Dear Editor,

Firstly, I will say that I really appreciate the thoughtful and constructive comments of the editor and reviewers. Secondly, while some of the text is changed, the largest change is the addition of two figures that I hope will clarify some of the confusion.
There were comments about the colloquial tone. I apologize for this. Much of the content was originally written as a review with a very different style, but was instead extracted as discussion for this paper. I have tried to reword this as suggested.

1 Round 1: Editor: Iker Irisarri

- COMMENT: As noted by one of the reviewers, one key publication in MBE covers several important aspects that are directly relevant to this preprint. It puts the evolution of sterol/hopanoid pathways in a broader context and tackles several points that are of relevance for your work, and therefore should be appropriately discussed.

- REPLY: I am familiar with this paper. Their objective and results are different, but complementary to mine, and I will try to incorporate this.

- COMMENT: A recurrent concern seems to be tree rooting, which is neither trivial nor arbitrary. One of the Reviewers suggests an interesting reference and the MBE paper by Santana-Molina et al. discusses also this topic. Could the presence of SHC/OSC homologs in Archaea help in this task? This possibility is not mentioned by the author, but I assume archaean homologs are unlikely based on what is said in Santana-Molina et al.? Regarding the hypothesis in Fig 1B, I also agree that the bacterial root probably makes most sense. However, the hypothesized evolutionary history of these enzymes rests on this particular rooting for which evidence seems to be scarce besides the phylogenetic distribution of paralogs that seem quite complex anyway. As pointed out by the Reviewers, this assumption should be made explicit and the real uncertainty of the hypothesized scenario better reflected by revising the tone of the relevant sections. The uncertainty could also be better reflected in other sections of the manuscript as well, such as when discussing biomarkers in fossils.

- REPLY: I agree that this is a major point, and I will try to clarify this in the paper. There is an unfortunate bias in interpreting trees based on how they are drawn, which is evident in the Takishita 2012 paper. Their tree was unrooted, but drawn with a root between SHC and OSC, that is, the root was based on the difference in activity of the enzymes rather than the evolutionary history. There was likewise no biological or scientific justification for this in their case either. Even for Santana-Molina 2020, they argue that the SHC (bacterial) part of the tree begins to look like the bacterial tree of life, so the root may be roughly the LBCA and therefore lacks an obvious point to root it.

- COMMENT: From my viewpoint, the manuscript contains too many expressions that e.g., personalize molecules, assign choices to organisms, or claim bizarre loss events. Such expressions fit very well into what has been called the night science language (Yanai & Lercher 2019; https://doi.org/10.1186/s13059-019-1800-6) but should be avoided in scientific publications. Note also Reviewers comments on specific sections where the writing could be improved for clarity and correctness. Finally, there are a few demonstrative pronouns without an object following them (this), which sometimes make it difficult to follow the argumentation (e.g. point 2.3).

- REPLY: Again, I apologize for this. I have tried to reword some of these.

- COMMENT: Regarding the general structure of the paper, I think the current narrative structure helps transmitting the Authors view, but the data behind the claims are not always clear. For example, when

laying out the evolutionary history in points 2.2.1 to 2.2.6, it could help to refer to specific parts of the tree that back up each claim, thereby better connecting the phylogenetic hypothesis and the biological interpretation.

- REPLY: I do this in Figure 1B, with numbers 1-4, but I will try to reference this better in the text.
- COMMENT: I appreciate the transparency of centrally depositing alignments and trees in Bitbucket. It might be good to add a fully annotated tree as a supplement, given that the current taxon names in the tree short and not very informative, and only a cartoon of the tree is presented in the manuscript.
- REPLY: I have added annotated and colored PDFs of both IQ-tree outputs to the repo, with BS values.
- COMMENT: Fig. 1 needs support values (at least for the key nodes) and perhaps a rewrite of the caption because the current one Recreated tree from [Takishita et al., 2012] might sound like the tree correspond to the old tree by Takishita et al.
- REPLY: The tree in Figure 1A **was** recreated from Takishita 2012. They had sent me their alignment, and I had remade the tree, and redrawn it with the same color scheme as part B, specifically to compare the rooting side by side. This was meant to highlight how our interpretation is heavily based on how we draw the tree or decide the root. This point may not have been clear. In any case, I have added a few support values to the figure in part B, and I have also changed the wording of the caption.
- COMMENT: Note also the comment that asks about how the new added sequenced improved our view on the story. Could the data of Santana-Molina et al. MBE paper change that? Please, take into account the considerations on phylogenetic methodology raised by two of the Reviewers. I wondered whether using HMM or psiBLAST could help identify more distant homologs that might be relevant. For BLAST searches, significance thresholds should be provided.
- REPLY: Human OSC will already blast to bacterial SHC with evalue of 1e-40. I doubt HMMs will do anything to improve that.
- COMMENT: Lastly, one of the Reviewers suggests adding a schematic figure setting up the stage for the evolutionary questions. This might clarify one doubt I had: Fig 1 of Santana-Molina et al. depicts SQMO as a step in the sterol but not in the hopanoid synthesis, whereas in the preprint I understand that this oxygen-dependent step of SQMO is common to both hopanoid and sterol production: Following the oxygen-dependent step, one of two enzymes then forms the multi-ring structure, either squalene-hopene cyclase (SHC) for hopanoids or oxidosqualene cyclase (OSC) for sterols. Which one is right or am I missing something here?
- REPLY: I have added a schematic. You are right, this was my mistake in the wording of the pathway. The pathway diverges after squalene, either to SHC, or to SQMO then OSC. I have also fixed this wording in the text.

