HAL open science

# Asgard archaea do not close the debate about the universal tree of life topology

Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre

## To cite this version:

Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre. Asgard archaea do not close the debate about the universal tree of life topology. PLoS Genetics, 2018, 14 (3), pp.e1007215. 10.1371/journal.pgen.1007215 . hal-02177609

**HAL Id: hal-02177609**

**https://hal.science/hal-02177609v1**

Submitted on 15 Oct 2019

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

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

Distributed under a Creative Commons Attribution 4.0 International License

PLOS | GENETICS

FORMAL COMMENT

# Asgard archaea do not close the debate about the universal tree of life topology

**Violette Da Cunha[1,2‡], Morgan Gaia[1‡], Arshan Nasir[3], Patrick Forterre[1,2]***

**1** Institut Pasteur, Unité de Biologie Moléculaire du Gène chez les Extrêmophiles (BMGE), Département de Microbiologie, Paris, France, **2** Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette cedex, France, **3** Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan

‡ These authors share first authorship on this work.
* forterre@pasteur.fr, patrick.forterre@i2bc.paris-saclay.fr

Check for updates

OPEN ACCESS

**Citation:** Da Cunha V, Gaia M, Nasir A, Forterre P (2018) Asgard archaea do not close the debate about the universal tree of life topology. PLoS Genet 14(3): e1007215. https://doi.org/10.1371/journal.pgen.1007215

**Editor:** Antonis Rokas, Vanderbilt University, UNITED STATES

**Received:** November 8, 2017

**Accepted:** January 22, 2018

**Published:** March 29, 2018

**Copyright:** © 2018 Da Cunha et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** VDC, MG, and PF are supported by a European Research Council (ERC) grant from the European Union's Seventh Framework Program (FP/2007-2013)/Project EVOMOBIL-ERC Grant Agreement no. 340440. AN is supported by the Higher Education Commission, Pakistan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

In Da Cunha et al. [1], we provided evidence that the results by Spang et al. [2] positioning the Lokiarchaea as the link between prokaryotes and eukaryotes were deeply influenced by the inclusion of likely contaminated sequences, the presence of fast-evolving species (FES), and the choice of phylogenetic markers. Our analyses revealed the presence of two phylogenetic signals within the concatenated universal markers: one supporting a two domains (2D) scenario, and the other supporting a three domains (3D) scenario. Our extensive RNA polymerase large subunits phylogenetic analyses strongly supported the latter [1]. Far from disregarding the other Asgards [3,4], Thorarchaea sequences were extensively used for reanalysing the universal markers, and most Asgard sequences were included in our RNA polymerase analyses. Altogether, our results suggested that the Asgards are not the ancestors of Eukarya but a sister group to Euryarchaeota.

## There is evidence of contamination in the lokiarchaeal genomes

On top of the very high heterogeneity detected in Loki's genome (78.21%), our quality analyses with CheckM (Parks, Skennerton, Imelfort, http://ecogenomics.github.io/CheckM/)) and Anvi'o (Meren, http://merenlab.org/software/anvio/) also revealed a high contamination index (between 45% and 57%), necessarily underestimated because quality analyses are limited to defined sets of markers [5]. Considering a good part of this contamination index as the result of the heterogeneity still leaves room for actual contamination from other sources. There is no need for important contamination to bias phylogenetic reconstruction if it is located in strong markers, as we have shown with the elongation factor 2 (EF2) [1].

We could identify such a likely contamination in EF2 by detecting long insertions matching to eukaryotic paralogs, essentially in Heimdall LC3 (formerly Loki 3) [1]. No other Asgard sequences [3] contain the LC3-specific insertions, reinforcing the contamination hypothesis. In an EF2 single-protein tree without bacteria to increase the signal, Heimdallarchaeota are not monophyletic, with LC3 still branching with eukaryotes, whereas all the other Asgards are sister groups to Euryarchaeota (Fig 1), suggesting that patches of contaminating sequences indeed remained in Heimdall LC3 after trimming. The complex evolutionary history of this protein could perhaps explain the accumulation of artificial insertions in the LC3 genome. We suspected that additional hidden patches of contamination could similarly be present elsewhere in other Asgard universal proteins, especially in Heimdall LC2 and LC3 reconstructed through a Multiple Displacement Amplification (MDA) process. This could explain why not only Asgards belonging to different phyla (Lokiarchaea, and Heimdallarchaea [represented by

