Corresponding author(s): Anja Spang (anja.spang@nioz.nl)
Thijs Ettema (thijs.ettema@icm.uu.se)

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
☒	☐ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The genomes of the 10 Asgard representatives were downloaded from NCBI.
Data analysis	Databases used to analyze data and or against which data was compared: - nr (ftp://ftp.ncbi.nlm.nih.gov/) - UniProtKB (https://www.uniprot.org/) - HydDB (https://services.birc.au.dk/hyddb/) - Carbohydrate-Active enZYmes (CAZYmes) database (http://www.cazy.org/) - MEROPS peptidase database (https://www.ebi.ac.uk/merops/) - esterases were predicted using HMM-profiles from the ESTHER database (https://www.re3data.org/repository/r3d100010542) - transporter database (http://www.tcdb.org/) - Metacyc Metabolic Pathway Database (https://metacyc.org/) - KEGG (https://www.genome.jp/kegg/) - COGs and arCOGs (ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG) Data was analyzed using the following published softwares: - dbCAN webtool75 - blastdbcmd version 2.6.0+ - Interproscan-5.22-61.0 - HMMer vs. 3.1b2

- PSORT v3.0
- Mafft-LINSi v7.305b
- BMGE-1.12
- IQ-TREE v. 1.5.5
- Phylobayes MPI 1.7
- Psort v3.0
- R v3.3.0 (including plyr v 1.8.4 and ggplot2 v3.0.0)
- DIAMOND v0.9.9.110
- CD-HIT version 4.6
- FastTree version 2.1.9
- TrimAL v1.4
- Jalview 2.10.2b2
- Seaview n.d.
- FigTree v1.4.2

Data Visualization:
- Adobe Illustrator CC 2015 (19.2.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data and code vailability
The genomes of the herein analysed Asgard archaea have been made publicly available on NCBI previously. Detailed annotations of the metabolic repertoire are provided in Suppl. Tables 1-3 accompanying this manuscript. Raw data files and custom scripts are made available via figshare under the following link: https://figshare.com/s/5f153d1dcacadd3b3ed6.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We performed comparative genomics and phylogenomic analyses of published (publicly available) genomes of the Asgard archaea to infer their metabolic potential. Based on this analysis we update hypotheses on the origin of the eukaryotic cell.
Research sample	We analysed the genomes of Asgard archaea available at NCBI in January 2017, i.e. 2 published Lokiarchaeota (strains GC14_75 and CR4), 4 published Thorarchaeota (strains SMTZ-45, SMTZ1-83, SMTZ1-45 and AB25), 3 published Heimdallarchaeota (strains LC3, AB125 and LC2) and one published Odinarchaeote (strain LCB_4). Files of the metagenome assembled genomes of these Asgard archaea were retrieved from the ftp server of NCBI (ftp://ftp.ncbi.nih.gov/genomes/genbank/archaea/).
Sampling strategy	We did not obtain any new samples, as these genomes have already been reconstructed in previous and published work.
Data collection	The genome data described in the present manuscript was downloaded from the ftp server of NCBI and originally deposited along with following references: Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173-+, doi:10.1038/nature14447 (2015). Seitz, K. W., Lazar, C. S., Hinrichs, K. U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. The ISME Journal, doi:10.1038/ismej.2015.233 (2016). Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature 541, 353-358, doi:10.1038/nature21031 (2017).
Timing and spatial scale	we herein analyzed genomes of Asgard archaea deposited at NCBI in January 2017. Timing and spacial scale were not studied.

Data exclusions	no data was excluded from our analysis
Reproducibility	we have in detail described our methodological approaches for genome annotations as well as phylogenetic analyses to ensure reproducibility. We calculated branch support values for the different clusters obtained in our phylogenies that provide an indication as to how stable a given monophyletic group is. Considering this, our phylogenetic analyses are reproducible (when the same phylogenetic strategy is applied such as the same model of evolution etc). Genome annotations are based on inferences from various different databases. While this provides some level of confidence, annotations may have to be updated when new information is generated and deposited in databases.
Randomization	Since central tendencies and deviations between different treatments and controls have not been studied here, randomization and blinding are not applicable
Blinding	Since central tendencies and deviations between different treatments and controls have not been studied here, randomization and blinding are not applicable

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging