nature portfolio

Peer Review File

Serial innovations by Asgard archaea shape the DNA replication machinery of early eukaryotic ancestor

Corresponding Author: Dr Fabai Wu

Version 0:

Decision Letter:

20th March 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Fabai,

Your manuscript entitled "Replisome innovations bridging archaea and eukaryotes" has now been seen by three reviewers, whose comments are attached. The reviewers have raised a number of concerns which will need to be addressed before we can offer publication in Nature Ecology & Evolution. We will therefore need to see your responses to the criticisms raised and to some editorial concerns, along with a revised manuscript, before we can reach a final decision regarding publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file in Microsoft Word format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a “Response to reviewers” document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

* Extended Data Figures - please ensure that any supplementary figures and tables that are crucial to the manuscript’s conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Yours sincerely,

[redacted]

Reviewer expertise:

Reviewer #1: eukaryogenesis, phylogenetics

Reviewer #2: archaeal DNA replication, molecular biology

Reviewer #3: Asgard archaea, eukaryogenesis, phylogenetics

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In this manuscript, Feng and colleagues report a range of analyses on the evolution of the replisome (DNA replication) during the transition from Asgards to eukaryotic cells. The paper brings together a broad range of new data (new MAGs) and analyses (phylogenetic and experimental), and draws several interesting conclusions about how the replisome functions in modern Asgards, and some of the steps of eukaryogenesis. The phylogenetic analyses (on which I am qualified to comment) are performed to a high standard, and the comparative genomic inferences seem reasonable. I have a few specific comments and criticisms below which I hope will be useful; overall, my major critique of the paper is that it does not (in my view) engage enough with the implications of the findings for eukaryotic origins, as I outline below. The organisation of the paper (and brevity) could also be improved, with - for example - the section on cdc6 either being shortened, or its connection to the rest of the paper clarified.

1. The different scaling of replisome gene content with total genome gene content in different lineages is an interesting observation, but the significance for the issue of eukaryogenesis is not very clear. If the correlation is driven by variation in cdc6 copy number, the question is then why cdc6 copy number varies with the size of the gene repertoire. The manuscript devotes a fair amount of space to this point but the insight provided seems limited.
2. I found the manuscript a little challenging to parse in terms of its organisation. There are a number of quite interesting points made - for example, inheritence by eukaryotes of horizontally acquired genes in Asgards; experimental work showing that protein complexes in some Asgards appear "eukaryote-like". However, these observations were not drawn together into a (relatively) simple conclusion about eukaryogenesis. One potential approach to "fix" this would be to begin the paper with a clear explanation of the set of genes that constitute the "replisome" (maybe with a diagram, to link to the final summary diagram), and to clearly outline the open questions about replisome evolution that the paper addresses, and the consequences (or stakes) for debates about eukaryogenesis. For example, a main recurring result is that some Asgard replisomes may be eukaryote-like in terms of the number and interactions of components. If so, how does this change our understanding of eukaryogenesis? Are there aspects of eukaryotic genome organisation that are correlated with having a certain kind of replisome / set of replisome proteins, such that the findings would allow inferences to be drawn about e.g. the genome size or structure of the archaeal ancestor? Without clear implications, the impact of these findings is limited, despite the novelty and interest of experimental work on Asgard genes.
3. Recombination vs. duplication - the paper uses "recombination" to mean "gene transfer across large phylogenetic distances" a number of times. I think this will confuse readers, who might typically think of recombination as something that happens between fairly closely related (and so homologous) chromosomes.
4. Lines 330-340 - very interesting to combine comparative genomics with experimental data, for example investigating the formation of Lokiarchaeum RFC complex. But need more context - does the complex do, what does this mean for the prokaryote-to-eukaryote transition.

Minor comments

While the term "replisome" is defined in the first couple of sentences, I think the concept could benefit from being expanded a bit in the introduction, or at beginning of the results. For example, what is the list of genes that make up the replisome gene set, and how do their numbers change with replisome "copy number"? The italicization of "rep" makes it look like a particular gene, but as far as I understand it, what is intended here is something like "the set of genes (including DNA polymerase subunits and various associated factors) that collectively carry out DNA replication" or similar.

Line 46 - "recombinants" - evolved/arose from recombination between…? (clarity/unpacking the concept).

69-71 - increase in complexity … mobile genetic elements … adaptive traits - not clear what is meant here. Maybe delete or revise to clarify the meaning? It seems the text may be referring to the somewhat fraught issue of which, if any, features of eukaryotic cells are specifically adaptive and contribute to eukaryotic "complexity", but it's too vague to really say much. For example, if the aim is to provide background for why the evolution of the replisome is interesting, I would maybe focus on that here instead.

Lines 115,122 - not sure what "phenomenological" and related words mean here. Could they be deleted? In the first instance, there is simply a pattern, as far as I can see.

209 - "mostly maintained"

246 - "C-terminal"

Reviewer #2 (Remarks to the Author):

Feng et al 2025 provide an interesting perspective on the evolution of the DNA replication proteins in archaea and how they might have evolved during eukaryogenesis, focussing on the number, type, interactions, and structural arrangement of these genes and their encoded proteins in Asgard archaea, which are considered the closest microbial relatives of Eukarya. They identify some interesting resemblances of various Asgard archaeal replication proteins to modern eukaryotic ones, although the replisomes appear generally more typical of archaea as could be expected based on previous work placing Asgard archaea in the tree of life. Several of the predicted interacting replisome domains were verified by yeast two-hybrid interaction tests. The mix of these eukaryotic-like Asgard replication proteins in various lines of Asgards creates a complex picture of how they may have ended up in LECA. Discussion could also include any evidence for whether they may have all been present in the archaeal progenitor of LECA and then some were more recently lost in modern Asgard lineages to create the patterns seen today in Asgards, or how they may have come together from different archaeal sources in early eukaryotes, where possible.

My main points the authors should consider are:

Line 102. It is not clear here what homologs were searched for (‘replisome genes’) and why. For example, were the searches based on accepted comprehensive lists of experimentally verified components of the archaeal, yeast or other replisome, or what defined a rep gene?

How could the poorer knowledge of replication in some clades confound the scaling analysis – please briefly explain this and how it was dealt with.

Fig. 1b. It would appear important to report correlation coefficients, and variance (SD) in rep gene count for these? And then interpret the data critically when observing the slopes (potential errors in) and their differences (Fig. 1c).

Line 125-6. Please provide statistical support for this noted difference in slope (if any?). It is not clear what the statistical significance means in Fig 1c – it is more important to compare differences in slope given the interpretation in the text, not just whether there is a significant slope or not.

Line 148. Could the presence of two homologous families (Cdc6 and Orc1) in eukaryotes be evidence against this statement (only 1 in LECA), given archaea do not appear to have distinct functional Orc1? If two were in the LECA, one may have developed as Orc1 in LECA?

Minor points:
Line 72. The meaning of this statement in the context of the paragraph is not very clear: “By contrast, the underlying evolutionary constraints or selective force for conserved essential functions such as DNA replication remain unclear.” Should this include reference to selective force that increases complexity?
Line 74-75. Fix grammar throughout this sentence.
Line 113. Haloarchaeotes – is this a correct term? Revise.
Line 115. How are ‘small-genomed archaea’ defined? Clarify please.
Line 246. Clarify which Pol-delta1 here is the eukaryotic one.
The whole paper would benefit from a thorough grammar and typographical edit, which would greatly improve readability.

Reviewer #3 (Remarks to the Author):

Review Wu, 2025

In this interesting article, Wu et al. combined phylogenetic, statistical, and biochemical approaches to investigate replisome evolution and the archaeal origin of replisome components in eukaryotes. Specifically, the authors discovered that ceratin

Asgard archaea encode homologs of some eukaryotic rep genes that were previously not identified in prokaryotes. Indeed, it seems that different eukaryotic rep genes appear most closely related to homologs of distinct extant asgardarchaeal lineages, which is similar to observations on EPSs made in previous work that indicates that Asgardarchaeal genome evolution has been shaped by gene loss.

While the work is extensive and interesting, I found the article a bit difficult to read and I am not entirely sure about some of the conclusions drawn by the authors: e.g. “Our study thus suggest an emergence of eukaryotic replisome signatures prior to eukaryogenesis via distributed innovations across Asgard archaea, implicating an important role of horizontal gene transfer in the assemblage of LECA’s core genetic system”.

In turn, below I provide suggestions that could potentially help to improve the presentation of the work:

Major Comments

Restructuring of paragraphs
The presented data is interesting and relevant but I had the impression that it is not yet very coherently presented: e.g. Figure 4 shows different snippets but misses to convey a key message. I wonder whether the authors could consider restructuring their article and figures a bit. Please see below a suggestion of how the data could be presented in a way that’s potentially more easy to follow. It would also be great if the result and discussion sections could be carefully proof-read with regard to typos and grammar and language more broadly as various passages are currently difficult to understand.

For example, in my opinion, the results section about copy number and scaling is a bit lengthy and could be shortened with parts of it (including aspects shown in Figure 1) being moved to a supplementary material. Instead, you could perhaps provide a bit more information about the inventory of the various replication proteins of archaea and eukaryotes (and of bacteria if they have homologs) and the phylogenetic origins of eukaryotic homologs based on phylogenetics analyses. Potentially, you could consider following this by discussing the placement of eukaryotes relative to archaea based on these rep genes. This would set up the framework a bit more clearly. And from this broad look, you could then focus on certain aspects in more detail that best support the key message of the article.
Perhaps it would also ease readability if paragraphs would have headers related to the respective major findings presented (as an example, headers like “Traceable evolutionary history” are not very informative). Likewise, in terms of Figures, I think it would be helpful, if the overviews are presented in Figure 1 (e.g. elements of Figure 2 a and Figure 4a), with subsequent Figures focusing on one functional aspect per Figure that support a certain observation based on a set of different analyses. Other results could perhaps be moved to the Supplements or Extended Data Figures? For example, the discussion of the syntheny analyses could potentially be discussed more concisely with details being moved to Supplements.

Finally, it would be great if you could then come back to the big picture in light of all the observations made. How does this all fit into our broader view on eukaryogenesis? E.g. regarding origins of eukaryotic rep gene homologs, two major aspects could be discussed in the conclusion:
1. Genes inherited from Asgards
It seems that different replication machinery has distinct (Asgard) archaeal origins: ie. that FECA had all these elements that have been retained in some of the extant Asgard archaea but lost in others. This patterns seems similar to what was observed for ESP distributions: while Asgards together combine many ESPs, different ESPs have been retained in different Asgard lineages.
2. Components that have been acquired (from where?) or evolved during eukaryogenesis (e.g. by duplication) from FECA to LECA. What is the evidence for duplications versus HGT versus vertical inheritance with loss in Asgard archaea taking into account the various results including topology testing.

Topology testing
The authors suggest that “phylogenetics indicates that these increased gene copies did not arise through intra-lineage duplication” and that “these eukaryotic gene copies do not correspond directly to the individual copies that emerged in the Asgard archaea”. What is the support for these inferences, i.e. the observed tree topologies and the relevant branches? Considering the uncertainty in single gene trees that are based on short alignments and are prone to comprise long-branch attraction artifacts (e.g. when distant paralogs that evolved under distinct evolutionary rates are included), it might be possible that the trees are poorly resolved or cannot distinguish among alternative placements (e.g. that (Asgard) archaeal paralogs cluster basal to respective eukaryotic homologs). To test this, it would be helpful if you could perform topology tests, i.e fix branches to the alternative hypothesis and test whether these can be rejected under the best fitting model?

Data availability
All identified rep genes for each archaeal and bacterial taxon should be listed in a table providing accessions and annotations to verify that they are true homologs and to ensure reproducibility

Minor comments:

- check for spelling errors, e.g. plural vs singular forms of verbs
- line 40: consider modifying “three billion years ago” since recent research indicates an earlier divergence between Archaea and Bacteria

- rephrase: “the gross conservation from archaea” for better clarity? E.g. “in comparison to Archaea…”
- line 62: “Asgardarchaeota phylum”. There are currently two validated taxonomic names for this phylum, potentially cite those in this context?

Tamarit D, et al. Description of Asgardarchaeum abyssi gen. nov. spec. nov., a novel species within the class Asgardarchaeia and phylum Asgardarchaeota in accordance with the SeqCode. Syst Appl Microbiol. 2024 Jul;47(4):126525. doi: 10.1016/j.syapm.2024.126525.

Imachi H, et al. Promethearchaeum syntrophicum gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage 'Asgard' archaea, and proposal of the new archaeal phylum Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. Int J Syst Evol Microbiol. 2024 Jul;74(7):006435. doi: 10.1099/ijsem.0.006435.

