	X	X	X
C-3 ketoreduction		ERG27	HSD17B7		!	!	!	!	!	!	!	!	!	!
ER anchoring		ERG28	C14orf1	ERG28	!	X	!	X	X	!	!	!	!	X
delta-8, delta-7 isomerization	ERG2	EBP	HYD1	X	X		X		!	!	X	X	XX	
C-5 desaturation	ERG3	SC5DL	STE1	X		X	X	X	!	!	X	X	X	
C-24 methylation or C-28 methylation	ERG6		SMT1, SMT2	X	XX	XX	X	X	XX	XXX	X	X	XX	
delta-7 reduction		DHCR7	DWF5	X	!				X	X		X	X	
delta-24 reduction	ERG4	DHCR24	DWF1	XX	X		!		X	XX	X	X	XX	
C-22 desaturation	ERG5		CYP710A1	!	X		X	X			X	!	X	
cyclopropylsterol isomerization			CPI1		X	X	X	X	X	X		X	X	

The same color indicates homology. Presence is indicated by an X symbol. A double or triple X indicates presence of two or three copies. In the case of the paralogous families ERG24/TM7SF2-FK, an X with asterisk indicates the presence of orthologues of TM7SF2-FK, an X without asterisk indicates the presence of orthologues of ERG24, and a double X with asterisk indicates the presence of both paralogues (i.e. in *Monosiga*). In the case of the non-homologous families ERG4/DHCR24-DWF1 and ERG2/EBP-HYD1, the presence of one or the other copy is indicated by green and violet boxes, respectively, whereas the presence of both copies is indicated by a dark yellow box. Absences consistent with experimental characterizations are indicated by empty boxes, while absences not consistent with experimental characterizations are indicated by red boxes with exclamation points (see text for discussion).

within the large family of cytochrome P450 (not shown). Interestingly, we found that the fungi *Aspergillus* and *Neurospora* (Pezizomycotina) harbor additional homologs of enzymes not belonging to the canonical fungi pathway (EBP and DHCR24), as well as an additional SMT copy (Supplementary Material 1, Supplementary Material online). The first two enzymes bring redundant functions to these fungi, whereas the additional SMT copy may produce C-29 sterols, differently from the classical fungi ergosterol (C-28 sterol).

A number of sterols have been characterized from organisms present in our data set other than fungi, vertebrates, and land plants, to the exception of the green algae *O. tauri* and the red algae *C. merolae* (fig. 2) (Raederstorff and Rohmer 1987; Nes et al. 1990; Giner and Boyer 1998; Veron et al. 1998; Salimova et al. 1999; Roberts et al. 2003; Kodner, Summons, et al. 2008). With the combined knowledge of these sterols and our phylogenomic data, we sought to tentatively predict some additional features of the sterol pathways in these organisms. One first general information on the nature of the pathway can be obtained from the characteristics of ERG7 orthologs, which catalyze the cyclization of squalene epoxide either to lanosterol or to cycloartenol (fig. 1A and C), one of the most complex reactions catalyzed by a single enzyme (Summons et al. 2006). Among all enzymes of the sterol pathway, ERG7 is one of the most conserved at the sequence level and orthologs are indeed present in all species capable of de novo synthesis (table 1). The active sites are conserved and have

been largely studied (Summons et al. 2006). In particular, the production of lanosterol or cycloartenol in fungi, vertebrates, and land plants has been linked to the presence of particular residues in the active site of ERG7: Three positions are in fact differentially conserved between lanosterol synthases and cycloartenol synthases (Summons et al. 2006; fig. 3). In fungi and vertebrates, lanosterol synthase is characterized by T381, C/Q449, and V453, whereas cycloartenol synthase is characterized by Y381, H449, and I453 (Summons et al. 2006). However, as shown for kinetoplastids, position 381 can be variable (they have an Y but make lanosterol; Buckner et al. 2000), and position 449 can also be variable because *Methylococcus* and *Gemmata* have an H while they make lanosterol (Summons et al. 2006) (fig. 3). Thus, only position 453 is indicative of lanosterol or cycloartenol production. An alignment of ERG7 homologs from our data set (fig. 3) shows that position 453 is consistent with the experimental characterizations available for a few additional phyla, such as in the amoebozoan *Dictyostelium* (I453, cycloartenol route; Nes et al. 1990), the heterolobosean *Naegleria* (I453, cycloartenol route; Raederstorff and Rohmer 1987), and the kinetoplastids (V453, lanosterol route; Roberts et al. 2003). No data are available on the other phyla, although based on this position it can be predicted that the choanoflagellate *Monosiga* makes lanosterol (V453), consistent with their affiliation with fungi and animals into the phylum Opisthokonta, whereas the brown algae *Aureococcus*, the diatom *Thalassiosira*, the red algae *Cyanidioschyzon*, and the green algae

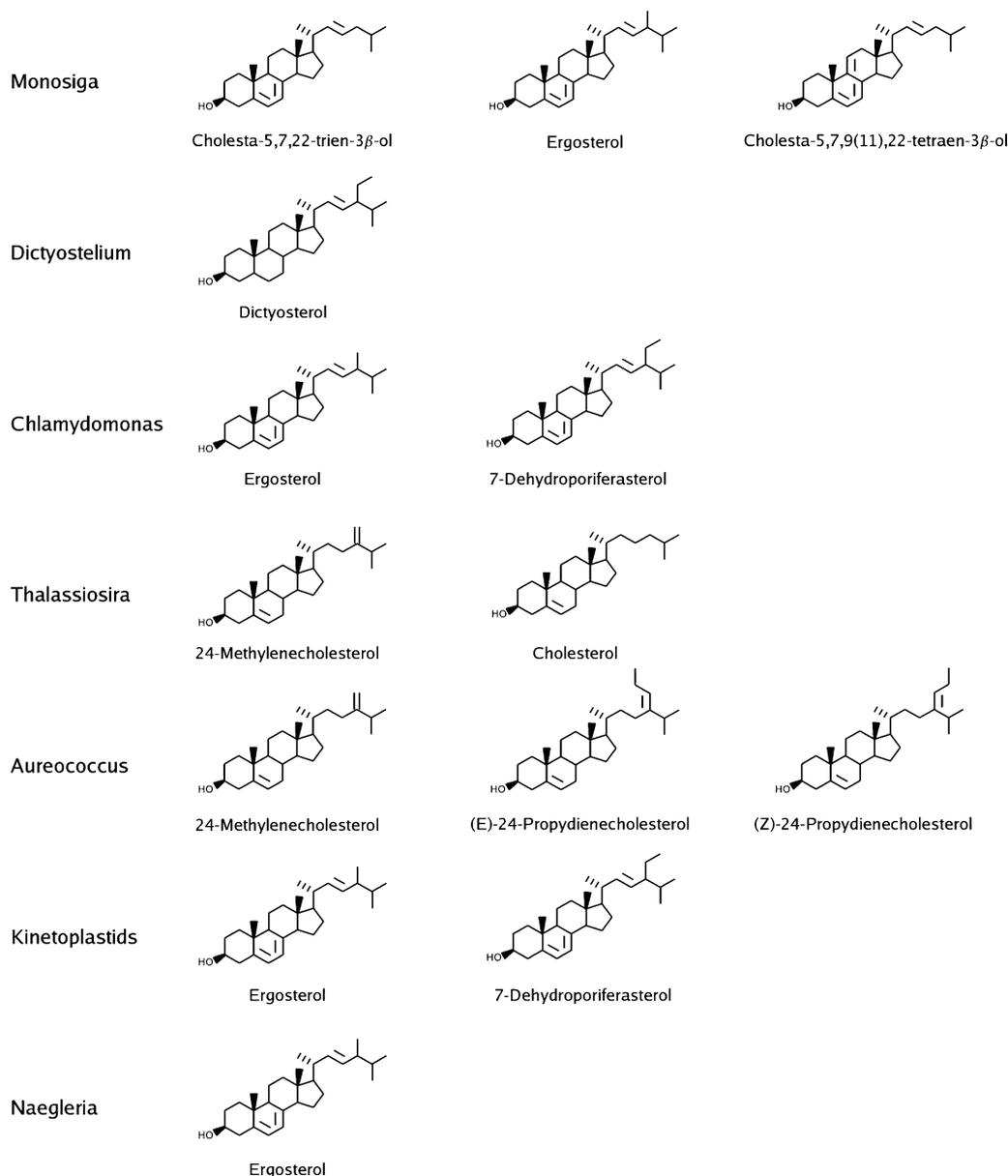


FIG. 2.—Sterols characterized from organisms other than fungi, vertebrates, and land plants present in our data set. For sterol modifications discussed in the text, refer to figure 1B for numbering.

*Chlamydomonas* and *Ostreococcus* all make cycloartenol (I453) (fig. 3).

Among the variety of the sterols shown in figure 2, a number of features are in common with the products of canonical pathways (fig. 1A and B): i) they are all derived from cyclization of squalene into lanosterol or cycloartenol; ii) final products have no methyl group in position 4 and position 14 compared with the first products of cyclization of squalene; iii) they are free of double bonds between carbon C-8 and C-9; and iv) they present a double bond between carbon C-5 and C-6, except for dictyosterol (fig. 2). These characteristics imply that the following steps must necessarily take place in the synthesis of all these sterols (fig. 1C): i) squalene cyclization (squalene monooxygenation + oxidosqualene cyclization); ii) loss of the methyl

group in position 14 (C-14 demethylation + C-14 reduction); ii) loss of both methyl groups at position 4 (C-4 methyl oxidation, C-3 dehydrogenation/C-4 decarboxylation, C-3 ketoreduction); iii) reduction of the delta-8 double bond (delta-8, delta-7 isomerization); and iv) formation of a double bond between C-5 and C-6 (C-5 desaturation). Enzymes catalyzing two of these steps are present in all these sterol-synthesizing organisms (table 1): oxidosqualene cyclization (ERG7) and C-14 demethylation + C-14 reduction (ERG11 + ERG24/FK). A particular case concerns C-3 dehydrogenation/C-4 decarboxylation (ERG26): In a phylogeny of ERG26 (Supplementary Material 1, Supplementary Material online) we identify a clear cluster of orthologs including the known enzymes of plants, fungi, and vertebrates. However, this cluster lacks *M. brevicollis*,

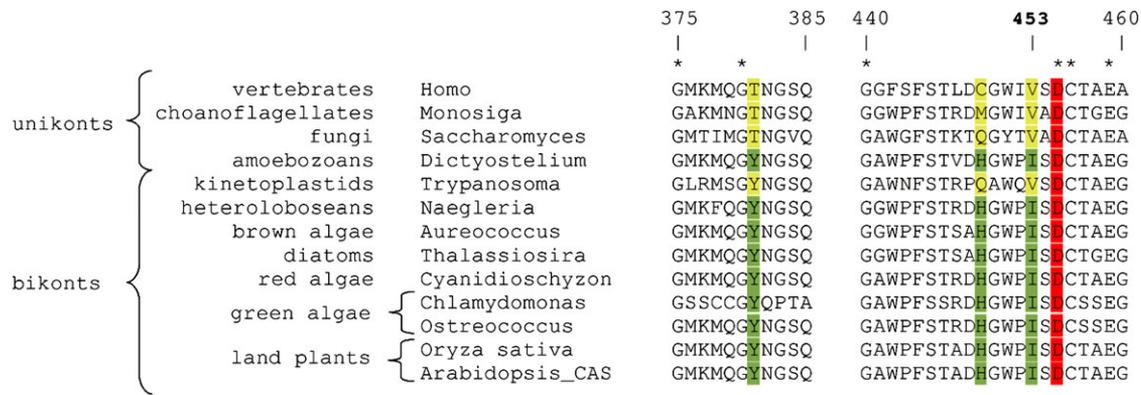


FIG. 3.—Alignment of ERG7 homologs from representative taxa. The catalytic residue D455 is highlighted in red. Positions 381, 449, and 453 are differentially conserved between lanosterol synthases (yellow) and cycloartenol synthases (green). The numbering refers to the *H. sapiens* ERG7 homolog (NP\_002331).

*O. tauri*, *T. pseudonana*, *A. anophagefferens*, and kinetoplastids. Interestingly, these genomes harbor homologs that form a separate group (Supplementary Material 1, Supplementary Material online). The perfect matching of this phylogenetic pattern and the requirement for C-3 dehydrogenation/C-4 decarboxylation indicates that these distant homologs are good candidates to perform an ERG26 function.

Intriguingly, several genomes lack clear homologs for the remaining steps that we infer to be necessary to produce the sterols that have been characterized in these organisms (table 1): lack of squalene monooxygenase (ERG1) in *M. brevicollis*, *C. reinhardtii*, *T. pseudonana*, and *A. anophagefferens*; lack of C-4 methyl oxidase (ERG25) in *M. brevicollis*, *C. reinhardtii*, *T. pseudonana*, *A. anophagefferens*, and kinetoplastids; lack of C-3 ketoreductase (ERG27) in all analyzed genomes apart from vertebrates and fungi; lack of ER-anchoring protein (ERG28) in *M. brevicollis*, *O. tauri*, *T. pseudonana*, *A. anophagefferens*, and kinetoplastids; lack of delta-8, delta-7 isomerase (ERG2 or EBP) in *T. pseudonana* and *A. anophagefferens*; and lack of C-5 desaturase (ERG3) in *T. pseudonana* and *A. anophagefferens*. Because these steps are essential, yet unknown enzymes must exist that fulfill these missing functions (to the possible exception of ERG28 that may be dispensable; Gachotte et al. 2001). In particular, it is possible that the enzyme performing C-3 ketoreduction in these organisms will turn out to be the same of that acting in land plants, which, as mentioned before, is still unknown. Importantly, we found orthologs of CPI1 in the organisms where we predicted a pathway following a cycloartenol route based on the active sites of their ERG7 (fig. 3 and table 1). Other steps are not universally present and are discussed in detail in Supplementary Material 2 (Supplementary Material online). In general, our analysis allowed us to infer the presence/absence of orthologs and the corresponding enzymatic steps in all these organisms (table 1). However, the precise nature of the corresponding pathways is difficult to predict because, as we have seen in the canonical pathways, the order of steps can be arranged in various ways. Nevertheless, we could point out a number of missing steps that are not compatible with the sterols known to be produced by a given organism, these steps being thus performed by yet-to-discover enzymes (red cases with ex-

clamation points in table 1). Finally, to our knowledge, two species of our data set (*O. tauri* and *C. merolae*) have not yet been studied for their sterol composition. Assuming that the enzymatic steps that take place in all other organisms analyzed also take place in these organisms, and given our phylogenomic data, we tentatively predicted the kind of sterols they may produce (Supplementary Material 2, Supplementary Material online).

To sum up, our phylogenomic data indicate a wide variety of pathways in different eukaryotic organisms. Interestingly, there is no clear separation between the pathway of photosynthetic organisms and nonphotosynthetic ones, differently to what commonly assumed. For example, the green algae *O. tauri* and *C. reinhardtii* do not appear to have the typical land plants pathway because C-14 reduction is performed by an ERG24 ortholog and not by a typical land plants FK ortholog (table 1). Interestingly, the early-emerging land plant lineage *Physcomitrella patens* harbors a set of enzymes that is identical to that of *O. sativa* and *A. thaliana* (data not shown), indicating that what is known as the typical land plant pathway can be generalized to the whole lineage. No orthologues of land plants SMO1 and SMO2 are present in the diatom *T. pseudonana*, the brown algae *A. anophagefferens*, and the red algae *C. reinhardtii*, suggesting that these organisms may harbor a pathway whose order of steps is somehow different from that of land plants. On the other hand, the nonphotosynthetic protists *N. gruberi* and *D. discoideum* synthesize sterol via cycloartenol and perform a C-14 reduction by an ortholog of typical land plants FK (table 1). Thus, phylogenomic analysis indicates that these different organisms harbor pathways that display a mixture of features of the three canonical pathways from fungi, vertebrates, and land plants. This leads to the question on what may have been the ancestral pathway in eukaryotes.

#### Did the LECA Already Make Sterols?

From the distribution of orthologs in table 1, we can tentatively infer the set of enzymes that were already present in LECA. Although the phylogenies of these enzymes are generally not fully resolved due to lack of signal, major

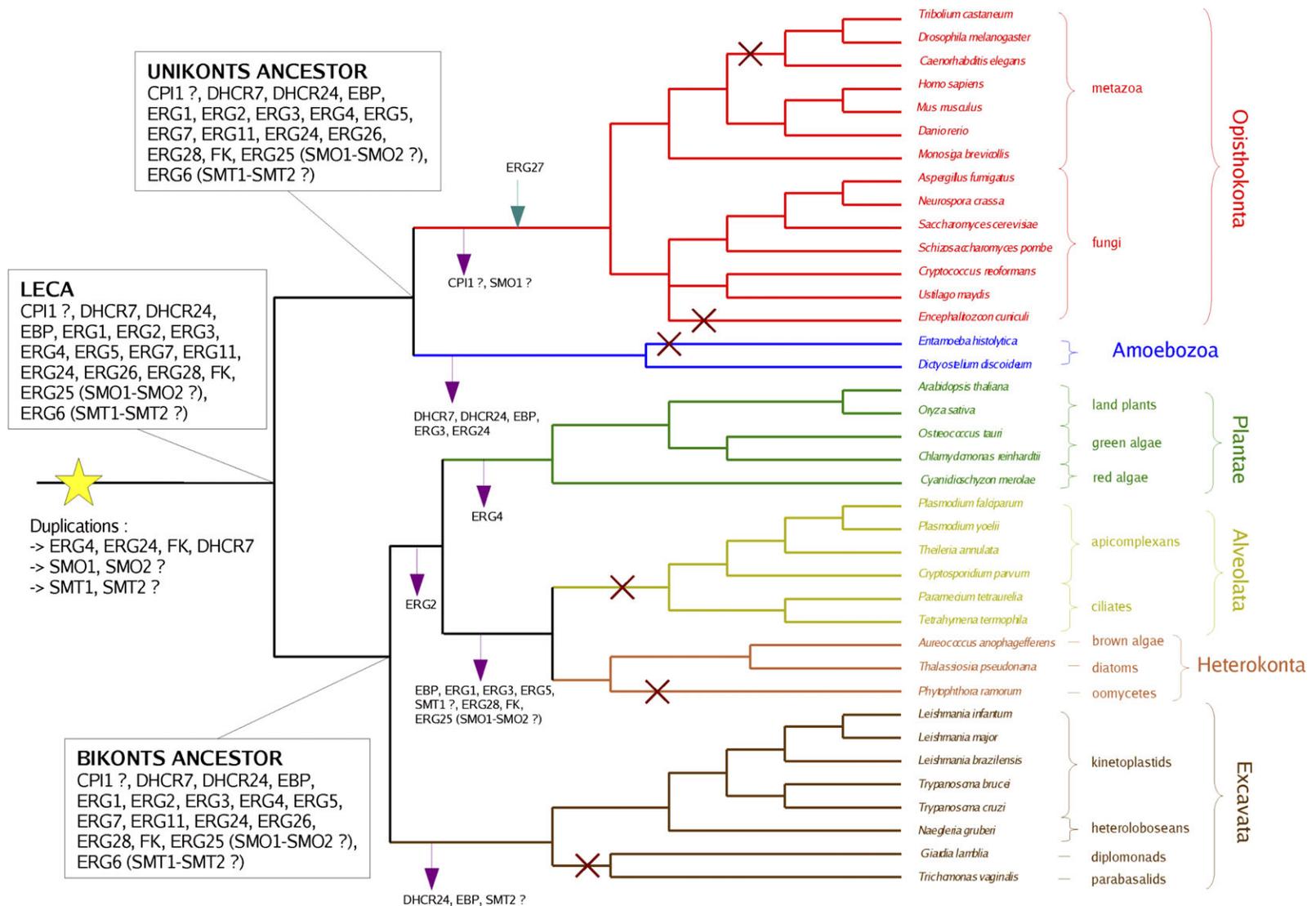


FIG. 4.—Inference of ancestral sets of sterol enzymes. These are inferred based on a consensus phylogeny of eukaryotes included in our data set according to the most recent data (Burki et al. 2008; Hampf et al. 2009) rooted in between unikonts and bikonts. Lineages where the ability to synthesize sterols has been lost are indicated by red crosses. We inferred maximal and minimal ancestral enzyme contents at the three most ancient nodes (LECA, Unikonts, and Bikonts). Additional proteins in the maximal content with respect to the minimal content are indicated by question marks (see text for details). Putative duplications inferred in the lineage leading to the LECA are indicated by a star. Losses and gains of proteins are indicated only for the most basal branches by purple and blue arrows, respectively.

eukaryotic groups are generally recovered. This allows inferring ancestral sets of proteins by mapping their presence on a consensus eukaryotic phylogeny (Burki et al. 2008; Hampl et al. 2009) rooted in between unikonts and bikonts (fig. 4). For example, an enzyme whose orthologs are present in fungi, amoebozoans, plants, and kinetoplastids and are well separated in the corresponding phylogeny (see, e.g., ERG5 in Supplementary Material 1, Supplementary Material online) lead us to infer the presence of this enzyme in the LECA and that its current absence in some phyla is due to subsequent gene loss. We inferred the ancestral sets of enzymes in the three most ancient nodes of the eukaryotic phylogeny (LECA, unikonts, and bikonts) (fig. 4) by minimizing HGT events among eukaryotic lineages. It has been put forward that HGT among eukaryotes may be more frequent than generally assumed (Andersson 2009). However, we could not observe clear cases of such HGT in our phylogenetic reconstructions, to one potential exception discussed below.

Phylogenetic analysis indicates that the duplications that gave rise to the gene families corresponding to ERG4-ERG24-FK-DHCR7 likely occurred before the divergence of major eukaryotic phyla (Supplementary Material 1, Supplementary Material online), indicating that these enzymes were already present in the LECA (fig. 4). The same may be said for the gene duplications giving rise to SMO1-SMO2, and to SMT1-SMT2, because the unikont *D. discoideum* has two gene copies of each family, similar to land plants and some other bikonts, which cluster within different orthologous groups (Supplementary Material 1, Supplementary Material online) (fig. 4). However, phylogenetic reconstruction does not allow excluding that *D. discoideum* obtained its additional SMO1 and SMT2 via HGT from a bikont lineage (Supplementary Material 1, Supplementary Material online). In this case, these two duplications would have occurred after the unikont/bikont bifurcation. Therefore, we indicated the occurrence of these duplications and the presence of the corresponding proteins in the LECA by question marks (fig. 4). The presence of a CYP11 ortholog in the unikont *D. discoideum* (table 1) may indicate that the LECA had this enzyme, although the poor resolution of the corresponding phylogenetic tree and the proximity with the ortholog from *P. pacifica* (Supplementary Material 1, Supplementary Material online) cannot exclude that it was acquired by HGT (therefore a question mark in fig. 4). Interestingly, this analysis suggests that the LECA may have already harbored at least 19 enzymes (fig. 4). Even if *D. discoideum* obtained its additional SMO1, SMT2, and CYP11 copies via HGT from a bikont lineage, a minimum set of 16 enzymes can still be inferred in the LECA (fig. 4). Among all enzymes, ERG27 only has specifically appeared in the Opisthokonta lineage. The opisthokont-specific ERG27 belongs to a gene family including bacterial short-chain dehydrogenases as well as other distant eukaryotic hypothetical proteins (not shown) and may have thus been recruited in the ancestor of Opisthokonta (fig. 4). However, because C-3 ketoreduction would have been essential in the pathway inferred in the LECA, this opisthokont-specific ERG27 may have replaced an ancestral enzyme performing the same function in the LECA. It is thus possible that the LECA had the non-

homologous equivalent of opisthokonts ERG27 that has yet to be discovered in land plants as well as in all other lineages (table 1). The place of the root of the eukaryotic tree in between unikonts and bikonts has been recently questioned based on the fact that a member of Amoebozoa (unikonts) harbors bikont characters (Minge et al. 2009; Roger and Simpson 2009). If we consider that the root of the eukaryotic tree is presently unknown and we replace the consensus phylogeny by a polytomy, we are forced to infer a maximal set of enzymes back to the LECA in order to minimize convergent protein gains. This leads to the same ancestral sets inferred with the unikont/bikont rooting.

Phylogenomic inference suggests that the LECA had a large potential of enzymatic functions and may have been capable of synthesizing a diverse panel of sterols. In fact, among the set of enzymes inferred to have been already present in the LECA, a few are nonorthologous, but perform redundant functions, such as C-14 reductases (ERG24/FK), delta-8 delta-7 isomerases (ERG2/EBP), and delta-24 reductases (ERG4/DHCR24) (table 1 and fig. 4). Our analysis does not allow inferring how the pathway was progressively assembled in the lineage leading to LECA. However, if LECA possessed a CYP11, this implies that it used a cycloartenol route rather than a lanosterol route because no known organisms that possess a CYP11 synthesize lanosterol. Consequently, the switch from a cycloartenol to a lanosterol route would have occurred at least twice independently in eukaryotic evolution (i.e., in opisthokonts and in kinetoplastids). An ancestral cycloartenol route is indeed consistent with previous hypotheses (Bloch 1991; Ourisson and Nakatani 1994).

Our analysis indicates that organisms that do not synthesize sterols have lost the pathway (fig. 4). Interestingly, the loss of sterol synthesis in insects and nematodes appears to be specific to these lineages because the placozoan *Trichoplax adhaerens* has a full set of typical animal enzymes, except for DHCR7 (data not shown). Sometimes, a few enzymes have been retained, possibly for the processing of sterols harvested from the environment, such as, for example, is the case of DHCR7, where a copy is present in the non-sterol-producing oomycete *P. ramorum* (Supplementary Material 1, Supplementary Material online). Our analysis also indicates that the present-day distribution of sterol enzymes in the different eukaryotic organisms analyzed is the consequence of differential losses that probably accompanied specialization of the pathways along with eukaryotic diversification. However, appearance of novel enzymes that are not yet known may have also taken place. This prompts to explore further the sterol-synthesizing abilities of a wider sampling of eukaryotic diversity, all the more so that the presence of a particular pathway in a given organism may not be generalized to the whole phylum it belongs to.

#### A Bacterial Origin?

Consistent with previous findings, we found orthologs of the first two genes of the pathway (ERG1 and ERG7) in *G. obscuriglobus* (Pearson et al. 2003), *M. capsulatus* (Lamb et al. 2007), and *S. aurantiaca* (Bode et al.