2 Reviewer 1: Samuel Abalde

- COMMENT: However, I do have one major concerns that is related to the general tone of the manuscript. Here, the author proposes a new hypothesys for the origin of SHC and OSC enzymes. After reading the manuscript a couple of times, I always have the feeling that the author presents this new hypothesys (LECA already had both enzymes) as a fact based on the evidence, but that is not true. I like this hypothesis and I think that is very likely, based on the results presented, but all the body of evidence relies in one simple assumption: that the tree must be rooted in one particular branch. Even if that rooting makes sense based in our current knowledge, the lack of an outgroup is a

major burden because rooting the tree in a different branch might end up in a different story and we cannot disregard that possibility. This limitation is of course addressed in the main text, but I think the manuscript should never leave that hypothetical ground, if I may say so. To make this point clear: an extended dataset suggests the presence of SHC and OSC in the LECA, instead of LECA already presented SHC and OSC.

- REPLY: I tried to reword this.
- COMMENT: I find confusing the use of so many acronyms along the text. SHC, STC, OCT, SQMO, HGT... I understand the use of these acronyms, its just that sometimes it requires an extra effort to understand what the author is talking about.
- REPLY: Again, I apologize. Sadly, much of cell biology is memorization of 3 or 4 letter acronyms. I have tried to make the names of the genes clear in the schematic figure. I could also add a list of abbreviations at the end. I understand the challenge here, as comparing across the tree of life will combined conventions for naming among very different organisms. OSC, oxido-squalene synthase, is the protein created by the genes *erg7* in yeast (for ergosterol biosynthesis 7) and *LSS* in human (for lanosterol synthase). I feel it is more intuitive or understandable to refer to OSC rather than to *erg7* or *LSS*.
- COMMENT: I wonder if hopanoids and steroids appear in the same rocky substrates and, if not, if there is evidence of one of them in more ancient rocks. In the conclusion the need of a thorough review of the fossil record is mentioned, and I agree that this would be a really valuable addition to this problem.
- REPLY: This was indeed a motivation for my study. There is some discussion of this already by Brocks (2017 Nature), which is cited in the manuscript.
- COMMENT: Page 7, some bacteria appear have: appear to have. - Page 8, the genomes of cnidarians (corals and jellyfish), ctenophores: and ctenophores.
- REPLY: Thank you, these have both been fixed.

3 Reviewer 2: Denis Baurain

- COMMENT: I don't have strong feelings against this specific piece of work, except the fact that it was released one month after the publication of a large study covering the same topic (and more) by Devos et al. in Mol Biol Evol (2020) [doi:10.1093/molbev/msaa054]. Therefore, I think that the latter study should be mentioned and its findings discussed in the present manuscript to be really useful to the community.
- REPLY: As mentioned above, I have tried to incorporate this, though that study focused mostly on bacteria.
- COMMENT: The introduction is just a bit too brief to fully understrand the study for someone not familiar with the pathways at play. Indeed, Fischer and Pearson (2007) do not provide an overview of the two pathways nor of their interplay, whereas Nes (2011) is too technical for the argument here while providing no details about hapanoids. Figure 1 in Devos et al. (2020) is much better in this respect, but still overly complex for what is needed here. Please consider adding a schematic figure (or panel to an existing figure) to efficiently set up the scentific background. This would also allow the author to precisely define what is STC, which is not yet achieved in the current version of the manuscript.
- REPLY: I see what you mean. I have added a brief schematic of the pathway to a new figure.

- COMMENT: I am convinced that the author knows what he is discussing. However, I am sometimes puzzled by the vocabulary. For example, this sentence is ambiguous to me: "All else being equal, the presence of OSC in some bacteria and stem eukaryotes is nonetheless best explained by primary inheritance of OSC by a pre-eukaryotic host from a bacterial endosymbiont." Similarly, "This is explained by vertical inheritance from bacteria at the origin of eukaryotes (probably endosymbiosis)" and "primary inheritance in the LECA from a bacterial endosymbiont at the origin of mitochondria" look confusing.

- REPLY: I tried to reword this.