- line 65: should this be past tense? “leading…”
- line s86: “so far” without dash?
- line 108: rep count (without s)
- line 112: which have?
- line 113 and 121: what do you mean by “bridge”? I am not sure that’s the best term in this context
- line 131-134: how sure can you be that you picked up all rep genes? Could it be that less well investigated archaea encode rep proteins that have so far been missed? If so, could this affect the relations between rep genes and total proteome size?
- perhaps check that you use consistent taxonomy, e.g. what do you mean by Haloarchaeotes? I think, in GTDB the group encompassing Haloarchaea, Methanosarcinales etc have been referred to as Halobacteriota? Or which taxonomy do you use (check for consistency)
- Figure legend 1, line 2. Archaea already include Asgard archaea so perhaps its not needed to mention those separate?
- Figure legend 1: what do you mean by monophyletic versus paraphylyetic? At which level was monophyly required? E.g. some of the trees labeled as paraphyletic still seem to have lineages that are monophyletic, e.g. the Thors and Lokis in C? Or do I misunderstand this?
- Figure legend 2, i.e. line 172: do you mean “replisome single-copy gene trees” or just one tree?
- Figure legend 2, i.e. line 187: do you mean “recombination” or “horisontal gene transfer”?
- line 195: with “Close examination of individual rep families” do you mean the close examination of the phylogenies of individual rep families? Or something else?
- line 242: “structural innovations bridge the archaeal and eukaryotic replisome”. This is an example of a sentence that has multiple meanings: e.g. in its literal sense, this sentence suggest there is a (physical) bridge between the archaeal and eukaryotic replisome established by structural innovations. I feel like it would be better if you could try to be more precise in language use to ease understanding.
- pay attention to use accurate taxonomic terms rather than abbreviations/slang, e.g. for Baldrarchaeia
- line 275: “relative subunit sizes”: what does this mean?
- Line 361: what do you mean “chimeric origin” and “chimerism”? Chimerism of what?
- what does this mean: “This coincides with their domain changes as described above”?
- the section on gene organisation patterns seems very speculative to me. I think its important to keep in mind that extant Asgard archaea do not reflect the proteome and gene arrangements that were present in the archaeal ancestor of eukaryotes. Furthermore, most of the asgard genomes are MAGs and don’t have complete genomes which could make it harder to identify synteny. More importantly however, I think ancestral genome inferences and the reconstruction of the proteome of the lineage leading to eukaryotes would be needed to investigate the scenarios. Also the wording in this section is often difficult to understand: e.g. “obligately, differentially lost”… And what do we know about the fitness cost?
- line 436: rephrase “bridge archaea and eukaryotes.”
- line 450: microbial evolution is known to be shaped by HGT so this is not very surprising?
- line 457-8: you suggested: “implicated that HGT-based gene recombination was essential at the early stage of replisome formation during eukaryogenesis”. While this might be possible, the data rather shows the ongoing evolution in Asgardarchaeota after their divergence from the lineage leading to eukaryotes and not what happened during eukaryogenesis? Since you did not infer the proteome set in the archaeal lineage leading to eukaryotes but in extant Asgardarchaea?
- line 460: which origin do you mean here?
- what is the reason to use LG+C60+F+G4 for the single gene trees while C10 was used for the concatenated tree? I would recommend using C60 also for the concatenated tree as the alignment is longer and it therefore makes more sense to have more categories. C60 might be too much on single gene alignments depending on the alignment length. Perhaps the authors could perform model test to compare C10-C60 and choose the best fitting model. (in addition to performing topolgy testing as suggested above)
- Data availability: it would be helpful if the authors can also make alignments files available (both before and after trimming). Furthermore, it would be valuable if all workflows and scripts for the various analyses were made available
- Figure 2B: how reliable are those topologies and the placement of the eukaryotic lineages relative to asgardarchaea? I think it would be very umportant to show support values and sister lineages more clearly as many of these rep genes do not seem to group eukaryotic homologs with Asgard: e.g. its hard to see the differences of the circles indicating bootstraps, maybe color-coding would be better. Considering that the placement of the eukaryotic branch is key to the interpretation of the work, I feel the single gene trees should be presented more accurately.
- Table S1: It would be helpful if you can add three columns to inform about genome/MAG completeness, contamination and redundancy for each of the archaeal taxa analysed
- for eukaryotes: how do you determine copy number? Are these all complete genomes or also transcriptomes? Does this include different splice variants?

- Figure S3: did you verify that your search did not pick up false positives or exclude false positives in case those were obtained? Furthermore, I think it would be important if you can provide IPR/PFAM/KO/COG annotations for all the rep proteins of all archaea for verification and reproducibility
- Table S3: can you provide a link to a database and protein accessions?
- can you provide a S-Table listing all the taxa that were used in phylogenetic analyses of rep genes?
- Did you add Bacteria whenever homologs were present? I think that would be helpful for rooting and to rule out that certain components may derive from Bacteria. If this is what has been done, perhaps mention it in the methods (unless I missed this). In general, it would be good if you can provide a table with all taxa and all rep genes as mentioned above.
- you mention “Cdc45 appear to be phylogenetically related to RecJ/Cdc45 proteins encoded by the Lokiarchaeia”: what is the support for this? The tree topology looks poorly resolved.
- Fig S8b: are the B6 Aenigma 1 sequences also contaminations from Baldr?

Reviewer #3 (Remarks on code availability):

I briefly checked the repository and it seems it mostly provides supplementary data

**********END**********

Version 1:

Decision Letter:

3rd July 2025

*Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors.

Dear Fabai,

Your revised manuscript entitled "Serial innovations by Asgard archaea shape the DNA replication machinery of early eukaryotic ancestor" has now been seen by the same three reviewers, whose comments are attached. The reviewers believe the study has improved but they still have a number of concerns which will need to be addressed before we can offer publication in Nature Ecology & Evolution. We will therefore need to see your responses to the criticisms raised and to some editorial concerns, along with a revised manuscript, before we can reach a final decision regarding publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file in Microsoft Word format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

* Include a “Response to reviewers” document detailing, point-by-point, how you addressed each reviewer comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the reviewers along with the revised manuscript.

* If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions at http://www.nature.com/natecolevol/info/final-submission. Refer also to any guidelines provided in this letter.

* Include a revised version of any required reporting checklist. It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review. A revised checklist is essential for re-review of the paper.

* Extended Data Figures - please ensure that any supplementary figures and tables that are crucial to the manuscript’s conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your

homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know. We will be happy to consider your revision so long as nothing similar has been accepted for publication at Nature Ecology & Evolution or published elsewhere.

Nature Ecology & Evolution is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Yours sincerely,

[redacted]

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

Feng and colleagues have substantially improved the readability and clarity of the manuscript, in line with the suggestions from all of the reviewers in the previous round. This is now an interesting contribution that, in my view, merits publication.

Reviewer #2 (Remarks to the Author):

The authors have carefully addressed the previous comments and revised their manuscript giving reason, and it's significantly improved technically and in organisation. Below are some points on the new aspects of the paper:

Line 32. For the abstract, I believe this sentence could be more specific and clearer or may not be needed. For example, how did the distributed innovations collectively drive protein complexes to 'evolve functional divisions?' it's not clear to me, at this stage, how this could have driven the evolution more than any other model, as this seems to imply. i.e. Why would a sub-complex innovation in one line of Asgards be different or special in driving evolution of complexity, compared to another potential model where all of the innovations occurred sequentially (vertically) in one line leading to LUCA?
And what is meant by 'functional divisions' here? Duplication and diversification? Clarify, otherwise it might initially sound like one protein evolving to do multiple functions, which is not what the authors are saying with the paper.
Should a sentence like this be at the end of the abstract instead (if even needed given the rest of the abstract)?

Line 33. Grammar – 'expansions'

Line 36. Grammar/meaning – 'be sustained by' (?)

Line 110. and throughout. Make consistent labelling of figs S1, 2 etc, or 'extended data..'

Line. 269. One/single same meaning – avoid duplication of meaning.

Fig. 3 and elsewhere. Much of the integrated text is extremely tiny. This should be enlarged so at least readable. Zooming can help but there is space to enlarge this text.

Fig 4C. How were the replicate data (n) obtained (enabling calculation of an SEM and p-values from t-tests)? Indicate in the legend. The use of t-tests appears unclear here where there is no obvious way to obtain replicates of the regressions unless random nonredundant subsamples of the data could be used for separate regressions and their slopes taken as the replicates. But a common test for different slopes of two regression datasets would be an analysis of covariance. The results of both approaches could be noted, if appropriate.

Line. 391. Grammar – 'includes'

Line 396. Clarify: 'more dynamic evolutionary path than previously assumed". What is defined as dynamic here, and who assumed?

The latter two sections of the 'Results', especially the last one, read more like discussion. Consider placing the last section discussing Fig 6 at the start of the discussion?

Reviewer #3 (Remarks to the Author):

The authors have addressed the comments raised by the reviewers and restructured the article according to suggestions. I think this has considerably improved the clarity and readability of the article. The authors have also put the results into a broader perspective on eukaryotic origins.

I very much appreciate the extensive work of the authors on this manuscript, yet I am concerned that some of the results are over-interpreted (e.g. regarding the hypothesis that the eukaryotic replication machinery has origins from different Asgard archaeal lineages) and that presented scenarios/hypotheses (e.g. in Fig 6) are not sufficiently supported by the current analyses. Single gene trees are prone to phylogenetic artifacts and the topologies presented do not always align with the proposed scenario, see for example the sections on the origin of Pol α/ Pol δ/ Pol ζ. While the author indicate that they have performed AU tests, those are not discussed in some of the relevant sections so that the reliability of certain placements is not yet clear to me (I provide some examples below to indicate where results from AU test would be valuable to be mentioned explicitly). Furthermore, the presence of homologs in the closest archaeal sister lineage (i.e. Heimdalls or a subclade within) followed by subsequent loss in currently sampled representatives (and retention in other lineages such as RFCs in Lokis) cannot be excluded. I would recommend the authors to to use a statistical framework for reconciling single gene trees with the species tree and then present results more objectively, potentially discussing different scenarios that could explain the data. See more specific comments below.

Main comments
I have some concerns regarding the interpretation of the results presented in this paragraph “Baldrarchaeia encodes a Polδ-like protein complex” and corresponding figures including Fig S6:
- The current tree topology (though, I do not know what the support is for the key nodes) would be most consistent with a scenario in which bacterial homologs were at the origin of eukaryotic Pol α/ Pol δ/ Pol ζ. Upon duplication, early eukaryotes would then have transferred a homolog of an ancestral Pol δ/ Pol ζ to Baldrarchaeia and Aenigmatarchaeota while a second duplication would have led to the origin of Pol δ and Pol ζ, respectively. Which bacteria are at the base of eukaryotic Pol α/ Pol δ/ Pol ζ?
- What is the reason to suggest that the Baldrarchaeia and Aenigmatarchaeota homologs are more closely related to Pol δ? Based on the tree topopogy, they are basal to a monophyletic clade composed of Pol δ and Pol ζ and the node could be rotated. E.g. you mention “this fusion resembles the catalytic subunit of eukaryotic Polδ named Polδ1.” Can you exclude resemblance to Pol α/ Pol ζ? The alignment illusration in S6a suggests that the Baldrarchaeia and Aenigmatarchaeota homologs have retained some indels similar to Pol α (which would be compatible with the transfer to Bald and Ainigma scenario)?
- Could you reject using AU test that the Heimdall clade as opposed to the bacteria is basal to the eukaryotic Pol α/ Pol δ/ Pol ζ including the Baldrarchaeia and Aenigmatarchaeota homologs (that could have been acquired by these archaea from eukaryotes later as suggested above)?
- Can you reject monophyly of eukaryotic Pol α/ Pol δ/ Pol ζ? What is the support of Baldrarchaeia and Aenigmatarchaeota homologs placing inside of this cluster? E.g. can you exclude the topology of DP1 being reflected by DP2?
- Taking all this into account, I am not convinced by the scenario presented in Fig. 2g. and the section “Functional division of Eukaryotic DNA polymerases emerged sequentially”, i.e. the origin of eukaryotic Pol α/ Pol δ/ Pol ζ

- it would be great if you can show support values (even if as dots indicating support above 80 and/or 90) for trees in the supplementary material. It is otherwise difficult to understand the level of support for the different placements of eukaryotic homologs relative to archaeal and bacterial homologs.

- The tree in Figure 3a is difficult to read, so I wasn’t able to see the support of Loki as opposed to Heimdall homologs being basal to eukaryotic homologs. Perhaps mention if there was any AU-testing done in this context and what it suggests.

- The authors suggest that “replisome gene paralogs emerged primarily via HGT across Asgard archaea”. However, it is not clear to me whether the authors have rigorously tested this hypothesis in a statistical framework (I did not see any indication in the methods). To determine whether homologs emerged via HGTs as opposed to duplications or losses, it would be important to perform gene-tree / species-tree reconciliations that allow comparing single gene trees against a species tree and estimating gene family history including not only transfers but also duplications, originations, and losses. For example, the assumptions regarding duplication versus HGT as illustrated in Fig. 4D are too simplistic. E.g. a gene may have duplicated on a much earlier node followed by subsequent differential losses across various Asgard lineages and transfer to other Archaea and Bacteria. Without a statistical framework that also takes into account tree uncertainty, it is difficult to determine where Asgard archaeal homologs are coming from and I am therefore not convinced that the hypothesis of Asgard replication proteins being shaped predominantly by HGT is justified.

- In line with this, for the scenario presented in Fig 6, the authors would need to perform reconciliations and ancestral reconstructions to for example be able to infer the replisome set present in Lokis/Bald/Hel/Hods/FECA and in subsequent ancestors and infer later losses versus gains. This would allow to statistically infer whether subunits originated via HGT or were inherited from a common ancestor with subsequent duplication and differential loss in various Asgard lineages.