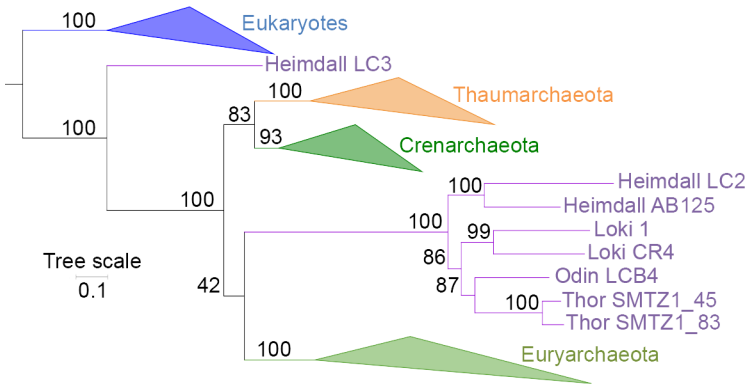


**Fig 1. Maximum-likelihood tree of EF2 without bacteria.** Maximum-likelihood single-protein tree of EF2 without bacteria (LG+R6 model) after inclusion of Asgard sequences. Eukaryotes, Thaumarchaeota, Crenarchaeota, and Euryarchaeota are indicated in blue, orange, dark green, and light green, respectively. Asgard sequences are indicated in purple. Odin and Thorarchaea seem to possess probable paralogs retrievable with BLASTp searches that were not included because they are different in length than the other Asgard sequences and contain specific indels. The scale bar represents the average number of substitutions per site. Values at nodes represent support calculated by nonparametric bootstrap (out of 100).

https://doi.org/10.1371/journal.pgen.1007215.g001

the formerly named Loki2 and Loki3]) but also to the same phylum (Heimdallarchaeota) were not always monophyletic in our single-protein analyses [1].

We suspected that patches of contamination could also be present in Heimdall LC3 RpoA since this subunit is encoded by a single gene in LC3 (like in Thaumarchaeota and the related Bathyarchaea and Aigarchaea), whereas all other Asgards have a dimeric version (like all Crenarchaeota, all Euryarchaeota, and most DPANN). Unlike Spang et al. [4], we never observed best hits to Lokiarchaea nor to the two other Heimdallarchaeotes using the complete LC3 sequence (best hits to Bathyarchaea only) or various portions of N- and C-terminal sequences (Loki3/Heimdall LC3 RpoA access numbers: AKC94880/OLS19521). In contrast, we recovered best hits to other Asgards when using Heimdallarchaeote LC2 or AB_125 sequences. The branching of LC3 at the base of Asgards in the RpoA tree [3] could indicate that LC3 RpoA contains Heimdallarchaeote sequences mixed with sequences from other archaea related to Bathyarchaea.

## The presence of FES in dataset favours 2D topologies

Spang et al. [4] argued that their 2D trees are not affected by Long-Branch Attraction (LBA), because LBA would have attracted eukaryotes outside the Archaea. However, the branching of Archaea in the well-known fast-evolving *Methanopyrus kandleri* in the 2D tree of Spang et al. [2] clearly reveals an LBA artefact. The Euryarchaea could then be possibly attracted outside Archaea because of their evolutionary relationship with DPANN [6–8]. This probably explains why removing FES dramatically increased (from 1 to 11) the number of trees recovering the 3D topology [1].