- COMMENT: In my opinion, it is important to distinguish the multiple possible cases here. Even if a pathway (or enzyme) is ancestral to all extant eukaryotes (i.e., present in LECA) and then vertically spread, it can be there for multiple reasons: 1) ancestral to all three domains (vertical evolution), 2) present in the archaeal host cell (if there was any such thing, again vertical evolution), 3) provided by a bacterial symbiont (e.g., the future mitochondrion, thus endosymbiosis), 4) "invented" along the eukaryotic stem (i.e., ESP, none of the above). Alternatively, the pathway (enzyme) can be introduced into a subset of extant eukaryotic lineages through H/LGT (horizontal/lateral evolution). These options should be clarified upfront. Regarding case (3), it may be useful to mention that, according to Devos et al. (2020), a mitochondrial origin of OSC is unlikely ("The scarcity of OSC in Alpha-proteobacteria indicates that this bacterial contribution is unlikely to be related to the mitochondria").

- REPLY: I thought that I had explained many of these possibilities in the paper already, but perhaps they were not explained well. Regarding case (3), a single, ancient loss can remove the gene from nearly all alpha proteobacteria, so this is not evidence against a mitochondrial origin.

- COMMENT: Another example is the lengthy discussion about rooting (split into multiple parts in the manuscript). While I of course agree with the fact that most phylogenetic reconstructions do not yield rooted trees, I am less fond of the claim that rooting is necessarily an "arbitrary decision". What is required is external evidence, and indeed distinct functions in case of multigene families can be such evidence. Here, I have the impression that ancestrally duplicated genes (i.e., prior to the evolution of the three domains) appear unlikely to the author ("this implies a parallel origin of the two enzymes relative to the unknown outgroup"). Considering that such genes do exist, I think this position should be better argued. For some ideas about rooting the tree of life, I humbly refer to our own piece: Gouy et al. (2015) in Philos Trans R Soc Lond B Biol Sci [doi: 10.1098/rstb.2014.0329].

- REPLY: The reviewer raises some good points, but the issue, and key point in time, is related to eukaryotes, not the root of life.

- COMMENT: Somewhat in opposition to the previous comment, Devos et al. (2020) argue that "the current data do not provide enough resolution to infer the presence of SHC in a common bacterial ancestor, although the results do indicate that the biosynthesis of hopanoid (defined by the SHC enzyme) precedes the diversification of the whole Gracilicutes group and some of the Terrabacteria taxa". This should be kept in mind for the first step of the scenario developed in the present study ("SHC was the original enzyme, distributed across many bacterial lineages"). Related to that, the following sentence is confusing ("nor is it clear which bacterial lineages may have possessed this enzyme due to horizontal gene transfer between prokaryotic lineages"). For me, the question is not well formulated: either SHC is ancestral to all existing bacteria and there is no point discussing extant lineages or it appeared in a lineage that was already distinct from other still-extant lineages and thus not ancestral to all. What is the author's hypothesis? Furthermore, the sentence about OSC emergence by duplication of SHC (or a SHC/OSC ancestor I would rather say if I reasoned as a pure cladist) draws a parallel with the distribution of SHC that is unjutified to me ("it is also currently unclear who

this might be", emphasis mine) because OSC is much less extensively distributed among bacteria and thus the question of the lineage of origin appears more legitimate than for SHC.

- REPLY: I see your point, I have tried to clarify this in the text.

- COMMENT: Co-emergence and/or co-evolution of OSC with SQMO ("It is very likely that this was coincident with the origin of SQMO") would make sense but is not strictly required since all these enzymes seem to exhibit some level of metabolic plasticity (see Devos et al., 2020). In the present work, SQMO distribution and evolution is not discussed at all. So it is difficult to take a stance about this issue. Related to that, understanding the "choice made in each [eukaryotic] lineage to keep either STC/SHC or OSC" would require to answer the following two questions: Can STC work after SQMO? Can OSC work without SQMO? All these points should be better discussed based on the current literature. In this respect, the explanatory ideas about the similarly patchy distribution of elongation factors (EF1a/EFL) might come handy (e.g., Keeling and Inagaki, 2004, PNAS; doi: 10.1073/pnas.0404505101).

- REPLY: These are all good questions. As you suggest, there is likely some metabolic plasticity between different species. In general, it appears that OSC cannot work without SQMO. There are antifungal compounds that block SQMO (e.g. terbinafine), suggesting that these pathogenic fungi cannot bypass this step.

- COMMENT: Regarding Methods and Figures, I think that they can be improved. Figure 1B should differentiate at least some bacterial groups if the origins of OSC (and eukaryotic SHC/STC) have to be discussed (as it appear to be the case in the main text, e.g., "Fungi have acquired SHC from an ancestor of Anaeromyxobacter, a delta-proteobacterium"). Moreover, some idea of the statistical support at important nodes is required to understand if there is phylogenetic resolution at all. For example, the dimensions of the alignment should be provided. Related to that, how were identified the MMETSP sequences (by TBLASTN searches against mRNA sequences or by BLASTP against predicted proteins)? How did the author deal with the contaminant sequences plaguing some samples of the original MMETSP dataset?

- REPLY: The MMETSP data were very messy, though this added almost entirely OSC sequences, which does not affect the root of OSC. I had to make multiple iterations of the tree, and removed contamination where I could obviously tell it was contamination, that is, when any given taxon had 2 full length proteins that ended up in different parts of the tree.