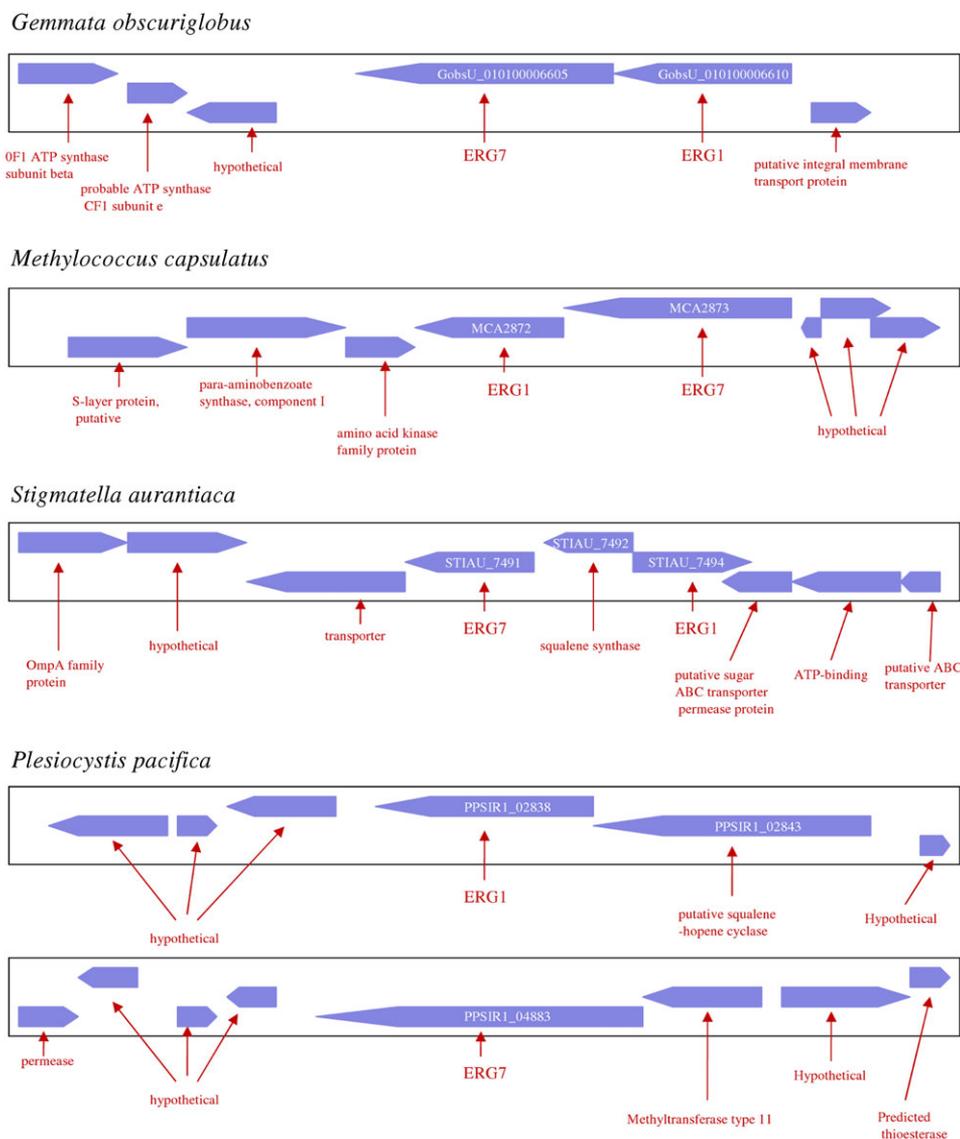


FIG. 5.—Genomic context of ERG1 and ERG7 genes in the genomes of the four sterol-synthesizing bacteria *G. obscuriglobus*, *M. capsulatus*, *S. aurantiaca*, and *P. pacifica*.

2003), and we report the presence of these two genes in *P. pacifica*. In *G. obscuriglobus* and *M. capsulatus*, the genomic context of ERG1 and ERG7 is known to be conserved and this has been taken as further indication for a common HGT involving these two genes (Pearson et al. 2003). However, the genome contexts in *S. aurantiaca* and *P. pacifica* are different (fig. 5). In *P. pacifica*, ERG1 is not close on the genome to ERG7, but to the bacterial ERG7 homolog SHC. This is puzzling because SHC is not involved in the synthesis of sterols and ERG1 is not involved in the synthesis of hopanoids. Interestingly, in *S. aurantiaca*, a gene coding for the enzyme responsible for the last step of the pathway leading from IPP to squalene (squalene synthase) is found in between ERG1 and ERG7 (fig. 5), leading to a conserved context for three enzymes acting consecutively.

The presence of the first two enzymes of the pathway in the few sterol-producing bacteria has been previously discussed as resulting from an ancient HGT (Pearson

et al. 2003). In general, previously published trees have been presented with only representatives of eukaryotes and these bacteria (Pearson et al. 2003). However, ERG1 homologs are not only present in the four sterol-producing bacteria but also have homologs in other bacteria within a large family of monooxygenases. In a tree including the whole family (fig. 6), the ERG1 homologues of the four sterol-producing bacteria branch basally, but do not appear to be more closely related to eukaryotic ERG1 than they are to the other bacterial monooxygenases, and they share no specific sequence signature with eukaryotic ERG1 (fig. 7). Therefore, the hypothesis of an ancient HGT involving the ERG1 homologues of the four sterol-producing bacteria and eukaryotic ERG1 remains open. Conversely, ERG7 and their bacterial SHC counterparts clearly separate into two clusters in the corresponding phylogeny (fig. 8). The four bacterial ERG7 appear to be more closely related to their eukaryotic homologs (fig. 8), indicating a specific

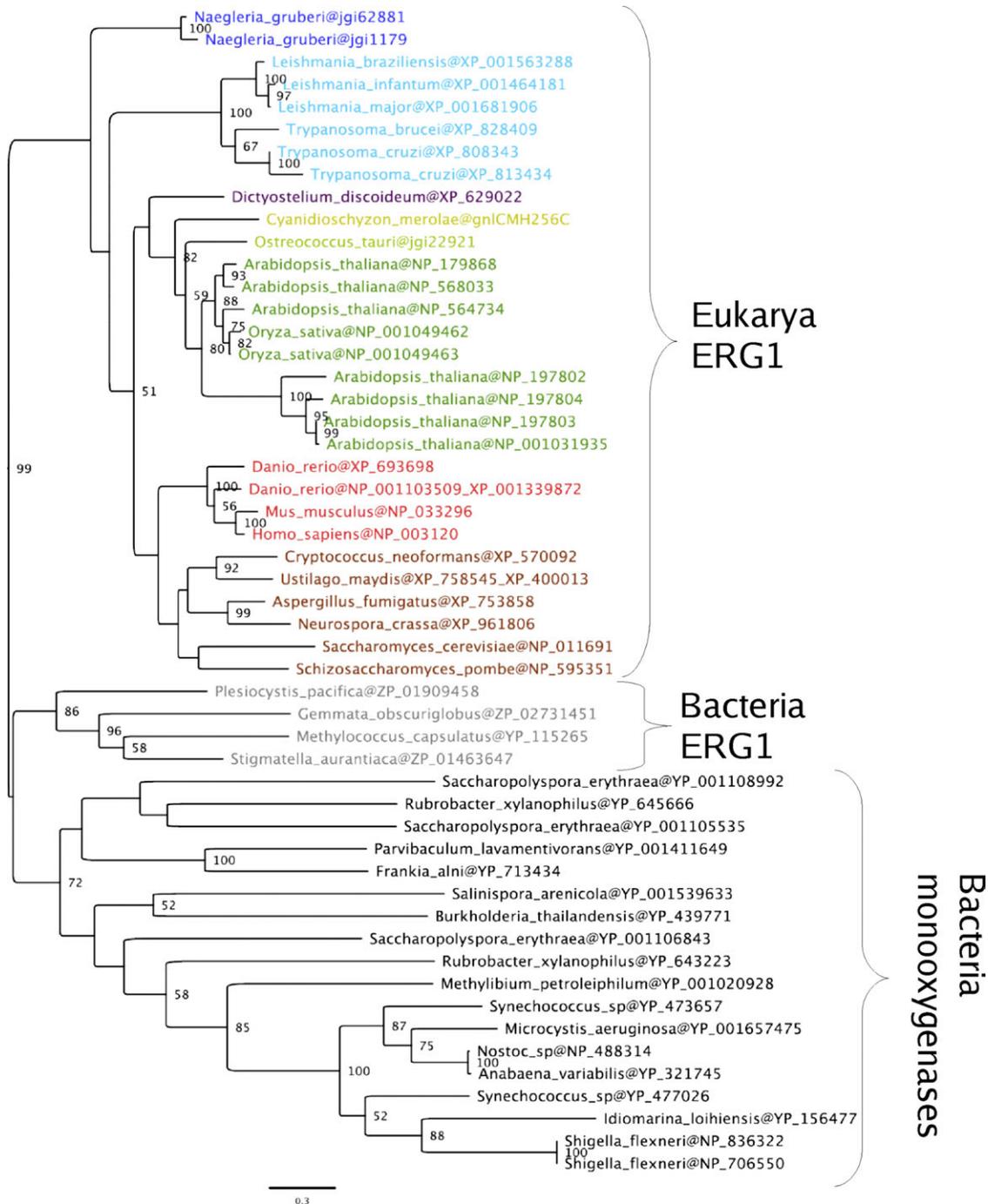


FIG. 6.—Maximum likelihood tree of ERG1 homologs. Only a subset of the most closely related bacterial monoxygenases are included. Following removal of ambiguously aligned regions, the final data set included 130 conserved positions. The tree was obtained by PHYML (Guindon and Gascuel 2003) as detailed in Materials and Methods. For clarity, only bootstrap values >50% are shown.