- The placement of the eukaryotic branch relative to Asgard archaea is currently unresolved. The authors infer one species tree based on just six homologs which are unlikely to present a final picture of the eukaryotic branch. Therefore, and due to the limitation in interpreting the single gene trees without statistical framework, the scenario in Fig. 6 is speculative. In light of

current literature and the conflicting results as to whether Heimdalls as a whole or specific sub-clades of Heimdalls are sister to the branch leading to eukaryotes, I would recommend doing reconciliations with both topologies and test which one is most supported in light of replisome gene repertoires.

Minor
- Line 84: I would delete environmental here: “environmental archaea” (aren’t all organisms environmental? And if used in opposition to cultivated, some Asgards are now in culture…) .

- Fig 1c: the absence of circles does not mean absence of homologs, right? If so, this might be a bit confusing. Could you indicate presence of homologs in grey and putative siser-relationship with eukarytoes in a nother color? To clearly differentiate Asgards that do not have a homolog at all versus lacking a homolog closely related to eukaryotes.

- can you refer to the underlying trees for the hypothesis presented in Fig 4e?

**********END**********

Version 2:

Decision Letter:

7th August 2025

Dear Fabai,

Thank you for submitting your revised manuscript "Serial innovations by Asgard archaea shape the DNA replication machinery of early eukaryotic ancestor" (NATECOLEVOL-25010164B). It has now been seen again by Reviewer #3 and their comments are below. The reviewer finds that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Ecology & Evolution, pending minor revisions to satisfy the reviewers' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Ecology & Evolution. Please do not hesitate to contact me if you have any questions.

Sincerely,

[redacted]

Reviewer #3 (Remarks to the Author):

I have read with interest the response letter by the authors to the recent round of revisions on their manuscript “Serial innovations by Asgard archaea shape the DNAx replication machinery of early eukaryotic ancestor”. I have also looked through the additions/changes to the new version of the manuscript.

I appreciate the authors efforts to address my concerns: I agree that doing gene-tree/species-tree reconciliations is time consuming and would indeed require to rerun single gene trees using a taxon set that corresponds to the taxa in the species tree. Considering that the authors have already put substantial effort and work in this, reconciliations may be something to implement in the future. I therefore understand that the authors prefer to focus on addressing my concerns using AU testing (even if not a substitute for gene tree/species tree reconciliations) and textual clarifications. I think that the topology testing that the authors implemented in this round was helpful and allowed clarifying the support for some of the key placements. However, while I am fully appreciative of the work, I am not yet convinced by the clarity of the textual presentation of some of the results and their discussion.

Unfortunately, I am currently not able to provide further detailed feedback on this submission, but point out a number of specific points regarding the recent changes below:

- The new additions need to be carefully checked for grammar and readability, in my view they do not always lead to improved clarity yet.

- Abstract, please rephrase “By reconciling events of gene gain and loss with an archaea-eukaryotic tree based on curated replisome markers” to something like: “By comparing curated replisome markers phylogenies with a archaea-eukaryotic tree” in order to avoid confusion. The phrasing which the authors chose here, is generally used when gene-tree species-tree reconciliation approaches have been applied. Since you did not perform such analyses, this could be a bit misleading.

- line 177: I would remove “yet” (since it could be the CTD was lost later)

- line 198: It would be great if you can specify what you mean by “less related homologs”. Can you describe any parameters/requirements that underlie the decision to call something “less related” for reproducibility?

- line 362-364: this sentence is a bit unclear to me, not sure its needed or adds anything.

- line 365: is this correct?: “cannot not be well resolved”

- line 378-390: some sentences may need some rephrasing, e.g. “burst of single-node expansion…”

- line 420-423: rephrase for clarity?

- line 513: “a complex evolutionary trajectory” instead of “divese evolutionary trajectories” (I think only one trajectory led to eukaryogenesis which included many events?)

Response to reviewers' comments

Based on reviewers' suggestions on the structure and emphasis of the paper, we made the following major changes to the structure:

1. To better guide the readers through the composition of the archaeal and eukaryotic replisomes, we added Figure 1 (the original Fig. S1) and associated introductory paragraphs. We also moved the original Fig. 4a to Fig. 1c to summarize the main proteins of interest in this study.
2. To make the structural innovations of protein complexes more accessible to readers, we split the original Fig. 3 into the current Fig. 2 and 3 and revised the texts. We also added illustrations and corresponding texts to more clearly hypothesize the emergence of functional division and hetero-multimerization of eukaryotic protein complexes.
3. Since the implication of the original Fig. 1 regarding gene scaling has limited implications for eukaryogenesis, we substantially shortened the texts, and now moved the panels to Fig. 4a-c. The original panel on Cdc6/ORC1 is dropped.
4. We added Fig. 6 to illustrate the main conclusion of the paper.

Below is the point-by-point response (in black) to reviewers' comments (in blue).

Reviewer #1 (Remarks to the Author):

In this manuscript, Feng and colleagues report a range of analyses on the evolution of the replisome (DNA replication) during the transition from Asgards to eukaryotic cells. The paper brings together a broad range of new data (new MAGs) and analyses (phylogenetic and experimental), and draws several interesting conclusions about how the replisome functions in modern Asgards, and some of the steps of eukaryogenesis. The phylogenetic analyses (on which I am qualified to comment) are performed to a high standard, and the comparative genomic inferences seem reasonable. I have a few specific comments and criticisms below which I hope will be useful; overall, my major critique of the paper is that it does not (in my view) engage enough with the implications of the findings for eukaryotic origins, as I outline below. The organisation of the paper (and brevity) could also be improved, with - for example - the section on cdc6 either being shortened, or its connection to the rest of the paper clarified.

Thank you for the positive comments on our manuscript. We also highly appreciate your reminder that the findings should be placed in the broader context of eukaryotic origin. We have now adjusted the structure of the paper to better emphasizing the main findings. We also removed the description on the Cdc6 genes which, while interesting on its own, is a sidetrack that does not connect to the eukaryotic origin. The current headings, the final figure, and the discussion section should to a better extent show the implication of our findings for eukaryogenesis.

1. The different scaling of replisome gene content with total genome gene content in different lineages is an interesting observation, but the significance for the issue of eukaryogenesis is not very clear. If the correlation is driven by variation in cdc6 copy number, the question is then why cdc6 copy number varies with the size of the gene repertoire. The manuscript devotes a fair amount of

space to this point but the insight provided seems limited.

We thank you for pointing this out. Indeed, we were intrigued by the scaling of the replisome genes, especially finding that Asgard archaea was not the lineage that expanded its replisome gene content most dramatically. We also acknowledge that its link to eukaryogenesis is quite implicit, for which we think will require future investigations. We now moved the figure 1 into Fig. 4a-c and shortened the relevant description (Line 329-339, under the section 'Asgard archaea show moderate replisome scaling'). Since the observation on Cdc6 is a sidetrack, we removed it from the figure and text.

2. I found the manuscript a little challenging to parse in terms of its organisation. There are a number of quite interesting points made - for example, inheritence by eukaryotes of horizontally acquired genes in Asgards; experimental work showing that protein complexes in some Asgards appear "eukaryote-like". However, these observations were not drawn together into a (relatively) simple conclusion about eukaryogenesis. One potential approach to "fix" this would be to begin the paper with a clear explanation of the set of genes that constitute the "replisome" (maybe with a diagram, to link to the final summary diagram), and to clearly outline the open questions about replisome evolution that the paper addresses, and the consequences (or stakes) for debates about eukaryogenesis.

Thank you for your kind advice in the structure of the paper. In the revised version, we now added a Figure 1 as you suggested to introduce the overall composition of the replisome, and the differences between archaea and eukaryotic replisomes. We also now clarify that the major challenge to understand the origin of eukaryotic replisome is the emergence of 'functional division' and 'heteromultimerization' (or homo- to heteromultier transition). We also added the final Figure 6 to summarize the main findings of the paper.

For example, a main recurring result is that some Asgard replisomes may be eukaryote-like in terms of the number and interactions of components. If so, how does this change our understanding of eukaryogenesis? Are there aspects of eukaryotic genome organisation that are correlated with having a certain kind of replisome / set of replisome proteins, such that the findings would allow inferences to be drawn about e.g. the genome size or structure of the archaeal ancestor? Without clear implications, the impact of these findings is limited, despite the novelty and interest of experimental work on Asgard genes.

Thank you for pointing out the importance of placing the various observations in the context of eukaryogenesis. At the most basic level, we found replisome features that were thought to only exist in eukaryotes. This should be of most apparent interests for cell biologists and biochemists. From evolution point of view, these features are least vulnerable to phylogenetic uncertainty and thus can be leveraged to examine the sequence of events around eukaryogenesis.

Since the literature of hypothesis regarding eukaryogenesis is broad, and our paper focuses specifically on the replisome, we would not necessarily argue that our findings 'changed' the understanding of eukaryogenesis. However, we are confident that the sequence of events we resolved can be used to constrain our understanding of eukaryogenesis. First, our curated replisome

marker suggested a deep origin of eukaryotes sister to the Heimdall clan (Heimdallarchaeia-Wukongarchaeia-Sifarchaeia); this is also consistent with the lack of evidence in any gene tree for Hodarchaeales except for the Hold PolB that is obviously prone to HGT. Second, we found that the gene copy expansion via HGT in the Asgard archaea is critical for its innovations. Third, HGT is also prevalent in the early evolution after the FECA-Heimdall split. The important involvement of HGT does come at a surprise given that replisome has been considered one of the most conserved apparatuses during evolution.

Finally, in terms of genome size of the archaeal ancestor, our data suggest that the early archaeal ancestor (after split from Heimdall clan) has a small genome. This stems from the following phenomenon: While this ancestral lineage imported replisome genes from other Asgard archaeal lineages, it did not maintain multiple copies of different origins. Instead, it lost its original copy after an import of a paralog, maintaining a relatively simple replisome, until massive duplication at the later stage of LECA.

These above implications are now summarized in the final 'Discussion' section.

3. Recombination vs. duplication - the paper uses "recombination" to mean "gene transfer across large phylogenetic distances" a number of times. I think this will confuse readers, who might typically think of recombination as something that happens between fairly closely related (and so homologous) chromosomes.

We thank you for pointing out the confusion that this term may incur. In the revised manuscript, we explicitly replaced "recombination" with the term "HGT" to describe "gene transfer across large phylogenetic distances". We also more explicitly explained the use in Line 342-344:

"An increase of replisome genes belonging to known gene families can arise through either a duplication of an existing gene, or through acquisition from another lineage via HGT, which would exhibit different topology in a phylogenetic tree (Fig. 4d)."

4. Lines 330-340 - very interesting to combine comparative genomics with experimental data, for example investigating the formation of Lokiarchaeum RFC complex. But need more context - does the complex do, what does this mean for the prokaryote-to-eukaryote transition.

We have now provided more extensive description for the functional division and hetero-multimerization of the RFC clamp loader during archaea-to-eukaryote transition in the second paragraph of introduction (Line 67-71, 74-76). We also moved the hypothetical evolutionary trajectory (previous Fig. 4) immediately next to the findings in Fig. 3i, with associated text under "RFC subunits differentiated sequentially during eukaryogenesis" (Line 306-319).

Minor comments

While the term "replisome" is defined in the first couple of sentences, I think the concept could benefit from being expanded a bit in the introduction, or at beginning of the results. For example, what is the list of genes that make up the replisome gene set, and how do their numbers change with

replisome "copy number"? The italicization of "rep" makes it look like a particular gene, but as far as I understand it, what is intended here is something like "the set of genes (including DNA polymerase subunits and various associated factors) that collectively carry out DNA replication" or similar.

Thank you for pointing out the lack of clarity. We have now specified the definition of the term "replisome" in the first paragraph (Line 51-53):

"The replisomes, namely the molecular machineries responsible for the above mechanical and enzymatic tasks"

We have also added the first section to specify that the gene families are collected from model organisms where functions have been determined. We also dropped the use of the italized *"rep"*, and instead used "replisome genes" throughout the manuscript.

Line 46 - "recombinants" - evolved/arose from recombination between…? (clarity/unpacking the concept).

We have rephrased it as follows (Line 63-65):

"their common ancestor likely arose through domain recombination between the canonical archaeal B-type and D-type polymerases named PolB and PolD".

69-71 - increase in complexity … mobile genetic elements … adaptive traits - not clear what is meant here. Maybe delete or revise to clarify the meaning? It seems the text may be referring to the somewhat fraught issue of which, if any, features of eukaryotic cells are specifically adaptive and contribute to eukaryotic "complexity", but it's too vague to really say much. For example, if the aim is to provide background for why the evolution of the replisome is interesting, I would maybe focus on that here instead.

We thank you for pointing this out. We have now removed these texts in the introduction and instead used more words to clarify the outstanding questions in replisome evolution. The introduction now ends with (Line 93-99):

"Systematically dissecting the molecular diversity in the newly expanded genetic repertoire of archaea and eukaryotes may provide new insights into the evolutionary history of supramolecular complexes comprising the extant eukaryotic replisomes. Conversely, resolving replisome evolution may help to clarify the sequence of events leading to the emergence of eukaryotes, as well as revealing general principles for how essential genetic modules innovate while conserving their core functions."