## There is no strong LBA affecting the position of Eukarya in the 3D trees

Spang et al. [4] argued that LBAs were affecting the 3D trees we obtained (i.e., with the concatenated 6 AU-relevant Woese and the 11 Woese proteins, and the RNA polymerase [1]), resulting in the misplacement of Eukaryotes between the bacterial outgroup and the Archaea. This seems unlikely for the RNA polymerase, since we have shown that Asgards remained a

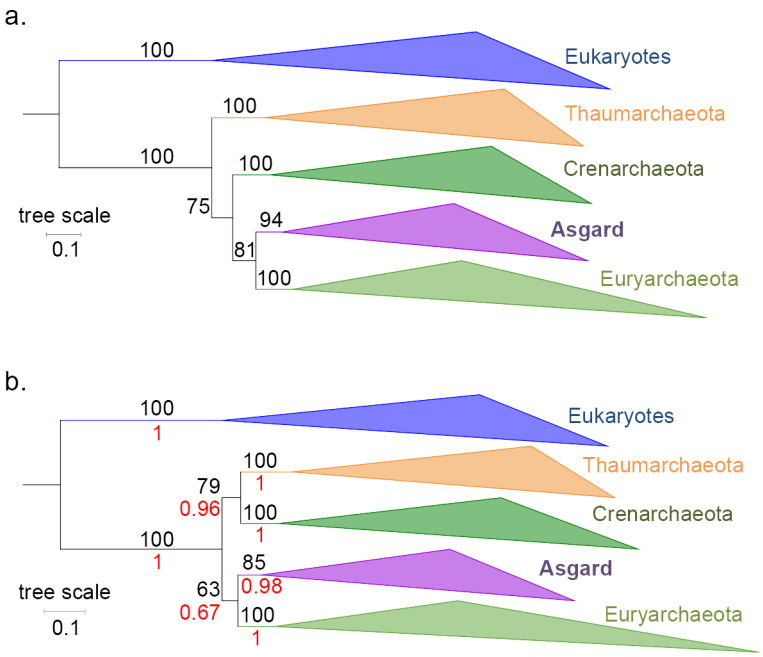


**Fig 2. Phylogenetic trees of the concatenations of the 11 Woese proteins and the 6 AU-relevant Woese proteins, without bacteria. (a)** Maximum-likelihood tree of the concatenated 11 Woese proteins with the Asgard sequences (LG +R7). No satisfactory convergence could be obtained in Bayesian inference with the CAT-GTR model at the time of this submission. **(b)** Phylogenetic tree of the concatenated 6 AU-relevant Woese proteins with the Asgard sequences. ML (LG+F+R8) and Bayesian inference (CAT-GTR model; maxdiff = 0.09) trees were identical, and supports from both were reported on the topology (nonparametric bootstrap, out of 100, in black, and posterior probabilities in red, respectively). All other combinations of chains (4 independent chains)—despite moderate convergences (maxdiff between 0.01 and 0.03)—yielded the same global topology, with the Asgards grouped with Euryarchaeota. For both trees, the scale bar represents the average number of substitutions per site.

https://doi.org/10.1371/journal.pgen.1007215.g002

sister group to Euryarchaeota (Archaea rooted in Thaumarchaea) in the absence of bacteria (Figure S35 in [1]). We have now performed similar analyses with the two concatenated Woese protein datasets after inclusion of all available Asgard sequences and obtained trees again displaying the grouping of Asgards with Euryarchaeota (Fig 2). These results strongly support the absence of strong LBA affecting the position of Eukarya in the 3D trees. If an LBA artefact was indeed misplacing the Eukaryotes, removing the bacteria would have dramatically impacted the trees. Our new analyses corroborate instead the position of the putative Asgard superphylum as a sister group to Euryarchaeota.

## Phylogenetic analyses excluding EF2 are compatible with the presence of conflicting signals within the universal markers

We never claimed that the Asgard/Eukarya affiliation could not be obtained without EF2. We ourselves obtained it when we removed EF2 from the concatenation of the eocyte proteins (Figures S26, S27 in [1]). Zaremba-Niedzwiedzka et al. indeed obtained significant support in a maximum-likelihood framework for a 2D universal proteins tree based on 48 universal markers without EF2 [3]. However, their species dataset still contained many FES that could introduce a bias favouring 2D trees [1]. Furthermore, they could not obtain any "good" (max-diff <0.1) nor "acceptable" (maxdiff <0.3) convergence in Bayesian framework with the CAT-GTR model (Supplementary Table 4 in [3]; see Phylobayes manual) and hence could not corroborate their results with this approach. Consequently, despite having more Asgard taxa,

Zaremba-Niedzwiedzka et al. [3] could not obtain results as congruent and robust as those obtained by Spang et al. [2] once they removed EF2. The lack of convergence in their Bayesian analysis is actually compatible with the presence of conflicting signals within the universal proteins, previously hidden by the presence of EF2 [1].

Finally, the confirmed presence of many Eukaryotic Signature Proteins (ESPs) in the genomes of the additional Asgards [3] cannot be seen as a confirmation of their grouping with Eukarya since their presence could also be explained by their losses in other archaeal lineages, as suggested for ESPs in Thaumarchaeota [9–11], or ancient gene transfers with proto-eukaryotes.

## Conclusions

Spang et al. [4] stated that we 1) used "inadequate methodology", without explaining, 2) "misinterpret[ed] data" while presenting some of our findings out of their context, and 3) "ignore [d] previous work", which our publication is actually based on. Results, criticism, and debates should be welcome in science and not trigger hostility. Although our analyses presently favour the 3D topology, we consider that the relationship between Archaea and Eukarya is still an open question requiring more studies.

## Author Contributions

**Conceptualization:** Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre.

**Formal analysis:** Violette Da Cunha, Morgan Gaia.

**Funding acquisition:** Patrick Forterre.

**Investigation:** Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre.

**Project administration:** Patrick Forterre.

**Supervision:** Patrick Forterre.

**Validation:** Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre.

**Visualization:** Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre.

**Writing – original draft:** Violette Da Cunha, Morgan Gaia, Arshan Nasir, Patrick Forterre.

## References


1. Da Cunha V, Gaia M, Gadelle D, Nasir A, Forterre P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. PLoS Genet. 2017; 13: e1006810. https://doi.org/10.1371/journal.pgen.1006810 PMID: 28604769
2. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. 2015; 521(7551): 173–9. https://doi.org/10.1038/nature14447 PMID: 25945739
3. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature. 2017; 541(7637): 353–8. https://doi.org/10.1038/nature21031 PMID: 28077874
4. Spang A, Eme L, Saw J, Caceres EF, Zaremba-Niedzwiedzka K, Lombard J et al. Asgard archaea are the closest prokaryotic relatives of eukaryotes. PLoS Genet. 2017
5. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25:1043–55. https://doi.org/10.1101/gr.186072.114 PMID: 25977477
6. Brochier C, Forterre P, Gribaldo S. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox. Genome Biol. 2004; 5(3):R17. https://doi.org/10.1186/gb-2004-5-3-r17 PMID: 15003120

7. Brochier C, Gribaldo S, Zivanovic Y, Confalonieri F, Forterre P. Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? Genome Biol. 2005; 6(5):R42. https://doi.org/10.1186/gb-2005-6-5-r42 PMID: 15892870

8. Brochier-Armanet C, Forterre P, Gribaldo S. Phylogeny and evolution of the Archaea: one hundred genomes later. Curr Opin Microbiol. 2011; 14(3):274–81. https://doi.org/10.1016/j.mib.2011.04.015 PMID: 21632276

9. Forterre P. The common ancestor of archaea and eukarya was not an archaeon. Archaea. 2013; 2013:372396. https://doi.org/10.1155/2013/372396 PMID: 24348094

10. Brochier-Armanet C, Gribaldo S, Forterre P. A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. Biol Direct. 2008; 3:54. https://doi.org/10.1186/1745-6150-3-54 PMID: 19105819

11. Spang A, Hatzenpichler R, Brochier-Armanet C, Rattei T, Tischler P, Spieck E, et al. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends Microbiol. 2010; 18(8):331–40. https://doi.org/10.1016/j.tim.2010.06.003 PMID: 20598889