evolutionary relationship, as previously put forward (Pearson et al. 2003; Chen et al. 2007). Three bacterial ERG7 branch basally with respect to their eukaryotic counterparts, and this may be interpreted in favor of a hypothesis where eukaryotic ERG7 originated from bacteria. However, the ERG7 of *S. aurantiaca* emerges from within eukaryotes (fig. 8), suggesting that it was obtained via HGT. Moreover, along with their ERG7 copies, *M. capsulatus*,

*G. obscuriglobus*, and *P. pacifica* all have also a typical bacterial SHC ortholog, whereas this is not the case for *S. aurantiaca* (fig. 8). Thus, the possibility remains that the three bacterial ERG7 were obtained from eukaryotes, either via an ancient HGT before eukaryotic diversification or that fast evolution following transfer leads to an artificial grouping of the three bacterial ERG7 at the base of eukaryotes. The SHC sequences of *G. obscuriglobus* and

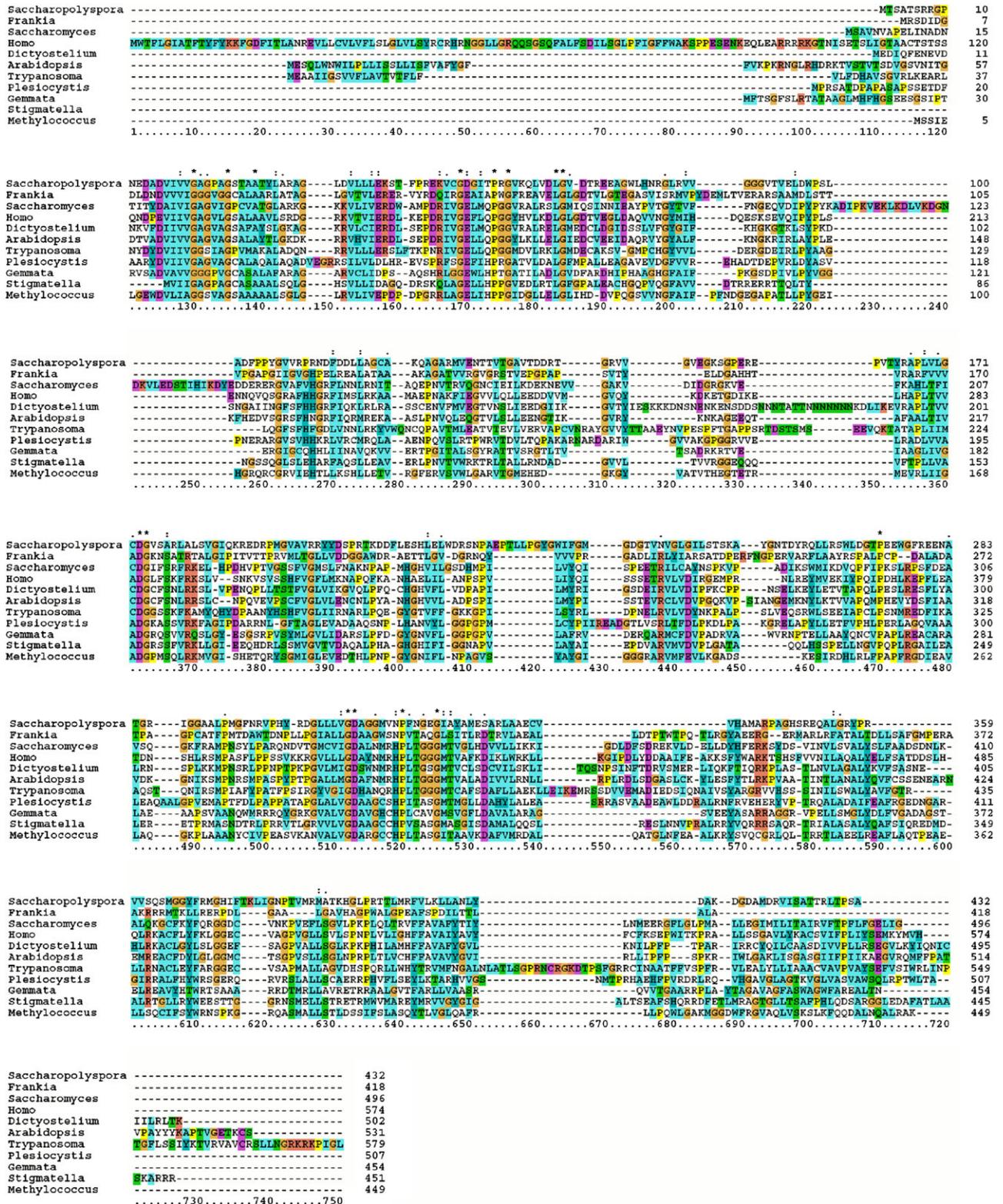


FIG. 7.—Alignment of ERG1 homologs including sequences of the four sterol-synthesizing bacteria (*G. obscuriglobus*, *M. capsulatus*, *S. aurantiaca*, and *P. pacifica*), of two other bacteria (*Saccharopolyspora erythraea* and *Frankia alni*), and of five representatives of eukaryotes (*T. brucei*, *S. cerevisiae*, *H. sapiens*, *D. discoideum*, and *A. thaliana*).

*P. pacifica* are very divergent, in particular that of *Gemmata*, which is split into three genes. It is possible that the presence of an ERG7 and therefore the synthesis of sterols have

relaxed the functional constraints of native SHC sequences. It would nevertheless be interesting to verify the activities of these SHC orthologs.

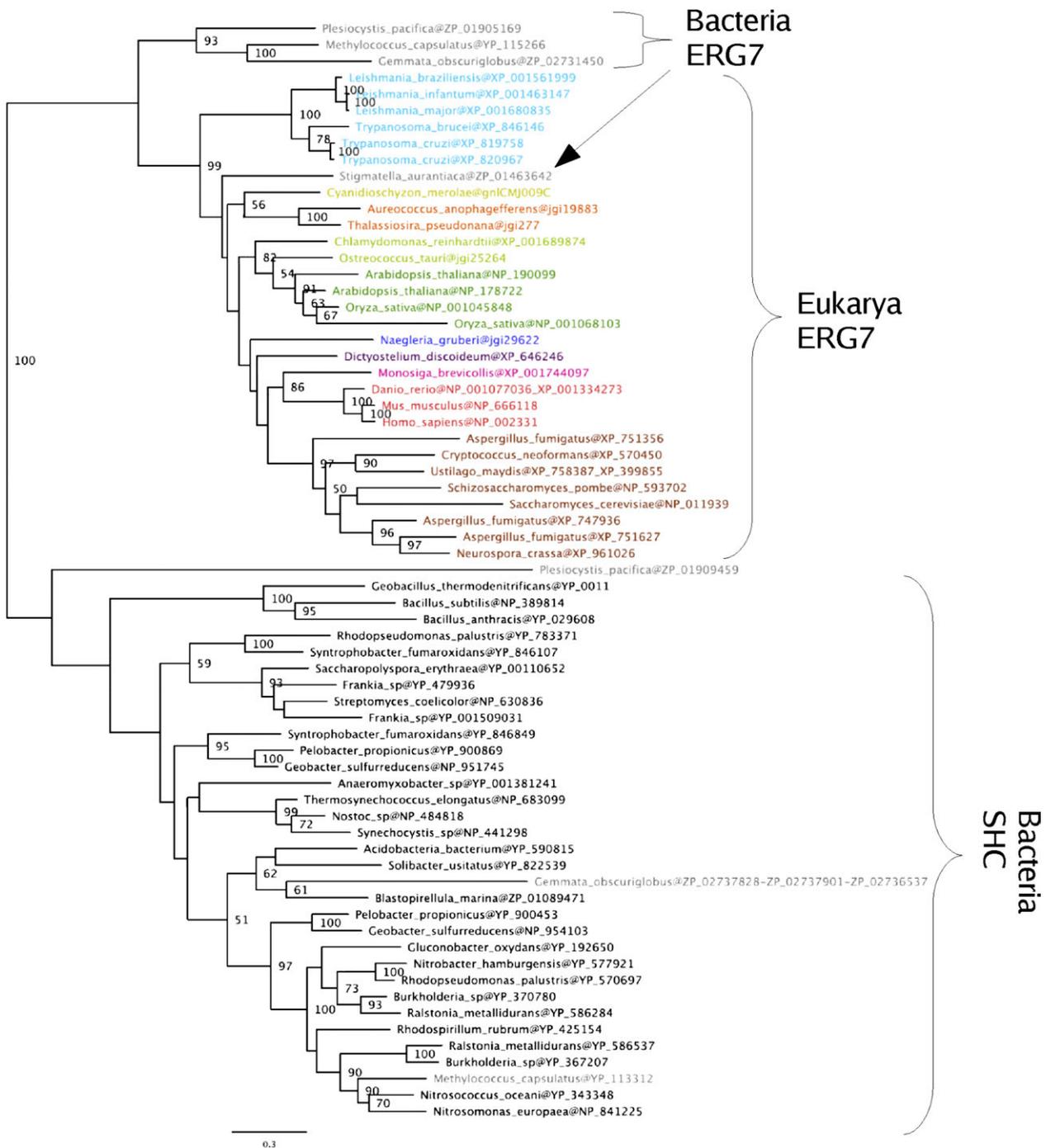


FIG. 8.—Maximum likelihood tree of ERG7 homologs. Following removal of ambiguously aligned regions, the final data set included 386 conserved positions. The tree was obtained by PHYML with the same criteria as detailed in Materials and Methods.

*Gemmata obscuriglobus* and *S. aurantiaca* have no other homologs of the pathway, whereas *M. capsulatus* has a homolog of ERG11, consistent with the fact that this bacterium is able to produce more complex sterols (Lamb et al. 2007). Interestingly, we found that *M. capsulatus* also has a homolog of DHCR24 that branches robustly with Metazoans and choanoflagellates (Supplementary Material 1, Supplementary Material online) and thus has likely originated via HGT from these eukaryotes. The presence of

DHCR24 responsible for delta-24 reduction (fig. 1) is consistent with the 4 $\alpha$ -methyl-5 $\alpha$ -cholest-8(14)-en-3 $\beta$ -ol and 4,4-dimethyl-5 $\alpha$ -cholest-8(14)-en-3 $\beta$ -ol produced by *M. capsulatus* (Bouvier et al. 1976). Similar to *M. capsulatus*, *P. pacifica* also has a homolog of ERG11 (Supplementary Material 1, Supplementary Material online). ERG11 homologs are also present in Mycobacteria (Supplementary Material 1, Supplementary Material online). It has been proposed that mycobacterial and *M. capsulatus*

ERG11 originated via HGT from plants (Režen et al. 2004). However, the ERG11 of *M. capsulatus* and mycobacteria group with the ERG11 homolog of *P. pacifica*. Thus, the eventual HGT would have occurred in one of these bacteria followed by HGT among them. Moreover, phylogenetic analysis does not indicate a closer proximity of plants ERG11 with these bacterial homologs (Supplementary Material 1, Supplementary Material online). By congruence with ERG7, it is possible that ERG11 was also acquired via an ancient HGT from an ancestor of Eukaryotes or a specific eukaryotic source that remains undetermined due to the high divergence of bacterial ERG11 sequences.

### The Case of *P. pacifica*

In addition to ERG1, ERG7, and ERG11, *P. pacifica* has six homologs of enzymes of the pathway catalyzing C-14 reduction (two copies of FK), C-4 demethylation (ERG25 and ERG26), delta-8-delta-7 isomerization (ERG2), and cyclopropylsterol isomerization (CpI1), consistent with the identification of very elaborated sterols in its close relative *Nannocystis* (Bode et al. 2003). The sterols produced by *P. pacifica* have not yet been characterized. However, the combination of these enzymes leads us to infer production of 7,24-cholestadien-3beta-ol in *P. pacifica*, similar to some characterized sterols from *Nannocystis* (Bode et al. 2003). The presence of a CPI1 homolog in *P. pacifica* is interesting because this is the first reported case of a bacterial homolog of this enzyme, which would indicate a cycloartenol route. It is thus possible that *P. pacifica* produces cycloartenol, although no cycloartenol has been identified in its close relative *Nannocystis* (Bode et al. 2003). Interestingly, we found that the ERG7 homolog of *P. pacifica* has a V453, which is characteristic of a lanosterol route (fig. 3). It will be interesting to characterize the *P. pacifica* CPI1 homolog in order to assess its precise function, as well as the types of sterols produced by *P. pacifica*. The presence of CPI1 uniquely in *P. pacifica* among bacteria strongly points toward a recent acquisition via HGT from eukaryotes, although its sequence does not appear particularly close to any precise eukaryotic source (Supplementary Material 1, Supplementary Material online). *Plesiocystis pacifica* has two clear orthologs of FK and one ortholog of ERG2, which cluster within eukaryotic sequences and thus likely originated via HGT from eukaryotes (Supplementary Material 1, Supplementary Material online). Concerning the step of C-4 demethylation, *P. pacifica* has homologs of ERG25 and ERG26. *P. pacifica* ERG25 emerges from within the large protein family that includes eukaryotic ERG3/ERG25 (not shown) and therefore also very likely originated via HGT from eukaryotes. As for ERG26, phylogenetic analysis does not show a clear orthology relationship between *P. pacifica* and its eukaryotic counterparts, although this may be due to its high divergence (Supplementary Material 1, Supplementary Material online). The ERG25 and ERG26 homologs from *P. pacifica* likely perform ERG25 and ERG26 functions, for two reasons: 1) the possibility that *P. pacifica* produces sterols as elaborated as *Nannocystis*, which would require the action of an ERG25 and an ERG26, and 2) the presence of an ERG2 which in principle requires the pre-

vious action of ERG25 and ERG26. Interestingly, these arguments also imply the presence of an ERG27. However, because *P. pacifica* has no homologs of fungi and vertebrates ERG27, it is possible that it harbors a homolog of the still-missing enzyme providing an ERG27 function in land plants and other eukaryotic lineages (table 1). If this is so, this putative ERG27 equivalent may display a closer relationship to eukaryotes (and be preferentially absent in Opisthokonta and non-sterol-producing lineages) than other bacteria. We extracted from our databank the genomes of eukaryotes lacking ERG27 plus 52 bacteria from all phyla, and we searched the whole genome of *P. pacifica* against it. Then, we filtered for those genes that had all these eukaryotes as first Blast hits within an *e* value  $< 1 \times 10^{-10}$ . We found eight genes harboring such a phyletic pattern. Four of these correspond to ERG11, ERG7, and the two copies of FK, what makes a good positive control. The remaining four genes are a member of the P450 family (distant relative of ERG 11), a hypothetical serine/threonine protein phosphatase (which is, however, also present in eukaryotes that do not make sterol), an aspartatecarbamoyl transferase catalytic subunit (but phylogenetic analysis indicates that it is not closer to eukaryotes than to other bacteria), and a succinate-semialdehyde dehydrogenase (NAD(P)<sup>+</sup>) (which has only distant bacterial homologues). The increase of the *e* value cut-off gave no other good candidates. The last protein (succinate-semialdehyde dehydrogenase [NAD(P)<sup>+</sup>]) is particularly interesting because its annotated activity is congruent with an ERG27 function (C-3 ketoreduction), and further analysis on all eukaryotic genomic data reveals that it is absent in eukaryotes that do not make sterols. Moreover, this protein is also specifically absent in eukaryotes that harbor an ERG27 copy, to the exception of fungi (fig. 9). If this were the real ancestral ERG27, it would mean that fungi have retained it and use it for a different function. In fact, *S. cerevisiae* mutants for ERG27 are sterol auxotrophs (Goffeau et al. 1996), indicating that this potential ERG27 analog (NP\_011904) cannot rescue the fungi ERG27 function. Interestingly, this protein appears to function in the endoplasmic reticulum in *S. cerevisiae*, precisely like ERG27, and mutants are defective in directing meiotic recombination events to homologous chromatids (Goffeau et al. 1996). It would be extremely interesting to verify the involvement of this enzyme in the sterol pathway of nonfungal lineages lacking a bona fide ERG27.

### Discussion and Conclusions

The emergence of specific features in the lineage leading to eukaryotes is one of the most intriguing issues in the domain of early evolution. Phylogenomics is a recent and powerful approach to dissect the emergence and subsequent evolution of cellular processes. However, its application to eukaryotes is just becoming to emerge thanks to the recent availability of a sufficient sampling of complete genomes. Recent analyses have investigated the emergence and evolution of different unique eukaryotic cellular structures (Baptiste et al. 2005; Eme et al. 2009; Field and Dacks 2009). However, the emergence and evolution of

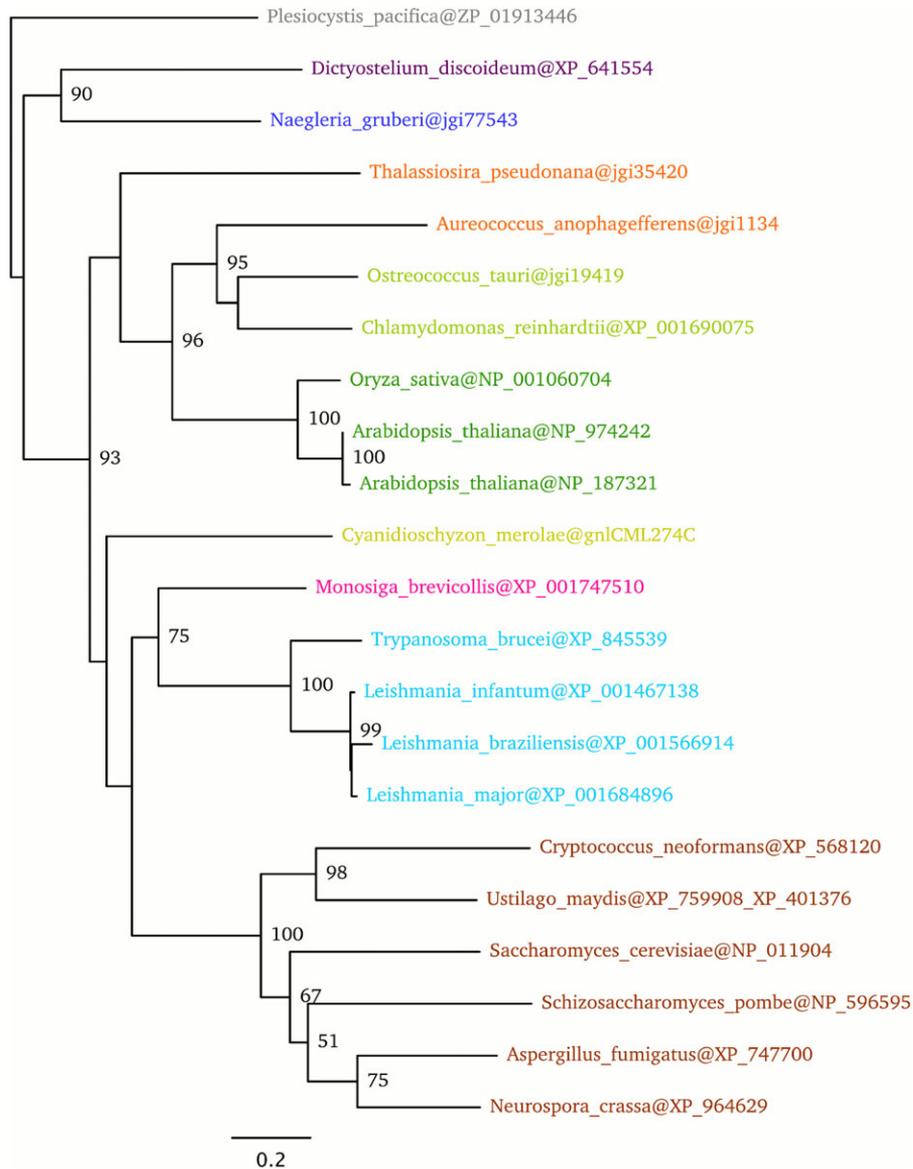


FIG. 9.—Maximum likelihood tree of putative ERG27 analogues. Following removal of ambiguously aligned regions, the final data set included 483 conserved positions. The tree was obtained by PHYML with the same criteria as in Materials and Methods.

eukaryotic-specific metabolic pathways is still poorly explored. Here we have reported a phylogenomics analysis of one such eukaryotic-specific metabolic pathway. The presence of sterols is an essential characteristic of all eukaryotic membranes and thus presumably accompanied the very emergence of this domain of life. In particular, the appearance of sterols may have allowed an increase fluidity of eukaryotic membranes and thus possibly represented an important step toward increasing cell size. Additionally, the emergence of sterols may have provided protection against oxidative stress (Galea and Brown 2009).

We have shown that eight enzymes of the sterol pathway belong to large gene families that include distant bacterial homologs (if the few sterol-making bacteria are not considered). These are ERG1 (monooxygenases), ERG7 (SHCs), ERG5-ERG11 (cytochrome P450), ERG3-ERG25

(oxidases), ERG26 (NADP-dependent dehydrogenases), and DHCR24 (bacterial FAD/FMN-containing dehydrogenases). These have been likely recruited from preexisting enzymes in parallel to the emergence of the sterol pathway in the lineage leading to the LECA. By contrast, four enzymes (ERG28 and EBP, ERG24, and ERG6) do not have any bacterial homologue and have thus presumably arisen in the eukaryotic lineage. The same is likely true for five additional enzymes: Three (ERG4, DHCR7, and ERG2) have very few bacterial homologs other than *P. pacifica* and these clearly derive from HGT from eukaryotes (ERG4, *Coxiella*) and DHCR7 (*Coxiella* and *Protochlamydia*), ERG2 (*Mycobacteria*), and two (CPI1 and FK) have only *P. pacifica* as bacterial homolog (Supplementary Material 1, Supplementary Material online).

We have highlighted that *P. pacifica* harbors the largest reported set of homologs of eukaryotic sterol-synthesizing

enzymes. It may be put forward that this myxobacterium is at the origin of these eukaryotic enzymes, which may lend support to the hypothesis that Myxobacteria might have been the potential partners of a syntrophic association at the origin of eukaryotes (Lopez-Garcia and Moreira 1999). Indeed, *P. pacifica* sequences appear to branch basally to eukaryotes, which may be consistent with this bacterium being the source of eukaryotic homologs. However, for all enzymes previously mentioned that belong to large gene families, *P. pacifica* sequences do not branch with their typical bacterial homologs, indicating that they may not have arisen from gene duplication of native genes followed by functional switch. Nevertheless, this may also be due to an acceleration of evolutionary rates that prevent assessing the real evolutionary relationships of these *P. pacifica* enzymes. However, we have shown that it is possible that *P. pacifica* has acquired its pathway for sterol synthesis via HGT from eukaryotes (either ancient or not). By the same reasoning, as proposed previously (Pearson et al. 2003) it is possible that ERG7 has been transferred to one of the few sterol-producing bacteria and then among them. However, we put forward the possibility that ERG1 in these bacteria does not derive from HGT from eukaryotes. Consequently, it is possible that the ERG1 function was recruited into the sterol pathway of these bacteria from a native enzyme only after HGT from eukaryotes bringing along the ERG7 copy. If this were so, ERG1 homologs would have been recruited independently in these bacteria and eukaryotes to function with ERG7. Interestingly, phylogenetic analysis of the eukaryotic enzymes involved in the pathway leading from IPP to squalene do not show any particular evolutionary relationships with the sterol-producing bacteria (data not shown), supporting the hypothesis that the capacity of making sterols in these bacteria arose via HGT from eukaryotes and that the eukaryotic sterol pathway originated in the LECA from the branching off on a preexisting native pathway leading to squalene. In this regard, the evolutionary proximity of *P. pacifica* and eukaryotic sterol enzymes should not be taken as support for an origin of eukaryotes from a symbiosis between myxobacteria and Archaea because in this case the previous steps should also show a similar pattern. Similarly, the proposal that actinobacteria are at the origin of the eukaryotic pathway (Cavalier-Smith 2006) is clearly weakened by the fact that the few homologs present in mycobacteria very likely derive from HGT from eukaryotes. More importantly, no actinobacteria have so far been shown to synthesize sterols (Režen et al. 2004) and these enzymes in mycobacteria are likely used to process sterols from their eukaryotic hosts.

Our analysis also indicates that the LECA had the potential to make a wide array of different sterols and that subsequent evolution occurred through tinkering via differential enzyme losses and specializations in the various eukaryotic lineages in parallel with their divergence from the LECA. This is consistent with the idea that the ability of synthesizing elaborated sterols would have paved the way toward specific eukaryotic characters such as cell signaling and multicellularity (Bloch 1991). Our analysis does not allow us to predict the order of steps leading to the assembly of the whole pathway in the lineage leading to the LECA. However, because ERG7 acts on squalene epoxide, which is in turn obtained by the oxygenation of squalene by

ERG1, we think that the emergence of ERG7 in an ancestor of eukaryotes was likely a key event in the assembly of the eukaryotic sterol pathway. Moreover, we tentatively infer that the LECA harbored a cycloartenol route rather than a lanosterol route, consistent with previous hypotheses (Bloch 1991; Ourisson and Nakatani 1994). Importantly, because the action of ERG7 and the subsequent steps in the sterol pathway are highly linked to oxygen and because we inferred the presence of a complex sterol pathway in the LECA, this suggests that this ancestor lived in a fully oxygenated environment.

Our study allowed reconstructing the potential capacities for making specific sterols in the eukaryotic organisms where the pathways have not yet been characterized. The emerging enzymatic landscape highlighted in this study indicates that the differences in the characterized pathways from fungi, land plants, and vertebrates, as well as those inferred in the noncharacterized organisms, arose by different organization of a common set of conserved enzymes, loss of specific enzymes, and the addition of a few novel enzymes. In particular, the different order of steps does not seem to be linked to a cycloartenol or lanosterol route because, for example, kinetoplastids have a set of enzymes that is very similar to that of fungi, whereas they use a lanosterol route with an order of steps very similar to that of land plants (e.g., C-4 demethylation occurring through two nonconsecutive steps; Lepesheva et al. 2004). Thus, the reorganization of a similar order of steps observed in fungi and animals may be specific to these two lineages and possibly linked to the recruitment of their unique ERG27, which replaced an ancestral analogous enzyme. Interestingly, this would have occurred twice independently in fungi and animals because choanoflagellates appear to have retained the ancestral C-3 ketoreductase and possibly an order of steps similar to that of land plants.