Lines 115,122 - not sure what "phenomenological" and related words mean here. Could they be deleted? In the first instance, there is simply a pattern, as far as I can see.

Indeed, it is merely a pattern! In the revised version, this section was shortened, and the full sentence was removed.

209 - "mostly maintained"

The typo is now corrected.

246 - "C-terminal"

The typo is now corrected.

Reviewer #2 (Remarks to the Author):

Feng et al 2025 provide an interesting perspective on the evolution of the DNA replication proteins in archaea and how they might have evolved during eukaryogenesis, focussing on the number, type, interactions, and structural arrangement of these genes and their encoded proteins in Asgard archaea, which are considered the closest microbial relatives of Eukarya. They identify some interesting resemblances of various Asgard archaeal replication proteins to modern eukaryotic ones, although the replisomes appear generally more typical of archaea as could be expected based on previous work placing Asgard archaea in the tree of life. Several of the predicted interacting replisome domains were verified by yeast two-hybrid interaction tests. The mix of these eukaryotic-like Asgard replication proteins in various lines of Asgards creates a complex picture of how they may have ended up in LECA. Discussion could also include any evidence for whether they may have all been present in the archaeal progenitor of LECA and then some were more recently lost in modern Asgard lineages to create the patterns seen today in Asgards, or how they may have come together from different archaeal sources in early eukaryotes, where possible.

We thank you for the advice on the discussion that puts our findings into context. In the revised manuscript, we now added Fig. 1 to show the components LECA's replisome and the various Asgard archaeal features related to it. We also added Fig. 6 to summarize the gene gain and gene loss of replisome leading to LECA.

My main points the authors should consider are:

Line 102. It is not clear here what homologs were searched for ('replisome genes') and why. For example, were the searches based on accepted comprehensive lists of experimentally verified components of the archaeal, yeast or other replisome, or what defined a rep gene?

Thank you for pointing out the lack of clarification in this regard. We have now added in the main text (line 102-104):
'In this study, we investigate 35 protein families involved in the DNA replication of model archaeal species belonging to Halobacteriales, Thermococcales and Sulfolobales, and/or model eukaryotes such as humans, animals, plants, and yeasts.'

How could the poorer knowledge of replication in some clades confound the scaling analysis – please briefly explain this and how it was dealt with.

Indeed, it is possible that certain proteins involved in DNA replication may be missed in lesser-characterized clades. This will to some extent influence the absolute number of replisome genes being quantified. We have now specified this scenario as well as tuning down the discussion on the scaling analyses.

First, now in the first introductory paragraph, we clarify that the essential steps of DNA replication in all domains of life are similar (Line 43-44), implicating that if we found the proteins responsible for all these steps, we most likely covered the main replisome components:

"DNA replication is an essential mechano-enzymatic process progressed in a similar fashion in all cellular life forms"

Second, we also clarify in the result section (line 104-107):
"These protein families are predicted to cover all universal steps of DNA replication across archaea and eukaryotes (as depicted in Fig. 1a), while unknown proteins that play regulatory roles in specific lineages may be unaccounted for."

Finally, in our paper, we primarily focused on the emergence of known eukaryotic protein complexes, particularly the functional division and hetero-multimerization of known proteins. Inaccuracy caused by potentially missing any additional components should not influence our overall conclusion.

Fig. 1b. It would appear important to report correlation coefficients, and variance (SD) in rep gene count for these? And then interpret the data critically when observing the slopes (potential errors in) and their differences (Fig. 1c).

Thank you for pointing out the importance of provide the statistical data for the fitting parameters. We have now provided all correlation coefficient, Standard Error and SD in replication gene counts in Supplementary Table 7. Standard Error of the Mean (SEM) is now also plotted in the updated Fig. 4c.

Line 125-6. Please provide statistical support for this noted difference in slope (if any?). It is not clear what the statistical significance means in Fig 1c – it is more important to compare differences in slope given the interpretation in the text, not just whether there is a significant slope or not.

We now provided statistical values for all groups in comparison to Asgard archaea in the Supplementary Table 7. In the text, we specifically compared the difference between Asgard and its sister group Thermoproteota, as well as Asgard and the Halobacteriota. We thus mark the significance specifically in the updated Fig. 4c as well. The p values related to the specific statements in the texts are also now specified.

Line 148. Could the presence of two homologous families (Cdc6 and Orc1) in eukaryotes be evidence against this statement (only 1 in LECA), given archaea do not appear to have distinct functional Orc1? If two were in the LECA, one may have developed as Orc1 in LECA?

Thank you for pointing this out. Due to a lack of strong support (after topological testing) for Cdc6 of any archaeal lineage to be most closely related to the eukaryotic Cdc6/Orc1, we decided to remove this section from the text.

Minor points:
Line 72. The meaning of this statement in the context of the paragraph is not very clear: "By contrast, the underlying evolutionary constraints or selective force for conserved essential functions such as DNA replication remain unclear." Should this include reference to selective force that increases

complexity?

Indeed, this sentence reads rather implicit. We have now removed the sentence. In the revised version, a related sentence is written as follows (Line 96-99):

"Conversely, resolving replisome evolution may help to clarify the sequence of events leading to the emergence of eukaryotes, as well as revealing general principles for how essential genetic modules innovate while conserving their core functions."

Line 74-75. Fix grammar throughout this sentence.
The line we removed due to structuring of the introduction section.

Line 113. Haloarchaeotes – is this a correct term? Revise.
Thank you for pointing it out. We have revised it to "Halobacteriota" throughout.

Line 115. How are 'small-genomed archaea' defined? Clarify please.
We meant archaea, which are small-genomed compared to the eukaryotes. This line has been removed due to restructuring of the introduction section.

Line 246. Clarify which Pol-delta1 here is the eukaryotic one.
We have named the Pol-delta1-related protein in Baldr as "Baldr PolB3" throughout, which we also depicted it in the Fig. 2a.

The whole paper would benefit from a thorough grammar and typographical edit, which would greatly improve readability.
Thank you for pointing it out. The manuscript has been revised substantially for readability.

Reviewer #3 (Remarks to the Author):

In this interesting article, Wu et al. combined phylogenetic, statistical, and biochemical approaches to investigate replisome evolution and the archaeal origin of replisome components in eukaryotes. Specifically, the authors discovered that ceratin Asgard archaea encode homologs of some eukaryotic rep genes that were previously not identified in prokaryotes. Indeed, it seems that different eukaryotic rep genes appear most closely related to homologs of distinct extant asgardarchaeal lineages, which is similar to observations on EPSs made in previous work that indicates that Asgardarchaeal genome evolution has been shaped by gene loss.

While the work is extensive and interesting, I found the article a bit difficult to read and I am not entirely sure about some of the conclusions drawn by the authors: e.g. “Our study thus suggest an emergence of eukaryotic replisome signatures prior to eukaryogenesis via distributed innovations across Asgard archaea, implicating an important role of horizontal gene transfer in the assemblage of LECA’s core genetic system”.

In turn, below I provide suggestions that could potentially help to improve the presentation of the work:

Major Comments

Restructuring of paragraphs
The presented data is interesting and relevant but I had the impression that it is not yet very coherently presented: e.g. Figure 4 shows different snippets but misses to convey a key message. I wonder whether the authors could consider restructuring their article and figures a bit. Please see below a suggestion of how the data could be presented in a way that’s potentially more easy to follow. It would also be great if the result and discussion sections could be carefully proof-read with regard to typos and grammar and language more broadly as various passages are currently difficult to understand.

For example, in my opinion, the results section about copy number and scaling is a bit lengthy and could be shortened with parts of it (including aspects shown in Figure 1) being moved to a supplementary material. Instead, you could perhaps provide a bit more information about the inventory of the various replication proteins of archaea and eukaryotes (and of bacteria if they have homologs) and the phylogenetic origins of eukaryotic homologs based on phylogenetics analyses. Potentially, you could consider following this by discussing the placement of eukaryotes relative to archaea based on these rep genes. This would set up the framework a bit more clearly. And from this broad look, you could then focus on certain aspects in more detail that best support the key message of the article.

Perhaps it would also ease readability if paragraphs would have headers related to the respective major findings presented (as an example, headers like “Traceable evolutionary history” are not very informative). Likewise, in terms of Figures, I think it would be helpful, if the overviews are presented in Figure 1 (e.g. elements of Figure 2 a and Figure 4a), with subsequent Figures focusing

on one functional aspect per Figure that support a certain observation based on a set of different analyses. Other results could perhaps be moved to the Supplements or Extended Data Figures? For example, the discussion of the syntheny analyses could potentially be discussed more concisely with details being moved to Supplements.

We highly appreciate your notes on the readability of the manuscript and your effort in helping us improve the flow of the paper. We have now made substantial adjustments to the paper accordingly and have texts checked carefully.

As suggested, we now added a general guide through the DNA replication process in the introduction section, with an accompanying new Figure 1. We also primed the main findings of the paper by summarizing the impact of gene copy number expansion in two aspects, namely functional division and heteromultimerization, as well as summarizing the eukaryotic-like features in Fig. 1c.

Also as suggested, we followed figure 1 with 4 figures each representing one main aspect of the paper. After communicating with microbiologists who are not necessarily familiar with phylogenetics, but who we do want to become our readers, we brought forth the original Fig. 3 and split it into the current Fig. 2 and 3; we also added illustrations to infer evolutionary implications. In this way, the major biological findings come first, while the more abstract analyses on gene copy and phylogenetics patterns came in Fig. 4 and 5. We hope that the current division of figures is an improvement from the previous one. With this, we also followed your suggestion to divide the original Figure 4 into Fig. 1, 3, and 5 to serve their different purposes in their respective contexts, instead of an assembly of snippets.

As suggested, we used more explicit wording in the headings. With these new headings, the readers should be able to grasp the flow and main conclusions of the paper.

As suggested, we have removed the paragraph on the synteny analyses. We now only used the rfc operons to support the evolutionary path in Fig. 3 but no longer referring to the change of gene synteny for primase.

Finally, it would be great if you could then come back to the big picture in light of all the observations made. How does this all fit into our broader view on eukaryogenesis? E.g. regarding origins of eukaryotic rep gene homologs, two major aspects could be discussed in the conclusion:
1. Genes inherited from Asgards
It seems that different replication machinery has distinct (Asgard) archaeal origins: ie. that FECA had all these elements that have been retained in some of the extant Asgard archaea but lost in others. This patterns seems similar to what was observed for ESP distributions: while Asgards together combine many ESPs, different ESPs have been retained in different Asgard lineages.
2. Components that have been acquired (from where?) or evolved during eukaryogenesis (e.g. by duplication) from FECA to LECA. What is the evidence for duplications versus HGT versus vertical inheritance with loss in Asgard archaea taking into account the various results including topology testing.

We thank you for the advice. We now added Fig. 6 that summarized the sequential emergence and gene flow that led to the replisome of the eukaryotic ancestor. The associated text is now in Line 460-476.

We now also substantially revised the 'Discussion' section to discuss the two main aspects (Line 478-530).

Topology testing

The authors suggest that "phylogenetics indicates that these increased gene copies did not arise through intra-lineage duplication" and that "these eukaryotic gene copies do not correspond directly to the individual copies that emerged in the Asgard archaea". What is the support for these inferences, i.e. the observed tree topologies and the relevant branches? Considering the uncertainty in single gene trees that are based on short alignments and are prone to comprise long-branch attraction artifacts (e.g. when distant paralogs that evolved under distinct evolutionary rates are included), it might be possible that the trees are poorly resolved or cannot distinguish among alternative placements (e.g. that (Asgard) archaeal paralogs cluster basal to respective eukaryotic homologs). To test this, it would be helpful if you could perform topology tests, i.e fix branches to the alternative hypothesis and test whether these can be rejected under the best fitting model?

We acknowledge the importance of topology testing and have now incorporated this analysis. The approximately unbiased (AU) test results enabled us to statistically reject certain topological hypotheses, with these cases were considered as potential HGT events. The complete set of tested hypotheses, along with their corresponding AU test p-values, is presented in Table S5, with detailed descriptions provided in the SI materials. All hypothesis trees have been deposited in FigShare folder under "(11) AU tests".

Data availability

All identified rep genes for each archaeal and bacterial taxon should be listed in a table providing accessions and annotations to verify that they are true homologs and to ensure reproducibility

We have uploaded all the related files onto FigShare (https://figshare.com/s/e7a2157c7391dd4df024), including our dataset, used HMM profiles, gene homologs, AU test hypothesis trees, single-gene and concatenated trees, used protein structures, and the original experimental gel images, etc. For each gene family, we provided a table with homologous sequences and EGGNOG mapper annotations found in the dataset. We also provided the score threshold values we used to select sequences for trees. We have also uploaded all the new MAGs to NCBI, and the accessions is provided in Table S2.

Minor comments:

- check for spelling errors, e.g. plural vs singular forms of verbs

Thank you for pointing it out. We have made a thorough check for verbs throughout the new draft.

- line 40: consider modifying "three billion years ago" since recent research indicates an earlier

divergence between Archaea and Bacteria

We have removed the use of the specific dating estimate in the revised version, and changed to 'since the ancient evolutionary split between bacteria and archaea' (Line 53).

- rephrase: "the gross conservation from archaea" for better clarity? E.g. "in comparison to Archaea…"

In the revised version, we have revised to 'In comparison to well-studied archaeal models…'.(Line 58)

- line 62: "Asgardarchaeota phylum". There are currently two validated taxonomic names for this phylum, potentially cite those in this context?

Tamarit D, et al. Description of Asgardarchaeum abyssi gen. nov. spec. nov., a novel species within the class Asgardarchaeia and phylum Asgardarchaeota in accordance with the SeqCode. Syst Appl Microbiol. 2024 Jul;47(4):126525. doi: 10.1016/j.syapm.2024.126525.

Imachi H, et al. Promethearchaeum syntrophicum gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage 'Asgard' archaea, and proposal of the new archaeal phylum Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. Int J Syst Evol Microbiol. 2024 Jul;74(7):006435. doi: 10.1099/ijsem.0.006435.

Thank you for pointing this out. We have now referred to both papers, and wrote (Line 86-87):

'Asgard archaea, which were recently proposed to be classified under Prometheoarchaeota or Asgardarchaeota phylum' .

Due to unsettled debate, we decided to use Asgard archaea throughout, and avoided the use of either phylum nomenclature.

- line 65: should this be past tense? "leading…"

In the revised manuscript, we now broke the two sentences, and dropped the use of 'leading'. Now it reads (Line 87):

'They encode various eukaryotic signature proteins (ESPs),…'

- line s86: "so far" without dash?

Revised.

- line 108: rep count (without s)

Revised.

- line 112: which have?

Revised.

- line 113 and 121: what do you mean by "bridge"? I am not sure that's the best term in this context

We wanted to use "bridge" to refer to an expectation Asgarding would show a property that is an intermediate between other archaea and eukaryotes. However, given the suggestion by you and other reviewers, we have substantially shortened the entire section regarding gene count scaling, where both sentences are now removed.

- line 131-134: how sure can you be that you picked up all rep genes? Could it be that less well investigated archaea encode rep proteins that have so far been missed? If so, could this affect the relations between rep genes and total proteome size?

Indeed, it is possible that certain proteins involved in DNA replication may be missed in lesser-characterized clades. This will to some extent influence the absolute number of replisome genes being quantified. We have now specified this scenario as well as tuning down the discussion on the scaling analyses.

First, now in the first introductory paragraph, we clarify that the essential steps of DNA replication in all domains of life are similar (Line 43-44), implicating that if we found the proteins responsible for all these steps, we most likely covered the main replisome components:
"DNA replication is an essential mechano-enzymatic process progressed in a similar fashion in all cellular life forms"

Second, we also clarify in the result section (line 104-107):
"These protein families are predicted to cover all universal steps of DNA replication across archaea and eukaryotes (as depicted in Fig. 1a), while unknown proteins that play regulatory roles in specific lineages may be unaccounted for."

Finally, in our paper, we primarily focused on the emergence of known eukaryotic protein complexes, particularly the functional division and hetero-multimerization of known proteins. Inaccuracy caused by potentially missing any additional components should not influence our overall conclusion.

- perhaps check that you use consistent taxonomy, e.g. what do you mean by Haloarchaeotes? I think, in GTDB the group encompassing Haloarchaea, Methanosarcinales etc have been referred to as Halobacteriota? Or which taxonomy do you use (check for consistency)

Thank you for pointing it out, we have now corrected it to 'Halobacteriota' throughout.

- Figure legend 1, line 2. Archaea already include Asgard archaea so perhaps its not needed to mention those separate?

Thank you for pointing out the confusion. The green fitted line represented all archaea (including Asgard archaea), while the black line was specific to Asgard archaea alone. However, we realized that using a distinct point color for Asgard archaea might cause confusion. To improve clarity, we have now changed the Asgard archaea points to green (matching other archaea) and added a black circle outline for distinction. Please see the new Fig. 4a.

- Figure legend 1: what do you mean by monophyletic versus paraphylyetic? At which level was monophyly required? E.g. some of the trees labeled as paraphyletic still seem to have lineages that are monophyletic, e.g. the Thors and Lokis in C? Or do I misunderstand this?

We meant to use “monophyletic” to describe whether the two gene copies from the same clade are sister to each other. It is indeed unclear as the definition of ‘clades’ is also context dependent. We have now removed the use of ‘monophyletic’ and ‘paraphyletic’. Instead, we directly define ‘duplication’ and ‘HGT’ at any evolutionary stage as indicated by the phylogenetic tree, which were verified via AU testing.

- Figure legend 2, i.e. line 172: do you mean “replisome single-copy gene trees” or just one tree?
Revised. We have now used ‘replisome tree’ to refer to our concatenated replisome tree.

- Figure legend 2, i.e. line 187: do you mean “recombination” or “horisontal gene transfer”?
Revised. We now use HGT here and throughout the text to replace ‘recombination’, as the latter could lead to confusion with DNA recombination. (See new Fig. 4d and e)

- line 195: with “Close examination of individual rep families” do you mean the close examination of the phylogenies of individual rep families? Or something else?

Yes. Indeed, we meant the detailed gene copy number and phylogenetics. Due to restructuring, this line we removed for brevity. The current related line reads (Line 353):
“Combined, these approaches resolved the origins of gene copy expansion in various lineages for a subset of gene families”.

- line 242: “structural innovations bridge the archaeal and eukaryotic replisome”. This is an example of a sentence that has multiple meanings: e.g. in its literal sense, this sentence suggest there is a (physical) bridge between the archaeal and eukaryotic replisome established by structural innovations. I feel like it would be better if you could try to be more precise in language use to ease understanding.

Thank you for pointing it out. We have now revised the line (now Line 131) to
‘Detailed structural comparisons further identified archaea-encoded proteins with distinct structural features characteristic to eukaryotes’.
We have also edited the texts throughout to make the language use more precise.

- pay attention to use accurate taxonomic terms rather than abbreviations/slang, e.g. for Baldrarchaeia

Thank you for pointing it out. We now changed to full taxonomic terms throughout the text and figures.

- line 275: “relative subunit sizes”: what does this mean?

Thank you for pointing out the lack of clarity. We meant the relative size different between PriL and PriS mentioned in the text above it. Now we have used more explicit description as follows (Line 197-199):
‘Unlike primase subunits encoded by other archaeal clades, where PriS is mostly larger than PriL,

the two gene products in the Heimdall clan show size features seen in eukaryotic primase subunits, with PriL larger than PriS.'

- Line 361: what do you mean "chimeric origin" and "chimerism"? Chimerism of what?

Thank you for pointing out the lack of clarification. 'Chimeric origin' has been used recent papers to refer to gene origins from a mixed number of lineages, which was used in titles such as:
'Santana-Molina et al *Chimeric origins and dynamic evolution of central carbon metabolism in eukaryotes* 2025 Nature Eco & Evo."
While it also suits our case, we realized that there was only limited use of it in our context, so in the revised manuscript, we replaced it with more explicit description as follows (Line 471-472):
'*various structural innovations sequentially occurred in different Asgard archaeal lineages were then horizontally transferred into the progenies of FECA*'

- what does this mean: "This coincides with their domain changes as described above"?

We meant that the synteny change at the emergence of the Heimdall clan coincides with the motif changes of Primase. As you pointed out below, such an inference from synteny change may be too speculative. Hence, we have removed this section.

- the section on gene organisation patterns seems very speculative to me. I think its important to keep in mind that extant Asgard archaea do not reflect the proteome and gene arrangements that were present in the archaeal ancestor of eukaryotes. Furthermore, most of the asgard genomes are MAGs and don't have complete genomes which could make it harder to identify synteny. More importantly however, I think ancestral genome inferences and the reconstruction of the proteome of the lineage leading to eukaryotes would be needed to investigate the scenarios. Also the wording in this section is often difficult to understand: e.g. "obligately, differentially lost"… And what do we know about the fitness cost?

Thank you for pointing out the genome organization that may differ at the early stage of eukaryogenesis and the influence of incomplete MAGs on the judgement of synteny change. We have now removed the claim on relation between the emergence of eukaryotic-like primase and gene arrangement. However, we do feel that gene synteny clarifies co-inheritance and co-evolution in specific contexts.

First, while MAGs can be fragmented, our large collection of the asgard archaea included high quality genomes where gene pairs of interests are not disrupted. Also, contigs containing DNA replication genes are often found in conserved regions of the genomes which are also less prone to fragmentation during assembly. To minimize potential artifacts from fragmented MAGs in our synteny analyses, we implemented stringent quality filters by using high-quality genomes and only considering genes having neighbors on the contigs.

Second, while there is great uncertainty regarding the gene organization and genome structure of FECA, we argue that it should not affect the proposals above. Our proposal of gene arrangements

is based on the lineages either formed pre-FECA, or in an Asgard archaea lineage outside the FECA clade. In both cases, the extrapolated ancestral lineages remained prokaryotic. Any changes in gene arrangements after the split of FECA from the Asgard archaeal ancestor, or after the HGT of gene operons into the FECA progeny, does not influence our proposals. Hence, based on the current dataset, we made the following mentions by combining gene synteny, phylogenetics, and structural comparisons:

1. The primase *priS-priL* gene pair have likely been vertically transmitted from the Asgard archaeal ancestor (Heimdall clan) into the FECA. (Line 205)
2. The *rfcS3-rfcS2-rfcL* operon has likely been horizontally transferred from Lokiarchaeales into the progeny of the FECA (Fig. 3g, Line 310).

- line 436: rephrase "bridge archaea and eukaryotes."

We have rephrased. Now the related sentence reads (Line 480-481):
"…revealed the compositional diversity of replisomes across archaea and eukaryotes, implicating a dynamic evolutionary path leading to the complex replisome of LECA."

- line 450: microbial evolution is known to be shaped by HGT so this is not very surprising?

While we acknowledge that HGT is a mechanism that generates genetic diversity in microbial evolution, we do note that the extent to which it infiltrates a core machinery such as the replisome is not as trivial. We still think that the extent it occurred as found in our study is quite surprising. Most notably, we found that the HGT between Asgard archaea and FECA is strong, while there appear to be a strong genetic barrier between the FECA and other archaeal lineages.

- line 457-8: you suggested: "implicated that HGT-based gene recombination was essential at the early stage of replisome formation during eukaryogenesis". While this might be possible, the data rather shows the ongoing evolution in Asgardarchaeota after their divergence from the lineage leading to eukaryotes and not what happened during eukaryogenesis? Since you did not infer the proteome set in the archaeal lineage leading to eukaryotes but in extant Asgardarchaea?

Thank you so much for pointing out the lack of clarification between the two different phases. In fact, we meant both. We hope that the processes are better clarified in Fig. 5 and 6, along with associated texts.

- line 460: which origin do you mean here?

We meant the single gene origin that became duplicated into multiple copies in the LECA. The updated text reads (Line 511-514):
"However, gene paralogs in the LECA appeared to have almost exclusively arose through duplication of a single gene origin, suggesting that the early eukaryotic ancestor (after branching from the Heimdall clan) had a highly constrained gene content with little tolerance for gene copy expansion."

- what is the reason to use LG+C60+F+G4 for the single gene trees while C10 was used for the concatenated tree? I would recommend using C60 also for the concatenated tree as the alignment is longer and it therefore makes more sense to have more categories. C60 might be too much on single gene alignments depending on the alignment length. Perhaps the authors could perform model test to compare C10-C60 and choose the best fitting model. (in addition to performing topolgy testing as suggested above)

We feel that this was a misunderstanding. We only used C10 for the initial classification of the Asgard archaeal clades. The final phylogenomic tree was already carried out using LG+C60+F+G4. It is clarified under the Method section 'Phylogeny of concatenated replisome proteins'.

- Data availability: it would be helpful if the authors can also make alignments files available (both before and after trimming). Furthermore, it would be valuable if all workflows and scripts for the various analyses were made available

All the alignments or original sequences now have been provided in FigShare, including those for HMM profile construction and trees. We also made a workflow about our analyses with used commander lines. The custom scripts were also uploaded onto FigShare.

- Figure 2B: how reliable are those topologies and the placement of the eukaryotic lineages relative to asgardarchaea? I think it would be very umportant to show support values and sister lineages more clearly as many of these rep genes do not seem to group eukaryotic homologs with Asgard: e.g. its hard to see the differences of the circles indicating bootstraps, maybe color-coding would be better. Considering that the placement of the eukaryotic branch is key to the interpretation of the work, I feel the single gene trees should be presented more accurately.

Thank you for emphasizing the clarity for the bootstrap values. Because we have used many colors to indicate taxonomy in the tree, it became too complex when more colors are used on the branches as well. We thus opted to only use circles to show bootstrap values above 75, and distinguish the moderate level and high level of bootstraps in light blue and light grey. We now also added the bootstrap number near the eukaryotic branch. Please see the updated Fig. 5. The original tree files can be found in Figshare.

- Table S1: It would be helpful if you can add three columns to inform about genome/MAG completeness, contamination and redundancy for each of the archaeal taxa analysed

For the 436 Asgard archaea, the information has been provided in Table S2 and the genomes have been deposited to FigShare. For the representative species of Asgard archaea, other archaea, and eukaryotes, the contigs, completeness, contamination, and taxonomy were saved in Table S3. Since the estimates of contamination and redundancy are interchangeable using the current analytical tools, we only used contamination in the column header.

- for eukaryotes: how do you determine copy number? Are these all complete genomes or also

transcriptomes? Does this include different splice variants?

We recognized that the EukProt dataset includes many transcriptomes and ESTs, and the alternative splicing isoforms were retained without clear distinction. While we utilized this resource for single-gene tree construction for the purpose of covering sequence diversity, we adopted a more stringent approach for gene copy number analysis (now moved to the current Fig. 4). Specifically, we separately downloaded the complete genomes of model eukaryotic species and removed alternative splicing variants by gene locations in the corresponding GFF files. The complete list of representative eukaryotes, along with their download sources and associated references, is provided in Table S3.

Besides the Methods and supplementary table, we clarified in the main text (Line 114-117):
"For eukaryotes, we leveraged the EukProt3 database11 to cover broad proteome diversity for phylogenetics, while using representative species with complete genomes for gene copy number analyses."

- Figure S3: did you verify that your search did not pick up false positives or exclude false positives in case those were obtained? Furthermore, I think it would be important if you can provide IPR/PFAM/KO/COG annotations for all the rep proteins of all archaea for verification and reproducibility

For all homology searches, we employed EGGNOG mapper annotations to verify gene identity and establish appropriate score cutoffs. Given the high conservation of DNA replication genes, the false positive results present as the low-scoring hits, which would be excluded by the threshold. The complete dataset, including all identified homologs (with IPR/PFAM/KO/COG annotations), score cutoffs, and final sequences used for phylogenetic reconstruction, has been deposited in FigShare "(8)(9) Homologs".

- Table S3: can you provide a link to a database and protein accessions?

We have updated the supplementary table (now S4), and all the used protein structures are now shared via FigShare "(12) Used protein structures ".

Commented [YF1]: 在这里，原来可能写错了

- can you provide a S-Table listing all the taxa that were used in phylogenetic analyses of rep genes?

In our study we used two datasets, (1) a Main dataset for single-gene trees, including 463 Asgard archaea (Table S2), all the non-Asgard archaea of GTDA archaea r207 (https://data.gtdb.ecogenomic.org/releases/release207/), and eukaryotes from EukProt v3 (https://evocellbio.com/eukprot/). For genes having homologs in bacteria, the bacterial dataset of GTDA r207 were also used. (2) For gene number count and the concatenated tree, we selected high quality representatives, the list was provided in Table S3.

- Did you add Bacteria whenever homologs were present? I think that would be helpful for rooting and to rule out that certain components may derive from Bacteria. If this is what has been done,

perhaps mention it in the methods (unless I missed this). In general, it would be good if you can provide a table with all taxa and all rep genes as mentioned above.

We included bacterial homologs as described in Methods. Since archaeal/eukaryotic DNA replication is less similar to bacteria, many analyzed genes lack bacterial counterparts. That might the reason they feel like absent in this study. Only PolB, RnhB, Ligase, and topoisomerases have bacterial homologs, which we visualized in yellow in Figs. S4-S6. However, these bacterial versions did not appear to influence the evolution leading to the eukaryotic replisome (they likely related to the DNA processes of mitochondrion or plastid). We thus wrote in the text (Line 365-366):
'Note that HGT from archaea to bacteria, as well as from bacteria into specific eukaryotes also occur, but only in limited cases (see Supplementary text)'.

- you mention "Cdc45 appear to be phylogenetically related to RecJ/Cdc45 proteins encoded by the Lokiarchaeia": what is the support for this? The tree topology looks poorly resolved.

The tree is shown in Fig. S5, and we have carried out AU testing to support that the specific clade of Cdc45 encoded by Lokiarchaeia is divergent from the conserved Lokiarchaeia copy, and is basal to the eukaryotic branch.

AU testing further showed that at least Lokiarchaeia, Thorarchaeia, and Hodarchaeales showed Cdc45 gene acquisition via HGT, but not duplication. However, because Cdc45 have different number of copies beyond the two copies we distinguished, even within the same order or family, which made it very difficult to depict in Fig. 4e. Hence we color coded it in grey in Fig. 4e.

The specific sentence referenced above is now removed from the paper for brevity. Instead, the mention of Cdc45 was done together with several other genes that showed similar patterns. Such as Line 440-442:
"Leveraging the detailed gene-level analysis, we precluded PolB, Cdc45, Top3, RfcS, and Fen1, due to their dynamic gene gain and loss that involved HGT-derived contribution to LECA"

- Fig S8b: are the B6 Aenigma 1 sequences also contaminations from Baldr?
The present of B6 Aenigma 1 in Aenigmatarchaeota is more common and it is not a contamination.

Reviewer #3 (Remarks on code availability):

I briefly checked the repository and it seems it mostly provides supplementary data

We have added a Code Availability section in the manuscript and the used custom scripts were saved in FigShare "Used_scripts". The used conman lines for the softwares were provided in FigShare "WorkFlow.docx".

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

Feng and colleagues have substantially improved the readability and clarity of the manuscript, in line with the suggestions from all of the reviewers in the previous round. This is now an interesting contribution that, in my view, merits publication.

We thank you for your help in improving the paper and approving the revised manuscript.

Reviewer #2 (Remarks to the Author):

The authors have carefully addressed the previous comments and revised their manuscript giving reason, and it's significantly improved technically and in organisation. Below are some points on the new aspects of the paper:
We thank you for your help in improving the paper and your positive comments on the revised manuscript.

Line 32. For the abstract, I believe this sentence could be more specific and clearer or may not be needed. For example, how did the distributed innovations collectively drive protein complexes to 'evolve functional divisions?' it's not clear to me, at this stage, how this could have driven the evolution more than any other model, as this seems to imply. i.e. Why would a sub-complex innovation in one line of Asgards be different or special in driving evolution of complexity, compared to another potential model where all of the innovations occurred sequentially (vertically) in one line leading to LUCA?
And what is meant by 'functional divisions' here? Duplication and diversification? Clarify, otherwise it might initially sound like one protein evolving to do multiple functions, which is not what the authors are saying with the paper.
Should a sentence like this be at the end of the abstract instead (if even needed given the rest of the abstract)?
Thank you for pointing out the need of clarification on this sentence. This line was meant to point out the origin of two features that is most prominent in eukaryotic replisome. However, indeed we realize that the abstract already had multiple summary-like sentences, we have removed this line to make it more concise.

Regarding 'functional division', we borrowed from the term 'division of labor' in ecology, where two different groups play complementary roles to together complete a task. We still think it cannot be simply replaced by other terms such as 'diversification'. We now marked this term in the main Fig. 1 to assist understanding.

Line 33. Grammar – 'expansions'

corrected

Line 36. Grammar/meaning – 'be sustained by' (?)

modified into 'be always transmitted vertically' for clarity.

Line 110. and throughout. Make consistent labelling of figs S1, 2 etc, or 'extended data..'
Thank you for pointing it out, we have now changed all the Fig. Sx and Table Sx into Extended Data Fig. x and Supplementary Table x, respectively.

Line. 269. One/single same meaning – avoid duplication of meaning.
modified into "a single'

Fig. 3 and elsewhere. Much of the integrated text is extremely tiny. This should be enlarged so at least readable. Zooming can help but there is space to enlarge this text.
We have enlarged the small texts in figures.

Fig 4C. How were the replicate data (n) obtained (enabling calculation of an SEM and p-values from t-tests)? Indicate in the legend. The use of t-tests appears unclear here where there is no obvious way to obtain replicates of the regressions unless random nonredundant subsamples of the data could be used for separate regressions and their slopes taken as the replicates. But a common test for different slopes of two regression datasets would be an analysis of covariance. The results of both approaches could be noted, if appropriate.

We apologize for the confusion. The errorbars refers to the standard error that is more generally called SER (standard error of regression, not SEM), that was used to indicate the uncertainty of the fitted slope. The sample size n is now indicated in the legend. With the SER of the slopes, it was sufficient to carry out two-sample t-test. The approach used for the calculation is in the Method section 'curve fitting and statistical analyses'.

Line. 391. Grammar – 'includes'
corrected. Now it reads:

'The expanded repertoire of HGT-derived gene gain in the Asgard archaea includes the previously designated single-copy essential genes'

Line 396. Clarify: 'more dynamic evolutionary path than previously assumed". What is defined as dynamic here, and who assumed?
We are sorry that it may be not clear here. "Dynamic" and "previously assumed" are based on a conventional hypothesis. It is generally assumed that the central genetics is highly conserved, and are thus used as phylogenomics markers. However, we do not mean to negatively refer to a specific paper, and thus we now rephrase (Line 513):

'implicating diverse evolutionary trajectories leading to the orchestration of LECA's complex replisome'

The latter two sections of the 'Results', especially the last one, read more like discussion. Consider placing the last section discussing Fig 6 at the start of the discussion?
We thank you for the suggestion. The section 'Curated replisome markers position eukaryotes sister to the Heimdall clan' is based on phylogenetic results, so we incline to keep this section in the results. While the section 'Serial innovations, HGT, and gene loss drove the formation of a prototypical eukaryotic replisome' described the hypothetical model, we feel that it does set itself apart from the remaining discussion section, we feel that it might benefit from a header for the readers, and will leave it for now and will ask for copy editor's advice.

Reviewer #3 (Remarks to the Author):

The authors have addressed the comments raised by the reviewers and restructured the article according to suggestions. I think this has considerably improved the clarity and readability of the article. The authors have also put the results into a broader perspective on eukaryotic origins.
We thank you for the generous effort in helping us improve the flow and clarity of the paper and are glad that you find it an improvement. We also thank you for the advice of adding AU testing to support the topologies of the phylogenetic trees.

I very much appreciate the extensive work of the authors on this manuscript, yet I am concerned that some of the results are over-interpreted (e.g. regarding the hypothesis that the eukaryotic replication machinery has origins from different Asgard archaeal lineages) and that presented scenarios/hypotheses (e.g. in Fig 6) are not sufficiently supported by the current analyses. Single gene trees are prone to phylogenetic artifacts and the topologies presented do not always align with the proposed scenario, see for example the sections on the origin of Pol α/ Pol δ/ Pol ζ. While the author indicate that they have performed AU tests, those are not discussed in some of the relevant sections so that the reliability of certain placements is not yet clear to me (I provide some examples below to indicate where results from AU test would be valuable to be mentioned explicitly). Furthermore, the presence of homologs in the closest archaeal sister lineage (i.e. Heimdalls or a subclade within) followed by subsequent loss in currently sampled representatives (and retention in other lineages such as RFCs in Lokis) cannot be excluded. I would recommend the authors to to use a statistical framework for reconciling single gene trees with the species tree and then present results more objectively, potentially discussing different scenarios that could explain the data. See more specific comments below.

We thank you for the pointing out the need for further clarification for several aspects of the paper, especially on the detailed results of the AU testing that we carried out. We also very much appreciate your suggestion to state the possibility of early emergence and differential loss more explicitly as an alternative to HGT in the text; we have now stated the possibility in two prominent spots:

'It is, however, also difficult to completely exclude a scenario where they arose through an even earlier duplication/acquisition followed by a loss in most lineages…….'

'Fen1-2 variant may have been acquired by the Asgard archaeal ancestor in additional to the canonical Fen1-1 and ultimately replacing the role of latter in the FECA (as depicted in Fig. 6), or a result of HGT from the Lokiarchaeia-Thorarchaeia common ancestor to FECA.'

We did not intentionally bias our interpretation towards HGT as we did acknowledge that some events occurred in the Heimdall clan ancestor, and we noted that the evolutionary paths of several genes are ambiguous without clear motif signatures and phylogenetic robustness. Regarding the Pol α/ Pol δ/ Pol ζ, it is a matter of subunit evolution and complex composition, which we explain below. Regarding RFC, we have carried out several AU tests on the phylogenetic trees, we also now added more AU tests as you indicated, which rejected Heimdall clan or Hod as ancestors; we also note that we highlighted RFC evolution in the paper also based on structural predictions and motif analyses.

As you have noted that single gene trees can be prone to artefacts, which was why we carried out extensive sequence alignments to find motif features, structural novelties, as well as experiments to test protein interactions. Throughout the paper, we in fact refrained from making conclusions on complexes that did not show clear structural or compositional innovations in archaeal clades that are predicted to be basal to eukaryotic proteins. While the overall picture is more complex and sometimes written to fit the interests of experimentalists, we hope that you also empathize with our effort to try to understand evolution using information beyond phylogenetics.