Importantly, we showed that phylogenomics approaches are a powerful tool to identify still-missing enzymes, such is the case of the elusive ERG27 equivalent in land plants and probably in other eukaryotic phyla. In this respect, more experimental data on a wider sampling of eukaryotic diversity are surely necessary to test our predictions.

Finally, because sterols are key markers for indicating the presence of specific eukaryotic lineages in the fossil record, our study may be a useful reference for palaeo geochemistry studies.

### Supplementary Material

Supplementary Material is available at *Genome Biology and Evolution* online ([http://www.oxfordjournals.org/our\\_journals/gbe/](http://www.oxfordjournals.org/our_journals/gbe/)).

### Acknowledgments

The authors declare that they have no competing interest. S.G. conceived the study. E.D. performed all analyses. E.D. and S.G. wrote the manuscript, which was read, edited, and approved by both authors. The authors would like to thank Roger Summons, Robin Kodner, and two anonymous referees for useful comments and suggestions.

## Literature Cited

- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res.* 25:3389–3402.
- Anderson RGW, Jacobson K. 2002. Cell biology—a role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains. *Science.* 296:1821–1825.
- Andersson JO. 2009. Horizontal gene transfer between microbial Eukaryotes. In: Gogarten MB, Gogarten JP, Olendzenski L, editors. *Horizontal gene transfer: genomes in flux.* Totowa (NJ): Humana Press Inc. p. 473–487.
- Baptiste E, Charlebois RL, Macleod D, Brochier C. 2005. The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biol.* 6(10):R85.
- Bloch K. 1991. Cholesterol: evolution of structure and function. In: Vance DE, Vance J, editors. *Biochemistry of lipids, lipoproteins and membranes.* Amsterdam: Elsevier. p. 363–381.
- Bode HB, et al. 2003. Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidisqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol Microbiol.* 47:471–481.
- Boucher Y, Kamekura M, Doolittle WF. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol.* 52:515–527.
- Bouvier F, Rahier A, Camara B. 2005. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog Lipid Res.* 44:357–429.
- Bouvier P, Rohmer M, Benveniste P, Ourisson G. 1976. Delta8(14)-steroids in bacterium *Methylococcus capsulatus*. *Biochem J.* 159:267–271.
- Brown DA, London E. 2000. Structure and function of sphingolipid- and cholesterol-rich membrane rafts. *J Biol Chem.* 275:17221–17224.
- Buckner FS, Nguyen LN, Joubert BM, Matsuda SPT. 2000. Cloning and heterologous expression of the *Trypanosoma brucei* lanosterol synthase gene. *Mol Biochem Parasitol.* 110:399–403.
- Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biol Lett.* 4:366–369.
- Cavalier-Smith T. 2006. Cell evolution and Earth history: stasis and revolution. *Philos Trans R Soc B Biol Sci.* 361:969–1006.
- Chen L-L, Wang G-Z, Zhang H-Y. 2007. Sterol biosynthesis and prokaryotes-to-eukaryotes evolution. *Biochem Biophys Res Commun.* 363:885–888.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Eme L, Moreira D, Talla E, Brochier-Armanet C. 2009. A complex cell division machinery was present in the last common ancestor of eukaryotes. *PLoS One.* 4:Article No. e5021
- Field MC, Dacks JB. 2009. First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr Opin Cell Biol.* 21:4–13.
- Gachotte D, et al. 2001. A novel gene conserved from yeast to humans is involved in sterol biosynthesis. *J Lipid Res.* 42:150–154.
- Galea AM, Brown AJ. 2009. The special relationship between sterols and oxygen: were sterols an adaptation to aerobic life? *Free Radic Biol Med.* 47:880–889.
- Giner JL, Boyer GL. 1998. Sterols of the brown tide alga *Aureococcus anophagefferens*. *Phytochemistry.* 48:475–477.
- Goffeau A, et al. 1996. Life with 6000 genes. *Science.* 274:546.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Hapl V, et al. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups.”. *Proc Natl Acad Sci USA.* 106:3859–3864.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.
- Jacobson K, Dietrich C. 1999. Looking at lipid rafts? *Trends Cell Biol.* 9:87–91.
- Kodner RB, Pearson A, Summons RE, Knoll AH. 2008. Sterols in red and green algae: quantification, phylogeny, and relevance for the interpretation of geologic steranes. *Geobiology.* 6:411–420.
- Kodner RB, Summons RE, Pearson A, King N, Knoll AH. 2008. Sterols in a unicellular relative of the metazoans. *Proc Natl Acad Sci USA.* 105:9897–9902.
- Lamb DC, et al. 2007. Lanosterol biosynthesis in the prokaryote *Methylococcus capsulatus*: insight into the evolution of sterol biosynthesis. *Mol Biol Evol.* 24:1714–1721.
- Lepesheva GI, Nes WD, Zhou WX, Hill GC, Waterman MR. 2004. CYP51 from *Trypanosoma brucei* is obtusifoliol-specific. *Biochemistry.* 43:10789–10799.
- London E. 2002. Insights into lipid raft structure and formation from experiments in model membranes. *Curr Opin Struct Biol.* 12:480–486.
- Lopez-Garcia P, Moreira D. 1999. On hydrogen transfer and a chimeric origin of eukaryotes. *Trends Biochem Sci.* 24:424.
- Melkonian KA, Ostermeyer AG, Chen JZ, Roth MG, Brown DA. 1999. Role of lipid modifications in targeting proteins to detergent-resistant membrane rafts—many raft proteins are acylated, while few are prenylated. *J Biol Chem.* 274:3910–3917.
- Minge MA, Silberman JD, Orr RJS, et al. 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc R Soc B Biol Sci.* 276:597–604.
- Nes WD, Norton RA, Crumley FG, Madigan SJ, Katz ER. 1990. Sterol phylogeny and algal evolution. *Proc Natl Acad Sci USA.* 87:7565–7569.
- Ourisson G, Nakatani Y. 1994. The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chem Biol.* 1:11–23.
- Pearson A, Budin M, Brocks JJ. 2003. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci USA.* 100:15352–15357.
- Philippe H. 1993. Must, a computer package of management utilities for sequences and trees. *Nucleic Acids Res.* 21:5264–5272.
- Raederstorff D, Rohmer M. 1987. Sterol biosynthesis via cycloartenol and other biochemical features related to photosynthetic phyla in the amebas *Naegleria lovaniensis* and *Naegleria gruberi*. *Eur J Biochem.* 164:427–434.
- Režen T, Debeljak N, Kordiš D, Rozman D. 2004. New aspects on lanosterol 14a-demethylase and cytochrome P450 evolution: lanosterol/cycloartenol diversification and lateral transfer. *J Mol Evol.* 59:51–58.
- Roberts CW, McLeod R, Rice DW, Ginger M, Chance ML, Goad LJ. 2003. Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Mol Biochem Parasitol.* 126:129–142.
- Roger AJ, Simpson AGB. 2009. Evolution: revisiting the root of the eukaryote tree. *Curr Biol.* 19:R165–R167.
- Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. 1993. Isoprenoid biosynthesis in bacteria—a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem J.* 295:517–524.

- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
- Salimova E, Boschetti A, Eichenberger W, Lutova L. 1999. Sterol mutants of *Chlamydomonas reinhardtii*: characterisation of three strains deficient in C24(28) reductase. *Plant Physiol Biochem*. 37:241–249.
- Schneider R. 2007. Intracellular sterol transport in eukaryotes, a connection to mitochondrial function? *Biochimie*. 89: 255–259.
- Sillo A, et al. 2008. Genome-wide transcriptional changes induced by phagocytosis or growth on bacteria in *Dictyostelium*. *BMC Genom*. 9:291.
- Simons K, Toomre D. 2000. Lipid rafts and signal transduction. *Nat Rev Mol Cell Biol*. 1:31–39.
- Summons RE, Bradley AS, Jahnke LL, Waldbauer JR. 2006. Steroids, triterpenoids and molecular oxygen. *Philos Trans R Soc B Biol Sci*. 361:951–968.
- Trigos A, Castellanos-Onorio O, Salinas A, Yanez-Morales MJ. 2005. Ergosterol from *Phytophthora drechsleri*, a unusual metabolite of a member of this genus. *Mycopathologia*. 159:469–471.
- Tyler KM, et al. 2009. Flagellar membrane localization via association with lipid rafts. *J Cell Sci*. 122:859–866.
- Veron B, Dauguet JC, Billard C. 1998. Sterolic biomarkers in marine phytoplankton. II. Free and conjugated sterols of seven species used in mariculture. *J Phycol*. 34:273–279.
- Volkman JK. 2003. Sterols in microorganisms. *Appl Microbiol Biotechnol*. 60:495–506.
- Volkman JK. 2005. Sterols and other triterpenoids: source specificity and evolution of biosynthetic pathways. *Organ Geochem*. 36:139–159.

Geoffrey McFadden, Associate Editor

Accepted September 4, 2009