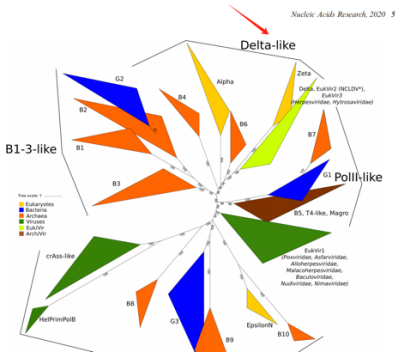
Main comments

I have some concerns regarding the interpretation of the results presented in this paragraph "Baldrarchaeia encodes a Polδ-like protein complex" and corresponding figures including Fig S6:

Reading your questions below on this section made us realize that we have not explain properly to the readers regarding the difference between the evolutionary relations between the major catalytic subunits - Pol α1/ Pol δ1/ Pol ζ1 and the compositional similarity (Pol α/ Pol δ/ Pol ζ). This, combined with insufficient descriptions of our previous AU tests beyond our supplementary table and dataset, can indeed lead to misunderstanding. We highly appreciate your detailed examination of the phylogenetic tree and your effort to help us improve clarity.

Before we address individual concerns below, we would like to first explain the rationale and literature basis for referring to Baldrarchaeia PolB3-DP1 complex as Polδ-like, instead of Pol α/ Pol δ/ Pol ζ-like:

First, the use of Polδ-like to refer to the entire clade of this eukaryotic DNA pol, encompassing Pol α1/ Pol δ1/ Pol ζ1 has been seen in literature – for example, in the most recent comprehensive analyses across the PolB family (Kazlauskas et al, NAR 2020, https://doi.org/10.1093/nar/gkaa760, see Figure below). In the current draft, we now avoided such an inheritance from literature, and instead using Polα1/δ1-like for this phylogenetic clade. However, Polδ1 does differ from Pol α1 in that the former retained the exonuclease activity of the ancestral PolB, which was lost by Pol α1. Hence the common ancestors of Polδ1 and Pol α1 have been assumed to be more Polδ1-like, and we suggested that the Baldrarchaeia PolB3 is also more Polδ1-like.



Second, Pol ζ1 (or named REV3L) is a subunit of the low-fidelity DNA synthase involved in the replication of damaged DNA. It was not necessarily considered an essential component of the replisome. Hence, we opted to only mention Pol ζ in the main text briefly in the text, which now reads:

"Phylogenetic analysis using the PolB domain alone also inferred a shared evolutionary origin between Baldrarchaeial PolB3 and Polα1/δ1, as well as the translesion DNA synthase subunit Polζ1, although topological testing cannot distinguish the order of emergence among these clades (Supplementary Table 5 and Supplementary text)." Line 148-151

Third, we refer to the Baldrarchaeia PolB3-DP1 complex as Polδ-like, rather than Pol α/δ-like, due to the compositional and probable functional similarity. Pol α has a very specific functional role and compositional makeup in eukaryotes. Compositionally, Polα1/α2 stably interacts with the eukaryotic PriL/PriS to form a heterotetramer. This was not seen for the Baldrarchaeia PolB3, as we suggest now in Fig. 2c with Y2H experiment, which we newly added. Functionally, it synthesizes a short DNA primer immediately after the RNA primer synthesized by the PriL/PriS subunits, a role so far not reported for any known archaeal DNA pol. On the other hand, Polδ synthesizes the lagging strand DNA, which is a necessary process in all domains of life. Additionally, Polδ plays the canonical function of lagging strand synthesis, and only has two core subunits Polδ1/ Polδ2 (some model eukaryotes have two additional small subunits but this is not universal), which is homologous to Baldrarchaeia PolB3/DP1. Hence, we referred to the Baldrarchaeia complex as Polδ-like.

Combined, we added the following text in Line 169-176:

Eukaryotes Polα is a stable hetero-tetrameric complex that also comprises two primase subunits, which synthesizes a short RNA sequence prior to the DNA primer synthesis by Polα1. Unlike the strong Polα1CTD-PriL interaction in budding yeast, our Y2H assay did not show an interaction between its canonical archaeal primase and PolB3CTD (Fig. 2c). In addition, Polδ inherited the exonuclease (exo) activity from both PolB and DP1 for proofreading, whereas Polα lost exo in both subunits{Pavlov, 2006 #188}. Overall, our data suggest that Baldrarchaeia encode a DNA polymerase that is most similar to the eukaryotic Polδ,……'

- The current tree topology (though, I do not know what the support is for the key nodes) would be most consistent with a scenario in which bacterial homologs were at the origin of eukaryotic Pol α/ Pol δ/ Pol ζ. Upon duplication, early eukaryotes would then have transferred a homolog of an ancestral Pol δ/ Pol ζ to Baldrarchaeia and Aenigmatarchaeota while a second duplication would have led to the origin of Pol δ and Pol ζ, respectively. Which bacteria are at the base of eukaryotic Pol α/ Pol δ/ Pol ζ? (Q1)

We have now added the support values for the nodes. Additionally, we also carried out AU testing for the multiple scenarios you mention here and below.

The bacterial PolB homolog that appear basal to all above mentioned homologs is in fact a singleton, encoded by a MAG of Acidiferrobacter sp. (GCA_002694065). The bootstrap is 75 for this branch, and AU test cannot reject the placement of this branch at another location away from the base (p value 0.5). Additionally, we did not find any domain resemblance of this sequence to any of the eukaryotic DNA pols. Based on the above observations, we would not place any speculations on this branch. We highly appreciate the reviewer for pointing this out; we now also added the above note in the SI text under the PolB section. In the main text, we wrote:

"Phylogenetic analysis using the PolB domain alone also inferred a shared evolutionary origin between Baldrarchaeial PolB3 and Polα1/δ1, as well as the translesion DNA synthase subunit Polζ1, although topological testing cannot distinguish the order of emergence among these clades (Supplementary Table 5 and Supplementary text)." Line 148-151

- What is the reason to suggest that the Baldrarchaeia and Aenigmatarchaeota homologs are more closely related to Pol δ? (Q2) Based on the tree topopogy, they are basal to a monophyletic clade composed of Pol δ and Pol ζ and the node could be rotated. E.g. you mention "this fusion resembles the catalytic subunit of eukaryotic Polδ named Polδ1." Can you exclude resemblance to Pol α/ Pol ζ? (Q3) The alignment illusration in S6a suggests that the Baldrarchaeia and Aenigmatarchaeota homologs have retained some indels similar to Pol α (which would be compatible with the transfer to Bald and Ainigma scenario)? (Q4)

(Q2) Regarding the use of Pol δ to refer to the specific protein complex, we have explained above.

(Q3, Q4) Indeed you are correct. From the large subunit phylogeny alone, we cannot distinguish which one is closer. We have now also specifically use Pol α1/δ1-like to refer to the large subunit PolB3. However, as described above, the retainment of exo activity from PolB by δ1 do give it more resemblance to the ancestral form. Modified texts are indicated above.

- Could you reject using AU test that the Heimdall clade as opposed to the bacteria is basal to the eukaryotic Pol α/ Pol δ/ Pol ζ including the Baldrarchaeia and Aenigmatarchaeota homologs (that could have been acquired by these archaea from eukaryotes later as suggested above)? (Q5)

While we used the PolB section to perform phylogenetic analysis, so as to make an overall classification of PolBs and to suggest an overall diversity, we note that Heimdall clade do not at all encode any PolBs that has the same domain composition as the eukaryotic Pol α1/ Pol δ1/ Pol ζ1. We have also experimentally shown that Heimdall DP1 and DP2 forms a canonical PolD (Fig. 2b), which is fundamentally different. On ther other hand, Baldrarchaeia and Aenigmatarchaeota have lost DP2. Based on these ensembles of evidence, we do not see a rationale to consider Heimdallarchaeial PolBs, which does not contain $DP2_{CTD}$, to be basal to eukaryotic DNA polymerase subunits in this context.

- Can you reject monophyly of eukaryotic Pol α/ Pol δ/ Pol ζ? (Q6) What is the support of Baldrarchaeia and Aenigmatarchaeota homologs placing inside of this cluster? (Q7) E.g. can you exclude the topology of DP1 being reflected by DP2? (Q8)

(Q6) We have carried out AU testing placing Baldrarchaeia PolB3/Aenigmatarchaeota PolB basal to Pol α1/ Pol δ1/ Pol ζ1, a scenario that cannot be rejected (p value 0.115). Hence, indeed we cannot reject that Polα1/Polδ1/Polζ1 could be monophyletic.

(Q7) The bootstrap support of this clade at the tree is 98 (now added in Fig. S7).

(Q8) In the DP1 tree, Baldrarchaeal and Aenigmatararchaeota DP1 is outside but close to the base of the eukaryotic branches (Baldrarchaea + Aenigmatararchaeota clade as the sister group of Polα1/δ1/ζ1, AU test p value = 0.131). Hence, combining the AU testing of PolB and the DP1 tree, we would speculate that the Baldrarchaeial and Aenigmatararchaeota PolB-$DP2_{CTD}$-DP1 complex is most likely ancestral to all Pol α/ Pol δ/ Pol ζ complexes. Additionally, we propose that it would be most parsimonious that Baldrarchaeial PolB3 is more ancestral to Aenigmatararchaeota PolB because the CTD of the former is homologous to full $DP2_{CTD}$, whereas CTD of the latter appeared to be a degenerate form of $DP2_{CTD}$.

- Taking all this into account, I am not convinced by the scenario presented in Fig. 2g. and the section "Functional division of Eukaryotic DNA polymerases emerged sequentially", i.e. the origin of eukaryotic Pol α/ Pol δ/ Pol ζ

In summary to the above contexts, we agree with you that we cannot resolve which clade among the Pol α1/ Pol δ1/ Pol ζ1 subunits came first purely based phylogenetics. However, our Pol δ-like, Pol α-like or Pol ε-like refers to their protein domain composition and subunit makeup, not the major subunit phylogeny alone. The Fig. 2g and associated section was also meant to emphasize the emergence of these complexes and functional divisions.

This section was added during the last revision in a response to your previous comment that our results were not placed into the context of eukaryogenesis. While we acknowledge that there are various aspects not being resolved phylogenetically or experimentally, we still think that this figure and associated section is faithful to literature and serves as a useful hypothetical for future studies, for either further experimental validation or debate. We have now modified the annotations in the figure where we omitted the use of 'Polδ-like' so as to avoid confusion and also wrote in the caption that it is a speculative model (Line 239). Note that we no longer indicate which lineage these features come from, such that this figure only focuses on the order of events that we speculate to have occurred from the structural and functional point of view.

In my personal discussion with eukaryotic biochemists who work on DNA polymerases, some have assumed that Polε were the earliest to emerge because it has the most domains, and the remaining ones would be a product of domain loss. This reflected the importance of showing such a schematic where the prokaryotic molecular ancestors are taken into account.

- it would be great if you can show support values (even if as dots indicating support above 80 and/or 90) for trees in the supplementary material. It is otherwise difficult to understand the level of support for the different placements of eukaryotic homologs relative to archaeal and bacterial homologs.
Thank you for the suggestion. We have now labeled all the bootstrap values for the nodes where the eukaryotic clades branch from. The phylogenetic trees are all provided in the supplementary material, where all bootstrap values are shown.

- The tree in Figure 3a is difficult to read, so I wasn't able to see the support of Loki as opposed to Heimdall homologs being basal to eukaryotic homologs. Perhaps mention if there was any AU-testing done in this context and what it suggests.

We have now added bootstrap values at the root positions. Based on your suggestion, AU tests are also updated. we conducted AU tests on both RfcL and RfcS trees by hypothesized that 1) move Lokiarchaeales into other Asgard archaea and 2) move the Heimdall clan as the sister group of eukaryotes. All the hypotheses were rejected (p value < 0.05), supporting Lokiarchaeales are closest to eukaryotes on all RFC subunits (Table S5).

We now added line 268-270 regarding AU testing:

"Topological testing rejected alternative scenarios where eukaryotic RFC subunits branched outside the Asgard archaea clade or sister to the Heimdall clan RFCs (Supplementary Table 5 and Supplementary text)."

We would like to note that the evolutionary relations between the phylogenetically suggested archaeal RfcS and the eukaryotic homologs are also supported by structural predictions and sequence alignments.

- The authors suggest that "replisome gene paralogs emerged primarily via HGT across Asgard archaea". However, it is not clear to me whether the authors have rigorously tested this hypothesis in a statistical framework (I did not see any indication in the methods). To determine whether homologs emerged via HGTs as opposed to duplications or losses, it would be important to perform gene-tree / species-tree reconciliations that allow comparing single gene trees against a species tree and estimating gene family history including not only transfers but also duplications, originations, and losses. For example, the assumptions regarding duplication versus HGT as illustrated in Fig. 4D are too simplistic. E.g. a gene may have duplicated on a much earlier node followed by subsequent differential losses across various Asgard lineages and transfer to other Archaea and Bacteria. Without a statistical framework that also takes into account tree uncertainty, it is difficult to determine where Asgard archaeal homologs are coming from and I am therefore not convinced that the hypothesis of Asgard replication proteins being shaped predominantly by HGT is justified.
We thank you for pointing out the requirement for further clarification and the need to discuss HGT vs duplication and loss more rigorously.

We have now modified the title into 'Asgard archaeal replisome gene paralogs have diverse origins'.

Thank you for pointing out the lable of 'HGT' in Fig. 4D have mixed assumption immediately with the representation of tree topology. We totally agree that tree topology should be described first before evolutionary scenarios are inferred. We had previously used 'paraphyletic' and 'recombination', as opposed to 'monophyletic' and 'duplication' in our first version, but was suggested by one of the reviewers to rephrase. We now modified 'duplication' and 'HGT' into 'shared root' and 'different roots'. And then in the context of 'different roots' group, we examine the possibility of 'HGT' and 'ancient duplication followed by differential loss'.

In terms of the methodology, in your previous review, you have suggested us to use AU testing for such a purpose, which we adopted to carry out a thorough analysis. We feel that a detailed examination via AU testing should be sufficient to address these concerns. While we understand the rationale for your new suggestion of using reconciliation approaches, we refrain from adopting for the following reasons:

1. Reconciliation methods are far less broadly practiced as AU testing, and a recent systematic comparison by Weiner et al (https://www.biorxiv.org/content/10.1101/2024.11.20.624597v5.full) suggested that each existing approach can yield artefacts due to different reasons.

2. Our phylogenetic trees were based on a suitable number of evenly sample protein sequence diversity combined with C60-based mixture model, and sometimes include bacteria; these selections do not correspond to the species selection for the species tree others and we construct. Hence, application of the readily available reconciliation packages is not feasible.

In the text, we now acknowledge that based on the trees, Fen1 could have undergone early acquisition and later loss, and we illustrated this now in Fig. 6. We also mention that HGT cannot be completely excluded. However, duplication was still not supported by AU testing.

The relevant changes in the main text are as follows (Line 381-390):

"Given the shared phylogenetic nodes for most protein paralogs in the eukaryotes, we postulate that they arose through duplication in the LECA; we cannot exclude ancient HGT from closely related lineages of LECA. On the other hand, given the distribution of polyphyletic multi-copy genes only in specific Asgard archaea lineages, including Fen1, Top6A, Top6B, RfcS, RnhB, Ligase, and Top3, we postulate that they arose through HGT prior to the diversification of these lineages (see Supplementary text for gene-specific descriptions). It is, however, also difficult to completely exclude a scenario where they arose through an even earlier duplication/acquisition followed by a loss in most lineages, especially for Fen1 (see Supplementary text)."

- In line with this, for the scenario presented in Fig 6, the authors would need to perform reconciliations and ancestral reconstructions to for example be able to infer the replisome set present in Lokis/Bald/Hel/Hods/FECA and in subsequent ancestors and infer later losses versus gains. This would allow to statistically infer whether subunits originated via HGT or were inherited from a common ancestor with subsequent duplication and differential loss in various Asgard lineages.

We have added Fig. 6 upon your previous request as a summary graph, which only included the limited elements that showed clear structural changes and/or compositional changes. We feel that the graph satisfies the 'phylogenetics combined with AU testing' as requested previously. The methological difficulty and uncertainty of using readily available reconciliation packages is mentioned above. We clearly stated in the caption that it is a 'proposed evolutionary trajectory', we added more detailed explanations in the SI text for each gene involved in the graph. We hope that these additions can provide a more balanced view for the readers. We do note that we used 'acquisition/duplication' instead of just 'duplication' because duplication scenario was rejected by AU testing for almost all hypothetical cases involving differential loss highlighted in this graph.

For the case of Fen1, we now wrote (Line 503-506):

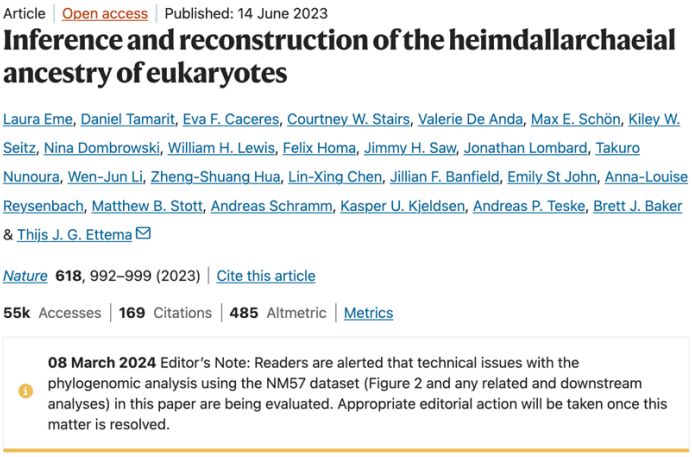
'Fen1-2 variant may have been acquired by the Asgard archaeal ancestor in additional to the canonical Fen1-1 and ultimately replacing the role of latter in the FECA (as depicted in Fig. 6), or a result of HGT from the Lokiarchaeia-Thorarchaeia common ancestor to FECA.'

- The placement of the eukaryotic branch relative to Asgard archaea is currently unresolved. The authors infer one species tree based on just six homologs which are unlikely to present a final picture of the eukaryotic branch. Therefore, and due to the limitation in interpreting the single gene trees without statistical framework, the scenario in Fig. 6 is speculative. In light of current literature and the conflicting results as to whether Heimdalls as a whole or specific sub-clades of Heimdalls are sister to the branch leading to eukaryotes, I would recommend doing reconciliations with both topologies and test which one is most supported in light of replisome gene repertoires.

Indeed, in our previous manuscript, we opted not to discuss the debate you mentioned in exquisite detail. The phylogenetic topology that our illustration is based on is consistent with most of the recent studies. Since you have not pointed to specific references, I can guess that you meant the Em et al Nature 2023 paper that placed Hodarchaeales next to the eukaryotes. The reason that we did not discuss such a topology was because it was a single study, which our single-gene evolution cannot be reconciled with. We forced this topology on our concatenated replisome tree, as well as our 6 single-copy gene trees. The concatenated tree test rejected the Hod origin, and among the 6 single gene trees, only RfcL did not reject Hod. Additionally, we did not observed any conserved eukaryotic replisome genes to branch from the Hodarchaeales in any other single replisome gene trees, except for the partial PolB; PolB exist in high copy numbers and is known to be subject to HGT. Combined, these

observations made it difficult to reconcile the evolution of single replisome genes in the context of this above topology with respect to the paths and timing of both vertical and horizontal transmissions. We opted to refrain from discussing this paper because we did not want to shift the focus of the paper towards a negative discussion on one specific study.

Secondly, we would like to note that the Eme et al paper was followed by the following note below on Mar 8, 2024, where their NM57 dataset and downstream phylogenomic analyses have to be re-evaluated. We also note that this study did not use the best quality genomes that was available; NM57 contained RadA homologs 3 times and included RfcS and Fen1, which we show to have undergone complex gain and loss around eukaryotic origin. We did not think it would be helpful to negatively discuss upon a single unsettled scenario.



Article | Open access | Published: 14 June 2023

Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes

Laura Eme, Daniel Tamarit, Eva F. Caceres, Courtney W. Stairs, Valerie De Anda, Max E. Schön, Kiley W. Seitz, Nina Dombrowski, William H. Lewis, Felix Homa, Jimmy H. Saw, Jonathan Lombard, Takuro Nunoura, Wen-Jun Li, Zheng-Shuang Hua, Lin-Xing Chen, Jillian F. Banfield, Emily St John, Anna-Louise Reysenbach, Matthew B. Stott, Andreas Schramm, Kasper U. Kjeldsen, Andreas P. Teske, Brett J. Baker & Thijs J. G. Ettema

Nature **618**, 992–999 (2023) | Cite this article

55k Accesses | **169** Citations | **485** Altmetric | Metrics

08 March 2024 Editor's Note: Readers are alerted that technical issues with the phylogenomic analysis using the NM57 dataset (Figure 2 and any related and downstream analyses) in this paper are being evaluated. Appropriate editorial action will be taken once this matter is resolved.

We note that while increasing the number of markers can potential increase phylogenomic resolution, it is important to examine the markers carefully and remove those that are apparently influenced by HGT, or those that underwent early duplication and sporadic differential loss. Our use of 6 genes, involved in a central genetic process, have undergone careful examination to best avoid HGT influence. While it is shorter than the NM57 markers, which as we mentioned above to have included artefactual markers, its matrix length is 2280 sites, comparible to previously used ribosomal markers (e.g. 2295 sites from 16 ribosomal genes by Seitz ISME J 2016 and 2684 sites from 15 ribosomal genes by Zaremba-Niedzwiedzka 2017 Nature). Furthermore, our illustration uses a topology that is in line of most other recent studies (except for Em et al 2023) and our own phylogenomic analyses based on highest quality genomes and a larger marker set (Wu et al 2022).

While we inclined not to discuss the Hodarchaeales scenario, to address the issue you raised, we now added:

"Whereas a recent study inferred a preferrable placement of eukaryotes sister to the Hodarchaeales within the Heimdall clan with a mixed set of markers (ref 21), forcing such a topology on phylogenetic trees of the six single-copy replisome genes or their concatenation receives no statistical support (Supplementary Table 5); it is thus difficult to reconcile replisome evolution with such a scenario in this study."

We do respect the efforts that the Eme et al paper has made in trying to improve the resolution at the eukaryotic origin, and we hope that this above comment in the main text would be sufficient to excuse us from a more aggressive discussions in the main text. We added some additional discussions to the Supplementary text section 'Replisome tree topology'.

Minor

- Line 84: I would delete environmental here: "environmental archaea" (aren't all organisms environmental? And if used in opposition to cultivated, some Asgards are now in culture…) .

Microbial ecologists often distinguish 'environmental' microbes from 'host-associated' microbes, hence the use of 'environmental'. Indeed, it can be confusing. We now modified 'environmental archaea' to 'archaeal lineages'.

- Fig 1c: the absence of circles does not mean absence of homologs, right? If so, this might be a bit confusing. Could you indicate presence of homologs in grey and putative siser-relationship with eukarytoes in a nother color? To clearly differentiate Asgards that do not have a homolog at all versus lacking a homolog closely related to eukaryotes.

We have added in the caption ‘Other less closely related homologs are not indicated.’

- can you refer to the underlying trees for the hypothesis presented in Fig 4e?

We have added in the caption: ‘The detailed phylogenetic trees are shown in Extended Data Figs. S3-S7.’

Reviewer #3:

Remarks to the Author:

I have read with interest the response letter by the authors to the recent round of revisions on their manuscript “Serial innovations by Asgard archaea shape the DNAx replication machinery of early eukaryotic ancestor”. I have also looked through the additions/changes to the new version of the manuscript.

I appreciate the authors efforts to address my concerns: I agree that doing gene-tree/species-tree reconciliations is time consuming and would indeed require to rerun single gene trees using a taxon set that corresponds to the taxa in the species tree. Considering that the authors have already put substantial effort and work in this, reconciliations may be something to implement in the future. I therefore understand that the authors prefer to focus on addressing my concerns using AU testing (even if not a substitute for gene tree/species tree reconciliations) and textual clarifications. I think that the topology testing that the authors implemented in this round was helpful and allowed clarifying the support for some of the key placements. However, while I am fully appreciative of the work, I am not yet convinced by the clarity of the textual presentation of some of the results and their discussion.

Thank you for your thorough review process that substantially improved the clarity and rigor of the paper.

Unfortunately, I am currently not able to provide further detailed feedback on this submission, but point out a number of specific points regarding the recent changes below:

- The new additions need to be carefully checked for grammar and readability, in my view they do not always lead to improved clarity yet.

Thank you for your notes. We have revised the manuscript to improve the word choice and grammar.

- Abstract, please rephrase “By reconciling events of gene gain and loss with an archaea-eukaryotic tree based on curated replisome markers” to something like: “By comparing curated replisome markers phylogenies with a archaea-eukaryotic tree” in order to avoid confusion. The phrasing which the authors chose here, is generally used when gene-tree species-tree reconciliation approaches have been applied. Since you did not perform such analyses, this could be a bit misleading.

Revised as suggested. Now it reads:

‘Placing the captured events of gene gain and loss in the context of archaea-eukaryote evolution inferred by the phylogeny of concatenated single-copy replisome genes, we propose a hypothetical model for the emergence of the complex eukaryotic replisome.’

- line 177: I would remove “yet” (since it could be the CTD was lost later)

Corrected.

- line 198: It would be great if you can specify what you mean by “less related homologs”. Can you describe any parameters/requirements that underlie the decision to call something “less related” for reproducibility?

We initially used the term “less closely related homologs” to describe replisome components that are not structurally or phylogenetically closely related to eukaryotes (such as Cdc6, MCM, etc.).

However, we recognized that this phrasing could be misleading. The sentence has therefore been removed.

- line 362-364: this sentence is a bit unclear to me, not sure its needed or adds anything.
Thank you for the note. The line was mere an introductory sentence for the detailed descriptions that followed. We intend to keep this line so that the following lines do not appear abrupt.

- line 365: is this correct?: “cannot not be well resolved”
Thank you for pointing out the typo. Now revised as “cannot be well resolved”.

- line 378-390: some sentences may need some rephrasing, e.g. “burst of single-node expansion…”
Revised to ‘burst of duplication’.

- line 420-423: rephrase for clarity?
Rephrased. Now it reads: ‘Besides a vertical transmission as depicted in Fig. 5a, a horizontally acquired gene copy also could become the sole copy that was retained in the LECA’

- line 513: “a complex evolutionary trajectory” instead of “divese evolutionary trajectories” (I think only one trajectory led to eukaryogenesis which included many events?)
Revised as suggested.